BAYESBurst?

Reed Essick
ressick@mit.edu

May 25, 2014

- Work out priors on signal carefully. These include
  - localized in time in wave-frame (transient) : $h_j(t) = (\theta(t+\Delta t) - \theta(t-\Delta t))\, h_j(t)$
  - localized in frequency in wave-frame (band-limitted) : $h_j(f) = (\theta(f+\Delta f) - \theta(t-\Delta f))\, h_j(f)$
  - signals are distributed uniformly in volume, and all observed strain comes from a single event (astrophysically distributed)
- Propose implementation
  - position-space (θ, ϕ) or spherical-harmonics (Y_{lm})?
  - time-domain $(h_j(t))$ or frequency domain $(h_j(f))$?

Contents

1 Notation

Throughout these notes I adopt the following notation.

- all sky positions will be given in Earth-fixed coordinates (θ, ϕ). These are the standard polar coordinates, with polar angle θ and azimuthal angle ϕ.
- antenna patterns are functions of souce position $\vec{\Omega} \equiv (\theta, \phi) \in \mathbb{R}^2$ as well as frequency. Most analyses treate the antenna patterns as independent of frequency, but we want to wrap the time-shifts caused by different times-of-arival at different points on the Earth into the antenna patterns. In the frequency domain, this is simlpy a shift in phase $e^{-2\pi i f \Delta t}$.
- we work in the frequency domain because the antenna patterns, including time delays, are trivial in that basis. We avoid shifting data streams by applying phases in the frequency domain.
- lower case greek indicies will represent interferometers (ifos). eg: $\beta \in \{L, H, V\}$
- latin indicies will represent polarization states. eg: $j \in \{+, \times\}$
- upper case greek indicies will represent sky positions. eg: $\Omega = (\theta, \phi)$ or $\Omega = (l, m)$ as appropriate.
- we adopt the einstein summation notation for repeated indicies unless otherwise noted. If there is any ambiguity, we'll explicitly write the sums with Σ notation

2 preliminaries

We begin our analysis with several basic assumptions about interferometric gravitational wave detectors. These include characterizing the noise in the instruments and the detectors' sensitivities to different polarizations and source positions.

2.1 uncorrelated gaussian noise

We assume that the probability for observing a set of complex noise amplitudes in the fourier domain is

$$p(n_\beta(f)) = \frac{1}{N} \exp\left(- \int \mathrm{d}f \sum_\beta \frac{n_\beta \cdot n_\beta^*}{S_\beta} \right) \tag{1}$$

where we define the gaussian noise power spectrum as

$$\langle n_\beta(f) \rangle = 0 \tag{2}$$

$$\langle n_\beta(f) n^*_\alpha(f') \rangle = \frac{1}{2} S_\beta(f) \delta_{\beta\alpha} \delta(f - f') \tag{3}$$

Note that the probability is simply the noise weighted inner product of the noise realization.

2.2 antenna patterns and time delays

We follow the conventions used in [1] when defining our antenna patterns. Their equation B7 describes the analytic form for the antenna patterns, which are complicated, but they depend on three parameters.[1]

$$\begin{aligned} \mathcal{F}_+ &= \mathcal{F}_+(\theta, \phi, \psi) \\ \mathcal{F}_\times &= \mathcal{F}_\times(\theta, \phi, \psi) \end{aligned} \tag{4}$$

where θ and ϕ are the normal spherical coordinates in an Earth-fixed frame. ψ is the "polarization angle" and is poorly defined for bursts. Physically, it reflects the relative rotation of the coordinate systems in which we define the gravitational wave and the Earth-fixed frame. Therefore, a rotation of ψ simply mixes the different polarization states and does not affect the signal physically. If we have not model for the signal a priori, we can reconstruct our signal in any coordinate system we choose. This means that we can select ψ arbitrarily without affecting our reconstruction.

Furthermore, because detectors are spatially separated and gravitational waves are expected to travel at the speed of light, true signals will arrive in different detectors at different times. If a source comes from the direction (θ, ϕ), we expect

$$\begin{aligned} \Delta t_{12} \equiv t_{ifo1} - t_{ifo2} &= \hat{\Omega} \cdot (\vec{r}_{ifo1} - \vec{r}_{ifo2}) \\ &= \sin\theta\cos\phi(x_{ifo1} - x_{ifo2}) + \sin\theta\sin\phi(y_{ifo1} - y_{ifo2}) + \cos\theta(z_{ifo1} - z_{ifo2}) \end{aligned} \tag{5}$$

Therefore, when accounting for the different times of arrival at various detectors, we can either shift the observed data in time or shift the reconstructed signal as needed. We choose the latter, and furthermore we work in the frequency domain to obviate the shifts in time. One can consider this to be reconstructing the signal as it arrives at the Earth's center (geocenter) rather than at any particular detector.

[1] Antenna patterns can be defined for other polarizations that are not predicted by General Relativity. Our analysis is general enough to encompass these cases, but we only explicitly consider the case of 2 polarizations.

For each detector, we define the *total* antenna pattern to be the combination of these two effects: sensitivity changes based on relative orientations and time of arrival differences based on relative locations. Therefore, throughout this note we refer to the antenna patterns as

$$F_{\beta i}(\theta,\phi,\psi) = \mathcal{F}_{\beta i}(\theta,\phi,\psi)\cdot e^{-2\pi i f(t_\beta(\theta,\phi)-t_{geo})} = \mathcal{F}_{\beta i}\cdot e^{-2\pi i f \Delta t_{\beta\oplus}} \tag{6}$$

3 Point sources

Here we examine a non-parametric bayesian approach to signal reconstruction for *point sources.* Extended sources are considered in Section 5.

3.1 maximum likelihood estimators

We can define the likelihood ratio as

$$\begin{aligned}
\mathcal{L} = \frac{p(d_\beta - F_{\beta j}h_j)}{p(d_\beta)} &= \exp\left(\int \mathrm{d}f \sum_\beta \frac{|d_\beta|^2 - |d_\beta - F_{\beta j}h_j|^2}{S_\beta}\right) && (7)\\
&= \exp\left(\int \mathrm{d}f \sum_\beta \frac{d_\beta F^*_{\beta j}h^*_j + d^*_\beta F_{\beta j}h_j - h_k F_{\beta k}F^*_{\beta j}h^*_j}{S_\beta}\right) && (8)
\end{aligned}$$

To obtain our maximum likelihood estimator, we vary this functional with respect to $h^*_j(f)$, treating h_j and h^*_j as independent variables. Euler-Lagrange equations yield

$$\begin{aligned}
\frac{\delta}{\delta h^*_m}\log\mathcal{L} &= \frac{\delta}{\delta h^*_m}\int \mathrm{d}f \sum_\beta \frac{d_\beta F^*_{\beta j}h^*_j + d^*_\beta F_{\beta j}h_j - F_{\beta k}h_k F^*_{\beta j}h^*_j}{S_\beta} = 0 && (9)\\
\Rightarrow 0 &= \frac{\mathrm{d}}{\mathrm{d}f}\left(\frac{\partial}{\partial(\mathrm{d}h^*_m/\mathrm{d}f)}\sum_\beta \frac{d_\beta F^*_{\beta j}h^*_j + d^*_\beta F_{\beta j}h_j - F_{\beta k}h_k F^*_{\beta j}h^*_j}{S_\beta}\right)\\
&\quad -\frac{\mathrm{d}}{\mathrm{d}h^*_m}\left(\sum_\beta \frac{d_\beta F^*_{\beta j}h^*_j + d^*_\beta F_{\beta j}h_j - F_{\beta k}h_k F^*_{\beta j}h^*_j}{S_\beta}\right) && (10)\\
&= -\sum_\beta \frac{\left(d_\beta F^*_{\beta k} - h_j F_{\beta j}F^*_{\beta k}\right)\delta_{km}}{S_\beta} && (11)\\
\Rightarrow \sum_\beta \frac{d_\beta F^*_{\beta k}}{S_\beta} &= \sum_\beta \frac{F^*_{\beta k}F_{\beta j}h_j}{S_\beta} && (12)
\end{aligned}$$

We now make the following definitions of convenience

$$A_{kj} \equiv \sum_{\beta} \frac{F^*_{\beta k} F_{\beta j}}{S_\beta} = A^*_{jk} \tag{13}$$

$$B_{j\beta} \equiv \frac{F^*_{\beta j}}{S_\beta} \qquad \text{(no sum over } \beta) \tag{14}$$

where A_{kj} is an $N_{polarizations} \times N_{polarizations}$ matrix and $B_{\beta j}$ is a $N_{ifos} \times N_{polarizations}$ matrix. With this in hand, we can write the Euler-Lagrange equations in the simple form

$$A_{kj} h_j = B_{k\beta} d_\beta \tag{15}$$

$$\Rightarrow \hat{h}_j = \left(A^{-1}\right)_{jk} B_{k\beta} d_\beta \tag{16}$$

We should note that A_{kj} is singular for at certain sky locations for a single interferometer, but it should never be singular for two (even-slightly) mis-aligned interferometers. Let us now consider the properties of this estimator. If the data contains a true signal $d_\beta = F_{\beta j} h_j + n_\beta$, we have

$$\hat{h}_j = \left(A^{-1}\right)_{jk} B_{k\beta} \left(F_{\beta m} h_m + n_\beta\right) \tag{17}$$

$$= \left(A^{-1}\right)_{jk} \left(B_{k\beta} F_{\beta m}\right) h_m + \left(A^{-1}\right)_{jk} B_{k\beta} n_\beta \tag{18}$$

We note that

$$B_{k\beta} F_{\beta m} = \sum_{\beta} \frac{F^*_{\beta k}}{S_\beta} F_{\beta_m} = A_{km} \tag{19}$$

which yields the pleasant simplification

$$\hat{h}_j = \left(A^{-1}\right)_{jk} A_{km} h_m + \left(A^{-1}\right)_{jk} B_{k\beta} n_\beta \tag{20}$$

$$= \delta_{jm} h_m + \left(A^{-1}\right)_{jk} B_{k\beta} n_\beta \tag{21}$$

$$\Rightarrow h_j - \hat{h}_j \equiv \epsilon_j = \left(A^{-1}\right)_{jk} B_{k\beta} n_\beta \tag{22}$$

and we see that the estimator is *unbiased* with gaussian errors around the actual signal.[2] Explicitly, we can compute the expected distributions of the reconstructed errors as

[2] The errors in the reconstructed polarization are functions of the noise in each detector, so they must also be gaussian distributed.

$$\begin{aligned}
\langle \epsilon_j \rangle &= \left\langle \left(A^{-1}\right)_{jk} B_{k\beta} n_\beta \right\rangle && (23)\\
&= \left(A^{-1}\right)_{jk} B_{k\beta} \langle n_\beta \rangle && (24)\\
&= 0 && (25)\\
\langle \epsilon_j^* \epsilon_k \rangle &= \left\langle n_\alpha^* B_{n\alpha}^* \left(A^{-1}\right)_{jn}^* \left(A^{-1}\right)_{km} B_{m\beta} n_\beta \right\rangle && (26)\\
&= B_{n\alpha}^* \left(A^{-1}\right)_{jn}^* \left(A^{-1}\right)_{km} B_{m\beta} \langle n_\alpha^* n_\beta \rangle && (27)\\
&= B_{n\alpha}^* \left(A^{-1}\right)_{jn}^* \left(A^{-1}\right)_{km} B_{m\beta} \left(\frac{1}{2} S_\beta \delta_{\alpha\beta}\right) && (28)\\
&= \frac{1}{2} \left(A^{-1}\right)_{jn}^* \left(A^{-1}\right)_{km} \sum_\beta \left\{ B_{n\beta}^* B_{m\beta} S_\beta \right\} && (29)\\
&= \frac{1}{2} \left(A^{-1}\right)_{jn}^* \left(A^{-1}\right)_{km} A_{mn} && (30)\\
&= \frac{1}{2} \left(A^{-1}\right)_{jn}^* \delta_{kn} && (31)\\
&= \frac{1}{2} \left(A^{-1}\right)_{jk}^* && (32)
\end{aligned}$$

Note that the standard deviation increases when the antenna patterns are small. This is because the network is less sensitive to the actual strain signal and we should expect larger errors in the reconstruction.

This also means that we can write the likelihood as

$$\begin{aligned}
\mathcal{L} &= \exp\left(\int \mathrm{d}f \sum_\beta \frac{|d_\beta|^2 - \left| d_\beta - F_{\beta j}\left(\hat{h}_j + \epsilon_j\right)\right|^2}{S_\beta} \right) && (33)\\
&= \exp\left(\int \mathrm{d}f\, d_\beta B_{\beta j} \left(\hat{h}_j + \epsilon_j\right)^* + d_\beta^* B_{\beta j}^* \left(\hat{h}_j + \epsilon_j\right) - \left(\hat{h}_j + \epsilon_j\right)^* A_{jk} \left(\hat{h}_k + \epsilon_k\right) \right) && (34)\\
&= \exp\Big(\int \mathrm{d}f\, d_\beta B_{\beta j} \left(A^{-1}\right)_{jk}^* B_{k\alpha}^* d_\alpha^* + d_\beta^* B_{\beta j}^* \left(A^{-1}\right)_{jk} B_{k\alpha} d_\alpha - d_\beta^* B_{m\beta}^* \left(A^{-1}\right)_{mj}^* A_{jk} \left(A^{-1}\right)_{kn} B_{n\alpha} d_\alpha \\
&\quad - \epsilon_j^* A_{jk} \epsilon_k + d_\beta B_{\beta j} \epsilon_j^* + d_\beta^* B_{\beta j}^* \epsilon_j - d_\beta^* B_{m\beta}^* \left(A^{-1}\right)_{jm}^* A_{jk} \epsilon_k - \epsilon_j^* A_{jk} \left(A^{-1}\right)_{kn} B_{n\alpha} d_\alpha \Big) && (35)
\end{aligned}$$

Exchanging dummy indicies and using the fact that A_{jk} is Hermitian shows that many terms cancel, including all linear terms in ϵ_j, and the final likelihood ratio can be written as

$$\mathcal{L} = \frac{p(d_\beta - F_{\beta j} h_j)}{p(d_\beta)} = \exp\left(\int \mathrm{d}f\, d_\beta^* B_{\beta j}^* \left(A^{-1}\right)_{jk} B_{k\alpha} d_\alpha - \epsilon_j^* A_{jk} \epsilon_k \right) \qquad (36)$$

which has a pleasing form. We see that the likelihood ratio is gaussian distributed around it's maximum value.

3.2 dominant polarization frame

As we've written A_{jk}, the off-diagnal components will most likely be non-zero. However, we also have freedom to choose the coordinate system in which we reconstruct the signal. This is controlled by the choice of polarization angle (ψ) in the definition of the antenna patterns. It can be shown that changing ψ is equivalent to rotating the wave-frame coordinate system and therefore mixing $h_j \to h'_j = U(1)h_j$. By appropriate choice of ψ, we can diagonalize A_{jk}. This is appealing for several reasons, not the least of which is lower computational complexity. It also allows us to easily identify some interesting featurs of networks of interferometric gravitational wave-detectors. In general, we can write

$$A_{jk} = s_j \delta_{jk} \tag{37}$$

where some of the s_j may be vanishingly small, but the set satisfies

$$\sum_i s_i = \sum_{i,\beta} \frac{F^*_{\beta i} F_{\beta i}}{S_\beta} \tag{38}$$

This choice of coordinate system, in which A_{jk} is diagonal, is often called the *dominant polarization frame.* We should also note that this is a feature of the relative orientations of the detectors and not their relative locations. The time of arrival difference cancel out because of the complex conjugation involved in the definition of A_{jk}, and therefore this matrix only depends on the detectors' orientations.

3.2.1 singular A_{jk} and effective number of polarizations

We note that s_j may not be non-zero for all source positions. In fact, for a single detector and two polarizations, it is straightforward to show that

$$s_0 = \frac{|F_0(\theta, \phi, \psi = \psi_o)|^2 + |F_1(\theta, \phi, \psi = \psi_o)|^2}{S} \tag{39}$$

$$s_1 = 0 \tag{40}$$

$$\tag{41}$$

where we've explicitly stated the "original" antenna patterns in a frame defined by ψ_o. Here, we notice that A_{jk} is *singular.* Whenever A_{jk} is singular, it means the detector network is effectively insensitive to one or more polarization channels. Therefore, a single detector is effectively sensitive to only a single polarization channel for all source positions. These eigenvalues may vary over the sky for more complicated detector networks.

An interesting case is a two-detector network with slightly mis-aligned detectors. In this case, $s_1 \sim |F_{1j} - F_{2j}|^2$ to lowest order in mis-alignment. If the detectors are only slightly mis-aligned, the network's sensitivity to the second polarization is very small but *non vanishing.*

In practice, if A_{jk} is singular for a set of points on the sky, we simply restrict our reconstruction to a subset of polarization channels to which the network is sensitive. This defines the *effective number of polarizations* to which the network is sensitive from any direction in the sky.

3.3 posterior probabilities

3.3.1 sky position

The useful distributions are the posteriors for the signal parameters. In this case, the full posterior can be written as

$$p(h_j, \theta, \phi | d_\beta) = \frac{p(d_\beta | h_j, \theta, \phi) p(h_j, \theta, \phi)}{p(d_\beta)} \tag{42}$$

In general, this is a very difficult function to compute. However, if we restrict ourselves to the posterior for the sky position

$$p(\theta, \phi | d_\beta) = \int \mathcal{D}h_j \, \frac{p(d_\beta | h_j, \theta, \phi) p(h_j, \theta, \phi)}{p(d_\beta)} \tag{43}$$

where the integral is taken over all possible signals $h_j = \hat{h}_j + \epsilon_j$. This means that, at a given (θ, ϕ), we can exchange the measure for something tractable : $\mathcal{D}h_j = \mathcal{D}\epsilon_j$. Furthermore, each frequency is independent in this integral, so we can exchange the order of the marginalization over h_j and the integration over f. This means we can analytically compute the posterior sky map

$$\begin{aligned} p(\theta, \phi | d_\beta) &= \int \mathcal{D}\epsilon_j \exp\left(\left(\int \mathrm{d}f \, d^*_\beta B^*_{\beta j} \left(A^{-1}\right)_{jk} B_{k\alpha} d_\alpha - \epsilon^*_j A_{jk} \epsilon_k \right) p(h, \theta, \phi) \right. & (44) \\ &= \prod_f \exp\left(d^*_\beta B^*_{\beta j} \left(A^{-1}\right)_{jk} B_{k\alpha} d_\alpha \right) \int_{-\infty}^{\infty} \mathrm{d}^{N_p} \epsilon_j \exp\left(-\epsilon^*_j A_{jk} \epsilon_k \right) p(\hat{h}_j + \epsilon_j, \theta, \phi) & (45) \end{aligned}$$

where N_p is the number of polarization states. The only barrier to evaluating these integrals analytically is the prior $p(h_j, \theta, \phi)$, which we can reasonably assume will take the form

$$p(h_j, \theta, \phi) = p(h_j) p(\theta, \phi) = p(h_j) \frac{1}{4\pi} \tag{46}$$

where we've assumed the prior on (θ, ϕ) is uniform across the sky (constant probability per steradian)[3] Furthermore, we can assume the most uninformative prior on h_j, so that

$$p(h_j) = \mathit{constant} \tag{47}$$

[3]This assumption is easily relaxed and does not affect the marginalization. Any prior is allowed on (θ, ϕ) and it will simply come outside the integral.

Under these assumptions[4], we have a very simple form for the posterior

$$p(\theta,\phi|d_\beta) = \frac{constant}{4\pi}\prod_f \exp\left(d_\beta^* B_{\beta j}^* \left(A^{-1}\right)_{jk} B_{k\alpha} d_\alpha\right)\int_{-\infty}^{\infty} \mathrm{d}^{N_p}\epsilon_j \exp\left(-\epsilon_j^* A_{jk}\epsilon_k\right) \quad (48)$$

$$= \frac{constant}{4\pi}\prod_f \exp\left(d_\beta^* B_{\beta j}^* \left(A^{-1}\right)_{jk} B_{k\alpha} d_\alpha\right)\sqrt{(2\pi)^{N_p}\left|\left(A^{-1}\right)_{jk}\right|} \quad (49)$$

Importantly, the marginalization preferentially gives more posterior probability to locations with *lower* antenna patterns, and therefore locations that are *less sensitive* to true gravitaitonal wave signals. This is counter intuitive (one expects there to be more posterior where the network is more sensitive), and is an artifact of the assumption $h_j = constant$. If all signals are equally likely, then the marginalization will select those regions with more allowed "volume" in the space of possible signals. This corresponds to locations with larger errors in the reconstructed signals, which are locations with *smaller* antenna patterns. However, applying even a somewhat arbitrary prior on h_j that favors smaller signals can fix this problem.

We should note that the only dependence on the data streams d_β comes from the maximum likelihood estimate, which is quadratic in the data. Everything else can be computed *exactly once* for all sky positions (θ,ϕ) and then used to filter the data.

We also note that the antenna patters *naturally* modulate the posterior through the marginalization. This means that when $N_p = N_{ifo}$ (and for any (θ,ϕ) the maximum likelihood estimator exactly reproduces the data streams) and the likelihood ratio is unity for all (θ,ϕ), *the posterior will not be uniform.* This non-uniformity is independent of the data streams and reflects the different sensitivities of the detector network at different points in the sky. Without a prior, this marginalization favors locations with *low* antenna patters and very little sensitivity to actual signals. Adding a realistic prior on h_j should allow us to include effects like triangulation by assigning higher priors to signals with less total energy. However, the prior will contain terms which depend on the data and may complicate the simple form of our current posterior. Furthermor, the exact form for this prior is uncertain. However, to obtain different posteriors for different data streams, we will have to impose some prior when $N_p = N_{ifo}$.

3.3.2 gaussian priors on h_j

A prior on h_j is required to give posteriors that depend on the data in 2-detector networks. Some possible examples are

$$p(h_j) \propto \exp\left(-h_k^* Z_{kj} h_j\right) \quad (50)$$

[4]Other assumptions may render this integral untractable analytically, but we could, for example, choose a gaussian on h_j and still evaluate the integral analytically.

where $Z_{jk} = Z^*_{kj}$. This has been called a *white-noise prior*. Choice of Z_{kj} is arbitrary. This prior is not particularly well motivated beyond the fact that the marginalization is still tractable. With such a prior, we can write

$$
\begin{aligned}
p(\theta, \phi | d_\beta) &\propto p(\theta, \phi) \int \mathcal{D}h_j \exp\left(\int \mathrm{d}f\, \hat{h}^*_j A_{jk} \hat{h}_k - \epsilon^*_j A_{jk} \epsilon_k - \hat{h}^*_j Z_{jk} \hat{h}_k - \hat{h}^*_j Z_{jk} \epsilon_k - \epsilon^*_k Z_{jk} \hat{h}_k - \epsilon^*_j Z_{jk} \epsilon_k \right) \\
&= p(\theta, \phi) \int \mathcal{D}h_j \exp\left(\int \mathrm{d}f\, \hat{h}^*_j \left(A - Z\right)_{jk} \hat{h}_k - \left[\epsilon^*_j \left(A + Z\right)_{jk} \epsilon_k + \hat{h}^*_j Z_{jk} \epsilon_k + \epsilon^*_k Z_{jk} \hat{h}^*_k \right] \right) \\
&= p(\theta, \phi) \int \mathcal{D}h_j \exp\left(\int \mathrm{d}f\, \hat{h}^*_j \left(A - Z\right)_{jk} \hat{h}_k + \zeta^*_j \Phi_{jk} \zeta_k - \left(\zeta_j + \epsilon_j\right)^* \Phi_{jk} \left(\zeta_k + \epsilon_k\right) \right)
\end{aligned} \tag{51}
$$

where $\zeta_k = \left(\Phi^{-1}\right)_{kj} Z_{jm} \hat{h}_m$ and $\Phi_{jk} = \left(A + Z\right)_{jk}$. WE can shift the marginalization measure to integrate over the gaussian terms independent of $\hat{h}_j$ to obtain

$$
p(\theta, \phi | d_\beta) = p(\theta, \phi) \prod_f \exp\left(\hat{h}^*_j \left(A - Z\right)_{jk} \hat{h}_k + \zeta^*_j \Phi_{jk} \zeta_k \right) \sqrt{(2\pi)^{N_p} \left|\Phi^{-1}\right|} \tag{52}
$$

$$
\begin{aligned}
&= p(\theta, \phi) \prod_f \exp\left(\hat{h}^*_j \left(A - Z\right)_{jk} \hat{h}_k + \hat{h}^*_m Z_{mj} \left(A + Z\right)^{-1}_{jk} Z_{kn} \hat{h}_n \right) \sqrt{(2\pi)^{N_p} \left|(A + Z)^{-1}_{jk}\right|} \\
&= p(\theta, \phi) \prod_f \exp\left(\hat{h}^*_j \left(\left(A - Z\right)_{jk} + Z_{jm} \left(A + Z\right)^{-1}_{mn} Z_{nk} \right) \hat{h}_k \right) \sqrt{(2\pi)^{N_p} \left|(A + Z)^{-1}_{jk}\right|}
\end{aligned} \tag{53}
$$

Similarly, if we assume a gaussian prior on h_j with some non-zero mean H_j, so that

$$
p(h_j) \propto \exp\left(-(h_k - H_k)^* Z_{kj} (h_j - H_j) \right) \tag{54}
$$

we obtain the following

$$
p(\theta, \phi | d_\beta) = p(\theta, \phi) \prod_f \exp\left(\hat{h}^*_j \left(A\right)_{jk} \hat{h}_k - \left(\hat{h}_j - H_j\right)^* \left(Z_{jk} - Z_{jm} \Phi^{-1}_{mn} Z_{nk} \right) \left(\hat{h}_k - H_k\right) \right) \sqrt{(2\pi)^{N_p} \left|\Phi^{-1}\right|} \tag{55}
$$

When we include the proper normalization for the prior on h_j as well as the *effective number of polarizations* defined by $N'_p = \mathrm{rank}\{A\}$, this posterior takes the following form

$$
p(\theta, \phi | d_\beta) = p(\theta, \phi) \left(2\pi\right)^{N'_p} \prod_f \exp\left(\hat{h}^*_j \left(A\right)_{jk} \hat{h}_k - \left(\hat{h}_j - H_j\right)^* \left(Z_{jk} - Z_{jm} \Phi^{-1}_{mn} Z_{nk} \right) \left(\hat{h}_k - H_k\right) + \frac{1}{2} \log\left(\left|\Phi^{-1}\right| |Z \right. \right. \tag{56}
$$

where A, Φ, and Z are $N'_p \times N'_p$ matrices, restricted to the non-singular subspace of A. However, this also assumes that $Z = \sigma^{-2} \mathbb{I}_{N'_p \times N'_p}$. By including the prior's normalizations, marginalization over extra degrees of freedom ($N_p > N'_p$) contribute a factor of unity.

Furthermore, we may consider linear combinations of gaussians

$$p(h_j) = \sum_N C_N \cdot \exp\left(- \left(h_k - H_k^{(N)} \right)^* Z_{kj}^{(N)} \left(h_j - H_j^{(N)} \right) \right) \tag{57}$$

where N indexes the gaussian. Again, choice of the $Z_{kj}^{(N)}$ are somewhat arbitrary, but each term in this sum can be evaluated through the marginalization. We might be able to decompose more general priors into this form, which will then give us an analytic forumla for the posterior in terms of the decomposition. Explicitly, this is[5]

$$\begin{aligned} p(\theta, \phi | d_\beta) \quad &= \quad p(\theta, \phi) \sum_N C_N \prod_f \exp\left(\hat{h}_j^* \left(A\right)_{jk} \hat{h}_k - \left(\hat{h}_j - H_j^{(N)} \right)^* \left(Z_{jk}^{(N)} - Z_{jm}^{(N)} \left(A + Z^{(N)} \right)^{-1}_{mn} Z_{nk}^{(N)} \right) \left(\hat{h}_k - \right.\right. \\ &\qquad \times \sqrt{(2\pi)^{N'_p} \left| \left(A + Z^{(N)} \right)^{-1}_{jk} \right| \left| Z^{(N)} \right|} \end{aligned}$$

Importantly, if we choose many narrow gaussians we can approximate an arbitrary prior (sum of δ-functions). We can also endow $Z_{jk}^{(N)}$, $H_j^{(N)}$ with frequency dependence without any major modifications. This means we can demand that there is equal energy in each frequency bin so that $h \propto 1/f$, or something similar, with appropriate definitions for $Z_{jk}^{(N)}$ and/or $H_j^{(N)}$. Note that this looks like the prior with modifications to the coefficients C_N based on the marginalization over ϵ_j. It may not be possible to re-sum these terms analytically to explicitly state the posterior in closed form (assuming we've expanded a closed form prior into gaussians). *Importantly*, this gives us a way to explicitly compute the posterior without sampling the parameters space of possible signals. All computations are done analytically assuming constant (θ, ϕ), which could provide large speed-ups computationally over Monte-Carlo Markov-Chain algorithms.

In fact, a simple error minimization argument allows us to determine the optimal coefficients for an arbitrary set of gaussian priors when they are used to approximate an arbitrary function $f(h)$. If we define each gaussian by it's mean and standard deviation as $|\mu_n, \sigma_n\rangle$, then we have the overlap matrix

$$M_{nm} = \langle \mu_n, \sigma_n | \mu_m, \sigma_m \rangle \tag{59}$$

where the inner product is defined as $\langle a|b\rangle = \int\limits_{-\infty}^{\infty} \mathrm{d}h \; a \cdot b$.

Minimizing the sum-square error of the approximation yields a set of coefficients

[5]With clever algebra, we may be able to cast this into something more recognizable. For instance, the sum of the products should be the product of the sums and we can exchange the order of the $\sum_N$ and $\prod_f$.

$$C_k = \left(M^{-1}\right)_{km} \langle \mu_m, \sigma_m | f \rangle \tag{60}$$

We can also derive the optimal placement of a fixed number of gaussians using a similar method. However, the result is not partitularly illuminating and simply selecting a reasonable set of gaussians by hand should suffice.

For the special case of zero-centered gaussians ($\mu_m = 0 \,\forall\, m$) and $f \propto h^{-4}$, then we have

$$\begin{aligned}
M_{nm} = M_{mn} &= \sqrt{\frac{2\sigma_n\sigma_m}{\sigma_n^2 + \sigma_m^2}} && (61)\\
\langle \sigma_m | f \rangle &\sim 2 \int\limits_{\epsilon << \sigma_m}^{\infty} \mathrm{d}h \; (h/h_o)^{-4} \left(\pi\sigma_m^2\right)^{-1/4} \exp\left(-\frac{h^2}{2\sigma_m^2}\right) && (62)\\
&\sim 2 \int\limits_{\epsilon/\sigma_m << 1}^{\infty} \mathrm{d}x\sigma_m \; (x\sigma_m/h_o)^{-4} \left(\pi\sigma_m^2\right)^{-1/4} \exp\left(-\frac{x^2}{2}\right) && (63)\\
&\sim \sigma_m^{-13/4} \left[2h_o^4(\pi)^{-1/4} \int\limits_{\varepsilon << 1}^{\infty} \mathrm{d}x \, x^{-4} \exp\left(-\frac{x^2}{2}\right) \right] && (64)\\
C_k &= \left(M^{-1}\right)_{km} \sigma_m^{-13/4} \cdot \mathit{constant} && (65)\\
& && (66)
\end{aligned}$$

3.3.3 other priors on h_j

Alternatively, we can write down some astrophysically motivated prior, such as uniform in co-moving volume. However, for burst signals, we do not immediately have a good estimate for the distance D. We can relate this to the observed data through

$$\frac{E_{GW}}{D_L^2} \propto \int \mathrm{d}f \, f^2 h_j^* h_j \tag{67}$$

To obtain a prior on h_j, we should marginalize over all possible D_L and E_{GW}[6]

[6] This assumes all signals have the same intrinsic *standard candle* E_{GW} regardless of their frequency content, which may not be true.

$$\begin{aligned}
p(h_j) &\propto \int p(h_j|D_L, E_{GW})p(D_L)p(E_{GW})\mathrm{d}D_L\mathrm{d}E_{GW} && (68)\\
&= \int \delta\left(D_L - \sqrt{\frac{E_{GW}}{\int \mathrm{d}f\, f^2 h_j^* h_j}}\right) p(D_L)p(E_{GW})\mathrm{d}D_L\mathrm{d}E_{GW} \\
&= \int \delta\left(D_L - \sqrt{\frac{E_{GW}}{\int \mathrm{d}f\, f^2 h_j^* h_j}}\right) (4\pi D_L^2)p(E_{GW})\mathrm{d}D_L\mathrm{d}E_{GW} \quad (\text{for } z \ll 1) \\
&= \int 4\pi \frac{E_{GW}}{\int \mathrm{d}f\, f^2 h_j^* h_j} p(E_{GW})\mathrm{d}E_{GW} \\
&= \frac{4\pi \langle E_{GW}\rangle}{\int \mathrm{d}f\, f^2 h_j^* h_j} && (69)
\end{aligned}$$

Notice that this does the something reasonable in that it prefers signals with smaller h_j, because they're likely to have come from farther away. However, this renders the posterior nearly impossible to compute analytically:

$$p(\theta, \phi|d_\beta) = p(\theta, \phi) \int \mathcal{D}\epsilon_j \exp\left(\left(\int \mathrm{d}f\, d_\beta^* B_{\beta j}^* \left(A^{-1}\right)_{jk} B_{\alpha k} d_\alpha - \epsilon_j^* A_{jk}\epsilon_k\right) \frac{4\pi \langle E_{GW}\rangle}{\int \mathrm{d}f\, f^2(\hat{h}_j^* + \epsilon_j^*)(\hat{h}_j + \epsilon_j)}\right. \tag{70}$$

In fact, we can do a little better than this by expanding the denomenator of the prior in a power series and noting that all odd powers of ϵ_j will vanish? At least this is true of the linear terms. Depending on the width A_{jk} and the energy contained in the maximum likelihood estimate, we may be able to truncate the series after a few terms, which we can evaluate analtyically.

- separate into different frequency components? $\Rightarrow$ make the measure a product?
- Can we Taylor expand the integral with respect to each frequency component?

Using this as motivation, we may be able to expand this prior into a sum of gaussians and use our previous result to evaluate the marginalization analytically.

3.3.4 singular A_{jk} and priors on h_j

If A_{jk} is singular, then the network is insensitive to at least one polarization. We can simply restrict ourselves to a subset of polarizations to which the detector network is sensitive if we properly normalize our priors so that

$$\int \mathcal{D}h_j\ p(h_j) = 1 \tag{71}$$

This means that we can simply neglect the polarizations we cannot detect without worrying about differences in normalizations between different parts of the sky. The marginalization of the polarizations to which the network is insensitive will simply contribute a factor of unity.

3.3.5 signal morphology

Alternatively, we can attempt to calculate the posterior for the actual signal h_j, which we obtain through marginalization over (θ, ϕ). This means computing the following

$$p(h_j|d_\beta) = \int \mathrm{d}\cos\theta \mathrm{d}\phi \, \frac{p(d_\beta|h_j, \theta, \phi) p(h_j, \theta, \phi)}{p(d_\beta)} \tag{72}$$

This is a much more difficult problem, and unfortunately it may not be tractable analytically. However, for each (θ, ϕ) we can compute the mean and variance of the gaussian likelihood function. Marginalization can be done numerically by simply summing over pixels. This can be done by recording the mean, covariance matrix, and amplitude of the likelihood associated with each pixel. We can then compute the posterior for many values of h_j by simply computing the gaussians and summing. Arbitrary priors can be applied to h_j in a straightforward manner a postiori. Furthermore, priors on (θ, ϕ) can also be incorporated as part of the summation without much computational complexity.

We should be careful that we add likelihoods using the same wave-frame coordinate system for all (θ, ϕ). Furthermore, we will reconstruct the distribution over the complex variables h_j, when we are most likely interested in $|h_j|$ which may require marginalization over the phase.

With this posterior in hand, we can make definite statements about the p-value associated with an observed strain that is inconsistent with no-signal.

3.4 Example: one polarization, two detectors

For a single polarization, $A_{jk} = A \rightarrow (A^{-1}) = 1/A$. Therefore, we can write our estimator as

$$\hat{h} = \frac{1}{A} B_\beta d_\beta = \left(\sum_\alpha \frac{F_\alpha F_\alpha^*}{S_\alpha} \right)^{-1} \sum_\beta \frac{F_\beta^* d_\beta}{S_\beta} \tag{73}$$

If we assume there are only two detectors with identical noise (H,L), then we have

$$\hat{h} = \frac{F_H^* d_H + F_L^* d_L}{|F_H|^2 + |F_L|^2} = h + \frac{F_H^* n_H + F_L^* n_L}{|F_H|^2 + |F_L|^2} \tag{74}$$

This means that the maximum likelihood statistic can be written as

$$\log \mathcal{L} = \frac{|F_H^* d_H + F_L d_L^*|^2}{S\left(|F_H|^2 + |F_L|^2\right)} \tag{75}$$

which is weird. We notice that there may be a very strong dependence on the source direction, which comes from amplitude-consistency checks between H and L. These checks are possible because $N_{ifos} > N_p$. Perhaps more interestingly, we can consider the marginalization with a uniform prior on h. The posterior for (θ, ϕ) in this case is

$$p(\theta, \phi|d_\beta) = p(\theta, \phi)\exp\left(\frac{|F_H^* d_H + F_L d_L^*|^2}{S\left(|F_H|^2 + |F_L|^2\right)}\right)\left[2\pi\frac{S/2}{|F_H|^2 + |F_L|^2}\right] \tag{76}$$

Notice that the term in the brackets decreases when the antenna patterns increase. This means that with a uniform prior in h, the marginalization prefers positions with low antenna patterns. This is because the errors are larger in those regions, so the marginalization picks up more weight. If we instead use a gaussian prior on h such that

$$p(h) \propto \exp\left(h^* Z h\right) = \exp\left(|h|^2/2\sigma^2\right) \tag{77}$$

we can write the posterior as

$$\begin{aligned} p(\theta, \phi|d_\beta) \quad &= \quad p(\theta, \phi)\exp\left(\frac{|F_H^* d_H + F_L d_L^*|^2}{S\left(|F_H|^2 + |F_L|^2\right)}\right)\left[2\pi\frac{S/2}{|F_H|^2 + |F_L|^2}\right] \\ &\quad \times \exp\left(-h^* Z h + h^* \frac{Z^2}{A+Z} h\right)\sqrt{\frac{A}{A+Z}} \end{aligned} \tag{78}$$

and the modification to the poserterior is

$$\exp\left(-\frac{|\hat{h}|^2}{2\sigma^2}\frac{|F_H|^2 + |F_L|^2}{|F_H|^2 + |F_L|^2 + S/2\sigma^2}\right)\sqrt{\frac{|F_H|^2 + |F_L|^2}{|F_H|^2 + |F_L|^2 + S/2\sigma^2}} \tag{79}$$

Importantly, we see that this factor seems reasonable. For each location, the numerator in the exponential's argument should be roughly the same. This means that for larger antenna patterns, the gaussian widthd increases and there is more weight assigned to that location. This gaussian weighting should overwhelm the contribution from marginalization without a prior with appropriate σ. Therefore, we expect the posterior to follow the antenna patterns for appropriate choice of σ.[7]

If we additionally assume that there is only one detector, then this further simplifies to

$$\exp\left(-\frac{|d|^2}{2\sigma^2}\frac{|F|^2}{|F|^2 + S/2\sigma^2}\right)\sqrt{\frac{|F|^2}{|F|^2 + S/2\sigma^2}} \tag{80}$$

[7]Note that in the limit of $\sigma \to \infty$, we recover the posterior obtained with a uniform prior on h, as expected.

Here, it is entirely clear that regions with large antenna patterns (with respect to $S/2\sigma^2$) are favored. This gives us modulation along the antenna patterns controlled by one parameter: σ.

3.5 Example: N polarizations, N detectors

In this example we show that the likelihood contribution $(\hat{h}_k^* A_{kj} \hat{h}_j)$ is degenerate across the entire sky when $N_{ifo} = N_p$, regardless of whether A is singular. This is a proof by example with the $N = 2$ case.

Here, we have

$$d_\beta B_{\beta j} = d_A \frac{F_{Aj}}{S_A} + d_B \frac{F_{Bj}}{S_B} = (S_A S_B)^{-1}(d_A S_B F_{Aj} + d_B S_A F_{Bj}) \tag{81}$$

This means that

$$\begin{aligned}
\hat{h}_k^* A_{kj} \hat{h}_j &= d_\beta B_{\beta j} \left(A^{-1}\right)_{jk} B_{\alpha k} d_\alpha && (82)\\
&= (S_A S_B)^{-2} \left[|d_A|^2 S_B^2 F_{Aj} \left(A^{-1}\right)_{jk} F_{Aj}^* + d_A d_B^* S_A S_B F_{Aj} \left(A^{-1}\right)_{jk} F_{Bk}^* + (A \leftrightarrow B)\right] && (83)\\
& && (84)
\end{aligned}$$

Now, we can consider the implicit sums separately. First, we have

$$\begin{aligned}
F_{Aj} \left(A^{-1}\right)_{jk} F_{Aj}^* &= (A_{00}A_{11} - A_{01}A_{10})^{-1} \left(F_{A0}A_{11}F_{A0}^* + F_{A1}A_{00}F_{A1}^* - F_{A0}A_{01}F_{A1}^* - F_{A1}A_{10}F_{A0}^*\right) && (85)\\
&= (A_{00}A_{11} - A_{01}A_{10})^{-1}(S_A S_B)^{-1} \left(|F_{A0}|^2(S_B|F_{A1}|^2 + S_A|F_{B1}|^2)\right. && (86)\\
&\qquad +|F_{A1}|^2(S_B|F_{A0}|^2 + S_A|F_{B0}|^2) && (87)\\
&\qquad -F_{A0}F_{A1}^*(S_B F_{A0}^* F_{A1} + S_A F_{B0}^* F_{B1}) && (88)\\
&\qquad \left.-F_{A1}F_{A0}^*(S_B F_{A1}^* F_{A0} + S_A F_{B1}^* F_{B0})\right) && (89)\\
&= (A_{00}A_{11} - A_{01}A_{10})^{-1}(S_B)^{-1} \left(|F_{A0}|^2|F_{B1}|^2 + |F_{A1}|^2|F_{B0}|^2\right. && (90)\\
&\qquad \left.-F_{A0}F_{A1}^*F_{B0}^*F_{B1} - F_{A0}^*F_{A1}F_{B0}F_{B1}^*\right) && (91)\\
& && (92)
\end{aligned}$$

Now, expanding the determinant of A yields

$$A_{00}A_{11} - A_{01}A_{10} = (S_A S_B)^{-1} 1 \left[|F_{A0}|^2|F_{B1}|^2 + |F_{A1}|^2|F_{B0}|^2 - F_{A0}F_{A1}^*F_{B0}^*F_{B1} - F_{A0}^*F_{A1}F_{B0}F_{B1}^*\right] \tag{93}$$

so we have

$$\begin{aligned}
F_{Aj} \left(A^{-1}\right)_{jk} F_{Aj}^* &= (S_A S_B)^{-1} 1 \left[|F_{A0}|^2|F_{B1}|^2 + |F_{A1}|^2|F_{B0}|^2 - F_{A0}F_{A1}^*F_{B0}^*F_{B1} - F_{A0}^*F_{A1}F_{B0}F_{B1}^*\right] && (94)\\
&\quad \times (S_B)^{-1} \left(|F_{A0}|^2|F_{B1}|^2 + |F_{A1}|^2|F_{B0}|^2 - F_{A0}F_{A1}^*F_{B0}^*F_{B1} - F_{A0}^*F_{A1}F_{B0}F_{B1}^*\right) && (95)\\
&= S_A && (96)\\
& && (97)
\end{aligned}$$

which is a remarkable simplification. Furthermore, we can expand the second implicit sum as

$$
\begin{aligned}
F_{Aj}\left(A^{-1}\right)_{jk}F^{*}_{Bk} &= (A_{00}A_{11}-A_{01}A_{10})^{-1}\left(F_{A0}A_{11}F^{*}_{B0}+F_{A1}A_{00}F^{*}_{B1}-F_{A0}A_{01}F^{*}_{B1}-F_{A1}A_{10}F^{*}_{B0}\right) && (98)\\
&= (A_{00}A_{11}-A_{01}A_{10})^{-1}(S_AS_B)^{-1}\left(F_{A0}F_{B0}(S_B|F_{A1}|^2+S_A|F_{B1}|^2)\right. && (99)\\
&= \qquad +F_{A1}F^{*}_{B1}(S_B|F_{A0}|^2+S_A|F_{B0}|^2) && (100)\\
&\qquad -F_{A0}F^{*}_{B1}(S_BF^{*}_{A0}F_{A1}+S_AF^{*}_{B0}F_{B1}) && (101)\\
&\qquad \left.-F_{A1}F^{*}_{B0}(S_BF^{*}_{A1}F_{A0}+S_AF^{*}_{B1}F_{B0})\right) && (102)\\
&= 0 && (103)
\end{aligned}
$$

and we see that this term simplifies even more. We can then write the total term as

$$
\begin{aligned}
\hat{h}^{*}_{k}A_{kj}\hat{h}_{j} &= d_\beta B_{\beta j}\left(A^{-1}\right)_{jk}B_{\alpha k}d_\alpha && (104)\\
&= (S_AS_B)^{-2}\left[|d_A|^2S_B^2S_A+|d_B|^2S_A^2S_B\right] && (105)\\
&= \frac{|d_A|^2}{S_A}+\frac{|d_B|^2}{S_B} && (106)
\end{aligned}
$$

and we see that the maximum likelihood estimator exactly reproduces the observed data, rendering the likelihood to a statement about the probability of gaussian noise generating the observed data. This is, clearly, uniform over the entire sky and uniformative.

While this is only proof for the $N_{ifo}=N_p=2$ case, it should be generic whenever $N_{ifo}=N_p$.

3.6 Example: 2 polarizations, 2 (nearly aligned) detectors

If we have two nearly aligned detectors and work in the dominant polarization frame, we can write

$$
A_{ij}=s_i\delta_{ij}\mid s_0\gg s_1\cap s_0+s_1=\frac{F_{A0}^2+F_{A1}^2}{S_A}+\frac{F_{B0}^2+F_{B1}^2}{S_B} \qquad (107)
$$

Notice that as long as $s_1\geq 0$ the network is still sensitive to *both* polarizations. This means that without a prior on the singal (h_j), the likelihood will be degenerate across the entire sky. This is because we can always find some combination of reconstructed strains that will exactly reproduce the observed data streams.

However, if we apply even a simple gaussian prior with zero-mean, we can break this degeneracy. This is equivalent to claiming that small-amplitude signals are more likely than large-amplitude signals, which is not an unreasonable statment.

Let us assume a simple gaussian as the prior on h_j:

$$Z_{jk} = z\delta_{ij} = \frac{1}{2\sigma^2}\delta_{ij} \tag{108}$$

and note that this prior is independent of the coordinate system chosen, ie: it is invarient under changes in ψ. If this is the case, then marginalization yields a simple posterior

$$\begin{aligned}
p(\theta,\phi|d_\beta) &= p(\theta,\phi)\prod_f \exp\left(\hat{h}_j^*\left((A-Z)_{jk} + Z_{jm}\left(A+Z\right)^{-1}_{mn} Z_{nk}\right)\hat{h}_k\right)\sqrt{(2\pi)^{N_p}\left|(A+Z)^{-1}_{jk}\right|} && (109)\\
&= p(\theta,\phi)\prod_f \exp\left(\hat{h}_j^*\left((s_j - z) + \left(\frac{z^2}{s_j+z}\right)\right)\delta_{jk}\hat{h}_k\right)\sqrt{(2\pi)^{N_p}\prod_j (s_j+z)^{-1}} && (110)\\
&= p(\theta,\phi)\prod_f \exp\left(\hat{h}_j^*\left(\frac{s_j^2}{s_j+z}\right)\delta_{jk}\hat{h}_k\right)\sqrt{(2\pi)^{N_p}\prod_j (s_j+z)^{-1}} && (111)
\end{aligned}$$

We can separate this into separate components, each depending on only a single polarization. We then have, writing the estimators explicitly in terms of the data

$$\begin{aligned}
p(\theta,\phi|d_\beta)_j &= \prod_f \exp\left(\left|\frac{B_{j\beta}d_\beta}{s_j}\right|^2\left(\frac{s_j^2}{s_j+z}\right)\right)\sqrt{(2\pi)^{N_p}(s_j+z)^{-1}} && (112)\\
&= \prod_f \exp\left(\frac{\left|B_{j\beta}d_\beta\right|^2}{s_j+z}\right)\sqrt{(2\pi)^{N_p}(s_j+z)^{-1}} && (113)\\
&= \prod_f \exp\left(\left|\frac{F_{jA}}{S_A}d_A + \frac{F_{jB}}{S_B}d_B\right|^2 (s_j+z)^{-1}\right)\sqrt{(2\pi)^{N_p}(s_j+z)^{-1}} && (114)
\end{aligned}$$

Furthermore, if we assume the detectors have similar noise curves so that $S_A = S_B = S$, we can simplify further.

$$\begin{aligned}
p(\theta,\phi|d_\beta)_j &= \prod_f \exp\left(\left|F_{jA}\frac{d_A}{\sqrt{S}} + F_{jB}\frac{d_B}{\sqrt{S}}\right|^2 \frac{1}{S\,(s_j+z)}\right)\sqrt{(2\pi)^{N_p}(s_j+z)^{-1}} && (115)\\
&= \prod_f \exp\left(\frac{\left|F_{jA}\frac{d_A}{\sqrt{S}} + F_{jB}\frac{d_B}{\sqrt{S}}\right|^2}{\mathbb{F}_j + S/2\sigma^2}\right)\sqrt{(2\pi)^{N_p}\frac{S}{\mathbb{F}_j + S/2\sigma^2}} && (116)
\end{aligned}$$

where $\mathbb{F}_j \equiv S\cdot s_j \sim O(1)$ for the large eigenvalue and $\mathbb{F}_j \ll 1$ for the small eigenvalue. We can now, naturally, explore what happens when we turn this prior on and off, which corresonds to $2\sigma^2 \leq S$ and $2\sigma^2 \gg S$, respectively.

If the *prior is turned off* ($2\sigma^2 \gg S$), we obtain

$$p(\theta,\phi|d_\beta)_j = \prod_f \exp\left(\frac{\left|F_{jA}\frac{d_A}{\sqrt{S}} + F_{jB}\frac{d_B}{\sqrt{S}}\right|^2}{\mathbb{F}_j}\right)\sqrt{(2\pi)^{N_p}\frac{S}{\mathbb{F}_j}} \tag{117}$$

and if $\mathbb{F}_j \ll 1$, then we can also expect $F_{j\beta} \ll 1$ in this coordinate system. Similarly if $\mathbb{F}_j \gg 1$. This means that the exponential factor will stay roughly of order unity and there is a strong degeneracy between all sky postions, modulated only by the marginalization which prefers regions with low antenna patterns.

However, if the *prior is turned on* ($\mathbb{F}_0 \gg S/2\sigma^2 \gg \mathbb{F}_1$), we obtain

$$p(\theta,\phi|d_\beta)_0 = \prod_f \exp\left(\frac{\left|F_{0A}\frac{d_A}{\sqrt{S}} + F_{0B}\frac{d_B}{\sqrt{S}}\right|^2}{\mathbb{F}_0}\right)\sqrt{(2\pi)^{N_p}\frac{S}{\mathbb{F}_0}} \tag{118}$$

$$p(\theta,\phi|d_\beta)_1 = \prod_f \exp\left(\frac{\left|F_{1A}\frac{d_A}{\sqrt{S}} + F_{1B}\frac{d_B}{\sqrt{S}}\right|^2}{S/2\sigma^2}\right)\sqrt{(2\pi)^{N_p}2\sigma^2} \tag{119}$$

$$\tag{120}$$

and the degeneracy between the two polarizations is clearly broken. We see that the large eigenvalue behaves essentially as if the prior was turned off, but the probability of having signal in the small eigenvalue is surpressed *exponentially* because $|F_{1\beta}|^2 \ll S/2\sigma^2$.

What this means is that the total posterior will favor locations in which the reconstructed signal is consistent with a *single polarization*, rather than two polarizations. The weight which would have come from reconstructed signal in the small eigenvalue polarization is surpressed by the prior. [8]

Therefore, we expect the network to behave as if there were only a single polarization with two detectors when we apply this prior. If the small eigenvalue is small over most of the sky, as is the case for nearly aligned detectors, then it's contribution to the total posterior is a factor near unity over the entire sky and any modulation in the posterior will be due to the reconstruction of the large eigenvalue polarization *only*. In effect, we should get the triangulation and amplitude consistency checks that naturally occur whenever $N_{ifo} > N_p$ without this being strictly true.[9]

We should note that we specified a rather specific prior: $\mathbb{F}_0 \gg S/2\sigma^2 \gg \mathbb{F}_1$. This prior says that the majority of signals are *not detectable* because they are well below the noise floor. However, the wings of the distribution do extend

[8]This is related to one of the original motivations for Coherent WaveBurst "regulators," although this formulation is somewhat better motivated than the ad hoc implementation of that algorithm.

[9]A more realistic prior decomposed into a sum of gaussians should give a similar result, although the algebra may be significantly more complicated.

to large enough strains where they are significantly larger than the minimum eigenvalue. Because the minimum eigenvalue can be made to vanish exactly for all source locations (aligned detectors), there should be some wiggle-room when defining the prior in practice. We can also define a variable width depending of frequency ($\sigma(f)$) without complicating this derivation.

We should also keep in mind that this analysis concerned nearly aligned detectors, and 2 mis-aligned detectors may behave differently.

4 triangulation and effective priors

We now discuss how some "common sense" sky localization algorithms fall out of this formalism. These include *triangulation* and *effective priors* on the source's location (θ, ϕ).

4.1 triangulation

Most "back-of-the-envelop" estimates of sky localization rely on *triangulation* using spatially separated detectors. Essentially, this is the most basic form of amplitude consistency between the detectors. An observer assumes they can associate a wiggle in one detector with a wiggle in another detector, and the difference in the times at which those wiggle occured corresponds to the time-of-flight for the wave-front to travel from one detector to the other. This is indeed a useful concept, but it is not always applicable. In particular, it relies on the fact that an observe *can confidently associate* a wiggle in one detector with a wiggle in another detector. This is equivalent to an assumption that they can identify and associate parts of the gravitational-wave signal as seen in different detectors. Also, because triangulation is essentially an amplitude check (time-of-arrivals of an identifiable portion of the gravitational wave), it behaves like a *likelihood* in the bayesian sense.

However, this can only be done when $N_{ifo} > N_p$, which is the general requirement for amplitude consistency. If $N_{ifo} < N_p$, the network will be insensitive to some of the polarizations and we should expect $N_{ifo} = (N_p)_{\text{eff}}$. When $N_{ifo} = N_p$, there is a one-to-one transformation that can recreate the observed data streams with some combination of polarizations *for each point in the sky*. Therefore, observers cannot confidently determine whether the wiggle observed in one detector comes from *the same part of the gravitational wave* as a wiggle in another detector. That means triangulation is not possible and leads to a degenerate likelihood over most of the sky.

Fortunately, the LIGO detectors are nearly aligned, which means they are effectively sensitive to a single polarization over most of the sky. Therefore, for a LHO-LLO network, $N_{ifo} = 2 > 1 = (N_p)_{\text{eff}}$. Similarly, for the LHO-LLO-Virgo network, we have 3 detectors and the condition is satisfied as well. We see that the naive assumption that triangulation will always work is fortuitously true for the actual detector networks.

We should also note that triangulation gives very different error regions depending on N_{ifo} and $(N_p)_{\rm eff}$. For the LHO-LLO network, the locus of points that gives a consistent time-of-arrival using a single polarization is an annulus. This annulus can be a great circle and will typically contain a large portion of the sky. The annulus's width is controlled by the accuracy to which observers can determine the time-of-flight between detectors. For the LHO-LLO-Virgo netowrk, assuming all three detectors participate, triangulation gives 2 points in the sky, which reflects a symmetry with respect to reflection about the plane defined by the three detectors. We expect the error region to consist of blobs centered around these points, with the blob radius determined by the accuracy to which we measure time-of-flight between the various pairs of detectors.

4.2 effective priors on (θ, ϕ)

Triangulation behaves like a likelihood, but we can apply further knowledge about the population of sources to compute a prior. The simplest prior on (θ, ϕ) would come from the assumption that the sources live in the galactic plane, or they live near galaxies. This is a direct prior on (θ, ϕ) and is not difficult to interpret.

However, we can also derive an *effective* prior on (θ, ϕ) using knowledge of the network sensitivity and the distribution of signal amplitudes. The concept is simple: if quiet signals are more likely than loud signals, then we are more likely to detect events where the network's large eigenvalue is in fact large. What this means is that the population of *detected* events will follow the antenna patterns in a predictable way, more detected signals when the large eigenvalue is large and fewer detected signals when the large eigenvalue is small. Of course, the particular functional form the distribution of detected events takes is not obvious, but a simple estimate goes as follows:

Sources are uniformly distributed in volume, independently of their morphology or intrinsic strength. However, we can relate the observed strain to the energy in the wave and the distance through

$$\frac{E_{GW}}{/} D_L^2 \int \mathrm{d}f\,(f h_j^*)(f h_j) \approx f_o^2 \int \mathrm{d}f\, h_j^* h_j \tag{121}$$

where our approximation assumes a narrow-band signal. Furthermore, working in the dominant polarization frame (so things are diagonal)

$$\begin{aligned} h_j^* h_j &\approx \hat{h}_j^* \hat{h}_j && (122)\\ &= d_\beta^* B_{\beta k}^* \left(A^*\right)^{-1}_{jk} \left(A\right)^{-1}_{ji} B_{\alpha i} d_\alpha && (123)\\ &= d_\beta^* \frac{F_{\beta j}^*}{S_\beta} s_j^{-2} \frac{F_{\alpha j}}{S_\alpha} d_\alpha && (124)\\ & && (125) \end{aligned}$$

If we assume the detectors have similar noise curves, we can approximate this further as

$$h_j^* h_j \approx d_\beta^* \frac{F_{\beta j}^*}{S} \left(\frac{S}{Ss_j}\right)^2 \frac{F_{\alpha j}}{S} d_\alpha \quad (126)$$

$$= d_\beta^* \frac{F_{\beta j}^* F_{\alpha j}}{(Ss_j)} d_\alpha \quad (127)$$

$$(128)$$

Furthermore, $s_j \sim F^*F/S \rightarrow Ss_j \sim F^*F$, and we can make the last leap to the statement that

$$h_j^* h_j \sim \frac{1}{Ss_j} \quad (129)$$

Now, we if we expect the signals to be distributed uniformly in volume, we require

$$p(D_L) \propto D_L^2 \approx \frac{E_{GW}/f_o^2}{\int \mathrm{d}f \, h_j^* h_j} \quad (130)$$

$$\sim Ss_j \sim F^*F \quad (131)$$

$$= p_{\mathrm{eff}}(\theta, \phi) \quad (132)$$

There is a considerable amount of hand-waving in this argument, and it really holds closely when $(N_p)_{\mathrm{eff}} = 1$, and there is only a single polarization. This is the case for the LHO-LLO network, and indeed we observe a distribution of detected signals that closely follows the predicted form.

More complicated and/or careful arguments may produce slightly modified effective priors, but the combination of *triangulation* and *effective priors* actually captures a good fraction of the physics associated with localizing generic gravitational wave signals. Fully coherent methods, such as the rest of this note, can perform more thorough amplitude checks (rather than just time-of-arrival) and improve upon this estimate, and it is not clear to what extent that information can improve localization.

5 Extended sources

Point sources were described in Section 3. Here we consider extended sources which may come from distant parts of the sky simultaneously. This modifies the way in which the strain enters our data streams, although all we really have to do is add a few more indicies.

5.1 maximum likelihood estimators

Consider the position-space integral

$$\int \mathrm{d}\cos\theta \mathrm{d}\phi\, F_{\beta j}\left(\theta,\phi\right) h_j\left(\theta,\phi\right) = \int \mathrm{d}\cos\theta \mathrm{d}\phi \left(\sum_{lm} Y_{lm} F_{\beta j(lm)}\right)\left(\sum_{l'm'} Y_{l'm'} h_{j(l'm')}\right) \quad (133)$$

$$= \sum_{lml'm'} F_{\beta j(lm)} h_{j(l'm')} \int \mathrm{d}\cos\theta \mathrm{d}\phi Y_{lm} Y_{l'm'} \quad (134)$$

$$= \sum_{lm} F_{\beta j(lm)} h_{j(lm)} \quad (135)$$

which is the equivalent of parseval's theorem for spherical harmonics. Now, this summation is technically over an infinite series, but we can always truncate the series at high order (l, m) to make this tractable. Furthermore, a similar summation is reasonable if we pixelate the sky in to a set of discrete points. The mathematics that follows is independent of the decomposition (position-space or spherical harmonics), so we simply refer to the postion tuple with a single *upper case greek letter*.

Now, our likelihood functional is modified as follows

$$\mathcal{L} = \frac{p(d_\beta - F_{\beta j} h_j)}{p(d_\beta)} = \exp\left(\int \mathrm{d}f \sum_\beta \frac{|d_\beta|^2 - |d_\beta - F_{\beta j\Omega} h_{j\Omega}|^2}{S_\beta}\right) \quad (136)$$

$$= \exp\left(\int \mathrm{d}f \sum_\beta \frac{d_\beta F^*_{\beta j\Omega} h^*_{j\Omega} + d^*_\beta F_{\beta j\Omega} h_{j\Omega} - h_{k\Omega} F_{\beta k\Omega} F^*_{\beta j\Psi} h^*_{j\Psi}}{S_\beta}\right) \quad (137)$$

where summation over sky positions Ω, Ψ are implied.

If we vary this functional with respect to $h^*_{j\Omega}$, we obtain the following Euler equations.

$$\sum_\beta \frac{F^*_{\beta\Omega j} d_\beta}{S_\beta} = \sum_{\beta\Psi} \frac{F^*_{\beta\Omega j} F_{\beta\Psi k} h_{k\Psi}}{S_\beta} \quad (138)$$

and we can define analogous matricies as before

$$A_{\Omega\Psi jk} \equiv \sum_\beta \frac{F^*_{\beta\Omega j} F_{\beta\Psi k}}{S_\beta} \quad (139)$$

$$B_{\beta\Omega j} \equiv \frac{F^*_{\beta\Omega j}}{S_\beta} \quad (140)$$

which allows us to solve explicitly for the maximum likelihood estimator

$$\hat{h}_{\Psi k} = \left(A^{-1}\right)_{\Psi\Omega kj} B_{\beta\Omega j} d_\beta \quad (141)$$

Note, the size of these matricies depends on the number of sky positions included. This means that the matrix inversion may be non-trivial computationally, or poorly defined in the continuum limit. However, if we limit ourselves to a finite number of sky locations this should be reasonable, if expensive.[10]

Now, the errors associated with this estimator are

$$\hat{h}_{\Psi k} = \left(A^{-1}\right)_{\Psi\Omega kj} B_{\beta\Omega j}\left(F_{\beta\Upsilon i} h_{i\Upsilon} + n_\beta\right) \tag{142}$$

$$= \left(A^{-1}\right)_{\Psi\Omega kj}\left(B_{\beta\Omega j}F_{\beta\Upsilon i}\right) h_{i\Upsilon} + \left(A^{-1}\right)_{\Psi\Omega kj} B_{\beta\Omega j} n_\beta \tag{143}$$

$$= \left(A^{-1}\right)_{\Psi\Omega kj} A_{\Omega\Upsilon ji} h_{i\Upsilon} + \left(A^{-1}\right)_{\Psi\Omega kj} B_{\beta\Omega j} n_\beta \tag{144}$$

$$\Rightarrow \hat{h}_{\Psi k} - h_{\Psi k} \equiv \epsilon_{\Psi k} = \left(A^{-1}\right)_{\Psi\Omega kj} B_{\beta\Omega j} n_\beta \tag{145}$$

and we see that the errors are gaussian around the maximum likelihood estimator.

NOTE THAT THE SKY POSTION AND THE POLARIZATION INDICIES LIKE TO GO TOGETHER. In particular, with this inversion we have a summation over both which yields the kroniker delta. This implies that we should group them together into a tuple of three, indexed by a single letter rather than two indicies.

If we plug this into our likelihood functional, we obtain the following

$$\log\mathcal{L} = \int \mathrm{d}f\, d^*_\beta B^*_{\beta\Omega j}\left(A^{-1}\right)_{\Omega\Psi jk} B_{\alpha\Psi k} d_\alpha - \epsilon^*_{k\Omega} A_{\Omega\Psi kj}\epsilon_{\Psi k} \tag{146}$$

and indeed, the probability is gaussian in the errors. Note, we are summing over all sky positions implicitly with repeated capital greek indicies.

5.2 posterior probabilities

We now ask the question, how do we compute the posterior using this likelihood, and in what way is this most efficient.

We can decompose the posterior in either position-space or spherical harmonics. This may yield simplification of the formula to compute the posterior, but it is not immediately apparent how this comes into play.

Again, we can write the posterior as

$$p(\Omega|d_\beta) = \int \mathcal{D}h_{j\Omega}\,\frac{p(d_\beta|h_{j\Omega},\Omega)p(h_{j\Omega},\Omega)}{p(d_\beta)} \tag{147}$$

and note that Ω is *fixed* in the integration? Can we represent the likelihood as a sum over independent sky position/polarization channels? If that's the case, we simply have to decompose the posterior into this diagonal basis and

[10] For any practical search, we're only interested in the approximate posterior anyway. We won't be able to resolve the source perfectly in any case, so this slight pixelization should not affect the search in any meaningful way.

compute each term. The total posterior will come from re-summing the decomposition. When we write $p(\Omega|d_\beta)$, we mean the posterior that the signal came from that sky postion and only that sky position? In which case, my nice summations over source location should be broken and we use a single sky position. Indeed, this seems reasonable. We can then use the *point estimate* forumation from this point onward. The crucial step will be when we consider $p(\theta, \phi|d_\beta)$ after computing $p(l, m|d_\beta)$, which should simply be a re-summation weighted by spherical harmonics. There should also be strong reality constraints that should help us reduce the number of spherical harmonics that need to be considered.

Actually, the summation in the integrand comes from considering all possible sky postions. This is appropriate when computing $p(h_{j\Omega}|d_\beta)$. **WE NEED TO DETERMINE THE EXTENT TO WHICH THERE IS MIXING BETWEEN POLARIZATION CHANNELS CAUSED BY THE SUMMATION OVER SKY POSITIONS.** If there is a better basis, we should work in that. What we want is to separate the sums, which makes this analysis much more straightforward.

References

[1] Warren G. Anderson, Patrick R. Brady, Jolien D. E. Creighton, and Éanna É. Flanagan. Excess power statistic for detection of burst sources of gravitational radiation. *Phys. Rev. D*, 63:042003, Jan 2001.