Inferring parameters for binary black hole mergers using normalizing flows

Stephen R. Green (Albert Einstein Institute Potsdam)

based on arXiv:2008.03312 with Jonathan Gair

[https://www.github.com/stephengreen/lfi-gw]

Third Workshop on Machine Learning and the Physical Sciences (NeurIPS 2020), Vancouver, Canada

Summary:

We train a neural conditional density estimator to perform fast Bayesian inference for gravitational waves. Using normalizing flows, we learn posterior probability distributions over the full 15D space of binary black hole system parameters, given detector strain data from multiple detectors. We apply the method to the first gravitational-wave detection, GW150914, obtaining results consistent with standard sampling algorithms.

This is the first demonstration that deep-learning can be to infer all 15 binary black hole parameters from real gravitational-wave strain data, including accurate estimates of uncertainties.

Context:

Since the first discovery in 2015, the LIGO and Virgo gravitational-wave observatories have published details of **50** compact binary mergers, all of which have been analyzed using Bayesian inference. Using standard methods, this is computationally expensive and time consuming, requiring many waveform model evaluations.

Need for new and faster approaches:

- Rapid multi-messenger follow up
- Higher event rate with improved detectors
- Enable use of waveform models with more physics

Normalizing flows:

Model the Bayesian posterior $p(\theta\,|\,s)$, for system parameters θ given strain data s, using a neural conditional density estimator $q(\theta\,|\,s)$.

A normalizing flow f_s is an invertible mapping on a sample space with simple Jacobian determinant. Define

$$q(\theta\,|\,s) \equiv \pi(f_s^{-1}(\theta)) \left| \det J_{f_s}^{-1} \right|$$

where $\pi = \mathcal{N}(0,1)^D$. Since $\pi(u)$ is easy to evaluate and sample so is $q(\theta\,|\,s)$.

- Use a neural spline coupling flow to define a sufficiently flexible $q(\theta\,|\,s)$. [Durkan *et al*, 2019]

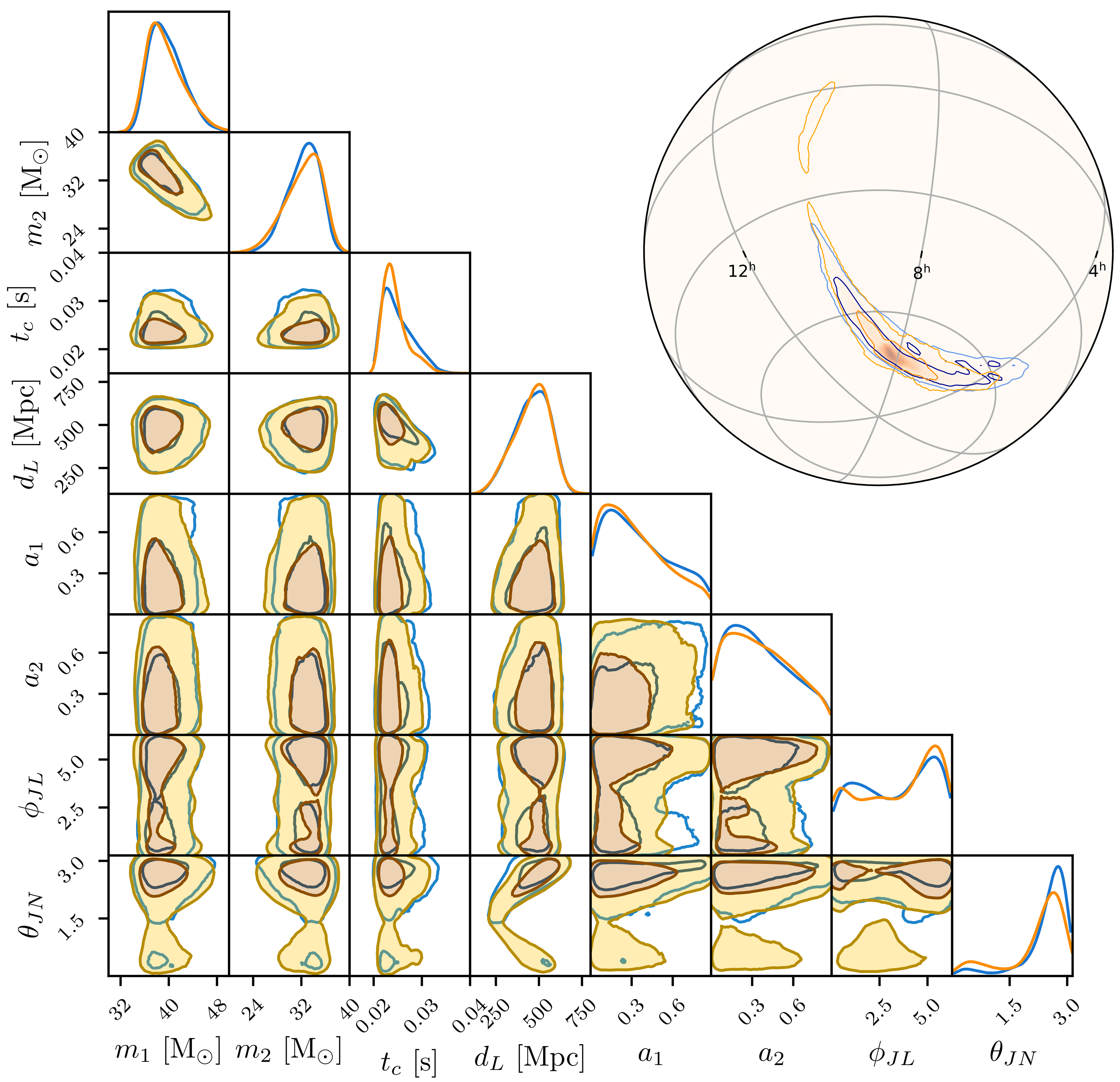

Figure: Comparison of posteriors for GW150914 generated by the neural network (orange) and the standard bilby dynesty sampler (blue). 50% and 90% credible regions indicated.

Parameter	Description	Prior	Extrinsic
(m_1, m_2)	component masses	$[10\ \mathrm{M}_\odot, 80\ \mathrm{M}_\odot]$, $m_1 \geq m_2$	No
ϕ_c	reference phase	$[0, 2\pi]$	No
$t_{c,\mathrm{geocent}}$	time of coalescence	$[-0.1\ \mathrm{s}, 0.1\ \mathrm{s}]$	Yes
d_L	luminosity distance	$[100\ \mathrm{Mpc}, 1000\ \mathrm{Mpc}]$	Yes
(a_1, a_2)	dimensionless spin magnitudes	$[0, 0.88]$	No
$(\theta_1, \theta_2, \phi_{12}, \phi_{JL})$	spin angles	standard	No
θ_{JN}	inclination relative to line-of-sight	$[0, \pi]$, uniform in sine	No
ψ	polarization angle	$[0, \pi]$	Yes
(α, δ)	sky position	uniform over sky	Yes

Table: Binary black hole system parameters θ. Training set consists of 10^6 elements, with 10% reserved for validation. Parameters labelled "extrinsic" were chosen at train time, as were noise realizations, effectively enlarging the training set.

Likelihood-free training:

Train $q(\theta\,|\,s) \longrightarrow p(\theta\,|\,s)$. Requires simulated data, but no likelihood evaluations or posterior samples:

$$L = -\int \mathrm{d}s\mathrm{d}\theta\, p(\theta, s) \log q(\theta|s)$$

$$\approx -\frac{1}{N}\sum_{i=1}^{N} \log q\left(\theta^{(i)}|s^{(i)}\right) \qquad \text{(Monte Carlo approximation)}$$

$$\text{where } \theta^{(i)} \sim p(\theta) \text{ and } s^{(i)} \sim p(s|\theta^{(i)})$$

1. Sample prior, $\theta^{(i)} \sim p(\theta)$.
2. Simulate a waveform, $h^{(i)} = h(\theta^{(i)})$.
3. Add noise, $s^{(i)} = h^{(i)} + n^{(i)}$, where $n^{(i)} \sim p_{S_n}(n)$. Detector noise is assumed stationary Gaussian, with power spectral density estimated prior to event.
4. Evaluate $q(\theta^{(i)}\,|\,s^{(i)})$, and minimize L.

- IMRPhenomPv2 precessing waveform model.
- Waveforms compressed using a singular value decomposition.

- **Flow details:**
  - 15 coupling transforms, with rational-quadratic spline functions.
  - Each coupling transform defined by fully-connected residual network, with 10 blocks of two 512-unit hidden layers.

- **Training details:** 500 epochs @ batch size 512, Adam optimizer.

Conclusions:

- We performed accurate parameter estimation on GW150914 strain data from multiple detectors in the full 15D space.
- Network learns global set of posteriors $p(\theta\,|\,s)$ for all strain data consistent with training distribution. We evaluated performace across parameter space using a P−P plot test.
- Trained network generates 5,000 posterior samples per second.

Next steps:

- Condition flow also on detector noise characteristics, which vary slightly from event to event. This would allow to fully amortize training time over many detections.
- Extend to treat longer waveforms (e.g., binary neutron stars)
- Move beyond idealization of stationary Gaussian noise.