Use of conditional variational auto encoder to analyze ringdown gravitational waves

Takahiro S. Yamamoto\textsuperscript{1} and Takahiro Tanaka\textsuperscript{1,2}

\textsuperscript{1}Department of Physics, Kyoto University, Kyoto 606-8502, Japan
\textsuperscript{2}Center for Gravitational Physics, Yukawa Institute for Theoretical Physics, Kyoto University, Kyoto 606-8502, Japan

(Dated: February 28, 2020)

Recently, several deep learning methods are proposed for the gravitational wave data analysis. One is conditional variational auto encoder (CVAE), proposed by Gabbar et al.\textsuperscript{11}. We study the accuracy of a CVAE in the context of the estimation of the QNM frequency of the ringdown. We show that the accuracy of the estimation by the CVAE is better than the matched filtering. The areas of confidence regions are also compared and it is shown that the CVAE can return smaller confidence regions. Also, we assess the reliability of the confidence regions estimated by the CVAE. Our work confirms that the deep learning method has ability to compete with or overcome the matched filtering.

I. INTRODUCTION

In 2015, Laser Interferometer Gravitational-Wave Observatory (LIGO) detected gravitational waves from a binary black hole (BBH) merger\textsuperscript{2}. In Observation run 1 and 2, ten BBH merger events were confirmed\textsuperscript{3}. Currently, advanced LIGO and advanced Virgo are operating and KAGRA will join this detector network in 2020\textsuperscript{4}. Besides the improvement of detectors, the improvement of data analysis methods can contribute to accelerate the gravitational wave physics and astronomy.

Recently, the use of deep learning methods is proposed for various purposes, e.g., the detection of gravitational waves\textsuperscript{5,6}, the parameter estimation\textsuperscript{1,7,8}, the noise subtraction\textsuperscript{9,10}, and the classification of glitch noises\textsuperscript{11}. Our work is devoted to investigating the accuracy of the parameter estimation. Our question is how accurately deep learning methods can estimate physical parameters, or whether deep learning methods can estimate parameters more accurately than the standard method.

In this paper, we focus on the analysis of ringdown gravitational waves. The ringdown is the last stage of a BBH merger. The remnant black hole is largely perturbed just after the merger and the perturbation decays as gravitational waves are emitted. Time lapse perturbations of the black hole is dominated by the black hole quasi-normal modes (QNMs). The ringdown gravitational waves can be modeled by the damped sinusoidal waveforms having the complex-valued QNM frequencies predicted by the black hole perturbation theory\textsuperscript{12,13}. In general relativity (GR), the QNM frequencies are determined by the black hole mass and spin. Because of this property, the ringdown gravitational waves are useful for the test of GR\textsuperscript{14,15}. One way to estimate the QNM frequencies is the matched filtering using the inspiral-merger-ringdown gravitational waves\textsuperscript{17,18}. The posterior distribution of the binary masses and the spins is estimated and it can be converted into the mass and the spin of the remnant black hole by the fitting formula obtained from numerical relativity simulations\textsuperscript{19}. This method relies on GR and the inference of these parameters is mainly governed by the inspiral part. If the effects caused by exotic theories (e.g. modified theories of gravity, black hole mimickers) modify the merger-ringdown part without changing the inspiral part, bias would be introduced in the posterior in this method. Thus, we need a method to estimate QNM frequencies using only the merger-ringdown part.

There are two possible directions of investigation: improving the matched filtering and implementing alternative methods. In Ref.\textsuperscript{20}, comparison of various methods for the analysis of ringdown was done using test mock data. The result shows that the deep learning method is competitive with the matched filtering. The deep learning method used in this challenge was the one constructed for the point estimation, that is, the neural network returns only a single estimated value for each parameter that we want to estimate. Despite of this shortcoming, deep learning methods are still expected to be a useful method complementary to the matched filtering.

Recently, the authors of Ref.\textsuperscript{1} proposed the use of the conditional variational auto encoder (CVAE) for gravitational data analysis. In addition to that the computational speed of the CVAE is much faster than that of the matched filtering, the CVAE can estimate the posterior probability distributions of parameters. Although the purpose of Ref.\textsuperscript{1} was the rapid inference, we apply the CVAE for the off-line analysis and assess the accuracy of the inference of the CVAE.

This paper is organized as follows. In Sec.\textsuperscript{II} we present the construction of the waveforms. In Sec.\textsuperscript{III} we briefly review the matched filtering. In Sec.\textsuperscript{IV} the idea and the implementation of CVAE are explained. In Sec.\textsuperscript{V} we introduce convolutional neural networks (CNNs) as another competitors to the CVAE. In Sec.\textsuperscript{VI} the results obtained by the CVAE are compared with the matched filtering and the CNN. We focus on the accuracy of the maximum posterior estimations and the area of the confidence regions. We also confirm that the confidence regions obtained by the CVAE have the frequentist
meaning by making the P-P plot, with evaluation of the magnitude of the error. We summarize our results and future works in Sec. VII. Throughout this paper, we set $G = c = 1$.

II. PREPARING MOCK TEMPLATES

As explained in Introduction, the situation we consider is that only the merger-ringdown part is modified from that of GR, and we compare deep learning methods and the matched filtering in such a situation. For this purpose, we need to generate a test dataset by modifying only the merger-ringdown part of the waveform. In some modified theories of gravity, gravitational waves from inspiraling BBHs can be calculated in the post-Newtonian approximation. But consistent simulations throughout the inspiral-merger-ringdown phases have not been done so far. In addition, it is a highly speculative assumption that only the merger-ringdown part might be modified. Therefore, what we can do for generating modified templates is to modify the merger-ringdown phase of GR templates in a phenomenological manner. Using the modified templates, we prepare a mock test data for comparison of the deep learning methods and the matched filtering. These templates are used not only for preparing a test dataset, but also for training neural networks and for constructing the template bank of the matched filtering.

The precise modeling of the transition from the inspiral phase to the post-merger phase is difficult, but we would be able to roughly assume that the gravitational waves of the merger-ringdown phase have the following properties,

- The amplitude after the peak monotonically decreases. At a later time, the amplitude decays exponentially.
- The frequency monotonically increases and converges to a certain QNM frequency at a later time.

We focus on the case where the waveforms are modified only after the time $t_p$, at which the amplitude of GR template reaches its peak. Therefore, the inspiral part of the modified waveform coincides with GR one. In this work, we focus only on $l = m = 2$ mode and ignore overtones as they are much weaker especially for nearly equal-mass binaries. The importance of the multi-modes and overtones has been studied in Refs. [21,22].

We denote the QNM frequencies for GR templates and for modified templates by $\omega_{R,1}$ and $\omega_{R,1}$, respectively. The modified templates are constructed by modifying the complex-valued templates in GR, $h^{GR}(t)$. First, we decompose the strain $h^{GR}(t)$ into the amplitude $A^{GR}(t)$ and the frequency $\omega^{GR}(t)$ as

$$h^{GR}(t) = A^{GR}(t) e^{i \phi^{GR}(t)}, \quad \phi^{GR}(t) = \int_0^t dt' \omega^{GR}(t').$$

From $A^{GR}(t)$ and $\omega^{GR}(t)$, the modified amplitude and frequency, $A(t)$ and $\omega(t)$, are generated. Our modified templates are characterized by two parameters, $\delta_\omega$ and $\delta_\omega$. The real and imaginary parts of the QNM frequency, $\omega_R$ and $\omega_I$, are specified by the fractional deviation from the GR values as

$$\omega_{R,1} = \omega_{R,1}(1 + \delta_R).$$

In our work, the modifications of the frequencies are assumed to be small. The deviations of the real part and the imaginary part of QNM frequencies are assumed to be less than 30% and 50%, respectively (i.e. $|\delta_R| < 0.30$, $|\delta_I| < 0.50$).

Modified amplitudes are constructed from two parts, before and after the peak. After the peak, the amplitudes are modified from GR as

$$A'(t) = \frac{A^{GR}(t)}{1 + e^{4M\omega^{GR}_I x}} + \frac{A^{RD}(t)}{1 + e^{-4M\omega^{GR}_I x}},$$

with

$$A^{RD}(t) = \frac{1.18}{1 + e^{-M\omega^{GR}_I x} + e^{M\omega x}}.$$ (2.4)

where $M$ is the total mass of the binary, $x$ is the normalized time defined as $x := (t - t^{GR}_p)/M$, and $t^{GR}_p$ is the time when the GR amplitude $A^{GR}(t)$ reaches its peak. The time when the modified amplitude $A'(t)$ reaches its maximum is denoted by $t'_p$ and can differ from $t^{GR}_p$. We connect the GR amplitude before $t^{GR}_p$ and the modified amplitude after $t'_p$ with an appropriate normalization. Namely, the modified amplitude $A(t)$ is obtained as

$$A(t) = \begin{cases} 
A^{GR}(t) & (t \leq t^{GR}_p), \\
\alpha A'(t + t'_p - t^{GR}_p) & (t > t^{GR}_p), 
\end{cases}$$

with $\alpha := A^{GR}(t^{GR}_p)/A'(t'_p)$.

The GW frequency $\omega(t)$ of the modified waveform is specified as

$$\omega(t) = \frac{\omega^{GR}(t)}{1 + e^{4M\omega^{GR}_I x}} + \frac{\omega^{RD}(t)}{1 + e^{-4M\omega^{GR}_I x}},$$

with

$$\omega^{RD}(t) = \omega^{GR}_p + (\omega_R^{GR} - \omega^{GR}_p) \tanh(0.85M\omega^{GR}_I x),$$ (2.7)

and $\omega^{GR}_p := \omega^{GR}(t^{GR}_p)$.

Finally, we generate the gravitational wave strain, $h(t)$, by

$$h(t) = A(t)e^{i\phi(t)}, \quad \phi(t) = \int_0^t dt' \omega(t').$$ (2.8)

The waveform of the modified model having $\delta_R = \delta_I = 0$ coincide with that of GR.

As a seed for modified templates, we use the waveform SXS:0305 [24] and the total mass is fixed to $M =$
72.158M⊙. The GR values of QNM frequency is calculated from the fitting formula in Ref. [15]. Examples of the modified templates are shown in Fig. 1.

In the following analysis, the frequency $f$ is used rather than $\omega$. They are related with each other by $\omega_{R,1} = 2\pi f_{R,1}$. The sampling rate is 4096Hz.

III. MATCHED FILTERING

When the waveforms can be theoretically modeled and generated rapidly, the matched filtering is a powerful method for the parameter estimation (see [25] as a standard textbook). The detection statistic is the signal-to-noise ratio (SNR) and it can be calculated by the noise-weighted inner product between the observational data $s(t)$ and a template $h(t)$,

$$\text{SNR} = 4\text{Re} \int_{f_{\text{min}}}^{f_{\text{max}}} df \frac{\tilde{s}(f)\tilde{h}^*(f)}{S_n(f)},$$  

(3.1)

where $S_n(f)$ is the noise power spectral density, $\tilde{s}(f)$ and $\tilde{h}(f)$ are the Fourier transforms of $s(t)$ and $h(t)$, respectively. We use the LIGO O1 noise power spectral density,

$$S_n(f) = 10^{-44} \times \left( \frac{18.0}{0.1 + f} \right)^4 + 0.49 \times 10^{-46} \times \left( \frac{f}{2000.0} \right)^2 \times 16.0 \times 10^{-46}[\text{strain}^2/\text{Hz}],$$  

(3.2)

given in Ref. [26].

We do not optimize the coalescence time in the present matched filtering analysis. Instead, we fix it to the value of the injected templates, assuming that it can be easily guessed from the inspiral part of the gravitational wave data. Therefore, our templates are parameterized by the deviation of the QNM frequency, $\{\delta_R, \delta_I\}$, and the initial phase, $\phi_0$. Since the initial phase can be marginalized analytically, the parameter search is done on the parameter space of $\{\delta_R, \delta_I\}$. With the uniform prior, the posterior distribution of the real and imaginary parts of the QNM frequency $\{f_R, f_I\}$ can be obtained by

$$p(f_R, f_I|s) \propto \exp \left[ \frac{\text{SNR}^2(\delta_R, \delta_I)}{2} \right].$$  

(3.3)

For the post-merger analysis, we set the boundaries of the integration range of frequency to $f_{\text{min}} = 160\,\text{Hz}$ and $f_{\text{max}} = 512\,\text{Hz}$. The lower cutoff frequency, $f_{\text{min}}$, is the frequency at which the amplitude of the template reaches the maximum.

In our work, the template bank is constructed to form a uniform grid in the $(\delta_R, \delta_I)$ plane. The parameter $\delta_R$ is varied in the range $[0.7, 1.3]$ with the step of 0.006, while $\delta_I$ in the range $[0.5, 1.5]$ with the step of 0.01. The template bank consists of 10,201 templates.

IV. CONDITIONAL VARIATIONAL AUTO ENCODER

A. Idea of CVAE

In this subsection, we explain the idea of CVAE [1].

In Bayesian inference, the existence of the true posterior $p(y|x)$, the distribution of the physical parameters $y$ under the assumption that a signal $x$ is given, is assumed. Here, the parameterized distributions $p_d(y|x)$ are used as an approximation of $p(y|x)$. The parameter $\theta$ depends on the input signal $x$. The neural network is trained to estimate the relation between $x$ and $\theta$ using a training dataset, that is, a lot of pairs of input data and the true values of the physical parameters, $\{(x_i, y_i)\}_{i=1...N}$. The Kullback-Leibler (KL) divergence,

$$KL[p(y|x)|p_d(y|x)] := \int dy \hat{p}(y|x) \log \frac{\hat{p}(y|x)}{p_d(y|x)},$$  

(4.1)

is one of the natural choices for quantifying the mismatch between two probability distributions. Here, we consider
the minimization of the expected value of the KL divergence,

$$E_{\hat{p}(x)} [KL(\hat{p}(y|x) || p_\theta(y|x))].$$ (4.2)

Because only the terms including $p_\theta(y|x)$ are essential for optimization, the minimization of (4.2) is equivalent to the maximization of the average of the cross entropy:

$$E_{\hat{p}(x)} [H(\hat{p}(y|x) || p_\theta(y|x))] := \int dx dy \hat{p}(x,y) \log p_\theta(y|x)$$

$$= \int dx dy \hat{p}(x,y) \log p_\theta(y|x).$$ (4.3)

This can be approximated by the sample mean,

$$E_{\hat{p}(x)} [H(\hat{p}(y|x) || p_\theta(y|x))] \simeq \frac{1}{N} \sum_{i=1}^{N} \log p_\theta(y_i|x_i).$$ (4.4)

For example, Gaussian distribution can be used as $p_\theta(y|x)$. However, it would be too simple to approximate the posterior. In order to enhance the flexibility of the approximant, the hidden variable model is often employed. The approximated distributions are given as a superposition of simple distributions,

$$p_\theta(y|x) = \int dz \ p_{\theta_0}(y|x,z)p_{q_\phi}(z|x).$$ (4.5)

The additional variables $z$, so-called hidden variables, inherit compressed information of the data $x$. With the hidden variable model, $\log p_\theta(y|x)$ appeared in R.H.S of Eq. (4.4) is bounded by the evidence lower bound (ELBO),

$$\log p_\theta(y|x) \geq \mathcal{L}_{ELBO} := E_{q_\phi(z|x,y)}[\log p_{\theta_0}(y|x,z)] - KL[q_\phi(z|x,y) || p_{q_\phi}(z|x)]$$ (4.6)

for an arbitrary distribution $q_\phi(z|x,y)$. The negative ELBO, $-\mathcal{L}_{ELBO}$, is employed as the loss function to be minimized.

A CVAE estimates the relation between the parameters of distributions and the conditioning variables. As an example, the distribution $p_{q_\theta}(z|x)$ presents the probability of $z$ conditioned by $x$. The neural network corresponding to $p_{q_\theta}(z|x)$ takes $x$ as an input and predicts the plausible value of $\theta_\phi$. In Eq. (4.6), three distributions, $p_{\theta_0}$, $p_{q_\phi}$ and $q_\phi$, appear. Therefore, we need three networks for emulating these distributions.

Further simplification of Eq. (4.6) can be done as follows. First, the first term of the R.H.S of Eq. (4.6) can be approximated by the sample average,

$$E_{q_\phi(z|x,y)} \log p_{\theta_0}(y|x,z) \simeq \frac{1}{N_k} \sum_{j=1}^{N} \log p_{\theta_0}(y|x,z_j),$$ (4.7)

where $z_j$ is the $j$-th sample of $z$ following $q_\phi(z|x,y)$. In this work, we set $N_k = 1$. Second, we adopt multivariate Gaussian distributions with diagonal covariance matrices as $p_{\theta_0}, p_{q_\theta}$ and $q_\phi$. We denote the mean and covariance matrix of $p_{q_\theta}(z|x)$ by

$$\vec{\mu}_E = (\mu_{E,1}, \mu_{E,2}, \ldots, \mu_{E,D_x})$$

$$\Sigma_E = \text{diag}(\sigma_{E,1}^2, \sigma_{E,2}^2, \ldots, \sigma_{E,D_x}^2),$$ (4.8a)

those of $p_{\theta_0}(y|x,z)$ by

$$\vec{\mu}_D = (\mu_{D,1}, \mu_{D,2}, \ldots, \mu_{D,D_x})$$

$$\Sigma_D = \text{diag}(\sigma_{D,1}^2, \sigma_{D,2}^2, \ldots, \sigma_{D,D_x}^2),$$ (4.8c)

and those of $q_\phi(z|x,y)$ by

$$\vec{\mu} = (\mu_1, \mu_2, \ldots, \mu_{D_x})$$

$$\Sigma = \text{diag}(\sigma_1^2, \sigma_2^2, \ldots, \sigma_{D_x}^2),$$ (4.8f)

where $D_x$ and $D_y$ are the dimensions of the hidden variable $z$ and the physical parameters $y$, respectively. Thus, the parameters $\theta_E, \theta_D$ and $\phi$ denoted abstractly so far are $\theta_E = \{\vec{\mu}_E, \Sigma_E\}, \theta_D = \{\vec{\mu}_D, \Sigma_D\}$ and $\phi = \{\vec{\mu}, \Sigma\}$. Then, the loss function for one training data is obtained as

$$\text{Loss} = \frac{D_y}{2} \log 2\pi + \sum_{i=1}^{D_y} \log \sigma_{D,i} + \frac{1}{2} \sum_{i=1}^{D_y} \frac{(y_i - \mu_{D,i})^2}{\sigma_{D,i}^2} + \frac{D_x}{2} - \frac{1}{2} \sum_{k=1}^{D_x} \left\{ \log \frac{\sigma_{E,k}^2}{\sigma_{\hat{\phi}}^2} + \frac{(\mu_k - \mu_{E,k})^2}{\sigma_{E,k}^2} + \frac{\sigma_{\phi,k}^2}{\sigma_{E,k}^2} \right\}.$$ (4.9)

Figure 2 shows the schematic picture of the CVAE we use in this work. The neural networks corresponding to $p_{q_\theta}(z|x)$, $q_\phi(z|x,y)$ and $p_{\theta_0}(y|x,z)$ are called Encoder1, Encoder2 and Decoder, respectively. Each neural network returns the mean and the diagonal elements of the covariance matrices of each distribution. At the training (the left figure of Fig. 2), all networks are simultaneously trained with the loss function (4.9). When the trained the CVAE is applied to a test data (the right figure of Fig. 2), we use the networks corresponding to $p_{\theta_0}$ and $p_{q_\theta}$ for estimating a posterior. Estimating the posterior for a test event is based on the following sampling method. First we sample one value of $z$ from the distribution $p_{q_\theta}(z|x)$. Next, with the sampled $z$, a sample of
FIG. 2. The schematic picture of the CVAE. Encoder1, Encoder2 and Decoder represent neural networks corresponding to the probability distributions $p_{\theta_E}(z|x)$, $q_{\phi}(z|x,y)$ and $p_{\theta_D}(y|x,z)$, respectively. Here, we adopt multivariate Gaussian distributions for all three distributions. The parameters characterizing these distributions are $\theta_D = \{\mu_D, \Sigma_D\}$, $\theta_E = \{\mu_E, \Sigma_E\}$ and $\phi = \{\mu, \Sigma\}$.

At the training (left), three networks are optimized so that the loss function is minimized. The Kulback-Leibler divergence are calculated with the output of the Encoder1 and the Encoder 2. The output of the Decoder is used for assessing the negative log posterior term. For test events (right), the Encoder 1 and the Decoder are employed for sampling predicted values.

the parameter $y$ is obtained from $p_{\theta_D}(y|z, x)$. Repeating these sampling processes, we finally obtain many samples of predicted values $y$ that follow the estimated posterior $p_\theta(y|x)$.

**B. Implementation**

In this subsection, the implementation of the CVAE that we use is described. We use PyTorch [27] for the implementation.

1. **Structure**

As explained in the subsection [V A], the CVAE consists of three neural networks, that is, two encoders and one decoder. Each of them has six layers and each internal layer has 512 units. We put a ReLU layer after each fully-connected layer except for the last layer of each neural network. Encoder1 and Encoder2 will output the mean and the diagonal elements of the covariance matrix of the hidden variables. We set the dimension of the hidden variables as $D_z = 16$. The input of Decoder is the sampled variables from the multi-variate Gaussian distribution having the mean and covariance matrix estimated by the encoder. Decoder returns the mean and the covariance matrix of the distribution $p_{\theta_D}(y|z, x)$. The entire structure of the CVAE we use in this work is shown in Table I.

2. **Dataset for training**

For the training, we use the same templates contained in the template bank for the matched filtering. Each template is labeled by $\{\delta_R, \delta_I\}$. The input signals as training data are generated as

$$x(t) = Ah_{\text{whitened}}(t) + n(t) \quad (4.10)$$

| Network | \# of units of respective layers |
|---------|---------------------------------|
| Encoder1 | [128, 512, 512, 512, 512, 512, 32] |
| Encoder2 | [130, 512, 512, 512, 512, 512, 32] |
| Decoder | [144, 512, 512, 512, 512, 512, 4] |
TABLE II. The training schedule for the CVAE. In the last stage of training, input signals have SNR varying from 8 to 30. After 45000 epochs, the training is terminated when decreasing of training loss saturates.

| epoch     | the range of $A$ | learning rate |
|-----------|------------------|---------------|
| 1 - 1000  | [8.0, 10.0]      | $1.0 \times 10^{-5}$ |
| 10001 - 15000 | [6.0, 10.0]     | $1.0 \times 10^{-5}$ |
| 15001 - 20000 | [4.0, 10.0]     | $1.0 \times 10^{-5}$ |
| 20001 - 25000 | [3.0, 10.0]     | $1.0 \times 10^{-5}$ |
| 25001 - 45000 | [2.0, 10.0]     | $1.0 \times 10^{-5}$ |
| 45001 -    | [2.0, 10.0]     | $1.0 \times 10^{-6}$ |

where $h_{\text{whitened}}$ is a template whitened with Eq. (3.2), the noise $n(t)$ is generated from the standard normal distribution, and the amplitude $A$ is chosen to realize a specified SNR. To prevent overfitting to a specific noise pattern, the noise realizations are generated and the whitened templates are injected into them for each iteration. From these simulated signals, we pick up 128 points starting from the amplitude peak, which is used as the input data of the CVAE.

3. Training and inference scheme

The Adam procedure [28] is used for the optimization algorithm. The learning rate is set to $10^{-5}$ initially and decreased to $10^{-6}$ on the later stage of the training. The scheduled training is employed, i.e., the amplitude of the signal is gradually decreased from a large initial amplitude. The training schedule is shown in Table II. The batch size is 256.

When the trained CVAE is applied to a test data, the sampling process to estimate the distribution is repeated until $4 \times 10^6$ samples are collected.

V. CONVOLUTIONAL NEURAL NETWORK

In this work, an ordinary neural network, which returns a single value for each parameter that we want to estimate, is also implemented as one of competitors to the CVAE. Convolutional neural networks (CNNs) are used for various research of the gravitational wave data analysis (e.g. [5]). Our CNN has three convolutional and four fully-connected layers. Each of them, except for the last layer, is followed by a ReLU layer. The output of the last layer is the estimated values of $\{\delta_R, \delta_I\}$. For respective convolutional layers, the numbers of filters are 128, 256 and 512, and the sizes of filters are 32, 8 and 8. All of fully connected layers have 512 units. We use mean square loss for the loss function. Also for the training of the CNN, scheduled training is employed. The training schedule is shown in Table III. The CNN is also implemented by PyTorch. The training dataset is the same as the CVAE.

VI. RESULTS

A. Dataset for comparison

We prepare the mock test data in the same way as the training data. The real-valued template $h_{\text{inj}}$ is generated from a complex-valued modified template $h = h_+ + ih_\times$ with the randomly sampled phase $\phi_0$, i.e.,

$$h_{\text{inj}} = h_+ \cos \phi_0 + h_\times \sin \phi_0.$$  

We use the noise curve of LIGO O1 for generating the Gaussian noise (Eq. (3.2)). Three datasets with SNR of the merger-ringdown part 30.0, 15.0 and 8.0 are prepared (the definition of the merger-ringdown SNR is Eq. (3.1)). Each dataset consists of 500 simulated data whose $\delta_R$ and $\delta_I$ are randomly sampled from the region satisfying our assumptions, i.e., $|\delta_R| < 0.3$ and $|\delta_I| < 0.5$.

B. Comparison of the point estimation

To quantify the accuracy of the estimates, we define the following two quantities,

$$\Delta Q := \frac{1}{N_{\text{data}}} \sum_{i=1}^{N_{\text{data}}} (Q_{i}^{\text{est}} - Q_{i}^{\text{true}}),$$

$$\sigma(Q) := \frac{1}{N_{\text{data}}} \left[ \sum_{i=1}^{N_{\text{data}}} (Q_{i}^{\text{est}} - Q_{i}^{\text{true}})^2 \right]^{1/2}.$$  

Here, $Q_{i}^{\text{est}}$ is given by the estimated value that maximizes the posterior distribution for the matched filtering and the CVAE, while it is given by the output value for the CNN. The comparison of the errors is shown in Table IV. From this table, we can conclude that

- For both $f_R$ and $f_I$, the means of the errors $\Delta Q$ are much smaller than the standard deviations $\sigma(Q)$. Therefore, the estimates of both $f_R$ and $f_I$ are not significantly biased in all methods.
TABLE IV. The comparison of the estimation errors. The quantities \( \Delta Q \) and \( \sigma(Q) \) are defined in Eqs. (6.2) and (6.3). The estimation by the CVAE has no significant bias for both of \( f_R \) and \( f_I \) and for any values of SNR. The matched filtering and the CNN also estimate QNM frequency with small bias for most cases. Comparing the values of \( \sigma(f_{R,1}) \) for the CVAE, we find that the CVAE takes the smallest values for all cases, except for imaginary part of the dataset having SNR=8. For this case, the CVAE has a smaller value of \( \Delta f_I \) than the CVAE. However, the CNN derives a slightly larger value of \( \Delta f_I \) than the CVAE. This means that the estimation by the CNN is more biased.

| SNR RD | \( \Delta f_R \) [Hz] | \( \sigma(f_R) \) [Hz] | \( \Delta f_I \) [Hz] | \( \sigma(f_I) \) [Hz] |
|--------|----------------|----------------|----------------|----------------|
| 30.0   | MF -0.1607 3.5243 | -0.1805 2.7237 |            |            |
|        | CNN 0.9732 8.2192 | -1.1812 3.0875 |            |            |
|        | CVAE 0.0267 3.1180 | -0.2528 2.4311 |            |            |
| 15.0   | MF -0.4015 7.4448 | -0.5448 5.4256 |            |            |
|        | CNN -0.0432 9.5206 | -0.6411 4.9630 |            |            |
|        | CVAE -0.4253 6.2759 | -0.2109 4.8657 |            |            |
| 8.0    | MF -0.1755 15.2181 | -1.7824 9.6581 |            |            |
|        | CNN 0.9783 14.2067 | 1.7371 7.7085 |            |            |
|        | CVAE -0.2350 12.4485 | 0.4289 8.9368 |            |            |

Because the standard deviations of the CVAE are smaller than those of the matched filtering and the CNN, we can say that the CVAE estimates the QNM frequencies more accurately than the other two methods.

C. Reliability of the confidence regions

An example of the predictions of posterior distributions by the CVAE and the matched filtering is shown in Fig. 3. Before comparing the posterior estimations by the CVAE and the matched filtering, we assess the reliability of the posterior distributions estimated by the CVAE. If the estimation of posterior distribution is reliable, the fraction of events whose true values are located within the \( x\)-% confidence region should be \( x\)-%. For visualization, a P-P plot is useful. In a P-P plot, we take the confidence level as horizontal axis and the fraction of events as vertical axis. If the posterior distribution is reliable, the P-P plot reduces to the diagonal line. We show the P-P plots obtained by the CVAE and the matched filtering in Fig. 4. It is found that the error estimation by the matched filtering includes no significant bias. On the other hand, the P-P plot for the CVAE seems to deviate from the 45° line only slightly. In order to quantify the systematic error, we generate additional 9,500 test events for each SNR. Figure 5 shows the deviation from 45° line for SNR=8.0 events. It is found that the estimation by the CVAE contains the systematic error less than 2%. A similar feature can be seen for the events having SNR 15.0 and 30.0.

D. Comparison of areas of confidence regions

Taking into account the existence of bias at a few percent level, we compare the confidence regions obtained by the CVAE and the matched filtering. To compare them quantitatively, we define

\[
\Delta S_i(x) = S_i^{\text{CVAE}}(x) - S_i^{\text{MF}}(x), \quad \Delta S(x) = \frac{1}{N_{\text{data}}} \sum_{i=1}^{N_{\text{data}}} \Delta S_i(x),
\]

where \( S_i^{\text{CVAE/MF}}(x) \) is the area of the \( x\)-% confidence region estimated by the CVAE/matched filtering for the \( i\)-th test event. When \( \Delta S_i(x) \) is negative, the constraint of the CVAE is tighter than that of the matched filtering. The comparison of the area of the confidence region is shown in Table V. For all datasets, the CVAE leads to more stringent constraint than the matched filtering.

FIG. 3. An example of posterior estimations for a test data whose SNR is 8.0. Blue and orange contours are confidence regions estimated by the CVAE and the matched filtering, respectively. The contours show (50, 90, 99)% confidence regions. Blue circle and orange square are the predicted values of the QNM frequency obtained by the CVAE and the matched filtering, respectively. Black cross shows the injected value of the QNM frequency.

TABLE V. The comparison of the areas of confidence regions. The quantity \( \Delta S(x) \) is defined in Eq. (6.5). For all datasets having different SNRs, the CVAE gives tighter constraint than the matched filtering.

| SNR RD | \( \Delta S(99) \) [Hz²] | \( \Delta S(90) \) [Hz²] | \( \Delta S(50) \) [Hz²] |
|--------|----------------|----------------|----------------|
| 30.0   | -10.8893 6.6020 | -2.3531 5.4256 |
| 15.0   | -119.521 64.5984 | -20.1443 54.256 |
| 8.0    | -415.235 185.065 | -46.8837 54.256 |
FIG. 4. The P-P plots of the matched filtering (left) and of the CVAE (right). The SNR of the test dataset is 8.0. The horizontal axis shows percentages of the confidence regions. The vertical axis shows the fraction of events whose true values are located within the confidence regions. If estimated confidence region has the frequentist meaning, the plot (blue line) is consistent with the diagonal line (black dotted line). The orange region is 1-σ error of the binomial distribution. The error estimation by the CVAE seems to be slightly biased. A similar feature can be seen for the datasets having SNR 15.0 and 30.0.

FIG. 5. The deviation of the P-P plot from the diagonal line. The SNR of the dataset is 8.0. Blue circles and orange squares are obtained with 500 and 10,000 test events, respectively. The CVAE estimates the posterior distributions with < 2% systematic error.

VII. CONCLUSION

In this paper, we investigated how accurately a CVAE can estimate the QNM frequencies using only merger-ringdown waveforms. To do this, we generated modified waveforms by changing the merger-ringdown part of the GR template and constructed a test dataset by injecting the waveforms into simulated Gaussian noise data. We compared the accuracies of the CVAE and the matched filtering, and showed the CVAE can predict the QNM frequencies with a higher accuracy than the matched filtering. Next, we evaluated the reliability of the confidence regions estimated by the CVAE, making a P-P plot. The estimated confidence levels have the systematic error less than 2%. The areas of 50%, 90% and 99 % confidence regions obtained by the CVAE and the matched filtering were compared and it was found that the CVAE can give more stringent constraint to the QNM frequencies than the matched filtering.

In this work, we only focused on the case of the Gaussian noise. To make the deep learning method applicable to the real event analysis, the case with the noise having non-Gaussianity need to be investigated. The higher modes of the ringdown signal were also neglected. The importance of the multi-mode analysis is indicated by several authors [21, 22]. Application to the black hole spectroscopy is remaining for future work.

CVAE is not the only method for estimating posteriors (e.g. Bayesian neural network [7], NN with reduced order modeling [8]). Comparison (or integration) with these methods would be insightful.

In this work, the merger-ringdown waveforms modified from those of GR were employed for training the CVAE. In this sense, our method is model-dependent. Although the post-merger templates based on the specific theory of modified gravity are not obtained so far, the result of our work is insightful when they can be constructed. On
the other hand, exploring model independent methods is a possible direction of future work. Even in non-GR theories, the ringdown gravitational waves would be expected to have the properties that the frequency is constant and the amplitude decays exponentially. Neural networks would be useful to detect these features from noisy signals and estimate the QNM frequencies independently of the way of modification.

ACKNOWLEDGMENTS

This work was supported by JSPS KAKENHI Grant Number JP17H06358 (and also JP17H06357), A01: Testing gravity theories using gravitational waves, as a part of the innovative research area, “Gravitational wave physics and astronomy: Genesis”. We thank the members of the A01 group for useful discussions. Some part of calculation has been performed by using GeForce 2080Ti GPU at Nagaoka University of Technology.

[1] H. Gabbard, C. Messenger, I. S. Heng, F. Tonolini and R. Murray-Smith, arXiv:1909.06296
[2] B. P. Abbott et al. (LIGO Scientific and Virgo Collaborations), Phys. Rev. X 6, 041015 (2016)
[3] B. P. Abbott et al. (LIGO Scientific and Virgo Collaborations), Phys. Rev. X 9, 031040 (2019)
[4] T. Akutsu et al. (KAGRA Collaboration), Nature Astronomy 3, 35 (2019)
[5] D. George and E. A. Huerta, Phys. Rev. D97, 044039 (2018)
[6] P. Astone, P. Cerdá-Durán, I. Di Palma, M. Drago, F. Muciaccia, C. Palomba, and F. Ricci, Phys. Rev. D98, 122002 (2018)
[7] H. Shen, E. A. Huerta, Z. Zhao, E. Jennings and H. Sharma, arXiv:1903.01998 (2019)
[8] A. Chua and M. Vallinseri, arXiv:1909.05966 (2020)
[9] W. Wei and E. A. Huerta, Physics Letters B 800, 135081 (2020)
[10] H. Shen, D. George, E. A. Huerta and Z. Zhao, 2019 IEEE International Conference on Acoustics, Speech and Signal Processing (2019)
[11] D. George, H. Shen and E. A. Huerta, Phys. Rev. D97, 101501 (2018)
[12] T. Regge and J. A. Wheeler, Phys. Rev. 108, 1063 (1957)
[13] F. J. Zerilli, Phys. Rev. D2, 2141 (1970)
[14] S. A. Teukolsky, Astrophysical Journal 185, 635 (1973)
[15] E. Berti, V. Cardoso and C. M. Will, Phys. Rev. D73, 064030 (2006)
[16] B. P. Abbott et al, (LIGO Scientific and Virgo Collaborations) Phys. Rev. Lett. 116, 221101 (2016)
[17] A. Ghosh, A. Ghosh, N. K. Johnson-McDaniel, C. K. Mishra, P. Ajith, W. Del Pozzo, D. A. Nichols, Y. Chen, A. B. Nielsen, C. P. L. Berry and L. London, Phys. Rev. D94, 021101(R) (2016)
[18] A. Ghosh, N. K. Johnson-McDaniel, A. Ghosh, C. K. Mishra, P. Ajith, W. Del Pozzo, C. P. L. Berry, A. B. Nielsen and L. London, Classical and Quantum Gravity 35, 014002 (2017)
[19] J. Healy, C. O. Lousto and Y. Zlochower, Phys. Rev. D90, 104004 (2014)
[20] H. Nakano, T. Narikawa, K. Oohara, K. Sakai, H. Shinohara, H. Takahashi, T. Tanaka, N. Uchikata, S. Yamamoto and T. S. Yamamoto, Phys. Rev. D99, 124032 (2019)
[21] E. Berti, J. Cardoso, V. Cardoso and M. Cavaglià, Phys. Rev. D76, 104044 (2007)
[22] M. Cabero, J. Westerweck, C. D. Capano, S. Kumar, A. B. Nielsen and B. Krishnan, arXiv: 1911.01361 (2019)
[23] M. Giesler, M. Ili, M. A. Scheel and S. A. Teukolsky, Phys. Rev. X 9 041060 (2019)
[24] https://data.black-holes.org/waveforms/catalog.html
[25] J. D. E. Creighton and W. G. Anderson, Gravitational-Wave Physics and Astronomy, Wiley-VCH (2011)
[26] https://www.gw-openscience.org/GW150914data/LOSC_Event_tutorial_GW150914.html
[27] A. Paszke, S. Gross, S. Chintala, G. Chanan, E. Yang, Z. DeVito, Z. Lin, A. Desmaison, L. Antiga and A. Lerer, NIPS 2017 workshop (2017)
[28] D. P. Kingma and J. Ba, a conference paper at the 3rd International Conference for Learning Representations, San Diego (2015)