NOTES ON LIKELIHOOD HETERODYNING

A PREPRINT

Alex Leviyev[1] and Aaron Zimmerman[1]

[1]Center for Gravitational Physics, University of Texas at Austin

May 29, 2023

ABSTRACT

This is a set of notes which explores in more detail the strategy of reducing the computational cost of evaluating the gravitational wave likelihood with a strategy known as heterodyning. Our aim is to provide concrete results and algorithms to supplement the ideas presented in the original paper.

1 Relative binning

1.1 Motivation

In GW data analysis, we work with a complex inner product space $(\mathbb{C}^n, \langle\cdot,\cdot\rangle_{\text{GW}})$, where for $h, g \in \mathbb{C}^n$, $\langle\cdot,\cdot\rangle_{\text{GW}} : \mathbb{C}^n \times \mathbb{C}^n \to \mathbb{C}$ is a sesquilinear inner product defined by:

$$\langle h, g\rangle_{\text{GW}} \coloneqq 4\sum_{i=1}^{n} \frac{\tilde{h}(f_i)^* \tilde{g}(f_i)}{S(f_i)} \Delta f, \tag{1}$$

where $\Delta f = 1/T$, where T is the duration of the signal, $\tilde{h} \coloneqq \text{DFT}(h)$, and $S(f)$ is a function which weighs each contribution to the sum (called the *power spectral density.*) Recall that a sesquilinear inner product satisfies the following three axioms:

- $\langle f, g\rangle = \langle g, f\rangle^*$, (conjugate symmetry)
- $\langle f, g + \alpha h\rangle = \langle f, g\rangle + \alpha\langle f, h\rangle$, (linearity in second argument)
- $\langle h, h\rangle > 0$ for $h \neq 0_{\mathbb{C}^n}$. (postive definiteness)

Remark 1.1. In physics it is typical to require linearity in the second argument. Hence we suggest sticking to this convention.

The likelihood is defined in terms of the inner product by

$$\mathcal{L}(\theta) \coloneqq e^{-\frac{1}{2}\mathfrak{Re}\left\langle h(\cdot;\theta)-d(\cdot), h(\cdot;\theta)-d(\cdot)\right\rangle_{\text{GW}}}, \tag{2}$$

where $d \in \mathbb{C}^n$ is the detector strain, and h is a waveform model parameterized by θ (e.g, IMRPhenomD). Using the linearity and conjugate symmetry of the inner product, we may show that[1]

$$\begin{aligned}
-\ln\mathcal{L}(\theta) &= \frac{1}{2}\mathfrak{Re}\big\langle h(\theta) - d, h(\theta) - d\big\rangle_{\text{GW}} \\
&= \frac{1}{2}\overbrace{\langle h(\theta), h(\theta)\rangle_{\text{GW}}}^{(a)} - \mathfrak{Re}\overbrace{\langle h(\theta), d\rangle_{\text{GW}}}^{(b)} + \frac{1}{2}\overbrace{\langle d, d\rangle_{\text{GW}}}^{(c)}.
\end{aligned}$$

Hence, for every $\mathcal{L}(\theta)$ evaluation, we must evaluate (a) and (b), which require $\mathcal{O}(n)$ evaluations of the waveform. Note that (c) can be precomputed once and stored. Relative binning aims to reduce the number of evaluations to $\mathcal{O}(m)$, where $m \ll n$, providing significant computational advantage.

[1] We use the notation $h(\cdot, \theta)$ to indicate a $\mathbb{C}^n$ vector paramerized by θ. In the following, we suppress the first slot which indexes the vector when convenient.

1.2 The heterodyning strategy

Heterodyning is a technique in signal processing which combines two high frequency signals to yield another signal with lower frequency. If we can modify the integrand of Eq. (1) such that it turns from a highly oscillating function into a slowly oscillating one, then we may in principle use a coarser grid in the sum (yielding computational benefits) while accumulating a minimal amount of error.

Consider the following. Suppose that θ_0 is a fiducial set of parameters, and that we are interested in a signal with support $[f_{\text{min}}, f_{\text{max}}]$. Then we define the *heterodyned signal* as

$$\begin{aligned} r(f;\theta) &:= \frac{h(f;\theta)}{h(f;\theta_0)} \\ &= \frac{A}{A_0} e^{i \overbrace{\big(\psi(f;\theta) - \psi(f;\theta_0)\big)}^{:=\Psi(f;\theta)}}, \end{aligned}$$

where $\Psi(f;\theta)$ is referred to as the *differential phase*. This expression is convenient since it decouples oscillations in the envelope A/A_0 and broad oscillations controlled by the phasor. Indeed, this function lives up to its name if both the envelope and differential phase are "slowly varying" functions with respect to f.

Example 1.2 (TaylorF2). In the simple case of TaylorF2, the envelope A/A_0 is not a function of frequency. On the other hand, it may be shown that the differential phase evolves as a polynomial. Hence, we expect the oscillatory behavior of r to be much milder than that of the original strain h.

Remark 1.3. Since r is parameterized by θ, this slow variation condition must hold for a reasonably large subset of parameter space Θ in order to be useful. However, analytically verifying this condition is inconvenient as it depends on the particular waveform family used to generate h. This assumption is therefore usually justified in retrospect by numerical experiements.

In Section 2, we will describe a scheme to construct a substantially sparser grid (compared to the original one) to automatically resolve r.

If this condition holds, then one may efficiently compress the real and imaginary parts of r with a *linear spline*.

Definition 1. Let $r : [f_{\text{min}}, f_{\text{max}}] \to \mathbb{C}$, and consider the cover $[f_{\text{min}}, f_{\text{max}}] = \cup_{i=1}^{m-1} [f_-^{(i)}, f_+^{(i)}) \cup [f_-^{(m)}, f_+^{(m)}]$, where $f_+^{(i)} = f_-^{(i+1)}$ for every $i \in \{1, \dots, m-1\}$, with $f_-^{(1)} = f_{\text{min}}$ and $f_+^{(m)} = f_{\text{max}}$. Then the *linear spline approximation* of r is given by

$$r(f;\theta) \approx r_0^{(b)}(\theta) + r_1^{(b)}(\theta)(f - f_-^{(b)}), \quad f \in \text{bin } b \tag{3}$$

where

$$r_0^{(b)}(\theta) := r(f_-^{(b)};\theta), \qquad r_1^{(b)}(\theta) := \frac{r(f_+^{(b)};\theta) - r(f_-^{(b)};\theta)}{f_+^{(b)} - f_-^{(b)}},$$

Proposition 1.4. *Suppose $r(f;\theta)$ is successfully heterodyned with fiducial parameters θ_0. Then with a first order spline the following approximations hold*

$$\begin{aligned} \langle h(\cdot,\theta), d(\cdot)\rangle &\approx \sum_b \big[A_0^{(b)} r_0^{(b)}(\theta)^* + A_1^{(b)} r_1^{(b)}(\theta)^*\big], \\ \langle h(\cdot,\theta), h(\cdot,\theta)\rangle &\approx \sum_b \Big[B_0^{(b)} |r_0^{(b)}(\theta)|^2 + 2B_1^{(b)} \mathfrak{Re}\big(r_0^{(b)}(\theta)^* r_1^{(b)}(\theta)\big)\Big], \end{aligned}$$

where for every bin b, the coefficients $A_0^{(b)}, A_1^{(b)} \in \mathbb{C}$ and $B_0^{(b)}, B_1^{(b)} \in \mathbb{R}$ are defined by

$$A_0^{(b)} := 4 \sum_{i: f_i \in \textit{bin } b} \frac{h(f_i;\theta_0)^* d(f_i)}{S(f_i)} \Delta f, \qquad A_1^{(b)} = 4 \sum_{i: f_i \in \textit{bin } b} \frac{h(f_i;\theta_0)^* d(f_i)(f_i - f_-^{(b)})}{S(f_i)} \Delta f,$$

$$B_0^{(b)} := 4 \sum_{i: f_i \in \textit{bin } b} \frac{|h(f_i;\theta_0)|^2}{S(f_i)} \Delta f, \qquad B_1^{(b)} := 4 \sum_{i: f_i \in \textit{bin } b} \frac{|h(f_i;\theta_0)|^2 (f_i - f_-^{(b)})}{S(f_i)} \Delta f.$$

Proof. (i) We have

$$\begin{aligned}
\langle h(\cdot,\theta), d(\cdot)\rangle &= \langle h(\cdot,\theta_0) r(\cdot,\theta), d(\cdot)\rangle \\
&\coloneqq 4\sum_b \sum_{i:f_i\in\text{bin } b} \frac{h(f_i;\theta_0)^* r(f_i;\theta)^* d(f_i)}{S(f_i)}\Delta f \\
&\approx 4\sum_b \sum_{i:f_i\in\text{bin } b} \frac{h(f_i;\theta_0)^* d(f_i)}{S(f_i)}\big[r_0^{(b)}(\theta)^* + r_1^{(b)}(\theta)^*(f_i - f_-^{(b)})\big]\Delta f \\
&= \sum_b \big[A_0^{(b)} r_0^{(b)}(\theta)^* + A_1^{(b)} r_1^{(b)}(\theta)^*\big].
\end{aligned}$$

(ii) On the other hand, we have

$$\begin{aligned}
\langle h(\cdot,\theta), h(\cdot,\theta)\rangle &= \langle h(\cdot,\theta_0) r(\cdot,\theta), h(\cdot,\theta_0) r(\cdot,\theta)\rangle \\
&\approx 4\sum_b \sum_{i:f_i\in\text{bin } b} \frac{|h(f_i;\theta_0)|^2 \big|r_0^{(b)}(\theta) + r_1^{(b)}(\theta)(f_i - f_-^{(b)})\big|^2}{S(f_i)}\Delta f \\
&= 4\sum_b \sum_{i:f_i\in\text{bin } b} \frac{|h(f_i;\theta_0)|^2 \Big(|r_0^{(b)}(\theta)|^2 + |r_1^{(b)}(\theta)(f - f_-^{(b)})|^2 + 2(f_i - f_-^{(b)})\mathfrak{Re}\big(r_0^{(b)}(\theta)^* r_1^{(b)}(\theta)\big)\Big)}{S(f_i)}\Delta f \\
&\approx \sum_b \Big[B_0^{(b)} |r_0^{(b)}(\theta)|^2 + 2B_1^{(b)} \mathfrak{Re}\big(r_0^{(b)}(\theta)^* r_1^{(b)}(\theta)\big)\Big],
\end{aligned}$$

where in the second to last line we've used the fact that for $z_1, z_2 \in \mathbb{C}$, $|z_1 + z_2|^2 = |z_1|^2 + |z_2|^2 + 2\mathfrak{Re}(z_1^* z_2)$, and in the last line we've ignored terms of $\mathcal{O}(|f - f_-^{(b)}|^2)$. $\square$

1.3 Heterodyning the gradient and Fisher

We may also investigate whether whether heterodyning produces accurate gradient and Fisher matrix evaluations. The expressions for this are given by

$$\begin{aligned}
-\ln \mathcal{L}_{,i}(\theta) &= \mathfrak{Re}\langle h_{,i}(\cdot,\theta), h(\cdot,\theta)\rangle - \mathfrak{Re}\langle h_{,i}(\cdot,\theta), d(\cdot)\rangle && \text{(heterodyne)} \\
&= \mathfrak{Re}\langle h_{,i}(\cdot,\theta), h(\cdot,\theta) - d(\cdot)\rangle, && \text{(standard)}
\end{aligned}$$

and

$$\begin{aligned}
-\ln \mathcal{L}_{,ij}(\theta) &= \mathfrak{Re}\langle h_{,ij}(\cdot,\theta), h(\cdot,\theta) - d(\cdot)\rangle + \mathfrak{Re}\langle h_{,i}(\cdot,\theta), h_{,j}(\cdot,\theta)\rangle \\
&\approx \mathfrak{Re}\langle h_{,i}(\cdot,\theta), h_{,j}(\cdot,\theta)\rangle.
\end{aligned}$$

We define

$$\Gamma_{ij}(\theta) \coloneqq \mathfrak{Re}\langle h_{,i}(\cdot,\theta), h_{,j}(\cdot,\theta)\rangle \tag{4}$$

as the *Fisher information matrix*. From the previous derivation, we see that the Fisher matrix acts as a surrogate for the Hessian of the potential. In fact, we may use it as a replacement in Newton optimization procedures due to the following

Proposition 1.5. *The Fisher information matrix Γ is positive-semi definite.*

Proof. Let $h(f;\theta) = h_1(f;\theta) + ih_2(f;\theta)$. Then by definition, we have

$$\begin{aligned}
\Gamma_{ij}(\theta) &\coloneqq \mathfrak{Re}\langle h_{,i}(\cdot,\theta), h_{,j}(\cdot,\theta)\rangle \\
&= 4\mathfrak{Re}\sum_i \frac{h_{,i}(f;\theta)^* h_{,j}(f;\theta)}{S(f_i)}\Delta f \\
&= 4\mathfrak{Re}\sum_i \frac{(h_{1,i} - ih_{2,i})(h_{1,j} + ih_{2,j})}{S(f_i)}\Delta f \\
&= 4\mathfrak{Re}\sum_i \frac{h_{1,i}h_{1,j} + h_{2,i}h_{2,j} + i(h_{1,i}h_{2,j} - h_{2,i}h_{1,j})}{S(f_i)}\Delta f
\end{aligned}$$

$$= 4\sum_i \frac{h_{1,i}h_{1,j} + h_{2,i}h_{2,j}}{S(f_i)}\Delta f.$$

Since $S > 0$ and both matricies in the numerator as PSD, so too is the sum. □

Recall that $h(f;\theta) = A(f;\theta)e^{i\psi(f;\theta)}$, and observe that

$$\begin{aligned}
r_{,j}(f;\theta) &\coloneqq \frac{h_{,j}(f;\theta)}{h(f;\theta_0)} \\
&= \frac{A_{,j}(f;\theta)e^{i\psi(f;\theta)} + i\psi_{,j}(f;\theta)A(f;\theta)e^{i\psi(f;\theta)}}{A(f;\theta_0)e^{i\psi(f;\theta_0)}} \\
&= \frac{A_{,j}(f;\theta)}{A(f;\theta_0)}e^{i(\psi(f;\theta)-\psi(f;\theta_0))} + \frac{\psi_{,j}(f;\theta)A(f;\theta)}{A(f;\theta_0)}e^{i(\psi(f;\theta)-\psi(f;\theta_0)+\frac{\pi}{2})},
\end{aligned}$$

and should also be compatible with our binning scheme. We now go a step further and derive heterodyned evaluation of the gradient of the potential, and the Fisher matrix.

Proposition 1.6. *The terms we need to calculate for the derivative and Fisher may be approximated as follows:*

$$\begin{aligned}
\langle h_{,j}(\theta), d\rangle &\approx \sum_b \big[A_0^{(b)} r_{0,j}^{(b)}(\theta)^* + A_1^{(b)} r_{1,j}^{(b)}(\theta)^*\big], \\
\langle h_{,j}(\theta), h(\theta)\rangle &\approx \sum_b \Big\{B_0^{(b)} r_{0,j}^{(b)}(\theta)^* r_0^{(b)}(\theta) + B_1^{(b)}\big[r_{0,j}^{(b)}(\theta)^* r_1^{(b)}(\theta) + r_{1,j}^{(b)}(\theta)^* r_0^{(b)}(\theta)\big]\Big\}, \\
\langle h_{,j}(\theta), h_{,k}(\theta)\rangle &\approx \sum_b \Big\{B_0^{(b)} r_{0,j}^{(b)}(\theta)^* r_{0,k}^{(b)}(\theta) + B_1^{(b)}\big[r_{0,j}^{(b)}(\theta)^* r_{1,k}^{(b)}(\theta) + r_{1,j}^{(b)}(\theta)^* r_{0,k}^{(b)}(\theta)\big]\Big\}.
\end{aligned}$$

Proof. (i) Follows immediately from the previous result:

$$\begin{aligned}
\langle h_{,j}(\theta), d\rangle &= \langle h(\theta), d\rangle_{,j} \\
&\approx \sum_b \big[A_0^{(b)} r_{0,j}^{(b)}(\theta)^* + A_1^{(b)} r_{1,j}^{(b)}(\theta)^*\big].
\end{aligned}$$

(ii) Similarly,

$$\begin{aligned}
\langle h_{,j}(\theta), h(\theta)\rangle &= \langle h(\cdot,\theta_0)r_{,j}(\cdot,\theta), h(\cdot,\theta_0)r(\cdot,\theta)\rangle \\
&\approx 4\sum_b \sum_{i:\, f_i \in \text{bin } b} \frac{|h(f_i;\theta_0)|^2}{S(f_i)}\Delta f\big[r_{0,j}^{(b)}(\theta) + r_{1,j}^{(b)}(f_i - f_-^{(b)})\big]^*\big[r_0^{(b)}(\theta) + r_1^{(b)}(f_i - f_-^{(b)})\big] \\
&\approx \sum_b \Big\{B_0^{(b)} r_{0,j}^{(b)}(\theta)^* r_0^{(b)}(\theta) + B_1^{(b)}\big[r_{0,j}^{(b)}(\theta)^* r_1^{(b)}(\theta) + r_{1,j}^{(b)}(\theta)^* r_0^{(b)}(\theta)\big]\Big\},
\end{aligned}$$

where in the last line we've ignored terms of $\mathcal{O}(|f_i - f_-^{(b)}|^2)$.

(iii) Follows a similar proof to (ii). Answer may be read off by replacing $r_0^{(b)}, r_1^{(b)} \leftarrow r_{0,k}^{(b)}, r_{1,k}^{(b)}$ in the rightmost square brackets of the second to last line. □

1.4 Reduced variance scheme

[Cornish, 2021] suggests using the following

Proposition 1.7. *The following expression holds:*

$$-\ln\mathcal{L}(\theta) = -\ln\mathcal{L}(\theta_0) + \frac{1}{2}\mathfrak{Re}\langle h(\cdot,\theta) - h(\cdot,\theta_0), h(\cdot,\theta) - h(\cdot,\theta_0)\rangle + \mathfrak{Re}\langle h(\cdot,\theta) - h(\cdot,\theta_0), h(\cdot,\theta_0) - d(\cdot)\rangle. \quad (5)$$

Proof. The result follows directly from the definition

$$-\ln\mathcal{L}(\theta) \coloneqq \frac{1}{2}\mathfrak{Re}\langle h(\cdot,\theta) - d(\cdot), h(\cdot,\theta) - d(\cdot)\rangle$$

$$
\begin{aligned}
&= \frac{1}{2}\mathfrak{Re}\big\langle \big[h(\cdot,\theta) - h(\cdot,\theta_0)\big] + \big[h(\cdot,\theta_0) - d(\cdot)\big], \big[h(\cdot,\theta) - h(\cdot,\theta_0)\big] + \big[h(\cdot,\theta_0) - d(\cdot)\big]\big\rangle \\
&= \frac{1}{2}\Big[\mathfrak{Re}\big\langle h(\cdot,\theta) - h(\cdot,\theta_0), h(\cdot,\theta) - h(\cdot,\theta_0)\big\rangle + 2\mathfrak{Re}\big\langle h(\cdot,\theta) - h(\cdot,\theta_0), h(\cdot,\theta_0) - d(\cdot)\big\rangle\Big] - \ln\mathcal{L}(\theta_0) \\
&= -\ln\mathcal{L}(\theta_0) + \frac{1}{2}\mathfrak{Re}\langle h(\cdot,\theta) - h(\cdot,\theta_0), h(\cdot,\theta) - h(\cdot,\theta_0)\rangle + \mathfrak{Re}\langle h(\cdot,\theta) - h(\cdot,\theta_0), h(\cdot,\theta_0) - d(\cdot)\rangle.
\end{aligned}
$$

□

This appears to be a good expression to approximate the likelihood from because both integrals now have small values, and hence the total error is smaller (i.e, the variance of the overall estimate should be improved). This idea can be extended to the idea of derivatives as well

Corollary 1.8. *The derivative may be expressed as*

$$
-\ln\mathcal{L}(\theta)_{,j} = \mathfrak{Re}\langle h_{,j}(\cdot,\theta), h(\cdot,\theta) - h(\cdot,\theta_0)\rangle + \mathfrak{Re}\langle h_{,j}(\cdot,\theta), h(\cdot,\theta_0) - d(\cdot)\rangle. \tag{6}
$$

Proof. Follows immediately from the previous result. □

Observe that this is the end of the line. The Fisher matrix does not benefit from a reduced variance form following this line of reasoning. Since we are currently using a zero-noise injection, the second term is exactly zero, and hence all we need to focus on in the first term.

2 Bin selection algorithm

Definition 2. Let $h(f;\theta)$ be a complex valued signal. Then the *post-Newtonian ansatz* asserts that $\psi(f) := \arg(h(f))$ takes the form

$$
\psi(f;\theta) := \sum_i \alpha_i(\theta) f^{\gamma_i}, \tag{7}
$$

where $\gamma := (-5/3, -2/3, 1, 5/3, 7/3)$.

We now have that

$$
\begin{aligned}
\Psi(f;\theta) &:= \psi(f;\theta) - \psi(f;\theta_0) \\
&= \sum_i \big(\alpha_i(\theta) - \alpha_i(\theta_0)\big) f^{\gamma_i} \\
&= \sum_i \delta\alpha_i(\theta) f^{\gamma_i}
\end{aligned}
$$

The heterodyne approximation is valid if Ψ varies slowly within a bin, which translates to a constraint on the $\delta\alpha_i$ terms. This motivates the following definition:

Proposition 2.1. *Suppose that the parameters θ are restricted to*

$$
\Theta := \big\{\theta \in \chi : \forall i\ |\delta\alpha_i(\theta)| \le 2\pi\chi f_{*,i}^{-\gamma_i}\big\}, \tag{8}
$$

where

$$
f_{*,i} := \begin{cases} f_{max} & \text{if } \gamma_i > 0 \\ f_{min} & \text{else.} \end{cases} \tag{9}
$$

Then the differential phase change over the interval $[f_-, f_+]$ obeys the following bound:

$$
|\Psi(f_+;\theta) - \Psi(f_-;\theta)| \le 2\pi\chi \sum_i \Big|\Big(\frac{f_+}{f_{*,i}}\Big)^{\gamma_i} - \Big(\frac{f_-}{f_{*,i}}\Big)^{\gamma_i}\Big|. \tag{10}
$$

Proof.

$$
\begin{aligned}
|\Psi(f_+;\theta) - \Psi(f_-;\theta)| &= \Big|\sum_i \delta\alpha_i(\theta)\big[f_+^{\gamma_i} - f_-^{\gamma_i}\big]\Big| \\
&\le \sum_i |\delta\alpha_i(\theta)||f_+^{\gamma_i} - f_-^{\gamma_i}| && \text{(triangle inequality)} \\
&\le 2\pi\chi \sum_i \Big|\Big(\frac{f_+}{f_{*,i}}\Big)^{\gamma_i} - \Big(\frac{f_-}{f_{*,i}}\Big)^{\gamma_i}\Big|.
\end{aligned}
$$

□

References

[Cornish, 2021] Cornish, N. J. (2021). Heterodyned likelihood for rapid gravitational wave parameter inference. *Physical Review D*, 104(10).