Coherence as an Organizing Principle in Statistical Signal Processing[1]

Louis Scharf

2014 IEEE Workshop on Statistical Signal Processing (SSP)

Gold Coast, Australia,
29 June, 2014-02 July, 2014

[1] National Science Foundation, CCF-1018472, Air Force Office of Scientific Research, FA 9550-10-1-0241 and FA 9550-10-C-0090.

The Pre-History

Coherence, and even multiple coherence, has a storied history in mathematics and statistics, where it may be attached to the concepts of correlation, principal angles between subspaces, canonical coordinates, and so on.

In electrical engineering, frequency selectivity in filters and wavenumber selectivity in antenna arrays is really a story of coherence between lagged or derivative components in the passband and interference between them in the stopband.

Coherence is perhaps better appreciated in physics, where it is used to describe phase alignment in time and/or space, as in *coherent* light. In fact, if you read Richard Feynmann's delightful book, QED, you might draw the conclusion that coherence describes a great many phenomena in classical and modern physics.

The Pre-History, contd

But in signal processing the history for coherence is relatively short. Perhaps it began in the $70's$ and $80's$.

The work that has most influenced our thinking is the work on generalized coherence by Cochran, et al. between 1987 and 1995[2], and the recent work of Ramirez, et al. and Klausner, et al.[3].

[2] H. Gish and D. Cochran, "Generalized coherence, in Proc. IEEE Int. Conf Acoust, Speech, Signal Process, vol 5, 1987, pp. 2745-2748; D. Cochran and H. Gish, "Multiple-channel detection using generalized coherence, in Proc. IEEE ICASSP, vol 5, 2883-2886, 1989; D. Cochran, H. Gish, D. Sinno, "A geometric approach to multiple-channel signal detection," IEEE Trans Signal Proc, vol 43, no 9, 2049-2057, Sept 1995.

[3] D. Ramirez, I. Santamaria, J. Via, and LS,"Detection of spatially-correlated Gaussian time series," vol 58, no 10, pp 5006-5015, 2010; N. Klausner, M. Azimi-Sadjadi, and LS, "Detection of spatially-correlated time series from a network of sensor arrays," IEEE Trans Signal Processing, vol 62, no 6, pp 1396-1407, 2014

Rough Outline

- A brief discussion of motivating problems.
- A few classical results from multivariate statistical analysis.
- A general discussion and defense of coherence as we wish to define it.
  - Establish contact with KL divergence, canonical coordinates, principal angles between subspaces.
  - Establish the Hadamard, Hilbert, and Euclidean geometries of coherence.
- Examples of coherence in detection, estimation, and time series analysis.
- Concluding remarks regarding the role of geometry and invariance in statistical signal processing.

Motivating Problems

We are motivated by space-time problems, natural or manufactured:

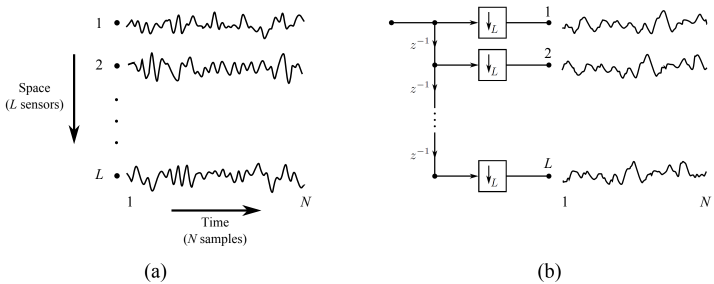


Figure 1: Vector valued time series, generated in (a) as a multisensor array of time series and in (b) as a polyphase representation of a scalar time series.

- sensor arrays for radar, sonar, acoustics, structural health monitoring, ad-hoc spectrum access, and geophysics,
- networks of PMUs in the smart grid,
- financial time series,
- analysis of scalar PC time series as vector-valued WSS time series, for communications and ad-hoc spectrum sensing.

Multivariate Statistical Analysis and Multiple Correlation Coefficient

Let's begin at the beginning, with the statistician's *multiple correlation coefficient*.

Consider the random variable $u \in \mathcal{C}$ and the random vector $\mathbf{v} \in \mathcal{C}^p$. The *composite covariance matrix* is

$$\mathbf{R} = \mathbb{E}\begin{bmatrix} u \\ \mathbf{v} \end{bmatrix}\begin{bmatrix} u^* & \mathbf{v}^H \end{bmatrix} = \begin{bmatrix} r_{uu} & \mathbf{r}_{uv}^H \\ \mathbf{r}_{uv} & \mathbf{R}_{vv} \end{bmatrix}$$

where $r_{uu} = \mathbb{E}uu^*$ is 1×1, $\mathbf{r}_{uv}^H = \mathbb{E}u\mathbf{v}^H$ is $1 \times p$, and $\mathbf{R}_{vv} = \mathbb{E}\mathbf{v}\mathbf{v}^H$ is $p \times p$. Each of these terms is a *Hilbert space inner product*.

The coherence we have in mind is

$$\rho^2(\mathbf{R}) = 1 - \frac{\det[\mathbf{R}]}{\det[r_{uu}]\det[\mathbf{R}_{vv}]} = \frac{\mathbf{r}_{uv}^H \mathbf{R}_{vv}^{-1} \mathbf{r}_{uv}}{r_{uu}}$$

Multivariate Statistical Analysis and Multiple Correlation Coefficient, contd

Multiple correlation coefficient, or coherence, is the cosine-squared of the angle between the 1-dimensional subspace spanned by the random variable u and the p-dimensional subspace spanned by the random variables $\mathbf{v} = (v_1, \cdots, v_p)$.

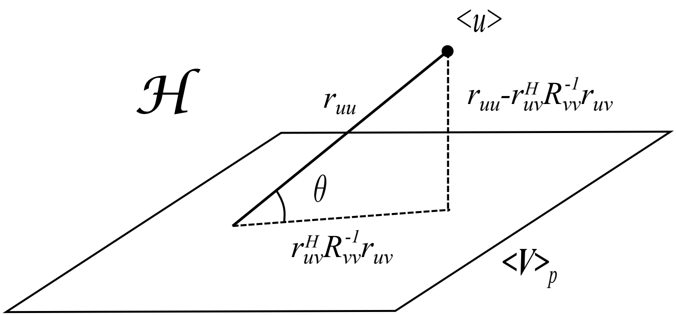


Figure 2: Coherence and Hilbert space.

Multivariate Statistical Analysis and Multiple Correlation Coefficient, contd

Now suppose in place of the random variables u and $\mathbf{v}$ we have only M iid realizations of them, organized into the *data matrix* $\mathbf{X}$ and *sample covariance matrix* $\mathbf{S}$:

$$\mathbf{X} = \begin{bmatrix} u[1] & v_1[1] & \cdots & v_p[1] \\ u[2] & v_1[2] & \cdots & v_p[2] \\ \vdots & \vdots & \vdots & \vdots \\ u[M] & v_1[M] & \cdots & v_p[M] \end{bmatrix} = \begin{bmatrix} \mathbf{u} & \mathbf{V} \end{bmatrix}$$

$$\mathbf{S} = \mathbf{X}^H\mathbf{X} = \begin{bmatrix} \mathbf{u}^H\mathbf{u} & \mathbf{u}^H\mathbf{V} \\ \mathbf{V}^H\mathbf{u} & \mathbf{V}^H\mathbf{V} \end{bmatrix} \text{ approxs } \begin{bmatrix} r_{uu} & \mathbf{r}_{uv}^H \\ \mathbf{r}_{uv} & \mathbf{R}_{vv} \end{bmatrix} = \mathbf{R}$$

Each of the *Euclidean inner products* in this sample covariance matrix is a sample estimator of its corresponding Hilbert space inner product in the composite covariance matrix.

Multivariate Statistical Analysis and Multiple Correlation Coefficient, contd

The *sample estimator* of the multiple correlation coefficient is

$$\mathbf{S} = \mathbf{X}^H\mathbf{X} = \begin{bmatrix} \mathbf{u}^H\mathbf{u} & \mathbf{u}^H\mathbf{V} \\ \mathbf{V}^H\mathbf{u} & \mathbf{V}^H\mathbf{V} \end{bmatrix}$$

$$\rho^2(\mathbf{S}) = 1 - \frac{\det[\mathbf{S}]}{\det[\mathbf{u}^H\mathbf{u}]\det[\mathbf{V}^H\mathbf{V}]} = \frac{\mathbf{u}^H\mathbf{P}_V\mathbf{u}}{\mathbf{u}^H\mathbf{u}},$$

where $\mathbf{P}_V$ is the projection onto the p-dimensional subspace $\langle\mathbf{V}\rangle$. This is the statistician's sample multiple correlation coefficient.

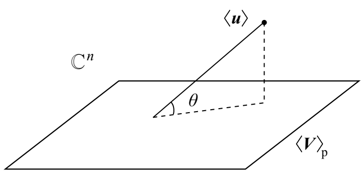
There is a connection:

$$\frac{\mathbf{r}_{uv}^H\mathbf{R}_{vv}^{-1}\mathbf{r}_{uv}}{r_{uu}} \longleftrightarrow \frac{\mathbf{u}^H\mathbf{P}_V\mathbf{u}}{\mathbf{u}^H\mathbf{u}}$$

Multivariate Statistical Analysis and Multiple Correlation Coefficient, contd

So, the sample estimator of Hilbert space coherence is Euclidean space coherence. The geometry and null distribution ($\mathbf{r}_{uv} = \mathbf{0}$) are these:

$$\text{Beta}[p, (M-p)] \sim \frac{\Gamma(M)}{\Gamma(p)\Gamma(M-p)} b^{p-1}(1-b)^{M-1}, 0 < b < 1.$$



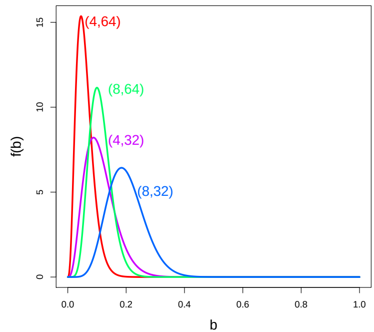


Figure 3: Geometry of Multiple Coherence and its Beta pdfs for various parameters, p and M.

Multivariate Statistical Analysis and Multiple Correlation Coefficient, contd

This result is often derived for the case where all random variables are jointly proper complex normal. But in fact, the result holds if

- $\mathbf{u} \in \mathcal{C}^n$ and $\mathbf{V} \in \mathcal{C}^{n \times p}$ are independently drawn, and
- one or both of $\mathbf{u}$ and $\mathbf{V}$ are invariantly distributed with respect to unitary transformations $\mathbf{Q} \in U(n)$
- Why? $\mathbf{u}^H \mathbf{P}_V \mathbf{u} / \mathbf{u}^H \mathbf{u} = \mathrm{tr}[\mathbf{P}_V \mathbf{P}_u]$.

For example, $\mathbf{u}$ may be white Gaussian and $\mathbf{V}$ fixed, or $\mathbf{u}$ may be white Gaussian and $\mathbf{V}$ may be independently drawn, or $\mathbf{u}$ may be fixed and $\mathbf{V}$ may be drawn as p iid white Gaussian vectors. But more generally, $\mathbf{u}$ may be spherical, and $\mathbf{V}$ may be spherically contoured.

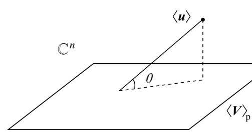


Figure 4: Invariantly-distributed subspaces.

Coherence

Hadamard Coherence. To analyze the more general space-time problems we began with, we need to generalize our notion of coherence. To this end, consider the positive definite Hermitian matrix

$$\mathbf{R} = \begin{bmatrix} \mathbf{R}_{11} & \mathbf{R}_{12} & \cdots & \mathbf{R}_{1L} \\ \mathbf{R}_{21} & \mathbf{R}_{22} & \cdots & \mathbf{R}_{2L} \\ \vdots & \vdots & \vdots & \vdots \\ \mathbf{R}_{L1} & \mathbf{R}_{L2} & \cdots & \mathbf{R}_{LL} \end{bmatrix}$$

This is a matrix of puzzle pieces, which is to say if $\mathbf{R}_{11}$ is $n_1 \times n_1$ and $\mathbf{R}_{22}$ is $n_2 \times n_2$, then $\mathbf{R}_{12}$ is $n_1 \times n_2$, and so on. The puzzle pieces fit.

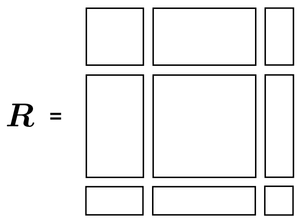


Figure 5: Puzzle pieces

Hadamard Coherence, contd. Each of the ideas of coherence that we will encounter and use may be written as a function of the *Hadamard coherence*

$$\rho^2(\mathbf{R}) = 1 - \frac{\det[\mathbf{R}]}{\det[\mathbf{R}_{11}]\det[\mathbf{R}_{22}]\cdots\det[\mathbf{R}_{LL}]}$$

$\boldsymbol{R}$ =

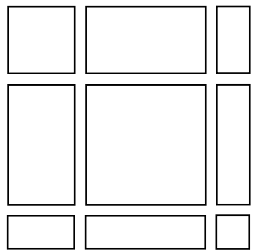


Figure 6: Puzzle pieces

This function is invariant to non-singular, block-diagonal, transformations $\mathbf{T}$, with action $\mathbf{TRT}^H$.

Coherence, contd

Kullback-Leibler. Suppose each of the $\mathbf{R}_{mn}$ in this block-structured matrix is the cross-correlation between two random vectors $\mathbf{x}_m$ and $\mathbf{x}_n$. In this case, the Kullback-Leibler distance between the distributions $P \sim CN[0, \mathbf{R}]$ and $Q \sim CN[0, \mathrm{diag}[\mathbf{R}_{11}, \cdots, \mathbf{R}_{LL}]$ is

$$D_{KL}(P||Q) = E_P \log \frac{P(X)}{Q(X)} = -\log \frac{\det \mathbf{R}}{\prod_{\ell=1}^{L} \det \mathbf{R}_{\ell\ell}}.$$

The connection between *coherence* as we have defined it and the Kullback-Leibler distance is then

$$\rho^2(\mathbf{R}) = 1 - \frac{\det[\mathbf{R}]}{\prod_{\ell=1}^{L} \det[\mathbf{R}_{\ell\ell}]} = 1 - e^{-D_{KL}(P||Q)}$$

Coherence, contd

Hilbert Coherence. Consider the Hilbert space of second-order random variables. Define the random vectors $\mathbf{u} = (u_1, \cdots, u_q)$ and $\mathbf{v} = (v_1, \cdots, v_p)$ and their covariance matrix

$$\mathbf{R} = \begin{bmatrix} \mathbf{R}_{uu} & \mathbf{R}_{uv}^H \\ \mathbf{R}_{uv} & \mathbf{R}_{vv} \end{bmatrix}$$

where $\mathbf{R}_{uu} = \mathbb{E}\mathbf{u}\mathbf{u}^H, \mathbf{R}_{uv} = \mathbb{E}\mathbf{u}\mathbf{v}^H, \mathbf{R}_{vv} = \mathbb{E}\mathbf{v}\mathbf{v}^H$ are respectively $q \times q$, $q \times p$, and $p \times p$. The Hadamard coherence between the q-dimensional subspace $\langle \mathbf{u} \rangle$ and the p-dimensional subspace $\langle \mathbf{v} \rangle$ is

$$\rho^2(\mathbf{R}) = 1 - \frac{\det[\mathbf{R}]}{\det[\mathbf{R}_{uu}]\det[\mathbf{R}_{vv}]} = 1 - \frac{\det[\mathbf{R}_{uu} - \mathbf{R}_{uv}^H \mathbf{R}_{vv}^{-1} \mathbf{R}_{uv}]}{\det[\mathbf{R}_{uu}]}$$

So coherence compares the volumes of the error concentration ellipses for $\mathbf{u}$, before and after filtering.

When $\langle \mathbf{u} \rangle$ is the one-dimensional subspace $\langle u \rangle$ spanned by the random variable u, then Hilbert coherence is the multiple correlation coefficient we began with.

Coherence, contd

Hilbert Coherence, contd. Rewrite *Hilbert coherence* as

$$\rho^2(\mathbf{R}) = 1 - \frac{\det[\mathbf{R}_{uu} - \mathbf{R}_{uv}\mathbf{R}_{vv}^{-1}\mathbf{R}_{vu}]}{\det[\mathbf{R}_{uu}]} = 1 - \prod_{i=1}^{min(p,q)} (1 - k_i^2)$$

The k_i are singular values of the *coherence matrix* $\mathbf{C} = \mathbf{R}_{uu}^{-1/2}\mathbf{R}_{uv}\mathbf{R}_{vv}^{-H/2}$, and they are called *canonical correlations* between the *canonical coordinates* of $\mathbf{u}$ and $\mathbf{v}$. That is, $\text{diag}[k_1, \cdots, k_{min(p,q)}] = \mathbf{F}^H\mathbf{C}\mathbf{G}$, and each of the k_i is the cross-correlation between a canonical coordinate pair, $\mu_i = \mathbf{f}_i^H\mathbf{R}_{uu}^{-1/2}\mathbf{u}$ and $\nu_i = \mathbf{g}_i^H\mathbf{R}_{vv}^{-1/2}\mathbf{v}$.

So, in order to talk about coherence in a Hilbert space, it is necessary to talk about the *canonical correlations* between *canonical coordinates*.The squared-canonical correlations k_i^2 are *fine-grained* coherences between canonical coordinates of the subspaces $\langle\mathbf{u}\rangle$ and $\langle\mathbf{v}\rangle$.

Coherence, contd

Euclidean Coherence. Begin with rectangular matrices $\mathbf{U} \in \mathcal{C}^{n \times q}$ and $\mathbf{V} \in \mathcal{C}^{n \times p}$, $(p, q) < n$, They span the respective subspaces $\langle \mathbf{U} \rangle$ and $\langle \mathbf{V} \rangle$. Construct the Grammian

$$\mathbf{R} = \begin{bmatrix} \mathbf{U}^H\mathbf{U} & \mathbf{U}^H\mathbf{V} \\ \mathbf{V}^H\mathbf{U} & \mathbf{V}^H\mathbf{V} \end{bmatrix}$$

Hadamard coherence is then *Euclidean coherence* between the q-dimensional subspace $\langle \mathbf{U} \rangle$ and the p-dimensional subspace $\langle \mathbf{V} \rangle$,

$$\rho^2(\mathbf{R}) = 1 - \frac{\det[\mathbf{U}^H(\mathbf{I} - \mathbf{P}_V)\mathbf{U}]}{\det[\mathbf{U}^H\mathbf{U}]} = 1 - \prod_{i=1}^{min(p,q)} (1 - \rho_i^2),$$

where $\mathbf{P}_V$ is the projection onto $\langle \mathbf{V} \rangle$. This is a *bulk* definition of coherence, based on *fine-grained* coherences ρ_i^2. These fine-graineds are coherences (in fact cosine-squareds of principal angles between subspaces). When $\langle \mathbf{U} \rangle$ is the one-dimensional subspace $\langle \mathbf{u} \rangle$, then Euclidean coherence is the sample multiple correlation coefficient we began with.

Coherence, contd

So *multiple coherence* seems to stand on firm ground:

- In the MVN case it is a monotone function of Kullback-Leibler distance between the covariance model $\mathbf{R}$ and the covariance model $\text{diag}[\mathbf{R}_{11} \cdots \mathbf{R}_{LL}]$.
- It compares volumes of error concentration ellipses, before and after filtering.
- In signal-plus-noise models, it is an increasing function of the volume of the signal-to-noise ratio matrix.
- It is a Schur convex function of principal angles and canonical correlations, which are maximal invariants to the group of non-singular, block-diagonal, linear transformations.
- For concrete examples, it has the Hilbert space or Euclidean geometries we expect to see.

The Cramér-Rao Bound

Suppose unknown parameters $\boldsymbol{\theta} \in \mathcal{R}^p$ modulate the mean of a multivariate normal measurement. The measurement model is $\mathbf{y} = \mathbf{x}(\boldsymbol{\theta}) + \mathbf{w} \sim N_n[\mathbf{x}(\theta), \sigma^2 \mathbf{I}_{n \times n}]$, $\mathbf{x} \in \mathcal{R}^n$ and $\boldsymbol{\theta} \in \mathcal{R}^p$, $p \leq n$. The Fisher matrix for estimating $\boldsymbol{\theta}$ from $\mathbf{y}$ is the Gramian

$$\sigma^2 \mathbf{J} = \mathbf{G}^T \mathbf{G} = \begin{bmatrix} \mathbf{g}_i^T \mathbf{g}_i & \mathbf{g}_i^T \mathbf{G}_i \\ \mathbf{G}_i^T \mathbf{g}_i & \mathbf{G}_i^T \mathbf{G}_i \end{bmatrix}$$

$$\mathbf{G} = [\mathbf{g}_1, \cdots, \mathbf{g}_p] : \text{an } n \times p \text{ matrix of sensitivities; } \mathbf{g}_i = \frac{\partial \mathbf{x}}{\partial \theta_i} \in \mathcal{R}^n :$$

Euclidean coherence is

$$\rho^2(\sigma^2 \mathbf{J}) = 1 - \frac{\det[\mathbf{J}]}{\det[\mathbf{g}_i^T \mathbf{g}_i] \det[\mathbf{G}_i^T \mathbf{G}_i]} = \frac{\mathbf{g}_i^T \mathbf{P}_{\mathbf{G}_i} \mathbf{g}_i}{\mathbf{g}_i^T \mathbf{g}_i}$$

The Cramér-Rao Bound, contd

The ii-th element of $\mathbf{J}^{-1}$ is the Cramér-Rao bound on MSE for an unbiased estimator of the i-th element of $\boldsymbol{\theta}$ from $\mathbf{y}$. Let's normalize this by the CRB in the case where all parameters but θ_i are known:

$$\frac{(\mathbf{J}^{-1})_{ii}}{1/J_{ii}} = \frac{1}{1-\rho^2(\mathbf{J})}; \qquad \rho^2(\mathbf{J}) = \frac{\mathbf{g}_i^T \mathbf{P}_{\mathbf{G}_i} \mathbf{g}_i}{\mathbf{g}_i^T \mathbf{g}_i}.$$

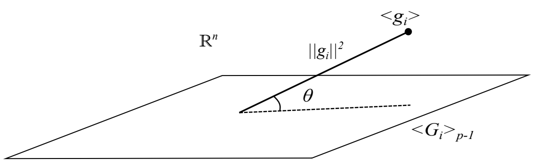


Figure 7: The Euclidean geometry of the CRB.

Coherence again. You do not want an instrument for which a sensitivity $\mathbf{g}_i$ is coherent with the the sensitivities $\mathbf{G}_i$.

The Cramér-Rao Bound, contd

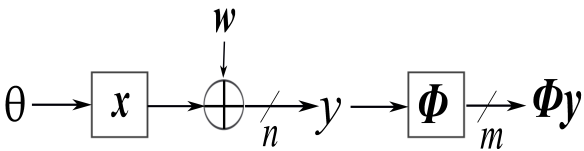


Figure 8: Compression and CRB.

What happens when $\mathbf{y} = \mathbf{x}(\theta) + \mathbf{w}$ is compressed as $\mathbf{\Phi y}$ with a random Gaussian $\mathbf{\Phi} \in \mathcal{C}^{m \times n}$, $\mathbf{\Phi}_{ij} : N[0, 1/m]$, or by sub-sampling in a co-prime pattern, say? Not much. The formulas and geometry remain unchanged, with $\mathbf{G}, \mathbf{g}_i, \mathbf{G}_i$ replaced by $\mathbf{P}_{\mathbf{\Phi}^T}\mathbf{G}$, etc, which are projections of sensitivities onto the subspace of compression vectors, $\langle \mathbf{\Phi}^T \rangle$. But of course performance is impacted, negatively.

CAUTION: You must have an excess of SNR, or low ambitions for resolution, or a lot of apriori information about your parameters to get away with compression (as for example in monopulse radar).

Matched Subspace Detectors

Let's consider three subspace detectors:

- MSD: matched subspace detector for detecting signals lying at unknown locations in a known subspace.
- MDD: matched direction detector for detecting signals known to lie at unknown locations on an unknown line in a known subspace.
- MLD: matched location detector for detecting signals lying at unknown locations on a known line.

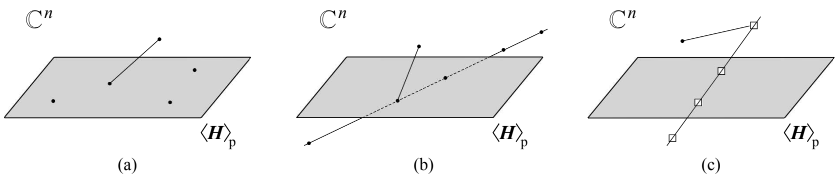


Figure 9: From left to right: subspace model for signals trapped in a subspace, for signals trapped on an unknown line in a subspace, and for signals trapped on a known line.

Matched Subspace Detector

A signal **Hx** lies in a subspace $\langle \mathbf{H} \rangle$, $\mathbf{H} \in \mathcal{C}^{n \times p}$, at an unknown location. Think of columns of **H** as the modes of the signal, and $\mathbf{x} \in \mathcal{C}^p$ as the mode weights; the measurement is $\mathbf{y} = \mathbf{Hx} + \mathbf{n}$.

The CFAR matched subspace detector for testing $H_0 : \mathbf{y} : CN_n[0, \sigma^2 \mathbf{R}]$ vs $H_1 : \mathbf{y} : CN_n[\mathbf{Hx}, \sigma^2 \mathbf{R}], \mathbf{x} \neq \mathbf{0}$, σ^2 unknown, measures the *Euclidean coherence* between the measurement **z** and the subspace $\langle \mathbf{G} \rangle$:

$$\Lambda = \frac{\mathbf{z}^H \mathbf{P}_G \mathbf{z}}{\mathbf{z}^H \mathbf{z}}, \qquad \mathbf{G} = \mathbf{R}^{-1/2}\mathbf{H}, \qquad \mathbf{z} = \mathbf{R}^{-1/2}\mathbf{y}$$

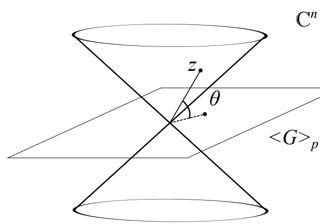


Figure 10: Geometry of the CFAR matched subspace detector.

Matched Subspace Detector, contd

This UMP-Invariant statistic is invariant to scaling of $\mathbf{z}$, and to rotations in $\langle \mathbf{G} \rangle$. Its null distribution is $B[p,(n-p)]$.

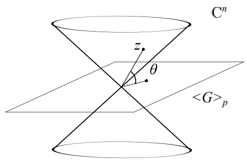


Figure 11: Geometry of the CFAR matched subspace detector.

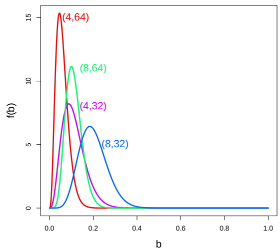


Figure 12: Null distribution of the CFAR MSD for various values of p, n, independent of R. $B[p,(n-p)]$

Matched Subspace Detectors, contd

Suppose the covariance matrix $\mathbf{R}$ is unknown. Then consider the two sample experiment $[\mathbf{X}, \mathbf{y}]$, where $\mathbf{X} = [\mathbf{x}[1], \mathbf{x}[2], \cdots, \mathbf{x}[M]]$ is a sequence of M realizations of $\mathbf{x} \sim CN_n[0, \gamma^2\mathbf{R}]$.

Then the CFAR adaptive subspace detector (also called ACE for Adaptive Coherence Estimator) is identical with the MSD, with the sample covariance matrix $\mathbf{X}\mathbf{X}^H$ replacing $\mathbf{R}$.

The distribution of this statistic is complicated, but may be written as a mixture of betas, independent of $\mathbf{R}$. The geometry remains unchanged from that of the MSD.

Matched Subspace Detector, contd

Could this result be obtained from a coherence argument? Begin with a measurement vector $\mathbf{z} \in \mathcal{C}^n$ and a linear space of dimension p, spanned by the columns of the matrix $\mathbf{G} \in \mathcal{C}^{n \times p}$. Construct the data matrix $\mathbf{Z}$ and its Grammian:

$$\mathbf{Z} = \begin{bmatrix} \mathbf{z} & \mathbf{G} \end{bmatrix}; \quad \mathbf{Z}^H\mathbf{Z} = \begin{bmatrix} \mathbf{z}^H\mathbf{z} & \mathbf{z}^H\mathbf{G} \\ \mathbf{G}^H\mathbf{z} & \mathbf{G}^H\mathbf{G} \end{bmatrix}$$

Coherence is then

$$\rho^2(\mathbf{Z}^H\mathbf{Z}) = 1 - \frac{\det[\mathbf{Z}^H\mathbf{Z}]}{\det[\mathbf{z}^H\mathbf{z}]\det[\mathbf{G}^H\mathbf{G}]} = \frac{\mathbf{z}^H\mathbf{P}_G\mathbf{z}}{\mathbf{z}^H\mathbf{z}}$$

This method and result extends to multiple measurements $\{\mathbf{z}[m], m = 1, \cdots, M\}$, by stacking the $\mathbf{z}[m]$ into the vector $\mathbf{z} \in \mathcal{C}^{Mn}$ and replacing the matrix $\mathbf{G}$ by $\mathbf{I}_M \otimes \mathbf{G} \in \mathcal{C}^{Mn \times Mp}$. The resulting coherence statistic is distributed as Beta$[Mp, M(n-p)]$ under the null hypothesis.

Detection of Spatially-Correlated Time Series

Let's generalize the original experiments of Cochran, Gish, and Sinno (1987, 1989, 1995); and Leshem and van der Veen (2001).

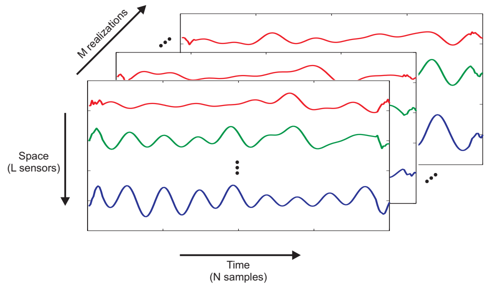


Figure 13: Array's worth of time series.

- L-eLement array; n-sample time series at each array element ℓ, denoted $\mathbf{x}_\ell \in \mathcal{C}^n$;
- These time series vectors may be organized into a long space-time vector, by stacking them top-to-bottom: $\mathbf{x} = (\mathbf{x}_l^T, \cdots, \mathbf{x}_L^T)^T \in \mathcal{C}^{Ln}$.

Detection of Spatially-Correlated Time Series, contd

The covariance matrix of this space-time vector $\mathbf{x}$ is $\mathbf{R} = \mathbb{E}\mathbf{x}\mathbf{x}^H = \{\mathbf{R}_{\ell m}\} \in \mathcal{C}^{Ln \times Ln}$. Each of the $n \times n$ blocks $\mathbf{R}_{\ell m} = \mathbb{E}\mathbf{x}_\ell \mathbf{x}_m^H \in \mathcal{C}^{n \times n}$ is the cross correlation between time series at sensors ℓ, m.

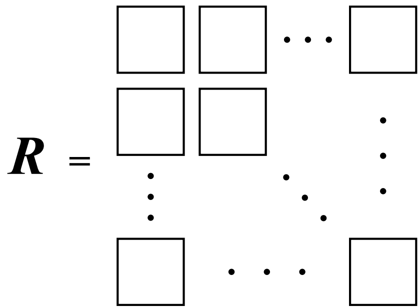


Figure 14: Covariance matrix of space-time vector consists of $n \times n$ blocks, each the auto- or cross- covariance matrix between the n-sample time series at two difference sensors. There are L^2 blocks for an L-eLement array.

Detection of Spatially-Correlated Time Series, contd

Hadamard coherence for this problem is

$$\rho^2(\mathbf{R}) = 1 - \frac{\det[\mathbf{R}]}{\prod_{\ell=1}^{L} \det[\mathbf{R}_{\ell\ell}]}$$

where the $\mathbf{R}_{\ell\ell} \in \mathcal{C}^{n\times n}$ are the matrix blocks of $\mathbf{R}$ lying on the diagonal. When the time series at all L sensors are jointly WSS, bandlimited to Ω, then in the limit as the length n of each time series increases without bound, then this Hadamard coherence converges as a l.i.m. to

$$\rho^2(\mathbf{R}) = 1 - \frac{\det[\mathbf{R}]}{\prod_{\ell=1}^{L} \det[\mathbf{R}_{\ell\ell}]} \longrightarrow 1 - \exp\{\int_{\Omega} \log \frac{\det[\mathbf{K}(e^{j\theta})]}{\prod_{\ell=1}^{L} K_{\ell\ell}(e^{j\theta})} \frac{d\theta}{2\pi}\}$$

where $\mathbf{K}(e^{j\theta}) = \{K_{\ell m}(e^{j\theta})\} \in \mathcal{C}^{L\times L}$ is *multi-channel spectral correlation* and $K_{\ell m}(e^{j\theta})$ is the discrete-time Fourier transform of the cross correlation sequence $\{r_{\ell m}[n]\}$ between the times series at sensors ℓ, m, at lag n. The RHS we call *broadband coherence*.

Detection of Spatially-Correlated Time Series, contd

We would now like to give ourselves multiple experimental copies of the space-time field, and its sample correlation matrix, namely

- $\mathbf{X} = [\mathbf{x}[1], \cdots \mathbf{x}[M]]$, M iid realizations of $\mathbf{x} \in \mathcal{C}^{Ln}$,
- $\mathbf{S} = \mathbf{X}\mathbf{X}^H$, sample space-time correlation matrix, L blocks of $n \times n$.

and test $H_0 : \mathbf{x} : CN_L[\mathbf{0}, \mathbf{R} = \text{diag}[\mathbf{R}_{\ell\ell}]]$ vs $H_1 : \mathbf{x} : CN_L[\mathbf{0}, \mathbf{R} > \mathbf{0}]$. The GLRT is

$$\Lambda = \frac{\det[\mathbf{S}]}{\prod_{\ell=1}^{L} \det[\mathbf{S}_{\ell\ell}]} = 1 - \rho^2(\mathbf{S})$$

This statistic is invariant to channel-by-channel linear transformations, making it robust to linear channel impairments. Moreover, there is no reason the time series in this problem cannot be replaced with Doppler-shifted and time-delayed versions, in which case coherence is a multichannel ambiguity statistic, $\Lambda(\tau, \nu)$, suitable for source localization in passive and active radar, sonar, and acoustics, in multi-sensor and multi-platform arrays.

Detection of Spatially-Correlated Time Series, contd

The null distribution of this statistic is

$$\Lambda = \frac{\det[\mathbf{S}]}{\prod_{\ell=1}^{L} \det[\mathbf{S}_{\ell\ell}]} \overset{d}{=} \prod_{\ell=2}^{L} \prod_{k=1}^{n} B[M-(\ell-1)n-k,(\ell-1)n]$$

$$-\log \Lambda \longrightarrow \chi^2_\nu, \;\; \nu = \dim R - \dim R_0 = L^2 n^2 - L n^2,$$

independently of the underlying **R**, provided it is block diagonal.

Moreover this distribution result holds for arbitrary elliptically-distributed random vectors at each sensor. The distribution at any one sensor may be chosen arbitrarily.

This histogram of randomly generated coherence $\rho^2(\mathbf{S})$, and randomly generated beta products is convincing numerical evidence for a mathematical result (as if one were needed).

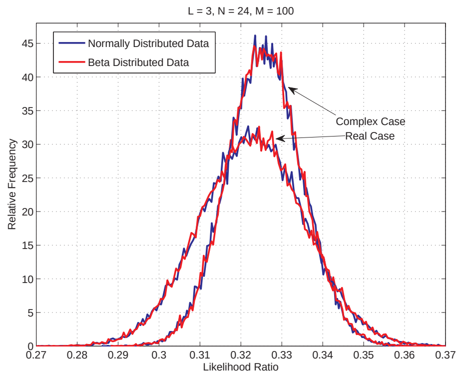


Figure 15: Histograms for test statistic Λ and for product of betas; $L = 3$, $n = 24$, $M = 100$. Independent of R.

Detection of Spatially-Correlated Time Series, contd

There are many variations on this experiment. For example: is the time series at sensor 1 linearly independent of the time series at sensors 2 through L, without regard for whether these time series are linearly independent. This kind of question arises in many contexts, including the construction of links in graphs of measurement nodes.

The GLRT, or geometrically inspired, test of such linear independence is the statistic

$$\Lambda = \frac{\det[\mathbf{S}]}{\det[\mathbf{S}_{11}]\det[\mathbf{S}^{11}]} \stackrel{d}{=} \prod_{k=1}^{n} B[M-(L-1)n-k,(L-1)n]$$

where $\mathbf{S}_{11}$ is the $n \times n$ NW block of the sample covariance matrix, containing second-order information about channel 1, and $\mathbf{S}^{11}$ is the $(L-1)n \times (L-1)n$ SE block of the sample covariance matrix, containing second-order information about channels 2 through L.

Detection of Spatially-Correlated Time Series, contd

Could these statistic have been obtained from geometrical reasoning, alone. Of course, provided we adopt multiple coherence as our definition of linear independence, based on a generalization of the Hadamard ratio. Are there other examples? Yes:

- Beamforming in Generalized Sidelobe Cancellers, giving a geometric comparison of Capon and Conventional Beamformers.
- Greedy filtering, as in the multistage or conjugate-direction Wiener filter.
- Time-frequency analysis as a story in coherence, resolving the discomfort over non-negativity.
- SNR loss in adaptive filtering.
- Compression of noisy measurements for transmission over a noisy channel, under a transmit power constraint.

Conclusions

This raises the question of a more *geometric* approach to signal processing, wherein functions of the data are derived from geometrical reasoning alone. To some extent this is the idea behind invariance arguments, which are often geometrical. But even these arguments are usually preceded by distribution assumptions and statistical reasoning, rather than by geometrical reasoning.

Such a methodology would contrast with our current methodology, which is to start with distribution assumptions for the data, followed with statistical reasoning (which sometimes includes invariance constraints), and as an afterthought ask about the geometry.

Perhaps this would widen the scope of statistical signal processing to machine learning and signal analytics, where there is often no theoretical or experimental basis for statistical assumptions.

Conclusions, contd

If you are a frequentist, or you believe in arguments of sufficiency and invariance, you might be happy. But what if you are a Bayesian?

- Some of my best friends are Bayesians.
- Maybe I am a closet Bayesian. You never know. (Actually, I am agnostic.)

You will want to assign prior distributions to unknown parameters and marginalize out the resulting joint distribution for the marginal distribution of measurements under H_0 and H_1. You might want to choose your priors to produce statistics with invariances you like. You will need to get the null distribution for detection statistics, and you will need to persuade us that your composite experiment, where measurements are drawn from your joint distribution of unknown parameters and noise, is the composite experiment that Mother Nature actually runs when she generates measurements.

If you like kernel methods, you are thinking that second-order may be replaced by kernels, followed by geometrical reasoning. I see no reason why this cannot be done. To what effect? I do not know.