7 Examples of frequentist statistics in gravitational wave astronomy

In this section we will describe some of the applications of frequentist statistical methods to gravitational wave detection. Fundamental to frequentist statistics is the likelihood. As described in the previous chapter, for gravitational wave detectors, we assume that the output of the detector, $s(t)$, is a linear combination of a signal, $h(t|\vec{\lambda})$, determined by a finite set of (unknown) parameters, $\vec{\lambda}$, and instrumental noise, $n(t)$. We assume in addition that the noise is Gaussian with a (usually known) power spectral density $S_h(f)$

$$s(t) = n(t) + h(t|\vec{\lambda}), \qquad \langle \tilde{n}^*(f)\tilde{n}(f')\rangle = S_h(f)\delta(f-f').$$

The signal is deterministic, but the noise is a random process. The likelihood, for parameters $\vec{\lambda}$, is therefore the probability that the observed noise realisation would take the value $n(t) = s(t) - h(t|\vec{\lambda})$, which can be seen to be

$$\mathcal{L}(s|\vec{\lambda}) = p(n(t) = s(t) - h(t|\vec{\lambda})) \propto \exp\left[-\frac{1}{2}(s - h(\vec{\lambda})|s - h(\vec{\lambda}))\right] \tag{100}$$

where the noise weighted overlap is as given in the last lecture

$$(a|b) = \int_{-\infty}^{\infty} \frac{\tilde{a}^*(f)\tilde{b}(f) + \tilde{a}(f)\tilde{b}^*(f)}{S_h(f)} \mathrm{d}f.$$

7.1 The Fisher Matrix

We introduced the Fisher Matrix in the discussion of the Cramer-Rao bound on the variance of an estimator, which, for a multivariate unbiased estimator, $\hat{\lambda}$, is given by

$$\mathrm{cov}(\hat{\lambda}_i, \hat{\lambda}_j) \geq [\mathbf{\Gamma}_\lambda]^{-1}_{ij}$$

where

$$(\mathbf{\Gamma}_\lambda)_{ij} = \mathbb{E}\left[\frac{\partial l}{\partial \lambda_i}\frac{\partial l}{\partial \lambda_j}\right].$$

In the above l denotes the log-likelihood. For the gravitational wave log-likelihood in Eq. (100), the derivative is

$$\frac{\partial l}{\partial \lambda_i} = \left(\frac{\partial h}{\partial \lambda_i}\middle| s - h(\vec{\lambda})\right) = \left(\frac{\partial h}{\partial \lambda_i}\middle| \mathbf{n}\right).$$

It therefore follows, from the result given in Eq. (93), that

$$(\mathbf{\Gamma}_\lambda)_{ij} = \mathbb{E}\left[\frac{\partial l}{\partial \lambda_i}\frac{\partial l}{\partial \lambda_j}\right] = \left\langle \left(\frac{\partial h}{\partial \lambda_i}\middle| \mathbf{n}\right)\left(\frac{\partial h}{\partial \lambda_i}\middle| \mathbf{n}\right)\right\rangle = \left(\frac{\partial h}{\partial \lambda_i}\middle| \frac{\partial h}{\partial \lambda_j}\right).$$

The Fisher Matrix gives a lower bound on the variance of any unbiased estimator of the parameters of the signal, and hence it provides a guide to how accurately the parameters can be measured. We know that the maximum likelihood estimator is asymptotically efficient, i.e., it achieves this Fisher Matrix bound, which is why it might be expected to provide a

good guide to parameter measurement precision. However, asymptotic efficiency refers to making many repeated measurements of the same parameter, which we do not typically do in gravitational wave observations. But it can be seen that the Fisher Matrix provides a good guide to measurement precision even for a single observation, as follows. We suppose that the true parameters of the signal are given by $\vec{\lambda}_0$, and expand to leading order about those parameters

$$\vec{\lambda} = \vec{\lambda}_0 + \Delta\vec{\lambda}, \qquad h(t|\vec{\lambda}) = h(t|\vec{\lambda}) + \partial_i h(t|\vec{\lambda})\Delta\lambda^i$$

where ∂_i denotes the derivative with respect to λ_i and the last term employs Einstein summation convention. This approximation is known as the **linear signal approximation**. The likelihood can then be expanded as

$$\begin{aligned}
\mathcal{L}(s|\vec{\lambda}) &\propto \exp\left[-\frac{1}{2}(n - \partial_i h(t|\vec{\lambda})\Delta\lambda^i | n - \partial_j h(t|\vec{\lambda})\Delta\lambda^j)\right] \\
&= \exp\left\{-\frac{1}{2}\left[(n|n) - 2(n|\partial_i h(t|\vec{\lambda}))\Delta\lambda^i + (\partial_i h(t|\vec{\lambda})|\partial_j h(t|\vec{\lambda}))\Delta\lambda^i\Delta\lambda^j\right]\right\} \\
&= \exp\left[-\frac{1}{2}(n|n)\right] \exp\left[-\frac{1}{2}\left(\Delta\lambda^i - (\Gamma^{-1})_{ik}(n|\partial_k h(t|\vec{\lambda}))\right)\Gamma_{ij}\left(\Delta\lambda^j - (\Gamma^{-1})_{jl}(n|\partial_l h(t|\vec{\lambda}))\right)\right] \\
&\qquad \times \exp\left[-\frac{1}{2}(n|\partial_i h(t|\vec{\lambda}))(\Gamma^{-1})_{ij}(n|\partial_j h(t|\vec{\lambda}))\right]. \qquad (101)
\end{aligned}$$

The latter term is sub-dominant since it is $O(1)$ compare to the middle term which is of order of the signal amplitude, or SNR. The middle term is a Gaussian, centred at $\Delta\lambda^i = (\Gamma^{-1})_{ik}(n|\partial_k h(t|\vec{\lambda}))$, and with covariance matrix given by the Fisher Matrix. The latter therefore provides an estimate of the width of the likelihood distribution and hence can be used as a guide to the uncertainty. In addition, the maximum likelihood estimator

$$\widehat{\Delta\lambda}^i = (\Gamma^{-1})_{ik}(n|\partial_k h(t|\vec{\lambda}))$$

has mean and variance

$$\mathbb{E}\left(\widehat{\Delta\lambda}^i\right) = 0, \qquad \mathrm{cov}\left(\widehat{\Delta\lambda}^i, \widehat{\Delta\lambda}^j\right) = \Gamma_{ij},$$

which again confirms the interpretation of the Fisher Matrix as the uncertainty in the parameter estimate. The fractional corrections to the Fisher Matrix estimate scale like the inverse of the signal-to-noise ratio and therefore the Fisher Matrix is a good approximation in the high signal-to-noise ratio limit.

The Fisher Matrix has been widely used in a gravitational wave context to assess the measurability of parameters using observations with present or future detectors. While the Fisher Matrix is only an approximation, it can be directly calculated by evaluating a small number of waveforms, rather than requiring samples to be obtained all over the waveform parameter space, and so it is much cheaper computationally. This makes it a good tool for Monte Carlo simulations over parameter space, to survey parameter estimation accuracies over a wide parameter range.

7.2 Matched filtering

In the previous chapter we introduced the idea of matched filtering, motivated by maximising the signal to noise ratio of a filtered data stream. The optimal filter has a frequency-domain

kernel $\tilde{K}(f) \propto \tilde{h}(f)/S_h(f)$. The use of the output of the optimal filter as a test statistic for a search can also be motivated by the frequentist concepts that we encountered in previous chapters. Suppose that we write $\mathbf{h}(\lambda) = A\hat{\mathbf{h}}(\lambda)$, where $(\hat{\mathbf{h}}(\lambda)|\hat{\mathbf{h}}(\lambda)) = 1$, to separate out the amplitude of the gravitational wave source from the other parameters. The log-likelihood can be written

$$\begin{aligned} l(\lambda) &= -\frac{1}{2}(\mathbf{s} - A\hat{\mathbf{h}}(\lambda)|\mathbf{s} - A\hat{\mathbf{h}}(\lambda)) = -\frac{1}{2}\left[(\mathbf{s}|\mathbf{s}) - 2A(\mathbf{s}|\hat{\mathbf{h}}) + A^2\right] \\ &= -\frac{1}{2}\left[(\mathbf{s}|\mathbf{s}) + (A - (\mathbf{s}|\hat{\mathbf{h}}))^2 - (\mathbf{s}|\hat{\mathbf{h}})^2\right]. \end{aligned} \tag{102}$$

Fora given λ, this is maximized by the choice $A = (\mathbf{s}|\hat{\mathbf{h}})$, for which the log-likelihood $\propto (\mathbf{s}|\hat{\mathbf{h}})^2 - (\mathbf{s}|\mathbf{s})$. The maximum likelihood estimator for parameters other than the amplitude is thus given by the maximum of the optimal filter output over the parameter space. So, optimal filtering is just maximum likelihood estimation. To do this in practice, the optimal filter must be evaluated over the whole parameter space. In the analysis of gravitational wave data, from LIGO in particular, this is achieved using a **template bank**, which is a set of templates that cover the whole parameter space. The overlap of each template with the detector data is evaluated, and the maximum of those template overlaps is used as a test statistic to identify whether or not there is a signal in the data.

The question that we want to ask is "Is there a gravitational wave signal in the data?". Assuming that the parameters λ are fixed, this can formulated as a hypothesis test on the signal amplitude

$$H_0 : A = 0, \qquad \text{vs.} \qquad H_1 : A > 0.$$

From the Neyman-Pearson lemma the optimal statistic for testing the simple hypothesis $A = 0$ versus $A = A_1$ is the likelihood ratio, which is

$$\exp\left[A_1(\mathbf{s}|\hat{\mathbf{h}}(\lambda)) - \frac{1}{2}A_1^2\right].$$

This is large for large values of the optimal filter $(\mathbf{s}|\hat{\mathbf{h}}(\lambda))$ and so we deduce that the optimal filter is also the most powerful detection statistic. As the detection statistic does not depend on A_1, this test is uniformly most powerful for the composite hypothesis $A > 0$. In the more usual case that λ is unknown, although the maximum of the optimal filter statistic is still the maximum likelihood estimator, this is no longer a uniformly most powerful test, although it remains quite close to being so.

LIGO matched filtering searches typically use a large number of templates, distributed throughout the parameter space in a **template bank**. The matched filter output is evaluated for all of these templates, and the maximum filter output over the template bank is used as a detection statistic. Template banks are typically characterised by their **minimal match**, MM. This is defined as the *minimum* over all *possible signals* of the **maximum** overlap of that signal with one of the templates in the bank

$$\min_{\vec{\lambda}}\left[\max_{h_{\mathrm{temp,i}}:i=1,\ldots,N}(h(\vec{\lambda})|h_{\mathrm{temp,i}})\right] \gtrsim \mathrm{MM}$$

where $\{h_{\mathrm{temp,i}} : i = 1, \ldots, N\}$ are the N templates in the template bank. The minimal match is the worst possible detection statistic that a randomly chosen signal could have. Setting this minimal match to some value close to 1 ensures that very few signals will be missed. A

typical value of the minimal match used in practice would be 0.97. For a uniform distribution of sources in a Euclidean Universe, the fraction of sources that would be missed is $1-0.97^3 = 0.087$.

Template banks can be constructed analytically using the Fisher Matrix as a metric. This follows from expanding the overlap of two normalised templates, $\hat{h}(\vec{\lambda}) = h(\vec{\lambda})/\sqrt{(h(\vec{\lambda})|h(\vec{\lambda}))}$,

$$(\hat{h}(\vec{\lambda})|\hat{h}(\vec{\lambda}+\Delta\vec{\lambda})) = (\hat{h}(\vec{\lambda})|\hat{h}(\vec{\lambda})) + \left(\hat{h}(\vec{\lambda})\left|\frac{\partial\hat{h}}{\partial\lambda_i}(\vec{\lambda})\right.\right)\Delta\lambda^i + \frac{1}{2}\left(\hat{h}(\vec{\lambda})\left|\frac{\partial^2\hat{h}}{\partial\lambda_i\partial\lambda_j}(\vec{\lambda})\right.\right)\Delta\lambda^i\Delta\lambda^j + \cdots .$$

The first term is 1 because of the normalisation. The second term vanishes since

$$(\hat{h}(\vec{\lambda})|\hat{h}(\vec{\lambda})) = 1 \quad \Rightarrow \quad \frac{\partial}{\partial\lambda_i}(\hat{h}(\vec{\lambda})|\hat{h}(\vec{\lambda})) = 0 \quad \Rightarrow \quad \left(\hat{h}(\vec{\lambda})\left|\frac{\partial\hat{h}}{\partial\lambda_i}(\vec{\lambda})\right.\right) = 0.$$

The third term can be simplified using

$$\begin{aligned}\frac{\partial}{\partial\lambda_j}\left(\hat{h}(\vec{\lambda})\left|\frac{\partial\hat{h}}{\partial\lambda_i}(\vec{\lambda})\right.\right) = 0 \quad &\Rightarrow \quad \left(\frac{\partial\hat{h}}{\partial\lambda_i}(\vec{\lambda})\left|\frac{\partial\hat{h}}{\partial\lambda_j}(\vec{\lambda})\right.\right) + \left(\hat{h}(\vec{\lambda})\left|\frac{\partial^2\hat{h}}{\partial\lambda_i\partial\lambda_j}(\vec{\lambda})\right.\right) = 0 \\ &\Rightarrow \quad \left(\frac{\partial\hat{h}}{\partial\lambda_i}(\vec{\lambda})\left|\frac{\partial\hat{h}}{\partial\lambda_j}(\vec{\lambda})\right.\right) = -\left(\hat{h}(\vec{\lambda})\left|\frac{\partial^2\hat{h}}{\partial\lambda_i\partial\lambda_j}(\vec{\lambda})\right.\right). \qquad (103)\end{aligned}$$

We deduce

$$(\hat{h}(\vec{\lambda})|\hat{h}(\vec{\lambda}+\Delta\vec{\lambda})) = 1 - \frac{1}{2}\Gamma_{ij}\Delta\lambda^i\Delta\lambda^j.$$

The Fisher Matrix (of normalised templates) thus provides a metric on parameter space, which can be used to place templates. This is only practical in low numbers of dimensions. In higher numbers of dimensions, it is easier to use **stochastic template banks**. A stochastic bank is constructed as follows

1. At step 1, choose the first template, $\hat{h}(\lambda_1)$, randomly from parameter space. Add it to the template bank, $\mathcal{T}$.

2. At step $i \geq 2$, set the counter to 1 and then repeat the following steps:

   (a) Draw a random set of parameter values, $\vec{\lambda}_i$, and evaluate the match, M, with the current template bank

   $$M = \left[\max_{h_{\rm temp}\in\mathcal{T}}(h(\vec{\lambda}_i)|h_{\rm temp})\right].$$

   (b) If $M < MM$, add $h(\vec{\lambda}_i)$ to the template bank and advance to step $i+1$. Otherwise, increment the counter. If the counter has reached $N_{\rm max}$, stop. Otherwise return to step (a).

7.3 LIGO searches

LIGO employs two different matched filtering algorithms to search for signals, Pycbc and Gstlal. They differ in various details, including how the template overlaps are computed. We will not discuss these in detail here, but refer the interested reader to relevant publications. For *gstlal* these are

- Cannon, K., Cariou, R., Chapman, A., et al. (2012), *Astrophys. J.* **748**, 136, doi: 10.1088/0004-637X/748/2/136.
- Privitera, S., Mohapatra, S. R. P., Ajith, P., et al. (2014), *Phys. Rev.* D **89**, 024003, doi: 10.1103/PhysRevD.89.024003
- Messick, C., Blackburn, K., Brady, P., et al. (2017), *Phys. Rev.* D **95**, 042001, doi: 10.1103/PhysRevD.95.042001
- Sachdev, S., Caudill, S., Fong, H., et al. (2019), arXiv:1901.08580
- Hanna, C., Caudill, S., Messick, C., et al. (2019), arXiv:1901.02227

For *pycbc* the relevant references are

- Nitz, A., Harry, I., Brown, D., et al. (2019), gwastro/pycbc: PyCBC Release v1.15.2, doi: 10.5281/zenodo.3596447
- Nitz, A. H., Dal Canton, T., Davis, D., & Reyes, S. (2018), *Phys. Rev.* D **98**, 024050, doi: 10.1103/PhysRevD.98.024050
- Usman, S. A., Nitz, A. H., Harry, I. W., et al. (2016), *Class. Quantum Grav.* **33**, 215004, doi: 10.1088/0264-9381/33/21/215004

Both searches adopt a traditional frequentist framework, in that the output of the pipeline is used as a detection statistic. If the detection statistic exceeds a threshold then the data is flagged as interesting, i.e., potentially containing a signal. The threshold is determined based on the behaviour of the search pipeline in the absence of any signals in the data. This background distribution is estimated using **time slides**. Both searches rely on consistency between triggers in two or more detectors. Any astrophysical gravitational wave signal must pass through both detectors within an interval of 10ms. If the data of one detector is time shifted relative to the other by more than this amount, then any coincident triggers in the two instruments must be due to instrumental noise only. By doing many different time shifts in this way, the background distribution can be estimated for much longer effective observation times.

In hypothesis testing, we discussed the notion of a significance or p-value. This makes sense if the size of the data set is fixed, but gravitational wave detectors are continuously taking data. Therefore it makes sense to quantify significance instead by a *false alarm rate* or FAR, which is the frequency at which triggers as extreme as the one observed, or more extreme, occur in the data. LIGO quotes FARs for all events that are distributed publicly.

We will now give an overview of a few techniques that are used in LIGO searches to improve their speed and efficiency.

7.3.1 Waveform consistency

The assumptions that lead to the optimal filter assume that the noise is stationary. This is approximately true for gravitational wave detectors, but they are also observed to have large glitches quite often. While the glitches do not match any of the templates well, there is often sufficient power in the glitch that they can trigger the detection statistic to exceed the threshold. To mitigate for this problem, LIGO searches use **waveform consistency** checks. These verify that after subtracting the best-fit template signal from the data, the resulting time series is consistent with being stationary Gaussian noise with the estimated PSD. If the template $\hat{h}$ coincides with the true signal, the quantity

$$\chi^2 = \sum_{k=1}^{N} \frac{|\hat{s}_k - \hat{h}_k|^2}{S_h(f_k)}$$

is the sum of squares of $N(0,1)$ distributed random variables, and hence follows a chi-squared distribution with N degrees of freedom. The mean of a χ^2_N random variable is N, so χ^2/N should be expected to be close to 1 if the template is a good match to the data, and much bigger otherwise. LIGO uses something called *effective SNR* as a detection statistic. This is defined as

$$\hat{\rho} = \frac{\rho}{(1 + (\chi^2/N)^3)^{\frac{1}{6}}}.$$

For real signals, this is close to the true SNR, while for glitches it is much smaller. The effective SNR is used as the detection statistic by *pycbc*.

7.3.2 Marginalisation over phase and time

A template bank requires templates in all parameters, so it is useful to reduce the dimensionality of the parameter space whenever possible. This can be done straightforwardly for the *initial phase* and *time of coalescence*. For a monochromatic signal

$$h(t|A, f_0, t_c, \phi_0) = A\cos(2\pi f_0(t-t_c)+\phi_0) = A\cos(2\pi f_0(t-t_c))\cos\phi_0 - A\sin(2\pi f_0(t-t_c)\sin\phi_0$$

the matched filter overlap is

$$(s|h) = A\cos\phi_0 O_c - A\sin\phi_0 O_s, \qquad \text{where } O_c = (s|\cos(2\pi f_0(t-t_c))), \quad O_s = (s|\sin(2\pi f_0(t-t_c))).$$

Differentiating with respect to ϕ_0 and equating it to zero, we find that the value of ϕ_0 that maximises the overlap is

$$\tan\phi_0 = -\frac{O_s}{O_c} \qquad \Rightarrow \quad \max_{\phi_0}(s|h)^2 = A^2(O_c^2 + O_s^2).$$

If this is used instead of the standard overlap, then the template bank automatically maximises over phase and this parameter direction does not need to be covered by templates.

To maximize over the unknown coalescence time we use

$$\tilde{h}(f|A, f_0, t_c, \phi_0) = \tilde{h}(f|A, f_0, 0, \phi_0)\exp(-2\pi i f t_c)$$

and observe that

$$(s|h(t|A, f_0, t_c, \phi_0)) = 2\Re\int_{-\infty}^{\infty} \frac{\tilde{s}^*(f)\tilde{h}(f|A, f_0, 0, \phi_0)}{S_h(f)}\exp(-2\pi i f t_c)\mathrm{d}f.$$

This is just the inverse Fourier transform of

$$\frac{\tilde{s}^*(f)\tilde{h}(f|A, f_0, 0, \phi_0)}{S_h(f)}.$$

Inverse Fourier transforms can be computed cheaply (in $n \log n$ time) using the fast Fourier transform. Therefore, the time of coalescence can be efficiently maximized over by computing the quantity above, taking its inverse fast Fourier transform, and then finding the maximum of the components of the resulting vector.

7.3.3 The F-statistic

The F-statistic is an extension of the above ideas to more of the extrinsic parameters of the signal. It is not used so much for LIGO, but has been used extensively in LISA data analysis work (see for example Cornish & Porter (2007), *Phys. Rev.* D**75**, 021301; *Class. Quantum Grav.* **24**, 5729). The idea is to write the signal as a sum of modes, such that the coefficients depend only on a (subset of) the extrinsic parameters, and then analytically maximise over those coefficients. For SMBH binaries in LISA the decomposition takes the form

$$h(t) = \sum_{i=1}^{4} a_i(\iota, \psi, D_L, \phi_c) A^i(t|M_c, \mu, t_c, \theta, \phi)$$

where

$$\begin{aligned}
a_1 &= \Lambda[(1+\cos^2\iota)\cos 2\psi \cos\phi_c - 2\cos\iota \sin 2\psi \sin\phi_c] \\
a_2 &= -\Lambda[(1+\cos^2\iota)\sin 2\psi \cos\phi_c + 2\cos\iota \cos 2\psi \sin\phi_c] \\
a_3 &= \Lambda[(1+\cos^2\iota)\cos 2\psi \sin\phi_c + 2\cos\iota \sin 2\psi \cos\phi_c] \\
a_4 &= -\Lambda[(1+\cos^2\iota)\sin 2\psi \sin\phi_c - 2\cos\iota \cos 2\psi \cos\phi_c] \\
A_1 &= M\eta x(t) D^+ \cos(\Phi) \\
A_2 &= M\eta x(t) D^\times \cos(\Phi) \\
A_3 &= M\eta x(t) D^+ \sin(\Phi) \\
A_4 &= M\eta x(t) D^\times \sin(\Phi).
\end{aligned} \tag{104}$$

Here the waveform parameters are inclination ι, polarization angle, ψ, luminosity distance, D_L, phase at coalescence, ϕ_c, chirp mass, M_c, reduced mass ratio, μ, time of coalescence, t_c, colatitude, θ, and azimuth, ϕ. We denote the waveform phase by $\Phi(t)$ and $x = (GM\omega/c^3)^{2/3}$, where ω is the orbital frequency and $M = m_1 + m_2$ is the total mass. The quantities D^+ and D^x are the two components of LISA's time-dependent response function.

Writing $N^i = (s|A^i)$, the matched filter overlap is

$$(s|h) = a_j N^j$$

and we want to maximise this subject to the constraint that the waveform is normalised which becomes

$$a_i M^{ij} a_j = 1,' \qquad \text{where } M^{ij} = (A^i|A^j).$$

This is a standard optimisation problem with solution

$$a_i = (M^{-1})_i j N^j = M_{ij} N^j.$$

The maximized value of the log-likelihood is the F-staistic

$$\mathcal{F} = \frac{1}{2} M_{ij} N^i N^j.$$

This can be used to automatically maximise over extrinsic parameters in a search, reducing the dimensionality of the parameter space to just that of the intrinsic parameters. Note that in the above we have taken the coefficients, a_i, to be independent of one another and unconstrained, while in practice they are correlated and take a potentially limited range of values because they all depend on the same set of four extrinsic parameters. Thus, we are finding the maximum over a space that is somewhat larger than the true space, and contains some unphysical values. If there is a signal in the data, then the maximization must nonetheless still give the right extrinsic parameter values (in the absence of noise).

7.3.4 Power spectral density estimation

The likelihood contains the spectral density of noise in the detector, which is usually not known precisely. LIGO searches (and parameter estimation codes) need to use a PSD that has been estimated from the data. This is accomplished by considering a number of other sections of data, distributed either side of the section of data that is of interested because it is believed to contain a signal. The power spectrum (i.e., the norm squared of the Fourier transform) is computed for each of the empty segments, $\sigma_i^2(f)$, and then these can be combined to give an estimate of the PSD in the segment of interest. The averaging can be done by taking the mean

$$\sigma_0^2(f) = \frac{1}{2N} \sum_{k=1}^{N} (s_k^2 + s_{-k}^2)$$

but in LIGO analyses it is more usual to use the median. The median is less susceptible to outliers in the data arising from non-stationary features in the noise.

7.4 Unmodelled searches

For burst sources matched filtering cannot be used, as it is not possible to build templates of potential signals. LIGO uses a number of different searches for unmodelled sources. Again, we won't describe these in detail, but refer to papers that give full details on the algorithms:

- **Coherent Wave Burst (CWB)**:
  - S. Klimenko et al. (2016), *Phys. Rev.* D **93**, 042004, arXiv:1511.05999.
- **MBTA**:
  - Adams, T., Buskulic, D., Germain, V., et al. (2016), *Class. Quantum Grav.* **33**, 175012, doi: 10.1088/0264-9381/33/17/175012
- **SPIIR**:
  - Luan, J., Hooper, S., Wen, L., & Chen, Y. (2012), *Phys. Rev.* D **85**, 102002, doi: 10.1103/PhysRevD.85.102002
  - Hooper, S., Chung, S. K., Luan, J., et al. (2012), *Phys. Rev.* D **86**, 024012, doi: 10.1103/PhysRevD.86.024012

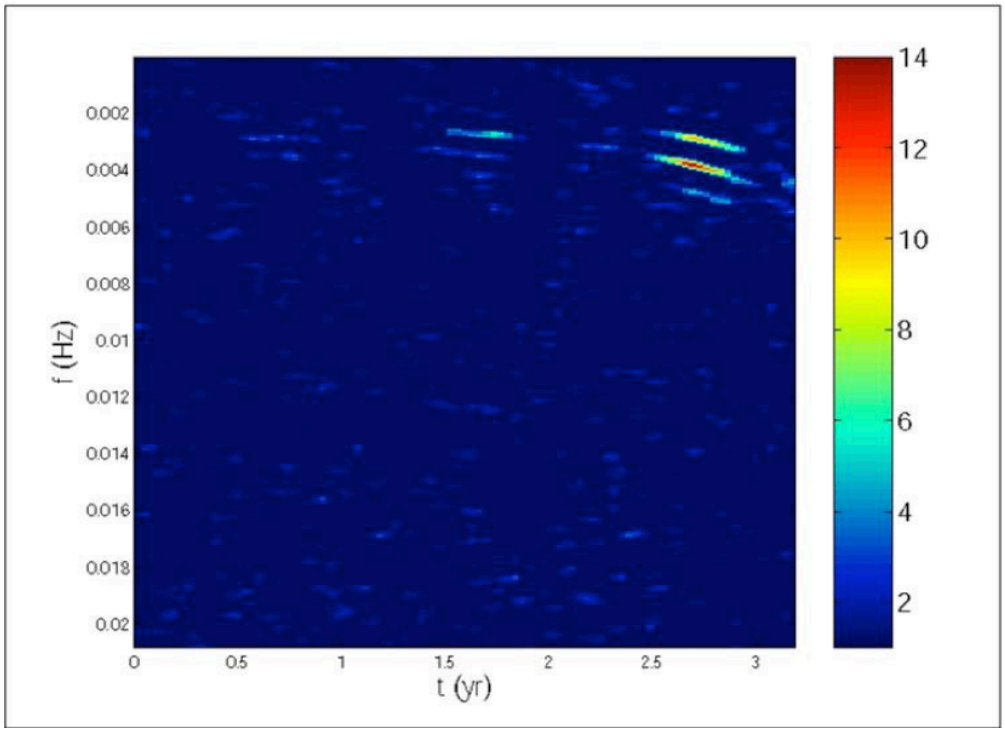

Figure 27: Example of a time-frequency spectrogram. Reproduced from Wen & Gair (2005).

  - Chu, Q. (2017), PhD thesis, University of Western Australia
  - Guo, X., Chu, Q., Chung, S. K., et al. 2018, *Co. Phys.* C **231**, 62, doi: 10.1016/j.cpc.2018.05.002

- **X-pipeline**:

  - Sutton, P. J., Jones, G., Chatterji, S., et al. (2010), *N J Phys.* **12**, 053034
  - Was, M., Sutton, P. J., Jones, G., & Leonor, I. (2012), *Phys. Rev.* D **86**, 022003

All of these algorithms search for clusters in **time-frequency spectrograms** of the data. The full data stream is divided into (usually overlapping) time segments, windowed and Fourier-transformed to obtain a frequency-domain representation of that chunk of data. The norm of these spectra is computed and they are then arranged next to one another in a grid. An example of a spectrogram is shown in Figure 27. Real astrophysical sources tend to produce coherent groups of bright pixels, or tracks, in these spectrograms. The patterns will be similar in different detectors in the network. The various time-frequency algorithms typically first evaluate bright pixels in the spectrograms, by thresholding on the power or some derived quantities. Then they cluster the pixels into groups, apply consistency criteria for the location of groups in two or more detectors in the network, and hence identify triggers of interest.

Time-frequency methods have also been applied to analysis of simulated LISA data, in the context of the LISA Mock Data Challenges (e.g., Gair, J.R. and Jones, G.J. (2007), *Class. Quantum Grav.* **24**, 1145; Gair, J.R., Mandel, I. and Wen, L. (2008), *Class. Quantum Grav.* **25**, 184031; Gair, J.R. and Wen, L. (2005), *Class. Quantum Grav.* **22**, S1359; Wen, L. and Gair, J.R. (2005), Detecting extreme mass ratio inspirals with LISA using time-frequency

methods, *Class. Quantum Grav.* **22**, S445.). While these algorithms were successful in simplified situations (i.e., with many fewer sources in the data than we would expect to see in practice) they are unlikely to be very effective when applied to real LISA data, due to the very large number of expected sources that will be overlapping in both time and frequency.

7.5 Semi-coherent searches

For continuous gravitational wave signals, e.g., rotating neutron stars in LIGO data, or very long-lived inspiral signals, e.g., extreme-mass-ratio inspirals in LISA data, matched filtering is possible in the sense that templates of the signals can be generated. however, it is computationally impossible, because the number of templates required to ensure a dense coverage of parameter space is extremely large. In these cases, it is possible to use **semi-coherent** search methods. These involve dividing the data stream into shorter segments, analysing each of those segments with matched filtering, and them adding up the power in the matched filter outputs along trajectories through the segments that correspond to physical inspirals. This approach is summarised in Figure 28. The semi-coherent approach is more computationally efficient, because the number of templates required to densely cover the parameter space for shorter observation times is much smaller.

A discussion of the use of a semi-coherent technique for detection of extreme-mass-ratio inspirals may be found in Gair, J.R. et al. (2004), *Class. Quantum Grav.* **21**, S1595. In that context, the coherent phase used 2 week segments of data, out of 1 year long LISA data sets. The coherent phase also employs the $\mathcal{F}$-statistic described above to automatically maximize over some of the extrinsic parameters.The impact of using the semi-coherent method rather than fully coherent matched filtering is to increase the estimated matched-filtering signal-to-noise ratio threshold for detection from $\rho = 14$ to $\rho = 30$.

In the context of the ground-based detectors, similar methods are used to search for continuous gravitational wave signals from rotating pulsars. The most recent LIGO results from the O2 science run are described in this paper

- Abbott, B.P. et al. (2019), *All-sky search for continuous gravitational waves from isolated neutron stars using Advanced LIGO O2 data*, *Phys. Rev.* D **100**, 024004.

LIGO uses two primary search methods. The **time-domain F-statistic** uses the same technique as the EMRI search described above. In fact, the latter was based on the former. Further details can be found in

- Aasi, J. et al. (2014), *Class. Quantum Grav.* **31**, 165014
- Jaranowski, P., Królak, A. and Schutz, B.F. (1998), *Phys. Rev.* D **58**, 063001
- Astone, P., Borkowski, K.M., Jaranowski, P., Pietka M. and Królak, A. (2010), *Phys. Rev.* D **82**, 022005
- Pisarski, A. and Jaranowski, P. (2015), *Class. Quantum Grav.* **32**, 145014

LIGO also employs a second method, called **the Hough transform**. The first stage of this algorithm is the same as the stack-slide method, i.e., coherent matched filtering on shorter segments of data. The second stage is slightly different, using the Hough transform, which is a technique for edge-detection in images, to identify tracks through the coherent template overlaps that might correspond to true signals. Further details can be found in

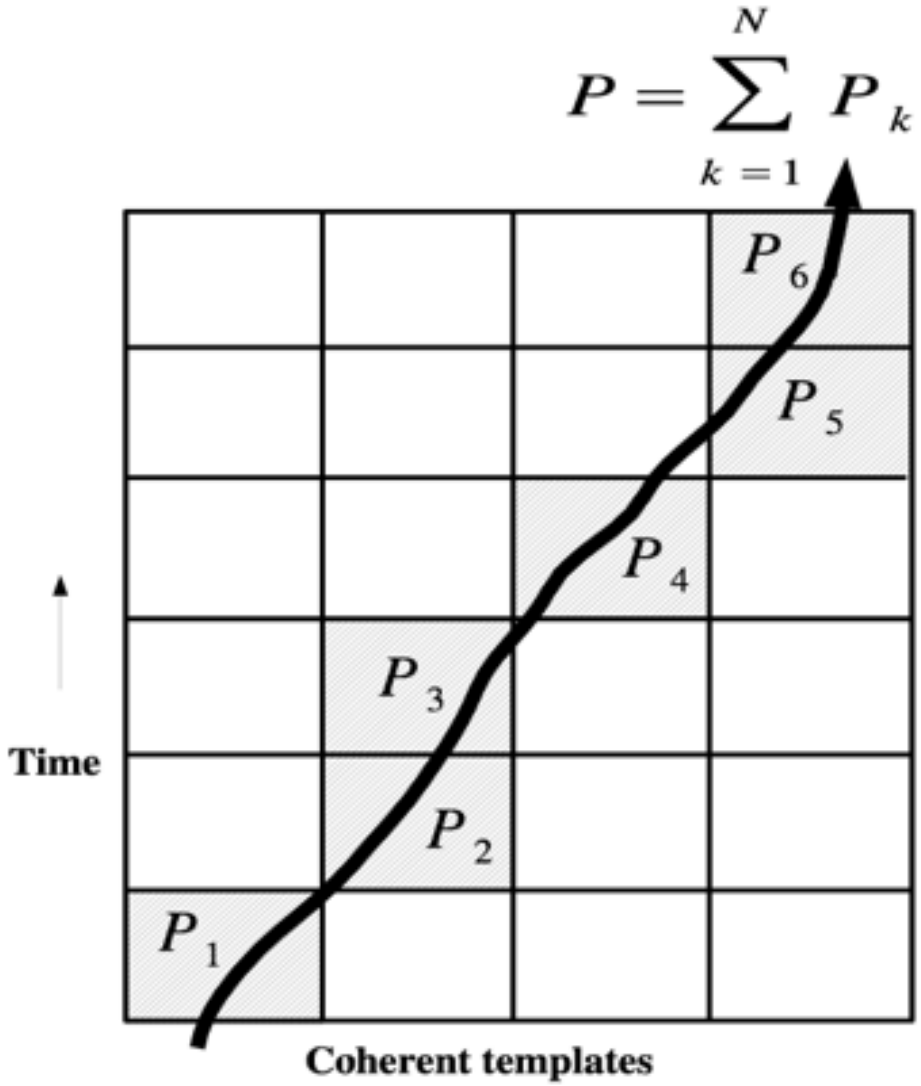

Figure 28: Illustration of the semi-coherent search method. The data is divided into shorter segments, which are searched coherently using waveform templates. The power in the templates is then summed incoherently along trajectories through the templates that correspond to EMRI inspiral trajectories. Reproduced from Gair et al. (2005).

- Astone, P., Colla, A., D?Antonio, S., Frasca, S. and Palomba, C. (2014), *Phys. Rev.* D **90**, 042002
- Antonucci, F., Astone, P., D?Antonio, S., Frasca, S. and Palomba, C. (2008), *Class. Quantum Grav.* **25**, 184015
- Krishnan, B., Sintes, A.M., Papa, M.A., Schutz, B.F., Frasca, S. and Palomba, C. (2004), *Phys. Rev.* D **70**, 082001

7.6 Searches for stochastic backgrounds

Stochastic backgrounds require different search techniques again. It is difficult to identify a background in a single detector, as it is essentially a noise source which is therefore challenging to distinguish from instrumental noise. Instead, background searches make use of multiple detectors and cross-correlate them to identify the common component of the noise. A typical detection statistic takes the form

$$\begin{aligned} Y_Q &= \int_0^T \mathrm{d}t_1 \int_0^T \mathrm{d}t_2 \; s_1(t_1) Q(t_1 - t_2) s_2(t_2) \\ &= \int_{-\infty}^{\infty} \mathrm{d}f \int_{-\infty}^{\infty} \mathrm{d}f' \, \delta_T(f - f') \tilde{s}_1^*(f) \tilde{Q}(f') \tilde{s}_2(f'). \end{aligned} \tag{105}$$

In the above, $Q(t)$ is a filter, which is analogous to the filter introduced in the single source detection case discussed earlier. The function $\delta_T(f)$ is a finite time approximation to the Dirac delta function

$$\delta_T(f) = \int_{-T/2}^{T/2} \mathrm{e}^{-2\pi i f t} \mathrm{d}t = \frac{\sin(\pi f T)}{\pi f}.$$

A generic gravitational wave background can be decomposed into a superposition of plane waves and a sum over polarisation states

$$h_{ij}(t, \vec{x}) = \int_{-\infty}^{\infty} \mathrm{d}f \int_{S^2} \mathrm{d}^{\Omega}_{\hat{k}} \mathrm{e}^{2\pi i f (t - \hat{k}\cdot\vec{x})} \mathcal{H}_A(f, \hat{k}) \mathbf{e}^A_{ij}(\hat{k}).$$

Here A labels the polarisation state, which for gravitational waves in general relativity is either plus or cross, $A = \{+, \times\}$, but in general metric theories could also include scalar and vector modes. The quantities $\mathbf{e}^A_{ij}(\hat{k})$ are the polarisation basis tensors for the individual polarisation modes

$$\mathbf{e}^+_{ij}(\hat{k}) = \hat{l}_i \hat{l}_j - \hat{m}_i \hat{m}_j, \qquad \mathbf{e}^\times_{ij}(\hat{k}) = \hat{l}_i \hat{m}_j + \hat{m}_i \hat{l}_j$$

where

$$\begin{aligned} \hat{k} &= \sin\theta\cos\phi\,\hat{x} + \sin\theta\sin\phi\,\hat{y} + \cos\theta\,\hat{z} \\ \hat{l} &= \cos\theta\cos\phi\,\hat{x} + \cos\theta\sin\phi\,\hat{y} - \sin\theta\,\hat{z} \\ \hat{m} &= -\sin\phi\,\hat{x} + \cos\phi\,\hat{y} \end{aligned} \tag{106}$$

are the standard spherical-polar coordinate basis vectors on the sky at colatitude θ and longitude ϕ. The quantities $\mathcal{H}^A(f, \hat{k})$ are the amplitudes of the various modes. For an unpolarised, stationary and statistically isotropic gravitational wave background, the expectation value of pairs of these amplitudes is given by

$$\left\langle \mathcal{H}^A(f, \hat{k}) \mathcal{H}^{A'*}(f', \hat{k}') \right\rangle = H(f) \delta(f - f') \delta^2(\hat{k}, \hat{k}') \delta_{AA'}, \tag{107}$$

where $H(f)$ is a real-valued function that depends on the energy density in the gravitational wave background and can be related to $\Omega_{\rm GW}(f)$, as introduced in the previous chapter, by

$$H(f) = \frac{3H_0^2}{32\pi^3}\frac{\Omega_{\rm GW}(f)}{|f|^3}.$$

The response of a particular gravitational wave detector, labelled by I, to a gravitational wave field can be written in the form

$$\begin{aligned} s_I(t) &= \int_{-\infty}^{\infty} {\rm d}\tau \int_{R^3} {\rm d}^3\vec{y}\, h_{ij}(t-\tau, \vec{x}-\vec{y}) R_I^{ij}(\tau, \vec{y}) \\ &= (2\pi)^3 \int_{-\infty}^{\infty} {\rm d}f \int_{R^3} {\rm d}^3\vec{k}\, \tilde{h}_{ij}(f,\vec{k}) \tilde{R}_I^{ij}(f,\vec{k}) {\rm e}^{i(2\pi f t - \vec{k}\cdot\vec{x}_I)} \end{aligned} \tag{108}$$

where $R^{ij}(t,\vec{x})$ is the impulse response of the detector, and the integral is over the spatial extent of the detector. Combining Eq. (108) with Eq. (107) we obtain

$$\langle Y_Q \rangle = \frac{T}{2}\int_{-\infty}^{\infty} \gamma_{12}(|f|)\tilde{Q}(f)H(f){\rm d}f$$

where $\gamma(|f|)$ is the **overlap reduction function**, which depends on the relative separation and orientation of the two detectors and is defined by

$$\gamma_{12}(|f|) = \int_{S^2} {\rm d}\Omega_{\hat{k}}\, \tilde{R}_1^A(f,\hat{k})\tilde{R}_2^{A*}(f,\hat{k}){\rm e}^{-2\pi i f \hat{k}\cdot(\vec{x}_1-\vec{x}_2)}$$

where

$$\tilde{R}_I^A(f,\hat{k}) = (2\pi)^e \mathbf{e}_{ij}^A(\hat{k})\tilde{R}_I^{ij}(f, 2\pi f\hat{k}).$$

The overlap reduction function for various combinations of ground-based interferometers and resonant bar detectors is shown in Figure 29. Stochastic backgrounds generated by large numbers of supermassive black hole binary inspirals are also the primary source for pulsar timing arrays. In that case, the "detector" is the measured redshift of a pulsar. The overlap reduction function for the detection of an isotropic stochastic background by cross-correlation of the measured redshifts of two different pulsars must be a function of only the angular separation between the pulsars on the sky. The resulting overlap reduction function curve is called the Hellings and Downs curve and is shown in Figure 30. Overlap reduction functions for non-isotropic backgrounds, for example anisotropic or correlated backgrounds, of backgrounds with non-GR polarisations, look different, providing a diagnostic for these physical properties of any observed stochastic background.

As in the case of the optimal filter, it is possible to maximise the signal-to-noise ratio of the filtered output. This takes a similar form to the optimal filter result

$$\tilde{Q}(f) \propto \frac{\gamma(|f|)\Omega_{\rm GW}(|f|)}{|f|^3 S_1(|f|)S_2(|f|)}$$

where $S_1(|f|)$ and $S_2(|f|)$ are the power spectral densities of the noise in the two detectors.

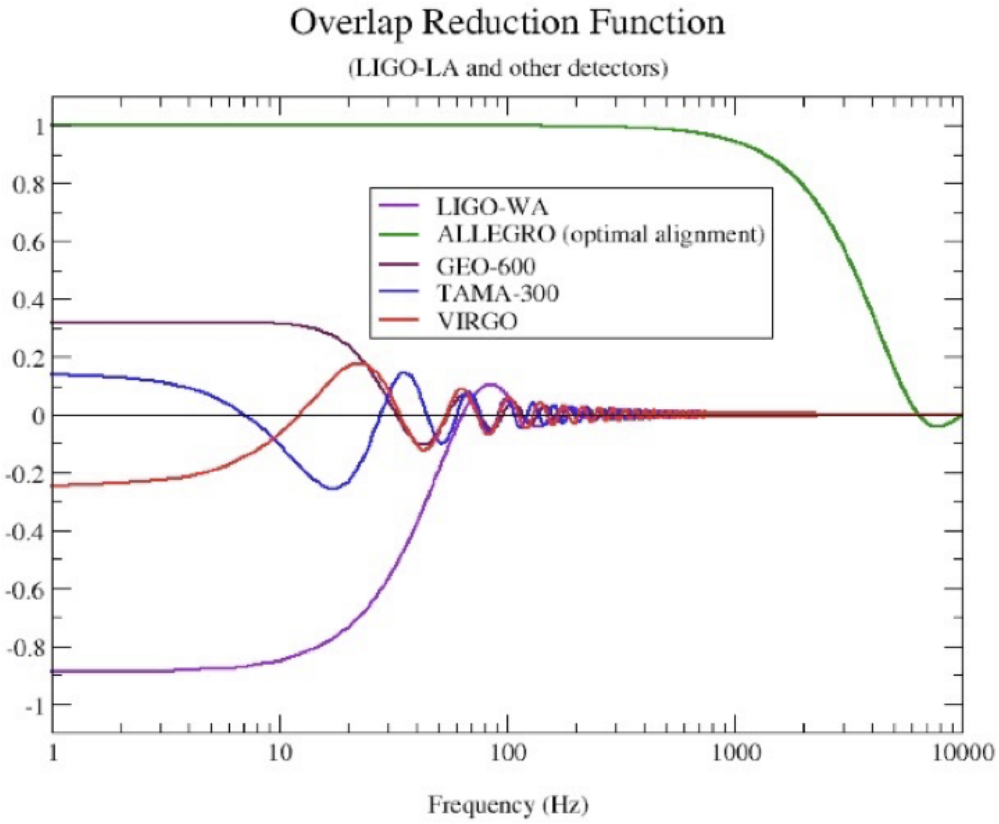

Figure 29: Overlap reduction function of the LIGO Livingston detector with LIGO Hanford (lower purple curve), Virgo (red curve), GEO (upper purple curve), TAMA (now obsolete) (blue curve) and the resonant bar detector Allegro (green curve), which was also sited in Louisiana. This was the network of detectors operating at the time of initial LIGO's science runs.

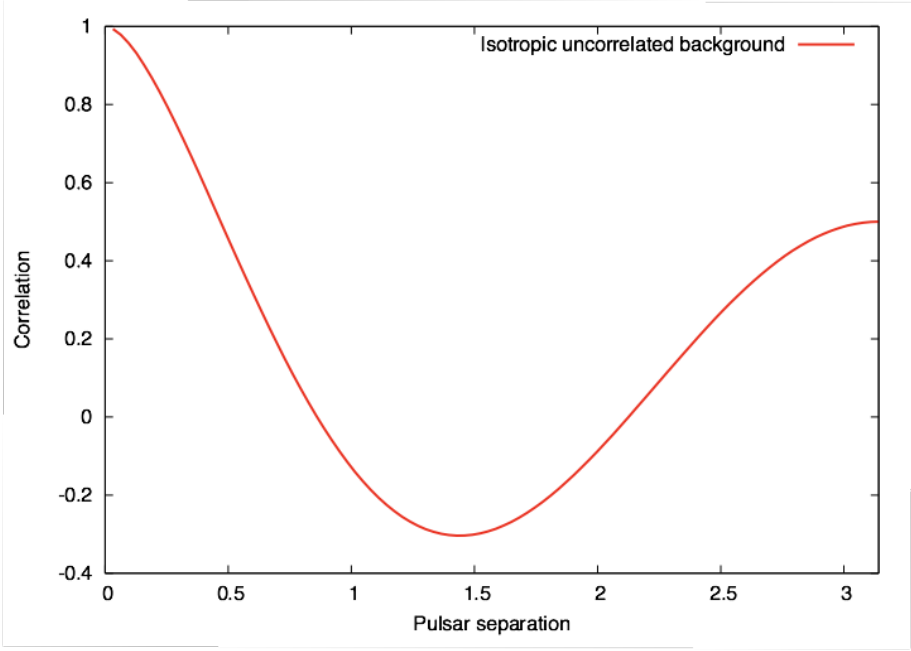

Figure 30: Overlap reduction function for the cross-correlation of the redshifts of two pulsars observed in a pulsar timing array, as a function of the angular separation of the two pulsars on the sky. This is known as the Hellings and Downs curve and the observation of a cross-correlation pattern that matches with this expectation is critical for the pulsar timing detection of gravitational waves.