Total-variation methods for gravitational-wave denoising: Performance tests on Advanced LIGO data

Alejandro Torres-Forné,[1] Elena Cuoco,[2,3,4] Antonio Marquina,[5] José A. Font,[1,6] and José M. Ibáñez[1,6]

[1]*Departamento de Astronomía y Astrofísica, Universitat de València,*
Dr. Moliner 50, 46100 Burjassot (València), Spain
[2]*European Gravitational Observatory (EGO), Via E Amaldi, I-56021 Cascina, Italy*
[3]*Scuola Normale Superiore (SNS), Piazza dei Cavalieri, 7, 56126 Pisa PI, Italy*
[4]*Istituto Nazionale di Fisica Nucleare (INFN)*
Sez. Pisa Edificio C - Largo B. Pontecorvo 3, 56127 Pisa, Italy
[5]*Departamento de Matemática Aplicada, Universitat de València,*
Dr. Moliner 50, 46100 Burjassot (València), Spain
[6]*Observatori Astronòmic, Universitat de València,*
C/ Catedrático José Beltrán 2, 46980 Paterna (València), Spain

We assess total-variation methods to denoise gravitational-wave signals in real noise conditions by injecting numerical-relativity waveforms from core-collapse supernovae and binary black hole mergers in data from the first observing run of Advanced LIGO. This work is an extension of our previous investigation in which only Gaussian noise was used. Since the quality of the results depends on the regularization parameter of the model, we perform a heuristic search for the value that produces the best results. We discuss various approaches for the selection of this parameter, based on the optimal, mean, or multiple values, and compare the results of the denoising upon these choices. Moreover, we also present a machine-learning-informed approach to obtain the Lagrange multiplier of the method through an automatic search. Our results provide further evidence that total-variation methods can be useful in the field of gravitational-wave astronomy as a tool to remove noise.

I. INTRODUCTION

The observation of gravitational waves from coalescing binary black holes (BBHs) during the first Advanced LIGO [1] observing run (O1) marked the commencement of gravitational-wave astronomy [2,3]. After a period of commissioning, the two LIGO detectors started the second observing run (O2) by the end of 2016, with the European detector Advanced Virgo [4] joining on August 2017. O2 was an overwhelming success. In addition to the observation of three new BBH mergers [5–7], the latter simultaneously observed by the three-detector network, it also accomplished the first observation of gravitational waves from a binary neutron star (BNS) merger [8]. Unlike BBH events, BNS mergers emit electromagnetic signals across the entire spectrum. Those were detected by dozens of telescopes, opening the field of multimessenger astronomy [9].

During O2, the BNS observational range of Advanced LIGO was as large as ∼100 Mpc. However, due to their intrinsic weakness, signals from most astrophysical sources within such a large volume are likely to remain in the limit of detectability. A careful analysis of the collected gravitational-wave data is therefore essential to ensure progress in spite of the conspicuous instrumental noise of the detectors (see Ref. [10] and references therein). Actual gravitational-wave signals may be misinterpreted as artificial noise transients (glitches), which requires their precise identification and eventual veto.

Noise removal is a long-standing, major effort in gravitational-wave data analysis, and specific algorithms have been developed for every type of signal. For coalescing compact binary (CBC) signals, such as the existing sample of observations, the inspiral signal can be observed by either targeting a broad range of generic transient signals or by correlating the data with analytic waveform templates from general relativity and maximizing such correlation with respect to the waveform parameters [11,12]. Identification is challenged for events with a low signal-to-noise ratio (SNR) due to the nonstationarity and non-Gaussanity of the detector noise. Matched filtering is impractical for well-modeled but continuous sources, like spinning (isolated) neutron stars, due to the large computational resources it would require. Cross-correlation and coherent detection methods [10,13,14] are the choice for these sources. In addition to CBC and continuous sources, there also exist other potential sources that produce

084013-1

gravitational-wave transients (or bursts), with core-collapse supernova (CCSN) signals being a paradigmatic example. Those can only be modeled imperfectly, as the computational requirements for obtaining their wave-forms from numerical-relativity simulations are significant and the intrinsic parameter space is larger than for CBC signals. Recently, coherent approaches over a network of detectors have proven very effective [15,16], increasing the detection confidence of long-duration burst signals (above several seconds). In contrast, short-duration bursts are more affected by detector glitches, and specific pipelines have been developed to differentiate between signals and noise transients, namely, BayesWave [17], either standing alone or in combination with coherentWaveBurst [18], and oLIB [19]. Approaches to estimate physical parameters and to reconstruct burst signal waveforms from noisy environments have also been put forward by Refs. [20–27].

Methods based on machine learning offer a promising alternative to current approaches, having already shown optimal performance for many tasks, like classification and regression, and in many scientific disciplines [28]. Machine-learning techniques have been recently applied for gravitational-wave astronomy as well [29–33]. In a previous paper [34], we assessed total-variation (TV) algorithms for denoising gravitational-wave signals. TV methods are based on L_1-norm minimization and have been mainly employed in the context of image processing, in which they constitute the best approach to solve the so-called Rudin-Osher-Fatemi denoising model [35]. Our first investigation [34] was limited to denoising gravitational waves embedded in additive *Gaussian* noise. We showed that noise can be successfully removed with TV techniques, irrespective of the signal morphology or astrophysical origin. In the current paper, we take a further step in the assessment of TV methods for gravitational-wave astronomy, using actual noise from the detectors instead of the idealized nonwhite Gaussian noise employed in our previous study. To this aim, we inject numerically generated signals (from BBH mergers and CCSN) into the data collected by the Advanced LIGO detector during the O1 observing run that extended from September 12, 2015, to January 19, 2016. Our goal is to test whether TV methods can effectively reduce noise in the conditions found with real data.

The paper is organized as follows. In Sec. II, we summarize the mathematical framework on which TV methods are based. Section III briefly explains the whitening method we employ to remove noise lines and other artifacts and the waveform catalogs we use to test our algorithm. In Sec. IV, we discuss the determination of the regularization parameter that produces the best results for the sources considered. The main results of our study are presented in Sec. V. Finally, a summary is provided in Sec. VI.

II. OVERVIEW OF THE METHOD

TV methods are based on the concept of TV-norm regularization, introduced by Ref. [35] in 1992 as a procedure to clean noisy signals. Starting from the classical linear degradation model, $f = u + n$, in which a noisy signal f is built from a signal u and some additive noise n, we assume white Gaussian noise with zero mean and variance σ^2.

The variational approach to u about the noise, is based on the minimization of f (the observed signal) and some information about the noise is to prioritize signals u, through the minimization of $R(u)$ a convex energy, called regularizer, subject to the constraint that the square of the $\|f - u\|_{L_2}^2$ matches the variance of the noise, σ^2. Applying Tikhonov's theorem, the constrained variational problem can be written in general as an unconstrained minimization problem by introducing the positive Lagrange multiplier $\mu > 0$:

$$u = \underset{u}{\mathrm{argmin}}\left\{ R(u) + \frac{\mu}{2}\|f - u\|_{L_2}^2 \right\}. \qquad (2.1)$$

This energy has two main terms. $R(u)$ is called the *regularization term* that rules out the signals with large values of $R(u)$. The second term $\|f - u\|_{L_2}^2$ is called the *fidelity term* and controls the degree of similarity between the solution u and the noisy signal f by computing the square of the L_2-norm. Both terms are weighted by a Lagrange multiplier $\mu > 0$ so that when it has a small value the relative weight of the fidelity term is small and the solution is highly regularized. In contrast, when the value of μ is high, the solution is dominated by the fidelity term, and u is similar to f.

In this paper, we shall use as regularizing energy $R(u)$ either the L_1-norm of the signal (see Ref. [36]) or the L_1-norm of its gradient (see Ref. [35]), which favors *sparse* solutions, i.e., very few nonzero components of the solution or its gradient. In addition, the algorithm to find L_1-norm minimizers is extremely efficient despite the fact that this norm is not differentiable.

Rudin *et al.* in their pioneering paper [35] proposed the use of the L_1-norm of the gradient for the regularizing energy. This specific formulation of the variational problem (2.1) is called the ROF model (after the authors' names) and reads

$$u = \underset{u}{\mathrm{argmin}}\left\{ \mathrm{TV}(u) + \frac{\mu}{2}\|f - u\|_{L_2}^2 \right\}. \qquad (2.2)$$

Since the ROF model uses the TV norm, the solution is the only one with the *sparsest* gradient. Thus, the ROF model reduces noise by making the gradient of the signal *more sparse* and avoiding spurious oscillations (ringing) on the solution signal.

The associated Euler-Lagrange equation of the ROF model is given by

$$\nabla \cdot \frac{\nabla u}{|\nabla u|} + \mu(f - u) = 0. \qquad (2.3)$$

This equation becomes singular when $|\nabla u| = 0$. To avoid this, the following regularized TV norm was used in Ref. [34], and we will call it the regularized ROF (rROF) algorithm, which allows us to obtain an approximate solution of the ROF model by smoothing the total-variation energy. The TV functional is slightly perturbed as

$$\text{TV}_\beta(u) := \int \sqrt{|\nabla u|^2 + \beta}, \qquad (2.4)$$

where β is a small positive parameter. Therefore, the rROF model reads

$$u = \underset{u}{\text{argmin}} \left\{ \text{TV}_\beta(u) + \frac{\mu}{2} \|f - u\|_{\text{L}_2}^2 \right\}. \qquad (2.5)$$

Assuming homogeneous Neumann boundary conditions, Eq. (2.5) becomes a nondegenerate second-order nonlinear elliptic differential equation of which the solution is smooth (for details, see Sec. II in Ref. [34]).

In this paper, we use a regularized ROF algorithm by solving the associated Euler-Lagrange equation of an energy that includes a smoothed TV norm (see Ref. [34] for details). The novelty here consists of using the rROF algorithm as a building block of an iterative procedure, called the Bregman iterative procedure (see Ref. [37]), which runs in the scale space from the solution of the regularizer TV model, using a very small value of the Lagrange multiplier, to the processed signal. Roughly speaking, we first choose the regularization parameter μ equal to a constant value μ_0, which is smaller than the optimal value needed to obtain a denoised signal by direct application of the rROF algorithm. The value of μ_0 is kept fixed through all the iterations. Next, we compute u_1 by solving

$$u_1 = \underset{u}{\text{argmin}} \left\{ \text{TV}_\beta(u) + \frac{\mu_0}{2} \|f - u\|_{\text{L}_2}^2 \right\}, \qquad (2.6)$$

$$f = u_1 + v_1, \qquad (2.7)$$

where v_1 is the residual. Then, we again apply the rROF algorithm using the same μ_0 and taking as an input signal $f + v_1$ to obtain u_2. We thus have

$$f + v_1 = u_2 + v_2. \qquad (2.8)$$

Applying this procedure for an arbitrary number of times n, we obtain a sequence of signals u_n for $n = 1, \ldots$ such that

$$f + v_{n-1} = u_n + v_n. \qquad (2.9)$$

The iteration stops when some discrepancy principle is satisfied, i.e., when the square of the L_2-norm of the residual matches the variance of the noise. In practice, however, the variance of the noise is not available, and

TABLE I. Values of the SSIM index for CCSN waveforms when using the optimal value of the regularization parameter for each signal, μ_{opt}; the mean value for all signals, $\bar{\mu}$; and multiple values, μ_{m}. The final column, Ref, indicates the SSIM index computed for the signal obtained after the whitening and the corresponding template.

Signal	Distance (kpc)	SSIM index			
		$[\mu_{\text{opt}}]$	$[\bar{\mu}]$	$[\mu_{\text{m}}]$	[Ref]
A	5	0.89	0.83	0.84	0.39
	10	0.74	0.68	0.69	0.14
	20	0.54	0.44	0.43	0.03
B	5	0.71	0.61	0.65	0.21
	10	0.51	0.33	0.38	0.06
	20	0.31	0.08	0.11	0.003
C	5	0.64	0.60	0.69	0.06
	10	0.40	0.46	0.51	0.012
	20	0.23	0.24	0.29	0.002

we have to resort to some other termination criterion. We refer to our original paper [34], in which a heuristic procedure is shown to estimate the optimal value of the Lagrange multiplier, for details (see Sec. IV, Table I, and Fig. 4).

Our test over the signals examples shows that tolerances of 10^{-3} for both the rROF algorithm and 10^{-2} for the iterative step are a good compromise between the accuracy on the results and computation speed. The number of Bregman iterations is set to be at most a couple of iterations for the same reasons. These algorithm parameters will remain the same for the cases considered in this paper.

III. ALGORITHM PIPELINE AND DATA CONDITIONING

Our previous work [34] has shown that the rROF algorithm leads to satisfactory results for signals embedded in Gaussian noise. However, the noise of gravitational-wave detectors is non-Gaussian and nonstationary [25, 38–40]. For example, there are well-known, modeled sources of narrow-band noise, such as the electric power (at 60 Hz and higher harmonics), mirror suspension resonances, or calibration lines (see Fig. 3 of Ref. [2]). For this reason, data must first be preconditioned to make the noise flat in frequency (a process known as whitening). To do so, we make as few assumptions as possible. In this work, we employ ten chunks of data of 5 s each from the Advanced LIGO Livingston detector to inject the different signals. The GPS times are selected randomly over the entire O1 period. The sampling frequency is 16 384 Hz.

We preprocess the data using the whitening procedure developed by Refs. [41,42]. This procedure uses an autoregressive (AR) model to transform the colored noise into white noise (see Ref. [41] for details). First, we obtain

the 3000 coefficients of the AR filter required by the whitening using 300 s of data at the beginning of the corresponding science segment of every signal injection. The whitening is applied in the time domain to every block of 5 s of data we use in order to avoid the border problems associated with the transformations in the frequency domain.

As in Ref. [34], we apply the TV method to two different types of gravitational-wave signals. The first type of waveforms is bursts from CCSN. We employ the waveform catalog of Dimmelmeier *et al.* Ref. [43], who built a catalog of 128 waveforms from general-relativistic simulations of rotating stellar core collapse to neutron stars. The simulations considered progenitors with a high rotation rate and two tabulated, microphysical equations of state (EoS). The second type of waveforms is based on the 174 numerical simulations from the inspiral and merger of BBHs of Mroué *et al.* [44], 167 of which cover more than 12 orbits and 91 of which represent precessing binaries.

The rROF algorithm is coded in FORTRAN combined with a PYTHON interface for plotting purposes. The algorithm is very efficient; the average time of 1000 runs of 3 s of data takes ∼16 ms, computed in a single processor 3.5 GHz Intel Core i7 with 16 Gbytes of RAM. The iterative procedure, including the Bregman iteration, takes on average 0.5 s to perform the denoising of 3 s of data.

IV. ESTIMATION OF THE REGULARIZATION PARAMETER

As already discussed in Ref. [34], the denoising results strongly depend on the value of the regularization parameter μ. If this value is too large, the fidelity term in Eq. (2.2) dominates, and the denoised signal is comparable to the original noisy signal y. On the contrary, if the value of μ is too small, it is the regularization term in Eq. (2.2) the dominant one and the amplitude of the resulting signal tends to zero. The existence of an *optimal* value of μ can be proven theoretically. However, this unique value is not equally appropriate for all possible cases one may consider (involving differences in the noise and/or in the signals) and must therefore be set empirically in practice.

In this section, we determine the interval of values of μ in which satisfactory results are expected. This is similar to the analysis we performed in Ref. [34], apart from the fact that we now use a logarithmic scale because the minimizer function converges faster than with a linear scale. For this reason, the regularization parameter used in the rROF algorithm is 10^{μ}. The optimal value, μ_{opt}, is the one that gives the best results according to some suitable metric function applied to the denoised signal and to the original one. This function is used to measure the quality of the recovered signal. In Ref. [34], we chose the peak signal-to-noise ratio as our quality estimator. In the present paper, we assess the results of the iterative rROF algorithm using the structural similarity (SSIM) index, motivated by the quality

assessment based on error sensitivity reported by Ref. [45]. This estimator deviates from the traditional measures of error, which are based on the calculation of the absolute error, because it takes into account the structural information of both the original and the reconstructed signals. The SSIM index varies between 0 (minimum similarity) and 1 (maximum similarity) and is defined as

$$\text{SSIM}(x, y) = \frac{(2\mu_x\mu_y + c_1)(2\sigma_{xy} + c_2)}{(\mu_x^2 + \mu_y^2 + c_1)(\sigma_x^2 + \sigma_y^2 + c_2)}, \quad (4.1)$$

where c_1 and c_2 are constants, μ_x (μ_y) is the average of x (y), σ_x^2 (σ_y^2) is the variance of x (y), and σ_{xy} is the covariance of x and y. The error provided by the SSIM index is used to determine the optimal value of the regularization parameter μ in each case.

We search for the optimal value of μ, i.e., the one that maximizes the SSIM, injecting numerical-relativity signals from the CCSN and BBH catalogs into O1 data. For the former, we employ 30 different CCSN signals at three different distances, namely, 5, 10, and 20 kpc. These distances are the same as were used in Ref. [26] and represent a reasonable example of distance for the signals in the Dimmelmeier catalog [43]. The injections are performed at ten different random GPS times over the Advanced LIGO O1 data. With all these data, we obtain the histogram of optimal values of μ shown in the left panel of Fig. 1. We follow the same procedure for the BBH signals of Ref. [44], but in this case, the chosen distances are 400, 800, and 1000 Mpc, which are similar to the distances of the detected BBH events [2,3,5–7]. The corresponding histogram is shown in the right panel of Fig. 1.

The mean value of μ for CCSN signals is $\bar{\mu}_{opt} = -0.28$ with a standard deviation of $\sigma_{opt} = 0.58$. In the case of BBH signals, the values are $\bar{\mu}_{opt} = -0.95$ and $\sigma_{opt} = 0.27$, respectively. The mean values of μ_{opt} are different for both types of signals. This is expected since the two signals are very different and the conditions that apply for one type do not apply for the other. Specifically, for the BBH signals, we have centred our analysis in the denoising of the very last cycles of the inspiral, the merger, and the ringdown parts. This selection produces unsatisfactory denoising of the inspiral part of the signal, as we show later.

Although the mean value of μ_{opt} is different for both catalogs, Fig. 1 also shows that partial overlap exists between the two distributions. This is expected since if we knew the variance of the noise it would be possible to determine the most appropriate value of the Lagrange multiplier for the fidelity term that corresponds to that variance. On the other hand, in a realistic situation, the template is unknown, which renders it impossible to obtain μ_{opt}. Therefore, other strategies are required to determine the values of the regularization parameter that produce
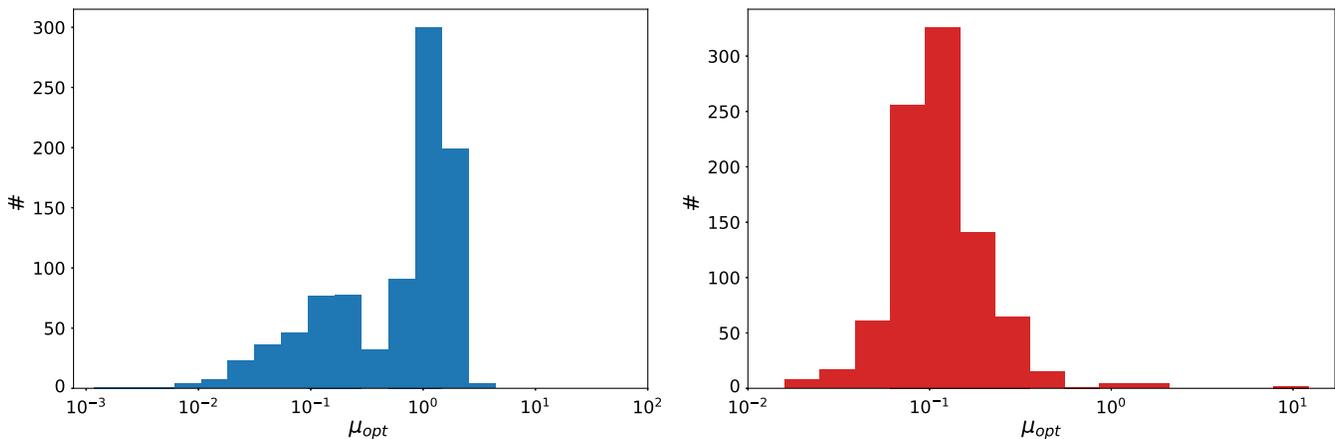
FIG. 1. Histograms of the optimal values of μ for the training signals from the core-collapse catalog (left panel) and BBH catalog (right panel) at three different distances in ten different GPS times.

good results. For this reason, in this work, we also try out and compare two different approaches based on the information provided by the histograms of Fig. 1. The first one is based on the mean value of μ for all waveforms. The second approach is to use the average of 20 different values sampled from a Gaussian distribution with the same mean and variance as the corresponding histograms.

V. RESULTS

A. Core-collapse supernova signals

We first assess our method with three signals from the CCSN catalog placed at a distance of 10 kpc and using the optimal value of μ for each case. The particular signals are s20a1o05 (signal A), s20a2o09 (signal B), and s20a3o15 (signal C) of the Dimmelmeier catalog [43]. The physical properties of the CCSN models leading to these waveform signals are reported in the Appendix. With the source at 10 kpc, the signal is visible over the noise. However, despite its simplicity, this is the first test the method shall pass. The results are displayed in Fig. 2. As the distance is fixed, the strain (amplitude) of the signal depends on the particular simulation; i.e., the SNR is different for each case. Figure 2 shows that signal A has the largest amplitude (top panel). The morphology of these signals is characterized by the presence of a negative peak linked to the time of the core bounce followed by damped oscillations of the newborn proto-neutron star.

All denoised signals are very similar to the whitened templates. Signal A is the strongest, and the denoised signal fits the template almost perfectly. The other two signals are weaker at 10 kpc, and the denoising procedure leads to more oscillatory signals. The amplitude and phase of the main positive and negative peaks and the first secondary peak in all signals are well recovered. In contrast, the low-amplitude damped oscillations that follow the burst (associated with the oscillation of the proto-neutron star) are lost

since, due to their low amplitude, they are more affected by noise. This is inherent to the rROF model, as it preserves large gradients and disfavors small ones.

The denoised signal shown in the middle panel of Fig. 2 is more oscillatory than the other two. In this case, a higher value of μ is required to recover its peaks properly, which leads to the presence of more noise than in the other cases. This fact is something to take into account in a realistic case in which the real signal is unknown because these small oscillations could be disregarded in favor of a more regular signal and more noise removal. However, it is always possible to use a larger value of μ to recover these parts of the signal.

To complete the analysis, we also compute the denoising using the mean value of the regularization parameter and the multiple regularization for all cases, in which the distance is different (5, 10, and 20 kpc). The resulting values of the SSIM index are shown in Table I. In particular, the last column of this table shows the SSIM index computed for the signal obtained after the whitening and the corresponding numerical template. Therefore, it provides a measure of the improvement obtained with the rROF method. The values of the SSIM index are computed in a 256 window centered at the position of the negative peak of the signal. These values are computed before applying the TV method in order to illustrate its performance. As expected, the values of the SSIM index become worse as the distance increases, irrespective of the type of regularization parameter employed. The comparison shows that the optimal value of the regularization parameter produces the best results in all cases. The denoising worsens when using the mean value of μ; however, the results are similar to those obtained with the optimal one. It thus seems possible to use the mean value for all signals and still obtain good results. Likewise, the use of multiple regularization values seems to be a good alternative, too, because the values of the SSIM index are very similar to the other cases. For the case of μ_{m}, the values
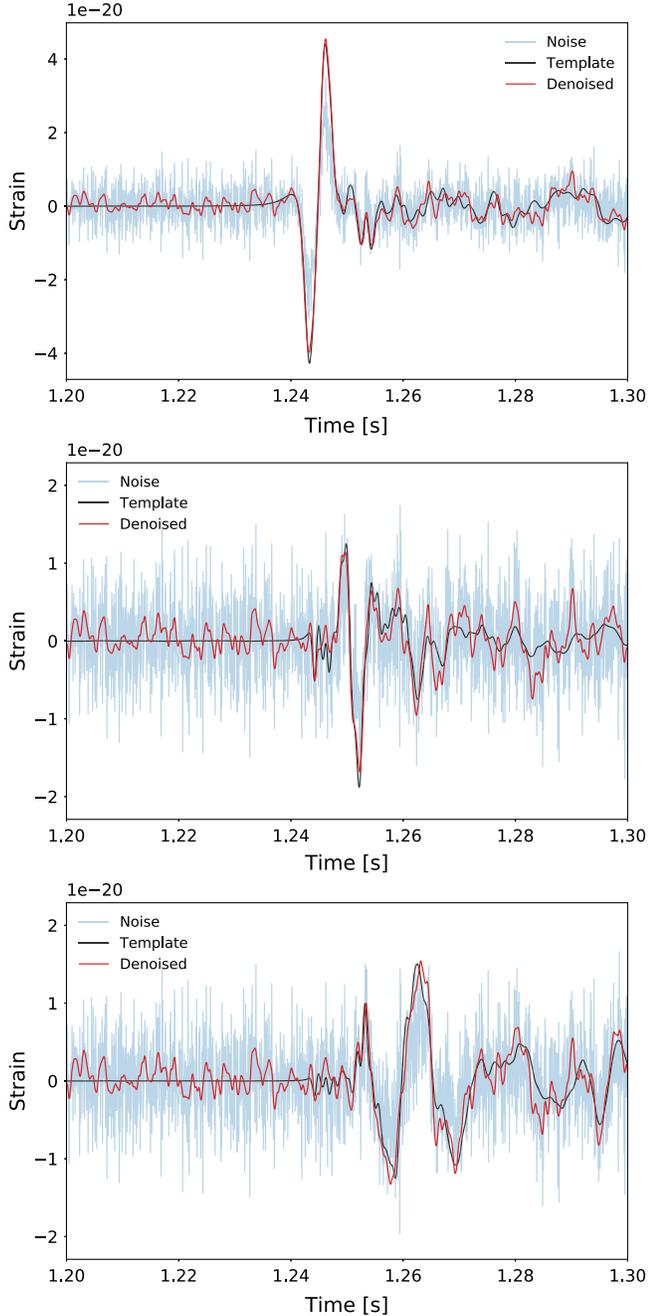
FIG. 2. Results of applying the rROF model to three CCSN signals from the Dimmelmeier catalog [43] at a distance for 10 kpc, namely, A (upper panel), B (middle panel), and C (lower panel). Black lines indicate the original signals (after whitening), and red lines correspond to the denoised ones.

of the SSIM index depend on the sampling of the Gaussian distribution. We have repeated the sampling several times, finding similar results. It is expected that as we increase the number of samples obtained from the distribution of μ the value of the SSIM index will converge to that obtained with $\bar{\mu}$. However, multiple regularization has the advantage that the results can be analyzed separately.

The results of Table I also show that there is not a very strong dependence with μ; i.e., if the chosen value is of the same order of magnitude than μ_{opt}, the results are quite similar. This is a different behavior with respect to the results we found in Ref. [34], in which this dependence was more critical. The reason is the use of the iterative procedure, which allows us to choose larger initial values of μ. Slightly different values of μ will require a different number of iterations to reach convergence, but the result will be similar.

To further test the performance for signals at 20 kpc, we make a complementary test. We compute the spectrogram of each signal and integrate the power for each temporal channel. Then, we determine the time at which the maximum power is achieved. We make this calculation for the noisy and the denoised signals and compare the results with the waveform template. In all cases considered, the time given by the denoised signal matches the one given by the template even if the noisy signal (after whitening) does not.

B. Binary black hole signals

We turn next to perform the same analysis on BBH waveform signals. These signals are significantly longer than those from CCSN and are composed by three parts: inspiral, merger, and ringdown. During the inspiral, the signal amplitude and frequency increase up until merger. For our tests, as we did in Ref. [34], we employ signal BBH0001 from the catalog of Ref. [44].

The results of denoising this BBH signal placed at a distance of 400 Mpc are shown in Fig. 3. The last few cycles before merger (at $t \sim 2.4$ s) are the part of the signal the TV algorithm recovers best, as expected, since the value of μ is adapted to this region. Previous cycles of the inspiral are less smooth because, in general, the merger requires a
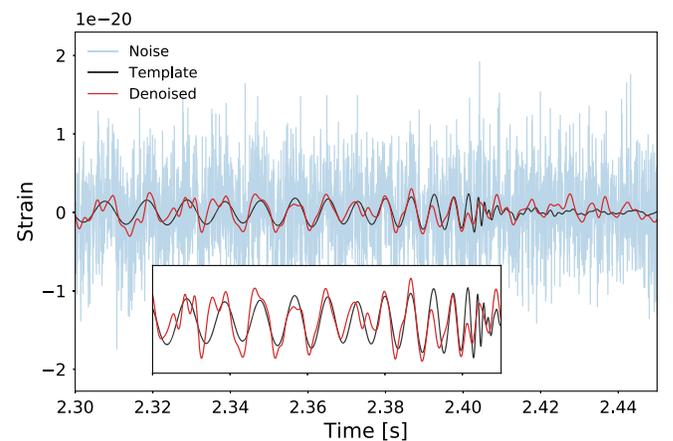

FIG. 3. Denoising of signal BBH0001 of the Mroué catalog [44] at a distance of 400 Mpc. The black line is the original template (after whitening), and the red line corresponds to the denoised signal. In the inset, the noise is not plotted, and the area of interest is enlarged to facilitate the comparison.

TABLE II.   Values of the SSIM index for a BBH waveform when using the optimal value of the regularization parameter for each signal, μ_{opt}; the mean value for all signals, $\bar{\mu}$; multiple values, μ_m; and the reference value computed for the signal obtained after the whitening and the corresponding template.

Signal	Distance (Mpc)	SSIM index			
		$[\mu_{opt}]$	$[\bar{\mu}]$	$[\mu_m]$	[Ref]
0001	400	0.43	0.25	0.25	4×10^{-4}
	800	0.26	0.14	0.10	1×10^{-4}
	1000	0.11	0.10	0.10	4×10^{-5}

higher value of μ, a choice that is not optimal for the rest of the signal.

The dependence of our results with other distances and other possible choices of the regularization parameter are reported in Table II. Both $\bar{\mu}$ and μ_m produce worse results than the optimal regularization parameter as the distance increases. This may happen if the optimal value for a given distance lies at the tail of the distribution shown in the histogram, as, e.g., in the case of 800 Mpc, in which $\mu_{opt} = -2.45$, which is very different from $\bar{\mu}$. Also note that for a distance of 1 Gpc all three choices of μ lead to similar fairly low values of the SSIM index.

As the waveform produced by a BBH coalescence has significant length and variations in frequency and amplitude, the optimal value of μ is different for different parts of the signal. In Fig. 3 and Table II, we have selected the values of μ that best fit the last cycles and the merger, choosing this part as the temporal window in which the SSIM index is computed. However, there is no guarantee that this value will produce the best results in other parts of the signal.

To determine the optimal value of μ in different parts of the signal, we split the waveform into pieces of 256 samples and search for μ_{opt} for each of the resulting windows. The denoised signal for a BBH merger at 400 Mpc is shown in
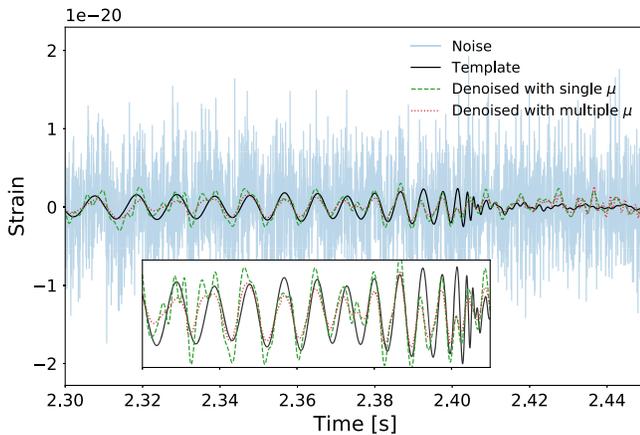
Fig. 4, both with a single value of μ_{opt} and with multiple values. If we compare the results for the SSIM index restricting the comparison to the last cycles of the inspiral and the merger, the value is similar to that obtained with a single μ_{opt} procedure. However, when considering the whole signal, the global value of the SSIM index significantly improves, increasing from 0.12 to 0.40. The comparison of Figs. 3 and 4 shows that the merger part is similar in both cases but the inspiral part is recovered better in the latter, in which the value of μ has been chosen to fit each part separately.

C. Automatic regularization with an artificial neural network

From the results of the previous section, we can devise an automatic way to find the optimal value of μ depending solely on the data at the input. The goal of this section is to present a simple way to obtain a value of μ closer to the optimum one in a realistic case, when the latter cannot be computed. This is a typical problem for machine-learning methods, in which the result is determined by the data. Machine-learning algorithms have been applied to a wide (and growing) range of fields (see Ref. [28] and references therein), and a large variety of approaches is available. For our case, we implement a nonlinear regressor that maps for each input window of the signal the optimal value of μ required to achieve the best denoising results. A more comprehensive analysis with different configurations of neural networks will be presented elsewhere.

We set up a very simple configuration with 40 neurons in one layer. The detailed structure of the network is shown in Fig. 5. Each neuron performs a linear calculation,

$$z_i = w_i^l X_i + b_i^l, \tag{5.1}$$

where w_i^l are the weights, b_i^l is the bias parameter, X_i is the input data, and z_i is the output of neuron i at layer l.

FIG. 4.   Same as Fig. 3, but including the curve corresponding to using multiple regularization parameters.
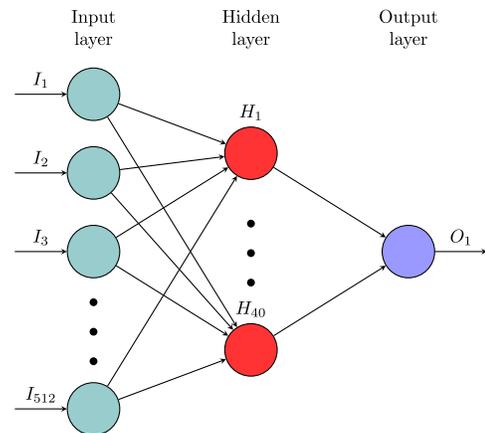

FIG. 5.   Structure of the artificial neural network used to perform the regression.

Nonlinearity is achieved using the so-called activation function (see Ref. [28] for details). In this case, we use the well-known Relu activation, which is given by

$$f(x) = \max(0, x). \tag{5.2}$$

The best values of the weight matrix of each layer is achieved during the training step, in which for each input example the network computes the output and compares the result with the input value μ_{opt} using the mean squared error (MSE) as an error quantifier. The network changes the weights iteratively to reduce the MSE. To perform this optimization procedure, we use the Adam optimizer [46].

We first consider the CCSN catalog employed in Sec. IV. For the training, we use 10^4 random examples of the set of 30 CCSN signals. The length of each signal example is 512 samples. The top panel of Fig. 6 displays the results of applying the regularization parameter determined by the regressor to the CCSN signal 60 of Ref. [43] (our signal A) at 10 kpc. As the figure shows, the results achieved with the optimal value of μ and with the one given by the regressor are very similar (both curves almost completely overlap).
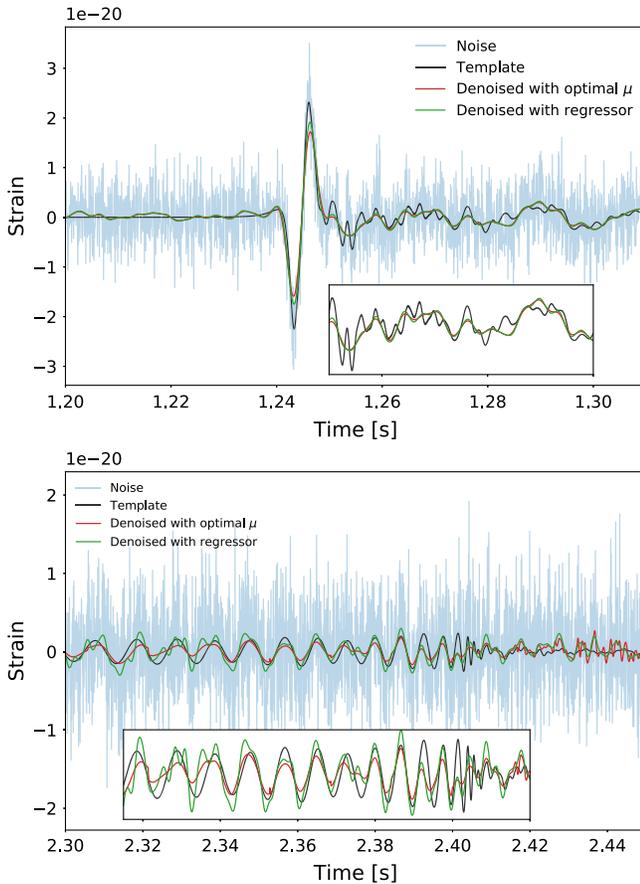

TABLE III. Comparative results between using the optimal value of μ and the value given by the regressor μ_{reg}, for three CCSN signals.

Signal	Distance (kpc)	SSIM index	
		$[\mu_{\mathrm{opt}}]$	$[\mu_{\mathrm{reg}}]$
A	5	0.78	0.73
	10	0.64	0.63
	20	0.46	0.38
B	5	0.25	0.26
	10	0.16	0.16
	20	0.10	0.05
C	5	0.62	0.51
	10	0.55	0.51
	20	0.46	0.46

In addition, Table III reports the values of the SSIM index after applying the regression procedure to three CCSN signals at three different distances and the values for μ_{opt} for reference. The results are different than the ones shown in Table I because of the different window used. Here, we use a sliding window to denoise the entire signal, and thus the windows are not exactly the same. For most cases, the values of the SSIM index given by the regressor are similar to the optimals.

Finally, we train the regressor with signals from the BBH catalog, and we repeat the analysis done in Sec. V B. Specifically, we train the network with 10^4 random examples of the set of 30 BBH signals, at five different GPS times and three distances, and a window of 512 samples. The value of the SSIM index is 0.41 for the optimal regularization parameter, 0.36 for the regressor in each window, and 0.2 if we use for all the signals μ_{opt} computed at merger. The comparison is shown in the bottom panel of Fig. 6.

VI. SUMMARY

This paper has extended the work we initiated in Ref. [34] to denoise gravitational-wave signals using total-variation methods. We have assessed these techniques in real noise conditions, injecting numerical-relativity waveforms from CCSN and BBH mergers in data from the first observing run of Advanced LIGO. We have shown that TV methods remove noise irrespective of the type of signal. The denoising procedure is performed in two steps. First, we apply a whitening procedure to remove lines and to flatten the noise spectrum, and next we apply the TV method. The quality of the results depends on the value of the regularization parameter μ of the ROF model. Therefore, we have to perform a heuristic search for the Lagrange multiplier that produces the best results. To improve the statistics, we have computed the optimal value of μ for 30 different signals from a CCSN catalog [43] and

FIG. 6. Top panel: denoising of CCSN signal A at a distance of 10 kpc, showing the comparison between μ_{opt} and μ_{reg}. Bottom panel: same comparison, but for the BBH signal.

for another 30 from a BBH merger catalog [44], placing the signals at various distances. The histograms have shown that the interval of optimal values of μ is not very wide. However, the use of an iterative procedure reduces somewhat the denosing dependence on μ and has allowed us to obtain similar results with a larger span of values of μ than in our first paper [34].

In a realistic situation, however, the original signal is not known, and it is not possible to determine the optimal value of the regularizer. Therefore, in this work, we have expanded the analysis by testing additional ways to perform the denoising using a general value of μ. The first ideas are based on the results of the histograms. We have shown that using the mean value can suffice in most cases. Another approach uses 20 different values of μ to compute the mean, yielding similar results. Multiple-μ denoising allows us to compare the results at different scales, to process them separately, and to combine them to obtain the best results. However, even though the method is very fast (on average, it takes about 0.5 s to denoise 3 s of signal with the iterative procedure), it requires multiple blind selections of μ, which in some cases might not be adequate. For this reason, we have also tested the use of a neural network to determine the value of the regularization parameter. We plan to further explore this approach in a future investigation.

For the case of long-duration signals such as BBH waveforms, our results have shown that a single value of the regularization parameter does not provide a good enough denoising across the entire signal. Instead, combining the results using optimal values of μ adapted to different parts of the signal improves the results. We have generalized this procedure by employing a regressor implemented with an artificial neural network of 40 neurons in one layer. We have shown that this machine-learning approach leads to results similar to those obtained with the optimal regularization parameter. Therefore, it is worth it to combine TV methods with machine-learning techniques to improve the results and obtain the Lagrange multiplier μ through an automatic search.

This paper provides further evidence that TV methods can be useful in the field of gravitational-wave astronomy as a tool to remove noise. They can be used in a preprocessing step before applying other common techniques of gravitational-wave data analysis. However, on their own, they cannot constitute a stand-alone pipeline since, as they do not use any information about sources, they cannot detect, classify, or extract physical information. We plan to investigate a combined strategy and test if the application of TV denoising can improve the results of other approaches, e.g., by reducing the uncertainties in Bayesian methods or reducing the false-alarm rate. Furthermore, we want to keep exploring the combination of TV methods with machine-learning techniques. More precisely, we will further explore the determination of μ via machine-learning regressors, and we will work on improving the denoising algorithm itself by using multilayer structures typical of deep learning. Our findings will be presented elsewhere.

APPENDIX: PHYSICAL PARAMETERS OF THE CCSN SIGNALS

To facilitate the identification of the three CCSN waveforms used in this study, Table IV reports the physical properties of the corresponding core-collapse models simulated by Ref. [43].

TABLE IV. Precollapse rotation properties of our three models of the core-collapse catalog [43]. A is the differential rotation length scale, $\Omega_{c,i}$ is the precollapse angular velocity at the center, and β_i is the precollapse rotation rate.

Signal name	Mass $M_\odot$	A [10^8 cm]	$\Omega_{c,i}$ [rad s^{-1}]	β_i [%]	EoS
s20a1o05	20	50.0	0.45	0.05	Shen
s20a2o09	20	1.0	4.56	0.90	Shen
s20a3o15	20	0.5	13.31	2.50	Shen

[1] J. Aasi *et al.* (LIGO Scientific Collaboration), Classical Quantum Gravity **32,** 115012 (2015).

[2] B. P. Abbott *et al.*, Phys. Rev. Lett. **116,** 061102 (2016).

[3] B. P. Abbott *et al.*, Phys. Rev. Lett. **116,** 241103 (2016).

[4] The Virgo Collaboration, Technical Report VIR-027A-09, 2009.

[5] B. P. Abbott *et al.*, Phys. Rev. Lett. **118,** 221101 (2017).

[6] B. P. Abbott *et al.*, Astrophys. J. Lett. **851,** L35 (2017).

[7] B. P. Abbott *et al.*, Phys. Rev. Lett. **119,** 141101 (2017).

[8] B. P. Abbott *et al.*, Phys. Rev. Lett. **119,** 161101 (2017).

[9] B. P. Abbott *et al.*, Astrophys. J. Lett. **848,** L12 (2017).

[10] N. J. Cornish and T. B. Littenberg, Classical Quantum Gravity **32,** 135012 (2015).

[11] B. S. Sathyaprakash and B. F. Schutz, Living Rev. Relativity **12**, 2 (2009).

[12] S. A. Usman *et al.*, Classical Quantum Gravity **33**, 215004 (2016).

[13] S. Bose, A. Pai, and S. Dhurandhar, Int. J. Mod. Phys. D **09**, 325 (2000).

[14] C. Palomba (LIGO Scientific and Virgo Collaborations), arXiv:1201.3176.

[15] E. Thrane and M. Coughlin, Phys. Rev. Lett. **115**, 181102 (2015).

[16] S. Klimenko *et al.*, Phys. Rev. D **93**, 042004 (2016).

[17] T. B. Littenberg, J. B. Kanner, N. J. Cornish, and M. Millhouse, Phys. Rev. D **94**, 044050 (2016).

[18] J. B. Kanner, T. B. Littenberg, N. Cornish, M. Millhouse, E. Xhakaj, F. Salemi, M. Drago, G. Vedovato, and S. Klimenko, Phys. Rev. D **93**, 022002 (2016).

[19] R. Lynch, S. Vitale, R. Essick, E. Katsavounidis, and F. Robinet, Phys. Rev. D **95**, 104046 (2017).

[20] I. S. Heng, Classical Quantum Gravity **26**, 105005 (2009).

[21] C. Röver, M.-A. Bizouard, N. Christensen, H. Dimmelmeier, I. S. Heng, and R. Meyer, Phys. Rev. D **80**, 102004 (2009).

[22] J. Logue, C. D. Ott, I. S. Heng, P. Kalmus, and J. H. C. Scargill, Phys. Rev. D **86**, 044023 (2012).

[23] W. J. Engels, R. Frey, and C. D. Ott, Phys. Rev. D **90**, 124026 (2014).

[24] M. C. Edwards, R. Meyer, and N. Christensen, Inverse Probl. **30**, 114008 (2014).

[25] M. C. Edwards, R. Meyer, and N. Christensen, Phys. Rev. D **92**, 064011 (2015).

[26] J. Powell, S. E. Gossan, J. Logue, and I. S. Heng, Phys. Rev. D **94**, 123012 (2016).

[27] J. Powell, M. Szczepanczyk, and I. S. Heng, Phys. Rev. D **96**, 123013 (2017).

[28] I. Goodfellow, Y. Bengio, and A. Courville, *Deep Learning* (MIT Press, Cambridge, MA, 2016), http://www.deeplearningbook.org.

[29] A. Torres-Forné, A. Marquina, J. A. Font, and J. M. Ibáñez, Phys. Rev. D **94**, 124040 (2016).

[30] M. Razzano and E. Cuoco, Classical Quantum Gravity **35**, 095016 (2018).

[31] H. Gabbard, M. Williams, F. Hayes, and C. Messenger, Phys. Rev. Lett. **120**, 141103 (2018).

[32] D. George, H. Shen, and E. A. Huerta, Phys. Rev. D **97**, 101501 (2018).

[33] D. George and E. A. Huerta, Phys. Lett. B **778**, 64 (2018).

[34] A. Torres, A. Marquina, J. A. Font, and J. M. Ibáñez, Phys. Rev. D **90**, 084029 (2014).

[35] L. I. Rudin, S. Osher, and E. Fatemi, Phys. D (Amsterdam, Neth.) **60**, 259 (1992).

[36] R. Tibshirani, J. R. Stat. Soc. Ser. B **58**, 267 (1996).

[37] S. Osher, M. Burger, D. Goldfarb, J. Xu, and W. Yin, Multiscale Model. Simul. **4**, 460 (2005).

[38] N. Christensen (LIGO Scientific and Virgo Collaborations), Classical Quantum Gravity **27**, 194010 (2010).

[39] J. Aasi *et al.*, Phys. Rev. D **88**, 122004 (2013).

[40] T. B. Littenberg and N. J. Cornish, Phys. Rev. D **91**, 084034 (2015).

[41] E. Cuoco, G. Calamai, L. Fabbroni, G. Losurdo, M. Mazzoni, R. Stanga, and F. Vetrano, Classical Quantum Gravity **18**, 1727 (2001), http://stacks.iop.org/0264-9381/18/i=9/a=309.

[42] E. Cuoco, G. Losurdo, G. Calamai, L. Fabbroni, M. Mazzoni, R. Stanga, G. Guidi, and F. Vetrano, Phys. Rev. D **64**, 122002 (2001).

[43] H. Dimmelmeier, C. D. Ott, A. Marek, and H.-T. Janka, Phys. Rev. D **78**, 064056 (2008).

[44] A. H. Mroué *et al.*, Phys. Rev. Lett. **111**, 241104 (2013).

[45] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, IEEE Trans. Image Process. **13**, 600 (2004).

[46] D. P. Kingma and J. Ba, arXiv:1412.6980.