8-15-2022

Acquisition Strategies for Gravitational Wave Surrogate Modeling

Karl Daningburg
kd1956@rit.edu

Acquisition Strategies for Gravitational Wave Surrogate Modeling

by

Karl Daningburg

A Thesis Submitted in Partial Fulfillment of the Requirements
for the Degree of Master of Science in Applied Mathematics
School of Mathematical Sciences, College of Science

Rochester Institute of Technology
Rochester, NY
August 15, 2022

Committee Approval:

R·I·T
College of
SCIENCE

Richard O'Shaughnessy, Ph.D. Date

School of Mathematical Sciences

Thesis Advisor

Carlos Lousto, Ph.D. Date

School of Mathematical Sciences

Committee Member

Nathan Cahill, D.Phil. Date

School of Mathematical Sciences

Director of Graduate Programs

Abstract

Gravitational wave science is dependent upon expensive numerical simulations, which provide the foundational understanding of binary merger radiation needed to interpret observations of massive binary black holes. The high cost of these simulations limits large-scale campaigns to explore the binary black hole parameter space. Surrogate models have been developed to efficiently interpolate between simulation results, but these models require a sufficiently comprehensive sample to train on. Acquisition functions can be used to identify points in the domain for simulation. We develop a new acquisition function which accounts for the cost of simulating new points. We show that when applied to a 3D domain of binary mass ratio and dimensionless spins, the accumulated cost of simulation is reduced by a factor of about 10.

Contents

I. Introduction

Einstein's theory of General Relativity (GR), a geometric description of the interaction between matter and spacetime, has been used to produce numerous testable hypotheses since its formulation in 1915. Prominent hypotheses range from gravitational lensing to the existence of black holes, regions of mass so dense that light itself is unable to escape. Although hypothesized to exist by Schwarzschild in 1916, it wasn't until 1971 with the discovery of Cygnus X-1 that physical evidence of black holes was found [33].

Certain effects of gravity in the GR paradigm are only significant in extremely strong fields or with extremely high velocities. While these conditions are rare, both of them are produced during the merger of two black holes. Such mergers emit tremendous amounts of energy in the form of gravitational waves. These waves are typically the only measurable output of binary mergers, making them a uniquely valuable source of information about the universe's most violent conditions. After decades of research, the first gravitational wave signal was detected in 2015 by the LIGO observatory [1].

The Advanced LIGO [23] and Virgo [2] instruments are now regularly observing gravitational waves [1, 37, 38, 36]. However, these waveforms are inscrutable without some reference as to the nature of the binary systems which produce them. To determine the parameters of the binaries responsible, each measurement is compared directly or indirectly with predictions for the gravitational radiation produced when binaries merge [40, 22, 3]. Numerical simulations provide the authoritative (and often only) foundational insight into radiation generated during the late stages of mergers [9]. However, these simulations are extremely expensive, making complete exploration of the domain infeasible.

Some scientists have turned to surrogate modeling, a class of mathematical methods which cheaply emulate an expensive objective function. Surrogate models have been widely applied to gravitational waveform modeling [8, 16, 6, 34, 39, 4, 17, 7, 19, 30, 29, 14], both to accelerate existing approximations and to interpolate between detailed numerical relativity simulations. Given a gravitational wave model evaluated on candidate binary parameters x_k, a surrogate model interpolates the waveform's asymptotic amplitude and phase across the binary parameters x (and time). However, to date these calculations all adopt a simplified way to select the initial training points x_k: they are usually selected with a greedy algorithm based on a *fast approximation* to the true waveform [4], independent of the cost of simulating those parameters x_k.

The process of collecting data to use in training a surrogate model is referred to as *acquisition*. The

acquisition process is completely dependent on the scientific domain. For example, a geologist estimating the distribution of a given mineral over an area could acquire new data by drilling boreholes at specific sites. Analogous to this experimental example is the computational case: a numerical researcher must make decisions about what parameter values to simulate to best characterize a system of interest. If the cost of acquiring data is low, the acquisition strategy is relatively unimportant. However, if data is costly to obtain, intelligent acquisition has high potential to reduce the cost of building a surrogate model. Researchers commonly encode their acquisition strategies into *acquisition functions*. These functions assign a value to each point in the domain. The decision about what point to acquire next is made by maximizing (or minimizing) the acquisition function.

In the case of numerical relativity, simulation cost is extremely high; furthermore, cost is heavily dependent on the initial system parameters [24, 25]. Our novel contribution is to incorporate both interpolator uncertainty and acquisition cost into our acquisition function. This approach balances the goals of (i) accurately modeling the objective function over the entire domain by picking points of maximum variance and (ii) minimizing total acquisition cost by minimizing the cost of selected points. Importantly, our goal is not to create a surrogate model which rivals existing surrogate models in accuracy or speed; it is only to demonstrate the potential of cost-weighted acquisition when it comes to training a surrogate model with numerical relativity as cheaply as possible.

This thesis is organized as follows. In Section II we introduce surrogate modeling with an example from the gravitational wave domain. We define the objective function we have chosen to learn, then define acquisition functions to be tested. We introduce a workflow for acquiring data and training a Gaussian process. In Section III we examine the performance of the acquisition functions and resulting surrogate models. In Section IV we summarize the work and conclude with future prospects for this research in V. Note that this thesis is based on the work in the following paper, which at the time of this writing has been accepted for publication in Physical Review D following revision [13].

II. Method

II.1 Surrogate Review by Example

To be concrete, we introduce surrogate modeling for gravitational waves by summarizing the specific techniques adopted by Blackman et al. in [4] to generate a surrogate gravitational wave model for precessing binaries. The complex dimensionless gravitational wave strain

$$h(t,\theta,\phi;\boldsymbol{\lambda}) = h_+(t,\theta,\phi;\boldsymbol{\lambda}) - ih_\times(t,\theta,\phi;\boldsymbol{\lambda}) \tag{II.1}$$

is expanded in terms of its polarizations h_+ and $h_\times$, where t is time, θ and λ are polar and azimuthal angles of wave propogation from the binary system, and $\boldsymbol{\lambda}$ is the set of chosen parameters to describe the system. Blackman et al. select black hole spin vectors and initial mass ratio as parameters for their surrogate model; however, the method applies for any parameter of interest. The angular dependence can be efficiently represented with a spin-weighted spherical harmonic decomposition

$$h(t,\theta,\phi;\boldsymbol{\lambda}) = \sum_{l=2}^{\infty}\sum_{m=-l}^{l} h^{lm}(t;\boldsymbol{\lambda})\,{}^{-2}Y_{lm}(\theta,\phi) \tag{II.2}$$

Following Blackman et al., we will focus only on the $l = 2,3$ modes, since higher-order modes have minimal impact on systematic error. Their surrogate models seek to represent each mode function $h_{lm}(t,\boldsymbol{\lambda})$ versus time and $\boldsymbol{\lambda}$.

Blackman et al.'s approach relies on discovering a finite set of greedy parameters

$$G \equiv \{\boldsymbol{\Lambda}_i \in \mathcal{T}\}_{i=1}^{N} \tag{II.3}$$

where $\mathcal{T}$ is a compact region of parameter space. They then perform NR simulations at each greedy parameter, resulting in the greedy solutions $\{W(t;\boldsymbol{\Lambda}_i)\}_{i=1}^{N}$. They transcribe these solutions into a regression model first by constructing an orthonormal linear basis $B_n = \{e^i(t)\}_{i=1}^{N}$ spanning the greedy solutions so that

$$W(t;\boldsymbol{\lambda}) \approx \sum_{i=1}^{n} c_i(\boldsymbol{\lambda})e^i(t) \tag{II.4}$$

where coefficient c_i is the inner product of $W(t;\boldsymbol{\lambda})$ with $e^i(t)$. In principle the greedy parameters G may be found by iteratively comparing the approximation to the true function $W(t;\boldsymbol{\lambda})$ and selecting the point of maximum error:

$$E_n(\boldsymbol{\lambda}) = \| W(\cdot;\boldsymbol{\lambda}) - \sum_{i=1}^{n} c_i(\boldsymbol{\lambda}) e^i(\cdot) \| \tag{II.5}$$

However, due to the high cost of full numerical simulation, dense evaluation of this error is not feasible. Blackman et al. circumvent this by building a mock surrogate from a cheaper data source: post-Newton waveform models. They assume that the greedy parameters found in construction of the mock surrogate, G^{PN}, approximate the greedy parameters for the NR surrogate, and thus are able to select training data at a reasonable cost.

Gravitational waves presented as functions of time have complicated dependence on time and binary parameters, which adds additional artificial hurdles when attempting to interpolate them. A powerful technique is to decompose the waveforms obtained by NR simulations into data pieces which vary more slowly and thus are more easily modeled. These disparate models can be combined to form a single surrogate model. Commonly the phase and amplitude of the waves are used as the data pieces; however, for the precessing waveforms which Blackman et al. are targeting, a more sophisticated decomposition is needed. First, each waveform is translated into a coprecessing coordinate frame. Next, a Gaussian filter is used to remove the effect of nutation of the frame's rotational axis. Finally, the waveforms are decomposed into symmetric and antisymmetric amplitudes and phases

$$A_{\pm}^{l,m}(t) = \frac{1}{2}\left(\mid \tilde{h}^{l,m}(t) \mid \pm \mid \tilde{h}^{l,-m}(t) \mid \right) \tag{II.6}$$

$$\varphi_{\pm}^{l,m}(t) = \frac{1}{2}\left(\varphi(\tilde{h}^{l,m}(t)) \pm \varphi(\tilde{h}^{l,-m}(t)) \right) \tag{II.7}$$

where $\tilde{h}$ is the waveform in the coprecessing frame. The surrogate training data and output are always expressed using this coprecessing-frame amplitude-phase decomposition. Similarly, surrogate models can be constructed for the precessional dynamics themselves, needed to translate between the simulation and coprecessing frame. Figure 1 illustrates the rationale for using an amplitude-phase decomposition for the simple case of a nonprecessing binary black hole's dominant mode: in this amplitude-phase decomposition, the two salient functions involved in $h(t)$ are slowly varying in time (and versus binary parameters).

Now that amplitudes and phases for the decomposed waveforms have been extracted, interpolation on the data can be performed. Blackman et al. use an empirical interpolant via Singular Value Decomposition to unify the $\tilde{h}(t)$ waveforms into a single surrogate model $W_S(t;\boldsymbol{\lambda})$. Despite relying on a SVD algorithm to find basis functions, Blackman et al. use a greedy algorithm to select time points over which to interpolate with these functions. For a detailed discussion on

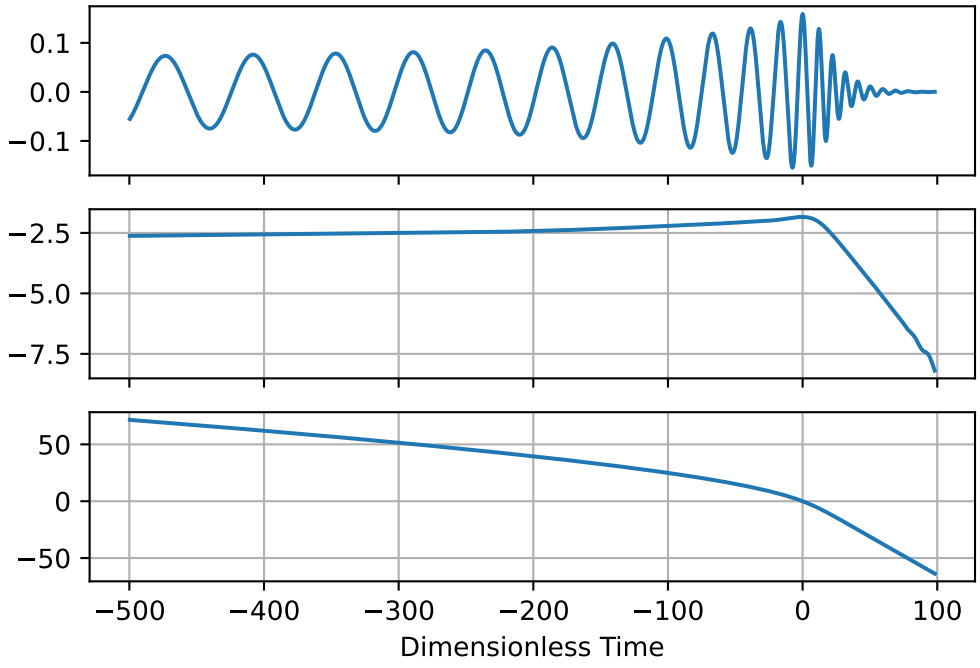


Figure 1: Strain wave h(t) is depicted, as well as the wave's decomposition into log amplitude and phase. These smooth functions are more easily interpolated than the raw wave data and thus are used as input to the surrogate model.

building reduced bases with SVD versus greedy algorithms, see Appendix B of [4]. Blackman et al. use the so-called "empirical time" algorithm developed in [17]. Given a reduced basis $\{e_i\}_{i=1}^{m}$ whose span approximates the objective function (in our case, gravitational wave phase and amplitude), the Empirical Interpolation Method (EIM) yields a set of m empirical times T_i chosen from the complete set of times t_i. These times are selected by comparing the value of the empirical interpolant

$$\mathcal{I}_j[h](t;\boldsymbol{\lambda}) = \sum_{i=1}^{j} C_i(\boldsymbol{\lambda}) e_i(t) \tag{II.8}$$

to the basis functions evaluated at all times, where the C_i coefficients are the solutions to the interpolation problem at the empirical times. At each step, one more basis function is added to $\mathcal{I}_j[h](t;\lambda)$; the time of maximum error is found and added to the set of empirical times T_i. This process continues until the number of times chosen is equal to the number of basis functions m.

Blackman et al. acquire data for their surrogate model by maximizing error between the true function and the surrogate. Since NR data is far too expensive to obtain in the quantities needed for this method, they evaluate a cheaper objective function to acquire new points. Another type of acquisition function which may be more easily used with an expensive reference function is one based on minimizing uncertainty rather than error. This formulation of the interpolation problem lends itself to Gaussian processes, which provide quantification of uncertainty throughout the domain. Therefore, we propose to focus on Gaussian processes as interpolators and measures of uncertainty as a main ingredient in our acquisition functions.

II.2 Gaussian Process Interpolation

Gaussian processes fall under a broader class of methods known as *stochastic processes*. Simplistically, stochastic processes are a probabilistic representation of possible functions which may be the source of data observations [32]. Given a data set $\mathcal{D} = (\mathbf{x}_i, y_i)$, we seek a succinct representation of all functions which map from vector inputs $\mathbf{x}_i$ to scalar outputs y_i. A stochastic process is initialized as a prior distribution, predicated on the idea that little is known about the underlying function before data is taken. It is then conditioned on data $\mathcal{D}$ using Bayes' Theorem. A trained stochastic process can be evaluated to provide new data in emulation of the underlying (objective) function. For a straightforward 1-D example, see Figure 2.

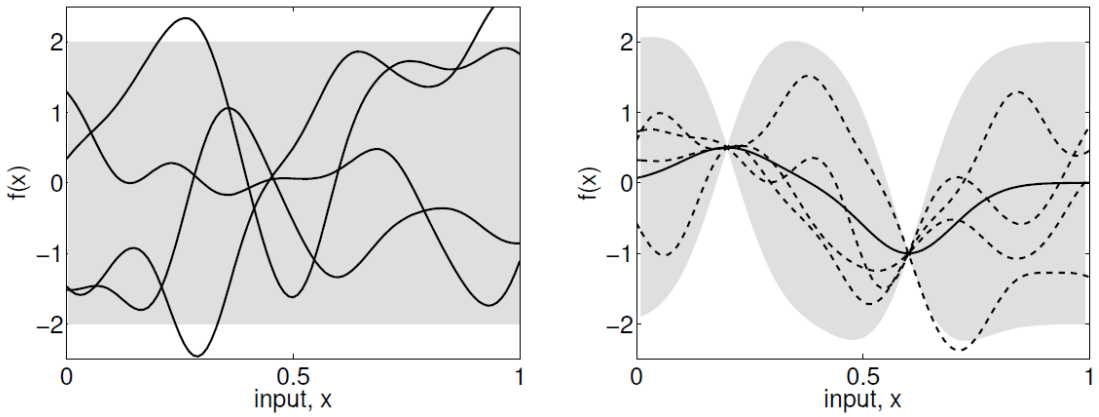


Figure 2: A simple example from [32]. On the left, a stochastic process before training is depicted. All functions contained by the prior distribution (gray) could explain the underlying function. On the right, training data has been provided and the possible functions restricted. Note the mean function (solid line) traveling through the data points, as well as the reduction in uncertainty near those data points.

Gaussian processes have been previously employed to perform the interpolation needed to construct surrogate models and identify candidate points for followup analysis; see, e.g., [14], which is similar to our cost-neutral approach. To define a Gaussian process we require a mean function μ and a kernel function k:

$$\mu(\mathbf{x}) = \mathbb{E}[f(\mathbf{x})] \tag{II.9}$$

$$k(\mathbf{x}, \mathbf{x}') = \mathbb{E}[(f(\mathbf{x}) - \mu(\mathbf{x}))(f(\mathbf{x}') - \mu(\mathbf{x}'))] \tag{II.10}$$

where $\mathbb{E}$ denotes the expected value and $f(\mathbf{x})$ is the objective function to be learned.

The choice of a kernel function is not a straightforward one. Since kernel functions encapsulate assumptions about the function to be learned, prior knowledge of this function is useful. For this introductory work, we use the simple and commonly-used squared exponential (SE) function with added white noise, with the understanding that further investigation is worthwhile:

$$K(x_p, x_q) = \sigma_f^2 \exp\left\{-\frac{(x_p - x_q)^2}{2l^2}\right\} + \sigma_n^2 \tag{II.11}$$

where x_p and x_q are two inputs to the objective function, σ_f^2 is the variance of the signal describing f, l is the characteristic length scale, and σ_n^2 is the white noise variance associated with the data. The SE function has the advantage of being easy to evaluate and interpret, and it yields smooth interpolation between data points.

Consider the squared exponential kernel introduced above in eqn. II.11. σ_f, l, and σ_n are termed *hyperparameters* because they are parameters of a non-parametric model [32]. These hyperparameters are tuned during the training process to provide the best fit to the data. The signal variance σ_f^2 captures the uncertainty about the correlation between inputs. The white noise term σ_n^2 allows for variability in the function even at points in the domain for which we have data; it also improves the numerical stability of the training process in the absence of significant noise, as is typical for computer simulation data. At high separation between inputs, the total variance will be $\sigma_f^2 + \sigma_n^2$. The remaining hyperparameter l scales the correlation between nearby function values. In short, a low-l data set will have rapidly changing target values, while a high l indicates a slowly changing function.

Having chosen mean and kernel functions, the hyperparameters of the GP must be tuned to provide the best fit to the data. These values are chosen by maximizing the marginal likelihood, or the likelihood that the training data was produced by a process with a given set of

hyperparameters:

$$\mathrm{P}(\mathbf{y}|\mathbf{x},\boldsymbol{\theta}) = \int \mathrm{P}(\mathbf{y}|\mathbf{f},\mathbf{x},\boldsymbol{\theta})d\mathbf{f} \tag{II.12}$$

where $\boldsymbol{\theta}$ is a vector containing the hyperparameters, $\mathbf{x}$ is a point in the domain, $\mathbf{y}$ is the corresponding output data, and $\mathbf{f}$ is the GP function [26].

The evaluation of a GP at a test point x_* provides us with two quantities: the expected value $\bar{f}_*$ of the objective function, and the variance $\mathbb{V}$ of that value:

$$\bar{f}_* = \mathbf{k}_*^T(K + \sigma_n^2 I)^{-1}\mathbf{y} \tag{II.13}$$

$$\mathbb{V}[f_*] = k(x_*, x_*) - \mathbf{k}_*^T(K + \sigma_n^2 I)^{-1}\mathbf{k}_* \tag{II.14}$$

Here $\mathbf{k}_*$ is the kernel evaluated on the test point and the training data, K is the kernel evaluated on the training data only, $\mathbf{y}$ is the target values of the training data, and I is the identity matrix. This dual output is the primary motivation for using GPs in our simulation targeting: we are provided with an estimate of the function value with $\bar{f}_*$, the desired output of a trained surrogate, but also with the uncertainty measure $\mathbb{V}$. We will discuss how the variance is used to acquire new data points in II.4.

II.3 Simulation Cost

Although Numerical Relativity (NR) simulations were not used for training this surrogate model, we develop this model with the intent of directly targeting and interpolating NR simulations. As our goal is a reduction of simulation and training cost, we develop a cost function to estimate the expense of simulating points in the domain. Our colleagues estimate [24] the length of a simulation scales approximately with mass ratio q, as does the resolution required to maintain accuracy; therefore the cost dependence on mass ratio can be modeled as q^2. Meanwhile, dimensionless spins of the two bodies have very little effect on the simulation cost until spin of 0.6 is reached. Thereafter, the cost increases rapidly and diverges as spin approaches 1. After consulting with an NR group [24], the following piecewise function models the cost-spin relationship:

$$\mathcal{S}(\chi) = \begin{cases} 1 & \chi \leq 0.6 \\ \frac{0.4}{1-\chi} & 0.6 < \chi \leq 0.9 \\ \frac{0.04}{(1-\chi)^2} & 0.9 < \chi < 1 \end{cases} \tag{II.15a}$$

For our chosen domain of mass ratio and spins for bodies 1 and 2, the cost function is as follows:

$$C(q, \chi_1, \chi_2) = q^2 + \mathcal{S}(\chi_1) + \mathcal{S}(\chi_2) \tag{II.15b}$$

where q is mass ratio $\frac{m_1}{m_2}$, $m_1 > m_2$, and χ_1 (χ_2) is dimensionless spin magnitude for body 1 (2).

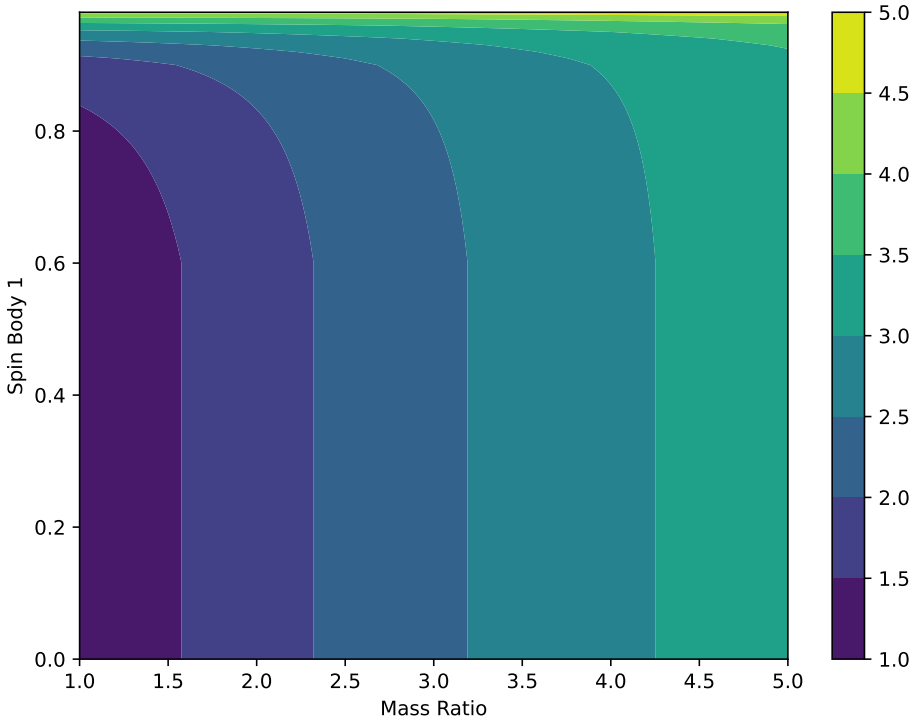


Figure 3: The estimated cost of a nonprecessing NR simulation versus binary mass ratio q and primary spin $\chi_{1,z}$ as estimated by Eq. (II.15). The color scale shows log(cost). For concreteness in this plot we adopt $\chi_{2,z} = 0$.

II.4 Acquisition

Acquisition refers to the collection of a data point, whether by experimentation or simulation. In order to train our surrogate model most efficiently, we develop acquisition functions which

assign a value to each point in the domain. Optimizers are used to maximize these acquisition functions, and the resulting point in the domain is selected for the next simulation. An example of acquisition function design for use with Gaussian processes is summarized.

In [35], Srinivas et al. formulate the Gaussian optimization problem in the context of selecting locations for temperature sensors. It is desirable to find areas of highest temperature with the minimal amount of sensors; this is analogous to choosing new points of an expensive objective function to evaluate in search of maxima. The group chooses to focus on the kernel function used in a Gaussian Process Upper Confidence Bound algorithm. The metric Srinivas et al. use is termed *regret*: the cost of not knowing a priori the objective function (i.e. temperature map) maxima. The group uses a greedy algorithm to pick a new data point x to maximize information gain like so:

$$x_t = \arg\max_{x \in \mathcal{D}} \sigma_{t-1}(x) \tag{II.16}$$

This acquisition function attempts to reduce uncertainty across the whole domain. Given this model's intent to find maxima in the temperature field, Srinivas et al. modify this function as follows:

$$x_t = \arg\max_{x \in \mathcal{D}} \mu_{t-1}(x) + \beta_t^{1/2} \sigma_{t-1}(x) \tag{II.17}$$

where β_t is a chosen constant. The first term drives the function to acquire points with expected high reward (existing suspected maxima) and the second term pushes towards points with high variance (read: uncertainty). In this way the acquisition function regulates the tradeoff between exploitation of current suspected maxima and exploration of new areas in the domain.

Whereas Srinivas et al. balanced interpolator uncertainty (variance) against the goal of finding peaks in temperature, we wish to formulate an acquisition function which balances the need to find points of maximum uncertainty with the requirement of keeping simulation costs at a minimum. Having defined variance and cost in Equations (II.14) and (II.15b), we can now construct an acquisition function which quantifies our learning per unit cost:

$$A = \frac{\mathbb{V}}{C} \tag{II.18}$$

Maximizing this function at each step constitutes a greedy algorithm for learning as much about the objective function as possible for the lowest cost.

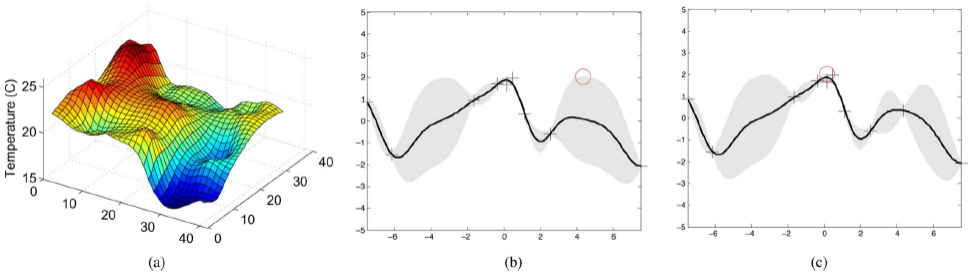


Figure 4: Example figure from [35] demonstrating data acquisition. (a) Underlying function: temperature data. (b) and (c) Two iterations of data acquisition for the Gaussian Process. The dark line represents the current posterior mean, while the gray regions indicates the confidence bounds on the GP. The red circles show the selection of (b) an area of high variance and (c) an area near a maximum.

For the purpose of comparison, we test both a variance maximizing routine (see, e.g., [14]) and a variance-to-cost ratio maximizing routine. For brevity, we will refer to these as $\mathbb{V}$ and $\mathbb{V} : C$ routines, respectively.

II.5 Time Selection

The data available for a given point in parameter space is comprised of a complete gravitational waveform which may be expressed in the frequency domain or the time domain. It is desireable to select an optimal set of times (or frequencies) at which to measure the amplitude or phase of each available GW and interpolate that data through parameter space, allowing us to reconstruct any waveform in the domain. Splines are piecewise polynomials which interpolate across data points, termed knots, by using constraints to ensure continuity and smoothness. We will rely on the commonly-used natural cubic splines, which are comprised of cubic polynomials. The basic constraints for natural cubic splines are as follows:

1. Each polynomial must be equal to the data at the knots surrounding it
2. The first and second derivatives of each polynomial must be equal to its adjacent polynomials at interior knots
3. The second derivative of the first and last polynomials must be equal to zero at the first and last knots

The algorithm for time selection proceeds as follows. Given a set of GWs associated with a set of parameters, we seek a minimal subset of dimensionless time points τ to fully characterize the behavior of any waveform in the domain; we will therefore call this set *characteristic time* τ_c. To

initialize the algorithm, five evenly-spaced time points are selected. The data corresponding to these times are fed to SciPy's interp1d routine, which trains a cubic spline interpolator for each available GW. We use this spline to interpolate across 50000 evenly spaced time points from -500 to +75 τ, with 0 τ corresponding to the moment of merger. The root-mean-square error

$$\text{RMSE}\left(\phi_j\right) = \frac{1}{n}\sum_{i=1}^{n}\sqrt{\left(\frac{\phi_f^{(ij)} - \phi_c^{(ij)}}{\max|\phi_f^{(i)}|}\right)^2} \tag{II.19}$$

is computed at each time step across all splines. Here ϕ is the GW phase, j indicates one of the 50000 time steps, n is the number of GWs available, ϕ_f is the fiducial spline, and ϕ_c is the spline trained only on τ_c. The time of maximum error is chosen to be added to the spline training data. The process iterates until the maximum RMSE $\left(\phi_j\right)$ falls below a desired threshold. These characteristic time points and the corresponding values of phase or amplitude are used to train Gaussian processes for interpolation across parameter space. New GWs can be constructed by training new splines on the data interpolated by the GPs.

II.6 Error Metric

We can evaluate the model's performance at any point in the space $\mathcal{T}$ by comparing it to the source of our data, IMRPhenomD. A brief discussion of *mismatch*, a common metric for quantifying the difference between waveforms, is necessary. Given two waveforms, a fiducial wave h_1 and a surrogate approximation h_2, we require an objective measure of how well h_2 approximates h_1. For a single mode we compute the *match* [10]:

$$\mathcal{O} = \frac{\max_t \left|\int_{-\infty}^{\infty} h_2^*(f) h_1(f) e^{i2\pi f t} df\right|}{||h_1||\ ||h_2||} \tag{II.20}$$

This expression is evaluated using a white-noise detector power spectrum; alternatively, this expression corresponds to the standard equally-weighted Hilbert space inner product in time or frequency. Empirically we estimate the overlap from our discrete data using an inverse Fourier transform

$$\mathcal{O} = \max_t \text{IFT}\left[N\left(\frac{\tilde{h}_2^*}{||\tilde{h}_2||}\right)\left(\frac{\tilde{h}_1}{||\tilde{h}_1||}\right)\right] \tag{II.21}$$

where N is the number of samples, $\tilde{h}$ indicates a single wave mode in the frequency domain, and * represents the complex conjugate. Two waves identical to a constant factor will have an overlap of unity; therefore to reframe as an error metric we compute the mismatch as

$$\mathcal{E} = 1 - \mathcal{O} \tag{II.22}$$

We chose this approach since this surrogate model is intended to be generic, independent of any detector's power spectrum. For a more complete discussion of waveform mismatch, see [4].

II.7 Testing With Fast Aligned Model

Our surrogate modeling methods are similar in principle to those outlined in [4]. Given a physical system parametrized by $\boldsymbol{\lambda} \in \mathcal{T}$, where $\mathcal{T}$ is a compact region in parameter space, we seek cheaply evaluated functions of time $W(t; \boldsymbol{\lambda})$ that describe the system. For our model $\boldsymbol{\lambda}$ includes mass ratio q and spins (χ_{1z} and χ_{2z}) of black holes in a binary black hole system and $\mathcal{T}$ encompasses a range of mass ratios and spins for which an existing model is available.

Although NR is the eventual data source for this type of model, we begin with a more rapidly solved method: IMRPhenomD [21]. IMRPhenomD is a phenomenological model which evaluates the gravitational wave signal of black hole binaries throughout the inspiral, merger, and ringdown phases. This model is capable of producing data for non-precessing black holes of mass ratio up to 18:1 and dimensionless spin of 0.85, or 0.98 for 1:1 mass sytems. IMRPhenomD is a hybrid model which incorporates both Effective-One-Body data and NR simulations for tuning. It has demonstrated error of under 1% versus NR test waveforms [21]. We selected this model to enable more rapid training of our surrogate model and thorough error analysis of the methods developed. Its close approximation of NR results ensure translation of our methods to NR data acquisition with similar performance.

Once our parameter space is populated following the method outlined in Sec. II.4, we turn to the task of interpolating the data across both time and the parameter space. We seek a set of dimensionless times $\tau_0, \ldots, \tau_k$ for each value to be interpolated across parameter space which characterizes the behavior of these values over time for all systems in the target region $\mathcal{T}$. Using the method described in Sec. II.5, we choose characteristic times τ_c for both phase and amplitude of the GWs. Importantly, the times for phase and the times for amplitude are selected separately, since the behavior of these two values is disparate. Once the characteristic times are selected, Gaussian processes are trained on the value of amplitude or phase at each characteristic time for all available gravitational waves.

To solve the surrogate model at any point in parameter space, first we evaluate the Gaussian processes at all characteristic times for both amplitude and phase. Then we reconstruct the full amplitude and phase curves in time using the natural cubic spline interpolant. These curves can be used to reconstruct the original strain waveform $h(t)$. To get a succinct understanding of

competing models' performance, we evaluate the wave mismatch [see Eqs. (II.20) and (II.22)] at select points distributed evenly through parameter space and take the L2 norm to produce a single performance metric. We compute this error metric at strategically chosen intervals of acquisition to understand the overall error trend as more points are acquired.

III. Results

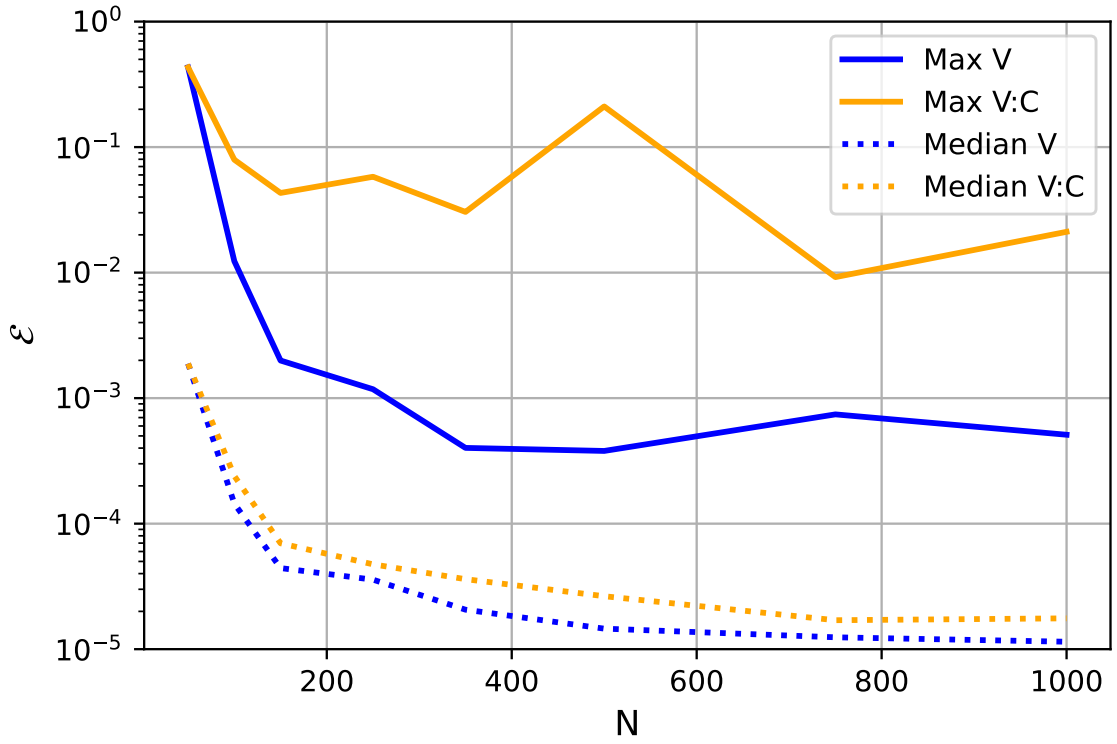


Figure 5: The mismatch (II.22) of two surrogate models over data acquisition. In blue the model acquires new points based only on GP variance; in orange, the ratio of variance to cost. The variance-to-cost acquisition function maximum error remains high in a small number of extremely high cost areas; however, it is competitive across most of the parameter space.

To compare the performance of two acquisition functions of (a) variance maximization (hereafter V) and (b) variance-to-cost ratio maximization (hereafter $V : C$), we devised the following test. For parameter space $\mathcal{T}$ we chose the three dimensions upon which simulation cost relies: mass ratio q and the spins χ_1 and χ_2. Mass ratio q we allowed to vary from 1 to 5, while each χ ranged from 0 to 0.99. For the purposes of characteristic time, we chose a maximum L2 norm error of 0.0001 across all acquired waveforms, with a maximum number of characteristic time points of 200. Recall that for each characteristic time point, a Gaussian process will be trained on the data from that moment in time for all acquired waveforms (see subsection II.5); therefore the cap of 200 is chosen to keep runtime reasonably low. We initialize the model with 50 data points selected randomly through Latin Hypercube Sampling, and then begin the iterative process of training and acquiring new data.

III.1 Global Performance

A comparison of the two surrogate models is presented in Figure 5. For this figure, the mismatch was measured at 1089 points evenly spaced in the three dimensional parameter space. Mass ratio q was sampled in intervals of 0.5 from 1 to 5, and spins χ_1 and χ_2 were sampled in intervals of 0.1 from 0 to 0.99. We use the median as a summary statistic of performance due to high error at a small number of points dominating metrics such as mean or root-mean-square. For completeness we also present the maximum mismatch within the parameter space. By the metric of median mismatch vs number of acquired points, the V scheme slightly outperforms the $V : C$ scheme.

Reframing the problem as maximizing learning per unit cost rather than per data point acquired reveals the advantage of the $V : C$ acquisition function. Figure 6 shows the relationship between mismatch and relative cost of acquired data points. While performance per cost is similar at higher errors, as the error is driven lower by further data acquisitions the $V : C$ routine strongly outperforms the V routine. The $V : C$ routine achieves comparable median mismatch across the described domain for about an order of magnitude smaller cost.

III.2 Performance Over Parameter Space

The $V : C$ methods achieve their improved global performance by sacrificing exploration and thus accuracy in the most costly regions. To highlight this tendency, Figure 7 shows the mismatch versus binary parameters after a fixed number of acquisitions. Notably, the maximum error has actually increased when increasing training data from 350 to 500 points. This is caused by overtraining in low-cost areas without the acquisition of any additional data in the highest-cost region. While the $V : C$-derived surrogate model is competitive with the V model in most of parameter space, it does not as reliably reproduce extreme-cost waveforms.

To date, nature has provided principally low-spin compact binaries, so a focus on low-spin exploration seems warranted. However, high-spin binaries have considerable discovery potential and interesting physics. A more thorough review of the science return of a given GW model based on its accuracy in different regions of parameter space is required to better quantify this tradeoff.

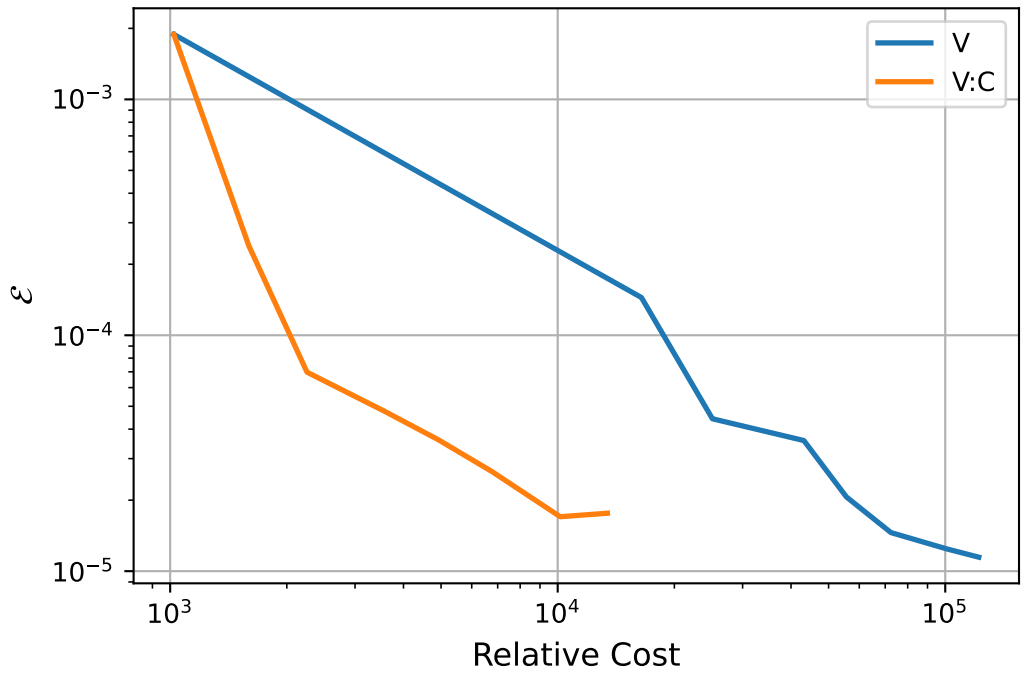
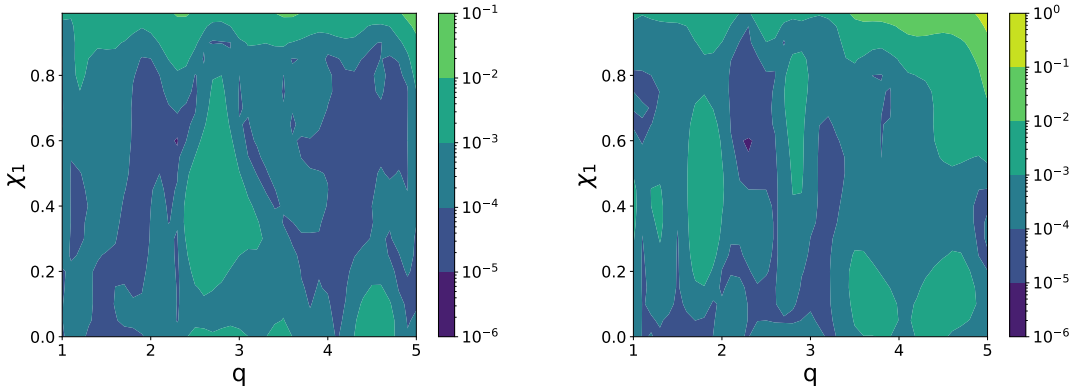


Figure 6: The relative acquisition cost of two surrogate models over median mismatch [Eq, (II.22)]. In blue the model acquires new points based only on GP variance; in orange, the ratio of variance to cost. The $V : C$ acquisition function achieves the same median match error across the majority of the parameter space with nearly an order of magnitude smaller cost once mismatch is driven below 10^{-4}.

III.3 Discussion

There are several key distinctions between the surrogate model developed here and the model developed by Blackman et al. [4]. Perhaps the most important distinction is the use of a Gaussian process as an interpolant rather than an SVD-based empirical interpolant. Using a Gaussian process opened up the possibility of using the interpolator's uncertainty (variance) about the objective function in order to place simulations to acquire more data efficiently. This approach is simpler and more direct than the method used in [4]; Blackman et al. had to resort to building a mock surrogate to evaluate error between the interpolator and the data source, since NR simulations are too expensive for this type of acquisition. This complicates the model training process, and the use of variance as a stand-in for error circumvents this step and simplifies our algorithm. However, as a non-spectral method, the specific GP interpolation adopted in this work



(a) Model mismatch with 350 training points.

(b) Model mismatch with 500 training points.

Figure 7: A comparison of the mismatch of the $V : C$ routine at two steps in the acquisition process. This 2D slice of the 3D parameter space is taken with χ_2 is set to 0.99 to include the highest cost regions. Note that for most of the space, mismatch has decreased; however, at the highest cost point in the domain (top right), mismatch has increased significantly, from $\mathcal{E} = 0.03$ to $\mathcal{E} = 0.21$.

converges relatively slowly versus the number of new simulations acquired. We expect that our technique can be applied to scenarios with other empirically-estimated error metrics like the SVD approach.

The second most important difference between the surrogate described in [4] and the model presented here is incorporation of simulation cost. While both approaches seek to minimize the total number of NR simulations required to train a model, our method gives consideration to the variability in cost of the simulations themselves. While we did not directly compare data acquisition cost between Blackman et al.'s model and our $V : C$ model, we did use the variance maximization routine V as a stand-in for this comparison. We found that a cost savings of about tenfold was gained by using the $V : C$ acquisition function. Of course, due to its extremely rapid (exponential) convergence, the asymptotic behavior of a spectral interpolation should be superior. Our goal in this study is simply to demonstrate the utility of incorporating cost estimates into surrogate simulation targeting.

IV. Conclusion

In this paper we have discussed the utility of surrogate models in the gravitational wave domain. Surrogate models have great potential to reduce the cost of astrophysical parameter inference. Not only are they powerful in the context of astrophysics, but in many other fields where data is expensive to acquire do surrogate models offer an optimal route to learning. Acquisition functions, when tailored to the use case, can make the training process for a surrogate model much cheaper and more effective. In our case, deriving an acquisition function which maximized learning per unit cost lead to an savings of about 10 times over a routine which only maximized variance.

Here we demonstrated how to incorporate simulation cost into a concrete strategy for iteratively building a surrogate for gravitational waves generated by black hole binaries. By minimizing the estimated (Gaussian process) variance per unit cost, we show how to target synthetic GW simulations to assemble a comparable-accuracy surrogate at lower cost than a similar cost-neutral approach. At the mismatch scales to 10^{-4} and beyond, we demonstrated our method should be about an order of magnitude less costly than a comparable cost-neutral approach over the same parameter space.

Our results were demonstrative of the specific situations in which cost minimization will yield significant savings. When an application is data-sparse due to high expense and there is high variability in cost over the domain, the method we developed here will offer great utility. Problems which are high-dimensional compound the effect of cost-minimizing acquisition. Further work is needed on acquisition functions in conjunction with other classes of surrogate models. A more rigorous study of the science return of a surrogate model depending on its accuracy could (a) help quantify the downsides of cost-weighted acquisition and (b) open the possiblity of more mature acquisition functions. More complete domains, and correspondingly more complex cost functions, may be investigated. Section V lays out other potential avenues for research in the vein of cost-effective surrogate modeling.

Though we focused on synthetic gravitational wave simulations specifically, our methods could be directly transferable to other data-sparse domains with high-cost simulations, including as a wide variety of applications in astrophysics involving compact object mergers.

V. Future Work

V.1 MOGPs

There are several promising areas for potential follow-up to the work presented here. One closely-related method to pursue is that of multi-output Gaussian processes (MOGPs) [41]. For simplicity, in this model we used a combination of many single-output GPs independently trained on different time steps. However, by doing so we neglect to consider correlation in value between neighboring time steps. MOGPs are an avenue by which we could improve this surrogate model by learning from the relationship between nearby output values for a given waveform, rather than only from a single instant in time. In [11], Conti and O'Hagan extend single-output Gaussian process emulators to multi-dimensional outputs with a particular focus on time series data. They discuss three strategies for doing so:

1. **Ensemble of Single-output GPs** Model each time step as a separate GP, the method used in this thesis. Conti and O'Hagan note that this method provides great flexibility due to the independence of each GP, but does not capture the relationship between nearby time steps, and thus they leave this option out of their testing.

2. **Time input (TI) emulator** Use time as an additional input to a single-output GP. This approach expands the input space by one dimension, associating each output with a specific time in addition to the other inputs. This method will encode the relationship between time and the outputs of interest, but the time to train the GP will be greatly increased due to the multiplication of the number of training points by a factor of the number of time samples.

3. **Multi-output (MO) emulator** Train a multi-output GP, with each output representing the value of interest at time t.

In testing the TI and MO strategies, Conti and O'Hagan find that an MO emulator is able to provide narrower and more reliable confidence intervals when trained on a dynamic mechanistic model. MOGPs are a natural and potentially powerful extension to the work in this thesis.

V.2 Objective Function-Informed Acquisition

Our base acquisition strategy was predicated on the idea that a Gaussian process' variance was a suitable predictor of error. Notably, for any Gaussian process the sites of the variance maxima depend only on the relative positions of available data points, not on the target output value. It is likely that there is a relationship between the shape of the interpolant output and its error;

for example, we may expect more error in a region where the interpolant is changing more rapidly than in an area where the function is nearly flat. This information could be used to place simulations in an even better estimation of error than variance alone. One simple way of incorporating this into our current strategy would be to compute the GP's gradient across the parameter space and multiply its norm into the acquisition function (II.18):

$$A = \frac{\mathbb{V}\,\|\nabla f_{GP}\|}{C} \tag{V.1}$$

In this way areas of more rapid change would be targeted, in conjunction with the preference for high variance and low cost.

V.3 Error Targeting and Resource Allocation

For the purpose of this work, NR was treated as a black box with predetermined cost for a given simulation with desired BBH parameters. Furthermore, no consideration was given to the accuracy of the NR data, nor to the requirements of downstream applications of the surrogate model (i.e. parameter estimation). Decisions about simulation placement would benefit from a more thorough analysis of (a) what accuracy is required and (b) what accuracy is available for training. Presently, the mismatch $\mathcal{E}$ (eqn. II.22) between NR waveforms h_{NR} and the true waveform h_{exact} ranges from 10^{-6} to 10^{-2}, with a median mismatch of around 10^{-4} [5]. In regions of the parameter space where NR is accurate only to $\mathcal{E} = 10^{-2}$, for example, it would be a waste of resources to place enough simulations to emulate h_{NR} to $\mathcal{E} = 10^{-3}$, since our goal is to emulate the true waveform h_{exact}. On the other hand, in regions where NR is able to achieve higher accuracies than needed, it would be possible to run simulations with coarser grids, thereby lowering simulation cost to the point where the mismatch between h_{NR} and h_{exact} is only as small as is needed. Fundamental to this proposal is the need for reliable location-specific error bounds on surrogate models constructed from NR. One simple way for estimating the error would be to leverage a *leave one out* strategy, in which individual training data points are left out, the surrogate model is trained on the remaining points, and the error at the left-out point measured [20]. This method could easily be added to our acquisition strategy and provides location-specific error bounds; furthermore, the bounds would be conservative since the final surrogate will contain all available training data. However, training a surrogate model for every data point at every acquisition step would be quite expensive. There are a number of possible strategies for reducing this cost (e.g. computing error at a sparse selection of data points, or only at a subset of acquisition steps).

There may be more rigorous and computationally inexpensive methods for obtaining error bounds on our surrogate model. For Gaussian processes specifically, recent work has been done to

compute practical error bounds on regression of an underlying function with noise [15]. Here Fiedler et al. argue the need for *frequentist* error bounds in safety-critical applications. While the Gaussian process is a *Bayesian* statistical construct, Bayesian error bounds (in this case, variance) are insufficient for these applications, since they provide no guarantee about worst performance, only a probability distribution. While our application is not safety-critical, hard limits on surrogate model error would allow us to construct a model to deliberate specifications meeting the needs of downstream users, while saving cost by preventing overtraining. We note that Fiedler et al.'s result, too involved to reproduce here, would not be straightforward to apply to our work. The relationship between error in the Gaussian process' estimation of waveform values (amplitude and phase) and the mismatch $\mathcal{E}$ is not immediately apparent and would require its own investigation.

Further complicating this analysis is the fact that an NR-based surrogate model will likely be trained on hybrid waveforms rather than pure NR waveforms. Hybrid waveforms are a combination between GWs produced by a cheaper method (e.g. post-Newtonian/PN) and GWs produced by NR. Hybrids take advantage of the fact that early-inspiral GWs are well-approximated by cheap methods to reduce the NR computation required to produce a full waveform. However, unifying NR and PN waveforms leads to a more involved error analysis [27]. It is possible to make decisions about which stage of the inspiral to unify the two waveforms, with a heavy impact on cost. A more holistic understanding of the mismatch requirements and error propogation through the GW data pipeline–NR $\rightarrow$ hybrid $\rightarrow$ surrogate $\rightarrow$ parameter estimation–could reveal more areas for cost savings.

V.4 Future Error Requirements

While current NR and its surrogates are sufficient for parameter estimation of the data measured at current-generation GW detectors, these methods as they stand are not accurate enough for the next generation of detectors planned for the 2030s. Pürrer and Haster [31] estimate that to prevent bias in parameter estimation with third-generation detectors, NR must improve mismatch by an order of magnitude over present performance, and semianalytical models by at least three orders of magnitude. Based on the convergence rates of surrogate models which rely on local learning (e.g. Gaussian processes), error three orders of magnitude lower than present will require an infeasible amount of training data. Future GW science would benefit greatly from interpolation schemes with faster convergence. We use the term *local* as a shorthand for the concept that when training a local model on data, the goodness of fit only improves in the neighborhood of a given

added datum. In contrast, *global* methods enjoy improved fit across the entire domain as data is added. An example of globally learning interpolators is a class of functions known as spectral methods. Spectral methods involve summing smooth basis functions to approximate an objective function:

$$u(x) = \sum_{k=0}^{N} a_k \phi_k(x) \tag{V.2}$$

where u is the approximating function, x is a point in the domain, N is the number of data points (nodes), ϕ_k are the basis functions, and a_k are coefficients [18].

The relative convergence in terms of root-mean-square error from the objective function is demonstrated in Figure 8. Note that at low data densities, even a suboptimal GP training routine performs better than the global spectral method. However, as more data becomes available, the spectral method achieves far superior convergence. It is noteworthy that spectral methods require a predetermined placement of simulations in a grid, and cannot function without complete population of this grid with data. Furthermore, to avoid poor behavior near the boundaries, spectral methods require more dense sampling at the edges of the domain. This is a burdensome requirement for NR, where cost increases near the borders of high spin and high mass ratio, leading to an inordinately high training cost (for recent work on circumventing this requirement, see [28]). The accuracy achievable with densely trained spectral methods is not yet necessary for GW science (see sec. V.3), and for now the flexibility of GPs and cost-minimizing acquisition mean that local methods are superior. However, eventually spectral methods will have the advantage in the era of GW detectors like LISA [12]. Therefore, it is imperative to quantify the capabilities and shortcomings of globally-learning methods, and to determine what cost-saving measures may be possible.

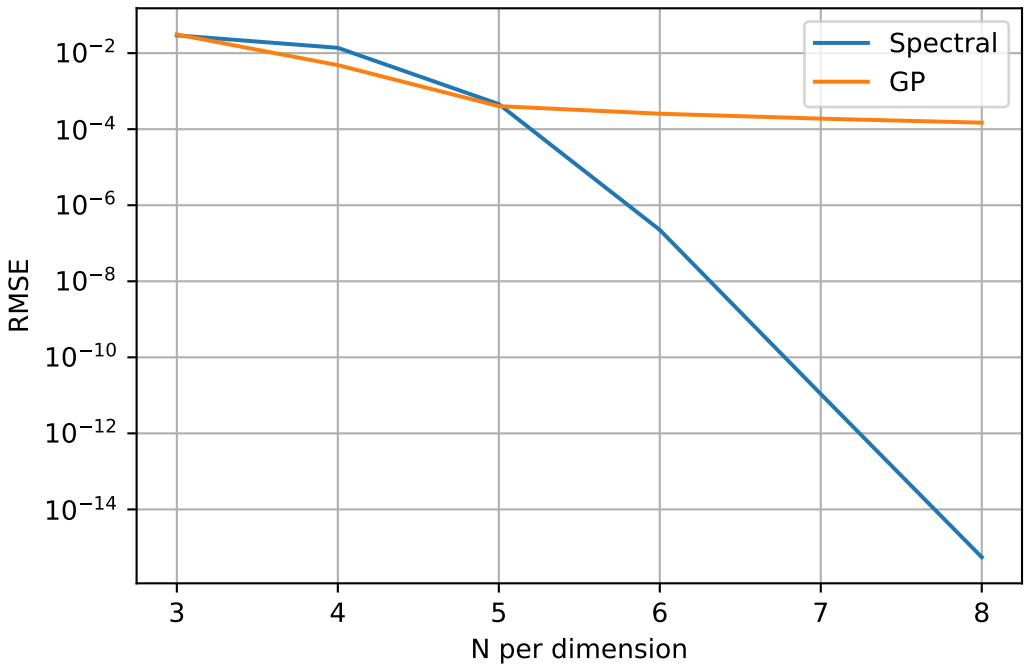


Figure 8: A convergence comparison between a globally learning method (spectral) vs locally learning (GP). [Eq, (II.22)]. These models were trained on a 2D Bessel function as a toy function of smoothness comparable to the functions we interpolate in GW applications. Note that this GP was trained on evenly spaced data, and better performance is achieved by our variance-maximizing routines.

VI. Acknowledgments

I am grateful for the patient and insightful mentorship of my advisor, Dr. Richard O'Shaughnessy. I also thank my family for their lifelong support. My Uncle David deserves a special thanks for (mostly) understanding my work and for his thought-provoking conversations.

VII. Appendix

VII.1 Equations for Gradient Methods

The Jacobian of the variance of a Gaussian Process using a squared exponential kernel was derived for use in a Python optimization routine:

$$J_m(X_*) = \frac{1}{l^2}\left\{\left(K^{-1} + (K^{-1})^T\right)K_*\right\}^T K_*\left(X_*^{(m,:)} - X^{(m,:)}\right) \tag{VII.1}$$

where J_m is the m*th* element of the Jacobian, K is the kernel evaluated on the training data, $K(X, X)$, K_* is the kernel evaluated at a given point, $K(X_*, X)$, and the superscript $(m, :)$ represents the entire m*th* row of a given matrix.

VIII. Bibliography

References

[1] B. P. Abbott, R. Abbott, T. D. Abbott, M. R. Abernathy, F. Acernese, K. Ackley, C. Adams, T. Adams, P. Addesso, R. X. Adhikari, and et al., *Observation of Gravitational Waves from a Binary Black Hole Merger*, Phys. Rev. Lett**116** (2016), 061102–+.

[2] F Acernese, M Agathos, K Agatsuma, D Aisa, N Allemandou, A Allocca, J Amarni, P Astone, G Balestri, G Ballardin, F Barone, J-P Baronick, M Barsuglia, A Basti, F Basti, Th S Bauer, V Bavigadda, M Bejger, M G Beker, C Belczynski, D Bersanetti, A Bertolini, M Bitossi, M A Bizouard, S Bloemen, M Blom, M Boer, G Bogaert, D Bondi, F Bondu, L Bonelli, R Bonnand, V Boschi, L Bosi, T Bouedo, C Bradaschia, M Branchesi, T Briant, A Brillet, V Brisson, T Bulik, H J Bulten, D Buskulic, C Buy, G Cagnoli, E Calloni, C Campeggi, B Canuel, F Carbognani, F Cavalier, R Cavalieri, G Cella, E Cesarini, E Chassande-Mottin, A Chincarini, A Chiummo, S Chua, F Cleva, E Coccia, P-F Cohadon, A Colla, M Colombini, A Conte, J-P Coulon, E Cuoco, A Dalmaz, S D'Antonio, V Dattilo, M Davier, R Day, G Debreczeni, J Degallaix, S Delé glise, W Del Pozzo, H Dereli, R De Rosa, L Di Fiore, A Di Lieto, A Di Virgilio, M Doets, V Dolique, M Drago, M Ducrot, G Endrőczi, V Fafone, S Farinon, I Ferrante, F Ferrini, F Fidecaro, I Fiori, R Flaminio, J-D Fournier, S Franco, S Frasca, F Frasconi, L Gammaitoni, F Garufi, M Gaspard, A Gatto, G Gemme, B Gendre, E Genin, A Gennai, S Ghosh, L Giacobone, A Giazotto, R Gouaty, M Granata, G Greco, P Groot, G M Guidi, J Harms, A Heidmann, H Heitmann, P Hello, G Hemming, E Hennes, D Hofman, P Jaranowski, R J G Jonker, M Kasprzack, F Kéfélian, I Kowalska, M Kraan, A Królak, A Kutynia, C Lazzaro, M Leonardi, N Leroy, N Letendre, T G F Li, B Lieunard, M Lorenzini, V Loriette, G Losurdo, C Magazzù, E Majorana, I Maksimovic, V Malvezzi, N Man, V Mangano, M Mantovani, F Marchesoni, F Marion, J Marque, F Martelli, L Martellini, A Masserot, D Meacher, J Meidam, F Mezzani, C Michel, L Milano, Y Minenkov, A Moggi, M Mohan, M Montani, N Morgado, B Mours, F Mul, M F Nagy, I Nardecchia, L Naticchioni, G Nelemans, I Neri, M Neri, F Nocera, E Pacaud, C Palomba, F Paoletti, A Paoli, A Pasqualetti, R Passaquieti, D Passuello, M Perciballi, S Petit, M Pichot, F Piergiovanni, G Pillant, A Piluso, L Pinard, R Poggiani, M Prijatelj, G A Prodi, M Punturo, P Puppo, D S Rabeling, I Rácz, P Rapagnani, M Razzano, V Re, T Regimbau, F Ricci, F Robinet, A Rocchi, L Rolland, R Romano, D Rosińska, P Ruggi, E Saracco, B Sassolas, F Schimmel, D Sentenac, V Sequino, S Shah, K Siellez, N Straniero, B Swinkels, M Tacca, M Tonelli, F Travasso, M Turconi, G Vajente, N van Bakel, M van Beuzekom, J F J van den

Brand, C Van Den Broeck, M V van der Sluys, J van Heijningen, M Vasúth, G Vedovato, J Veitch, D Verkindt, F Vetrano, A Viceré, J-Y Vinet, G Visser, H Vocca, R Ward, M Was, L-W Wei, M Yvert, A Zadro żny, and J-P Zendri, *Advanced virgo: a second-generation interferometric gravitational wave detector*, Classical and Quantum Gravity **32** (2014), no. 2, 024001.

[3] Gregory Ashton, Moritz Hübner, Paul D. Lasky, Colm Talbot, Kendall Ackley, Sylvia Biscoveanu, Qi Chu, Atul Divakarla, Paul J. Easter, Boris Goncharov, Francisco Hernandez Vivanco, Jan Harms, Marcus E. Lower, Grant D. Meadors, Denyz Melchor, Ethan Payne, Matthew D. Pitkin, Jade Powell, Nikhil Sarin, Rory J. E. Smith, and Eric Thrane, *BILBY: A User-friendly Bayesian Inference Library for Gravitational-wave Astronomy*, ApJS**241** (2019), no. 2, 27.

[4] J. Blackman, S. E. Field, M. A. Scheel, C. R. Galley, D. A. Hemberger, P. Schmidt, and R. Smith, *A Surrogate model of gravitational waveforms from numerical relativity simulations of precessing binary black hole mergers*, Phys. Rev. D**95** (2017), no. 10, 104023.

[5] J. Blackman, S. E. Field, M. A. Scheel, C. R. Galley, C. D. Ott, M. Boyle, L. E. Kidder, H. P. Pfeiffer, and B. Szilágyi, *Numerical relativity waveform surrogate model for generically precessing binary black hole mergers*, Phys. Rev. D**96** (2017), no. 2, 024058.

[6] Jonathan Blackman, Scott E. Field, Chad R. Galley, Béla Szilágyi, Mark A. Scheel, Manuel Tiglio, and Daniel A. Hemberger, *Fast and accurate prediction of numerical relativity waveforms from binary black hole coalescences using surrogate models*, Phys. Rev. Lett. **115** (2015), 121102.

[7] A. Bohé, L. Shao, A. Taracchini, A. Buonanno, S. Babak, I. W. Harry, I. Hinder, S. Ossokine, M. Pürrer, V. Raymond, T. Chu, H. Fong, P. Kumar, H. P. Pfeiffer, M. Boyle, D. A. Hemberger, L. E. Kidder, G. Lovelace, M. A. Scheel, and B. Szilágyi, *Improved effective-one-body model of spinning, nonprecessing binary black holes for the era of gravitational-wave astrophysics with advanced detectors*, Phys. Rev. D**95** (2017), no. 4, 044028.

[8] P. Canizares, S. E. Field, J. Gair, V. Raymond, R. Smith, and M. Tiglio, *Accelerated Gravitational Wave Parameter Estimation with Reduced Order Modeling*, Phys. Rev. Lett**114** (2015), no. 7, 071104.

[9] Joan Centrella, John G. Baker, Bernard J. Kelly, and James R. van Meter, *Black-hole binaries, gravitational waves, and numerical relativity*, Reviews of Modern Physics **82** (2010), no. 4, 3069–3119.

[10] H. S. Cho, E. Ochsner, R. O'Shaughnessy, C. Kim, and C. H. Lee, *Gravitational waves from BH-NS binaries: Phenomenological Fisher matricies and parameter estimation using higher harmonics*, Phys. Rev. D**87** (2013), 024004.

[11] Stefano Conti and Anthony O'Hagan, *Bayesian emulation of complex multi-output and dynamic computer models*, Journal of Statistical Planning and Inference **140** (2010), no. 3, 640–651.

[12] Curt Cutler and Michele Vallisneri, *Lisa detections of massive black hole inspirals: Parameter extraction errors due to inaccurate template waveforms*, Phys. Rev. D **76** (2007), 104018.

[13] Karl Daningburg and Richard O'Shaughnessy, *Cost minimization in acquisition for gravitational wave surrogate modeling*, 2022.

[14] Zoheyr Doctor, Ben Farr, Daniel E. Holz, and Michael Pürrer, *Statistical gravitational waveform models: What to simulate next?*, Phys. Rev. D**96** (2017), no. 12, 123011.

[15] Christian Fiedler, Carson W Scherer, and Sebastian Trimpe, *Practical and rigorous uncertainty bounds for gaussian process regression*, Proceedings of the AAAI Conference on Artificial Intelligence, vol. 35, AAAI, 2021, pp. 7439–7447.

[16] S. E. Field, C. R. Galley, J. S. Hesthaven, J. Kaye, and M. Tiglio, *Fast Prediction and Evaluation of Gravitational Waveforms Using Surrogate Models*, Physical Review X **4** (2014), no. 3, 031006.

[17] Scott E. Field, Chad R. Galley, Jan S. Hesthaven, Jason Kaye, and Manuel Tiglio, *Fast prediction and evaluation of gravitational waveforms using surrogate models*, Phys. Rev. X **4** (2014), 031006.

[18] Bengt Fornberg, *A practical guide to pseudospectral methods*, Cambridge Monographs on Applied and Computational Mathematics, Cambridge University Press, 1996.

[19] M. Hannam, P. Schmidt, A. Bohé, L. Haegel, S. Husa, F. Ohme, G. Pratten, and M. Pürrer, *Simple Model of Complete Precessing Black-Hole-Binary Gravitational Waveforms*, Phys. Rev. Lett**113** (2014), no. 15, 151101.

[20] Gareth James, Daniela Witten, Trevor Hastie, and Robert Tibshirani, *An introduction to statistical learning: With applications in r*, second ed., Springer, 2021.

[21] S. Khan, S. Husa, M. Hannam, F. Ohme, M. Pürrer, X. J. Forteza, and A. Bohé, *Frequency-domain gravitational waves from nonprecessing black-hole binaries. II. A phenomenological model for the advanced detector era*, Phys. Rev. D**93** (2016), no. 4, 044007.

[22] J. Lange, R. O'Shaughnessy, and M. Rizzo, *Rapid and accurate parameter inference for coalescing, precessing compact binaries*, Submitted to PRD; available at arxiv:1805.10457 (2018).

[23] LIGO Scientific Collaboration, J. Aasi, B. P. Abbott, R. Abbott, T. Abbott, M. R. Abernathy, K. Ackley, C. Adams, T. Adams, P. Addesso, and et al., *Advanced LIGO*, Class. Quant. Grav. **32** (2015), no. 7, 074001.

[24] Carlos Lousto, personal communication, 2021.

[25] Carlos O. Lousto and James Healy, *Exploring the Small Mass Ratio Binary Black Hole Merger via Zeno's Dichotomy Approach*, Phys. Rev. Lett. **125** (2020), no. 19, 191102.

[26] S. Prince M. O. Ahmed, *Tutorial #8: Bayesian optimization*, June 2020, Online; accessed 15-August-2022.

[27] Ilana MacDonald, Samaya Nissanke, and Harald P Pfeiffer, *Suitability of post-newtonian/numerical-relativity hybrid waveforms for gravitational wave detectors*, Classical and Quantum Gravity **28** (2011), no. 13, 134002.

[28] Akil Narayan and Dongbin Xiu, *Stochastic collocation methods on unstructured grids in high dimensions via interpolation*, SIAM Journal on Scientific Computing **34** (2012), no. 3, A1729–A1752.

[29] Serguei Ossokine, Alessandra Buonanno, Sylvain Marsat, Roberto Cotesta, Stanislav Babak, Tim Dietrich, Roland Haas, Ian Hinder, Harald P. Pfeiffer, Michael Pürrer, Charles J. Woodford, Michael Boyle, Lawrence E. Kidder, Mark A. Scheel, and Béla Szilágyi, *Multipolar effective-one-body waveforms for precessing binary black holes: Construction and validation*, Phys. Rev. D**102** (2020), no. 4, 044055.

[30] Geraint Pratten, Cecilio García-Quirós, Marta Colleoni, Antoni Ramos-Buades, Héctor Estellés, Maite Mateu-Lucena, Rafel Jaume, Maria Haney, David Keitel, Jonathan E. Thompson, and Sascha Husa, *Computationally efficient models for the dominant and subdominant harmonic modes of precessing binary black holes*, Phys. Rev. D**103** (2021), no. 10, 104056.

[31] Michael Pürrer and Carl-Johan Haster, *Gravitational waveform accuracy requirements for future ground-based detectors*, Phys. Rev. Research **2** (2020), 023151.

[32] C.E. Rasmussen and C.I.K. Williams, *Gaussian Processes for Machine Learning*, The MIT Press, 2006.

[33] H. L. Shipman, *The Implausible History of Triple Star Models for Cygnus X-1: Evidence for a Black Hole*, Astrophysical Letters**16** (1975), 9.

[34] R. Smith, S. E. Field, K. Blackburn, C. Haster, M. P?rrer, V. Raymond, and P. Schmidt, *Fast and Accurate Inference on Gravitational Waves from Precessing Compact Binaries*, (arXiv:1604.08253) (2016).

[35] Niranjan Srinivas, Andreas Krause, Sham M. Kakade, and Matthias W. Seeger, *Information-theoretic regret bounds for gaussian process optimization in the bandit setting*, IEEE Transactions on Information Theory **58** (2012), 3250–3265.

[36] The LIGO Scientific Collaboration, the Virgo Collaboration, B. P. Abbott, R. Abbott, T. D. Abbott, S. Abraham, F. Acernese, K. Ackley, C. Adams, V. B. Adya, and et al., *GWTC-3: Compact Binary Coalescences Observed by LIGO and Virgo During the Second Half of the Third Observing Run*, Available as LIGO-P2000318.

[37] The LIGO Scientific Collaboration, The Virgo Collaboration, B. P. Abbott, R. Abbott, T. D. Abbott, F. Acernese, K. Ackley, C. Adams, T. Adams, P. Addesso, and et al., *GWTC-1: A Gravitational-Wave Transient Catalog of Compact Binary Mergers Observed by LIGO and Virgo during the First and Second Observing Runs*, Phys. Rev. X**9** (2019), no. 3, 031040.

[38] The LIGO Scientific Collaboration, the Virgo Collaboration, R. Abbott, T. D. Abbott, S. Abraham, F. Acernese, K. Ackley, A. Adams, C. Adams, R. X. Adhikari, V. B. Adya, C. Affeldt, and et al., *GWTC-2: Compact Binary Coalescences Observed by LIGO and Virgo during the First Half of the Third Observing Run*, Physical Review X **11** (2021), no. 2, 021053.

[39] Vijay Varma, Scott E. Field, Mark A. Scheel, Jonathan Blackman, Davide Gerosa, Leo C. Stein, Lawrence E. Kidder, and Harald P. Pfeiffer, *Surrogate models for precessing binary black hole simulations with unequal masses*, Physical Review Research **1** (2019), no. 3, 033015.

[40] J. Veitch, V. Raymond, B. Farr, W. M. Farr, P. Graff, S. Vitale, B. Aylott, K. Blackburn, N. Christensen, M. Coughlin, W. D. Pozzo, F. Feroz, J. Gair, C. Haster, V. Kalogera, T. Littenberg, I. Mandel, R. O'Shaughnessy, M. Pitkin, C. Rodriguez, C. Röver, T. Sidery, R. Smith, M. V. D. Sluys, A. Vecchio, W. Vousden, and L. Wade, *Robust parameter estimation for compact binaries with ground-based gravitational-wave observations using LALInference*, Phys. Rev. D**91** (2015), 042003.

[41] Mauricio A. Álvarez, Lorenzo Rosasco, and Neil D. Lawrence, *Kernels for vector-valued functions: A review*, 2012.