Towards Unbiased Gravitational-Wave Parameter Estimation using Score-Based Likelihood Characterization

Ronan Legin [*1 2 3] **Maximiliano Isi** [*4] **Kaze Wong** [*4] **Alexandre Adam** [1 2 3]
Laurence Perreault-Levasseur [1 2 3 4 5] **Yashar Hezaveh** [1 2 3 4 5]

Abstract

Gravitational wave (GW) parameter estimation has conventionally relied on the assumption of Gaussian and stationary noise. However, noise from real-world detectors, such as LIGO, Virgo and KAGRA, often deviates considerably from these assumptions. In this paper, we use score-based diffusion models to learn an empirical noise distribution directly from detector data, which can then be combined with the forward simulator of the physical model to provide an unbiased model of the likelihood function. We validate the method by performing inference on a simulated gravitational wave event injected in real detector noise from LIGO, demonstrating its potential for providing accurate and scalable GW parameter estimation.

1. Introduction

Gravitational-wave (GW) detectors, like LIGO (Aasi et al., 2015) Virgo (Acernese et al., 2015) and KAGRA (Aso et al., 2013), record noisy time series encoding astrophysically-valuable signals from black hole and neutron star collisions in far away galaxies (Abbott et al., 2016b; 2017b; 2021b). Extracting science from these data requires using Bayesian inference to estimate the source parameters (like black hole masses and spins) (Veitch et al., 2015; Ashton et al., 2019; Romero-Shaw et al., 2020; Biwer et al., 2019). As in many other areas of astronomy, GW parameter estimation has traditionally assumed that instrumental noise is Gaussian and stationary (Abbott et al., 2020). This idealization is approximately justified for short segments of data by the central limit theorem, and has the additional advantage of having a tractable and inexpensive likelihood.

However, in reality, the statistics of noise often deviate significantly from a stationary Gaussian: most notably, the instruments evolve over time, and the data are contaminated by both transient non-Gaussian excursions ("glitches") and the nonlinear evolution of narrow spectral features ("lines") (Abbott et al., 2016c;a; Buikema et al., 2020); in principle, the data are also contaminated by subthreshold astrophysical signals (Abbott et al., 2021c). Since no generative models exist for most of these contaminants, they can usually only be handled by bespoke treatment of the affected data segments, which can be computationally expensive and can result in biases. The most prominent example of this was the first detection of a binary neutron star merger (Abbott et al., 2017b), which was famously contaminated by a loud glitch in one of the LIGO detectors.

In this work, we showcase a new inference framework to carry out GW parameter estimation without assuming stationary Gaussian noise, and without sacrificing the advantages of deterministic signal models on which standard GW inference relies. This would make it possible to analyze signals contaminated by noise artifacts without the need for special data treatments, and reduce biases in the analysis of large collections of signals which would otherwise be sensitive even to small departures from the Gaussian assumption.

We propose to do this by adapting the method used by Legin et al. (2023), originally focused on modelling noise in image data from the Hubble Space Telescope (HST) and the James Webb Space Telescope (JWST). In a similar spirit, we train a score-based diffusion model on samples of LIGO noise to learn the gradient of the log probability density of the noise distribution with respect to data elements; combined with the Jacobian of the physical signal model, we use Langevin sampling to produce draws from the likelihood without assuming any specific noise properties besides additivity.

[*]Equal contribution [1]Department of Physics, Université de Montréal, Montréal, Canada [2]Mila - Quebec Artificial Intelligence Institute, Montréal, Canada [3]Ciela - Montreal Institute for Astrophysical Data Analysis and Machine Learning, Montréal, Canada [4]Center for Computational Astrophysics, Flatiron Institute, New York, USA [5]Perimeter Institute for Theoretical Physics, Waterloo, Ontario, Canada. Correspondence to: Ronan Legin <ronan.legin@umontreal.ca>, Maximiliano Isi <misi@flatironinstitute.org>, Kaze Wong <kwong@flatironinstitute.org>.

Proceedings of the 40^{th} International Conference on Machine Learning, Honolulu, Hawaii, USA. PMLR 202, 2023.

The structure of the paper is as follows. In Section 2 we summarize the inference method and describe our specific forward model and data. In Section 3 we present examples of our setup applied to recover parameters for a simulated gravitational wave event in real detector noise. We discuss our results and future work in Section 4.

2. Method

2.1. Score-based models for non-Gaussian likelihoods

In this section, we present a concise overview of the technique introduced by Legin et al. (2023) to perform inference using non-Gaussian likelihoods. The setting is that of an inverse problem with additive noise. With this single assumption, we can model the likelihood, or more precisely the gradient of the log (aka, score) of the likelihood, by learning the score of the noise distribution. Given samples $\mathbf{x} \sim Q(\mathbf{x})$ from the underlying distribution of noise, Q, the score function, $\boldsymbol{\nabla}_{\mathbf{x}} \log Q(\mathbf{x})$, is learned with a neural network using denoising score matching (Hyvärinen, 2005; Vincent, 2011; Song et al., 2020). The trained score model can then be used to construct the score of the likelihood

$$\boldsymbol{\nabla}_{\theta} \log p(\mathbf{x}_O \mid \theta) = -\boldsymbol{\nabla}_{\mathbf{x}} \log Q(\mathbf{x}) \boldsymbol{\nabla}_{\theta} M(\theta)\,, \quad (1)$$

where $\mathbf{x}_O = M(\theta) + \mathbf{x}$ is an observed GW signal and where $M(\theta)$ is the physical model, which takes as input the parameters of interest, θ. During inference, $\mathbf{x}$ are residuals between the observation and the model, whose statistics should follow the noise if the signal model is accurate.

To generate samples from this likelihood, we employ sampling techniques based on Langevin dynamics which explicitly make use of the score function to sample from the underlying distribution. Specifically, we use a Metropolis-adjusted Langevin algorithm (MALA) to sample the target distribution. Following Legin et al. (2023), we refer to the setup as Score-based LIkelihood Characterization (SLIC).

2.2. Forward model

For our physical model $M(\theta)$, we use the differentiable version of the IMRPHENOMD waveform model (Khan et al., 2016) implemented in the RIPPLE package (Edwards et al., 2023). The ability to automatically differentiate the waveform is crucial to compute the required Jacobian. For a binary system with masses $m_{1/2}$ and dimensionless spins $\chi_{1/2}$, this waveform model is parameterized in terms of the chirp mass, $\mathcal{M} = (m_1 m_2)^{3/5}/(m_1 + m_2)^{1/5}$, the symmetric mass ratio, $\eta = m_1 m_2/(m_1 + m_2)$, and the spin magnitudes; the spins are assumed to be aligned with the orbital angular momentum of the binary. We sample in $\mathcal{M}$ and η, as well as in the source luminosity distance, d_L (a scaling factor for the signal amplitude), the coalescence (signal arrival) time, t_c, and a fiducial coalescence phase, ϕ_c. For simplicity, below we assume $\chi_1 = \chi_2 = 0$ and a sky location given by right ascension $\alpha = 1.95\,\mathrm{rad}$, declination $\delta = -1.27\,\mathrm{rad}$ and polarization angle $\psi = 0.82\,\mathrm{rad}$, consistent with the first ever GW detection, GW150914 (Abbott et al., 2016b;d).

2.3. Data

We train the score-based network on 11 h of actual LIGO data around GW150914 (GPS time 1126259462.423 s) obtained from the Gravitational Wave Open Science Center (Abbott et al., 2021a; Abbot et al., 2023). We split these data into 4 s segments for training, taking care to discard the one segment containing the true signal. Following standard LIGO-Virgo practice, we use data sampled at 4096 Hz; we apply no filters besides those pre-applied to the public data, but we begin the likelihood integration at 20 Hz. We train the network in the Fourier domain, after applying a Tukey window with shape parameter $\alpha_T = 0.1$ to each 4 s segment. For this demonstration, we only make use of the LIGO-Hanford instrument.

To test our inference framework, we add a simulated signal into noise data occurring *after* the end of our training set (GPS time 1126310000 seconds). The simulated signal had parameters consistent with GW150914 ($\mathcal{M} = 29\,M_{\odot}, \eta = 0.2495, \chi_1 = \chi_2 = 0, d_L = 400\,\mathrm{Mpc}$) and was placed at the center of a 4 s segment of noise, mimicking the standard setup for a real analysis.

3. Results

We demonstrate the potential of SLIC for GW astronomy in two ways: (1) by showing it can generate realistic synthetic noise, and (2) that it can be used to recover source parameters from a simulated signal in real LIGO data.

For (1), we sample 1024 noise realizations from SLIC and compare their Fourier amplitudes to the noise power spectral density (PSD) estimated from actual LIGO data using Welch's method. Figure 1 shows that the PSD estimated from both the model samples and Welch's method are in close agreement.[1] In particular, the model can reproduce key features of LIGO noise, like narrow spectral lines in the PSD, which are highly localized features that are difficult to learn in general. In addition to reproducing an overall power spectrum, the SLIC model can also encapsulate phase correlations encoding non-Gaussian features, like glitches. Figure 1 demonstrates the model's ability to capture intricate details in the noise distribution and gives us confidence that the model can generate realistic noise samples for our analysis.

[1] Exact agreement is not expected given the differences in data (11 h of training data vs 1 h outside training window), and estimation techniques (mean over independent 4s segments vs Welch).

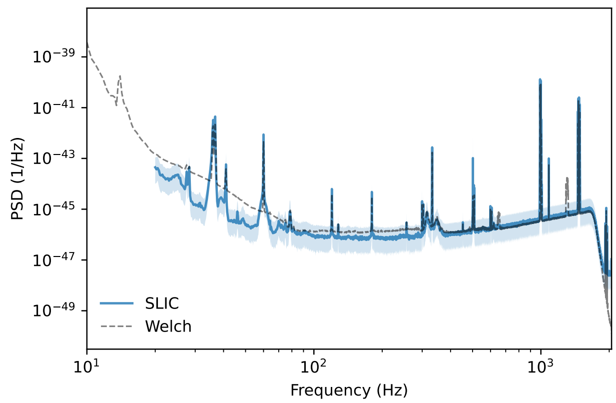

Figure 1. One-sided noise power spectral density (PSD) of LIGO noise as estimated from 1024 samples of synthetic 4s-long segments of noise generated by our score-based model trained on 11 h of LIGO noise (SLIC, blue), compared to a mean Welch estimate measured empirically from 1 h of real LIGO Hanford noise recorded after the data used in training and not overlapping with it (Welch, dashed gray). The blue band represents the 68% symmetric confidence interval around the median (solid blue line). SLIC is able to capture key features of the spectral properties of LIGO noise, resolving lines more finely than the Welch estimate thanks to its exposure to greater amounts of data (11 h vs 1 h).

Next (2), knowing SLIC can faithfully generate noise samples, we check that the SLIC likelihood can infer source parameters without bias. To this end, we inject simulated signals with known parameters into real LIGO noise not seen in training, and estimate the parameters using both our model and a standard Gaussian likelihood assuming a Welch PSD as in Fig. 1. Figure 2 shows the resulting posterior distributions for the five varied parameters under both methods. As shown, the SLIC posterior is capable of recovering the true injected values with high credibility. For this example, we have chosen a segment of real data without significant glitches, such that the Gaussian likelihood also recovers the true parameters.

4. Discussion

In this work, we trained a score-based diffusion model to learn the likelihood of real GW detector data. We coupled that to a deterministic forward model for the signal in order to carry out inference without the standard idealization of stationary Gaussian noise. The model is trained on actual noise samples and does not assume any knowledge specific to the GW detectors, which makes this a powerful tool for the analysis of current and future detectors, such as Cosmic Explorer (Abbott et al., 2017a) and Einstein Telescope (Punturo et al., 2010).

We are working on a number of improvements before applying our framework to real signals contaminated by instrumental glitches, such as GW170817. First, we need to expand the length of data the model can generate: currently, our model can only be used to analyze 4 s segments of data, suitable for binary black holes like GW150914 but not for binary neutron stars like GW170817, which typically require 128 s of data. Training a model for a longer segment requires more data, modifications in the architecture of the model such as dilating the convolution window, and longer training. Additionally, we currently only analyze data from a single detector at a time, rather than coherently modeling data from a network of detectors in our inference step. Coherent analysis will significantly improve parameter estimation because true signals appear coherently across detectors, while noise (even non-Gaussian) does not. Extending our pipeline to coherently analyze multi-detector data by training score models for different instruments is a straightforward extension of our setup that we are actively developing.

Currently, we are using MALA as our sampler. This could limit efficiency when we tackle the full problem compact-binary inference problem, which generally requires sampling over 15 dimensions with a nontrivial posterior structure. Integrating the likelihood as learned by our score-based model with more powerful sampling algorithms such as FLOWMC (Wong et al., 2023a) promises significant improvements in performance. Methods like FLOWMC can explore complex, high-dimensional spaces much more efficiently than MALA (e.g., Wong et al., 2023b). By combining the SLIC likelihood with an advanced sampler, we could achieve fast, accurate parameter estimation for real gravitational wave events even in the presence of non-Gaussianities and nonstationarity in the data.

In conclusion, we have demonstrated a method for learning the likelihood function for gravitational wave parameter estimation using score-based diffusion models that will allow us to analyze real data with fewer or no idealizations. While promising, this framework currently relies on simple samplers that limit its applicability to simplified problems in lower dimensions. Integrating advanced sampling techniques would be a major step towards applying this method for real gravitational wave inference in high-dimensional, multi-modal parameter spaces. With further development, machine learning models and sampling algorithms will enable fast, scalable and unbiased Bayesian inference for future gravitational wave observations.

Acknowledgements

This material is based upon work supported by NSF's LIGO Laboratory which is a major facility fully funded by the National Science Foundation. This work is partially supported by Schmidt Futures, a philanthropic initiative founded by

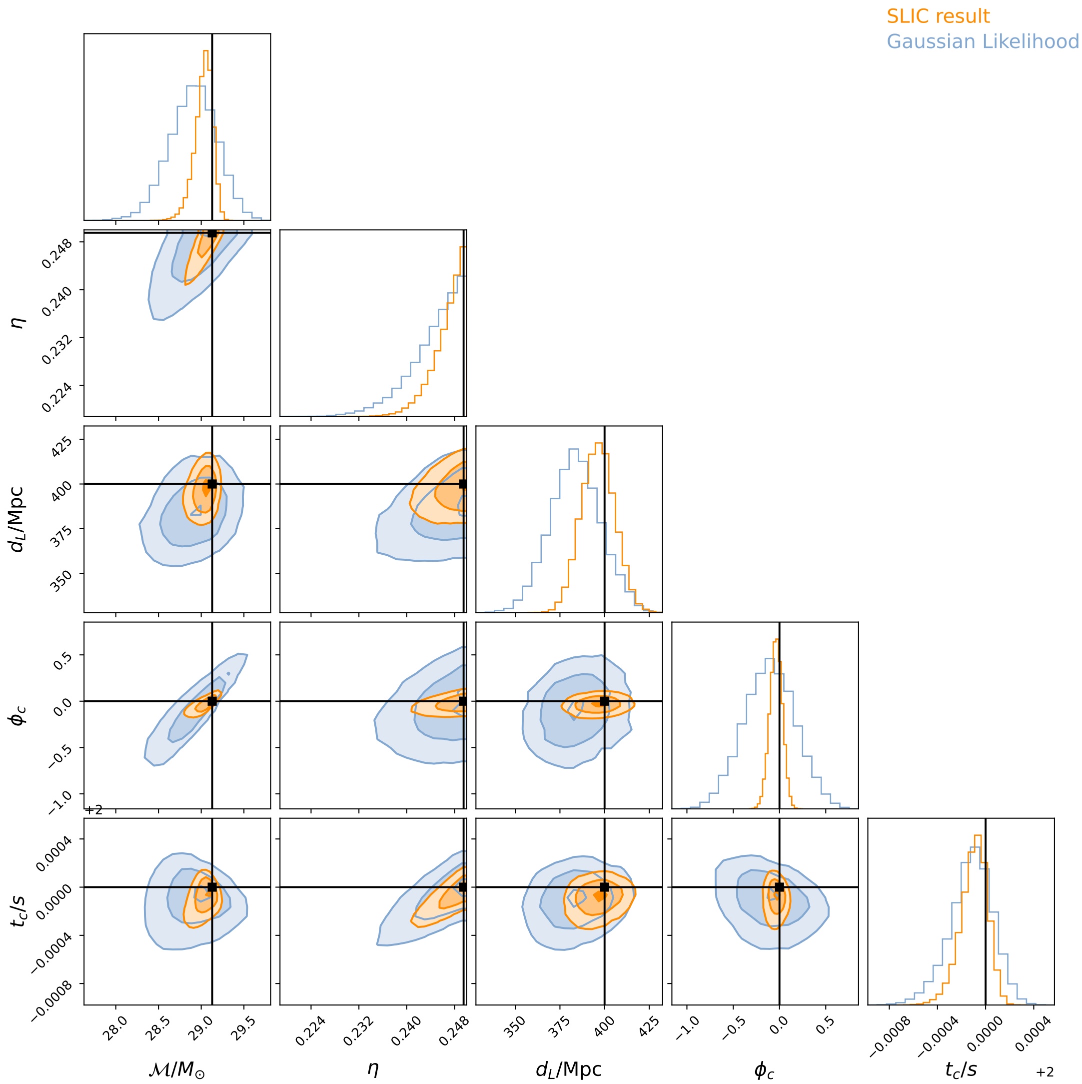

Figure 2. Posterior distribution of the five recovered parameters of a simulated signal using SLIC (orange) and the standard Gaussian likelihood (blue). The contours are the 10%, 39.35% and 90% interval. The true values are marked by the black lines. The posterior distribution using SLIC has a smaller variance compared to the Gaussian likelihood.

Eric and Wendy Schmidt as part of the Virtual Institute for Astrophysics (VIA). The Flatiron Institute is supported by the Simons Foundation. The work is in part supported by computational resources provided by Calcul Quebec and the Digital Research Alliance of Canada. Y.H. and L.P. acknowledge support from the Canada Research Chairs Program, the National Sciences and Engineering Council of Canada through grants RGPIN-2020-05073 and 05102, and the Fonds de recherche du Québec through grants 2022-NC-301305 and 300397.

References

Aasi, J. et al. Advanced LIGO. *Class. Quant. Grav.*, 32: 074001, 2015. doi: 10.1088/0264-9381/32/7/074001.

Abbot, R. et al. Gravitational wave open science center (1126216262-1126302662), 2023. URL https://gwosc.org/archive/links/O1/H1/1126216262/1126302662/simple/.

Abbott, B. P. et al. Sensitivity of the Advanced LIGO detectors at the beginning of gravitational wave astronomy. *Phys. Rev. D*, 93(11):112004, 2016a. doi: 10.1103/PhysRevD.93.112004. [Addendum: Phys.Rev.D 97, 059901 (2018)].

Abbott, B. P. et al. Observation of Gravitational Waves from a Binary Black Hole Merger. *Phys. Rev. Lett.*, 116(6): 061102, 2016b. doi: 10.1103/PhysRevLett.116.061102.

Abbott, B. P. et al. Characterization of transient noise in Advanced LIGO relevant to gravitational wave signal GW150914. *Class. Quant. Grav.*, 33(13):134001, 2016c. doi: 10.1088/0264-9381/33/13/134001.

Abbott, B. P. et al. Properties of the Binary Black Hole Merger GW150914. *Phys. Rev. Lett.*, 116(24):241102, 2016d. doi: 10.1103/PhysRevLett.116.241102.

Abbott, B. P. et al. Exploring the Sensitivity of Next Generation Gravitational Wave Detectors. *Class. Quant. Grav.*, 34(4):044001, 2017a. doi: 10.1088/1361-6382/aa51f4.

Abbott, B. P. et al. GW170817: Observation of Gravitational Waves from a Binary Neutron Star Inspiral. *Phys. Rev. Lett.*, 119(16):161101, 2017b. doi: 10.1103/PhysRevLett.119.161101.

Abbott, B. P. et al. A guide to LIGO–Virgo detector noise and extraction of transient gravitational-wave signals. *Class. Quant. Grav.*, 37(5):055002, 2020. doi: 10.1088/1361-6382/ab685e.

Abbott, R. et al. Open data from the first and second observing runs of Advanced LIGO and Advanced Virgo. *SoftwareX*, 13:100658, 2021a. doi: 10.1016/j.softx.2021.100658.

Abbott, R. et al. GWTC-3: Compact Binary Coalescences Observed by LIGO and Virgo During the Second Part of the Third Observing Run. 11 2021b.

Abbott, R. 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. 8 2021c.

Acernese, F. et al. Advanced Virgo: a second-generation interferometric gravitational wave detector. *Class. Quant. Grav.*, 32(2):024001, 2015. doi: 10.1088/0264-9381/32/2/024001.

Ashton, G. et al. BILBY: A user-friendly Bayesian inference library for gravitational-wave astronomy. *Astrophys. J. Suppl.*, 241(2):27, 2019. doi: 10.3847/1538-4365/ab06fc.

Aso, Y., Michimura, Y., Somiya, K., Ando, M., Miyakawa, O., Sekiguchi, T., Tatsumi, D., and Yamamoto, H. Interferometer design of the KAGRA gravitational wave detector. *Phys. Rev. D*, 88(4):043007, 2013. doi: 10.1103/PhysRevD.88.043007.

Biwer, C. M., Capano, C. D., De, S., Cabero, M., Brown, D. A., Nitz, A. H., and Raymond, V. PyCBC Inference: A Python-based parameter estimation toolkit for compact binary coalescence signals. *Publ. Astron. Soc. Pac.*, 131 (996):024503, 2019. doi: 10.1088/1538-3873/aaef0b.

Buikema, A. et al. Sensitivity and performance of the Advanced LIGO detectors in the third observing run. *Phys. Rev. D*, 102(6):062003, 2020. doi: 10.1103/PhysRevD.102.062003.

Edwards, T. D. P., Wong, K. W. K., Lam, K. K. H., Coogan, A., Foreman-Mackey, D., Isi, M., and Zimmerman, A. ripple: Differentiable and Hardware-Accelerated Waveforms for Gravitational Wave Data Analysis. 2 2023.

Hyvärinen, A. Estimation of non-normalized statistical models by score matching. *J. Mach. Learn. Res.*, 6:695–709, dec 2005. ISSN 1532-4435.

Khan, S., Husa, S., Hannam, M., Ohme, F., Pürrer, M., Jiménez Forteza, X., and Bohé, A. Frequency-domain gravitational waves from nonprecessing black-hole binaries. II. A phenomenological model for the advanced detector era. *Phys. Rev. D*, 93(4):044007, 2016. doi: 10.1103/PhysRevD.93.044007.

Legin, R., Adam, A., Hezaveh, Y., and Perreault Levasseur, L. Beyond Gaussian Noise: A Generalized Approach to Likelihood Analysis with non-Gaussian Noise. *arXiv e-prints*, art. arXiv:2302.03046, February 2023. doi: 10.48550/arXiv.2302.03046.

Punturo, M. et al. The Einstein Telescope: A third-generation gravitational wave observatory. *Class. Quant. Grav.*, 27:194002, 2010. doi: 10.1088/0264-9381/27/19/194002.

Romero-Shaw, I. M. et al. Bayesian inference for compact binary coalescences with bilby: validation and application to the first LIGO–Virgo gravitational-wave transient catalogue. *Mon. Not. Roy. Astron. Soc.*, 499(3):3295–3319, 2020. doi: 10.1093/mnras/staa2850.

Song, Y., Sohl-Dickstein, J., Kingma, D. P., Kumar, A., Ermon, S., and Poole, B. Score-Based Generative Modeling through Stochastic Differential Equations. *arXiv e-prints*, art. arXiv:2011.13456, November 2020. doi: 10.48550/arXiv.2011.13456.

Veitch, J. et al. Parameter estimation for compact binaries with ground-based gravitational-wave observations using the LALInference software library. *Phys. Rev. D*, 91(4): 042003, 2015. doi: 10.1103/PhysRevD.91.042003.

Vincent, P. A connection between score matching and denoising autoencoders. *Neural Comput.*, 23(7):1661–1674, 2011. doi: 10.1162/NECO_a_00142. URL https://doi.org/10.1162/NECO_a_00142.

Wong, K. W. k., Gabrié, M., and Foreman-Mackey, D. flowMC: Normalizing flow enhanced sampling package for probabilistic inference in JAX. *J. Open Source Softw.*, 8(83):5021, 2023a. doi: 10.21105/joss.05021.

Wong, K. W. K., Isi, M., and Edwards, T. D. P. Fast gravitational wave parameter estimation without compromises. 2 2023b.