Matrix Formulation of Visibility Statistics

Ludwig Schwardt

SKA South Africa (MeerKAT Project)

Calibration & Imaging Workshop
ASTRON, Netherlands
26 August 2010

The Mythical Thermal Noise

- "We want to integrate down to the thermal noise level"
- The ideal radiometer equation gives this level as

$$\sigma_{\mathrm{P}} \propto \frac{P}{\sqrt{K}} \quad \text{(Dicke, 1946)},$$

  where P is expected power and K is number of measurements (time-bandwidth product $B\tau$ for continuous-time signals)
- This is actually the *(standard) RMS error of the sample variance estimator* $\hat{\mathrm{P}} = \frac{1}{K}\sum_{k=1}^{K} x_k^2$ — size of statistical fluctuation of power measurement (Dicke)
- Purely a result of measuring on a finite sample

Studies of Visibility Statistics

Kulkarni (1989)

- *Self-noise:* Strong sources show same statistical fluctuations as strong receiver noise — it's the total received power that matters in radiometer equation
- For strong sources, visibilities are correlated (especially on baselines with shared antenna)
- Statistics of images and triple products

Gwinn (2001, 2004, 2006)

- PDF for product of correlated Gaussian variables (scalar)
- Effect of quantisation on this PDF

It's All About Covariance Matrices

- Describe instantaneous real samples from N antennas by N-dimensional zero-mean Gaussian random variable (RV) $\boldsymbol{x} \sim \mathcal{N}_N(\mathbf{0}, \boldsymbol{R})$ with *(true) covariance matrix* $\boldsymbol{R}$ and PDF

$$p(\boldsymbol{x}) = \frac{1}{(2\pi)^{-N/2}|\boldsymbol{R}|^{1/2}} \exp\left[-\frac{1}{2}\boldsymbol{x}^H\boldsymbol{R}^{-1}\boldsymbol{x}\right]$$

- Correlator calculates *sample covariance matrix*

$$\boldsymbol{S} \triangleq \frac{1}{K}\sum_{k=1}^{K}\boldsymbol{x}_k\boldsymbol{x}_k^T$$

  by multiplying and adding K of these vectors per dump

- Can also define unnormalised *scatter matrix*

$$\boldsymbol{W} \triangleq K\boldsymbol{S} = \sum_{k=1}^{K}\boldsymbol{x}_k\boldsymbol{x}_k^T$$

Sample Covariance Matrix

- Visibilities are not independent measurements, but elements of a matrix

$$\boldsymbol{S} = \begin{bmatrix} s_{11} & s_{12} & s_{13} & \cdots & s_{1N} \\ s_{21} & s_{22} & s_{23} & \cdots & s_{2N} \\ s_{31} & s_{32} & s_{33} & \cdots & s_{3N} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ s_{N1} & s_{N2} & s_{N3} & \cdots & s_{NN} \end{bmatrix}$$

  (measured visibilities in blue, autocorrelations in red)
- $\boldsymbol{S}$ has *structure* (symmetric, positive semi-definite)
- Instead of statistics of individual visibilities, rather consider statistics of $\boldsymbol{S}$ (or $\boldsymbol{W}$) as a whole

The Real Wishart Distribution

- It's old news... (Wishart, 1928)
- $\boldsymbol{W} \sim \mathcal{W}_N(K, \boldsymbol{R})$ has *real Wishart distribution*, with K degrees of freedom, scale matrix $\boldsymbol{R}$ and PDF

$$p(\boldsymbol{W}) = \frac{|\boldsymbol{W}|^{(K-N-1)/2}}{2^{NK/2}\Gamma_N(K/2)|\boldsymbol{R}|^{K/2}} \exp\left[-\frac{1}{2}\operatorname{tr}(\boldsymbol{R}^{-1}\boldsymbol{W})\right],$$

  where

$$\Gamma_N(K/2) = \pi^{N(N-1)/4} \prod_{i=1}^{N} \Gamma\left[(K-i+1)/2\right]$$

  is *multivariate gamma function*
- Wishart is *sampling distribution* of covariance matrix, just like χ^2 is sampling distribution of scalar variance

When Does This Matter?

Two important limits for radio astronomy:

- **The large-sample limit:** $K \gg 1$
  Wishart becomes Gaussian, i.e. $\operatorname{vec} \boldsymbol{S} \sim \mathcal{N}_N(\operatorname{vec} \boldsymbol{R}, \boldsymbol{R}_R)$
  with covariance between visibilities $\boldsymbol{R}_R \propto 1/K$
- **The weak-source limit:** $\boldsymbol{R} \approx$ diagonal
  Visibilities become uncorrelated
- "Thermal noise" is usually uncorrelated and Gaussian…

Wishart therefore more relevant for smaller K (fast dump rates, narrow-band channels) and stronger sources — low-frequency instruments?

The Bartlett Decomposition

- More efficient way to generate random matrices from Wishart distribution than brute force via Gaussian RVs
- If $\boldsymbol{W} \sim \mathcal{W}_N(\boldsymbol{R}, K)$, then $\boldsymbol{W} = \boldsymbol{L}\boldsymbol{T}\boldsymbol{T}^T\boldsymbol{L}^T$
- $\boldsymbol{L}$ is lower triangular matrix so that $\boldsymbol{R} = \boldsymbol{L}\boldsymbol{L}^T$ (*Cholesky decomposition* of $\boldsymbol{R}$)
- $\boldsymbol{T}$ is random lower triangular matrix with elements

$$\boldsymbol{T} = \begin{bmatrix} \sqrt{c_1} & 0 & 0 & \cdots & 0 \\ n_{21} & \sqrt{c_2} & 0 & \cdots & 0 \\ n_{31} & n_{32} & \sqrt{c_3} & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ n_{N1} & n_{N2} & n_{N3} & \cdots & \sqrt{c_N} \end{bmatrix}$$

where $c_i \sim \chi^2_{K-i+1}$ and $n_{ij} \sim \mathcal{N}(0, 1)$

From Real to Complex

- Classical optical coherence theory represents real field by complex analytic signal to simplify the maths
- View N-dimensional complex RV $\boldsymbol{x} = \operatorname{Re}\boldsymbol{x} + j\operatorname{Im}\boldsymbol{x}$ as $2N$-dimensional real RV $\boldsymbol{y} = [\operatorname{Re}\boldsymbol{x}, \operatorname{Im}\boldsymbol{x}]^T$
- Real covariance matrix $\mathrm{E}\left[\boldsymbol{y}\boldsymbol{y}^H\right]$ has $4N^2$ real elements, but complex covariance matrix $\boldsymbol{R} = \mathrm{E}\left[\boldsymbol{x}\boldsymbol{x}^H\right]$ only has $2N^2$ real elements — $\boldsymbol{y}$ is more general than $\boldsymbol{x}$!
- The missing elements are in the *complementary (pseudo) covariance matrix* (also known as *relation matrix*)

$$\boldsymbol{C} \triangleq \mathrm{E}\left[\boldsymbol{x}\boldsymbol{x}^T\right]$$

- $\boldsymbol{C}$ is $N \times N$, complex and *symmetric* (not Hermitian)

Proper Complex Random Vectors

- Complex RV $\boldsymbol{x}$ is *proper* (or *circular*) if $\boldsymbol{C} = \boldsymbol{0}$ — all second-order information contained in $\boldsymbol{R}$
- An equivalent definition is that real covariance matrix have the constrained form

$$\mathrm{E}\left[\boldsymbol{y}\boldsymbol{y}^{H}\right] = \begin{bmatrix} \operatorname{Re}\boldsymbol{R} & -\operatorname{Im}\boldsymbol{R} \\ \operatorname{Im}\boldsymbol{R} & \operatorname{Re}\boldsymbol{R} \end{bmatrix}$$

- Yet another way to state it:

$$\begin{aligned} \mathrm{E}\left[\operatorname{Re}\boldsymbol{x}\operatorname{Re}\boldsymbol{x}^{T}\right] &= \mathrm{E}\left[\operatorname{Im}\boldsymbol{x}\operatorname{Im}\boldsymbol{x}^{T}\right] \\ \mathrm{E}\left[\operatorname{Re}\boldsymbol{x}\operatorname{Im}\boldsymbol{x}^{T}\right] &= -\mathrm{E}\left[\operatorname{Re}\boldsymbol{x}\operatorname{Im}\boldsymbol{x}^{T}\right]^{T} \end{aligned}$$

- For each element of $\boldsymbol{x}$, real and imaginary parts have same variance and are uncorrelated

Proper vs Improper

Let $x(n)$ be a discrete-time real stationary bandpass signal.

The following complex signals are *proper*:

- Complex analytic representation of $x(n)$ (vCZ is OK!)
- Complex envelope of $x(n)$
- Discrete-time Fourier transform of $x(n)$

The following complex signals are *improper*:

- $x(n)$ itself (because $\boldsymbol{C} = \boldsymbol{R} \neq \mathbf{0}$)
- FFT of length-N vector $\boldsymbol{x}$ taken from $x(n)$ (*asymptotically* proper as $N \to \infty$ and FFT $\to$ DTFT)

FX Correlator Example

- F-step does FFT on blocks of $2N$ real voltage samples, producing N complex spectral samples per block
- Spectral channels 0 and N are always real and therefore improper — another reason to discard these…
- The rest of the channels are asymptotically proper as $M \to \infty$ — check propriety for small M though
- Polyphase filterbank in front of FFT does not change this

Complex Gaussian Distribution

- The N-dimensional RV $\boldsymbol{x} \sim \mathcal{N}_N^{\mathbb{C}}(\boldsymbol{m}, \boldsymbol{R})$ has a *complex Gaussian distribution* if the corresponding $2N$-dimensional RV $\boldsymbol{y}$ is Gaussian and $\boldsymbol{x}$ is proper
- PDF is straightforward extension of real one:

$$p(\boldsymbol{x}) = \frac{1}{\pi^N |\boldsymbol{R}|} \exp\left[-(\boldsymbol{x} - \boldsymbol{m})^H \boldsymbol{R}^{-1} (\boldsymbol{x} - \boldsymbol{m})\right]$$

  (only valid if $\boldsymbol{x}$ is proper!)
- Has maximum entropy among distributions with same $\boldsymbol{m}$ and $\boldsymbol{R}$

The Complex Wishart Distribution

- Let $\boldsymbol{x} \sim \mathcal{N}_N^{\mathbb{C}}(\mathbf{0}, \boldsymbol{R})$ and form scatter matrix $\boldsymbol{W} = \sum_{k=1}^{K} \boldsymbol{x}_k \boldsymbol{x}_k^H$
- $\boldsymbol{W} \sim \mathcal{W}_N^{\mathbb{C}}(K, \boldsymbol{R})$ has *complex Wishart distribution* with PDF

$$p(\boldsymbol{W}) = \frac{|\boldsymbol{W}|^{K-N}}{\tilde{\Gamma}_N(K)|\boldsymbol{R}|^K} \exp\left[-\operatorname{tr}(\boldsymbol{R}^{-1}\boldsymbol{W})\right],$$

where

$$\tilde{\Gamma}_N(K) = \pi^{N(N-1)/2} \prod_{i=1}^{N} \Gamma(K - i + 1)$$

is *complex multivariate gamma function*

- Similar properties to real Wishart, e.g. if $\boldsymbol{W} \sim \mathcal{W}_N^{\mathbb{C}}(K, \boldsymbol{R})$, then $\boldsymbol{AWA}^H \sim \mathcal{W}_N^{\mathbb{C}}(K, \boldsymbol{ARA}^H)$ for any full-rank matrix $\boldsymbol{A}$
- Even nicer: recursive formula to write down moments of any order (Letac & Massam, 2004)

Some Moments of Complex Wishart

- Let elements of $\boldsymbol{S} = \frac{1}{K}\boldsymbol{W}$ be s_{ij} and those of $\boldsymbol{R}$ be r_{ij}
- Mean visibility: $\mathrm{E}\left[s_{ij}\right] = r_{ij}$
- Correlation between arbitrary visibilities:

$$\mathrm{E}\left[s_{ij}{s_{kl}}^*\right] = r_{ij}{r_{kl}}^* + \frac{1}{K}r_{ik}{r_{jl}}^*$$

- Covariance between arbitrary visibilities:

$$\mathrm{E}\left[(s_{ij} - r_{ij})(s_{kl} - r_{kl})^*\right] = \frac{1}{K}r_{ik}{r_{jl}}^*$$

- Variance of visibility:

$$\mathrm{E}\left[|s_{ij} - r_{ij}|^2\right] = \frac{1}{K}r_{ii}r_{jj}$$

$\rightarrow$ Thermal noise level!

Closure Statistics

- Closure phase is angle of *triple product* $s_{ij}s_{jk}s_{ki}$
- Closure amplitude is $\left|(s_{ij}s_{kl})/(s_{ik}s_{jl})\right|$
- Wishart moments are useful to characterise these
- E.g. Mean of triple product (autocorrelations in red):

$$\begin{aligned} \mathrm{E}\left[s_{ij}s_{jk}s_{ki}\right] &= r_{ij}r_{jk}r_{ki} + \frac{1}{K}\left(r_{ii}r_{jk}r_{kj} + r_{jj}r_{ik}r_{ki} + r_{kk}r_{ij}r_{ji}\right) + \\ & \quad \frac{1}{K^2}\left(r_{ii}r_{jj}r_{kk} + r_{ji}r_{kj}r_{ik}\right) \end{aligned}$$

- Variance of triple product straightforward but tedious to write down…

Applications

- Good theoretical model
- Fancy thermal noise generator (full polarisation!)
- Detect deviation from visibility model via likelihood test
- MAP estimation: we have $p(\boldsymbol{S}|\boldsymbol{R})$, turn it around to form $p(\boldsymbol{R}|\boldsymbol{S})$ via $P(\boldsymbol{R})$, for which convenient conjugate prior is available

Appendix: Wishart Generator

```python
import numpy as np
import scipy.stats

def complex_wishart(R, df, size=1):
    """Generate random matrices from complex Wishart distribution.
    R -- True covariance matrix, of shape (dim, dim)
    df -- Degrees of freedom
    size -- Number of random matrices to generate
    Returns sequence of random matrices as complex array S, of shape (size, dim, dim)
    """
    # Obtain dimension of covariance matrix
    dim = R.shape[0]
    # Pre-allocate the sequence of output matrices
    S = np.zeros((size, dim, dim), dtype=np.complex128)
    # Do Cholesky decomposition of covariance matrix
    L = np.linalg.cholesky(R)
    # Load diagonal elements with square root of chi-square variates (a.k.a. chi variates)
    for n in xrange(dim):
        S[:, n, n] = scipy.stats.chi.rvs(2 * (df - n), size=size)
    # Obtain all lower triangle indices
    lower = np.tri(dim, k=-1).ravel() > 0
    nlow = len(lower)
    for m in xrange(size):
        # Load lower triangle with standard complex Gaussian RVs
        S[m].ravel()[lower] = np.random.randn(nlow) + 1.0j * np.random.randn(nlow)
        # Transform Bartlett decomposition factor by Cholesky factor
        LT = np.dot(L, S[m])
        # Finally form Hermitian matrix
        S[m] = 0.5 * np.dot(LT, LT.conj().transpose())
    return S
```