Coherence as an Organizing Principle in Statistical Signal Processing

Louis Scharf

2014 IEEE Workshop on Statistical Signal Processing (SSP)

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

\[^1\]

\[^1\]National Science Foundation, CCF-1018472, Air Force Office of Scientific Research, FA 9550-10-1-0241 and FA 9550-10-C-0090.
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.
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\textsuperscript{2}, and the recent work of Ramirez, et al. and Klausner, et al.\textsuperscript{3}.


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:

- 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.

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.
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

\[
R = \mathbb{E} \begin{bmatrix} u \\ \mathbf{v} \end{bmatrix} \begin{bmatrix} u^* & \mathbf{v}^H \end{bmatrix} = \begin{bmatrix} r_{uu} & r_{uv}^H \\ r_{uv} & R_{vv} \end{bmatrix}
\]

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

The coherence we have in mind is

\[
\rho^2(R) = 1 - \frac{\text{det}[R]}{\text{det}[r_{uu}] \text{det}[R_{vv}]} = \frac{r_{uv}^H R_{vv}^{-1} r_{uv}}{r_{uu}}
\]
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 $v = (v_1, \cdots, v_p)$.

Figure 2: Coherence and Hilbert space.
Now suppose in place of the random variables $u$ and $v$ we have only $M$ iid realizations of them, organized into the data matrix $X$ and sample covariance matrix $S$:

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

$$S = X^H X = \begin{bmatrix}u^H u & u^H V \\
V^H u & V^H V
\end{bmatrix} \approx \begin{bmatrix}r_{uu} & r_{uv}^H \\
r_{uv} & R_{vv}
\end{bmatrix} = R$$

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.
The sample estimator of the multiple correlation coefficient is
\[
S = X^H X = \begin{bmatrix} u^H u & u^H V \\ V^H u & V^H V \end{bmatrix}
\]
\[
\rho^2(S) = 1 - \frac{\det[S]}{\det[u^H u] \det[V^H V]} = \frac{u^H P_V u}{u^H u},
\]
where \(P_V\) is the projection onto the \(p\)-dimensional subspace \(\langle V \rangle\). This is the statistician’s sample multiple correlation coefficient.

There is a connection:
\[
\frac{r_{uv}^H R_{vv}^{-1} r_{uv}}{r_{uu}} \leftrightarrow \frac{u^H P_V u}{u^H u}
\]
So, the sample estimator of Hilbert space coherence is Euclidean space coherence. The geometry and null distribution ($r_{uv} = 0$) 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

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

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

\[ R = \begin{bmatrix} R_{11} & R_{12} & \cdots & R_{1L} \\ R_{21} & R_{22} & \cdots & R_{2L} \\ \vdots & \vdots & \ddots & \vdots \\ R_{L1} & R_{L2} & \cdots & R_{LL} \end{bmatrix} \]

This is a matrix of puzzle pieces, which is to say if $R_{11}$ is $n_1 \times n_1$ and $R_{22}$ is $n_2 \times n_2$, then $R_{12}$ 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

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

This function is invariant to non-singular, block-diagonal, transformations \( T \), with action \( TRT^H \).

**Figure 6:** Puzzle pieces
Kullback-Leibler. Suppose each of the $R_{mn}$ in this block-structured matrix is the cross-correlation between two random vectors $x_m$ and $x_n$. In this case, the Kullback-Leibler distance between the distributions $P \sim \mathcal{CN}[0, R]$ and $Q \sim \mathcal{CN}[0, \text{diag}[R_{11}, \ldots, R_{LL}]]$ is

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

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

$$\rho^2(R) = 1 - \frac{\det[R]}{\prod_{l=1}^{L} \det[R_{ll}]} = 1 - e^{-D_{KL}(P||Q)}.$$
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 $\mathbf{R}_{uu} = \mathbf{E}u\mathbf{u}^H$, $\mathbf{R}_{uv} = \mathbf{E}u\mathbf{v}^H$, $\mathbf{R}_{vv} = \mathbf{E}v\mathbf{v}^H$ 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.
Hilbert Coherence, contd. Rewrite Hilbert coherence as

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

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

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 \( \langle u \rangle \) and \( \langle v \rangle \).
Coherence, contd

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

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

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

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

where $P_V$ is the projection onto $\langle V \rangle$. This is a *bulk* definition of coherence, based on *fine-grained* coherences $\rho_i^2$. These fine-graineds are coherences (in fact cosine-squareds of principal angles between subspaces). When $\langle U \rangle$ is the one-dimensional subspace $\langle u \rangle$, then Euclidean coherence is the sample multiple correlation coefficient we began with.
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 $\theta \in \mathcal{R}^p$ modulate the mean of a multivariate normal measurement. The measurement model is $y = x(\theta) + w \sim \mathcal{N}_n[x(\theta), \sigma^2 I_{n \times n}]$, $x \in \mathcal{R}^n$ and $\theta \in \mathcal{R}^p$, $p \leq n$. The Fisher matrix for estimating $\theta$ from $y$ is the Gramian

$$
\sigma^2 J = G^T G = \begin{bmatrix}
    g_i^T g_i & g_i^T G_i \\
    G_i^T g_i & G_i^T G_i
\end{bmatrix}
$$

$G = [g_1, \ldots, g_p]$ : an $n \times p$ matrix of sensitivities; $g_i = \frac{\partial x}{\partial \theta_i} \in \mathcal{R}^n$:

Euclidean coherence is

$$
\rho^2(\sigma^2 J) = 1 - \frac{\text{det}[J]}{\text{det}[g_i^T g_i] \text{det}[G_i^T G_i]} = \frac{g_i^T P G_i g_i}{g_i^T g_i}
$$
The $ii$-th element of $J^{-1}$ is the Cramér-Rao bound on MSE for an unbiased estimator of the $i$-th element of $\theta$ from $y$. Let's normalize this by the CRB in the case where all parameters but $\theta_i$ are known:

$$
\frac{(J^{-1})_{ii}}{1/J_{ii}} = \frac{1}{1 - \rho^2(J)} \quad \rho^2(J) = \frac{g_i^T P G_i g_i}{g_i^T g_i}.
$$

![Figure 7: The Euclidean geometry of the CRB.](image)

Coherence again. You do not want an instrument for which a sensitivity $g_i$ is coherent with the the sensitivities $G_i$. 
What happens when $y = x(\theta) + w$ is compressed as $\Phi y$ with a random Gaussian $\Phi \in \mathbb{C}^{m \times n}$, $\Phi_{ij} : N[0, 1/m]$, or by sub-sampling in a co-prime pattern, say? Not much. The formulas and geometry remain unchanged, with $G, g_i, G_i$ replaced by $P_{\Phi^T} G$, etc, which are projections of sensitivities onto the subspace of compression vectors, $\langle \Phi^T \rangle$. 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.

![Diagram](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 $\langle H \rangle$, $H \in \mathbb{C}^{n \times p}$, at an unknown location. Think of columns of $H$ as the modes of the signal, and $x \in \mathbb{C}^p$ as the mode weights; the measurement is $y = Hx + n$.

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

$$\Lambda = \frac{z^H P_G z}{z^H z}, \quad G = R^{-1/2} H, \quad z = R^{-1/2} y$$

Figure 10: Geometry of the CFAR matched subspace detector.
This UMP-Invariant statistic is invariant to scaling of $z$, and to rotations in $\langle G \rangle$. 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 $R$. $B[p, (n - p)]$
Suppose the covariance matrix $R$ is unknown. Then consider the two sample experiment $[X, y]$, where $X = [x[1], x[2], \ldots, x[M]]$ is a sequence of $M$ realizations of $x \sim CN_n[0, \gamma^2 R]$.

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

The distribution of this statistic is complicated, but may be written as a mixture of betas, independent of $R$. The geometry remains unchanged from that of the MSD.
Could this result be obtained from a coherence argument? Begin with a measurement vector $z \in \mathbb{C}^n$ and a linear space of dimension $p$, spanned by the columns of the matrix $G \in \mathbb{C}^{n \times p}$. Construct the data matrix $Z$ and its Grammian:

$$Z = \begin{bmatrix} z & G \end{bmatrix}; \quad Z^HZ = \begin{bmatrix} z^HZ & z^HG \\ G^Hz & G^HG \end{bmatrix}$$

Coherence is then

$$\rho^2(Z^HZ) = 1 - \frac{\det[Z^HZ]}{\det[z^Hz] \det[G^HG]} = \frac{z^HP_Gz}{z^Hz}$$

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

**Figure 13:** Array’s worth of time series.

- $L$-element array; $n$-sample time series at each array element $\ell$, denoted $\mathbf{x}_\ell \in \mathbb{C}^n$;
- These time series vectors may be organized into a long space-time vector, by stacking them top-to-bottom:
  \[
  \mathbf{x} = (\mathbf{x}_1^T, \cdots, \mathbf{x}_L^T)^T \in \mathbb{C}^{Ln}.
  \]
The covariance matrix of this space-time vector $\mathbf{x}$ is

$$ \mathbf{R} = \mathbb{E} \mathbf{x} \mathbf{x}^H = \{ \mathbf{R}_{\ell m} \} \in \mathbb{C}^{N \times N}.$$ 

Each of the $n \times n$ blocks $\mathbf{R}_{\ell m} = \mathbb{E} \mathbf{x}_\ell \mathbf{x}_m^H \in \mathbb{C}^{n \times n}$ is the cross correlation between time series at sensors $\ell$, $m$.

\[
\begin{array}{cccc}
\mathbf{R} &=& \begin{array}{ccc}
\mathbf{C} & \cdots & \mathbf{C} \\
\mathbf{C} & \cdots & \mathbf{C} \\
\mathbf{C} & \cdots & \mathbf{C} \\
\end{array}
\end{array}
\]

**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.
Hadamard coherence for this problem is

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

where the \( R_{\ell\ell} \in \mathbb{C}^{n \times n} \) are the matrix blocks of \( R \) lying on the diagonal. When the time series at all \( L \) sensors are jointly WSS, bandlimited to \( \Omega \), 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

\[ \rho^2(R) = 1 - \frac{\det[R]}{\prod_{\ell=1}^{L} \det[R_{\ell\ell}]} \longrightarrow 1 - \exp\left\{ \int_{\Omega} \log \frac{\det[K(e^{i\theta})]}{\prod_{\ell=1}^{L} K_{\ell\ell}(e^{i\theta})} \frac{d\theta}{2\pi} \right\} \]

where \( K(e^{i\theta}) = \{K_{\ell m}(e^{i\theta})\} \in \mathcal{C}^{L \times L} \) is multi-channel spectral correlation and \( K_{\ell m}(e^{i\theta}) \) is the discrete-time Fourier transform of the cross correlation sequence \( \{r_{\ell m}[n]\} \) between the times series at sensors \( \ell, m \), at lag \( n \). The RHS we call broadband coherence.
We would now like to give ourselves multiple experimental copies of the space-time field, and its sample correlation matrix, namely

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

and test \( H_0 : \mathbf{x} : \mathcal{CN}_L[\mathbf{0}, \mathbf{R} = \text{diag}[\mathbf{R}_{\ell\ell}]] \) vs \( H_1 : \mathbf{x} : \mathcal{CN}_L[\mathbf{0}, \mathbf{R} > \mathbf{0}] \). 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.
The null distribution of this statistic is

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

\[-\log \Lambda \longrightarrow \chi^2_\nu, \ \nu = \dim R - \dim R_0 = L^2 n^2 - Ln^2,\]

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

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

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

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