GPU Accelerated Parameter Estimation of Gravitational Waves from Binary Black Hole Mergers

Brandon B. Miller, Masters Thesis

August 17, 2016

1 Background and Motivation

Gravitational waves are predicted as a mathematical consequence of the linearized Einstein Field Equations. The Einstein Field Equations themselves are a set of ten unique, coupled, nonlinear partial differential equations.

1.1 Pieces of the Einstein Equation

1.1.1 Reimann Tensor

The Reimann Tensor is a mathematical object that encapsulates the geometric *curvature* of spacetime. The concept of a curved spacetime is fundamentally no different from the concept of a curved space, and from examining the behavior of vectors in such a space we can address both the purpose and properties of the Reimann tensor without departing from the intuitive picture of spheres and arrows.

The Reimann tensor is traditionally derived by examining the local behavior of geodesics subject to a metric tensor that describes curved space. Here we motivate the concept with a visual explanation of parallel transport. Consider a vector pointing in some arbitrary direction σ somewhere on the sphere, as depicted in figure **FIGURE HERE**. If we maintain the orientation of the vector relative to the surface of the sphere as we traverse the path, we note that the vector behaves differently depending on the order in which we choose to traverse the directions μ and ν. This difference is representable by a vector that points between the two tips - one that is dependent on the initial direction σ and both the transport directions μ and ν. Let this vector be denoted $\vec{d}$. Then the Reimann tensor, $R_{\rho\sigma\mu\nu}$, is simply d_ρ, the ρ'th component of this difference vector. From this geometric picture some of the classical symmetries that are often introduced abreast the definition of the Reimann tensor are immediately intuitive, such as the skew symmetry depicted in figure **FIGURE HERE**. Since it is somewhat unfair (worse - *incorrect*) to characterize the entire geometry of the space with looping paths over large areas, the formal definition of the Reimann tensor is in fact the infinitesimal version of the concept above. Indeed, the concept of a single direction one may traverse is only valid in a small region of curved space, so the real Reimann tensor must describe the difference between two vectors as parallel transported around an infinitesimally small version of the original loop. It follows then how there must be a (possibly different) Reimann tensor at each point on the manifold in question. More remarkable however is the simplicity of the formal definition within this picture:

$$R^{\rho}_{\sigma\mu\nu}\partial_\rho = (\nabla_\mu\nabla\nu - \nabla_\nu\nabla_\mu)\partial_\sigma \tag{1.1}$$

It should be noted that the above formula is true only modulo a term containing $\nabla_{[\mu,\nu]}$ except in the special case of coordinate vector fields, where it is zero. Although it does little to redeem it as a first explanation of the concept, this also illuminates why the Reimann tensor is sometimes described as encoding the non-commutativity of covariant derivatives on a manifold. Somewhat more subtle is the explicit role of the Reimann tensor in

describing geodesic deviation. Crucial to a straightforward linearization of the Reimann tensor itself is the expression in terms of the Christoffel Symbols **AS DEFINED IN APPENDIX**

$$R^{\rho}_{\sigma\mu\nu} = \partial_{\mu}\Gamma^{\rho}_{\nu\sigma} - \partial_{\nu}\Gamma^{\rho}_{\mu\sigma} + \Gamma^{\rho}_{\mu\lambda}\Gamma^{\lambda}_{\nu\sigma} - \Gamma^{\rho}_{\nu\lambda}\Gamma^{\lambda}_{\mu\sigma} \tag{1.2}$$

The christoffel symbols contain some information about the curvature of space as a function of the coordinates, and thus in the framework of linearized gravity products of Christoffel symbols are second order:

$$R^{\rho}_{\sigma\mu\nu} = \partial_{\mu}\Gamma^{\rho}_{\nu\sigma} - \partial_{\nu}\Gamma^{\rho}_{\mu\sigma} \tag{1.3}$$

If $x^{\rho}(\tau)$ are the components of a geodesic parameterized by τ, then $\dot{x^{\rho}}(\tau) \equiv T^{\rho}$ are the components of a vector tangent to the geodesic. Suppose there is a family such nearby geodesics indexed by some other parameter, s, and that we can define *deviation vector* $S^{\rho}(\tau) = \partial_s x^{rho}(s, \tau)$ which characterizes the separation between nearby geodesics. It can be shown that the second derivative of this vector with respect to τ is related to the Reimann tensor through the differential equation

$$\ddot{S}^{\rho} = R^{\rho}_{\sigma\mu\nu}T^{\sigma}T^{\mu}S^{\nu} \tag{1.4}$$

2 Linearized Gravity

In linearized gravity we assume the spacetime is described by a flat spacetime $g_{\mu\nu}$ plus a small perturbation

$$g_{\mu\nu} = \eta_{\mu\nu} + h_{\mu\nu} \tag{2.1}$$

Since the other pieces of the Einstein field equations are directly derived from the metric, they must be individually linearized to form a description of spacetime around weak sources. The standard treatment of linearized gravity assumes that to first order we may raise and lower indices using the Minkowski metric. If we tried to use the full form of equation **EQUATION**, we would have

$$g^{\mu}_{\nu} = g^{\mu\alpha}g_{\alpha\nu} \tag{2.2}$$
$$= (\eta^{\mu\alpha} + h^{\mu\alpha})(\eta_{\alpha\nu} + h_{\alpha\nu}) \tag{2.3}$$
$$= \eta^{\mu\alpha}\eta_{\alpha\nu} + \eta^{\mu\alpha}h_{\alpha\nu} + h^{\mu\alpha}\eta_{\alpha\nu} + h^{\mu\alpha}h_{\alpha\nu} \tag{2.4}$$

The second two terms being second order in h, ther peturbation, are neglected in linearized gravity. We assume from here on that indices can be freely raised and lowered to first order using only the Minkowski metric $\eta_{\mu\nu}$.

$$h_\nu^\mu = \eta^{\alpha\mu} h_{\alpha\nu} \ \ \text{and} \ \ h^{\mu\nu} = \eta^{\alpha\mu} \eta^{\beta\nu} h_{\alpha\beta} \tag{2.5}$$

The next natural line of logic to pursue is to write down the Christoffel symbols for such a spacetime

$$\Gamma_{\alpha\beta}^\mu = \frac{1}{2} g^{\nu\mu} \left[\partial_\beta g_{\alpha\nu} + \partial_\alpha g_{\beta\nu} - \partial_\nu g_{\alpha\beta} \right] \tag{2.6}$$

$$= \frac{1}{2} \left(\eta^{\mu\nu} + h^{\mu\nu} \right) \left[\partial_\beta g_{\alpha\nu} + \partial_\alpha g_{\beta\nu} - \partial_\nu g_{\alpha\beta} \right] \tag{2.7}$$

Inserting the expanded form of the metric into the term within the brackets gives us six terms times two at the front, for twelve total. Thankfully, three correspond to the standard flat-space Christoffel symbols which can be made zero through a choice of coordinates (in fact, orthonormal Cartesian coordinates), three involve derivatives of $\eta_{\mu\nu}$ which are zero, and three are second order in $h_{\mu\nu}$, which we assume are zero. This leaves us with three:

$$\Gamma_{\alpha\beta}^\mu = \frac{1}{2} \eta^{\mu\nu} \left[\partial_\beta h_{\alpha\nu} + \partial_\alpha h_{\beta\nu} - \partial_\nu h_{\alpha\beta} \right] \tag{2.8}$$

$$= \frac{1}{2} \left[\partial_\beta h_\alpha^\mu + \partial_\alpha h_\beta^\mu - \partial^\mu h_{\alpha\beta} \right] \tag{2.9}$$

Where in the last term $\partial^\mu = \eta^{\mu\nu} \partial_\nu$ was used to raise the index on the actual partial derivative operator itself. It should be noted at this point that since we neglect terms that are second order in the Christoffel symbols themselves, in linearized gravity, the covariant derivative of a metric perturbation is equivalent to the normal partial derivate. The same assumption allows for simplification of the Reimann and Ricci tensors:

$$R_{\mu\sigma\nu}^\rho = \partial_\sigma \Gamma_{\mu\nu}^\rho - \partial_\nu \Gamma_{\mu\sigma}^\rho \tag{2.10}$$

$$R_{\mu\nu} = R_{\mu\rho\nu}^\rho = \partial_\rho \Gamma_{\nu\mu}^\rho - \partial_\mu \Gamma_{\mu\rho}^\rho \tag{2.11}$$

Plugging in equation **EQUATION** we have

$$R_{\mu\nu} = \frac{1}{2} \partial_\alpha \left[\partial_\nu h_\mu^\alpha + \partial_\mu h_\nu^\alpha - \partial^\alpha h_{\mu\nu} \right] - \frac{1}{2} \partial_\nu \left[\partial_\alpha h_\mu^\alpha + \partial_\mu h_\alpha^\alpha - \partial^\alpha h_{\mu\alpha} \right] \tag{2.12}$$

$$= \frac{1}{2} \left[\cancel{\partial_\alpha \partial_\nu h_\mu^\alpha} + \partial_\alpha \partial_\mu h_\nu^\alpha - \partial_\alpha \partial^\alpha h_{\mu\nu} - \cancel{\partial_\nu \partial_\alpha h_\mu^\alpha} - \partial_\nu \partial_\mu h_\alpha^\alpha + \partial_\nu \partial^\alpha h_{\mu\alpha} \right] \tag{2.13}$$

$$= \frac{1}{2} \left[\partial_\alpha \partial_\mu h_\nu^\alpha + \partial_\nu \partial^\alpha h_{\mu\alpha} - \partial_\alpha \partial^\alpha h_{\mu\nu} - \partial_\nu \partial_\mu h_\alpha^\alpha \right] \tag{2.14}$$

The quantity h_α^α is the trace of the pertubation metric and represents, in a sense, an overall measure of the *strength* or *amplitude* of the perturbation, this is often called just

h. The term $\partial^\alpha \partial_\alpha$ is an inner product that when formed through the Minkowski metric becomes the wave operator, $\Box = -\partial_t^2 + \nabla^2$. The Ricci tensor is obtained by contracting the Reimann tensor over the two free indices:

$$R \equiv \eta^{\mu\nu} R_{\mu\nu} = \frac{1}{2} \left[\partial_\alpha \partial^\mu h_\mu^\alpha + \partial_\mu \partial^\alpha h_\alpha^\mu - \Box h - \Box h \right] \tag{2.15}$$

Since α and μ are summed indices, the first two terms are the same. Dropping the $\frac{1}{2}$, the Ricci scalar curvature is

$$R = \partial_\alpha \partial_\mu h^{\mu\alpha} - \Box h \tag{2.16}$$

Many authors choose to simplify the expression for the Ricci Tensor by defining the quantity

$$V_\nu = \partial_\alpha h_\nu^\alpha - \frac{1}{2} \partial_\nu h \tag{2.17}$$

With which the first two terms of the linearized Ricci tensor can be reorganized as

$$\partial_\mu \partial_\alpha h_\nu^\alpha - \partial_\mu \partial_\nu h = \partial_\mu \left[\partial_\alpha h_\nu^\alpha - \partial_\nu h \right] \tag{2.18}$$

$$= \partial_\mu V_\nu - \frac{1}{2} \partial_\mu \partial_\nu h \tag{2.19}$$

Which upon substitution into the **EQUATION** for $R_{\mu\nu}$ yields

$$R_{\mu\nu} = \frac{1}{2} \left[\partial_\mu V_\nu - \frac{1}{2} \partial_\mu \partial_\nu h + \partial^\alpha \partial_\nu h_{\mu\alpha} - \Box h_{\mu\nu} \right] \tag{2.20}$$

The two unsimplified terms form the quantity $\partial_\nu \left[\partial^\alpha h_{\mu\alpha} - \frac{1}{2} \partial_\mu h \right]$ which we recognize as $\partial_\nu V_\mu$, leaving us with

$$R_{\mu\nu} = \frac{1}{2} \left[\partial_\mu V_\nu + \partial_\nu V_\mu - \Box h_{\mu\nu} \right] \tag{2.21}$$

Setting the right hand side of this equation gives us the linearized, vacuum Einstein Equation $R_{\mu\nu} = 0$:

$$\frac{1}{2} \left[\partial_\mu V_\nu + \partial_\nu V_\mu - \Box h_{\mu\nu} \right] = 0 \tag{2.22}$$

3 Implementation

The main weight of the computational work described above falls under an expensive Monte Carlo integration of the evidence as part of the Bayesian posterior. Zooming in it quickly becomes apparent that the main challenge to be overcome is efficient evaluation of the function to be integrated, in this case the factored likelihood. Prior to this work, the factored likelihood is was computed with a series of looping structures that individually compiled all the terms of the equation in an independent and serial manner. The process took substantial advantage of many high level operations available within Python to match up the correct terms across various harmonic mode time series, spherical harmonics, and antenna patterns. Thanks to clever optimization and gratuitous use of fast numerical Python libraries such as NumPy, the code was able to compute approximately 10^3 likelihood evaluations on the order of seconds, the limiting factor being the serial nature of loops and the inherent lethargy of high level languages such as Python. Although NumPy commonly passes target data to compiled routines written in faster languages such as C, some portions of the computation (in particular the marginalization over time) benefit little from this capability and form a bottleneck for efficient evaluation of the likelihood. Furthermore the original implementation heavily relied upon older structures within the existing LIGO Algorithms Library (LAL), a set of C routines bound to Python through Swig, to perform many intermediate calculations, such as computing spherical harmonics. The goal of the first phase of this work was thorough vectorization of the process used to build the terms of the likelihood utilizing the BLAS subroutines available through NumPy as often as possible, as well as reducing dependency on LAL. To that end, the components of the factored likelihod were reworked.

3.1 Complex Antenna Factor

The complex antenna factor is a complex number that modulates the gravitational wave signal as a function of incoming direction. This is necessary due to the non isotropic sensitivity of the LIGO interferometers themselves. F_+ and $F_\times$.

$$F_+ = [R_{0,0}X_0 + R_{0,1}X_1 + R_{0,2}X_2]X_0 - [R_{0,0}Y_0 + R_{0,1}Y_1 + R_{0,2}Y_2]Y_0 \quad (3.1)$$

$$+ [R_{1,0}X_0 + R_{1,1}X_1 + R_{1,2}X_2]X_1 - [R_{1,0}Y_0 + R_{1,1}Y_1 + R_{1,2}Y_2]Y_1 \quad (3.2)$$

$$+ [R_{2,0}X_0 + R_{2,1}X_1 + R_{2,2}X_2]X_2 - [R_{2,0}Y_0 + R_{2,1}Y_1 + R_{2,2}Y_2]Y_2 \quad (3.3)$$

Collecting all the positive terms of this expression gives

$$F_+ = R_{0,0}X_0X_0 + R_{0,1}X_0X_1 + R_{0,2}X_0X_2 \quad (3.4)$$

$$+ R_{0,0}X_0X_0 + R_{0,1}X_0X_1 + R_{0,2}X_0X_2 \quad (3.5)$$

$$+ R_{0,0}X_0X_0 + R_{0,1}X_0X_1 + R_{0,2}X_0X_2 \quad (3.6)$$

At which point we see that the same result is computable as an outer product of the vector $\vec{X}$ using

$$F_+ = \vec{X}\mathbf{R}\vec{X} - \vec{Y}\mathbf{R}\vec{Y} \tag{3.7}$$

And the complex part of the gravitational wave is thus

$$F_\times = \vec{X}\mathbf{R}\vec{Y} + \vec{Y}\mathbf{R}\vec{X} \tag{3.8}$$

One a sample-to-sample basis, the numbers that vary are the components of the vectors $\vec{X}$ and $\vec{Y}$. Thus we need a function that takes as input a *list* vectors (or *vector*) of vectors and produces a vector with the right components as output. To that end we define the tensor X_j^i where

$$X^i = \begin{bmatrix} X_0^i \\ X_1^i \\ X_2^i \end{bmatrix} \tag{3.9}$$

As well as the tensor R_{jk}^i where

$$R^i = \begin{bmatrix} R_{00}^i & R_{01}^i & R_{02}^i \\ R_{10}^i & R_{11}^i & R_{12}^i \\ R_{20}^i & R_{21}^i & R_{22}^i \end{bmatrix} \tag{3.10}$$

The tensor X is like a stack of all the different possible $\vec{X}$ coming out of the page. The Tensor R is like n copies of the matrix $\mathbf{R}$ stacked on top of each other. In this way the desired vector is obtainable with the tensor contraction

$$F_+^i = X^{lm}R_{lj}^i X_m^j - Y^{lm}R_{lj}^i Y_m^j \tag{3.11}$$

$$F_\times^i = X^{lm}R_{lj}^i Y_m^j + Y^{lm}R_{lj}^i X_m^j \tag{3.12}$$

$$\tag{3.13}$$

3.2 Vectorized Single Detector Log Likelihood

With the spherical harmonics and antenna factor in hand the following set of operations yield the factored likelihood for a single detector, the network likelihood being a simple sum over all of the detectors.

Equation 24 from Arxiv *15502.05370v1.pdf* for the network log likelihood reads

$$\ln \mathcal{L} = \frac{D_{ref}}{D} Re \sum_k \sum_{(l,m)} [F_k Y_{lm}]^* Q_{k,lm} \tag{3.14}$$

$$-\left[\frac{D_{ref}}{2D}\right]^2 \sum_k \sum_{(l,m),(l',m')} \left[|F_k|^2 Y_{l,m}^* Y_{l',m'} U_{k,(l,m),(l'm')}\right] \tag{3.15}$$

$$-\left[\frac{D_{ref}}{2D}\right]^2 \sum_k \sum_{(l,m),(l',m')} Re \left[F_k^2 Y_{l,m} Y_{l'm'} V_{k,(l,m),(l'm')}\right] \tag{3.16}$$

Consider a single detector, thus dropping the sum over k. The first term is of the form $\vec{A} \cdot \vec{B} = \sum_{i=0}^d A_i B_i$, so if $Q_{k,(l,m)}$ were a simple vector, we could write it as

$$-\left[\frac{D_{ref}}{2D}\right]^2 \sum_k \sum_{(l,m),(l',m')} \left[|F_k|^2 Y_{l,m}^* Y_{l',m'} U_{k,(l,m),(l'm')}\right] = -\left[\frac{D_{ref}}{2D}\right]^2 F * \vec{Y} \cdot \vec{Q} \tag{3.17}$$

However the $Q_{k,(l,m)}$ are actually harmonic mode time series and not single values. We desire a vector whose values are the likelihoods at each point in the time series.

Consider the case where we have only the $(2,-2), (2,0)$ and $(2,2)$ modes. If we write all the mode time series $Q^0, Q^1, Q^2...$ as the columns of a matrix , then the desired result is obtained with

$$F * \begin{bmatrix} Q_{2,-2}^0 & Q_{2,+0}^0 & Q_{2,+2}^0 \\ Q_{2,-2}^1 & Q_{2,+0}^1 & Q_{2,+2}^1 \\ Q_{2,-2}^2 & Q_{2,+0}^2 & Q_{2,+2}^2 \\ \vdots & \vdots & \vdots \end{bmatrix} \begin{bmatrix} (Y_{2,-2}) \\ (Y_{2,+0}) \\ (Y_{2,+2}) \end{bmatrix} = \begin{bmatrix} Q_{2,-2}^0 Y_{2,-2} + Q_{2,-2}^0 Y_{2,+0} + Q_{2,+2}^0 Y_{2,+2} \\ Q_{2,-2}^1 Y_{2,-2} + Q_{2,-2}^1 Y_{2,+0} + Q_{2,+2}^1 Y_{2,+2} \\ Q_{2,-2}^2 Y_{2,-2} + Q_{2,-2}^2 Y_{2,+0} + Q_{2,+2}^2 Y_{2,+2} \\ \vdots \end{bmatrix}$$
$$\tag{3.18}$$

With $\vec{Y}$ and $\mathbf{Q}$ defined as the matrix and vector above respectively, we have for the first term

$$\frac{D_{ref}}{D} Re \left[\mathbf{Q} \left(F\vec{Y}\right)^*\right] \tag{3.19}$$

The second term is a sum once over all the possible combinations of $(l,m), (l',m')$ pairs using the $U_{(l,m),(l',m')}$ cross terms. Its result is a scalar quantity made up of terms like

$$Y_{2,-2}^* Y_{2,-2} U_{(2,-2),(2,-2)} + Y_{2,-2}^* Y_{2,0} U_{(2,-2),(2,+0)} + Y_{2,-2}^* Y_{2,+2} U_{(2,-2),(2,+2)} \tag{3.20}$$

$$+ Y_{2,+0}^* Y_{2,-2} U_{(2,+0),(2,-2)} + Y_{2,+0}^* Y_{2,0} U_{(2,+0),(2,+0)} + Y_{2,+0}^* Y_{2,+2} U_{(2,+0),(2,+2)} \tag{3.21}$$

$$+ Y_{2,+2}^* Y_{2,-2} U_{(2,+2),(2,-2)} + Y_{2,+2}^* Y_{2,0} U_{(2,+2),(2,+0)} + Y_{2,+2}^* Y_{2,+2} U_{(2,+2),(2,+2)} \tag{3.22}$$

8

If we pack the $U_{(l,m),(l',m')}$ into the matrix $\mathbf{U}$ as defined below then the following set of matrix operations produces the same sum

$$\begin{bmatrix} Y_{2,-2}^* & Y_{2,+0}^* & Y_{2,+2}^* \end{bmatrix} \begin{bmatrix} U_{(2,-2),(2,-2)} & U_{(2,-2),(2,+0)} & U_{(2,-2),(2,+2)} \\ U_{(2,+0),(2,-2)} & U_{(2,+0),(2,+0)} & U_{(2,+0),(2,+2)} \\ U_{(2,+2),(2,-2)} & U_{(2,+2),(2,+0)} & U_{(2,+2),(2,+2)} \end{bmatrix} \begin{bmatrix} Y_{2,-2} \\ Y_{2,+0} \\ Y_{2,+2} \end{bmatrix}$$

because when you multiply $\mathbf{U}$ into $\vec{Y}$ this simplifies to

$$\begin{bmatrix} Y_{2,-2}^* & Y_{2,+0}^* & Y_{2,+2}^* \end{bmatrix} \begin{bmatrix} U_{(2,-2),(2,-2)}Y_{2,-2} + U_{(2,-2),(2,-2)}Y_{2,-2} + U_{(2,-2),(2,-2)}Y_{2,-2} \\ U_{(2,-2),(2,-2)}Y_{2,-2} + U_{(2,-2),(2,-2)}Y_{2,-2} + U_{(2,-2),(2,-2)}Y_{2,-2} \\ U_{(2,-2),(2,-2)}Y_{2,-2} + U_{(2,-2),(2,-2)}Y_{2,-2} + U_{(2,-2),(2,-2)}Y_{2,-2} \end{bmatrix}$$

Which becomes the desired scalar. This allows us to write the second term as

$$-\left[\frac{D_{ref}}{2D}\right]^2 \sum_k \sum_{(l,m),(l',m')} \left[|F_k|^2 Y_{l,m}^* Y_{l',m'} U_{k,(l,m),(l'm')}\right] = -\left[\frac{D_{ref}}{2D}\right]^2 |F^2|\vec{Y}^* \mathbf{U} \vec{Y} \quad (3.23)$$

We must always set up the spherical harmonic vectors based on the value of m and the cross terms in row-major form based first on m_2 and then on m_1. If we organize the matrix $\mathbf{V}$ in the same way then the same set of steps will lead us to conclude that

$$-\left[\frac{D_{ref}}{2D}\right]^2 \sum_k \sum_{(l,m),(l',m')} Re\left[F_k^2 Y_{l,m} Y_{l'm'} V_{k,(l,m),(l'm')}\right] = -\left[\frac{D_{ref}}{2D}\right]^2 Re\left[F^2 \vec{Y} \mathbf{V} \vec{Y}\right]$$

$$(3.24)$$

Combining the results the single detector log likelihood is

$$\ln\mathcal{L} = \frac{D_{ref}}{D}\Re\left[\mathbf{Q}\left(F\vec{Y}\right)^*\right] - \left[\frac{D_{ref}}{2D}\right]^2 \left[|F|^2 \vec{Y}^* \mathbf{U} \vec{Y} - \Re\left(F^2 \vec{Y} \mathbf{V} \vec{Y}\right)\right] \quad (3.25)$$

3.3 GPU Implementation

Vectorization of the code produces a performance improvement due to SIMD instruction sets available on many modern microprocessor architectures. This allows for some on-chip parallelism that the compiler may use to unroll loops within the machine code and increase efficiency. Beyond this however the process of recasting the computation into clear matrix and tensor operations illuminates a higher level of parallelism possible on the hardware level. A substantial amount of both academc and industrial research and development has centered around efficient parallel implementations of the equivalent operations, with impressive results.