BAYESBurst?

Reed Essick
ressick@mit.edu

May 1, 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)
- Work out carefully and consider the case of a single polarization. Show the dependence on the antenna patterns with and without a prior on h_j.
- Work out the case when A_{jk} is singular, and how we should approach that.
  - Project onto a subspace with lower rank than N_p when reconstructing the signal?
  - When marginalizing over the possible waveforms, we simply integrate the prior for the missing polarization(s)?
- 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: 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.

3 Point sources

Here we examine a non-parametric bayesian approach to signal reconstruction for *point sources*. Extended sources are considered in Section 4.

3.1 maximum likelihood estimators

We can define the likelihood ratio as

$$\mathcal{L} = \frac{p(d_\beta - F_{\beta j} h_j)}{p(d_\beta)} = \exp\left(\int \mathrm{d}f \sum_\beta \frac{|d_\beta|^2 - |d_\beta - F_{\beta j} h_j|^2}{S_\beta} \right) \quad (4)$$

$$= \exp\left(\int \mathrm{d}f \sum_\beta \frac{d_\beta F^*_{\beta j} h^*_j + d^*_\beta F_{\beta j} h_j - h_k F_{\beta k} F^*_{\beta j} h^*_j}{S_\beta} \right) \quad (5)$$

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

$$\frac{\delta}{\delta h^*_m} \log \mathcal{L} = \frac{\delta}{\delta h^*_m} \int \mathrm{d}f \sum_\beta \frac{d_\beta F^*_{\beta j} h^*_j + d^*_\beta F_{\beta j} h_j - F_{\beta k} h_k F^*_{\beta j} h^*_j}{S_\beta} = 0 \quad (6)$$

$$\Rightarrow 0 = \frac{\mathrm{d}}{\mathrm{d}f} \left(\frac{\partial}{\partial (\mathrm{d}h^*_m / \mathrm{d}f)} \sum_\beta \frac{d_\beta F^*_{\beta j} h^*_j + d^*_\beta F_{\beta j} h_j - F_{\beta k} h_k F^*_{\beta j} h^*_j}{S_\beta} \right)$$
$$- \frac{\mathrm{d}}{\mathrm{d}h^*_m} \left(\sum_\beta \frac{d_\beta F^*_{\beta j} h^*_j + d^*_\beta F_{\beta j} h_j - F_{\beta k} h_k F^*_{\beta j} h^*_j}{S_\beta} \right) \quad (7)$$

$$= -\sum_\beta \frac{\left(d_\beta F^*_{\beta k} - h_j F_{\beta j} F^*_{\beta k} \right) \delta_{km}}{S_\beta} \quad (8)$$

$$\Rightarrow \sum_\beta \frac{d_\beta F^*_{\beta k}}{S_\beta} = \sum_\beta \frac{F^*_{\beta k} F_{\beta j} h_j}{S_\beta} \quad (9)$$

We now make the following definitions of convenience

$$A_{kj} \equiv \sum_\beta \frac{F^*_{\beta k} F_{\beta j}}{S_\beta} = A^*_{jk} \quad (10)$$

$$B_{j\beta} \equiv \frac{F^*_{\beta j}}{S_\beta} \qquad \text{(no sum over } \beta\text{)} \quad (11)$$

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

$$\begin{aligned} A_{kj}h_j &= B_{k\beta}d_\beta && (12) \\ \Rightarrow \hat{h}_j &= \left(A^{-1}\right)_{jk} B_{k\beta}d_\beta && (13) \end{aligned}$$

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

$$\begin{aligned} \hat{h}_j &= \left(A^{-1}\right)_{jk} B_{k\beta}\left(F_{\beta m}h_m + n_\beta\right) && (14) \\ &= \left(A^{-1}\right)_{jk}\left(B_{k\beta}F_{\beta m}\right)h_m + \left(A^{-1}\right)_{jk} B_{k\beta}n_\beta && (15) \end{aligned}$$

We note that

$$B_{k\beta}F_{\beta m} = \sum_\beta \frac{F^*_{\beta k}}{S_\beta}F_{\beta_m} = A_{km} \tag{16}$$

which yields the pleasant simplification

$$\begin{aligned} \hat{h}_j &= \left(A^{-1}\right)_{jk} A_{km}h_m + \left(A^{-1}\right)_{jk} B_{k\beta}n_\beta && (17) \\ &= \delta_{jm}h_m + \left(A^{-1}\right)_{jk} B_{k\beta}n_\beta && (18) \\ \Rightarrow h_j - \hat{h}_j \equiv \epsilon_j &= \left(A^{-1}\right)_{jk} B_{k\beta}n_\beta && (19) \end{aligned}$$

and we see that the estimator is *unbiased* with gaussian errors around the actual signal.[1] Explicitly, we can compute the expected distributions of the reconstructed errors as

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

$$
\begin{aligned}
\langle \epsilon_j \rangle &= \left\langle \left(A^{-1}\right)_{jk} B_{k\beta} n_\beta \right\rangle && (20)\\
&= \left(A^{-1}\right)_{jk} B_{k\beta} \langle n_\beta \rangle && (21)\\
&= 0 && (22)\\
\langle \epsilon_j^* \epsilon_k \rangle &= \left\langle n_\alpha^* B_{n\alpha}^* \left(A^{-1}\right)_{jn}^* \left(A^{-1}\right)_{km} B_{m\beta} n_\beta \right\rangle && (23)\\
&= B_{n\alpha}^* \left(A^{-1}\right)_{jn}^* \left(A^{-1}\right)_{km} B_{m\beta} \langle n_\alpha^* n_\beta \rangle && (24)\\
&= B_{n\alpha}^* \left(A^{-1}\right)_{jn}^* \left(A^{-1}\right)_{km} B_{m\beta} \left(\frac{1}{2} S_\beta \delta_{\alpha\beta}\right) && (25)\\
&= \frac{1}{2} \left(A^{-1}\right)_{jn}^* \left(A^{-1}\right)_{km} \sum_\beta \left\{ B_{n\beta}^* B_{m\beta} S_\beta \right\} && (26)\\
&= \frac{1}{2} \left(A^{-1}\right)_{jn}^* \left(A^{-1}\right)_{km} A_{mn} && (27)\\
&= \frac{1}{2} \left(A^{-1}\right)_{jn}^* \delta_{kn} && (28)\\
&= \frac{1}{2} \left(A^{-1}\right)_{jk}^* && (29)
\end{aligned}
$$

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

$$
\begin{aligned}
\mathcal{L} &= \exp\left(\int \mathrm{d}f \sum_\beta \frac{|d_\beta|^2 - \left| d_\beta - F_{\beta j}\left(\hat{h}_j + \epsilon_j\right)\right|^2}{S_\beta} \right) && (30)\\
&= \exp\left(\int \mathrm{d}f\, d_\beta B_{\beta j} \left(\hat{h}_j + \epsilon_j\right)^* + d_\beta^* B_{\beta j}^* \left(\hat{h}_j + \epsilon_j\right) - \left(\hat{h}_j + \epsilon_j\right)^* A_{jk} \left(\hat{h}_k + \epsilon_k\right) \right) && (31)\\
&= \exp\left(\int \mathrm{d}f\, d_\beta B_{\beta j} \left(A^{-1}\right)_{jk}^* B_{k\alpha}^* d_\alpha^* + d_\beta^* B_{\beta j}^* \left(A^{-1}\right)_{jk} B_{k\alpha} d_\alpha - d_\beta^* B_{m\beta}^* \left(A^{-1}\right)_{mj}^* A_{jk} \left(A^{-1}\right)_{kn} B_{n\alpha} d_\alpha \right.\\
&\quad \left. - \epsilon_j^* A_{jk} \epsilon_k + d_\beta B_{\beta j} \epsilon_j^* + d_\beta^* B_{\beta j}^* \epsilon_j - d_\beta^* B_{m\beta}^* \left(A^{-1}\right)_{jm}^* A_{jk} \epsilon_k - \epsilon_j^* A_{jk} \left(A^{-1}\right)_{kn} B_{n\alpha} d_\alpha \right) && (32)
\end{aligned}
$$

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

$$
\mathcal{L} = \frac{p(d_\beta - F_{\beta j} h_j)}{p(d_\beta)} = \exp\left(\int \mathrm{d}f\, d_\beta^* B_{\beta j}^* \left(A^{-1}\right)_{jk} B_{k\alpha} d_\alpha - \epsilon_j^* A_{jk} \epsilon_k \right) \qquad (33)
$$

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

3.2 posterior probabilities

3.2.1 sky position

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

$$p(h_j, \theta, \phi | d_\beta) = \frac{p(d_\beta | h_j, \theta, \phi) p(h_j, \theta, \phi)}{p(d_\beta)} \tag{34}$$

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

$$p(\theta, \phi | d_\beta) = \int \mathcal{D}h_j \, \frac{p(d_\beta | h_j, \theta, \phi) p(h_j, \theta, \phi)}{p(d_\beta)} \tag{35}$$

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

$$\begin{aligned} p(\theta, \phi | d_\beta) &= \int \mathcal{D}\epsilon_j \exp\left(\left(\int \mathrm{d}f \, d^*_\beta B^*_{\beta j} \left(A^{-1}\right)_{jk} B_{k\alpha} d_\alpha - \epsilon^*_j A_{jk} \epsilon_k \right) p(h, \theta, \phi) \right. && (36) \\ &= \prod_f \exp\left(d^*_\beta B^*_{\beta j} \left(A^{-1}\right)_{jk} B_{k\alpha} d_\alpha \right) \int_{-\infty}^{\infty} \mathrm{d}^{N_p} \epsilon_j \exp\left(-\epsilon^*_j A_{jk} \epsilon_k\right) p(\hat{h}_j + \epsilon_j, \theta, \phi) && (37) \end{aligned}$$

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

$$p(h_j, \theta, \phi) = p(h_j) p(\theta, \phi) = p(h_j) \frac{1}{4\pi} \tag{38}$$

where we've assumed the prior on (θ, ϕ) is uniform across the sky (constant probability per steradian)[2] Furthermore, we can assume the most uninformative prior on h_j, so that

$$p(h_j) = constant \tag{39}$$

Under these assumptions[3], we have a very simple form for the posterior

[2]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.

[3]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.

$$p(\theta,\phi|d_\beta) = \frac{constant}{4\pi}\prod_f \exp\left(d_\beta^* B_{\beta j}^* \left(A^{-1}\right)_{jk} B_{k\alpha} d_\alpha\right) \int_{-\infty}^{\infty} \mathrm{d}^{N_p}\epsilon_j \exp\left(-\epsilon_j^* A_{jk}\epsilon_k\right) \tag{40}$$

$$= \frac{constant}{4\pi}\prod_f \exp\left(d_\beta^* B_{\beta j}^* \left(A^{-1}\right)_{jk} B_{k\alpha} d_\alpha\right) \sqrt{(2\pi)^{N_p}\left|\left(A^{-1}\right)_{jk}\right|} \tag{41}$$

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 $N_p = N_{ifo}$.

3.2.2 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

$$p(h_j) \propto \exp\left(-h_k^* Z_{kj} h_j\right) \tag{42}$$

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

$$
\begin{aligned}
p(\theta,\phi|d_\beta) &\propto p(\theta,\phi)\int \mathcal{D}h_j \exp\left(\int \mathrm{d}f\, \hat{h}_j^* A_{jk}\hat{h}_k - \epsilon_j^* A_{jk}\epsilon_k - \hat{h}_j^* Z_{jk}\hat{h}_k - \hat{h}_j^* Z_{jk}\epsilon_k - \epsilon_k^* Z_{jk}\hat{h}_k - \epsilon_j^* Z_{jk}\epsilon_k\right) \\
&= p(\theta,\phi)\int \mathcal{D}h_j \exp\left(\int \mathrm{d}f\, \hat{h}_j^* \left(A-Z\right)_{jk}\hat{h}_k - \left[\epsilon_j^* \left(A+Z\right)_{jk}\epsilon_k + \hat{h}_j^* Z_{jk}\epsilon_k + \epsilon_k^* Z_{jk}\hat{h}_k^*\right]\right) \\
&= p(\theta,\phi)\int \mathcal{D}h_j \exp\left(\int \mathrm{d}f\, \hat{h}_j^* \left(A-Z\right)_{jk}\hat{h}_k + \zeta_j^*\Phi_{jk}\zeta_k - \left(\zeta_j+\epsilon_j\right)^*\Phi_{jk}\left(\zeta_k+\epsilon_k\right)\right) \qquad (43)
\end{aligned}
$$

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

$$
\begin{aligned}
p(\theta,\phi|d_\beta) &= p(\theta,\phi)\prod_f \exp\left(\hat{h}_j^* \left(A-Z\right)_{jk}\hat{h}_k + \zeta_j^*\Phi_{jk}\zeta_k\right)\sqrt{(2\pi)^{N_p}\left|\Phi^{-1}\right|} \qquad (44) \\
&= p(\theta,\phi)\prod_f \exp\left(\hat{h}_j^* \left(A-Z\right)_{jk}\hat{h}_k + \hat{h}_m^* Z_{mj}\left(A+Z\right)^{-1}_{jk} Z_{kn}\hat{h}_n\right)\sqrt{(2\pi)^{N_p}\left|\left(A+Z\right)^{-1}_{jk}\right|} \\
&= p(\theta,\phi)\prod_f \exp\left(\hat{h}_j^* \left(\left(A-Z\right)_{jk} + Z_{jm}\left(A+Z\right)^{-1}_{mn} Z_{nk}\right)\hat{h}_k\right)\sqrt{(2\pi)^{N_p}\left|\left(A+Z\right)^{-1}_{jk}\right|} \qquad (45)
\end{aligned}
$$

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

$$
p(h_j) \propto \exp\left(-(h_k-H_k)^* Z_{kj}(h_j-H_j)\right) \qquad (46)
$$

we obtain the following

$$
p(\theta,\phi|d_\beta) = p(\theta,\phi)\prod_f \exp\left(\hat{h}_j^* \left(A\right)_{jk}\hat{h}_k - \left(\hat{h}_j-H_j\right)^*\left(Z_{jk} - Z_{jm}\Phi^{-1}_{mn}Z_{nk}\right)\left(\hat{h}_k+H_k\right)\right)\sqrt{(2\pi)^{N_p}\left|\Phi^{-1}\right|} \qquad (47)
$$

Furthermore, we may consider linear combinations of gaussians

$$
p(h_j) = \sum_N C_N \cdot \exp\left(-\left(h_k - H_k^{(N)}\right)^* Z_{kj}^{(N)}\left(h_j - H_j^{(N)}\right)\right) \qquad (48)
$$

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[4]

[4] 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 $\prod_f$.

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

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.

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

$$\frac{E_{GW}}{D_L^2} \propto \int \mathrm{d}f\, f^2 h_j^* h_j \tag{50}$$

To obtain a prior on h_j, we should marginalize over all possible D_L and E_{GW}[5]

$$\begin{aligned}
p(h_j) &\propto \int p(h_j|D_L, E_{GW})p(D_L)p(E_{GW})\mathrm{d}D_L\mathrm{d}E_{GW} \qquad (51)\\
&= \int \delta\left(D_L - \sqrt{\frac{E_{GW}}{\int \mathrm{d}f\, f^2 h_j^* h_j}}\right)p(D_L)p(E_{GW})\mathrm{d}D_L\mathrm{d}E_{GW}\\
&= \int \delta\left(D_L - \sqrt{\frac{E_{GW}}{\int \mathrm{d}f\, f^2 h_j^* h_j}}\right)(4\pi D_L^2)p(E_{GW})\mathrm{d}D_L\mathrm{d}E_{GW} \quad (\text{for } z \ll 1)\\
&= \int 4\pi\frac{E_{GW}}{\int \mathrm{d}f\, f^2 h_j^* h_j}p(E_{GW})\mathrm{d}E_{GW}\\
&= \frac{4\pi\left\langle E_{GW}\right\rangle}{\int \mathrm{d}f\, f^2 h_j^* h_j} \qquad (52)
\end{aligned}$$

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

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:

$$p(\theta,\phi|d_\beta) = p(\theta,\phi)\int \mathcal{D}\epsilon_j \exp\left(\int \mathrm{d}f\, d^*_\beta B^*_{\beta j}\left(A^{-1}\right)_{jk} B_{\alpha k} d_\alpha - \epsilon^*_j A_{jk}\epsilon_k\right)\frac{4\pi\left\langle E_{GW}\right\rangle}{\int \mathrm{d}f\, f^2(\hat{h}^*_j+\epsilon^*_j)(\hat{h}_j+\epsilon_j)} \tag{53}$$

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.

3.2.3 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

$$p(h_j|d_\beta) = \int \mathrm{d}\cos\theta \mathrm{d}\phi\, \frac{p(d_\beta|h_j,\theta,\phi)p(h_j,\theta,\phi)}{p(d_\beta)} \tag{54}$$

This is a much more difficult problem, and unfortunately it may not be tractable analytically. We can always compute a *p-value* for the null hypothesis : $p(h_j = 0\,\forall\, f|d_\beta)$, but this is simply a statement about the likelihood of a particular noise realization. The more interesting computation is to find the maximum (maxima?) of the posterior and set confidence regions around it (them). An all-sky search can then be performed by monitoring the lower bound on this confidence region. Furthermore, because we have not assumed anything about the *shape* of the waveform, this posterior should cover the entire signal-space and include every possible waveform.

3.3 Singular antenna matricies A_{jk} and reduction of number of polarizations

FIGURE OUT HOW TO DO THIS TRANSPARENTLY AND WRITE UP

3.4 Example: single polarization

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

$$\hat{h} = \frac{1}{A} B_\beta d_\beta = \left(\sum_\alpha \frac{F_\alpha F_\alpha^*}{S_\alpha} \right)^{-1} \sum_\beta \frac{F_\beta^* d_\beta}{S_\beta} \tag{55}$$

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

$$\hat{h} = \frac{F_H^* d_H + F_L^* d_L}{|F_H|^2 + |F_L|^2} = h + \frac{F_H^* n_H + F_L^* n_L}{|F_H|^2 + |F_L|^2} \tag{56}$$

This means that the maximum likelihood statistic can be written as

$$\log \mathcal{L} = \frac{|F_H^* d_H + F_L d_L^*|^2}{S\left(|F_H|^2 + |F_L|^2\right)} \tag{57}$$

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

$$p(\theta, \phi | d_\beta) = p(\theta, \phi) \exp\left(\frac{|F_H^* d_H + F_L d_L^*|^2}{S\left(|F_H|^2 + |F_L|^2\right)} \right) \left[2\pi \frac{S/2}{|F_H|^2 + |F_L|^2} \right] \tag{58}$$

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

$$p(h) \propto \exp\left(h^* Z h\right) = \exp\left(|h|^2/2\sigma^2\right) \tag{59}$$

we can write the posterior as

$$\begin{aligned} p(\theta, \phi | d_\beta) &= p(\theta, \phi) \exp\left(\frac{|F_H^* d_H + F_L d_L^*|^2}{S\left(|F_H|^2 + |F_L|^2\right)} \right) \left[2\pi \frac{S/2}{|F_H|^2 + |F_L|^2} \right] \\ &\quad \times \exp\left(-h^* Z h + h^* \frac{Z^2}{A + Z} h \right) \sqrt{\frac{A}{A + Z}} \end{aligned} \tag{60}$$

and the modification to the poserterior is

$$\exp\left(-\frac{|\hat{h}|^2}{2\sigma^2} \frac{|F_H|^2 + |F_L|^2}{|F_H|^2 + |F_L|^2 + S/2\sigma^2} \right) \sqrt{\frac{|F_H|^2 + |F_L|^2}{|F_H|^2 + |F_L|^2 + S/2\sigma^2}} \tag{61}$$

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.[6]

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

$$\exp\left(-\frac{|d|^2}{2\sigma^2}\frac{|F|^2}{|F|^2+S/2\sigma^2}\right)\sqrt{\frac{|F|^2}{|F|^2+S/2\sigma^2}} \tag{62}$$

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.

4 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.

4.1 maximum likelihood estimators

Consider the position-space integral

$$\begin{aligned}
\int \mathrm{d}\cos\theta\mathrm{d}\phi\, F_{\beta j}\left(\theta,\phi\right) h_j\left(\theta,\phi\right) &= \int \mathrm{d}\cos\theta\mathrm{d}\phi \left(\sum_{lm} Y_{lm}F_{\beta j(lm)}\right)\left(\sum_{l'm'} Y_{l'm'}h_{j(l'm')}\right) && (63)\\
&= \sum_{lml'm'} F_{\beta j(lm)}h_{j(l'm')}\int \mathrm{d}\cos\theta\mathrm{d}\phi Y_{lm}Y_{l'm'} && (64)\\
&= \sum_{lm} F_{\beta j(lm)}h_{j(lm)} && (65)
\end{aligned}$$

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.*

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

Now, our likelihood functional is modified as follows

$$\mathcal{L} = \frac{p(d_\beta - F_{\beta j} h_j)}{p(d_\beta)} = \exp\left(\int \mathrm{d}f \sum_\beta \frac{|d_\beta|^2 - |d_\beta - F_{\beta j\Omega} h_{j\Omega}|^2}{S_\beta}\right) \tag{66}$$

$$= \exp\left(\int \mathrm{d}f \sum_\beta \frac{d_\beta F^*_{\beta j\Omega} h^*_{j\Omega} + d^*_\beta F_{\beta j\Omega} h_{j\Omega} - h_{k\Omega} F_{\beta k\Omega} F^*_{\beta j\Psi} h^*_{j\Psi}}{S_\beta}\right) \tag{67}$$

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.

$$\sum_\beta \frac{F^*_{\beta\Omega j} d_\beta}{S_\beta} = \sum_{\beta\Psi} \frac{F^*_{\beta\Omega j} F_{\beta\Psi k} h_{k\Psi}}{S_\beta} \tag{68}$$

and we can define analogous matricies as before

$$A_{\Omega\Psi jk} \equiv \sum_\beta \frac{F^*_{\beta\Omega j} F_{\beta\Psi k}}{S_\beta} \tag{69}$$

$$B_{\beta\Omega j} \equiv \frac{F^*_{\beta\Omega j}}{S_\beta} \tag{70}$$

which allows us to solve explicitly for the maximum likelihood estimator

$$\hat{h}_{\Psi k} = \left(A^{-1}\right)_{\Psi\Omega kj} B_{\beta\Omega j} d_\beta \tag{71}$$

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.[7]

Now, the errors associated with this estimator are

$$\hat{h}_{\Psi k} = \left(A^{-1}\right)_{\Psi\Omega kj} B_{\beta\Omega j} \left(F_{\beta\Upsilon i} h_{i\Upsilon} + n_\beta\right) \tag{72}$$

$$= \left(A^{-1}\right)_{\Psi\Omega kj} \left(B_{\beta\Omega j} F_{\beta\Upsilon i}\right) h_{i\Upsilon} + \left(A^{-1}\right)_{\Psi\Omega kj} B_{\beta\Omega j} n_\beta \tag{73}$$

$$= \left(A^{-1}\right)_{\Psi\Omega kj} A_{\Omega\Upsilon ji} h_{i\Upsilon} + \left(A^{-1}\right)_{\Psi\Omega kj} B_{\beta\Omega j} n_\beta \tag{74}$$

$$\Rightarrow \hat{h}_{\Psi k} - h_{\Psi k} \equiv \epsilon_{\Psi k} = \left(A^{-1}\right)_{\Psi\Omega kj} B_{\beta\Omega j} n_\beta \tag{75}$$

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

[7]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.

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

$$\log \mathcal{L} = \int \mathrm{d}f \, d^*_\beta B^*_{\beta\Omega j} \left(A^{-1}\right)_{\Omega\Psi jk} B_{\alpha\Psi k} d_\alpha - \epsilon^*_{k\Omega} A_{\Omega\Psi kj} \epsilon_{\Psi k} \tag{76}$$

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

4.2 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

$$p(\Omega|d_\beta) = \int \mathcal{D}h_{j\Omega} \, \frac{p(d_\beta|h_{j\Omega}, \Omega) p(h_{j\Omega}, \Omega)}{p(d_\beta)} \tag{77}$$

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 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.