Detection of Multiple Correlated Time Series and its Application in Synthetic Aperture Sonar Imagery

Nick Klausner

Colorado State University
Department of Electrical and Computer Engineering

Advisor: Dr. Mahmood R. Azimi-Sadjadi
Co-Advisor: Dr. Louis Scharf

Committee Members:
Dr. Ali Pezeshki - ECE
Dr. Dan Cooley - Statistics

May 20, 2014
1 Introduction

2 Review of the GLRT

3 Null Distribution of the GLRT
   • Stochastic Representation
   • Asymptotic Null Distribution
   • Saddlepoint Approximations

4 Detection of WSS Processes
   • Frequency Domain GLRT
   • Extension to 2D WSS
   • Simulation

5 Application in Sonar Imagery
   • Data Description
   • False Alarm Studies
   • Detection Results

6 Conclusions
Introduction

- Detecting the presence of a common but unknown signal among two or more data channels is a problem that finds its uses in a variety of applications
  1. collaborative sensor networks
  2. geological monitoring of seismic activity
  3. both active and passive radar and sonar

- One possible solution to this problem is to construct a Generalized Likelihood Ratio Test (GLRT) that tests whether or not the composite covariance matrix of all the channels is block diagonal through the use of a generalized Hadamard ratio

- Often times, however, the most difficult part of hypothesis testing lies not so much in deriving the appropriate criteria, but rather in finding their exact distributions when the hypotheses are true and identifying the best critical region to adopt

- In this presentation, we'll take a close look at the probabilistic properties of this GLRT under the null hypothesis and discuss several methods that can be used to determine thresholds which approximately achieve a given false alarm probability

- The second subject is the detection of multiple WSS processes

- Asymptotic results on large Toeplitz matrices show that the generalized Hadamard ratio converges to a narrowband Hadamard ratio integrated over the Nyquist band, a broadband coherence statistic

- Ultimate goal is the application of these techniques to underwater target detection in pairs of SAS images of the seafloor
Multichannel Detection Using the GLRT

- Problem: test for the independence among $L$ random vectors \( \{x_i\}_{i=1}^{L} \) with $x_i = [x_i[0] \cdots x_i[N-1]]^T \in \mathbb{C}^N$
  - e.g. collection of multiple length-$N$ time series at distinct spatial locations

- Assuming this collection of random vectors to be zero mean, the composite vector $z = [x_1^T \cdots x_L^T]^T$ has space-time covariance matrix

\[
R = E[zz^H] = \begin{bmatrix}
R_{11} & R_{12} & \cdots & R_{1L} \\
R_{12}^H & R_{22} & \cdots & R_{2L} \\
\vdots & \vdots & \ddots & \vdots \\
R_{1L}^H & R_{2L}^H & \cdots & R_{LL}
\end{bmatrix} \in \mathbb{C}^{LN \times LN}
\]

- If $z \sim \mathcal{CN}(0, R)$, testing for independence is equivalent to the hypothesis test

\[
H_0 : R \in \mathcal{R}_0 \quad (\text{temporally correlated but spatially uncorrelated})
\]
\[
H_1 : R \in \mathcal{R} \quad (\text{both spatially and temporally correlated})
\]

where $\mathcal{R}$ denotes the space of all positive-definite Hermitian matrices and $\mathcal{R}_0 \subset \mathcal{R}$ denoting those matrices in $\mathcal{R}$ which are additionally block-diagonal.

- Given an experiment producing $M$ iid realizations $\{x_i[m]\}_{m=1}^{M}$ of the random vector from each channel

\[
x_i[m] = [x_i[0, m] \cdots x_i[N-1, m]]^T \in \mathbb{C}^N
\]

- Collection of composite random vectors $Z = [z[1] \cdots z[M]] \in \mathbb{C}^{LN \times M}$ has probability density function (PDF)

\[
f(Z; R) = \prod_{m=1}^{M} f(z[m]; R) = \frac{1}{\pi^{LN} M^{\frac{1}{2}}} \frac{1}{\det(R)^{\frac{1}{2}}} \exp \left\{ -M \text{tr} \left( R^{-1} \hat{R} \right) \right\}
\]
Multichannel Detection Continued

- $\hat{R}$ is the estimated composite covariance matrix

$$\hat{R} = \frac{1}{M} ZZ^H = \frac{1}{M} \sum_{m=1}^{M} z[m]z^H[m] = \begin{bmatrix}
\hat{R}_{11} & \hat{R}_{12} & \cdots & \hat{R}_{1L} \\
\hat{R}_{12} & \hat{R}_{22} & \cdots & \hat{R}_{2L} \\
\vdots & \vdots & \ddots & \vdots \\
\hat{R}_{1L} & \hat{R}_{2L} & \cdots & \hat{R}_{LL}
\end{bmatrix}$$

- Replacing the deterministic but unknown matrix $R$ with its ML estimate under each model, the likelihood ratio is a generalized Hadamard ratio

$$\Lambda = \left( \frac{\max_{R \in \mathcal{R}_0} f(Z; R)}{\max_{R \in \mathcal{R}} f(Z; R)} \right)^{1/M} = \frac{\det(\hat{R})}{\det(\hat{D})} = \frac{\det(\hat{R})}{\prod_{i=1}^{L} \det(\hat{R}_{ii})} = \det(\hat{C})$$

- Here, $\hat{R}$ and $\hat{D} = \text{blkdiag}\{\hat{R}_{11}, \ldots, \hat{R}_{LL}\}$ are the maximum likelihood estimators under the alternative and null hypotheses, respectively, and $\hat{C} = \hat{D}^{-1/2} \hat{R} \hat{D}^{-H/2}$ is a coherence matrix.

- Under suitable choices for the matrix $T$, the hypothesis testing problem and the likelihood ratio statistic remain unchanged upon replacing the random vector $z$ with the random vector $\tilde{z} = Tz$

1. The set of all matrices $T$ such that $T = \text{blkdiag}\{T_1, \ldots, T_L\}$ with $T_i$ any $N \times N$ invertible matrix

   $\Rightarrow$ There is no channel-by-channel invertible linear transformation, including scaling and filtering, that moves a covariance from $\mathcal{R}_0$ to $\mathcal{R}$ or vice versa.

2. The set of all matrices $T$ such that $T = P \otimes I_N$ with $P$ any $L \times L$ permutation matrix

   $\Rightarrow$ The ordering in channel index has no influence on likelihood.
Decomposition Using Gram Determinants

- For any $i \geq 2$ and any $n = 0, \ldots, N - 1$, data matrix $Z$ can be partitioned

$$Z = \begin{bmatrix}
Z_i \\
X_{in} \\
x_{in}^H \\
\vdots
\end{bmatrix} =
\begin{bmatrix}
x_1[1] & x_1[2] & \ldots & x_1[M] \\
\vdots & \vdots & \ddots & \vdots \\
x_{i-1}[1] & x_{i-1}[2] & \ldots & x_{i-1}[M] \\
x_i[0,1] & x_i[0,2] & \ldots & x_i[0,M] \\
x_i[1,1] & x_i[1,2] & \ldots & x_i[1,M] \\
\vdots & \vdots & \ddots & \vdots \\
x_i[n-1,1] & x_i[n-1,2] & \ldots & x_i[n-1,M] \\
x_i[n,1] & x_i[n,2] & \ldots & x_i[n,M] \\
\vdots & \vdots & \ddots & \vdots
\end{bmatrix}
$$

$Z_i \in \mathbb{C}^{(i-1)N \times M}$ : All $M$ realizations of the time series $x_1, \ldots, x_{i-1}$

$X_{in} \in \mathbb{C}^{n \times M}$ : All $M$ realizations of the time series at sensor $i$ up to temporal sample $n - 1$

$x_{in} \in \mathbb{C}^M$ : All $M$ realizations of the random variable $x_i[n]$

- With this, the northwest corner of the Gram matrix $ZZ^H$ obtains the following structure

$$ZZ^H = \begin{bmatrix}
\hat{R}_{ZZ} & \hat{R}_{ZX} & \hat{R}_{XX} & \ldots \\
\hat{R}_{Z} & \hat{R}_{X} & \hat{R}_{x} & \ldots \\
\hat{R}_{Z}^H & \hat{R}_{X}^H & \hat{R}_{x}^H & \ldots \\
\hat{R}_{Z}^{HH} & \hat{R}_{X}^{HH} & \hat{R}_{x}^{HH} & \ldots
\end{bmatrix}
\begin{align*}
\hat{R}_{ZZ} &= Z_i Z_i^H, \quad \hat{R}_{ZX} = Z_i X_{in}^H, \quad \hat{R}_{XX} = X_{in} X_{in}^H \\
\hat{R}_{Z} &= x_{in}^H, \quad \hat{R}_{X} = x_{in} x_{in}^H \\
\hat{R}_{x} &= x_{in}^H, \quad \hat{R}_{x} = x_{in} x_{in}^H
\end{align*}
Gram Determinants Continued

- Gram determinants are a technique used to test whether or not a collection of vectors are linearly independent.
- Using this theory, one can show that the determinant of the Gram matrix \( \mathbf{Z} \mathbf{Z}^H \) can be written

\[
M^{LN} \det \hat{\mathbf{R}} = \det(\mathbf{Z} \mathbf{Z}^H) = \det \left( X_{1N} X_{1N}^H \right) \prod_{i=2}^{N-1} \prod_{n=0}^{\infty} \sigma_{in}^2(\hat{\mathbf{R}})
\]

\[
\sigma_{in}^2(\hat{\mathbf{R}}) = \hat{\mathbf{r}}_{xx} - \left[ \hat{\mathbf{r}}_{Zx}^H \hat{\mathbf{r}}_{Xx}^H \right] \left[ \begin{array}{cc}
\hat{\mathbf{R}}_{ZZ} & \hat{\mathbf{R}}_{ZX} \\
\hat{\mathbf{R}}_{Zx} & \hat{\mathbf{R}}_{Xx}
\end{array} \right]^{-1} \left[ \begin{array}{c}
\hat{\mathbf{r}}_{Zx} \\
\hat{\mathbf{r}}_{Xx}
\end{array} \right] = \mathbf{x}_{in}^H P_{\mathbf{ZX}}^\perp \mathbf{x}_{in}
\]

- A similar result holds for the Gram matrix \( X_{iN} X_{iN}^H \)

\[
M^N \det \hat{\mathbf{R}}_{ii} = \det \left( X_{iN} X_{iN}^H \right) = \prod_{n=0}^{N-1} \sigma_{in}^2(\hat{\mathbf{R}}_{ii})
\]

\[
\sigma_{in}^2(\hat{\mathbf{R}}_{ii}) = \hat{\mathbf{r}}_{xx} - \hat{\mathbf{r}}_{Xx}^H \hat{\mathbf{R}}_{XX}^{-1} \hat{\mathbf{r}}_{Xx} = \mathbf{x}_{in}^H P_{X}^\perp \mathbf{x}_{in} = \mathbf{x}_{in}^H P_{\mathbf{ZX}}^\perp \mathbf{x}_{in} + \mathbf{x}_{in}^H P_{X_{iN}^H} \mathbf{x}_{in}
\]

- \( P_{\mathbf{ZX}} \) : Projection onto the \((i-1)N + n\) dimensional subspace \( \left[ \mathbf{Z}_{iN}^H \mathbf{X}_{iN}^H \right] \)
- \( P_{X} \) : Projection onto the \(n\) dimensional subspace \( \mathbf{X}_{iN}^H \)
- \( P_{P_{X}^\perp \mathbf{Z}} \) : Projection onto the \((i-1)N\) dimensional subspace \( P_{X_{iN}^H Z_{iN}^H} \)

These quadratic forms represent the squared residual from two linear regression problems that look to predict \( x_i[n] \)

- \( \sigma_{in}^2(\hat{\mathbf{R}}) \) uses \( x_1, \ldots, x_{i-1} \) and \( x_i[0], \ldots, x_i[n-1] \)
- \( \sigma_{in}^2(\hat{\mathbf{R}}_{ii}) \) uses \( x_i[0], \ldots, x_i[n-1] \) only

With these decompositions, the likelihood can be expressed as a product of scalars

\[
\Lambda = \frac{\det \hat{\mathbf{R}}}{\det \hat{\mathbf{D}}} = \frac{\det(\mathbf{Z} \mathbf{Z}^H)}{\prod_{i=1}^{L} \det \left( X_{iN} X_{iN}^H \right)} = \frac{\det \left( X_{1N} X_{1N}^H \right)}{\prod_{i=2}^{N-1} \prod_{n=0}^{\infty} \sigma_{in}^2(\hat{\mathbf{R}})} \cdot \frac{\prod_{i=2}^{N-1} \prod_{n=0}^{\infty} \mathbf{x}_{in}^H P_{\mathbf{ZX}}^\perp \mathbf{x}_{in} \cdot \prod_{i=2}^{N-1} \prod_{n=0}^{\infty} \mathbf{x}_{in}^H P_{\mathbf{X}^H_{1N} \mathbf{X}_{1N}^H} \mathbf{x}_{in}}{\prod_{i=2}^{N-1} \prod_{n=0}^{\infty} \mathbf{x}_{in}^H P_{\mathbf{ZX}}^\perp \mathbf{x}_{in} + \mathbf{x}_{in}^H P_{\mathbf{X}^H_{1N} \mathbf{X}_{1N}^H} \mathbf{x}_{in}}
\]
Stochastic Representation Under the Null Hypothesis

- Under $\mathcal{H}_0$, $z \sim \mathcal{CN}(0, D)$ for some $D = \text{blkdiag} \{R_{11}, \ldots, R_{LL}\} \in \mathcal{R}_0$

- By invariance, can always apply the linear transformation $D^{-1/2} = \text{blkdiag} \{R_{11}^{-1/2}, \ldots, R_{LL}^{-1/2}\}$ without any consequence to the likelihood ratio $\implies$ no loss in generality to assume that $D = I_{LN}$ (note that $I_{LN} \in \mathcal{R}_0$) or equivalently that $x_{in} \overset{iid}{\sim} \mathcal{CN}(0, I_M)$

- Orthogonality in $P_{ZX} x_{in}$ and $P_{P X} Z x_{in}$ implies that:
  \[ 2x_{in}^H P_{ZX} x_{in} \sim \chi^2_{2\alpha_i} \]
  \[ 2x_{in}^H P_{P X} Z x_{in} \sim \chi^2_{2\beta_i} \]

- If $X_1 \sim \chi^2_{\nu_1}$, $X_2 \sim \chi^2_{\nu_2}$, and $X_1 \perp \perp X_2$ then $X_1 \frac{X_1}{X_1+X_2} \sim \text{Beta} \left( \frac{\nu_1}{2} , \frac{\nu_2}{2} \right)$

Thus, the likelihood ratio under $\mathcal{H}_0$ is equal in distribution to a product of independent beta rv’s

\[ \Lambda | \mathcal{H}_0 \overset{d}{=} \prod_{i=2}^L \prod_{n=0}^{N-1} Y_{in} \]

\[ Y_{in} \sim \text{Beta} \left( \alpha_{in} , \beta_i \right) \]

- If $z \sim \mathcal{N}(0, R)$ then $Y_{in} \sim \text{Beta} \left( \frac{\alpha_{in}}{2} , \frac{\beta_i}{2} \right)$

- Two statistically equivalent means of simulating the null distribution:
Asymptotic Null Distribution

Wilk’s Theorem

Let \( \{x_i\}_{i=1}^M \) be an iid sample from \( f(x; \theta) \) where \( \theta \in \Theta \subseteq \mathbb{R}^{T+s} \). Furthermore, assume that the parameter vector is partitioned \( \theta = [\eta^T \xi^T] \) and consider the \( s \)-dimensional subspace \( \Theta_0 = \{(\eta, \xi) \in \Theta : \eta = \eta_0\} \). Then under the null hypothesis \( H_0 : \eta = \eta_0 \)

\[
-2 \ln \left( \frac{\max_{\theta \in \Theta_0} \prod_{i=1}^M f(x_i; \theta)}{\max_{\theta \in \Theta} \prod_{i=1}^M f(x_i; \theta)} \right) \to \chi^2_\nu \quad ; \quad \nu = \dim(\Theta) - \dim(\Theta_0)
\]

- Generalized Hadamard ratio has the following moments under the null hypothesis

\[
E[\Lambda^k | H_0] = \prod_{i=2}^L \prod_{n=0}^{N-1} E[Y^k_{in}] = \prod_{i=2}^L \prod_{n=0}^{N-1} \frac{B(\alpha_i n + k, \beta_i)}{B(\alpha_i n, \beta_i)} = \prod_{i=2}^L \prod_{n=0}^{N-1} \frac{\Gamma(M - n)\Gamma(M - (i - 1)N - n + k)}{\Gamma(M - (i - 1)N - n)\Gamma(M - n + k)}
\]

- If we define \( \xi_n = (1 - \rho)M - n \) with \( 0 \leq \rho \leq 1 \) an arbitrary real number, the characteristic function of the random variable \( Z = -2\rho M \ln \Lambda \) can be written

\[
\phi_Z(jt) = E[e^{jtZ} | H_0] = E[\Lambda^{-2j\rho Mt} | H_0] = \prod_{i=2}^L \prod_{n=0}^{N-1} \frac{\Gamma(\rho M + \xi_n)\Gamma(\rho M(1 - 2jt) + \xi_n - (i - 1)N)}{\Gamma(\rho M + \xi_n - (i - 1)N)\Gamma(\rho M(1 - 2jt) + \xi_n)}
\]

- Or, equivalently, its cumulant generating function \( \psi(jt) = \ln \phi_X(jt) \) is given by

\[
\psi_Z(jt) = \sum_{i=2}^L \sum_{n=0}^{N-1} \left[ \ln \Gamma(\rho M + \xi_n) - \ln \Gamma(\rho M + \xi_n - (i - 1)N) \right] + \sum_{i=2}^L \sum_{n=0}^{N-1} \ln \Gamma(\rho M(1 - 2jt) + \xi_n - (i - 1)N) - \sum_{i=2}^L \sum_{n=0}^{N-1} \ln \Gamma(\rho M(1 - 2jt) + \xi_n)
\]

- As \( |z| \to \infty \) and provided that \(|\arg z| < \pi\), the \( \ln \Gamma(\cdot) \) function has the asymptotic expansion where \( B_n(x) \) denotes an \( n^{th} \) order Bernoulli polynomial

\[
\ln \Gamma(z + a) = \frac{1}{2} \ln 2\pi + (z + a - \frac{1}{2}) \ln z - z + \sum_{n=1}^{\infty} (-1)^{n+1} \frac{B_{n+1}(a)}{n(n+1)z^n}
\]
Asymptotic Distribution Continued

- Using the 2\textsuperscript{nd} order Bernoulli polynomial \( B_2(x) = x^2 - x + \frac{1}{6} \), one obtains the asymptotic expression

\[
\psi_Z(t) = -\frac{\nu}{2} \ln(1 - 2jt) + \omega_1(\rho) \left[ (1 - 2jt)^{-1} - 1 \right] + \mathcal{O}(M^{-2})
\]

\[
\nu = 2 \sum_{i=2}^{L} \sum_{n=0}^{N-1} (i - 1)N = L^2 N^2 - LN^2
\]

\[
\omega_1(\rho) = \frac{1}{2\rho M} \sum_{i=2}^{L} \sum_{n=0}^{N-1} \left[ (i - 1)^2 N^2 + (1 - 2\xi_n)(i - 1)N \right] = \frac{1}{2\rho} \left( -\nu(1 - \rho) + \frac{L(L^2 - 1)N^3}{6M} \right)
\]

- Why scale by \( \rho \)? Set \( \rho \) so that \( \omega_1(\rho) \) vanishes \( \Rightarrow \) Achieve an approximation with error that is \( \mathcal{O}(M^{-2}) \) versus \( \mathcal{O}(M^{-1}) \) if one were to set \( \rho = 1 \)

- Setting \( \rho = \rho^* = 1 - \frac{(L+1)N}{6M} \) so that \( \omega_1(\rho^*) = 0 \) and letting \( M \) tend to infinity one obtains

\[
\phi_Z(t) = e^{\psi_Z(t)} M \to \infty (1 - 2jt)^{-\nu/2}
\]

- Carefully counting the number of independent parameters under each hypothesis, one finds that \( \dim(\mathcal{R}) = L^2 N^2 \) and \( \dim(\mathcal{R}_0) = LN^2 \) so that \( \nu = \dim(\mathcal{R}) - \dim(\mathcal{R}_0) \)

- Thus, for large \( M \) it follows that

\[
P \left[ -2\rho^* M \ln \Lambda \leq x \right] = P \left[ \left( \frac{1}{3} (L + 1)N - 2M \right) \ln \Lambda \leq x \right] \xrightarrow{M \to \infty} P \left[ \chi^2_\nu \leq x \right]
\]

\[\begin{array}{c}
L = 2, N = 24 \\
L = 3, N = 24 \\
L = 4, N = 24
\end{array}\]
Saddlepoint Approximations

- Saddlepoint approximations are powerful tools for obtaining accurate approximations for densities and distribution functions by approximating the inverse Fourier transform of a random variable’s known characteristic function.

- Given the random variable $X$ with CGF $\psi_X(t)$, the saddlepoint density uses the approximation

$$f_X(x) \approx \frac{1}{\sqrt{2\pi} \psi_X''(\hat{t}(x))} \exp \left\{ \psi_X(\hat{t}(x)) - \hat{t}(x)x \right\}$$

where the saddlepoint $\hat{t}(x)$ is the unique value of $t$ such that $\psi'_X(t) = x$ and $\psi''_X(\hat{t}(x)) > 0$.

- For any real-valued scalar $c$, the random variable $Z = c \ln \Lambda |\mathcal{H}_0$ has the MGF $\phi_Z(t)$

$$\phi_Z(t) = E \left[ e^{tZ} |\mathcal{H}_0 \right] = E \left[ \Lambda^{ct} |\mathcal{H}_0 \right] = \prod_{i=2}^{L} \prod_{n=0}^{N-1} \frac{\Gamma(M - n) \Gamma(M - (i - 1)N - n + ct)}{\Gamma(M - (i - 1)N - n) \Gamma(m - n + ct)}$$

- The CGF $\psi_Z(t) = \ln \phi_Z(t)$ has $k^{th}$-order derivative with polygamma function $\gamma_k(x) = \frac{d^{(k+1)} \ln \Gamma(x)}{dx^{(k+1)}}$

$$d^k \psi_Z(t) = c^k \sum_{i=2}^{L} \sum_{n=0}^{N-1} \left[ \gamma_{k-1}(M - (i - 1)N - n + ct) - \gamma_{k-1}(M - n + ct) \right]$$

- First step is to identify the saddlepoint, the value of $t$ such that

$$\psi'_Z(t) = c \sum_{i=2}^{L} \sum_{n=0}^{N-1} \left[ \gamma_{0}(M - (i - 1)N - n + ct) - \gamma_{0}(M - n + ct) \right] = x$$

- No simple closed-form solution $\Rightarrow$ Have to use numerical root-finding techniques, e.g. Newton-Raphson method.

$L = 2, N = 24$

$L = 3, N = 24$

$L = 4, N = 24$
Detection of WSS Processes

- Consider the original problem consisting of several sensors, each measuring the length-$N$ time series $x_i = [x_i[0] \cdots x_i[N-1]]$

- Goal is to develop methods that exploit the inherent Toeplitz structure in each block $R_{ik} = E[x_i x_k^H]$ when $\{x_i\}_{i=1}^L$ are jointly WSS

- Apply the linear transformation $T = F_N$ to each channel with $[F_N]_{\ell,k} = \frac{1}{\sqrt{N}} e^{-j2\pi\ell k/N}$ an $N \times N$ DFT matrix

- By invariance of the GLRT, the generalized Hadamard ratio remains unchanged and can be written

$$\Lambda = \det \left( \hat{C} \right) = \det \left( (F_N \otimes I_L)\hat{C} (F_N \otimes I_L)^H \right)$$

$$\hat{C} = \hat{D}^{-1/2} \hat{R} \hat{D}^{-H/2}$$

- Introduce a simple permutation to the rows and columns of the matrix inside this determinant to write the likelihood ratio as $\Lambda = \det \hat{C}$ with global coherence matrix $\hat{C}$ for frequencies $\theta_{\ell} = \frac{2\pi}{N} \ell$ for $\ell = 0, \ldots, N-1$

$$\hat{C} = \begin{bmatrix} \hat{C}(e^{j\theta_0}) & \hat{C}(e^{j\theta_0}, e^{j\theta_1}) & \ldots & \hat{C}(e^{j\theta_0}, e^{j\theta N-1}) \\ \hat{C}(e^{j\theta_1}, e^{j\theta_0}) & \hat{C}(e^{j\theta_1}) & \ldots & \hat{C}(e^{j\theta_1}, e^{j\theta N-1}) \\ \vdots & \vdots & \ddots & \vdots \\ \hat{C}(e^{j\theta N-1}, e^{j\theta_0}) & \hat{C}(e^{j\theta N-1}, e^{j\theta_1}) & \ldots & \hat{C}(e^{j\theta N-1}) \end{bmatrix}$$

- Every block $\hat{C}(e^{j\theta_{\ell}}, e^{j\theta_{m}}) \in \mathbb{C}^{L \times L}$ with $\hat{C}(e^{j\theta_{\ell}}) = \hat{C}(e^{j\theta_{\ell}}, e^{j\theta_{\ell}})$ has elements $\hat{C}_{ik}(e^{j\theta_{\ell}}, e^{j\theta_{m}})$ of the form

$$\hat{C}_{ik}(e^{j\theta_{\ell}}, e^{j\theta_{m}}) = f^H_N(e^{j\theta_{\ell}}) \hat{R}_{ii}^{-1/2} \hat{R}_{ik} \hat{R}_{kk}^{-H/2} f_N(e^{j\theta_{m}})$$

$$f_N(e^{j\theta}) = \frac{1}{\sqrt{N}} \left[ 1 \ e^{j\theta} \ e^{j2\theta} \ \ldots \ e^{j(N-1)\theta} \right]^T$$

- That is, if we define the DFT of the "whitened" random vector of each channel $w^{(m)}_i(e^{j\theta_{\ell}}) = f^H_N(e^{j\theta_{\ell}}) \hat{R}_{ii}^{-1/2} x_i[m]$ then

$$\hat{C}_{ik}(e^{j\theta_{\ell}}, e^{j\theta_{m}}) = \frac{1}{M} \sum_{m=1}^M w^{(m)}_i(e^{j\theta_{\ell}}) \left( w^{(m)}_k(e^{j\theta_{m}}) \right)^*$$
Any \( N \times N \) Toeplitz matrix \( T_N \) is asymptotically equivalent to a circulant matrix \( C_N \) in the sense that
\[
\frac{1}{N} \left\| T_N - C_N \right\|_F \xrightarrow{N \to \infty} 0
\]

When the channels are truly WSS, this asymptotic shows that the global coherence matrix is asymptotically equivalent to the block-diagonal matrix \( \tilde{C} \to \text{blkdiag} \{ \hat{C}(e^{j\theta_0}), \ldots, \hat{C}(e^{j\theta_N-1}) \} \) so that as \( M \) and \( N \) grow large
\[
\Lambda^{1/N} N \xrightarrow{\infty} \exp \left\{ \int_{-\pi}^{\pi} \ln \det \hat{C}(e^{j\theta}) \frac{d\theta}{2\pi} \right\} = \exp \left\{ \int_{-\pi}^{\pi} \ln \frac{\det \hat{S}(e^{j\theta})}{\prod_{i=1}^{L} \hat{S}_{ii}(e^{j\theta})} \frac{d\theta}{2\pi} \right\}
\]
\( \hat{S}(e^{j\theta}) \in \mathbb{C}^{L \times L} \) for \(-\pi < \theta \leq \pi\) is the estimated cross-spectral matrix
\[
\hat{S}(e^{j\theta}) = \begin{bmatrix}
\hat{S}_{11}(e^{j\theta}) & \hat{S}_{12}(e^{j\theta}) & \cdots & \hat{S}_{1L}(e^{j\theta}) \\
\hat{S}_{12}^*(e^{j\theta}) & \hat{S}_{22}(e^{j\theta}) & \cdots & \hat{S}_{2L}(e^{j\theta}) \\
\vdots & \vdots & \ddots & \vdots \\
\hat{S}_{1L}^*(e^{j\theta}) & \hat{S}_{2L}^*(e^{j\theta}) & \cdots & \hat{S}_{LL}(e^{j\theta})
\end{bmatrix}
\]
\( \hat{S}_{ik}(e^{j\theta}) = f_{N_k}^H(e^{j\theta}) \hat{R}_{ik} f_{N_k}(e^{j\theta}) \): quadratic estimate of the cross-spectrum at frequency \( \theta \)

Not the true GLRT for WSS processes but a computationally efficient alternative

Spectral analysis produces the composite vector
\[
Z^{(m)}(e^{j\theta \ell}) = \begin{bmatrix}
f_{N_k}^H(e^{j\theta \ell}) \times 1 \\
\vdots \\
f_{N_k}^H(e^{j\theta \ell}) \times L
\end{bmatrix}
\]

Average periodograms
\[
\hat{S}(e^{j\theta \ell}) = \frac{1}{M} \sum_{m=1}^{M} Z^{(m)}(e^{j\theta \ell}) (\cdot)^H
\]
Integrate a narrowband Hadamard ratio over the Nyquist band

Discussion
Extension to 2D WSS Processes

- In certain applications, one may collect multiple time series per channel so that the vector \( x_i \) contains \( N \) samples of a \( P \)-dimensional time series, e.g. collection of length-\( N \) time series in a network of \( P \)-element sensor arrays

\[
x_i = \left[ x_i^T [0] \cdots x_i^T [N-1] \right]^T \in \mathbb{C}^{NP}
\]

\[
x_i[n] = [x_i[n, 0] \cdots x_i[n, P-1]]^T \in \mathbb{C}^P
\]

- Goal: apply the same idea but for Toeplitz-block-Toeplitz structured matrices

- Apply the linear transformation \( T = F_N \otimes F_P \) to each channel, i.e. a 2D DFT applied both spatially and temporally

- Again, invariance of the GLRT guarantees that the generalized Hadamard ratio remains unchanged

\[
\Lambda = \det (\hat{C}) = \det \left( (T \otimes I_L)\hat{C}(T \otimes I_L)^H \right)
\]

- For any frequency \( \theta \), a permutation can be applied to the frequency-dependent coherence matrix \( \hat{C}(e^{j\theta}) \) such that

\[
\det \hat{C}(e^{j\theta}) = \det \tilde{C}(e^{j\theta})
\]

where

\[
\tilde{C}(e^{j\theta}) = \begin{bmatrix}
\hat{C}(e^{j\theta}, e^{j\phi_0}) & \cdots & \hat{C}(e^{j\theta}, e^{j\phi_0}, e^{j\phi P-1}) \\
\vdots & \ddots & \vdots \\
\hat{C}(e^{j\theta}, e^{j\phi P-1}, e^{j\phi_0}) & \cdots & \hat{C}(e^{j\theta}, e^{j\phi P-1})
\end{bmatrix}
\]

i.e. a global coherence matrix with wavenumbers \( \phi_\ell = \frac{2\pi}{P} \ell \) for \( \ell = 0, \ldots, P-1 \)

- When data is both temporally and spatially WSS, \( \tilde{C}(e^{j\theta}) \to \text{blkdiag} \{ \hat{C}(e^{j\theta}, e^{j\phi_0}), \ldots, \hat{C}(e^{j\theta}, e^{j\phi P-1}) \} \) so that as \( M, N \) and \( P \) grow large

\[
\Lambda \frac{1}{NP} \xrightarrow{N,P, \to \infty} \exp \left\{ \int_{-\pi}^{\pi} \int_{-\pi}^{\pi} \ln \det \hat{C}(e^{j\theta}, e^{j\phi}) \frac{d\theta d\phi}{4\pi^2} \right\} = \exp \left\{ \int_{-\pi}^{\pi} \int_{-\pi}^{\pi} \ln \frac{\det \hat{S}(e^{j\theta}, e^{j\phi})}{\prod_{i=1}^{L} \hat{S}_{ii}(e^{j\theta}, e^{j\phi})} \frac{d\theta d\phi}{4\pi^2} \right\}
\]

- Again, matrix \( \hat{S}(e^{j\theta}, e^{j\phi}) \in \mathbb{C}^{L \times L} \) is a cross-spectral matrix with elements

\[
\hat{S}_{ik}(e^{j\theta}, e^{j\phi}) = f_P^H (e^{j\phi}) \left( f_N^H (e^{j\theta} \ell) \otimes I_P \right) \hat{R}_{ik} \left( f_N (e^{j\theta} \ell) \otimes I_P \right) f_P (e^{j\phi})
\]
Multichannel Detection in Coregistered Images

Several applications where this test statistic may apply:
- Detection of correlated processes in a network of sensor arrays
- Detection of coherence among multiple images

Consider a pair \((L = 2)\) of real-valued images \(\{x_i[n, p]\}\) whose composite vector
\[ z[n, p] = [x_1[n, p] \ x_2[n, p]]^T \in \mathbb{R}^2 \]
follows the AR process
\[ z[n, p] = A_{10} z[n - 1, p] + A_{01} z[n, p - 1] + A_{11} z[n - 1, p - 1] + w[n, p] \]
\[ w[n, p] \overset{iid}{\sim} \mathcal{N}(0, \sigma^2 I_2) \]

Under \(H_0\): \(A_{jk} = \text{diag}(a_{11}^{jk}, a_{22}^{jk})\) for all \(j, k = 0, 1\)

Each image partitioned into \(8 \times 8\) blocks, “vectorized” and accumulated to form \(Z\)

Cross-spectral matrix \(\hat{S}(e^{j\theta}, e^{j\phi})\) estimated via Welch’s method

Plots give examples of each channel under both models, the estimated covariance matrix \(\hat{R}\), and the estimated magnitude coherence spectrum
\[ \hat{\gamma}(e^{j\theta}, e^{j\phi}) = \frac{|\hat{S}_{12}|}{\sqrt{\hat{S}_{11} \hat{S}_{22}}} (e^{j\theta}, e^{j\phi}) \]

When \(L = 2\), Hadamard Ratio \(= 1 - \hat{\gamma}^2(e^{j\theta}, e^{j\phi})\)

Example Under \(H_0\) 
Example Under \(H_1\)
Simulation Continued

- Generalized Hadamard ratio and broadband coherence both used to determine which model is in force
- By using unconstrained estimator $\hat{R}$, generalized Hadamard ratio is more generally applicable.
- However, broadband coherence statistic is better matched to WSS assumption.

- A pair of images are generated which, by and large, are uncorrelated ($H_0$).
- For several arbitrarily selected locations, data is correlated ($H_1$).
- Window scanned through each image and applied to both techniques.
  - Generalized Hadamard ratio fails to determine where correlation has been enforced.
  - Using broadband coherence, the locations of each region are clear.
Simulation Continued

- Generalized Hadamard ratio and broadband coherence both used to determine which model is in force.
- By using unconstrained estimator $\hat{R}$, generalized Hadamard ratio is more generally applicable.
- However, broadband coherence statistic is better matched to WSS assumption.

A pair of images are generated which, by and large, are uncorrelated ($H_0$).
For several arbitrarily selected locations, data is correlated ($H_1$).
Window scanned through each image and applied to both techniques.
Generalized Hadamard ratio fails to determine where correlation has been enforced.
Using broadband coherence, the locations of each region are clear.
Simulation Continued

- Generalized Hadamard ratio and broadband coherence both used to determine which model is in force
- By using unconstrained estimator \( \hat{R} \), generalized Hadamard ratio is more generally applicable
- However, broadband coherence statistic is better matched to WSS assumption

- A pair of images are generated which, by and large, are uncorrelated \((\mathcal{H}_0)\)
- For several arbitrarily selected locations, data is correlated \((\mathcal{H}_1)\)
- Window scanned through each image and applied to both techniques
- Generalized Hadamard ratio fails to determine where correlation has been enforced
- Using broadband coherence, the locations of each region are clear
Application in Sonar Imagery

- Navy is interested in automated detection and localization of underwater mines lying on the sea-floor
- A single Autonomous Underwater Vehicle (AUV) collects sonar returns from two frequency bands as the vehicle moves along-track and builds $L = 2$ (HF and BB) synthetic aperture sonar (SAS) images
- Underlying principle is that the presence of a target will lead to higher levels of coherence compared to background
- Images partitioned into overlapping ROIs, ROIs partitioned into blocks, all realizations used to estimate $\hat{R}$ and $\hat{S}(e^{j\theta}, e^{j\phi})$
- Difficult to argue that these $M$ realizations are truly iid given that they are spatially distributed across the ROI

- Applied to two datasets collected at different geographical locations
- Dataset 2 further partitioned into three environments

<table>
<thead>
<tr>
<th>Clutter Difficulty</th>
<th># Images</th>
<th># Targets</th>
</tr>
</thead>
<tbody>
<tr>
<td>Dataset 1</td>
<td>Medium/Hard</td>
<td>122</td>
</tr>
<tr>
<td>Day 1 Easy</td>
<td>180</td>
<td>4</td>
</tr>
<tr>
<td>Day 2 Hard</td>
<td>136</td>
<td>17</td>
</tr>
<tr>
<td>Day 3 Easy</td>
<td>142</td>
<td>47</td>
</tr>
<tr>
<td>Total</td>
<td>458</td>
<td>68</td>
</tr>
<tr>
<td>Dataset 2</td>
<td>Day 1 Easy</td>
<td>180</td>
</tr>
<tr>
<td>Day 2 Hard</td>
<td>136</td>
<td>17</td>
</tr>
<tr>
<td>Day 3 Easy</td>
<td>142</td>
<td>47</td>
</tr>
<tr>
<td>Total</td>
<td>458</td>
<td>68</td>
</tr>
</tbody>
</table>
Null Distribution in Sonar Imagery

- Several regions extracted throughout one image
  - Windows 1 – 3: Target
  - Windows 4 – 6: Cluttered Background
  - Windows 7 – 9: Uncluttered Background

- ROIs of size $80 \times 144$ partitioned into $8 \times 8$ blocks ($NP = 64, M = 180$)

- Figures show the ratio of squared residuals for $n = 0, \ldots, 63$
  \[ \eta_{in} = \frac{\sigma^2_{in}(\hat{R})}{\sigma^2_{in}(\hat{R}_{ii})} = \frac{x^H_{in} P_{\bot ZX} x_{in}}{x^H_{in} P_{\bot ZX} x_{in} + x^H_{in} P_{XZ} x_{in}} \]

- In this case ($L = 2$) Generalized Hadamard Ratio $= \prod_n \eta_{in}$

- Red lines signify a 95% bound under $\mathcal{H}_0$, i.e. the interval $[a, b] \subset [0, 1]$ such that
  \[ P[\eta_{in} \leq a | \mathcal{H}_0] = P[\eta_{in} \geq b | \mathcal{H}_0] = 0.025 \]

- Good agreement in uncluttered background (7 – 9), not as good in cluttered region (4 – 6)
False Alarm Studies

- Saddlepoint approximation used to determine threshold needed to theoretically achieve $P_{FA} = 10^{-2}$
- Top plots to right show HF images with targets outlined in red
- Middle images show the generalized Hadamard ratio for each ROI with regions above the threshold in red
- Bottom plots compare the histograms of the likelihood ratio to the saddlepoint approximation and asymptotic $\chi^2$ distribution
- Plots below give the empirical false alarm rate observed for both datasets using this threshold, dashed line depicts the desired $P_{FA}$
- In uncluttered backgrounds, one sees a fairly good agreement
- In environments with cluttered background, however, one will observe false alarms rates that are significantly higher than what is desired
Broadband Coherence in Sonar Imagery

- Examples from the same image studied before
  - Windows 1 – 3: Target
  - Windows 4 – 6: Cluttered Background
  - Windows 7 – 9: Uncluttered Background

- Plots display the HF and BB snippets for each example, the covariance estimate $\hat{R}$, and the estimated magnitude coherence $\hat{\gamma}(e^{j\theta}, e^{j\phi})$

- Generally observe a Toeplitz-block-Toeplitz behavior in the autocovariance matrices, especially for background

- Assumption that data is jointly WSS seems to be less applicable

- Be that as it may, one still observes a significant difference in the coherence spectrum when comparing target to background
Target Detection in Sonar Imagery

- Broadband coherence compared to the generalized Hadamard ratio with \( N = P = 8 \) and \( N = P = 1 \)
- Objects lying proud on the seafloor typically exhibit distinct highlight/shadow characteristic \( \Rightarrow \) search for regions in the magnitude of the HF image that match this behavior
- Let the magnitude of the HF ROI \( x \in \mathbb{R}^M \) follow the linear model
  \[
  x = \mu h + \phi 1 + w \quad \text{where} \quad w \sim \mathcal{N}(0, \sigma^2 I), \ 1_M = [1 \ \cdots \ 1]^T \in \mathbb{R}^M
  \]
  \( \Rightarrow \) Observation (\( x \)) = Target (\( \mu h \)) + Unknown Bias (\( \phi 1 \)) + Noise of Unknown Variance (\( w \))
- Testing the null hypothesis that \( \mu = 0 \) versus the alternative \( \mu > 0 \) yields the likelihood ratio
  \[
  \Lambda = \begin{cases} 
  0 & \hat{\mu} \leq 0 \\
  \frac{x^T P_1^\perp P_g P_1^\perp x}{x^T P_1^\perp P_g P_1^\perp x} & \hat{\mu} > 0
  \end{cases}
  \]
  where \( \hat{\mu} = \frac{h^T P_1^\perp x}{h^T P_1^\perp h} \), \( P_g = P P_1^\perp h \)
- Fundamentally different objective for this technique
  - Broadband coherence and generalized Hadamard ratio look for areas in HF and BB images which are highly correlated
  - Matched subspace detector looks for areas in the magnitude of the HF image which match the idealized template (\( h \))
- ROC curves compare performances on both datasets
  - BB coherence, generalized Hadamard ratio, matched subspace detector perform fairly similarly on Dataset 1
  - BB coherence exhibits improved performance on Dataset 2
  - Hadamard ratio seems to perform poorly on both datasets, intended for temporally white processes
Threshold empirically set to achieve $P_{FA} = 10^{-2}$

Figures compare likelihood ratio for BB coherence and matched subspace techniques on two images

Matched subspace can suffer model mismatch when lacking characteristic highlight/shadow structure

Nonparametric techniques can sometimes be an advantage for this problem

Detection Rates for Each Method

<table>
<thead>
<tr>
<th>Detection Method</th>
<th>Dataset 1</th>
<th>Dataset 2</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Day1</td>
<td>Day2</td>
</tr>
<tr>
<td>Broadband Coherence</td>
<td>68 (88%)</td>
<td>4 (100%)</td>
</tr>
<tr>
<td>Had. Ratio ($N = P = 1$)</td>
<td>29 (38%)</td>
<td>2 (50%)</td>
</tr>
<tr>
<td>Had. Ratio ($N = P = 8$)</td>
<td>61 (79%)</td>
<td>4 (100%)</td>
</tr>
<tr>
<td>Matched Subspace Detector</td>
<td>67 (87%)</td>
<td>4 (100%)</td>
</tr>
</tbody>
</table>

Average # False Alarms Per Image

<table>
<thead>
<tr>
<th>Detection Method</th>
<th>Dataset 1</th>
<th>Dataset 2</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Day1</td>
<td>Day2</td>
</tr>
<tr>
<td>Broadband Coherence</td>
<td>12.6</td>
<td>0.5</td>
</tr>
<tr>
<td>Had. Ratio ($N = P = 1$)</td>
<td>12.4</td>
<td>8.7</td>
</tr>
<tr>
<td>Had. Ratio ($N = P = 8$)</td>
<td>10.2</td>
<td>1.1</td>
</tr>
<tr>
<td>Matched Subspace Detector</td>
<td>13.2</td>
<td>2.3</td>
</tr>
</tbody>
</table>
Conclusions

- Detecting the presence of a common but unknown signal among multiple channels finds its uses in many applications, one of them being the detection of underwater targets in multiple sonar images.

- Testing the null hypothesis that a composite covariance matrix is block-diagonal is achieved using a generalized Hadamard ratio.

- Under the null hypothesis, the generalized Hadamard ratio is equal in distribution to a product of independent beta random variables.

- Allows one to define thresholds which approximately achieve a desired false alarm rate using techniques such as asymptotic distributions and saddlepoint approximations.

- Fairly capable of describing the probabilistic behavior in uncluttered environments, false alarm rates will be significantly higher when encountering clutter.

- When the data is WSS, the generalized Hadamard ratio is asymptotically given by a narrowband Hadmard ratio integrated over the Nyquist band, a broadband coherence.

- Even with finite support for measured signals, can provide improvements in detection performance.

- Although assumption of joint stationarity might be a bit of a stretch, broadband coherence proves to be an effective technique for underwater target detection in pairs of sonar images.

- Due to the wide variation in target conditions that can exist, methods that take advantage of general discriminative features in the data, as opposed to those that rely on specific target models, can in some cases be desirable.

- Broadband coherence detector transitioned and evaluated by a naval lab

\[ P_D = 86\%, \ 83\% \], high false alarm rate for one dataset.
Geometry of the GLRT

- As a simple example, consider the special case of real data with \( L = 2 \) and \( N = 1 \).
- In this case, every covariance matrix can be described as a point in \( \mathbb{R}^3 \)

\[
R = \begin{bmatrix}
\sigma_{11}^2 & \sigma_{12}^2 \\
\sigma_{12}^2 & \sigma_{22}^2
\end{bmatrix} \Leftrightarrow r = \begin{bmatrix}
\sigma_{11}^2 \\
\sigma_{22}^2 \\
\sigma_{12}^2
\end{bmatrix}
\]

- The set of points with \( \sigma_{11}^2, \sigma_{22}^2 > 0 \) and \( |\sigma_{12}^2| = \sqrt{\sigma_{11}^2 \sigma_{22}^2} \) describes the convex cone of positive semidefinite matrices.
- In this case \( \mathcal{R} \) denotes the 3D volume of points strictly inside this cone and \( \mathcal{R}_0 \) corresponds to a 2D plane that bisects the cone.
- Upon identifying the most likely point, \( \hat{R} \), we find \( \hat{D} \) by vertically dropping it into the plane which corresponds to the orthogonal projection of \( \hat{R} \) onto \( \mathcal{R}_0 \).

- It isn’t difficult to see that, in general, \( \hat{D} \) satisfies the principle of orthogonality. That is, for any \( R \in \mathcal{R}_0 \)

\[
\langle \hat{R} - \hat{D}, R \rangle = \text{tr} \left( \begin{bmatrix}
R_{11} & 0 & \cdots & 0 \\
0 & R_{22} & \cdots & 0 \\
\vdots & \vdots & \ddots & \vdots \\
0 & 0 & \cdots & R_{LL}
\end{bmatrix} \begin{bmatrix}
0 & \hat{R}_{12} & \cdots & \hat{R}_{1L} \\
\hat{R}_{12}^H & 0 & \cdots & \hat{R}_{2L} \\
\vdots & \vdots & \ddots & \vdots \\
\hat{R}_{1L}^H & \hat{R}_{2L}^H & \cdots & 0
\end{bmatrix} \right)
\]

\[
= \text{tr} \left( \begin{bmatrix}
0 & R_{11} \hat{R}_{12} & \cdots & R_{11} \hat{R}_{1L} \\
R_{22} \hat{R}_{12}^H & 0 & \cdots & R_{22} \hat{R}_{2L} \\
\vdots & \vdots & \ddots & \vdots \\
R_{LL} \hat{R}_{1L}^H & R_{LL} \hat{R}_{2L}^H & \cdots & 0
\end{bmatrix} \right) = 0
\]

- By simply nulling the off-diagonal blocks of \( \hat{R} \), we’ve effectively solved the optimization problem

\[
\hat{D} = \arg \min_{R \in \mathcal{R}_0} \left\| \hat{R} - R \right\|_F^2
\]
Chi Squared and Beta Random Variables

- \( X = [X_1 \cdots X_n]^T \in \mathcal{D} \Rightarrow \) random vector with joint PDF \( f_X(x) \) over domain \( \mathcal{D} = \{ x : f_X(x) > 0 \} \)
- \( T : \mathcal{D} \to \mathcal{R} \Rightarrow \) a one-to-one, continuously differentiable transformation that maps the domain \( \mathcal{D} \) to the range \( \mathcal{R} \)
- Joint PDF of the random vector \( Y = [Y_1 \cdots Y_n]^T = T(X) \) can be related to \( f_X(x) \)

\[
f_Y(y) = f_X \left[ T^{-1}(y) \right] \left| \det J_{T^{-1}} \right| 1_{\mathcal{R}}(y)
\]

where \( T^{-1} \) denotes the inverse transformation with Jacobian matrix

\[
J_{T^{-1}} = \begin{bmatrix}
\frac{\partial x_1}{\partial y_1} & \frac{\partial x_1}{\partial y_2} & \cdots & \frac{\partial x_1}{\partial y_n} \\
\frac{\partial x_2}{\partial y_1} & \frac{\partial x_2}{\partial y_2} & \cdots & \frac{\partial x_2}{\partial y_n} \\
\vdots & \vdots & \ddots & \vdots \\
\frac{\partial x_n}{\partial y_1} & \frac{\partial x_n}{\partial y_2} & \cdots & \frac{\partial x_n}{\partial y_n}
\end{bmatrix}
\]

and \( 1_{\mathcal{R}}(y) = 1 \) for \( y \in \mathcal{R} \), 0 otherwise.

- Suppose that \( X = [X_1 X_2]^T \) with \( X_1 \sim \chi^2_{\nu_1} \), \( X_2 \sim \chi^2_{\nu_2} \) and \( X_1 \perp X_2 \) with \( \mathcal{D} \in \mathbb{R}_+^2 \) and joint PDF

\[
f_{X_1 X_2}(x_1, x_2) = f_{X_1}(x_1)f_{X_2}(x_2) = \frac{1}{\Gamma \left( \frac{\nu_1}{2} \right) \Gamma \left( \frac{\nu_2}{2} \right)} \frac{\nu_1 + \nu_2}{2} \frac{1}{x_1^{\nu_1/2} x_2^{\nu_2/2}} e^{-\frac{x_1 + x_2}{2}}
\]

- Define the transform/inverse-transform pair

\[
T : Y_1 = X_1 + X_2; \quad Y_2 = \frac{X_1}{X_1 + X_2}
\]

\[
T^{-1} : X_1 = Y_1 Y_2; \quad X_2 = Y_1 (1 - Y_2)
\]

with range \( \mathcal{R} = (0, \infty) \times (0, 1) \) and Jacobian matrix

\[
J_{T^{-1}} = \begin{bmatrix}
\frac{\partial x_1}{\partial y_1} & \frac{\partial x_1}{\partial y_2} \\
\frac{\partial x_2}{\partial y_1} & \frac{\partial x_2}{\partial y_2}
\end{bmatrix} = \begin{bmatrix}
y_2 & y_1 \\
1 - y_2 & -y_1
\end{bmatrix}
\]

\[
\left| \det J_{T^{-1}} \right| = \left| -y_1 y_2 - y_1 (1 - y_2) \right| = y_1
\]
With this, the pair of random variables $Y_1$ and $Y_2$ has joint PDF

$$f_{Y_1 Y_2}(y_1, y_2) = f_{X_1 X_2}(y_1 y_2, y_1(1 - y_2)) y_1 1_{(0, \infty)}(y_1) y_1 1_{(0, 1)}(y_2)$$

$$= \left( \frac{1}{\nu_1 + \nu_2} \right)^{\nu_1 + \nu_2 - 1} \frac{1}{2} y_1^{\frac{\nu_1 + \nu_2}{2}} e^{-\frac{y_1}{2}} 1_{(0, \infty)}(y_1) \left( \frac{1}{\Gamma(\frac{\nu_1}{2}) \Gamma(\frac{\nu_2}{2})} y_2^{\frac{\nu_1}{2} - 1} (1 - y_2)^{\frac{\nu_2}{2} - 1} 1_{(0, 1)}(y_2) \right)$$

- Multiply and divide by $\Gamma(\frac{\nu_1 + \nu_2}{2})$ and use the fact that $B(x, y) = \frac{\Gamma(x) \Gamma(y)}{\Gamma(x + y)}$ to express the joint density as

$$f_{Y_1 Y_2}(y_1, y_2) = f_{Y_1}(y_1) f_{Y_2}(y_2)$$

$$f_{Y_1}(y_1) = \left( \frac{1}{\nu_1 + \nu_2} \right)^{\nu_1 + \nu_2 - 1} \frac{1}{2} y_1^{\frac{\nu_1 + \nu_2}{2}} e^{-\frac{y_1}{2}} 1_{(0, \infty)}(y_1) \right)$$

$$f_{Y_2}(y_2) = \left( \frac{1}{B(\frac{\nu_1}{2}, \frac{\nu_2}{2})} y_2^{\frac{\nu_1}{2} - 1} (1 - y_2)^{\frac{\nu_2}{2} - 1} 1_{(0, 1)}(y_2) \right)$$

- Three results for the price of one

1. $X_1 + X_2 \sim \chi^2_{\nu_1 + \nu_2}$ given the expression for $f_{Y_1}(y_1)$
2. $\frac{X_1}{X_1 + X_2} \sim \text{Beta} \left( \frac{\nu_1}{2}, \frac{\nu_2}{2} \right)$ given the expression for $f_{Y_2}(y_2)$
3. $\frac{X_1}{X_1 + X_2} \perp X_1 + X_2$ given that $f_{Y_1 Y_2}(y_1, y_2) = f_{Y_1}(y_1) f_{Y_2}(y_2)$

- This is but a special case of a more general result: if $X_1 \sim \Gamma(\alpha, \theta), X_2 \sim \Gamma(\beta, \theta)$, and $X_1 \perp X_2$ then $X_1 + X_2 \sim \Gamma(\alpha + \beta, \theta), \frac{X_1}{X_1 + X_2} \sim \text{Beta}(\alpha, \beta)$, and $\frac{X_1}{X_1 + X_2} \perp X_1 + X_2$
An $N \times N$ circulant matrix $C$ is a special Toeplitz matrix consisting of cyclic shifts of the coefficient sequence $\{c_k\}_{k=0}^{N-1}$

$$C = \begin{bmatrix}
c_0 & c_{N-1} & c_{N-2} & \cdots & c_1 \\
c_1 & c_0 & c_{N-1} & \cdots & c_2 \\
c_2 & c_1 & c_0 & \cdots & c_3 \\
& & & \ddots & \\
c_{N-1} & c_{N-2} & c_{N-3} & \cdots & c_0
\end{bmatrix}$$

Any matrix of this form is diagonalized by a DFT matrix $F_N = \left(\frac{1}{\sqrt{N}} e^{-j\frac{2\pi i k}{N}}\right)_{i,k=0}^{N-1}$ so that $F_N C F_N^H = \Lambda$ with $\Lambda = \text{diag}(\lambda_0, \cdots, \lambda_{N-1})$

The set of eigenvalues $\lambda_i$ and the coefficients $c_k$ form a DFT pair

$$\lambda_i = \sum_{k=0}^{N-1} c_k e^{-j\frac{2\pi i k}{N}}$$

$$c_k = \frac{1}{N} \sum_{i=0}^{N-1} \lambda_i e^{j\frac{2\pi i k}{N}}$$

If $C$ is Hermitian then $\lambda_i \in \mathbb{R}$ so that

$$c_{N-k} = \frac{1}{N} \sum_{i=0}^{N-1} \lambda_i e^{j\frac{2\pi i (N-k)}{N}} = \frac{1}{N} \sum_{i=0}^{N-1} \lambda_i e^{j2\pi i} e^{-j\frac{2\pi i k}{N}} = \left(\frac{1}{N} \sum_{i=0}^{N-1} \lambda_i e^{j\frac{2\pi i k}{N}}\right)^* = c_k^*$$

$\Rightarrow$ If $C$ is both Hermitian and circulant, the coefficients must satisfy $c_k = c_{N-k}^*, k = 1, \ldots, N-1$
Circulant Matrices II

Consider finding the most likely estimate of such a matrix, i.e. the matrix \( \hat{C} \) such that

\[
\hat{C} = \arg \min_C \ln \det(C) + \text{tr} \left( C^{-1} \hat{R} \right)
\]

Since \( C = F_N^H \Lambda F_N \), equivalent to finding the diagonal matrix \( \Lambda \) that minimizes \( \ell(\Lambda) \)

\[
\ell(\Lambda) = \ln \det(F_N^H \Lambda F_N) + \text{tr} \left( F_N^H \Lambda^{-1} F_N \hat{R} \right) = \sum_{i=0}^{N-1} \ln \lambda_i + \text{tr} \left( \Lambda^{-1} F_N \hat{R} F_N^H \right) = \sum_{i=0}^{N-1} \ln \lambda_i + \text{tr} \left( \Lambda^{-1} \hat{S} \right)
\]

\( \hat{S} \) is a global "spectral" matrix at frequencies \( \theta_\ell = \frac{2\pi}{N} \ell \) for \( \ell = 0, \ldots, N - 1 \)

\[
\hat{S} = \begin{bmatrix}
\hat{S}(e^{j\theta_0}) & \hat{S}(e^{j\theta_0}, e^{j\theta_1}) & \ldots & \hat{S}(e^{j\theta_0}, e^{j\theta N-1}) \\
\hat{S}(e^{j\theta_1}, e^{j\theta_0}) & \hat{S}(e^{j\theta_1}) & \ldots & \hat{S}(e^{j\theta_1}, e^{j\theta N-1}) \\
\vdots & \vdots & \ddots & \vdots \\
\hat{S}(e^{j\theta N-1}, e^{j\theta_0}) & \hat{S}(e^{j\theta N-1}, e^{j\theta_1}) & \ldots & \hat{S}(e^{j\theta N-1})
\end{bmatrix}
\]

where \( \hat{S}(e^{j\theta_\ell}, e^{j\theta_m}) = f_N^H(e^{j\theta_\ell}) \hat{R} f_N(e^{j\theta_m}) \) and \( \hat{S}(e^{j\theta_\ell}) = \hat{S}(e^{j\theta_\ell}, e^{j\theta_\ell}) \)

With this definition we have that

\[
\ell(\Lambda) = \sum_{i=0}^{N-1} \left( \ln \lambda_i + \frac{\hat{S}(e^{j\theta_i})}{\lambda_i} \right)
\]

\[
\frac{\partial \ell(\Lambda)}{\partial \lambda_i} = \frac{1}{\lambda_i} - \frac{\hat{S}(e^{j\theta})}{\lambda_i^2} = 0 \Rightarrow \hat{\lambda}_i = \hat{S}(e^{j\theta_i})
\]

\[
\frac{\partial^2 \ell(\Lambda)}{\partial \lambda_i^2} \bigg|_{\lambda_i = \hat{\lambda}_i} = \frac{1}{\hat{S}^2(e^{j\theta})} > 0
\]

Therefore, the maximum likelihood estimator is simply \( \hat{C} = F_N^H \text{diag} \left( F_N \hat{R} F_N^H \right) F_N \)

This process of forcing \( \hat{S} \rightarrow \text{diag} \left( \hat{S}(e^{j\theta_0}), \ldots, \hat{S}(e^{j\theta N-1}) \right) \), very similar to the development of broadband coherence
What does this mean for \( \hat{c}_k \)? Apply an inverse DFT to \( \hat{\lambda}_i \)

\[
\hat{c}_k = \frac{1}{N} \sum_{i=0}^{N-1} \hat{\lambda}_i e^{\frac{2\pi i k}{N}} = \frac{1}{N} \sum_{i=0}^{N-1} \left( \frac{1}{N} \sum_{\ell=0}^{N-1} \sum_{m=0}^{N-1} \hat{r}_{\ell m} e^{j (m-\ell) \theta_i} \right) e^{\frac{2\pi i k}{N}}
\]

\[
= \frac{1}{N^2} \sum_{\ell=0}^{N-1} \sum_{m=0}^{N-1} \hat{r}_{\ell m} e^{\frac{2\pi i (k+m-\ell)}{N}}
\]

Use the fact that \( \sum_{i=0}^{N-1} e^{\frac{2\pi i (k+m-\ell)}{N}} = N \) if \( k + m - \ell = 0 \) or \( k + m - \ell = N \), 0 otherwise to get

\[
\hat{c}_k = \frac{1}{N} \sum_{\ell=0}^{N-1} \sum_{m=0}^{N-1} \hat{r}_{\ell m} \left( \delta_{m-(\ell-k)} + \delta_{m-(\ell+N-k)} \right) = \frac{1}{N} \left( \sum_{\ell=0}^{k-1} \hat{r}_{\ell,\ell+N-k} + \sum_{\ell=k}^{N-1} \hat{r}_{\ell,\ell-k} \right)
\]

Fairly straightforward to show that these coefficients do indeed satisfy \( \hat{c}_k = \hat{c}_{N-k}^* \)

Estimator is very similar to the LS estimate of a Toeplitz matrix, the matrix \( \hat{T} = \arg \min_T \| \hat{R} - T \|_F^2 \)

Solution is to simply average down the diagonals of \( \hat{R} \)

\[
\hat{t}_k = \frac{1}{N-k} \sum_{\ell=k}^{N-1} \hat{r}_{\ell,\ell-k}
\]

The only fundamental difference: \( \hat{T} \) uses linear shifts while \( \hat{C} \) uses cyclic shifts

In the multichannel problem, one has to consider block-Toeplitz and block-circulant matrices

Very similar as before, every block-circulant matrix with blocks \( C_k \in \mathbb{C}^{L \times L} \) is block-diagonalized in a DFT basis, i.e. \( (F_N \otimes I_L) C \left( F_N^H \otimes I_L \right) = \text{blkdiag} \{ \Lambda_0, \ldots, \Lambda_{N-1} \} \)
A network of \( L = 3 \) ULA’s, each of \( P = 16 \) sensor elements, observes an autoregressive signal propagated as a planewave.

Here, \( x_i[n, p] \) is the \( n^{th} \) temporal sample collected by the \( p^{th} \) element of the \( i^{th} \) array.

Under \( H_1 \), the signal delayed to each array with a bulk propagation delay, propagated among each array as a planewave and corrupted with independent but colored sensor noise.

Under \( H_0 \), each array consists of independent sensor noise alone.

Data record partitioned into \( M \) copies of a length \( N = 24 \) time series and applied to the GLRT.

Three detection methods applied:
- Generalized Hadamard Ratio (denoted ‘Time Domain GLRT’)
- GLRT for 2D WSS Processes (denoted ‘Frequency/Wavenumber Domain GLRT’)
- GLRT for temporally WSS processes (denoted ‘Frequency Domain GLRT’) which employs the test statistic

\[
\Lambda = \exp \left\{ \int_{-\pi}^{\pi} \ln \frac{\det \hat{S}(e^{j\theta})}{\det \hat{S}_{ii}(e^{j\theta})} \frac{d\theta}{2\pi} \right\}
\]

Although the Time Domain and Frequency Domain GLRTs are more generally applicable, the Frequency/Wavenumber Domain GLRT is better matched to the assumptions of spatiotemporal stationarity.

\( M = 1200, \; SNR = -30 \) dB

\( M = 250, \; SNR = -26 \) dB

\( M = 50, \; SNR = -20 \) dB
Many ways to estimate the power spectrum of a signal, both parametric (e.g. ARMA modeling) and nonparametric (e.g. periodogram)

Trying to determine the spectrum of a random or noisy sequence \(\{x[n]\}_{n=0}^{MN-1}\) using the periodogram

\[
\hat{S}(e^{j\theta}) = \frac{1}{MN} \left| \sum_{n=0}^{MN-1} x[n] e^{-jn\theta} \right|^2
\]
yields an estimator that can suffer from high variance

One very intuitive way of reducing variance \(\Rightarrow\) partition the sequence into \(M\) copies of a length-\(N\) sequence \(x[n, m]\) and average periodograms

\[
\hat{S}(e^{j\theta}) = \frac{1}{M} \sum_{m=1}^{M} \hat{S}(m)(e^{j\theta}) ; \quad \hat{S}(m)(e^{j\theta}) = \frac{1}{N} \left| \sum_{n=0}^{N-1} x[n, m] e^{-jn\theta} \right|^2
\]

Trade-off between decreasing variance and increasing bias due to spectral leakage \(\Rightarrow\) window each sequence with function \(w[n]\) designed to suppress side-lobes

Welch’s method:
- Partition sequence into multiple segments with overlap
- Window each segment
- Average periodograms of each windowed segment

Same principle applied here using a separable Hamming window, i.e. \(w[n, p] = w[n]w[p]\) with \(w[n], w[p]\) two 1D Hamming windows

Steps in estimating the cross-spectral matrix \(\hat{S}(e^{j\theta}, e^{j\phi})\):
- Every ROI is partitioned into \(N \times P\) blocks with \(N = P = 32\) using 50\% overlap in both dimensions
- Each block is windowed with a separable Hamming window
- Apply 2D FFT and average periodograms
Several Images From Each Dataset

Dataset 1

Dataset 2
Day 1

Dataset 2
Day 2

Dataset 2
Day 3
Sequences of Toeplitz Matrices I

- Any $N \times N$ Hermitian Toeplitz matrix, $T_N$, can be described by a set of complex scalars $\{\gamma[k]\}_{k=0}^{N-1}$ or alternatively by its finite-length PSD

$$\tilde{S}(e^{j\theta}) = \sum_{k=-(N-1)}^{N-1} \gamma[k] e^{-j\theta k}$$

where we assume that $\sum_{k=0}^{\infty} |\gamma[k]| < \infty$ and we'll denote this Toeplitz matrix as $T_N(\tilde{S})$.

- Using this PSD, we'll build the circulant matrix $C_N(\tilde{S}) = F_N \hat{\Sigma} F_N^H$ where

$$\hat{\Sigma} = \text{diag}\left(\tilde{S}(e^{j\theta_0}), \ldots, \tilde{S}(e^{j\theta N-1})\right)$$

and is going to assume the form

$$C_N(\tilde{S}) = \begin{bmatrix}
\gamma[1] + \gamma[N-1]^* & \gamma[0] & \gamma[1]^* + \gamma[N-1] & \cdots & \gamma[N-2]^* + \gamma[2] \\
\gamma[2] + \gamma[N-2]^* & \gamma[1] + \gamma[N-1]^* & \gamma[0] & \cdots & \gamma[N-3]^* + \gamma[3] \\
& \ddots & \ddots & \ddots & \ddots \\
\gamma[N-1] + \gamma[1]^* & \gamma[N-2] + \gamma[2]^* & \gamma[N-3] + \gamma[3]^* & \cdots & \gamma[0]
\end{bmatrix}$$

- Two $N \times N$ matrices are asymptotically equivalent if $\lim_{N \to \infty} ||A_N - B_N|| = 0$ where we use the Hilbert-Schmidt norm

$$||A_N||^2 = \frac{1}{N} \sum_{i=0}^{N-1} \sum_{k=0}^{N-1} |[A_N]_{i,k}|^2$$
Sequences of Toeplitz Matrices II

- If $A_N$ and $B_N$ are two Hermitian matrices with eigenvalues $\{\lambda_i(A_N)\}$ and $\{\lambda_i(B_N)\}$, respectively, then asymptotic equivalence also implies that, for any continuous function $f$,

$$
\lim_{N \to \infty} \frac{1}{N} \sum_{i=0}^{N-1} f(\lambda_i(A_N)) = \lim_{N \to \infty} \frac{1}{N} \sum_{i=0}^{N-1} f(\lambda_i(B_N))
$$

- Assuming absolute summability of the Toeplitz coefficients, it can be shown that $T_N \left( \tilde{S} \right)$ and $C_N \left( \tilde{S} \right)$ are indeed asymptotically equivalent so that

$$
\lim_{N \to \infty} \ln \det \left( T_N \left( \tilde{S} \right) \right)^{\frac{1}{N}} = \lim_{N \to \infty} \frac{1}{N} \sum_{i=0}^{N-1} \ln \lambda_i(T_N) = \frac{1}{2\pi} \lim_{N \to \infty} \sum_{i=0}^{N-1} \ln \tilde{S} \left( e^{j\theta_i} \right) \Delta \theta
$$

$$
= \int_{-\pi}^{\pi} \ln S \left( e^{j\theta} \right) \frac{d\theta}{2\pi}
$$

- So the linear transformation $F_N$ asymptotically diagonalizes the matrix $T_N$ with eigenvalues $\{S \left( e^{j\theta_i} \right)\}_{i=0}^{N-1}$.

- Given a matrix $T_N \left( \tilde{S} \right)$ which is now block Toeplitz, it is really just a matter of formality to show that it is asymptotically equivalent to a block circulant matrix and that

$$
\lim_{N \to \infty} \ln \det \left( T_N \left( \tilde{S} \right) \right)^{\frac{1}{N}} = \int_{-\pi}^{\pi} \ln \det \left( S \left( e^{j\theta} \right) \right) \frac{d\theta}{2\pi}
$$

- So, again, the linear transformation $F_N \otimes I_L$ asymptotically block-diagonalizes matrix $T_N \left( \tilde{S} \right)$ with diagonal block $\{S \left( e^{j\theta_i} \right)\}_{i=0}^{N-1}$.