CARDIFF UNIVERSITY

MASTERS THESIS

Analytic Initial Conditions and Verifying Parameter Estimation for Black Hole Binaries

Author:
Talya Klinger

Supervisor:
Dr. Mark Hannam

This thesis is submitted for the degree of Master of Philosophy

in the

School of Physics and Astronomy
Gravity Exploration Institute

22/05/2023

CARDIFF UNIVERSITY

Summary

Gravity Exploration Institute
School of Physics and Astronomy

Master of Philosophy

Analytic Initial Conditions and Verifying Parameter Estimation for Black Hole Binaries

by Talya Klinger

The LIGO-Virgo Collaboration utilises nested sampling to infer the source properties of compact binaries, computing Bayesian evidences and posterior distributions. With poor sampling from the constrained prior, nested sampling algorithms may misbehave and fail to sample the posterior distribution faithfully. Fowlie et al. (2020) outlines a method of identifying pathologies such as plateaus in the parameter space, using likelihood insertion order statistics. We apply this method to nested sampling analyses of all events in the first and second gravitational wave transient catalogs. With a few exceptions that have negligible effect on the final posteriors, the data is consistent with uniform insertion order statistics and unbiased prior sampling. There is, however, weak evidence against uniformity at the catalog-level meta-test.

With improvements to the LIGO-Virgo-Kagra detector network and future gravitational-wave interferometers planned, the demand for accurate and efficient numerical relativity codes becomes more pressing. Black hole binaries (BBHs) with very asymmetric mass ratios create a significant speed bottleneck, requiring a small grid spacing to resolve. In the moving-puncture formalism of numerical relativity, black holes are represented by *trumpets*, with an asymptotically flat exterior surrounding an infinitely long cylinder. While there is an analytic expression for the stationary trumpet spacetime, its behavior in a realistic BBH simulation scenario–i.e., a Lorentz boosted trumpet in 1+log slicing–is less well understood. We derive and sketch the numerical solution of a two-dimensional boundary value problem for the boosted 1+log trumpet coordinates. Such a solution could be subtracted from binary data to capture the strong field features of the low-mass partner, allowing for lower resolution and faster evolution.

Acknowledgements

To my research supervisor, Prof. Mark Hannam, thank you for your wisdom, support, encouragement, and persistence throughout the past year. I am incredibly grateful for your confidence in me as a student and in the boosted trumpet cause, which I can only hope to earn. To Dr. Michalis Agathos, thank you for seeing the insertion order check through to the finish line. I appreciate the invaluable advice and feedback of Ulrich Sperhake, John Veitch, Will Handley, Michael Williams, Anthony Lasenby, Eugene Lim, and Roman Rafikov, my supervisors at Caltech Saul Teukolsky and Mark Scheel for supporting me as I finish this work, and my fellow students, Terri Pierce, Jordan Barker, Will Griffiths, and Sam Higginbotham for camaraderie and scientific community. I am grateful as well to my family and friends, whose unconditional support motivated me throughout this process. Lastly, I would like to thank the Marshall Aid Commemoration Commission for the scholarship that made this research possible.

Contents

Summary i

Acknowledgements ii

List of Figures iv

List of Tables vi

1 Introduction 1
1.1 Gravitational-Wave Astronomy and Parameter Estimation 1
1.2 Numerical Relativity and the Boosted Trumpet Solution 2

2 Nested Sampling and the Order Statistics Cross-Check 3
2.1 Bayesian Inference in Parameter Estimation 3
2.2 Summary of the Nested Sampling Algorithm 5
2.3 Insertion Order Cross-Check 6

3 Performing the Cross-Check 8
3.1 Gravitational Wave Data 8
3.1.1 First observing run (O1) 8
3.1.2 Second observing run (O2) 8
3.1.3 First half of third observing run (O3a) 9
3.2 Implementation 9
3.2.1 Data Parsing 9
3.2.2 Measuring Uniformity 10
3.2.3 Performing Insertion Order Crosscheck 10

4 Results 13

5 Trumpet Coordinates in Numerical Relativity 18
5.1 3+1 Numerical Relativity 18
5.2 Trumpet Coordinates 20
5.3 1+log slicing 20

6 Boosted Trumpet Coordinates 22
6.1 Initial Data 22
6.2 Boosted Trumpet Spacetime 24
6.3 Constructing the Boundary Value Problem 24
6.4 Next Steps 25

7 Conclusion 28

Bibliography 30

List of Figures

2.1 These illustrations depict likelihood insertion order plots from well-behaved (left) and pathological (right) nested sampling runs. In the first plot, the points, each representing a likelihood insertion index, are evenly distributed throughout the plane. In the second plot, some regions are densely populated with points or almost empty, signs of excessive sampling from a specific likelihood range. There is also a run of repeated indices, resulting from a plateau in prior space. 7

3.1 Top: Example of the insertion order distribution from a single chain from GW150914 analyzed with IMRPhenomPv2. The flat distribution of insertion indices is typical for our dataset, and indicates that no major errors in prior sampling occurred in this chain. Bottom: A less uniform-looking example of a single chain from GW170823 analyzed with SEOBNRv4. The histogram shows one very frequent value, potentially indicating a plateau. However, in this case, the fluctuations are most likely due to the small-sample statistics of this shorter chain. The deviation from uniformity can be quantified using the KS-statistic and the significance of such a deviation can be assessed by calculating the corresponding p-value, which takes into account the sample size. . 11

4.1 Cumulative plot of the event-level meta p-values associated with each event in the GWTC-1 and GWTC-2 dataset. Each point in this plot results from combining insertion orders of all chains from a single event and performing a KS test. If all insertion order data were completely uniform, the p-values would fall along a straight line from 0 to 1. The meta-p-value for this distribution is 0.00144. 14

4.2 A version of Figure 4.1–the cumulative plot of event-level meta p-values associated with each event in the GWTC-1 and GWTC-2 dataset–created without transforming the insertion index to a continuous variable. Compared to the “continuified" plot, there is a clear deficit at high p-values, with no event in the range $[0.8, 1.0]$ and a very high excess in the low p-value range $[0, 0.2]$. 16

4.3 Top: Typical example of a KS-statistic sequence obtained from rolling tests on data from GW190408, performed on chunks of $2 \times n_{\text{live}}$ points; data from each chain and log file is drawn in a different colour. Bottom: Histogram showing the distribution of the p-values of all rolling tests, conducted over all events and chains. Performing a catalog-level meta-test to compare this distribution to uniformity, we obtain a KS p-value of 0.871. 17

6.1 K computed from its definition minus K computed from the 1+log condition, i.e. $(\sqrt{-g}n^{\mu})_{,\mu} - \frac{\sqrt{-g}}{\alpha}(\partial_t - \beta^i\partial_i)\alpha$, for $v = 0.3$. The difference is particularly large near the origin; outside the pictured range, it approaches 0. 23
6.2 Coefficients for $h_{,z,z}$, $h_{,\rho,z}$, $h_{,\rho,\rho}$, and the constant term in Eqn. 6.10, evaluated with $v = 0.1$. Outside the pictured range, all of these coefficients approach 0 as $\rho, z \to \infty$. 26

List of Tables

4.1 Event-level meta-p values for each event in chronological order, together with the set of interferometric detectors (IFOs) that participated in the detection, the number of chain files per event and number of live points used in the analysis. 15

Chapter 1

Introduction

1.1 Gravitational-Wave Astronomy and Parameter Estimation

In 2015, the LIGO-VIRGO collaboration (LVC) observed the first gravitational wave signal from a black hole binary, ushering in a new era of gravitational wave astronomy [38, 1, 7]. This discovery–and about 80 more gravitational-wave detections that followed in the first three observing runs–enabled new measurements of the astrophysical population of compact objects [4, 25, 26], general relativity [46, 39], and cosmology [25]. For a comprehensive review of gravitational wave astronomy's applications to fundamental physics, see [16].

One of the most important techniques used by the LVC to find gravitational-wave signals in detector output is matched filtering, a statistical technique that identifies signals of known shape in noisy data by correlation with *templates*. For compact binaries, these waveform templates are constructed with post-Newtonian theory in early inspiral and full numerical relativity in late inspiral and merger, where the strong-field behavior is needed. Once a candidate signal has been identified with matched filtering, the next stage is parameter estimation.

In principle, the trigger and parameter estimation could be unified into one process, since the underlying physics models are the same. However, the most efficient computational methods for detection and parameter estimation are different, making a two-pass approach more effective. Parameter estimation uses Bayesian inference to produce confidence intervals for the parameters of the signal, including both extrinsic parameters, such as the distance from Earth to the source, and intrinsic parameters, such as the masses and spins of the two compact objects.

Since these parameters have so many scientific implications, ranging from the astrophysical population of compact objects to tests of general relativity, it is particularly important to make sure parameter estimation is not biased. To that end, we have implemented a statistical cross-check for one of the methods used in parameter estimation, *nested sampling*. Chapter 2 and Chapter 3 outline the nested sampling algorithm and the nature of the cross-check, and our results are presented in Chapter 4.

1.2 Numerical Relativity and the Boosted Trumpet Solution

Over the next several decades, many upgrades are planned to the LIGO-VIRGO-Kagra network, which has just expanded to include the Kagra interferometer. Entirely new gravitational-wave observatories are planned as well, including the Einstein Telescope and LISA, a space-based detector whose longer interferometer arms will allow for observation of lower frequency gravitational waves [41][9]. Interpreting these observations will require a wider range of templates and waveform models, so the influx of new experimental data intensifies the need for more efficient black hole binary simulations.

A significant bottleneck is the difficulty of modeling binaries with highly asymmetric mass ratios (written, by LVK convention, as the lower mass divided by the higher mass.) In numerical relativity, the size of the smallest grid cell determines the time step required to evolve the system. For low-mass-ratio binaries, a very dense grid is required to resolve the lighter partner's strong-field features. The LVK has already observed a potential event with a mass ratio of about 1/30–outside the range its waveform models were even calibrated for–and future detectors will raise this limit [24]. In particular, the LISA detector's low frequency range will allow it to observe events with total mass up to 10^7, allowing for extreme mass ratios.

With an exact solution for the spacetime around each black hole in the binary–or a very good guess–far lower resolution may be required, allowing for rapid evolution of low-mass-ratio inspirals. A starting point for this approach is the spacetime of the single black hole alone–an unexpectedly complicated problem after taking certain nuances of numerical relativity into account.

In numerical relativity codes using the moving-puncture method, each black hole is represented not by Schwarzschild coordinates, but rather by "trumpet coordinates," which represent a Schwarzschild black hole as an asymptotically flat spacetime funneling into a cylinder of finite radius and infinite length. Furthermore, the trumpet spacetime would have to be Lorentz boosted to capture the movement of the binary. Finally, the boosted trumpet would need to match the gauge conditions used for the evolution of black-hole binaries–in this case, 1+log slicing.

By allowing for motion and using the same gauge choice as numerical relativity codes, a solution for the boosted trumpet in 1+log slicing would provide insight into the moving puncture in its natural habitat. Currently, the trumpet spacetime is only well understood for stationary, non-spinning black holes, which fails to capture the intricacies of black hole binary inspirals. We provide one step toward a look under the hood of numerical relativity. Furthermore, this solution could be applied to fast and efficient evolution of asymmetric mass ratio binaries. After subtracting out an analytic solution for the lighter black hole, the remainder, encompassing weak-field features of the binary system, will likely be smooth enough to capture accurately at far lower resolution, ultimately leading to much faster evolutions.

Chapter 2

Nested Sampling and the Order Statistics Cross-Check

The research presented in this and the following two chapters has been published in Annalen der Physik, in collaboration with Dr. Michalis Agathos [35]. While the concept of this project was developed before October 2021, including a preliminary presentation of O1 and O2 results to the LIGO Parameter Estimation call, the O3 data and significant changes to the analysis were added over the past year during my studies at Cardiff University. All of the insertion order statistics, graphics, and meta-p-values presented here are new, and I was the primary data analyst on this project.

2.1 Bayesian Inference in Parameter Estimation

The LVC relies on Bayesian inference methods to fit waveform models to observed data, using algorithms designed for efficiently sampling high-dimensional parameter spaces. These parameter spaces encode all the astrophysical properties of the source: the mass and spin of each component, the distance to the source, its sky-location and orientation angles, time and phase of coalescence, orbital eccentricity, and any additional matter properties if a neutron star is present. One parameter estimation algorithm is nested sampling, a method for efficiently computing Bayesian evidences as well as posterior probability distributions, introduced by Skilling in 2006 [45] (for a review, see [14].) Here, we use a new method of statistically verifying nested sampling output, the *insertion order cross-check* developed by Fowlie et al. [29], to test for biased nested sampling in the LVC's gravitational-wave data analysis [35].

At time of writing, the LVC uses Nested Sampling alongside machine-learning based methods, Markov Chain Monte Carlo (MCMC) and occasionally RIFT [36] to obtain posterior distributions in parameter estimation. These algorithms serve necessary and complementary purposes. While the LVC's implementation of MCMC and the machine learning methods used for systems involving neutron stars converge faster for signals with long inspiral times, and RIFT allows for direct comparison to numerical relativity, they do not directly compute evidence. These methods deliver only the normalized posterior and require further calculations to estimate the evidence, introducing significant statistical errors. Nested Sampling computes the evidence directly, allowing for greater accuracy.

In the LSC Algorithm Library (LAL), nested sampling was originally implemented in the *LALInference* package [37, 47]. During the third observing run (O3),

the LVC (now LVK) Collaboration gradually shifted its main data analysis pipelines to a newer Bayesian inference library *bilby* [12], a Python-based modular code which combines LAL's libraries for data infrastructure and waveform modelling with third-party nested samplers. The insertion order cross-check defined later in this paper has already been implemented in many Bilby samplers, including *CPNest* [48], *nessai* [50], and *UltraNest* [23]. However, the first two observing runs, O1 and O2, and the first half of the third observing run, O3a, were analyzed using LALInference alone. In this thesis, we utilize the insertion order cross-check to perform a post-mortem analysis on all nested sampling output for GW events in O1, O2, and O3a, evaluating the validity of parameter estimation results for the LVC's event catalogs GWTC-1 and GWTC-2 [5].

For a GW event associated with the coalescence of a compact binary, the source properties are described by a parameter vector $\boldsymbol{\theta} \in \Theta$. Here, Θ denotes the corresponding parameter space, encompassing the astrophysical properties of the two sources and the time and location of the event. Given the observed data D, we seek to infer the parameters $\boldsymbol{\theta}$ of the source, assuming certain background information $\mathcal{I}$ about the nature of the source, the behaviour of the detectors and the validity of GR as the underlying theory–that is, to estimate the posterior distribution $P(\boldsymbol{\theta}|D,\mathcal{I})$.

We make use of Bayes' theorem to update our *prior* expectations, $P(\boldsymbol{\theta}|\mathcal{I})$:

$$\begin{aligned} P(D|\boldsymbol{\theta},\mathcal{I}) \times P(\boldsymbol{\theta}|\mathcal{I}) &= P(D|\mathcal{I}) \times P(\boldsymbol{\theta}|D,\mathcal{I}) \\ L(\boldsymbol{\theta}) \times \pi(\boldsymbol{\theta})\, d\boldsymbol{\theta} &= Z \times p(\boldsymbol{\theta})\, d\boldsymbol{\theta}. \end{aligned} \tag{2.1}$$

$L(\boldsymbol{\theta}) = P(D|\boldsymbol{\theta},\mathcal{I})$, known as the likelihood function, and $\pi(\boldsymbol{\theta}) = P(\boldsymbol{\theta}|\mathcal{I})$, give the desired quantities $Z = P(D|\mathcal{I})$, the evidence and $p = P(\boldsymbol{\theta}|D,\mathcal{I})$, the posterior.

Computing the likelihood function (the probability density for observing data D, given the model and the true values of the parameters) requires models for both the detector signal and noise.

In LIGO's case, the actual signal is the electrical current from a photodiode, which can be converted to strain–the deformation of spacetime–via calibration. Calibration error is determined by marginalizing over calibration. LALSimulation can generate a waveform model for the strain; these waveform models assume general relativity is the correct underlying theory.

The noise for each detector is assumed to be Gaussian and is characterized by a power spectral density (PSD) pre-estimated from a stretch of data around the time of the event [5].

For strain data d and source parameters $\boldsymbol{\theta}$, the likelihood is

$$\ln L(d|\boldsymbol{\theta}) = -\frac{1}{2}\sum_k \left(\frac{[d_k - \mu_k(\boldsymbol{\theta})]^2}{\sigma_k^2} + \ln\left(2\pi\sigma_k^2\right) \right) \tag{2.2}$$

where k is the frequency bin index, σ is the noise amplitude spectral density, and $\mu(\theta)$ is the waveform [13]. Information from all active detectors is combined into a *coherent network likelihood* by taking the product of individual detector likelihoods [47]. It is assumed that genuine astrophysical signals are coherent, that is, they are the

same between different detectors except for the expected difference in direction, arrival time, and noise. Since location-specific features such as seismic noise differ between detectors, demanding coherence makes it possible to exclude them.

The evidence Z–the probability of observing the measured data, given the model–is defined as

$$Z = \int L(\boldsymbol{\theta})\pi(\boldsymbol{\theta})d\boldsymbol{\theta}. \tag{2.3}$$

This is an important quantity in Bayesian data analysis, as the evidences produced by different models can be directly compared. Hence, the evidence can be used to rank competing hypotheses and quantify how well a given model is supported by the data.

$dX = \pi(\boldsymbol{\theta})d\boldsymbol{\theta}$ is known as the element of prior mass. If we have some way to estimate the prior mass contained by a likelihood contour $L(\boldsymbol{\theta}) > \lambda$,

$$X(\lambda) = \int_{L(\boldsymbol{\theta})>\lambda} \pi(\boldsymbol{\theta})d\boldsymbol{\theta} \tag{2.4}$$

, we can invert $X(\lambda)$ to compute the likelihood L on the surface enclosing prior mass X. Then, the evidence can be recast as a one-dimensional integral

$$Z = \int_0^1 L(X)dX. \tag{2.5}$$

This 1D form is more computationally tractable than integrating across a high-dimensional parameter space Θ.

Nested sampling is a method for computing evidence that takes advantage of this formulation, relying on the statistical properties of prior sampling to provide a fast and accurate estimate of the prior mass at each integration step.

2.2 Summary of the Nested Sampling Algorithm

Nested sampling relies on sampling from the *constrained prior*: points from the prior with likelihood higher than some minimum value. As points from the constrained prior are sampled and discarded throughout the algorithm, the samples used at each step are called *live points.*

The nested sampling algorithm proceeds as follows:

1. Choose the number of live points n_{live} and sample n_{live} initial points from the constrained prior. Also, set an evidence threshold ϵ.
2. Identify the live point with the lowest likelihood L_i^*. Discard the live point and record its likelihood.
3. Sample a new live point from $\pi(\boldsymbol{\theta})$ with $L > L_i^*$. At this stage, the prior volume compresses exponentially, giving prior volume $X_i \approx \exp(-1/n_{live})$ on the ith step (the proof is nontrivial, see [45]).
4. Integrate the evidence Z_i using L_i^* and X_i.

5. Repeat steps (2)-(4) until a stopping condition is reached: $L_{max}X_i/Z_i < e^{\epsilon}$, where L_{max} is the highest likelihood discovered so far, X_i is the prior volume inside the current iso-likelihood contour L_i^*, and Z_i is the current estimate of the evidence. For LALInference, $\epsilon = 0.1$; essentially, if all the live points were to have the maximum discovered likelihood, the evidence would only change by a factor of less than 0.1 [47].

Nested sampling requires faithful sampling from the constrained prior to produce accurate evidences and posteriors. In practice, sampling from the entire prior and accepting only points with high enough likelihood is impractically slow, because the volume of acceptable points decreases exponentially in time. So, most implementations of nested sampling sample from a restricted region of parameter space drawn around the live points. LALInference, in particular, generates samples by an MCMC chain from a randomly chosen previous livepoint, and choosing the length of the MCMC chain is a tradeoff between speed and accuracy [47].

If the restricted region is too small or the MCMC chains too short, the constrained prior may not fully cover the iso-likelihood contour, violating the fundamental assumptions of nested sampling. Plateaus–regions of constant $L(\boldsymbol{\theta})$–also violate the assumptions of nested sampling, causing live points to be non-uniformly distributed in X.

2.3 Insertion Order Cross-Check

The *insertion index* is the position where an element must be inserted in a sorted list to preserve order. (We index lists starting from 1 rather than 0.) More concretely, if x is a sorted list and there exists a sample y such that

$$x_{i-1} < y < x_i, \tag{2.6}$$

the insertion index of y in list x is i [29]. For example, in the list (1 3 5 7 9), $y = 4$ would have an insertion index of 3.

Fowlie et al. [29] noted that, if the assumptions of nested sampling are met, the insertion index of new live points into the list of likelihoods of current live points should follow a uniform distribution, *i.e.* new live points should have an arbitrary likelihood, only constrained to be higher than the lowest likelihood. The prior mass enclosed by a certain likelihood decreases monotonically as that likelihood increases, so sorting live points by likelihood is equivalent to sorting by prior mass.

Therefore, non-uniformity of likelihood insertion indices serves as an early warning for any irregularities in sampling the prior. For example, a likelihood plateau in the parameter space results in a stretch of repeated indices. Such a problem would be visible in plots of the likelihood insertion index or detectable with statistical tests of uniformity. Figure 2.1 illustrates what such non-uniformities might look like in a plot of insertion order.

For our purposes, the insertion order cross-check is the simplest and most flexible method for verifying nested sampling. Buchner's "shrinkage test" is limited to toy likelihood functions with certain analytic properties, and is designed for constrained

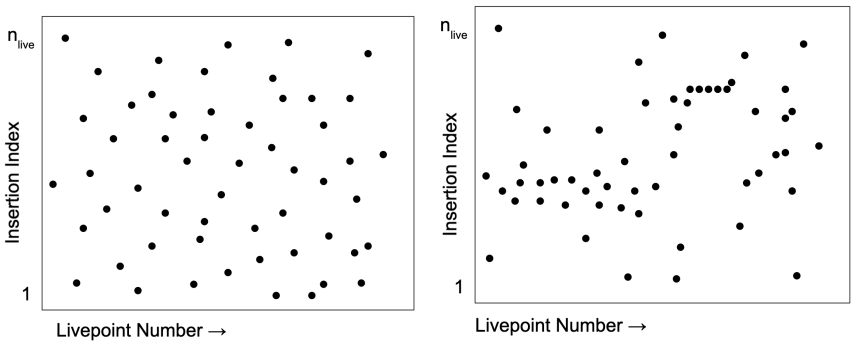


FIGURE 2.1: These illustrations depict likelihood insertion order plots from well-behaved (left) and pathological (right) nested sampling runs. In the first plot, the points, each representing a likelihood insertion index, are evenly distributed throughout the plane. In the second plot, some regions are densely populated with points or almost empty, signs of excessive sampling from a specific likelihood range. There is also a run of repeated indices, resulting from a plateau in prior space.

prior sampling using regions (such as ellipsoids) rather than LALInferenceNest's MCMC steps [22]. The diagnostic methods implemented in *Nestcheck* [32] are more applicable to LIGO parameter estimation, but they require multiple runs. One of the strengths of the insertion order crosscheck is that it can provide useful information on any scale, from an individual chain to an entire observing run.

Chapter 3

Performing the Cross-Check

3.1 Gravitational Wave Data

In this work, we analyse 45 events which were observed by two or more detectors, as listed in Table 4.1 (with the exception of GW190424A, a single-detector event). During the first three observing runs, two LIGO detectors (LIGO Hanford, LIGO Livingston) and the Virgo detector participated in the network. During O1, only the two LIGO detectors were operating, with Virgo beginning operation in August of 2017 during O2.

3.1.1 First observing run (O1)

In O1, the first observing run extending from 12 September 2015 to 19 January 2016, three GW events were detected with high confidence. All three were identified as signals from the coalescence of binary black holes (BBH): GW150914 (the first gravitational-wave detection), GW151012 (which was initially detected with low confidence and later passed the candidate threshold), and GW151226. For the catalog paper [2], all O1 events were analyzed using two waveform models: IMRPhenomPv2 [33], a phenomenological model calibrated to numerical relativity for gravitational waves from precessing BBH binaries, and SEOBNRv4, a model based on the Effective One-Body formalism [44]. We have analyzed the nested sampling results of all three events in O1.

3.1.2 Second observing run (O2)

O2 includes seven more BBH mergers and one binary neutron star (BNS) merger, all observed between 30 November 2016 and 25 August 2017. Some significant milestones from O2 are the first BNS event, GW170817, and the first GW signal to be detected by both LIGO interferometers and the Virgo interferometer, GW170814. The O2 events were also analyzed using IMRPhenomPv2 and SEOBNRv4, with a few exceptions. GW170729 was also analyzed with IMRPhenomD (a model similar to IMRPhenomPv2 for spinning but non-precessing binaries), and GW170809 was analyzed using IMRPhenomD and multiple runs of IMRPhenomPv2 with different priors.

The events from O1 and O2 together formed the first GW transient catalog GWTC-1, the results of which are detailed in Ref. [3]. The data cross-check includes all available output files for the primary nested sampling runs of the 10 binary black hole mergers. We omit the binary neutron star merger since it was not analyzed with nested sampling.

3.1.3 First half of third observing run (O3a)

The next upgrade to the instrumentation of the LIGO and Virgo detectors further improved their sensitivity, increasing the distance reach and event detection rate of the three-detector network. In O3a, the first half of the third observing run between 1 April 2019 and 1 October 2019, 39 new GW events were detected with high confidence. Of these, we consider the 35 that were analyzed with LALInferenceNest. Among the most interesting events in O3a are a couple of highly asymmetric (in mass) compact binaries, the first black-hole - neutron-star binary candidates and the most massive black-hole binary observed to date, reaching a total mass of $\sim 150 M_{\odot}$. Several different waveform models were employed in the analyses of these events, particularly for those with highly asymmetric mass ratios or weak signs of spin-induced precession. Both phenomenological (IMRPhenom) and effective-one-body (SEOBNR) models were used in all cases. Here, we focus on the nested sampling analyses with IMRPhenomPv2 as the underlying waveform model, since these are found in almost all O3a events. Further details about the data, the detection statistics and the properties of all O3a events can be found in Ref. [5]. The events from O3a, together with the ones from O1 and O2, form the second gravitational wave transient catalog, GWTC-2 [1].

3.2 Implementation

3.2.1 Data Parsing

The likelihood insertion indices defined in Sec. 2.3 can be computed either from the nested sampling iteration i and likelihood, or *birth* and *death contours* (initial and final likelihood), of each point. LALInference stores both the contours and the likelihoods themselves, but this information is distributed across 2 different types of output files. Log files contain the nested sampling iteration i, birth and death contours, and likelihood of each point. Since the likelihood is stored to higher precision than the birth and death contours, we use the likelihood and iteration to compute insertion indices. However, the initial pool of N_{live} (typically 1024 or 2048) points is missing from the log files, so it is impossible to compute the insertion indices of early live points from the logs alone.

On the other hand, the main data product of LALInferenceNest, the chain files, contain the initial live points but not the nested sampling iteration, making it impossible to compute insertion order from these files alone. Whenever possible, we match each log file with the associated chain file to find the initial live points, then replay the algorithmic process to compute the insertion indices exactly. In the absence of a match, we attempt to minimize the effect of the missing initial N_{live} live points by removing the first $5 * N_{\text{live}}$ points from our reconstructed chain (a cutoff which removed missing-point effects from insertion order plots effectively), then compute the insertion indices for the remaining points.

Log files have less strict I/O specifications than chain files and occasionally have some duplicate blocks of points, resulting from resubmitted jobs. When this is the case, we identify and remove the older duplicate live points.

[1] An extended and improved version of this catalog, known as GWTC-2.1, was more recently presented in Ref. [6]

3.2.2 Measuring Uniformity

To interpret the computed insertion order statistics, we must measure how uniformly they are distributed between 1 and N_{live}. Several statistical tests exist to determine whether two (or more) samples are drawn from the same underlying probability distribution, or compare a sample to a reference probability distribution. In particular, we use the Kolmogorov-Smirnov (KS) test [42], as implemented in *scipy* [49].

The KS test measures the distance between two cumulative distribution functions (CDFs). More precisely, for empirical CDF $F_{\text{data}}(x)$ and CDF of the uniform distribution F_U, the *Kolmogorov-Smirnov statistic* is defined as

$$D_n = \sup_x |F_{\text{data}}(x) - F_U(x)| \tag{3.1}$$

The KS test produces a test statistic between 0 and 1, with higher values corresponding to more distinct distributions. The KS statistic is independent of the number of samples, so the number of samples must be taken into account separately when interpreting the test results.

In the context of hypothesis testing, the p-value of a measurement or test statistic x is the probability of obtaining the observed value, or a more extreme value (for uniformity testing, a test statistic associated with a larger difference between the two distributions), assuming that the null hypothesis is correct. The KS test can be converted to p-values through the *Kolmogorov-Smirnov distribution*, which associates probabilities with test results for the hypothesis that two samples are drawn from different distributions. In cases where nested sampling has proceeded correctly, the insertion order distribution is uniform and the null hypothesis is true, leading to a uniform distribution of p-values over different runs. In isolation, small p-values do not necessarily mean that an entire nested sampling run is compromised, but if small p-values predominate, that could be a sign of systematic problems.

A seemingly insignificant subtlety that turns out to be important is that the standard KS test implementation is designed for comparing distributions of continuous variables. However, in this case our insertion order data is discrete [23]. The impact of this effect is discussed in Ch. 4. There are two ways of resolving the problem: we can either transform the data into an equivalent continuous distribution ranging in $[1, N_{\text{live}} + 1)$ by adding a random number in $[0, 1)$ to each insertion index (which will respect uniformity if the underlying discrete distribution is uniform); or we can implement a discrete version of the KS test or variations thereof, as described in [11, 28]. For simplicity, we have chosen to do the former.

3.2.3 Performing Insertion Order Crosscheck

LALinference is heavily parallelized, with each event's nested sampling analysis split into several parallel chains. The top and bottom panels of Figure 3.1 showcase two examples of the insertion order distribution of single chains, one very uniform, one less so.

For each event, we compute insertion orders and KS statistics over each individual chain. To assess the overall quality of nested sampling in each event, we perform

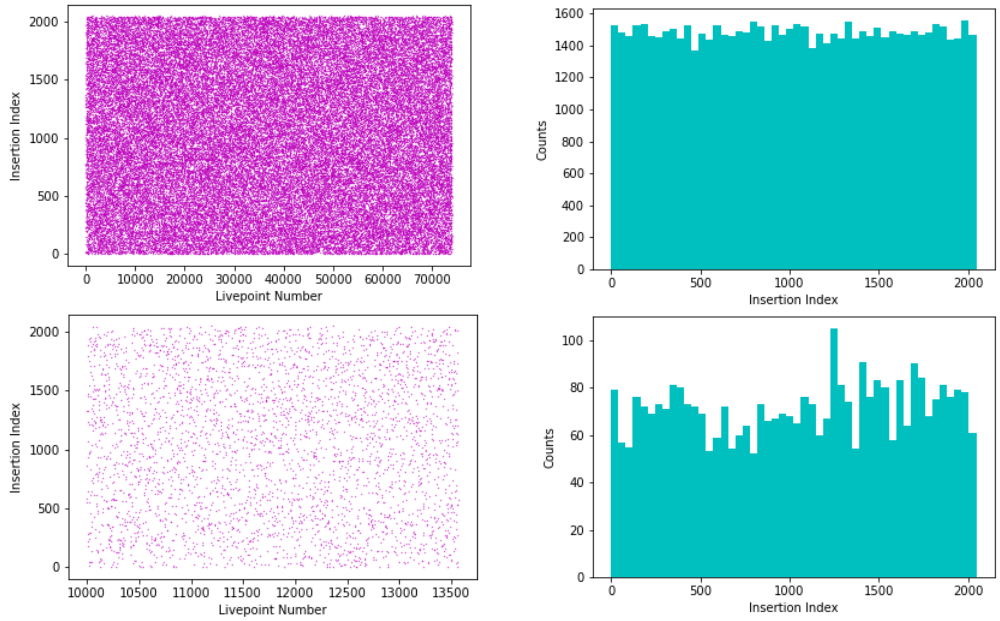


FIGURE 3.1: Top: Example of the insertion order distribution from a single chain from GW150914 analyzed with IMRPhenomPv2. The flat distribution of insertion indices is typical for our dataset, and indicates that no major errors in prior sampling occurred in this chain. Bottom: A less uniform-looking example of a single chain from GW170823 analyzed with SEOBNRv4. The histogram shows one very frequent value, potentially indicating a plateau. However, in this case, the fluctuations are most likely due to the small-sample statistics of this shorter chain. The deviation from uniformity can be quantified using the KS-statistic and the significance of such a deviation can be assessed by calculating the corresponding p-value, which takes into account the sample size.

a KS test on the combined insertion orders from all parallel chains which we will refer to as the “event-level meta test."

In each chain, we also perform “rolling tests:" a series of tests on each sequential stretch of $2 * N_{\text{live}}$ points. An example rolling test is shown in Figure 4.3. These rolling test values can be used to examine and compare the severity of local anomalies. In particular, we report the minimum p-value from all rolling tests of each event. To assess the uniformity of the entire GWTC-1 and GWTC-2 dataset, we perform a final KS test on the KS p-values from each individual rolling test, known as the “catalog-level meta test."

Chapter 4

Results

For each event in Table 4.1, we reconstruct the insertion index data and apply the insertion order cross-check. First, we perform rolling KS tests for each chain and verify that no individual chain shows significant bias in its insertion order statistics. For each of the events, we then pool together the insertion indices from all parallel chains and perform a single event-level KS test, by calculating the KS-statistic for that event and the corresponding p-value. The p-values for all the events analysed are given in Table 4.1. We find a seemingly healthy distribution that spans the entire range between 0 and 1, whose uniformity we will examine with a meta-p-value test for the entire catalog of events.

Before describing the final catalog-level results, we first discuss the impact of two subtle effects introduced in Sec. 3.2: the effect of the discrete nature of the insertion index and its "continuification," and the effect of missing data and the recovery process.

Although the insertion index is a discrete variable, it is possible that its wide range (N_{live} is typically 2048) would effectively make it continuous in practice. However, due to the very large number of data points, the underlying discreteness is significant from the perspective of a KS test. The KS statistic tests uniformity by measuring the maximum deviation of the empirical CDF from the diagonal. The empirical CDF of any discrete variable has a staircase-like structure, leading to some deviation from uniformity, and the ability of the KS test to discern this depends on the range of the discrete variable ($N_{\text{live}} \sim O(10^3)$) and the sample size. For large sample sizes (here $O(10^5)$), our results, illustrated in Figure 4.2, demonstrate that the discreteness has a significant impact. When treating the insertion index as a continuous variable, without transforming it to one, we found that there was a clear deficit at high p-values, with no event in the range $[0.8, 1.0]$ and a very high excess in the low p-value range $[0, 0.2]$. Accounting for the discrete nature of the variable significantly reduces this bias.

Next, we examine whether our attempt to correct for missing data, either in the initial N_{live} points or in mid-run, has significantly impacted the overall insertion order statistics. On a per-chain basis, the removal of $5 * N_{\text{live}}$ samples on either side of the gap seems to suppress the bias completely (even for a small number of missing chain points, the chain-level p-value is typically corrected by several orders of magnitude, back to $O(0.1)$). However, the catalog-wide statistics point in a different direction: in the population of all chains in the catalog, a residual bias remains. The size of this effect can be estimated by completely removing all chains that had missing data from the catalog-level analysis. Although these make up a small fraction of the total number of chains, we find that their removal leads to a systematic

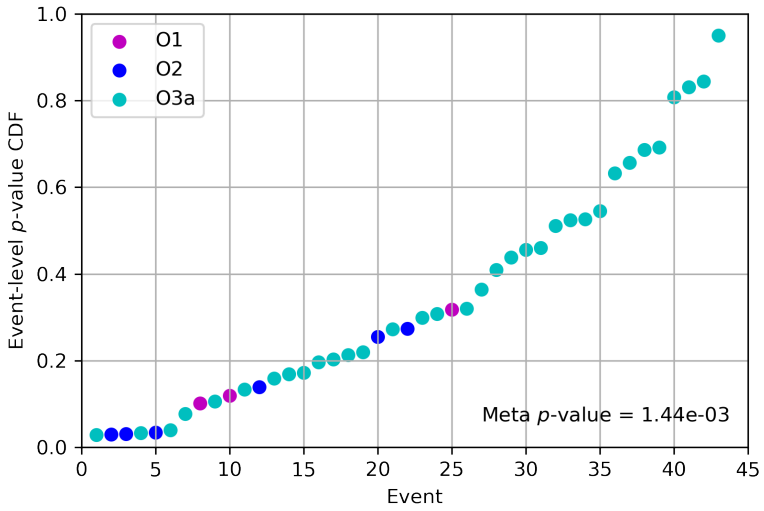


FIGURE 4.1: Cumulative plot of the event-level meta p-values associated with each event in the GWTC-1 and GWTC-2 dataset. Each point in this plot results from combining insertion orders of all chains from a single event and performing a KS test. If all insertion order data were completely uniform, the p-values would fall along a straight line from 0 to 1. The meta-p-value for this distribution is 0.00144.

improvement of the meta-p-value by a factor of ~ 4. A possible compromise would be to increase the number of points we truncate around missing data, however, in practice, this would render the heavily truncated chains virtually uninformative.

After removing all chains with missing data and transforming the insertion index to a continuous variable, we can now perform the final meta-test at the catalog level. Figure 4.1 shows the results of the event-level p-value test, each conducted over all insertion index samples for each event, arranged in increasing order of KS p-value. The event names and their test results are listed in chronological order in Table 4.1. If the insertion order data were uniform, the p-values would also follow a uniform distribution, and the points in Figure 4.1 would fall along a straight line from 0 to 1. However, we still observe an excess of low p-values, particularly among events from O1 and O2. A final test comparing the meta p-values to the uniform distribution results in a KS p-value of 1.44×10^{-3}.

We also perform a meta-p-value test on the rolling test results. Figure 4.3 shows the distribution of all rolling test p-values. The catalog-level meta-p-value, including all rolling tests and adjusted for the number, is 0.871, indicating no significant divergence from the uniform distribution and an overall healthy performance across all three runs.

Event	**IFOs**	**#chains**	N_{live}	**Meta p-value**
GW150914	HL	12	2048	0.2953
GW151012	HL	12	1024	0.0973
GW151226	HL	12	2048	0.1207
GW170104	HL	12	1024	0.1047
GW170608	HL	12	1024	0.0333
GW170729	HLV	20	2048	0.2519
GW170809	HLV	32	2048	0.0212
GW170814	HLV	26	2048	0.1603
GW170818	HLV	16	1024	0.1419
GW170823	HL	18	2048	0.0317
GW190408A	HLV	4	2048	0.5709
GW190412A	HLV	4	2048	0.8341
GW190413A	HLV	4	2048	0.4058
GW190413A	HLV	4	2048	0.0387
GW190421A	HL	8	2048	0.2037
GW190424A	L	4	2048	0.3008
GW190503A	HLV	4	2048	0.1350
GW190512A	HLV	4	2048	0.2218
GW190513A	HLV	16	2048	0.4650
GW190514A	HL	4	2048	0.2093
GW190517A	HLV	4	2048	0.5249
GW190519A	HLV	4	2048	0.6515
GW190521A	HL	4	2048	0.3208
GW190521B	HLV	4	2048	0.6330
GW190527A	HL	4	2048	0.0330
GW190602A	HLV	4	2048	0.9461
GW190620A	LV	4	2048	0.0321
GW190630A	LV	4	2048	0.5345
GW190701A	HLV	4	2048	0.0849
GW190706A	HLV	7	2048	0.1111
GW190707A	HL	4	2048	0.3056
GW190708A	LV	4	2048	0.1964
GW190719A	HL	4	2048	0.1297
GW190720A	HLV	4	2048	0.1349
GW190727A	HLV	4	2048	0.3458
GW190728A	HLV	4	2048	0.7932
GW190731A	HL	4	2048	0.4246
GW190803A	HLV	4	2048	0.8563
GW190828A	HLV	5	2048	0.1731
GW190828B	HLV	4	2048	0.5201
GW190909A	HL	4	2048	0.1586
GW190910A	LV	4	2048	0.4375
GW190915A	HLV	4	2048	0.6632
GW190929A	HLV	4	2048	0.6984
GW190930A	HL	4	2048	0.1731

TABLE 4.1: Event-level meta-p values for each event in chronological order, together with the set of interferometric detectors (IFOs) that participated in the detection, the number of chain files per event and number of live points used in the analysis.

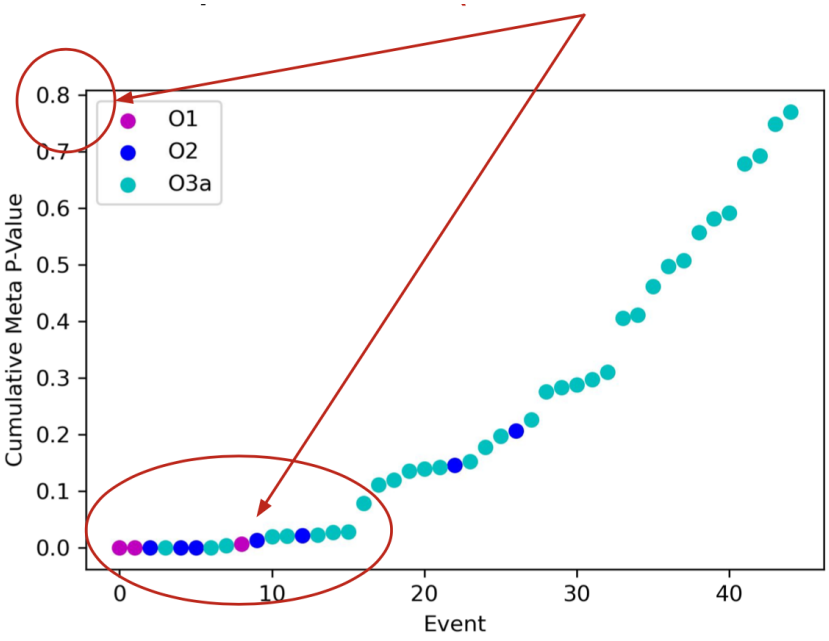


FIGURE 4.2: A version of Figure 4.1–the cumulative plot of event-level meta p-values associated with each event in the GWTC-1 and GWTC-2 dataset–created without transforming the insertion index to a continuous variable. Compared to the "continuified" plot, there is a clear deficit at high p-values, with no event in the range $[0.8, 1.0]$ and a very high excess in the low p-value range $[0, 0.2]$.

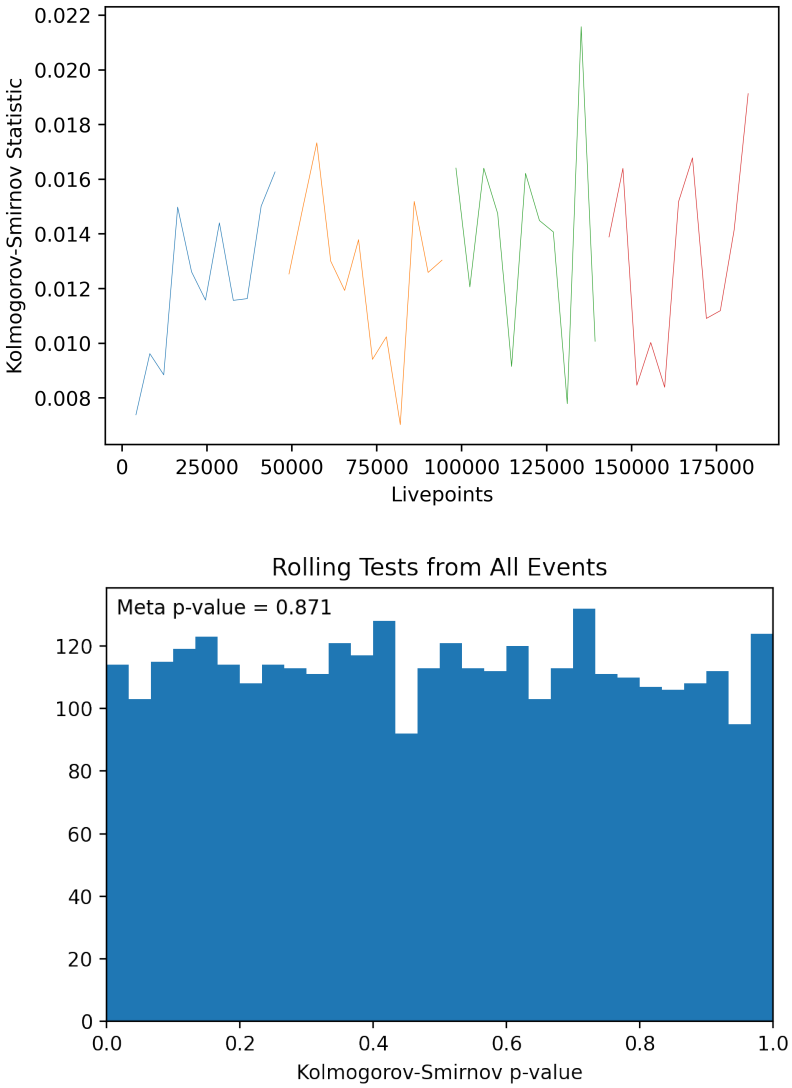


FIGURE 4.3: Top: Typical example of a KS-statistic sequence obtained from rolling tests on data from GW190408, performed on chunks of $2 \times n_{\text{live}}$ points; data from each chain and log file is drawn in a different colour. Bottom: Histogram showing the distribution of the p-values of all rolling tests, conducted over all events and chains. Performing a catalog-level meta-test to compare this distribution to uniformity, we obtain a KS p-value of 0.871.

Chapter 5

Trumpet Coordinates in Numerical Relativity

5.1 3+1 Numerical Relativity

The original, 4-dimensional formulation of Einstein's Equations is extremely difficult to solve for an arbitrary astrophysical system without much symmetry. Several different formalisms for numerical relativity have been developed to tackle this problem; the 3+1 formalism is a common approach used by codes including BAM [21], SPEC [34] (for initial conditions), GRChombo [10], and Einstein Toolkit [40].

3+1 numerical relativity relies on *foliation*: slicing the 4-dimensional spacetime into three-dimensional, spacelike *hypersurfaces*. Thus, the Einstein equations can be reformulated as an initial value problem with dynamical evolution between the 3d slices.

An overview of 3+1 numerical relativity can be found in Baumgarte and Shapiro [17], and our notation follows their conventions. First, the metric is decomposed into spatial hypersurfaces Σ_t. The simplest way to accomplish foliation is to choose spatial basis vectors tangent to the slices. With such basis vectors, any metric can be written in the form

$$ds^2 = -\alpha^2 dt^2 + \gamma_{ij}(dx^i + \beta^i dt)(dx^j + \beta^j dt). \tag{5.1}$$

On each time slice, the three-dimensional spacial metric is γ_{ij}. (For clarity, when γ_{ij} appears with indices it refers to the 3-metric on a single slice, and when γ appears without indices, it refers to the Lorentz factor.) The lapse function α is the proper time between time slices t and $t + dt$, and the shift vector β^i determines how the coordinates change between the two slices.

Together, the lapse, shift vector, and 3-metric suffice to describe the geometry of a single hypersurface. γ_{ij} and the *extrinsic curvature* constitute another complete description of one hypersurface.The extrinsic curvature K_{ij}, proportional to the Lie-derivative of γ_{ij} along the normal vector to the slices, is defined as

$$K_{ij} = -\frac{1}{2\alpha}\mathcal{L}_{\hat{n}}\gamma_{ij} = \frac{1}{\sqrt{-g}}(\sqrt{-g}n^{\mu})_{,\mu}. \tag{5.2}$$

Using the 3+1 formalism, we have

$$K_{ij} = \frac{1}{2\alpha}(\nabla_i\beta_j + \nabla_j\beta_i - \partial_t\gamma_{ij}) \tag{5.3}$$

where ∇_i is the covariant derivative with respect to the spatial metric γ_{ij}. (Throughout this text, trace(K) will be written as K.)

We also introduce a conformal factor ψ:

$$\gamma_{ij} = \psi^4 \tilde{\gamma}_{ij} \tag{5.4}$$

$$K = \tilde{K} \tag{5.5}$$

$$\beta^i = \tilde{\beta}^i \tag{5.6}$$

$$\tag{5.7}$$

Despite gauge freedom, a thoughtful choice of coordinate gauge–known as the *slicing condition*–is essential to the success of 3+1 numerical evolutions. The slicing must avoid singularities and be reasonably fast to compute.

The simplest slicing condition is *maximal slicing*, $tr(K) = 0$. However, to preserve maximal slicing during evolution, one must solve an elliptic equation at each step, which is too computationally expensive to be practical.

Bona et. al. [18] determine that the most general slicing condition invariant under any coordinate transformation on the slice is

$$Q - f(\alpha)\, tr(K) = 0 \tag{5.8}$$

where

$$Q = -\frac{1}{\alpha}(\partial_t - \beta^i \partial_i) \ln \alpha \tag{5.9}$$

Under *harmonic slicing*, which sets $f = 1$, Einstein's equations become a First Order, Flux Conservative, Hyperbolic (FOFCH) system, where spatial derivatives occur only in the flux term, not the source term. Many numerical and mathematical methods have been developed specifically for this form of partial differential equations in fluid dynamics, allowing for more flexible numerical relativity codes than the problem-specific methods used before. However, the harmonic slicing condition just barely avoids singularities, causing metric and curvature components to grow out of control near the horizon without exactly the right boundary conditions.

The *1 + log* slicing condition sets $f = 1/\alpha$. 1 + log is also consistent with the FOFCH formalism, but it behaves better near the horizon: as in maximal slicing, the lapse α approaches zero.

All of these slicing conditions produce physically equivalent results, but some are more suitable for numerical evolution than others. As all BBH codes in the moving-puncture formulation use 1+log slicing (or variations on 1+log slicing), we will ensure our boosted trumpet solution meets the following *1+log condition*:

$$(\partial_t - \beta^i \partial_i)\alpha = -n\alpha K \tag{5.10}$$

equivalent to

$$\mathcal{L}_{\hat{n}} \alpha = -nK \tag{5.11}$$

where $\hat{n}$ is the future-pointing normal vector.

5.2 Trumpet Coordinates

The initial data for black hole binaries in numerical relativity must use coordinates that do not reach the physical singularity. There are two methods for dealing with what happens at the event horizon $R = 2M$: punctures and excision. In the puncture approach, at the event horizon, the coordinates proceed through a wormhole to an extra copy of the asymptotically flat exterior spacetime [19]. These extra copies, each compactified to a single point in the grid, are referred to as the "punctures." The wormhole is described by a conformal factor ψ, which was held constant during evolution in early work with the "fixed-puncture" approach [8][20]. The "moving-puncture" approach, first demonstrated by Baker et al. and Campanelli et al., evolves ψ and allows the punctures to move across the grid [15][43]. Thus, the punctures are able to orbit each other and inspiral like a black hole binary.

However, as the punctures evolve dynamically, they are not fixed in this wormhole form. It turns out that the grid points, rather than remaining as 2 copies of an asymptotically flat spacetime, become infinitely long cylinders centered on each puncture [30]. [31] uses the gauge conditions to find an analytic expression for this spacetime–time-independent, asymptotically flat on one end, cylindrical on the other–known as "trumpet coordinates." Furthermore, it is possible to simply construct trumpet initial data, featuring only the region of spacetime relevant to numerical evolution, rather than using the entire wormhole.

The maximal slice of Schwarzschild spacetime

$$\gamma_{RR} = \left(1 - \frac{2M}{R} + \frac{C^2}{R^4}\right)^{-1} \tag{5.12}$$

$$K^i_j = \text{diag}\,(2C/r^3, -C/R^3, -C/R^3) \tag{5.13}$$

$$\beta^R = \frac{\alpha C}{R^2} \tag{5.14}$$

$$\alpha = \sqrt{1 - \frac{2M}{R} + \frac{C^2}{R^4}}, \tag{5.15}$$

$$\tag{5.16}$$

with $C = 3\sqrt{3}M^2/4$ and $R \in [1.5M, \infty)$, produces a trumpet [31][27]. Each constant coordinate time slice ends in a sphere of radius $R_0 = 3M/2$. Since the spatial metric diverges at R_0, the proper time between a point on this sphere and any point outside it is infinite. So, the data become an infinitely long cylinder with radius R_0. The lapse and β^R go to 0 as R approaches R_0.

5.3 1+log slicing

To use these trumpet initial conditions in numerical evolutions, they must be converted to 1+log slicing so the lapse follows an evolution equation. We can introduce a spherically symmetric height function $h(R)$ into Schwarzschild spacetime with the transformation,

$$t = T - h(R) \tag{5.17}$$

In these coordinates, the lapse, shift, and unit normal vector to the time slices are

$$\alpha = \sqrt{\frac{f}{1 - f^2 h'^2}} \tag{5.18}$$

$$\beta^R = \frac{f^2 h'}{f^2 h'^2 - 1} \tag{5.19}$$

$$n_\mu = (-\alpha, 0, 0, 0) \tag{5.20}$$

For a time-independent solution, the 1+log condition is

$$\beta^R \partial_R \alpha = n \alpha K. \tag{5.21}$$

Combining this with the trace of the extrinsic curvature

$$K = \frac{1}{\sqrt{-g}} (\sqrt{-g} n^\mu)_{,\mu}, \tag{5.22}$$

we obtain a first-order differential equation for the lapse (see [31] for details of its solution):

$$\alpha' = -\frac{n(3M - 2R + 2R\alpha^2)}{R(R - 2M + nR\alpha - R\alpha^2)} \tag{5.23}$$

This method of finding 1+log trumpet coordinates assumes that the trumpet metric is spherically symmetric and stationary. After a Lorentz boost, the trumpet spacetime violates both these conditions, with time-dependent metric coefficients and a direction of boosting. However, the basic method can be salvaged: with a two-dimensional height function, the 1+log condition becomes a second-order boundary value problem. In the next chapter, we derive a differential equation for the 2-dimensional height function and sketch its numerical solution.

Chapter 6

Boosted Trumpet Coordinates

To accurately describe motion in binaries, the trumpet coordinates need to be boosted. However, 1+log slicing is not preserved by a Lorentz boost. While either side of the 1+log equation $\mathcal{L}_{\hat{n}}\alpha = -nK$ may appear as an ordinary scalar, they are not in fact invariant; α and K are in fact made up of metric components and change under boosting. Fig. 6.1 illustrates the discrepancy between the extrinsic curvature computed from its definition and the 1+log condition. To reach the right trumpet coordinates, we will start with stationary 1+log trumpet initial data, introduce a 2-dimensional height function, and apply a Lorentz boost, before finally re-imposing 1+log slicing.

6.1 Initial Data

The initial data consists of a Schwarzschild black hole in trumpet coordinates and 1+log slicing. We generate this data following the approach used in [31] to put the trumpet solution in isotropic coordinates. These initial data include α, β_R, and the conformal factor ψ, as well as Trace(K). Throughout the code, Mathematica's *Interpolation* function is used to evaluate these quantities at a particular point.

We must choose the right 1+log coefficient (n in Eq. 5.10) to match the asymptotic behavior of the 1+log initial data with the asymptotic behavior of a boosted Schwarzschild spacetime. Regardless of the slicing condition, the behavior of the un-boosted solution should be the same for large r.

In this limit, starting from either Schwarzschild or isotropic coordinates, $\alpha \approx 1 - m/r$, $\psi \approx 1 + m/(2r)$, $\beta^i \approx 0$, and $\gamma_{ij} \approx (1 + 2m/r)\delta_{ij}$. The trace of the extrinsic curvature is given by

$$K = -\frac{1}{2\alpha}(\gamma^{ij}\partial_t\gamma_{ij} - 2\nabla_i\beta^i). \tag{6.1}$$

After boosting in the z direction, z only appears in functions of $z - vt$, allowing us to replace ∂_t with $-v\partial z$ in derivatives of the boosted metric. Applying this to the asymptotic 3-metric,

$$\partial_t\gamma_{ij} = -v\,\partial_z\left(1 + \frac{2m}{r}\right)\delta_{ij} = \frac{2mvz}{r^3}\delta_{ij}. \tag{6.2}$$

Taking the trace with γ^{ij}–which, to leading order is δ^{ij}, we have

$$\gamma^{ij}\partial_t\gamma_{ij} \approx \frac{6mvz}{r^3}. \tag{6.3}$$

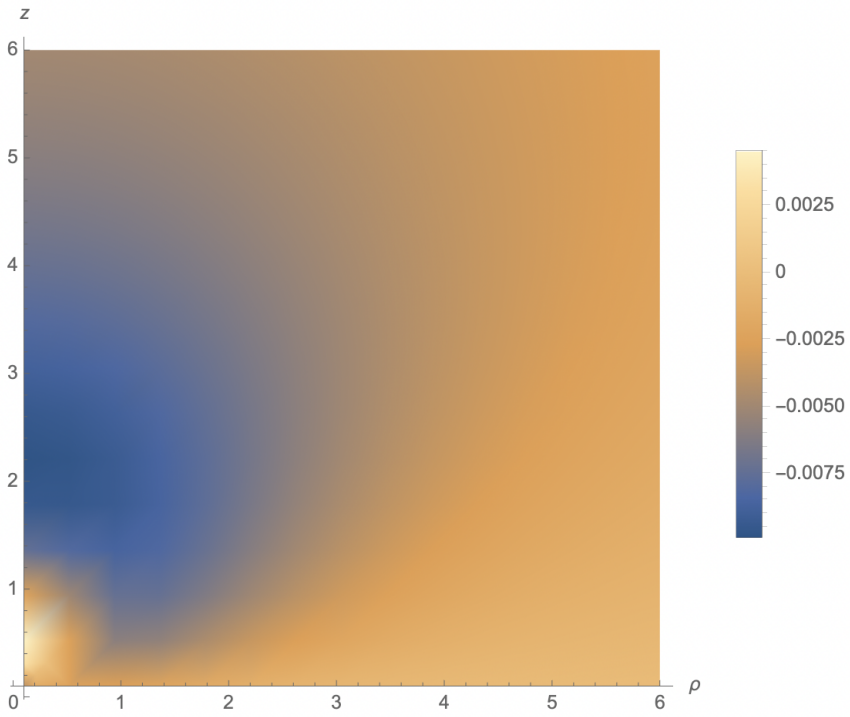


FIGURE 6.1: K computed from its definition minus K computed from the 1+log condition, i.e. $(\sqrt{-g}n^\mu)_{,\mu} - \frac{\sqrt{-g}}{\alpha}(\partial_t - \beta^i\partial_i)\alpha$, for $v = 0.3$. The difference is particularly large near the origin; outside the pictured range, it approaches 0.

Computing the shift vector of the boosted metric, we have

$$\beta^z \approx -v(\alpha^2 - \psi^4) \approx -4mv/r, \tag{6.4}$$

giving us $\nabla_i\beta^i \approx \partial_i\beta^i \approx 4mvz/r^3$.

Adding these terms, we obtain

$$\lim_{r\to\infty} K = \frac{mvz}{r^3}. \tag{6.5}$$

The 1 + log equation is

$$-n\alpha K = (\partial_t - \beta^i\partial_i)\alpha. \tag{6.6}$$

Using the advection relation, we see that the term $\partial_t\alpha = -v\partial_z\alpha$ goes as z/r^3. However, β^i goes as $1/r$, suppressing the $\beta^i\partial_i\alpha$ term to order z/r^4. So, the time derivative is the leading order term.

Dropping $\beta^i\partial_i\alpha$ and substituting in the asymptotic α, β, and γ, we obtain

$$K \approx \frac{1}{n}\frac{mvz}{r^3} \tag{6.7}$$

6.5 and 6.7 are only consistent if $n = 1$, so we choose a 1+log constant of 1 to generate initial data.

6.2 Boosted Trumpet Spacetime

Several transformations of the initial data are required before solving for the boosted trumpet system. First, we transform the Schwarzschild data from spherical to cylindrical coordinates.

In cylindrical coordinates, an arbitrary 3+1 spacetime with spherical symmetry takes the form

$$\begin{pmatrix} -\alpha(r)^2 + \psi(r)^4\beta^r(r)^2 & \frac{\rho}{r}\psi(r)^4\beta^r(r) & \frac{z}{r}\psi(r)^4\beta^r(r) & 0 \\ \frac{\rho}{r}\psi(r)^4\beta^r(r) & \psi(r)^4 & 0 & 0 \\ \frac{z}{r}\psi(r)^4\beta^r(r) & 0 & \psi(r)^4 & 0 \\ 0 & 0 & 0 & \rho^2\psi(r)^4 \end{pmatrix}. \tag{6.8}$$

We retain the functions $\alpha(r)$, β^r, and $\psi(r)$ from spherical coordinates since we have numerical data for them.

As with the stationary trumpet, a height function can be introduced into a metric with an arbitrary slicing condition to compute the correction to 1+log slicing. However, this height function needs to be 2-dimensional due to the lack of spherical symmetry. We apply the transformation $t \to T - h(\rho, z)$, doing so before boosting to avoid ambiguity between the two time coordinates and produce the simplest result. This way, only derivatives of h, not h itself, appear in the differential equation for the height function.

Finally, we apply a Lorentz boost along the $\hat{z}$ direction, producing a boosted trumpet metric. Since we will eventually set $T = 0$ for solution, and we can remove time derivatives using the advection condition $\partial_t = -v\partial_z$, we can simply apply the transformation $z \to \gamma z$ after contracting with the Lorentz transformation matrix twice.

After including the height function and boosting, the expressions for the metric components, α, and β^i–let alone K–become extremely long, taking up several pages. Throughout the process, we rely on Mathematica for symbolic calculations. Functions like *Simplify* and *Replace* do not work on such long nested expressions and can only be used on small pieces; inevitably further simplification is possible. Furthermore, different methods of computing K produce simpler or more complicated expressions. We found that

$$K = \frac{1}{\sqrt{-g}}\left(\sqrt{-g}n^\mu\right)_{,\mu} \tag{6.9}$$

produced the best results, using the fact that in general $\sqrt{-g} = \alpha\sqrt{\det(\gamma_{ij})}$.

6.3 Constructing the Boundary Value Problem

The definition of extrinsic curvature and the 1+log condition each produce a value for K. We can bring these two values into agreement by computing them for the boosted trumpet spacetime and iteratively solving for the height function. In particular, the equation to be solved is

$$\left(\sqrt{-g}n^\mu\right)_{,\mu} - \sqrt{-g}K_{1+\log} = 0 \tag{6.10}$$

where $K_{1+\log} = \frac{1}{\alpha}(\partial_t - \beta^i\partial_i)\alpha$.

h and up to second derivatives of h with respect to ρ and z appear in this equation. Our approach will be to automatically extract coefficients for these second derivatives, producing an equation of the form

$$a_{\rho\rho}h_{,\rho,\rho} + a_{\rho z}h_{,\rho,z} + a_{zz}h_{,z,z} + a_{\text{const}} = 0 \tag{6.11}$$

which can be addressed by finite differencing and root-finding.

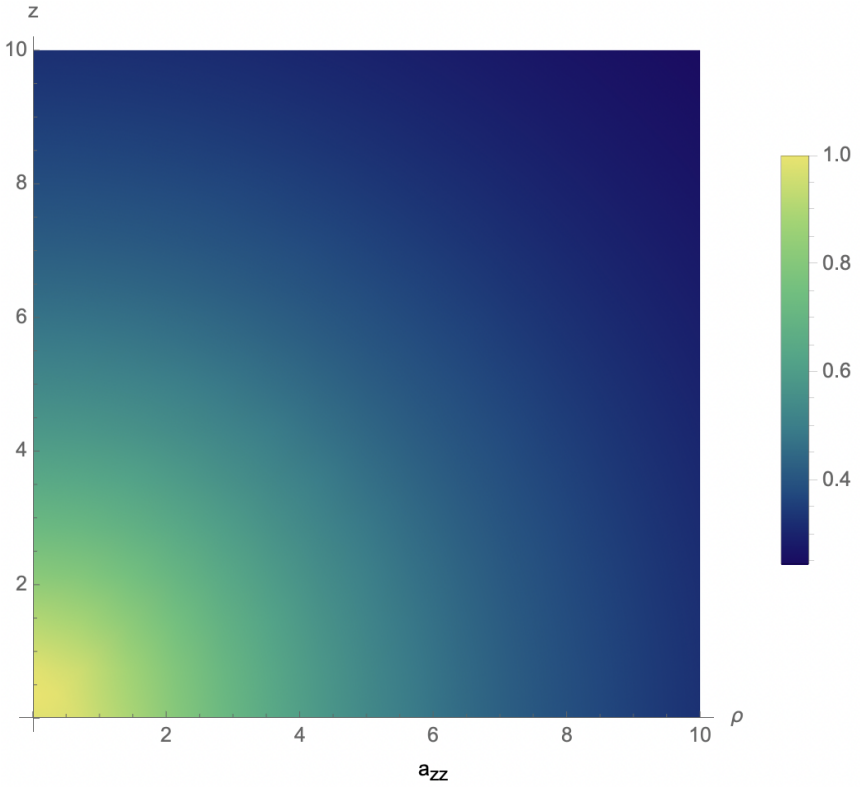
Solving the second-order system requires boundary conditions for h and its first derivatives, $h_{,\rho}$ and $h_{,z}$. From asymptotic flatness, $h = h_{,\rho} = h_{,z} = 0$ for large ρ and large z. Furthermore, $h' = 0$ on the boundaries $\rho = 0$, $z = 0$ to ensure axisymmetry and smoothness. These boundary conditions are confirmed by computing the difference from 1+log slicing seen in Fig. 6.1.

6.4 Next Steps

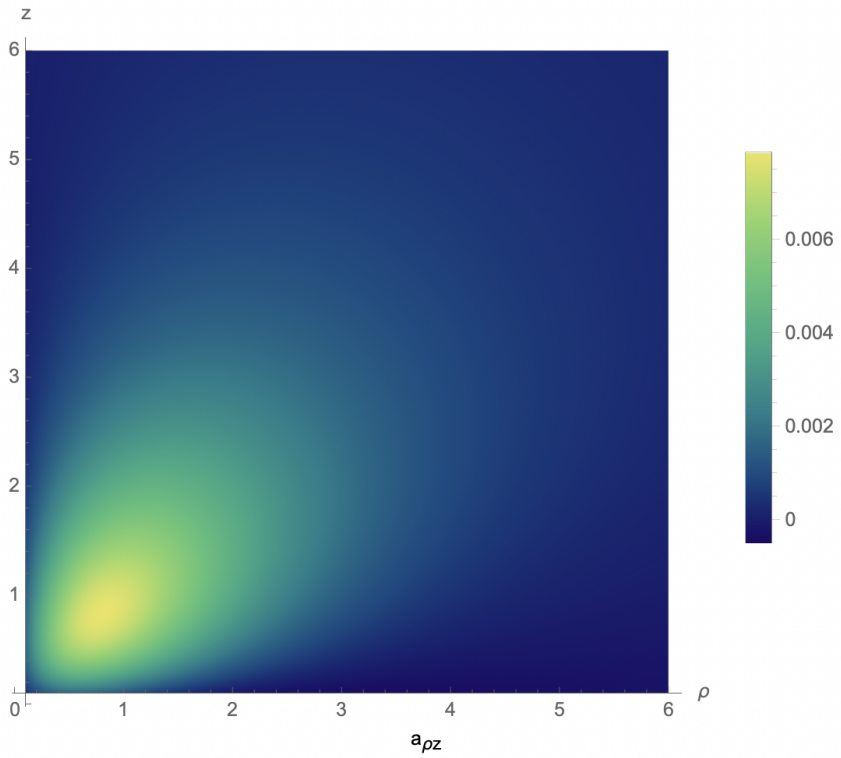
With the coefficients pictured in 6.2, the next step is to solve the equation. While the analytic methods used to solve for the height function and ultimately determine the lapse in [31] do not apply anymore, the problem is better suited to relaxation methods, a class of iterative methods for solving differential equations. Essentially, the equation can be linearized and discretized by finite differencing on a 2-dimensional grid in ρ and z. Starting from some initial guess, the equation is repeatedly solved to produce a correction to the current value of $h(\rho, z)$. Since only a small correction is expected, $h = 0$ everywhere is a suitable initial guess. On each step, we could obtain $h_{,\rho}$ and $h_{,z}$ from finite differencing and treat their values on each grid point as constants. Their values would then be used to compute the coefficients a_i to solve the second order equation.

The coefficients' odd behavior near the origin could lead to potential problems with numerical solution. In particular, a_{const} should go to 0 or some constant near the origin, not blow up. This likely results from Mathematica incorrectly evaluating the extremely complicated functions, and simplifying them further might help to make the limits more explicit. Furthermore, even with the correct limiting behavior, high resolution near the origin might be necessary for accurate strong field behavior. This could be accomplished either analytically by compactifying the coordinates, or numerically by adaptive mesh refinement.

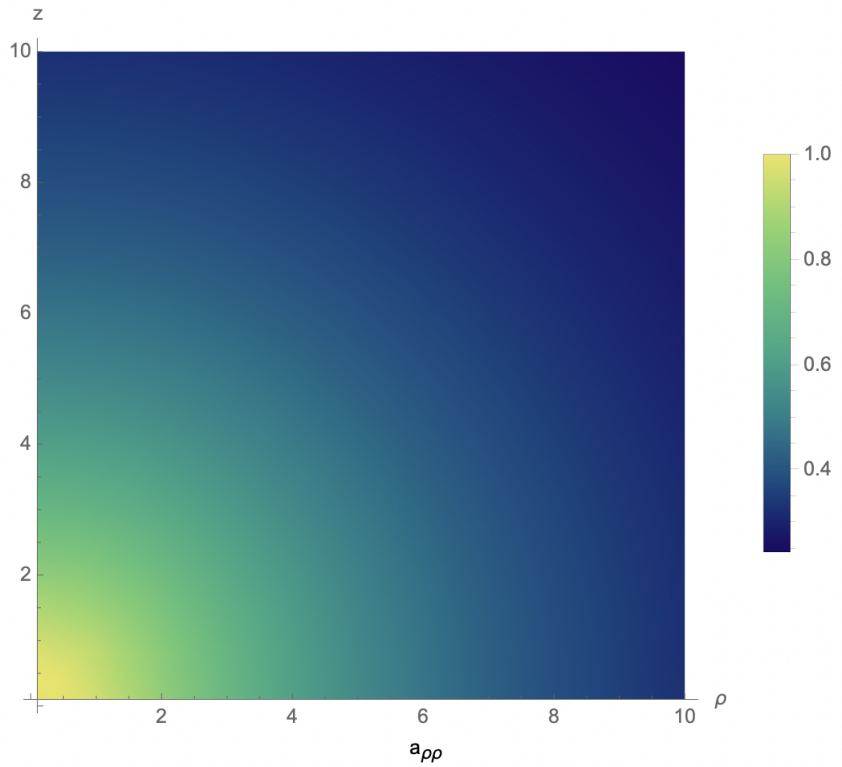
FIGURE 6.2: Coefficients for $h_{,z,z}$, $h_{,\rho,z}$, $h_{,\rho,\rho}$, and the constant term in Eqn. 6.10, evaluated with $v = 0.1$. Outside the pictured range, all of these coefficients approach 0 as $\rho, z \to \infty$.



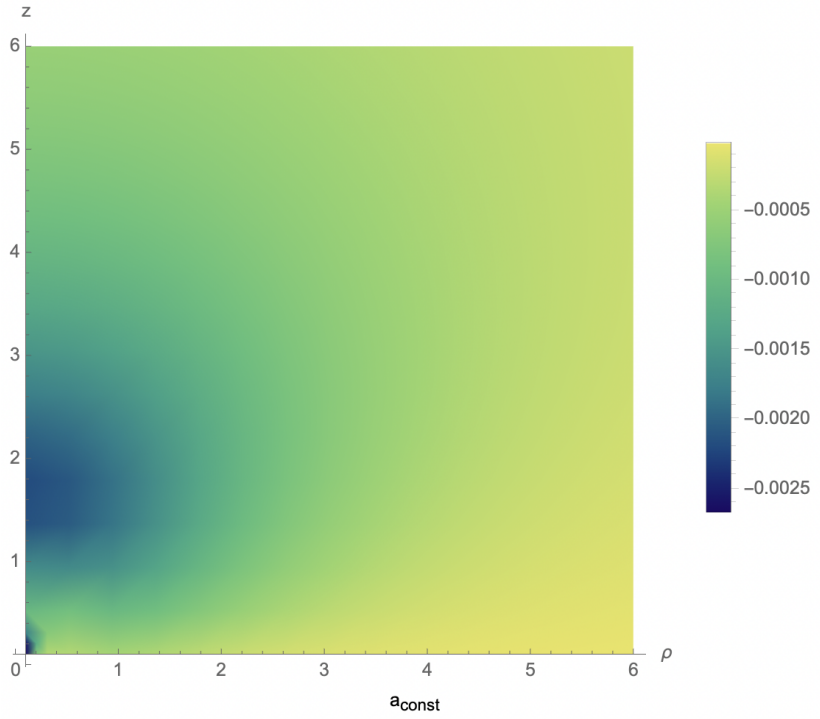
(A) Coefficient for $h_{,z,z}$ in the 1+log equation evaluated with $v = 0.1$.



(B) Coefficient for $h_{,\rho,z}$ in the 1+log equation evaluated with $v = 0.1$.



(C) Coefficient for $h_{,\rho,\rho}$ in the 1+log equation evaluated with $v = 0.1$.



(D) Constant term in the 1+log equation evaluated with $v = 0.1$. Some artifacts appear near the origin that might be resolved with simpler expressions for the coefficients.

Chapter 7

Conclusion

In Chapters 2, 3, and 4, we examined the hypothesis that the nested sampling analyses performed in GWTC-1 and GWTC-2 are well-behaved and unbiased, at least from the perspective of insertion order statistics. The event-level meta-p-values, shown in Figure 4.1 and listed in Table 4.1, follow a slightly non-uniform distribution at $p = 1.44 \times 10^{-3}$. In particular, although we could not identify an individual pathological case, Figure 4.1 shows that there is a slight excess of lower p-values, providing weak evidence for sampling misbehaviour. The vast majority of the posterior samples is collected towards the final stages of sampling, with dense exploration of the high-likelihood peak(s), so there is little reason to believe that weak signs of misbehaviour at random intervals would significantly impact the posterior distribution. However, such issues could effect evidence estimation.

While the few O1 and O2 events tend to have lower meta p-values, the events from O3a cover the entire range from 0 to 1. This may be (weak) evidence that the quality of nested sampling in LVC analysis has improved over time, with updates to the sampling algorithms and more computational resources. In particular, increasing the number of live points and the length of the MCMC chains used in prior sampling leads to more uniform insertion order statistics and, consequently, to more reliable nested sampling output.

The distribution of rolling test results pictured in Figure 4.3 is far more uniform than distribution of event-level meta test results, which is skewed toward lower p-values. The apparent contradiction between these results stems from the nature of the KS statistic. The KS test measures the supremum of distances from uniformity, so it measures whether any fault exists anywhere, not the average quality of sampling. While most individual segments are essentially uniform, it is unlikely that there would be no anomalies in multiple chains with tens of thousands of livepoints each. p-value adjustment for multiple tests reduces the impact of this effect, but does not eliminate it completely as long as the underlying distribution is non-uniform.

In combination, the rolling and meta tests indicate that sampling proceeds correctly in most small, local segments, but most events have at least one flaw in sampling somewhere. Moreover, we find that, however tempting it is to use continuous tests of uniformity (such as the Kolmogorov-Smirnov or Anderson-Darling tests) without transforming discrete data, or to try and recover partial results from chains with missing data points, these techniques lead to significant biases in the statistical results and should be avoided.

The location and p-value of the minimum rolling test can be used to identify and characterize these anomalous regions, and in practice, the combination of multiple

chains and the supplementation of nested sampling with MCMC posteriors further suppress their effects. Taking all complete parallel chains into account, the overall results are consistent with unbiased nested sampling.

In Chapters 5 and 6, we introduce the boosted trumpet spacetime in the context of moving-puncture numerical relativity. We began by generating 1+log sliced trumpet initial data, following the procedure outlined in [31]. By examining the leading-order contribution to the trace of the extrinsic curvature, we showed that the boosted 1+log spacetime will only have the right asymptotic behavior with 1+log coefficient $n = 1$.

We applied a Lorentz boost to the initial data and checked the 1+log condition. We verified that, without a height function or any other attempt to change the slicing, these boosted data do not satisfy the 1+log equation, as pictured in Fig. 6.1. At $T = 0$, the difference is particularly large for small radius in general and especially near the z-axis.

To remedy this disparity, we take a general spherically symmetric stationary spacetime and apply an as-yet unknown transformation to the time coordinate: the two-dimensional height function. We then boost in the $\hat{z}$ direction, reducing the system to axisymmetry. With these transformations, the 1+log condition becomes an equation for the height function, containing up to its second derivative with respect to ρ and z.

In terms of the lapse, shift, and conformal factor of the original spherical spacetime–the variables required to plug in the stationary trumpet initial data eventually–these expressions are extremely lengthy. However, the equation is linear in the second derivatives of h, though far from it in the first derivatives. We extracted coefficients for $h_{,\rho,\rho}$, $h_{,\rho,z}$, $h_{,z,z}$, and a constant term pictured in Fig. 6.2.

We propose treating this equation as a two-dimensional boundary problem using the constraints of axisymmetry and asymptotic flatness and solving it using relaxation methods. One limitation of the approach is that first derivative or h terms might turn out to be more important than expected, and computing them by finite differencing, rather than relaxing every possible variable at once, would not be sufficient. Another is the behavior near the puncture for ρ and z. Issues here could arise from the range and resolution of initial data, from how Mathematica evaluates long nested expressions, or the lack of compactification. With these matters resolved, an exact 1+log solution for the boosted trumpet spacetime could allow for faster evolution of asymmetric mass ratio binaries, ultimately expanding the horizons of numerical relativity.

Bibliography

[1] J. Aasi et al. "Advanced LIGO". In: *Class. Quant. Grav.* 32 (2015), 074001. DOI: 10.1088/0264-9381/32/7/074001. arXiv: 1411.4547 [gr-qc].

[2] B. P. Abbott et al. "Binary Black Hole Mergers in the first Advanced LIGO Observing Run". In: *Phys. Rev. X* 6.4 (2016), 041015. DOI: 10.1103/PhysRevX.6.041015. arXiv: 1606.04856 [gr-qc].

[3] B. P. Abbott 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". In: *Phys. Rev. X* 9.3 (2019), 031040. DOI: 10.1103/PhysRevX.9.031040. arXiv: 1811.12907 [astro-ph.HE].

[4] B. P. Abbott et al. "Binary Black Hole Population Properties Inferred from the First and Second Observing Runs of Advanced LIGO and Advanced Virgo". en. In: *The Astrophysical Journal Letters* 882.2 (Sept. 2019), L24. DOI: 10.3847/2041-8213/ab3800.

[5] R. Abbott et al. "GWTC-2: Compact Binary Coalescences Observed by LIGO and Virgo During the First Half of the Third Observing Run". In: *Phys. Rev. X* 11 (2021), 021053. DOI: 10.1103/PhysRevX.11.021053. arXiv: 2010.14527 [gr-qc].

[6] R. Abbott et al. "GWTC-2.1: Deep Extended Catalog of Compact Binary Coalescences Observed by LIGO and Virgo During the First Half of the Third Observing Run". In: *arXiv* 2108.01045 [gr-qc] (Aug. 2021).

[7] F. Acernese et al. "Advanced Virgo: a second-generation interferometric gravitational wave detector". In: *Class. Quant. Grav.* 32.2 (2015), 024001. DOI: 10.1088/0264-9381/32/2/024001. arXiv: 1408.3978 [gr-qc].

[8] Miguel Alcubierre et al. "Gauge conditions for long-term numerical black hole evolutions without excision". In: *Physical Review D* 67.8 (Apr. 2003), 084023. DOI: 10.1103/PhysRevD.67.084023.

[9] Pau Amaro-Seoane et al. *Laser Interferometer Space Antenna*. Feb. 2017. DOI: 10.48550/arXiv.1702.00786.

[10] Tomas Andrade et al. "GRChombo: An adaptable numerical relativity code for fundamental physics". en. In: *Journal of Open Source Software* 6.68 (Dec. 2021), 3703. DOI: 10.21105/joss.03703.

[11] Taylor B Arnold and John W Emerson. "Nonparametric goodness-of-fit tests for discrete null distributions." In: *R Journal* 3.2 (2011).

[12] Gregory Ashton et al. "BILBY: A user-friendly Bayesian inference library for gravitational-wave astronomy". In: *Astrophys. J. Suppl.* 241.2 (2019), 27. DOI: 10.3847/1538-4365/ab06fc. arXiv: 1811.02042 [astro-ph.IM].

[13] Gregory Ashton et al. "BILBY: A User-friendly Bayesian Inference Library for Gravitational-wave Astronomy". In: *The Astrophysical Journal Supplement Series* 241 (Apr. 1, 2019), 27. DOI: 10.3847/1538-4365/ab06fc.

[14] Greg Ashton et al. "Nested sampling for physical scientists". In: *Nature Reviews Methods Primers* 2.1 (May 2022). DOI: 10.1038/s43586-022-00121-x.

[15] John G. Baker et al. "Gravitational-Wave Extraction from an Inspiraling Configuration of Merging Black Holes". In: *Physical Review Letters* 96.11 (Mar. 2006), 111102. DOI: 10.1103/PhysRevLett.96.111102.

[16] Leor Barack et al. "Black holes, gravitational waves and fundamental physics: a roadmap". In: *Classical and Quantum Gravity* 36.14 (June 2019), 143001. DOI: 10.1088/1361-6382/ab0587.

[17] Thomas W. Baumgarte and Stuart L. Shapiro. *Numerical Relativity: Solving Einstein's Equations on the Computer*. Cambridge: Cambridge University Press, 2010. DOI: 10.1017/CBO9781139193344.

[18] Carles Bona et al. "New Formalism for Numerical Relativity". In: *Physical Review Letters* 75.4 (July 1995), 600–603. DOI: 10.1103/PhysRevLett.75.600.

[19] S. R. Brandt and B. Bruegmann. *BH Punctures as Initial Data for General Relativity*. Nov. 1997. DOI: 10.48550/arXiv.gr-qc/9711015.

[20] Bernd Brügmann, Wolfgang Tichy, and Nina Jansen. "Numerical Simulation of Orbiting Black Holes". In: *Physical Review Letters* 92.21 (May 2004), 211101. DOI: 10.1103/PhysRevLett.92.211101.

[21] Bernd Brügmann et al. "Calibration of moving puncture simulations". In: *Physical Review D* 77.2 (Jan. 2008), 024027. DOI: 10.1103/PhysRevD.77.024027.

[22] Johannes Buchner. "A statistical test for Nested Sampling algorithms". In: *Statistics and Computing* 26.1-2 (Jan. 2016), 383–392. DOI: 10.1007/s11222-014-9512-y.

[23] Johannes Buchner. "UltraNest – a robust, general purpose Bayesian inference engine". In: *arXiv* 2101.09604 [stat.CO] (2021). DOI: 10.48550/ARXIV.2101.09604.

[24] The LIGO Scientific Collaboration et al. *GWTC-3: Compact Binary Coalescences Observed by LIGO and Virgo During the Second Part of the Third Observing Run*. Nov. 2021. DOI: 10.48550/arXiv.2111.03606.

[25] The LIGO Scientific Collaboration et al. "Population Properties of Compact Objects from the Second LIGO-Virgo Gravitational-Wave Transient Catalog". In: *The Astrophysical Journal Letters* 913.1 (May 2021), L7. DOI: 10.3847/2041-8213/abe949.

[26] The LIGO Scientific Collaboration et al. *The population of merging compact binaries inferred using gravitational waves through GWTC-3*. Feb. 2022. DOI: 10.48550/arXiv.2111.03634.

[27] Kenneth A. Dennison and Thomas W. Baumgarte. "A simple family of analytical trumpet slices of the Schwarzschild spacetime". en. In: *Classical and Quantum Gravity* 31.11 (May 2014), 117001. DOI: 10.1088/0264-9381/31/11/117001.

[28] Jean-Marie Dufour and Abdeljelil Farhat. "Exact Nonparametric Two-Sample Homogeneity Tests". In: *Goodness-of-Fit Tests and Model Validity*. Ed. by C. Huber-Carol et al. Boston, MA: Birkhäuser Boston, 2002, 435–448. DOI: 10.1007/978-1-4612-0103-8_33.

[29] Andrew Fowlie, Will Handley, and Liangliang Su. "Nested sampling cross-checks using order statistics". In: *Mon. Not. Roy. Astron. Soc.* 497.4 (2020), 5256–5263. DOI: 10.1093/mnras/staa2345. arXiv: 2006.03371 [stat.CO].

[30] Mark Hannam et al. "Geometry and Regularity of Moving Punctures". In: *Physical Review Letters* 99.24 (Dec. 2007), 241102. DOI: 10.1103/PhysRevLett.99.241102.

[31] Mark Hannam et al. "Wormholes and trumpets: The Schwarzschild spacetime for the moving-puncture generation". In: *Phys. Rev. D* 78 (2008), 064020. DOI: 10.1103/PhysRevD.78.064020.

[32] Edward Higson et al. "Nestcheck: diagnostic tests for nested sampling calculations". In: *Mon. Not. Roy. Astron. Soc.* 483.2 (2019), 2044–2056. DOI: 10.1093/mnras/sty3090. arXiv: 1804.06406 [stat.CO].

[33] Sebastian Khan et al. "Frequency-domain gravitational waves from nonprecessing black-hole binaries. II. A phenomenological model for the advanced detector era". In: *Phys. Rev. D* 93.4 (2016), 044007. DOI: 10.1103/PhysRevD.93.044007. arXiv: 1508.07253 [gr-qc].

[34] Lawrence E. Kidder et al. "Black hole evolution by spectral methods". In: *Physical Review D* 62.8 (Sept. 2000), 084032. DOI: 10.1103/PhysRevD.62.084032.

[35] Talya Klinger and Michalis Agathos. "An Order Statistics Post-Mortem on LIGO–Virgo GWTC-2 Events Analyzed with Nested Sampling". en. In: *Annalen der Physik* n/a.n/a (), 2200271. DOI: 10.1002/andp.202200271.

[36] J. Lange et al. "Parameter estimation method that directly compares gravitational wave observations to numerical relativity". In: *Physical Review D* 96.10 (Nov. 2017). DOI: 10.1103/physrevd.96.104041.

[37] LIGO Scientific Collaboration. *LIGO Algorithm Library - LALSuite*. Free Software (GPL). 2018. DOI: 10.7935/GT1W-FZ16.

[38] LIGO Scientific Collaboration and Virgo Collaboration et al. "Observation of Gravitational Waves from a Binary Black Hole Merger". In: *Physical Review Letters* 116.6 (Feb. 11, 2016), 061102. DOI: 10.1103/PhysRevLett.116.061102.

[39] LIGO Scientific Collaboration and Virgo Collaboration et al. "Tests of general relativity with binary black holes from the second LIGO-Virgo gravitational-wave transient catalog". In: *Physical Review D* 103.12 (June 15, 2021), 122002. DOI: 10.1103/PhysRevD.103.122002.

[40] Frank Löffler et al. "The Einstein Toolkit: a community computational infrastructure for relativistic astrophysics". en. In: *Classical and Quantum Gravity* 29.11 (May 2012), 115001. DOI: 10.1088/0264-9381/29/11/115001.

[41] Michele Maggiore et al. "Science Case for the Einstein Telescope". In: *Journal of Cosmology and Astroparticle Physics* 2020.03 (Mar. 2020), 050–050. DOI: 10.1088/1475-7516/2020/03/050.

[42] Frank J. Massey. "The Kolmogorov-Smirnov Test for Goodness of Fit". In: *Journal of the American Statistical Association* 46.253 (1951), 68–78.

[43] James R. van Meter et al. "How to move a black hole without excision: Gauge conditions for the numerical evolution of a moving puncture". In: *Physical Review D* 73.12 (June 2006), 124011. DOI: 10.1103/PhysRevD.73.124011.

[44] Serguei Ossokine et al. "Multipolar Effective-One-Body Waveforms for Precessing Binary Black Holes: Construction and Validation". In: *Phys. Rev. D* 102.4 (2020), 044055. DOI: 10.1103/PhysRevD.102.044055. arXiv: 2004.09442 [gr-qc].

[45] John Skilling. "Nested sampling for general Bayesian computation". In: *Bayesian Analysis* 1.4 (2006), 833–859. DOI: 10.1214/06-BA127.

[46] The LIGO Scientific Collaboration and the Virgo Collaboration et al. "Tests of general relativity with the binary black hole signals from the LIGO-Virgo catalog GWTC-1". In: *Physical Review D* 100.10 (Nov. 20, 2019), 104036. DOI: 10.1103/PhysRevD.100.104036.

[47] J. Veitch et al. "Parameter estimation for compact binaries with ground-based gravitational-wave observations using the LALInference software library". In: *Phys. Rev. D* 91.4 (2015), 042003. DOI: 10.1103/PhysRevD.91.042003. arXiv: 1409.7215 [gr-qc].

[48] John Veitch et al. *johnveitch/cpnest: v0.11.4*. Version v0.11.4. Apr. 2022. DOI: 10.5281/zenodo.6460935.

[49] Pauli Virtanen et al. "SciPy 1.0–Fundamental Algorithms for Scientific Computing in Python". In: *Nature Meth.* 17 (2020), 261. DOI: 10.1038/s41592-019-0686-2. arXiv: 1907.10121 [cs.MS].

[50] Michael J. Williams, John Veitch, and Chris Messenger. "Nested sampling with normalizing flows for gravitational-wave inference". In: *Physical Review D* 103.10 (May 2021). DOI: 10.1103/physrevd.103.103006.