Fast Fisher Matrices and Lazy Likelihoods

Neil J. Cornish
Department of Physics, Montana State University, Bozeman, MT 59717

Theoretical studies in gravitational wave astronomy often require the calculation of Fisher Information Matrices and Likelihood functions, which in a direct approach entail the costly step of computing gravitational waveforms. Here I describe an alternative technique that sidesteps the need to compute full waveforms, resulting in significant computational savings. I describe how related techniques can be used to speed up Bayesian inference applied to real gravitational wave data.

Over the past two decades there have been literally hundreds of papers written describing parameter estimation studies in gravitational wave astronomy (we have to something to pass the time while waiting for the first detection). See Refs. [1–6] for some important early examples. The set-up is as follows: a waveform family \( h_k, h_x \) describing the radiation produced by some potential source of gravitational waves is chosen, and the detector response function is specified. From these inputs the waveform templates \( h(x) \) describing the signals produced in the detector for model parameters \( x \) can be computed. The output of the detector - or a network of detectors, it makes no difference - can then be written as \( s(t) = h(x, t) + n(t) \), where \( n(t) \) is the instrument noise. For theoretical studies it is usually assumed that the noise is Gaussian with a colored spectrum \( S_n(f) \), making it advantageous to shift the analysis to the Fourier domain where the noise samples are uncorrelated. The question is then asked, how well can the parameters \( x \) be constrained by the data? The answer follows from considering how well waveforms \( h(y) \) with parameters \( y \) are able to fit the data. The goodness of fit is found by taking the squared difference between the model and data, scaled by the noise level:

\[
\chi^2(y) = (s - h(y))s - h(y)) ,
\]

where the brackets denote the standard noise weighted inner product

\[
(a|b) = 2 \int \frac{a(f)b^*(f) + a^*(f)b(f)}{S_n(f)} df .
\]

The likelihood that the data would arise from a signal with parameters \( y \) is then [2]

\[
p(s|y) = C e^{-\chi^2(y)/2} ,
\]

where \( C \) is a constant that does not depend on the signal. The posterior probability, \( p(y|s) \), is simply the product of the prior probability, \( p(y) \), and the likelihood, \( p(s|y) \), divided by an overall normalization factor (the evidence). Notice that I never mentioned Mr. Wiener or matched filtering. Gravitational wave data analysis is no different than any other model fitting procedure: you compute a goodness of fit and turn the crank (but to confound astronomers in other fields we better keep on talking about matched filtering). Whatever you do, don’t start referring to the analysis as homodyne detection or the radio astronomers will descend on our field like a plague of locusts.

The parameter recovery accuracy is computed by looking at contours of the posterior. For strong signals the central limit theorem says that the posterior distribution is well approximated by a multivariate Gaussian distribution:

\[
p(y|s) \simeq \frac{1}{\text{det}(\Gamma / 2\pi)} \exp \left(-\Gamma_{ij} \Delta x^i \Delta x^j / 2\right) ,
\]

where \( \Gamma_{ij} \) is the Fisher information matrix and \( \Delta x^i = y^i - x^i \). It follows that the Fisher matrix is given by the expectation value of the negative Hessian of the log posterior density:

\[
\Gamma_{ij} = -\langle \partial_i \partial_j \ln p(x|s) \rangle = (h_{i|}h_{j}) - \partial_i \partial_j \ln p(x) .
\]

For sufficiently large signal-to-noise ratios, the variance-covariance matrix \( C^{ij} = \langle \Delta x^i \Delta x^j \rangle \) is given by the inverse of the Fisher information matrix. Better error estimates can be found by directly estimating the posterior distribution function using Markov Chain Monte Carlo [7, 8] or Nested Sampling [9, 10] techniques. Either way, these parameter estimation studies require that we compute a large number of noise weighted inner products, \( (a|b) \), which would seem to imply that we need to compute a large number of waveforms. But that turns out not to be the case.

Suppose that we want to map out the posterior using a Markov Chain or Nested Sampling. The first thing to realize is that you do not need to add simulated noise to the data [11]. This will only push the recovered recovered parameters off from the injected parameters in a way that depends on the particular noise realization (later on I will explain how to generalize the approach to handle data with noise). What we are really interested in is the allowed spread in recovered parameters, and that is set by the noise level \( S_n(f) \) in the noise weighted inner product. So the quantity we need to compute is

\[
\chi^2(y) = (h - h'[h - h']) \text{ where I'm using the shorthand} \ h = h(x) \text{ and } h' = h(y) .
\]

Suppose for the moment that we happen to have stationary phase approximation waveforms for \( h \) and \( h' \); we'll deal with time domain waveforms in a moment. Writing \( h = A(f) \exp(\Phi(f)) \) and \( h' = A'(f) \exp(\Phi'(f)) \) we get

\[
\chi^2(y) = 4 \int \frac{A^2 + A'^2 - 2AA' \cos(\Delta \Phi)}{S_n} df ,
\]
where it is understood that all of the quantities are frequency dependent. The first two terms in the integrand are always slowly varying function of frequency, and these terms can be integrated using a small number of function calls. The oscillatory term that comes from \((h|h')\) deserves more attention. So long as we are near maximum likelihood, the phase difference \(\Delta \Phi = \Phi(f) - \Phi'(f)\) will be a slowly vary function of \(f\), as will \(\cos(\Delta \Phi)\), and once again the integral can be computed to the desired accuracy with very few function calls. As we move away from maximum likelihood the phase difference grows and evaluating the integral takes more function calls. But we are not very interested in regions with low likelihood. It is simple to show that the variance of \(\chi^2\) is equal to the dimension \(D\) of the parameter space, and it follows that a Markov chain will rarely accept moves to places with \(\chi^2 > 3D\). Even there the phase evolution is not large, and the likelihood calculation remains inexpensive. Of course, moves will be proposed to locations with low likelihoods, and these locations do lead to very rapid oscillations in the \((h|h')\) terms. Which is why we simply set \((h|h') = 0\) whenever the phase change across the band exceeds some threshold (say tens of radians). The results of a MCMC or Nested Sampling analysis done in this way is indistinguishable from what you get when computing the likelihood directly.

The same technique can be used to compute the Fisher matrix if we write

\[
(h_{ij}|h_{ij}) \simeq \frac{(\Delta h_i|\Delta h_j)}{4\epsilon^i\epsilon^j}, \tag{7}
\]

with (no sum on \(i\))

\[
\Delta h_i = h(x + \epsilon^i e_i) - h(x - \epsilon^i e_i). \tag{8}
\]

Expanded out, the numerical central differences in equation (7) lead to \(2D^2 + D\) inner products that have to be evaluated. Since the \(\epsilon^i\) are by definition small, the integrands are all slowly varying and can be computed to the necessary accuracy with a small number of function evaluations. The calculation is even simpler if we work directly with the derivatives of the amplitude and phase:

\[
(h_{ij}|h_{ij}) = 4 \int \frac{A_{ij}A_{ij} + A^2 \Phi_i \Phi_j}{S_n} d\phi. \tag{9}
\]

Again, all the quantities in the integrand are slowly varying functions of frequency, and the integrand can be computed at little cost. This method of evaluating Fisher matrices has the added advantage that it is numerically far more stable than directly taking derivatives of the waveforms [12].

For the methods I have described to be useful the signals must have a discrete collection of instantaneous frequency components. When this condition is met, it is usually possible to derive stationary phase approximation waveforms [13]. In some instances the orbital timescale of the detector is comparable in duration to the gravi-
tational wave signal (e.g. most LISA, DECIGO or Einstein Telescope sources, signals from deformed neutron stars for LIGO/Virgo), and it is necessary to compute a stationary phase approximation to the detector response as well [4]. These calculations are straightforward, and including finite arm length effects is only a minor complication.

We can avoid the additional step of computing stationary phase approximations to the signal by invoking Parseval’s theorem and working directly in the time domain:

\[
(a|b) = 4 \int \frac{a(t)b(t)}{S_n(f(t))} dt, \tag{10}
\]

where the instantaneous frequency is given by the time derivative of the phase: \(f(t) = \partial_t \Phi(t)/(2\pi)\). Here I am assuming that the signal can be written in the form \(h(t) = A(t) \cos(\Phi(t))\), where \(A(t)\) is a slowly varying amplitude (in other words, the same condition that is required for the stationary phase approximation to work). Using the product formula:

\[
\cos \alpha \cos \beta = \frac{1}{2} [\cos(\alpha + \beta) + \cos(\alpha - \beta)]/2, \tag{11}
\]

and dropping the rapidly oscillating term involving the sum of the phases, we have

\[
(h|h') \simeq 2 \int \frac{A(t)A'(t)\cos(\Phi(t) - \Phi'(t))}{S_n(f(t))} dt. \tag{11}
\]

Again, for small phase differences, all the terms in the integrand are slowly varying, and the integral can be computed with a small number of function calls. Fisher matrix elements can be computed using the the time domain analog of equation (9).

In many cases the gravitational wave signals have power spread over multiple harmonics. For a Fisher matrix analysis it is enough to simply add together the contributions from each harmonic, but for likelihood calculations there is the possibility that different harmonics might overlap [14, 15]. To account for this possibility we need to compute

\[
\chi^2 = \sum_{ij} (h_i - h'_j|h_i - h'_j), \tag{12}
\]

where the \(i, j\) label the harmonics. Only those terms with small phase differences need to be computed.

The techniques described here are applicable to analytic or semi-analytic waveform models such as those derived using the post-Newtonian [16] or Effective One Body [17] approaches. These waveforms may require the numerical integration of energy fluxes and/or spin precession equations. An effective strategy for handling these cases is to include the calculation of the noise weighted inner products as differential equations, \(d(a|b)/dt\), and to solve the couple set of differential equations using a high order adaptive solver. This approach works very well for spinning systems since the precession timescale sets the amplitude evolution timescale. Because the orbital decay and spin precession timescales are generally much slower than the orbital timescale, there are substantial savings to be had from computing the likelihood and Fisher matrix elements in this way. If the evolution equations happen to be “stiff”, and require small step sizes, it may
be advantageous to compute the derivatives directly. For example, rather than integrating the equation

\[
\frac{d\alpha(x, t)}{dt} = a(x, t),
\]

for different values of \(x\) to estimate the derivatives \(a_x\), it is usually faster and more accurate to integrate the equations

\[
\frac{d\alpha_x(x, t)}{dt} = a_x(x, t).
\]

There are many applications for these techniques outside of theoretical parameter estimation studies. While the techniques described (so far) are not directly applicable when instrument noise is present, they can play a role in setting up template grids or driving stochastic searches. For example, in MCMC style searches it is important to have well chosen proposal distributions. The best are those that closely approximate the posterior distribution. The Fisher Information matrix provides a local approximation to the posterior, and multi-variate normal distributions of the form \(P = \mathcal{N}(\mu, \Sigma)\) have proven useful as proposal distributions \([14, 15, 18]\). However, if the Fisher matrix computation uses the full waveforms, it costs \(D(D + 1)/2\) times more to compute than the likelihood, and at that point you are better off using a cheaper proposal distribution and taking more steps in the chain. On the other hand, if the Fisher matrix elements can be computed at a fraction of the cost of the full waveforms, it makes sense to use \(P = \mathcal{N}(\mu, \Sigma)\) as a proposal distribution. Now suppose that the search has found a mode of the posterior at \(x\), which may or may not be the primary mode. To fully explore all the modes it helps to have a crude map of the global structure of the posterior. We can generate such a map at moderate computational cost. The procedure is to run a new MCMC search using the noise-free chi-squared \(\chi^2(y) = (h(x) - h(y))/(h(x) - h(y))\), which can be computed for a fraction of the cost of the true chi-squared. If \(h(x)\) happens to correspond to a secondary mode of the true posterior, one of the secondary modes of the approximate posterior will correspond to the primary mode of the true posterior. Thus, once the original search finds any hint of the signal, be it a secondary or tertiary mode, we can develop a map that points us towards the primary mode. Because noise will push the true posterior distribution away from the noise-free approximation, it is a good idea to temper the approximate posterior: \(\chi^2(y) \rightarrow \chi^2(y)/T\) using some “temperature” \(T > 1\). Increasing the temperature flattens the distribution, and we have to find a balance between making the approximate map uninformative \((T \rightarrow \infty)\), and being overly prescriptive \((T \rightarrow 1)\). Temperatures in the range \(T \in [3, 5]\) have been found to work quite well.

The same techniques used to speed up the likelihood and Fisher matrix calculations can also be used to speed up the computation of the template metrics \(g_{ij}\) for a grid based search. This should prove especially useful in high dimensional spaces where random template placement algorithms are required \([20]\). In particular, fast metric calculations are needed for hybrid grid-MCMC searches \([13]\) where the physical priors are replaced by \(p(x) \sim \sqrt{g(x)}\). This hybrid search is equivalent to a random grid search when no signals are present, which ensures full coverage of the search space, and has the advantage of being significantly faster than a grid search when signals are present (the saving grows with the signal strength).

In essence, the time saving techniques that I have described all amount to heterodyning of the data. That is, if we have signals \(h(x)\) and \(h(y)\) that differ by a small phase difference, their product yields a low frequency beat signal plus a high frequency signal that can be discarded without loss of information. The frequencies of the signals do not have to be constant for heterodyning to work. Heterodyning has been used in the search for gravitational wave signals from known radio pulsars \([21]\), and to simulate LISA observations of white dwarf binaries \([22]\). What apparently has not been realized before is that heterodyning can be used to significantly speed up MCMC and Nested Sampling explorations of the posterior for signals embedded in noisy instrument data. Suppose that the primary or a secondary mode of the posterior \(x\) has been located by some search algorithm and you would now like to fully map out the posterior distribution. Rather than work with the full signal \(s(t)\), first Fourier transform the data, whiten using the noise spectral density \(S_n(f)\), and heterodyne using the carrier phase \(\Phi(x, f)\). The high frequency components of the data can now be thrown away, and the noise weighted inner products can be computed using templates that are heterodyned against the carrier phase. The bandwidth of the heterodyned signal that needs to be kept depends on the details of the analysis, but the data volume will typically be reduced by many orders of magnitude. To gain the full benefit from this approach the heterodyned templates must be computed directly at low cadence using the phase difference \(\Delta \Phi(f)\) (or the equivalent in the time domain). Note that carrier phase that beats with the signal at the primary maximum will also beat with the signal at the secondary maxima. The likelihoods computed far from the maxima will not agree with those computed using the full signal, but if the heterodyned signal is given sufficient bandwidth, the differences will be small in regions with noticable posterior weight.

As a concrete example, consider a binary neutron star inspiral that might be observed by the LIGO/Virgo detectors. The gravitational wave signal in each detector can be written as

\[
h(t) = F^+ h_\times(t) + F^x h_\times(t)
= F^+ A_+(t) \cos \Phi(t) + F^x A_\times(t) \sin \Phi(t).
\]

The antenna patterns \(F^+\) and \(F^\times\) depend on the sky location and polarization angle. The amplitudes \(A_+(t)\) and \(A_\times(t)\) depend on the merger time, the orbital inclination, the masses and the distance to the binary. The
phase $\Phi(t)$ depends on the sky location (which determines the time delay between the signals seen at the different sites), merger time and the masses. Here I am using 2-PN waveforms without spin. The heterodyning can be done in the time domain by multiplying the signal by $\cos \Theta(t)$ and $\sin \Theta(t)$, where $\Theta(t)$ is the reference, or carrier phase corresponding to the best fit source parameters. For a 1.4$M_\odot$-1.3$M_\odot$ inspiral, the original waveforms reached a maximum frequency of 1628.5 Hz. The signal was whitened and heterodyned, then lowpass filtered to a maximum of 10 Hz. Low cadence templates for the sine and cosine streams can then be generated:

$$h_{\cos}(t) = \frac{1}{2} \left( F^+ A_+(t) \cos \Delta \Phi(t) + F^x A_x(t) \sin \Delta \Phi(t) \right)$$

$$h_{\sin}(t) = \frac{1}{2} \left( -F^+ A_+(t) \sin \Delta \Phi(t) + F^x A_x(t) \cos \Delta \Phi(t) \right),$$

(16)

where $\Delta \Phi(t) = \Phi(t) - \Theta(t)$. Figure 1 shows the whitened and heterodyned time domain signal and noise for a neutron star binary with network SNR = 15. We used the injected source parameters to generate the carrier phase $\Phi(t)$. Also shown is a template generated using ($10$) with the masses shifted by $-3\%$ and $+1\%$. The noise weighted inner products are computed directly in the time domain:

$$\langle h|s \rangle = 2 \int \frac{h_{\sin}(t) s_{\sin}(t) + h_{\cos}(t) s_{\cos}(t)}{S_n(f)} \, dt. \quad (17)$$

For the case shown, $\langle h_{\text{inj}}|h_{\text{inj}} \rangle = 225.0$, $\langle h|h \rangle = 231.2$, $\langle h_{\text{inj}}|n \rangle = -14.1$, and $\langle h|n \rangle = 12.5$. The inner products computed using the full cadence data agree with those computed using the heterodyned data to $\sim \pm 0.01$. The only difference is that the heterodyned inner products can be computed one thousand times faster.

Hopefully the techniques I have described will help reduce the high computational cost of Bayesian approaches to gravitational wave data analysis, and allow their wider adoption.

Acknowledgments

This work was supported by NASA Grant NNX10AH15G and NSF Award 0855407.

[1] F. Echeverria, Phys. Rev. D 40, 3194 (1989).
[2] L. S. Finn, Phys. Rev. D 46, 5236 (1992) [arXiv:gr-qc/9209010].
[3] C. Cutler and E. E. Flanagan, Phys. Rev. D 49, 2658 (1994) [arXiv:gr-qc/9402014].
[4] C. Cutler, Phys. Rev. D 57, 7089 (1998) [arXiv:gr-qc/9703068].
[5] E. E. Flanagan and S. A. Hughes, Phys. Rev. D 57, 4566 (1998) [arXiv:gr-qc/9710129].
[6] N. Christensen and R. Meyer, Phys. Rev. D 58, 082001 (1998).
[7] W. Hastings, Biometrika 57, 97 (1970).
[8] C. Andrieu, N. de Freitas, A. Doucet and M. I. Jordan, Machine Learning 50, 5 (2003).
[9] J. Skilling, AIP Conference Proceedings, 735, 395 (2004).
[10] F. Feroz and M. P. Hobson, MNRAS 384, 449 (2008). [arXiv:0704.3704 [astro-ph]].
[11] S. Nissanke, S. A. Hughes, D. E. Holz, N. Dalal and J. L. Sievers, [arXiv:0904.1017 astro-ph.CO].
[12] C. Cutler and S. Drasco have developed this approach independently, and are using it to compute Fisher matrices for Extreme Mass Ratio Inspirals (private communication).
[13] E. Droz, D. J. Knapp, E. Poisson and B. J. Owen, Phys. Rev. D 59, 124016 (1999) [arXiv:gr-qc/9901076].
[14] N. J. Cornish and J. Crowder, Phys. Rev. D 72, 043005 (2005) [arXiv:gr-qc/0506059].
[15] N. J. Cornish, arXiv:0804.3323 [gr-qc].
[16] L. Blanchet, Living Rev. Rel. 5, 3 (2002) [arXiv:gr-qc/0202016].
[17] A. Buonanno and T. Damour, Phys. Rev. D 59, 084006 (1999) [arXiv:gr-qc/9811091].
[18] N. J. Cornish and E. K. Porter, Phys. Rev. D 75, 021301 (2007) [arXiv:gr-qc/0605135].
[19] B. J. Owen, Phys. Rev. D 53, 6749 (1996) [arXiv:gr-qc/9511032].
[20] C. Messenger, R. Prix and M. A. Papa, Phys. Rev. D 79, 104017 (2009) [arXiv:0809.5223 [gr-qc]].
[21] R. J. Dupuis and G. Woan, Phys. Rev. D 72, 102002 (2005) [arXiv:gr-qc/0508096].

[22] N. J. Cornish and T. B. Littenberg, Phys. Rev. D 76, 083006 (2007) [arXiv:0704.1808 [gr-qc]].