Gravitational Waves Inference

Kyle Lee [1] **Yuechuan Yang** [1]

Abstract

Gravitational wave parameter estimation is an area of research that has been explored for many years. Researchers have long worked on developing methods for inferring the properties of gravitational wave sources from the signals they produce. In recent years, the application of neural networks as data-driven surrogates for replacing costly simulations and forward models has gained popularity. In this paper, we focus on two machine learning approaches, the Fourier Neural Mapping and an Autoencoder, to approximate the complex and computationally expensive mapping between physical parameters (e.g. masses and spins) of gravitational waves to observable strain signals. Our methodology aims to exploit the intrinsic low-dimensionality of this mapping and can be validated through formulating a Bayesian inverse problem, where a surrogate model is used to compute a likelihood function and derive a posterior distribution for parameter estimation. We find that the Fourier Neural Mapping performs effectively, showing a significant reduction in relative L_2 errors over time, whereas the autoencoder approach struggles more as a lot of relevant information is thrown away during dimensionality reduction. We conclude that Fourier representations excel at capturing such oscillatory signals, suggesting its potential for future surrogate modeling efforts in gravitational wave inference.

1. Introduction

As a direct consequence of Albert Einstein's general theory of relativity, gravitational waves are often described as "ripples" generated from the acceleration of massive objects with gravitational fields such as binary black hole systems (Barausse & Coauthors, 2024). These ripples propagate through the universe at the speed of light, carrying invaluable information about the fundamental nature of gravity. The gravitational wave signal from a binary black hole merger was first detected in 2015 by Laser Interferometer Gravitational-Wave Observatory (LIGO), a groundbreaking achievement that validated decades of theoretical predictions (Abbott et al., 2016).

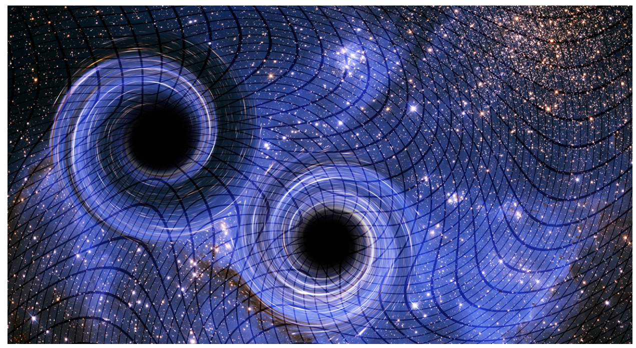

Figure 1. These ripples in spacetime propagate outward, carrying information about the mass, spin, and dynamics of the merging objects, as predicted by Einstein's general theory of relativity (Getty Images, 2024) (Abbott et al., 2016).

Gravitational waves carry a wealth of information about their sources, including the masses, spins, and orbital dynamics. This paper will primarily concentrate on binary black hole systems, in which the three most determining factors that govern the dynamics of the resulting waveforms are the two aligned spin magnitudes and the mass ratio. These factors directly affect the waveform's frequency, amplitude, and duration, which is crucial for understanding the properties of merging black holes (Tagawa et al., 2023). The mass ratio determines each object's relative contribution to the system's overall dynamics. A higher mass ratio implies that one black hole dominates the system's gravitational influence, whereas a smaller ratio means that both objects contribute equally to the dynamics. Meanwhile, the aligned spin magnitudes quantify the extent to which the black holes' spins are aligned with the orbital angular momentum. These aligned spins showcase the orbital evolution of the system and influence how quickly the black holes spiral toward each other (Mapelli, 2020).

In the past few decades, gravitational wave inference has become a popular area of research, as gravitational waves carry important information that light or electromagnetic

[1] Department of Computer Science, Brown University, Rhode Island, United States of America. Correspondence to: Kyle Lee (first author) and Yuechuan Yang <kyle_k_lee@brown.edu, yuechuan_yang@brown.edu>.

signals cannot provide (Escrig et al., 2024). However, extracting meaningful information from these signals remains a challenging and computationally expensive task because waveforms are highly non-linear and depend on a high-dimensional input space. For binary black hole systems, small changes in the input parameter such as mass can lead to significant variations in waveforms, requiring extensive computational power to explore the full parameter space (Xiong et al., 2024).

To address these challenges, machine learning and data-driven surrogate models provide an elegant yet robust approach to gravitational wave inference. Neural networks such as a nonlinear ResNet have shown great promise for learning in data-sparse environments (O'Leary-Roseberry et al., 2021), offering a powerful framework for modeling complex relationships in high-dimensional spaces. By exploiting the low-dimensionality representation of gravitational waves, these neural networks can dramatically accelerate the gravitational wave inference, making such inference more computationally feasible. Our primary contributions comprise of exploring two specific neural network-based approaches to gravitational wave inference: **Fourier Neural Mappings** (Tancik et al., 2020) and **Autoencoders** (Michelucci, 2022). Both methods are designed to address the computational bottlenecks associated with waveform generation and parameter estimation, while potentially maintaining accuracy and efficiency. Fourier Neural Mappings have been shown to efficiently solve a family of PDEs whose inputs and outputs may be continuous functions, and autoencoders are well-known for their dimensionality reduction properties, so they represent promising avenues to help improve surrogate models for gravitational wave inference where waveforms are highly complex.

2. Related Work

Current literature on gravitational wave surrogate modeling and the usage of machine learning consist of a wide variety of different approaches, ranging from regressions to neural differential equations. Although each approach has its own advantages and disadvantages, they all share the common goal of learning from complex, high-dimensional data and uncovering insights that may not be possible or are too computationally expensive to achieve through conventional modeling techniques. The models that perform the best might incur high computational costs, but exploiting samples from cheaper models has the potential of accelerating inference and extracting general insights that may be useful for more expensive models.

Numerical relativity (NR) simulations are widely recognized as one of the most accurate method for generating binary black hole gravitational waveforms. However, NR simulations are extremely costly, making them impractical for tasks such as parameter estimation. Thankfully, surrogate models based on NR waveforms have emerged as a promising solution, offering both efficiency and accuracy for gravitational wave inference. The intuition behind this method is to leverage a set of precomputed NR waveforms to extrapolate new waveforms according to post-Newtoninan theory (Varma et al., 2019). This method begins by employing a greedy reduced basis approach, which systematically selects a set of desired parameters. According to the training results, more greedy parameters implies higher accuracy. With more than 120 greedy parameters, the mismatch will be below 10^{-7}. In this way, the project errors for the entire dataset are guaranteed to be below the tolerance threshold.

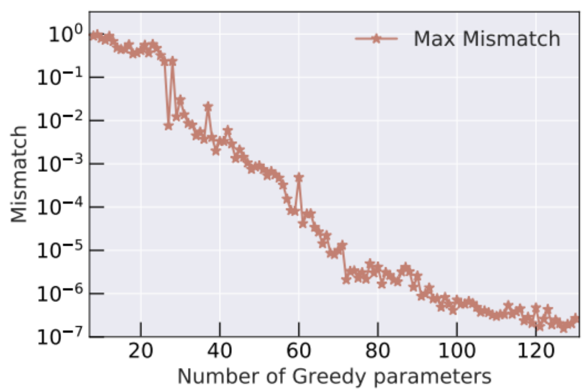

Figure 2. The surrogate model has shown improved convergence to the target waveforms as the number of greedy parameters increases. This behavior further showcases the power of the greedy reduced basis approach in selecting parameters that optimally represent the underlying waveform space. As more greedy parameters are incorporated, the projection errors across the dataset decrease significantly, ensuring that the surrogate accurately approximates the target waveforms. According to (Varma et al., 2019), with over 120 greedy parameters, the mismatch can be reduced to below 10^{-7}, demonstrating promising accuracy and reliability in reproducing numerical relativity waveforms.

Building on the efficiency of surrogate models, the development of efficient-one-body (**EOB**) models has further advanced the field of gravitational wave inference (Nagar et al., 2018). EOB models combine the analytical insights of post-Newtonian theory with the numerical precision of NR simulations, enabling accurate waveform generation over a wide range of binary black hole systems. Moreover, recent advancements, such as the inclusion of spin effects, tidal interactions, and post-merger dynamics, have enhanced the robustness of EOB frameworks for modeling complex systems. For example, TEOBResumS (an EOB-based model), exemplifies these innovations by incorporating improved descriptions of spin interactions and tidal effects, making it a powerful tool for gravitational wave data analysis.

Previously, researchers pioneered the use of machine learning with Convolutional Neural Networks (CNNs), which take time series input to be analyzed for detections and representation of gravitational wave signals. This approach is conventionally known as **Deep Filtering**, which was illustrated using simulated LIGO noise (George & Huerta, 2018). This method can be extended and repurposed for both detection and inference of gravitational waves from binary black hole mergers through continuous data streams from multiple LIGO detectors. CNNs for real-time gravitational wave analysis can extend the existing gravitational wave detection algorithm to higher dimensions and a deeper parameter space, as the intrinsic scalability of neural networks can overcome the curse of dimensionality. Predicting any number of parameters such as amplitude of waveforms is as simple as adding an additional neuron for each new parameter to the final layer and training the noisy waveforms with the corresponding labels (George & Huerta, 2018).

To handle the high dimensionality problem, researchers have previously constructed a time-domain model for gravitational waveforms from binary black holes merged with machine learning techniques, known as **mlgw** (Schmidt et al., 2021). In particular, the waveform's amplitude and phase evolution are captured using principal component analysis (PCA). Remarkably, a single waveform generation is sped up by a factor of around 10 to 50, depending on the binary mass and initial waveform's frequency. Moreover, **mlgw** displays a more concrete mathematical expression for the waveforms and its gradients with respect to the orbital parameters, which is useful for future gravitational wave data analysis.

DINGO (Deep Inference for Gravitational-Wave Observations) (Dax et al., 2021) creates a new standard for fast-and-accurate inference of physical parameters of detected gravitational waves, enabling efficient data analysis without compromising the accuracy. The main distinction between DINGO and traditional methods lies in that neural networks condition not only the event strain data d but also the PSD of the detector noise S_n; in other words, $q(\theta|d, S_n)$ represents a conditional distribution that relies on the data and the detector noise. Then with the GNPE (group equivariant neural posterior estimation) approach, explicit knowledge of time-translation symmetry can be used to simplify the data representation and allow the neural network to focus more on the nontrivial parameters.

Only a single time series of waveform data is necessary to help reconstruct the equations of motion governing a binary black hole system. With a class of differential equations parameterized by feed-forward neural networks, a space of mechanical models can be constructed and a physics-informed constrained optimization problem can then be created to minimize the waveform errors. With this data driven approach, the waveform data can be an indicator of

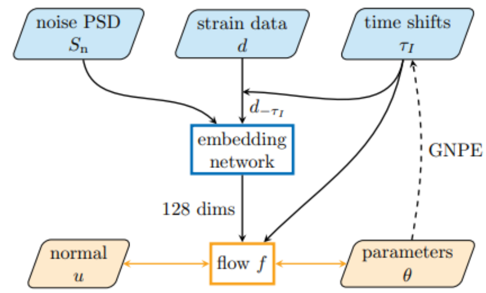

Figure 3. The posterior distribution is illustrated in terms of an invertible normalizing flow (orange), taking normally-distributed random variables u into posterior samples θ. The flow itself depends on a (compressed) representation of the noise properties S_n and the data d, as well as an estimate τ_1 of the coalescence time in each detector I. The data are time-shifted by τ_1 to simplify the representation. (Dax et al., 2021)

the dynamics of binary black hole systems (Keith et al., 2021). This approach is also applicable to various kinds of environments such as the extreme and comparable mass ratio systems in eccentric and non-eccentric orbits.

Another approach to deal with the high dimensionality issue is to use a **Mixture of Experts** framework that divides the parameter space of the binary black hole systems into smaller regions and trains specific models on each localized region (Schmidt et al., 2021; Jacobs et al., 1991). This is especially useful, as it can learn time-domain parameterized signals with high accuracy while reducing the computational complexity compared to more traditional approaches. In the most general setting (regression), Mixture of Experts performs a weighted combination of L linear regressions as follows (Schmidt et al., 2021):

$$y(\mathbf{x}) = \sum_{l=1}^{L} (W^T\mathbf{x})_l \cdot \frac{e^{(V^T\mathbf{x})_l}}{\sum_{l'=1}^{L} e^{(V^T\mathbf{x})'_l}}.$$

The former term represents a linear regression, whereas the latter term resembles the **Softmax function** (acts as a gated mechanism), which essentially normalizes the output to a probability distribution. Mixture of Experts is found to maintain a good balance between "simplicity" and "flexibility" and can be useful for future work where there are hundreds of parameters.

3. Methodology

3.1. Overview

Conventionally, researchers have relied heavily on Bayesian inference frameworks such as **Bilby** for gravitational wave analysis. Bilby, a Bayesian inference library for performing parameter estimation, operates within a physics-driven paradigm, leveraging parameterized waveform templates and Bayesian techniques to infer the properties of gravitational wave signals, such as the masses, spins, and distances of merging astrophysical objects. In recent years, driven by advancements in machine learning and deep learning techniques, data-driven surrogate models have emerged as a promising alternative for gravitational wave inference. Unlike traditional physics-based methods, data-driven approaches learn directly from the data source itself, bypassing the need for precomputed frameworks. By learning patterns directly from observed data, these models can generalize to signals with noise or features not explicitly accounted for in standard, structured methods. This flexibility has positioned data-driven models as a new trend in gravitational wave research, offering faster inference times and the potential to uncover previously undetectable signals. In our work, we only use Bilby to generate our training data instead of using its built-in models — rather, we use our surrogate model for training.

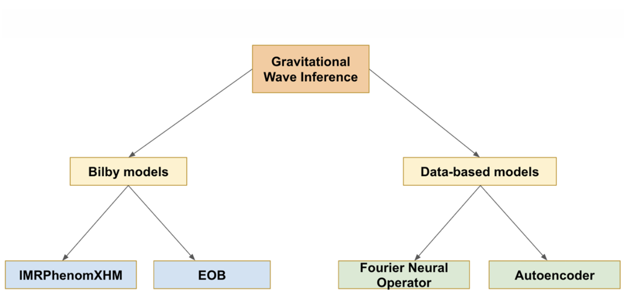

Figure 4. Comparison between Bilby models and data-based models for gravitational wave inference. Bilby models rely on post-Newtonian approximations and effective-one-body (EOB) frameworks to generate gravitational waveforms. These models incorporate physical insights and are often informed by numerical relativity (NR) results to improve accuracy. In contrast, data-based models leverage machine learning techniques and waveform datasets to infer gravitational wave signals efficiently. While Bilby models offer robustness grounded in theoretical physics, data-based models provide enhanced computational efficiency and the ability to approximate complex waveforms without relying on explicit physical properties. This comparison highlights the trade-offs between accuracy, interpretability, and computational cost when it comes to selecting models for gravitational wave data analysis.

3.2. Spectrograms

Researchers commonly create spectrograms to improve the visualization of gravitational waves, as they provide a clear time-frequency representation of the signal. In addition to plotting the time-domain and frequency-domain strains as seen in Figure 5, spectrograms can help them understand the dynamics of gravitational waves, specifically how the frequency of gravitational wave signals evolves over time (Dey et al., 2024). During the inspiral phase, for example, the frequency of the waveforms will increase as two objects spiral closer to each other. This phenomenon is often referred as a **chirp signal**, where both the frequency and amplitude will increase until the two objects merge. Spectrograms can accurately capture these fluctuations, providing a more intuitive understanding of the underlying physics (Higashino & Tsujikawa, 2023).

To generate the waveform representations with spectrograms, we leverage **GWpy**, a Python package for gravitational wave detection, and **IMRPhenomXHM**, a state-of-the-art frequency-domain model for the inspiral, merger, and ringdown of quasi-circular non-precessing black hole binaries (García-Quirós et al., 2020). The model is trained with up to 90,000 waveforms, and the training accuracy is calculated using the Wasserstein metric for each parameter. These spectrograms, which encode the time-frequency structure of the waveforms, can serve as the input to various machine learning models, including our autoencoder.

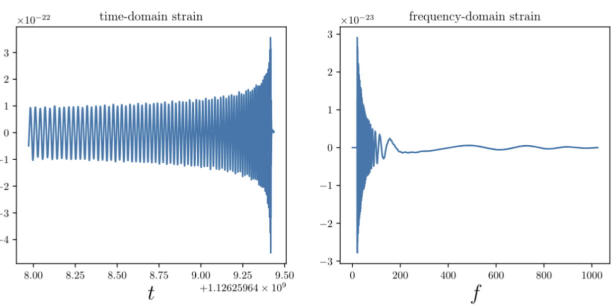

Figure 5. The left plot illustrates the **time-domain strain** of a gravitational wave signal as a function of time. The signal first begins with a low-frequency oscillation, corresponding to the inspiral phase of two objects. The sharp rise at the end corresponds to the coalescence and ringdown phases, where the system emits its most powerful gravitational waves. The right plot shows the **frequency-domain strain** of a gravitational wave signal. The sharp peak at low frequencies corresponds to the dominant contributions from the inspiral phase, where the frequency increases gradually as two compact objects spiral inward. The amplitude decreases at higher frequencies, displaying the late spiral and merger phases. Both figures are generated using **IMRPhenomXHM**.

3.3. Autoencoders

Autoencoders, consisting of an encoder and a decoder, provide a robust approach to handle high-dimensional, noisy datasets, making them particularly suitable for gravitational wave analysis. They can effectively denoise irrelevant information, reduce the dimensionality of the waveforms, and extract crucial features while preserving the essential structure of the data. In the context of gravitational wave inference, the encoder compresses the waveform data into a low-dimensional latent space. This latent space acts as a compact representation that captures only the fundamental features of the waveforms, effectively filtering out noise and redundancies. The decoder then reconstructs the original waveforms from this low-dimensional representation, ensuring that only the essential, noise-free features are retained.

Here, both the input and output of the autoencoder are the spectrogram representations of the gravitational waveforms, enabling a direct comparison between the original and reconstructed waveforms. By analyzing the differences between the input and output, we can assess how effectively the model represents the waveforms and identify whether the process of dimensionality reduction compromises any critical features of the signals. The key question we aim to consider is: "Does the autoencoder's compression of waveform data into a lower-dimensional space lead to the loss of any important physical information?"

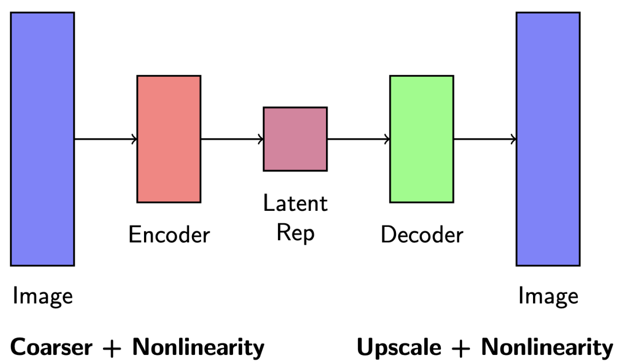

Figure 6. Starting with the spectrogram representation of a waveform, the model processes the data through an encoder to obtain a low-dimensional latent representation. This coarse, nonlinear representation is then passed through a decoder to reconstruct the original image in an upscaled, nonlinear form.

Algorithm 1: Encoder Architecture

```
Class Autoencoder
    Initialize the Autoencoder;
    Define encoder as a sequential layer stack;
      1. Conv2d(1, 16, kernel_size=3, stride=1, padding=1);
      2. ReLU();
      3. MaxPool2d(kernel_size=2, stride=2);
      4. Conv2d(16, 8, kernel_size=3, stride=1, padding=1);
      5. ReLU();
      6. MaxPool2d(kernel_size=2, stride=2);
```

Figure 7. This pseudocode illustrates the encoder architecture: a 2D convolution is applied to the spectrograms, followed by ReLU activation and a max-pooling layer. The implementation is adapted from the Udacity repository (Bank et al., 2021) and modified to suit our specific data representations.

Algorithm 2: Decoder Architecture and Forward Pass

```
Class Autoencoder (continued)
    Define decoder as a sequential layer stack;
      1. ConvTranspose2d(8, 16, kernel_size=2, stride=2,
     padding=0, output_padding=1);
      2. ReLU();
      3. ConvTranspose2d(16, 1, kernel_size=2, stride=2,
     padding=0, output_padding=0);
      4. Sigmoid();

Forward Pass:;
    Define forward(self, x);
      1. x = self.encoder(x);
      2. x = self.decoder(x);
      3. return x;
```

Figure 8. This pseudocode from the Udacity repository (Bank et al., 2021) illustrates the decoder architecture. To match the dimensionality of matrix operations, a ConvTranspose2D operation is used instead of Conv2D. In the final layer, a sigmoid activation replaces the ReLU activation. The code also details the forward pass, which first processes data through the encoder and then reconstructs the original representation via the decoder.

3.4. Fourier Neural Mapping (FNM)

It is important to note that the usage of neural networks intrinsically comes with some challenges, such as requiring an immense amount of training data. However, FNMs are designed to focus on exploiting the intrinsic structure of data more efficiently, so they have the potential of requiring less data compared to generic neural networks or autoencoders. In fact, FNMs are particularly effective for solving problems involving smooth, oscillatory, or periodic signals because they use Fourier features or operators to represent the underlying functions. In our experimental setting, our system parameters (x) are the mass ratio q and the aligned spin magnitudes χ_1 and χ_2, while the theoretical waveforms are the system outputs (y); we aim to learn the mapping from $x \to y$. Formally, for a sequence of layers $\{\mathcal{L}_t\}$, and mappings $\mathcal{Q}$ and $\mathcal{S}$, we define the Fourier Neural Mapping as the following composition:

$$\Psi^{(\mathrm{FNM})} := Q \circ \mathcal{G} \circ \mathcal{L}_{T-1} \circ \cdots \circ \mathcal{L}_2 \circ \mathcal{D} \circ S.$$

In practice, $\mathcal{G}$ is known as the *linear functional layer* (maps functions to vectors) and $\mathcal{D}$ is known as the *linear decoder layer* (maps vectors to functions); hence, $\Psi^{(\mathrm{FNM})}$ maps vectors to vectors. At a high level, the linear functional layer $\mathcal{G}$ takes a vector-valued function and integrates it against a fixed matrix-valued function to produce a finite vector output. In contrast, the linear decoder layer $\mathcal{D}$ takes a finite vector as input and multiplies it by a fixed matrix-valued function to produce an output function (Huang et al., 2024).

Meanwhile, each $\mathcal{L}_t$ represents a *nonlinear map* (maps functions to functions), where

$$(\mathcal{L}_t(v))(x) = \sigma_t \left(\underbrace{W_t v(x)}_{\text{weight}} + \underbrace{(\mathcal{K}_t v)(x)}_{\text{convolution operator}} + \underbrace{b_t(x)}_{\text{bias}} \right).$$

for any input x. We assume that all nonlinear layers $\{\mathcal{L}_t\}_{t=1}^T$ have the same non-polynomial and globally Lipschitz activation function $\sigma \in C^\infty(\mathbb{R};\mathbb{R})$. In our specific experiments, we have $\mathcal{S} : \mathbb{R}^3 \to \mathbb{R}^{d_1}$ and $\mathcal{Q} : \mathbb{R}^{d_T} \to \mathbb{R}^{2\cdot\text{data dimension}}$, where 3 represents the number of parameters (q, χ_1, χ_2) and data dimension refers to the output dimension, which can be calculated as

$$\text{int}\left(\frac{\text{duration} \cdot \text{sampling frequency}}{2}\right) + 1.$$

To generate our parameters q, χ_1, and χ_2 for training our FNM, we utilize Bilby (only for generating the data, not the model). The mass ratio is sampled from a uniform distribution with minimum 0.5 and maximum 1, while χ_1 and χ_2 are taken from a uniform prior distribution with minimum 0 and maximum 0.99 for the aligned component of the spins[1]. Our training set consists of 1000 triples (q, χ_1, χ_2); in the future, we hope to expand our analysis to include a larger training dataset.

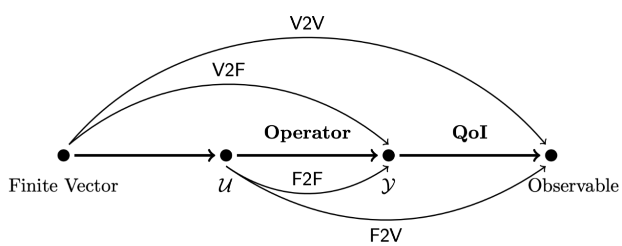

Figure 9. Fourier Neural Mappings (FNMs) generalize Fourier Neural Operators (FNOs) by allowing the input and output spaces to be finite-dimensional. In our experiment setting, we use the V2V mapping as our input space (parameters) and output space (spectrograms) are finite-dimensional. (Huang et al., 2024)

The forward pass of the implementation is shown in Figure 10. The way we train our surrogate FNM model is then by comparing $y_{\text{predicted}} = \Psi^{(\text{FNM})}(x)$ with the theoretical waveforms y_{true} generated using IMRPhenomXHM. Note that since the waveforms initially consist of complex values, we transform our outputs into a multi-channel image as follows:

$$\left[\overset{\text{Real}}{\begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{bmatrix}} \quad \overset{\text{Imaginary}}{\begin{bmatrix} b_{11} & b_{12} & \cdots & b_{1n} \\ b_{21} & b_{22} & \cdots & b_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ b_{m1} & b_{m2} & \cdots & b_{mn} \end{bmatrix}} \right].$$

[1] lscsoft.docs.ligo.org/bilby/api/bilby.gw.prior.AlignedSpin.html

Algorithm 3: Fourier Neural Mapping

Fourier Neural Mapping

```
Define forward(self, x);
  1. x = self.mlp0(x) # S;
  2. x = self.ldec0(x, self.s_latentspace) # D;
  3. for each (speconv, w) in zip(self.speconvs, self.ws);
       x = w(x) + speconv(x) # L_t;
       x = self.act(x);
  4. x_temp = self.lfunc0(x) # G;
  5. x = torch.cat((x_temp, x), dim=1);
  6. x = self.mlp1(x) # Q;
  7. return x;
```

Figure 10. Pseudocode for the FNM implementation. Each corresponding layer/mapping is color-coded in green. For the $\mathcal{Q}$ and $\mathcal{S}$ mappings, we have a pointwise single hidden layer of a fully-connected neural network. We have a linear decoder layer for $\mathcal{D}$ and a linear functional layer for $\mathcal{G}$. Note that self.speconvs refers to the Fourier integral operator layer. Our code is based off of https://github.com/nickhnelsen/fourier-neural-mappings (Huang et al., 2024) and modified to adapt to our data representation.

Lastly, since loss functions in the built-in torch.nn module do not support complex values, we developed our own custom relative L_2 loss metric as follows:

$$\text{relative L2 loss} = \frac{\sqrt{\sum_{i=1}^n |y_{\text{true},i} - y_{\text{pred},i}|^2}}{\sqrt{\sum_{i=1}^n |y_{\text{true},i}|^2}}$$

We chose the relative L_2 loss metric because it scales the error relative to the magnitude of the true waveform, which ensures that the loss does not disproportionately favor signals with much larger magnitudes. Additionally, since Fourier neural mappings are proficient in handling oscillatory signals, the relative L_2 loss can penalize deviations across all parts of the waveform while considering periodicity and structure.

3.5. Bayesian Inverse Problem (BIP)

As gravitational waves consist of noisy observations, we aim to infer the properties of the sources that produced such observations. This objective involves solving a Bayesian inverse problem (BIP), where the ultimate goal is to infer the source parameters. In other words, the objective is to estimate the unknown inputs of a system from its known outputs. Solving a BIP requires a forward model (such as the Fourier Neural Mapping) that maps the parameters of interest to the corresponding observable quantities.

A critical assumption in Bayesian inverse problems is that our observed data d, which in our case is assumed to be the waveforms, can be expressed as the sum of a forward model $f(\theta)$ evaluated at the model parameters θ and a Gaussian noise $\eta \sim \mathcal{N}(0, \Gamma_n)$, where Γ denotes the noise covariance matrix: $d = f(\theta) + \eta$. For simplicity, we assume that $\Gamma = \sigma^2 I$, where I is the identity matrix and σ^2 is a scalar representing the noise variance, although other works have considered Γ as a nontrivial covariance operator with different noise variances at different data points.

In contrast to a deterministic inverse problem (i.e. where we want to find $\theta^* = \text{argmin}_\theta ||d - f(\theta)||^2$), the ultimate goal of Bayesian inverse problems is to characterize the posterior distribution, which by Bayes' theorem can be formulated as:

$$\pi_{\text{post}}(\theta|d) = \frac{\mathcal{L}(d|\theta) \cdot \pi_{\text{pr}}(\theta)}{\pi(d)} = \frac{\mathcal{L}(d|\theta)\, \pi_{\text{pr}}(\theta)}{\int \mathcal{L}(d|\theta')\, \pi_{\text{pr}}(\theta') d\theta'},$$

where $\pi(d)$ is the evidence (i.e. marginal likelihood) and $\pi_{\text{pr}}(\theta)$ refers to the prior distribution reflecting any preliminary information regarding the model parameters (does not depend on d). We also have a likelihood function $\mathcal{L}(d|\theta)$, defined as follows:

$$\mathcal{L}(d|\theta) = \frac{1}{(2\pi)^{n/2}|\Gamma|^{1/2}} \exp\left(-\frac{1}{2}(d - f(\theta))^T \Gamma^{-1}(d - f(\theta))\right.$$

As our surrogate models act as efficient approximations of expensive-to-compute simulations, we can pair the observed waveform data with the model predictions to evaluate the likelihood function $\mathcal{L}(d|\theta)$ in the Bayesian inverse problem. Then, with the posterior distribution $\pi_{\text{post}}(\theta|d)$, we can perform standard sampling techniques such as Markov Chain Monte Carlo (MCMC), Nested Sampling, or Variational Transport to sample an independent and identically distributed (i.i.d.) set $\{\theta_i\}_{i=1}^N$, which can help us understand the source parameters better (Saleh et al., 2024). With multiple different surrogate models, we can then sample different sets from various posterior distributions and measure the statistical distance $\mathcal{D}$ of these distributions through metrics like Jensen-Shannon, Wasserstein, and KL-divergence:

$$\mathcal{D}(\{\theta_i\}^p, \{\theta_i\}^q) \to \mathbb{R} \qquad \forall\, 1 \leq p, q \leq n.$$

The primary computational challenge in Bayesian inference lies in the process of repeatedly evaluating the potentially costly model $f(\theta)$ many times during the sample generation process.

4. Results

Using our custom L_2 loss metric, we can see that our FNM surrogate model was able to train quite well as shown in Figure 12. The optimal set of hyperparameters we found for this model with the Adam optimizer was a learning rate of 0.001 and a batch size of 100. We set the duration and sampling frequency parameters to be 8 and 2048, respectively, which corresponds to a data dimension of $\frac{8\times2048}{2} + 1 = 8193$.

On the other hand, our autoencoder was able to train successfully, but there was minimal improvement after around 500 epochs (Figure 13), regardless of our choice of hyperparameters. In fact, by visually comparing the reconstructed output generated by the autoencoder with the original spectrograms (Figure 11), we can see that the the original spectrograms appear to have more detailed structures and patterns, whereas the reconstructions contain more blurred and noisy features.

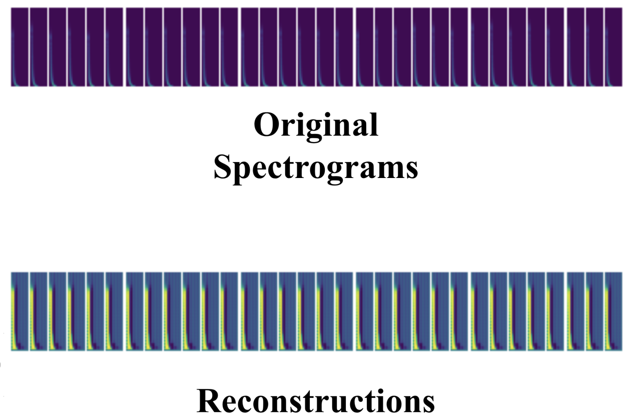

Figure 11. The autoencoder struggles to capture fine-grained features of the input, which is especially important for complex signals like gravitational waveforms. It appears that while the general shape of the spectrogram is retained, some features are lost in the reconstruction process.

Because our autoencoder pipeline involves the spectrogram representation, and the pixel values in the spectrograms are the power of each signal at each time, we concluded that the phase evolution is much more informative than the amplitude so a lot of relevant information must have been thrown away by the dimensionality reduction. Therefore, it might actually be beneficial to have multiple images per waveform to help capture all of the necessary information, and we can still use the autoencoder output to represent them.

5. Conclusion

In conclusion, we have found that the Fourier Neural Mapping works well as a surrogate model for Bayesian inference. Because the relative L_2 loss continues to decrease quite substantially after 1000 epochs, we are convinced that our FNM model predictions would pair well with the theoretically observed waveform data, which we can then use to evaluate the likelihood function in a BIP. However, we also discovered that autoencoders and convolutional neural networks in general can really struggle with the waveform data. Therefore, more robust dimensionality reduction techniques that can retain some information on the phase amplitude of waveforms will need to be explored.

Figure 12. The relative L_2 errors upon training our FNM. Although there appears to be some fluctuation due to the intrinsic noisiness of gravitational waves, the error is steadily decreasing. After 1000 epochs, the loss goes to around $10^{-1.7} \approx 0.02$.

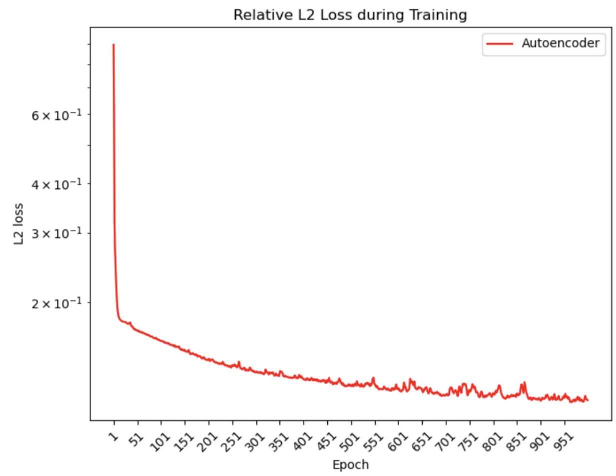

Figure 13. The relative L_2 loss during the training of our autoencoder. The loss decreases significantly within the initial epochs, indicating that the model quickly learns to capture key features of the data. However, after approximately 1000 epochs, the loss stabilizes and plateaus at a value of around 0.09. This behavior suggests that the autoencoder has reached its optimal learning capacity and further training would likely provide minimal improvement.

5.1. Limitations

Currently, our approach is constrained by the available computing resources, limiting the number of epochs we can train for and the dimension of the data we can handle. Additionally, we are focusing primarily on a small subset of parameters—namely, q, χ_1, and χ_2—which restricts the scope of our analysis. While these parameters are critical for gravitational wave inference, there are other important factors that have not been included in the current model. This limitation in the parameter space may hinder the model's overall ability to fully capture the complexity of real-world gravitational wave data.

5.2. Future Work

With increased computing resources, the next steps will involve training the model over a larger number of epochs, allowing for a more refined learning process, and potentially expanding the data dimension to improve model performance. As we found that the autoencoder can struggle with reconstructing the spectrograms, we are looking into the continuous short-time Fourier transform (which is what is used to build the spectrograms in the first place), so that we can keep the phase information and avoid taking absolute values (power of the signal which is what is stored in the spectrograms). We are also considering alternative formats besides spectrograms to represent and retain all the necessary information on our waveform data. We plan to explore alternative surrogate models that utilize more advanced methods for learning mappings between different vector spaces. Additionally, we aim to investigate whether other loss functions capable of handling complex values could improve the model's effectiveness. With more time, we also intend to actually implement a Bayesian inverse problem framework to validate the surrogate model's findings. Lastly, future work should extend our analysis to include all relevant parameters (right ascension, declination, luminosity distance, etc.) for gravitational wave inference, which, for most models, involves around 15 parameters in real-world data, providing a more comprehensive and accurate representation.

5.3. Applications

Data-based models of gravitational waves leverage machine learning and statistical techniques to analyze the complex signals detected by gravitational wave observatories such as LIGO. These models are particularly effective in extracting and learning the physical source parameters from noisy, real-world data, whereas traditional analytical models such as Bilby may struggle due to the complexity of the signals. By training on datasets of simulated gravitational waveforms, data-based surrogate models can accurately identify important features without the need of computationally expensive resources. These models can also help improve the sensitivity and efficiency of gravitational wave searches, enabling the detection of faint and rare events that might otherwise be unnoticed (Okounkova et al., 2023). Furthermore, data-based approaches like the FNM we explored can help refine waveform templates used for parameter estimation, facilitating the real-time analysis of gravitational wave detections.

References

Abbott, B. P., Abbott, R., Abbott, T. D., Abernathy, M. R., et al. Observation of gravitational waves from a binary black hole merger. *Physical Review Letters*, 116(6): 061102, 2016. doi: 10.1103/PhysRevLett.116.061102.

Bank, D., Koenigstein, N., and Giryes, R. Autoencoders, 2021. URL https://arxiv.org/abs/2003.05991.

Barausse, E. and Coauthors, N. The physics of gravitational waves. In *Proceedings of Winter School of Theoretical Physics, Second and Third Training School of COST Action CA18108 "Quantum gravity phenomenology in the multi-messenger approach" — PoS(QG-MMSchools)*, QG-MMSchools, pp. 002. Sissa Medialab, March 2024. doi: 10.22323/1.440.0002. URL http://dx.doi.org/10.22323/1.440.0002.

Dax, M., Green, S. R., Gair, J., Macke, J. H., Buonanno, A., and Schölkopf, B. Real-time gravitational wave science with neural posterior estimation. *Physical Review Letters*, 127(24), December 2021. ISSN 1079-7114. doi: 10.1103/physrevlett.127.241103. URL http://dx.doi.org/10.1103/PhysRevLett.127.241103.

Dey, R., Micchi, L. F. L., Mukherjee, S., and Afshordi, N. Spectrogram correlated stacking: A novel time-frequency domain analysis of the stochastic gravitational wave background. *Physical Review D*, 109(2), January 2024. ISSN 2470-0029. doi: 10.1103/physrevd.109.023029. URL http://dx.doi.org/10.1103/PhysRevD.109.023029.

Escrig, G., Campos, R., Qi, H., and Martin-Delgado, M. A. Quantum bayesian inference with renormalization for gravitational waves, 2024. URL https://arxiv.org/abs/2403.00846.

García-Quirós, C., Colleoni, M., Husa, S., Estellés, H., Pratten, G., Ramos-Buades, A., Mateu-Lucena, M., and Jaume, R. Multimode frequency-domain model for the gravitational wave signal from nonprecessing black-hole binaries. *Physical Review D*, 102(6), September 2020. ISSN 2470-0029. doi: 10.1103/physrevd.102.064002. URL http://dx.doi.org/10.1103/PhysRevD.102.064002.

George, D. and Huerta, E. Deep learning for real-time gravitational wave detection and parameter estimation: Results with advanced ligo data. *Physics Letters B*, 778: 64–70, March 2018. ISSN 0370-2693. doi: 10.1016/j.physletb.2017.12.053. URL http://dx.doi.org/10.1016/j.physletb.2017.12.053.

Getty Images. Gravitational waves, 2024. URL https://www.gettyimages.com/photos/gravitational-waves. Accessed: 2024-12-16.

Higashino, Y. and Tsujikawa, S. Inspiral gravitational waveforms from compact binary systems in horndeski gravity. *Physical Review D*, 107(4), February 2023. ISSN 2470-0029. doi: 10.1103/physrevd.107.044003. URL http://dx.doi.org/10.1103/PhysRevD.107.044003.

Huang, D. Z., Nelsen, N. H., and Trautner, M. An operator learning perspective on parameter-to-observable maps. *arXiv preprint*, 2024. doi: 10.48550/arXiv.2402.06031.

Jacobs, R. A., Jordan, M. I., Nowlan, S. J., and Hinton, G. E. Adaptive mixtures of local experts. *Neural Computation*, 3:79–87, 1991.

Keith, B., Khadse, A., and Field, S. E. Learning orbital dynamics of binary black hole systems from gravitational wave measurements. *Physical Review Research*, 3(4), November 2021. ISSN 2643-1564. doi: 10.1103/physrevresearch.3.043101. URL http://dx.doi.org/10.1103/PhysRevResearch.3.043101.

Mapelli, M. Binary black hole mergers: Formation and populations. *Frontiers in Astronomy and Space Sciences*, 7, July 2020. ISSN 2296-987X. doi: 10.3389/fspas.2020.00038. URL http://dx.doi.org/10.3389/fspas.2020.00038.

Michelucci, U. An introduction to autoencoders, 2022. URL https://arxiv.org/abs/2201.03898.

Nagar, A. et al. Time-domain effective-one-body gravitational waveforms for coalescing compact binaries with nonprecessing spins, tides, and self-spin effects. *Physical Review D*, 98(10):104052, 2018. doi: 10.1103/PhysRevD.98.104052.

Okounkova, M., Isi, M., Chatziioannou, K., and Farr, W. M. Gravitational wave inference on a numerical-relativity simulation of a black hole merger beyond general relativity. *Physical Review D*, 107(2), January 2023. ISSN 2470-0029. doi: 10.1103/physrevd.107.024046. URL http://dx.doi.org/10.1103/PhysRevD.107.024046.

O'Leary-Roseberry, T., Du, X., Chaudhuri, A., Martins, J., Willcox, K., and Ghattas, O. Adaptive projected residual networks for learning parametric maps from sparse data. 12 2021. doi: 10.48550/arXiv.2112.07096.

Saleh, B., Field, S. E., Kaye, J. P., Talbot, C., and Nissanke, S. Tempered multifidelity importance sampling for gravitational wave parameter estimation. *Physical Review D*, 110(10):104037, 2024. doi: 10.1103/PhysRevD.110.104037.

Schmidt, S., Breschi, M., Gamba, R., Pagano, G., Rettegno, P., Riemenschneider, G., Bernuzzi, S., Nagar, A., and Del Pozzo, W. Machine learning gravitational waves from binary black hole mergers. *Physical Review D*, 103 (4), February 2021. ISSN 2470-0029. doi: 10.1103/physrevd.103.043020. URL http://dx.doi.org/10.1103/PhysRevD.103.043020.

Tagawa, H., Kimura, S. S., Haiman, Z., Perna, R., and Bartos, I. Observable signature of merging stellar-mass black holes in active galactic nuclei. *The Astrophysical Journal*, 950(1):13, June 2023. ISSN 1538-4357. doi: 10.3847/1538-4357/acc4bb. URL http://dx.doi.org/10.3847/1538-4357/acc4bb.

Tancik, M., Srinivasan, P. P., Mildenhall, B., Fridovich-Keil, S., Raghavan, N., Singhal, U., Ramamoorthi, R., Barron, J. T., and Ng, R. Fourier features let networks learn high frequency functions in low dimensional domains, 2020. URL https://arxiv.org/abs/2006.10739.

Varma, V., Field, S. E., Scheel, M. A., Blackman, J., Kidder, L. E., and Pfeiffer, H. P. Surrogate model of hybridized numerical relativity binary black hole waveforms. *Physical Review D*, 99(6), March 2019. ISSN 2470-0029. doi: 10.1103/physrevd.99.064045. URL http://dx.doi.org/10.1103/PhysRevD.99.064045.

Xiong, C.-Y., Sun, T.-Y., Zhang, J.-F., and Zhang, X. Robust inference of gravitational wave source parameters in the presence of noise transients using normalizing flows, 2024. URL https://arxiv.org/abs/2405.09475.