THE PTA LIKELIHOOD:

Practical Usage of the Optimal *Cross-Correlation* Statistic

Joe Simon, UC Boulder

July 19, 2022

PTA GW Summer School - Vanderbilt University

NANOGrav Physics Frontiers Center

NSF

Quick Recap of the PTA Likelihood

$$\vec{\delta t} = M\vec{\epsilon} + \boxed{F\vec{a}} + U\vec{j} + \vec{n}$$

Red Noise Processes

$$\vec{\delta t}_{\mathrm{red}} \longrightarrow \langle \vec{\delta t}_{\mathrm{red}} \vec{\delta t}_{\mathrm{red}}^{\mathrm{T}} \rangle = \mathbf{F} \langle \vec{a}\,\vec{a}^{\mathrm{T}} \rangle \mathbf{F}^{\mathrm{T}} = \mathbf{F} \boxed{\phi} \mathbf{F}^{\mathrm{T}}$$

$$[\phi]_{ab} = \Gamma_{ab}\rho + \kappa_a \delta_{ab}$$

Overlap Reduction Function

Intrinsic Red Noise PSD

Correlated Signal (aka GWB) PSD

Quick Recap of the PTA Likelihood

$$C = \begin{pmatrix} \langle \delta t_1 \delta t_1^T \rangle & \langle \delta t_1 \delta t_2^T \rangle & \dots & \langle \delta t_1 \delta t_N^T \rangle \\ \langle \delta t_2 \delta t_1^T \rangle & \langle \delta t_2 \delta t_2^T \rangle & \dots & \langle \delta t_2 \delta t_N^T \rangle \\ \vdots & \vdots & \ddots & \vdots \\ \langle \delta t_N \delta t_1^T \rangle & \langle \delta t_N \delta t_2^T \rangle & \dots & \langle \delta t_N \delta t_N^T \rangle \end{pmatrix} = C_a + S_{ab}\,|_{a \neq b}$$

Auto-covariance
(WN + intrinsic RN)

Cross-covariance
(correlated signals aka GWB)

$$S_{ab} = F_a \, \Gamma_{ab} \, \rho(f) \, F_b^T \qquad \rho(f) = \frac{A_{\mathrm{corr}}^2}{12\pi^2} \left(\frac{f}{f_{\mathrm{yr}}} \right)^{-\gamma}$$

$\gamma = 13/3$
for SMBHBs

Constructing An "Optimal" Detection Statistic

From a frequentist standpoint, an "optimal" detection statistic is one that minimizes the false alarm probabilities at fixed false detection probabilities.

Allen, Creighton et al. (2001)

Two Different (Yet Equivalent Approaches):

1) The OS as a 2-point correlation statistic.

2) The OS as a log-likelihood ratio.

Constructing An "Optimal" Detection Statistic

<u>Some References:</u>

Anholm, et al., PRD 79 (2009) - *Optimal strategies for GW stochastic background searches in pulsar timing data*

Chamberlain, et al., PRD 91 (2015) - *Time-domain implementation of the optimal cross-correlation statistic for stochastic gravitational-wave background searches in pulsar timing data*

Vigeland, et al., PRD 98 (2018) - *Noise-marginalized optimal statistic: A robust hybrid frequentist-Bayesian statistic for the stochastic gravitational-wave background in pulsar timing arrays*

Constructing An "Optimal" Detection Statistic

Directly Fitting The Cross-Correlated Power

The optimal statistic $\hat{A}$ can by found by minimizing the following chi-squared:

$$\chi^2 = \sum_{ab} \left| \frac{\xi_{ab} - A^2_{\mathrm{corr}}\Gamma_{ab}}{\sigma_{ab}} \right|^2$$

Overlap Reduction Function

Correlated power in the residuals of a pulsar pair

Uncertainty in the correlated power

Amplitude of the correlated signal

Constructing An "Optimal" Detection Statistic

Directly Fitting The Cross-Correlated Power

The optimal statistic $\hat{A}$ can by found by minimizing the following chi-squared:

$$\chi^2 = \sum_{ab} \left| \frac{\xi_{ab} - A^2_{\mathrm{corr}}\Gamma_{ab}}{\sigma_{ab}} \right|^2$$

$$\xi_{ab} = \frac{\overrightarrow{\delta t_a^T} C_a^{-1} \hat{S}_{ab} C_b^{-1} \overrightarrow{\delta t_b}}{\mathrm{tr}[C_a^{-1}\hat{S}_{ab}C_b^{-1}\hat{S}_{ab}]}$$

$$\sigma_{ab} = (\mathrm{tr}[C_a^{-1}\hat{S}_{ab}C_b^{-1}\hat{S}_{ab}])^{-1/2}$$

$$A^2_{\mathrm{corr}}\Gamma_{ab}\hat{S}_{ab} = S_{ab}$$

Constructing An "Optimal" Detection Statistic

Directly Fitting The Cross-Correlated Power

The optimal statistic $\hat{A}$ can by found by fitting the correlated power, assuming some particular cross-correlation coefficients, and minimizing the resulting chi-squared.

$$\hat{A}^2 = \frac{\sum_{ab} \overrightarrow{\delta t_a^T} C_a^{-1} \tilde{S}_{ab} C_b^{-1} \overrightarrow{\delta t_b}}{\sum_{ab} \mathrm{tr}[C_a^{-1} \tilde{S}_{ab} C_b^{-1} \tilde{S}_{ab}]}, \; A^2_{\mathrm{corr}} \tilde{S}_{ab} = S_{ab}$$

This definition ensures that $\langle \hat{A}^2 \rangle = A^2_{\mathrm{corr}}$

Constructing An "Optimal" Detection Statistic

Directly Fitting The Cross-Correlated Power

$$\hat{A}^2 = \frac{\sum_{ab} \overrightarrow{\delta t_a^T} C_a^{-1} \tilde{S}_{ab} C_b^{-1} \overrightarrow{\delta t_b}}{\sum_{ab} \mathrm{tr}[C_a^{-1} \tilde{S}_{ab} C_b^{-1} \tilde{S}_{ab}]}$$

If $A_{\mathrm{corr}} = 0$, the variance of $\hat{A}^2$ is

$$\sigma_0 = \left[\sum_{ab} \mathrm{tr}\left(C_a^{-1} \tilde{S}_{ab} C_b^{-1} \tilde{S}_{ab} \right) \right]^{-1/2}$$

Constructing An "Optimal" Detection Statistic

Directly Fitting The Cross-Correlated Power

$$\hat{A}^2 = \frac{\sum_{ab} \overrightarrow{\delta t_a^T} C_a^{-1} \tilde{S}_{ab} C_b^{-1} \overrightarrow{\delta t_b}}{\sum_{ab} \mathrm{tr}[C_a^{-1} \tilde{S}_{ab} C_b^{-1} \tilde{S}_{ab}]}$$

$$\sigma_0 = \left[\sum_{ab} \mathrm{tr} \left(C_a^{-1} \tilde{S}_{ab} C_b^{-1} \tilde{S}_{ab} \right) \right]^{-1/2}$$

The SNR for the power in the cross-correlations is

$$\mathrm{SNR} = \frac{\hat{A}^2}{\sigma_0} = \frac{\sum_{ab} \overrightarrow{\delta t_a^T} C_a^{-1} \tilde{S}_{ab} C_b^{-1} \overrightarrow{\delta t_b}}{\left(\sum_{ab} \mathrm{tr}[C_a^{-1} \tilde{S}_{ab} C_b^{-1} \tilde{S}_{ab}] \right)^{1/2}}$$

Constructing An "Optimal" Detection Statistic

as a log-likelihood ratio

Returning to the calculation of the PTA log-likelihood $\ln \mathscr{L}$,
one can expand the covariance matrix C in a Taylor series expansion
in the ORF coefficients Γ_{ab} to obtain a "first order" likelihood function.

Ellis et al (2013)

$$\ln \mathscr{L} \approx -\frac{1}{2}\left[\sum_{a}\left(\mathrm{tr}\ \ln\ C_a + \overrightarrow{\delta t_a^T} C_a^{-1} \overrightarrow{\delta t_a}\right) - \sum_{ab} \overrightarrow{\delta t_a^T} C_a^{-1} S_{ab} C_b^{-1} \overrightarrow{\delta t_b}\right]$$

Constructing An "Optimal" Detection Statistic

as a log-likelihood ratio

Now we consider the log-likelihood ratio $\ln \Lambda$,
between a model that includes a *spatially correlated* process
and one with only *spatially uncorrelated* noise components ($\Gamma_{ab} = 0$).

$$\ln \Lambda = -\frac{1}{2}\left[\sum_a \left(\mathrm{tr}\,\ln\, C_a + \overrightarrow{\delta t_a^T} C_a^{-1}\overrightarrow{\delta t_a}\right) - \sum_{ab} \overrightarrow{\delta t_a^T} C_a^{-1} S_{ab} C_b^{-1} \overrightarrow{\delta t_b}\right]$$

$$+\frac{1}{2}\sum_a \left(\mathrm{tr}\,\ln\, C_a + \overrightarrow{\delta t_a^T} C_a^{-1}\overrightarrow{\delta t_a}\right)$$

$$= \frac{A_{\mathrm{corr}}^2}{2}\sum_{ab} \overrightarrow{\delta t_a^T} C_a^{-1} \tilde{S}_{ab} C_b^{-1} \overrightarrow{\delta t_b} \qquad A_{\mathrm{corr}}^2 \tilde{S}_{ab} = S_{ab}$$

Constructing An "Optimal" Detection Statistic

as a log-likelihood ratio

From this log-likelihood ratio $\ln \Lambda$, we can again define the optimal statistic $\hat{A}$.

$$\ln \Lambda = \frac{A_{\mathrm{corr}}^2}{2} \sum_{ab} \overrightarrow{\delta t_a^T} C_a^{-1} \tilde{S}_{ab} C_b^{-1} \overrightarrow{\delta t_b}$$

$$\hat{A}^2 = \mathcal{N} \sum_{ab} \overrightarrow{\delta t_a^T} C_a^{-1} \tilde{S}_{ab} C_b^{-1} \overrightarrow{\delta t_b}$$

Constructing An "Optimal" Detection Statistic

as a log-likelihood ratio

From this log-likelihood ratio $\ln\ \Lambda$, we can again define the optimal statistic $\hat{A}$, using a normalization factor, $\mathcal{N}$, which we choose so that on average $\langle \hat{A}^2 \rangle = A^2_{\mathrm{corr}}$.

$$\hat{A}^2 = \mathcal{N} \sum_{ab} \vec{\delta t}_a^T C_a^{-1} \tilde{S}_{ab} C_b^{-1} \vec{\delta t}_b$$

$$\langle \hat{A}^2 \rangle = \langle \mathcal{N} \sum_{ab} \vec{\delta t}_a^T C_a^{-1} \tilde{S}_{ab} C_b^{-1} \vec{\delta t}_b \rangle = \mathcal{N} \sum_{ab} C_a^{-1} \tilde{S}_{ab} C_b^{-1} \langle \vec{\delta t}_a^T \vec{\delta t}_b \rangle$$

$$= \mathcal{N} \sum_{ab} C_a^{-1} \tilde{S}_{ab} C_b^{-1} S_{ab} = \mathcal{N} A^2_{\mathrm{corr}} \sum_{ab} C_a^{-1} \tilde{S}_{ab} C_b^{-1} \tilde{S}_{ab}$$

$$\mathcal{N} = \left[\sum_{ab} C_a^{-1} \tilde{S}_{ab} C_b^{-1} \tilde{S}_{ab} \right]^{-1}$$

Constructing An "Optimal" Detection Statistic

Two Different (Yet Equivalent Approaches):

1) The OS as a 2-point correlation statistic.

2) The OS as a log-likelihood ratio.

$$\hat{A}^2 = \frac{\sum_{ab} \overrightarrow{\delta t_a}^T C_a^{-1} \tilde{S}_{ab} C_b^{-1} \overrightarrow{\delta t_b}}{\sum_{ab} \mathrm{tr}[C_a^{-1} \tilde{S}_{ab} C_b^{-1} \tilde{S}_{ab}]}$$

The Optimal Statistic In Practice

The Optimal Statistic In Practice

Assets:

- Computationally inexpensive

- Incredible flexibility
  - easy to dropout pulsars and/or switch noise models.

- Isolates the spatial correlations
  - easy to swap out different ORFs

- Easy to calculate null distributions
  - sky scrambles / phase shifts run quick and easily

Limitations:

- Poor parameter estimation
  - does not give a probability distribution (e.g. posterior)

- Can give biased results
  - especially if not careful with handling of intrinsic RN values

Frequentist analyses give sensible answers when they closely approximate the Bayesian approach.

- Neil Cornish's talk last Thursday