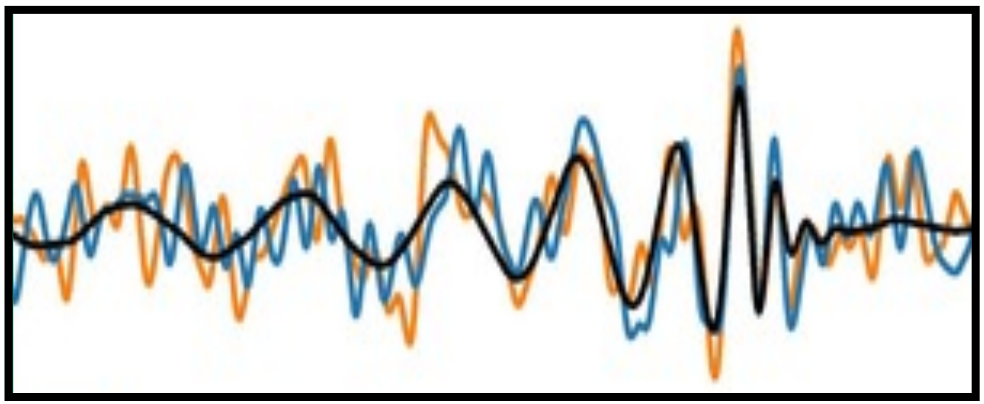
Total-variation and dictionary-learning methods for GW data analysis



José Antonio Font
Universitat de València
www.uv.es/virgogroup

VIRGO

G2NET COST ACTION CA17137 A NETWORK FOR GRAVITATIONAL WAVES, GEOPHYSICS AND MACHINE LEARNING

EGO, 14-16 January 2019

Outline of the talk

1. Total-variation methods for gravitational-wave ***denoising***

   Torres-Forné, Marquina, Font, & Ibáñez, PRD, 90 , 084029 (2014)

   Torres-Forné, Cuoco, Marquina, Font, & Ibáñez, PRD, 98 , 084013 (2018)

2. Gravitational-wave ***denoising*** with dictionary learning

   Torres-Forné, Marquina, Font, & Ibáñez, PRD, 94 , 124040 (2016)

3. ***Classification*** of gravitational-wave glitches via dictionary learning

   Llorens-Monteagudo, Torres-Forné, Font, & Marquina, CQG (2019)

Sources of noise

The sensitivity of GW detectors is limited by diverse sources of noise.

<u>Fundamental Noises:</u>

I. Displacement Noises

$\Delta L(f)$

- Seismic noise
- Radiation Pressure
- Thermal noise

  Suspensions

  Optics

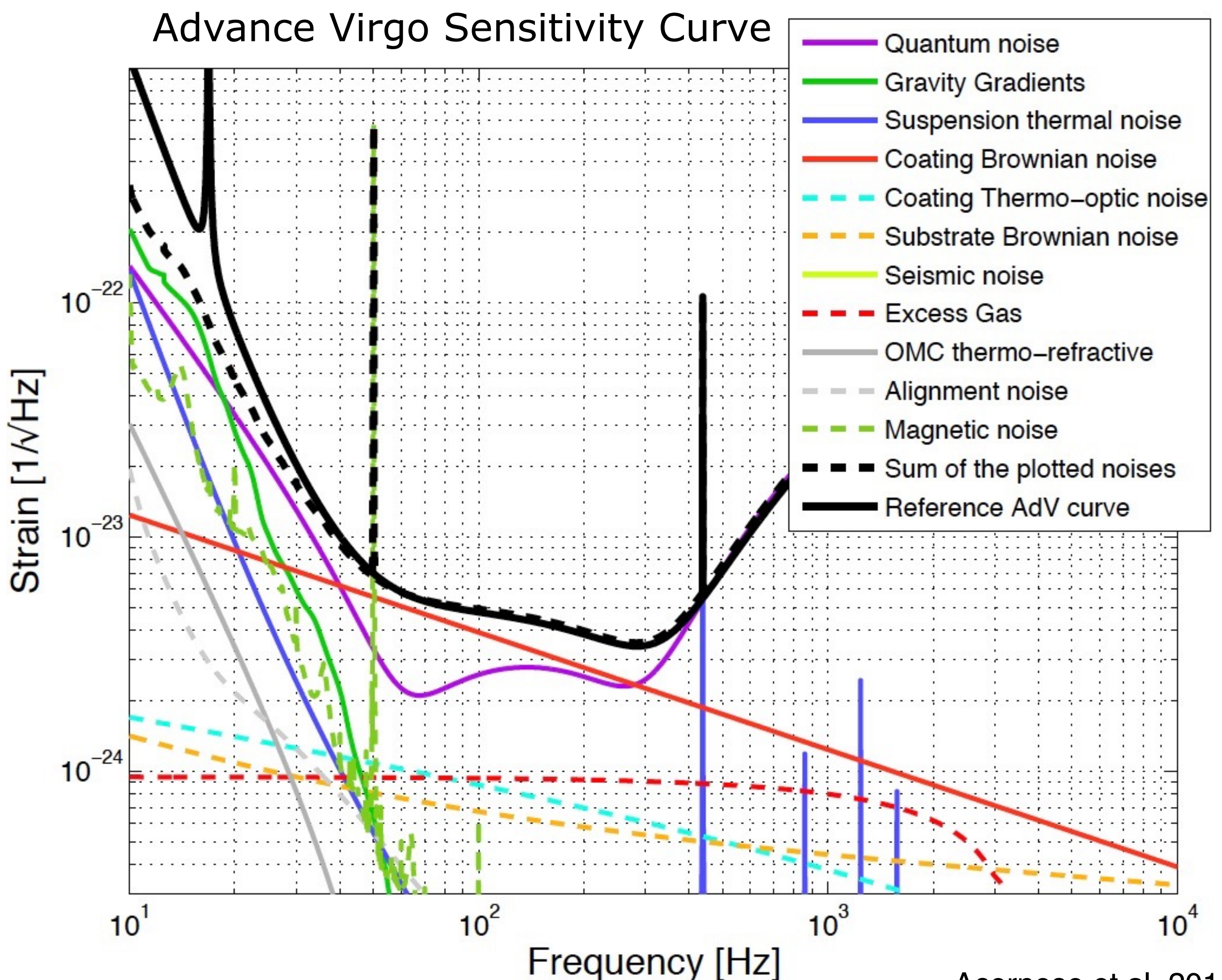
II. Sensing Noises

$\Delta t_{photon}(f)$

- Shot Noise
- Residual Gas

<u>Technical Noises:</u>

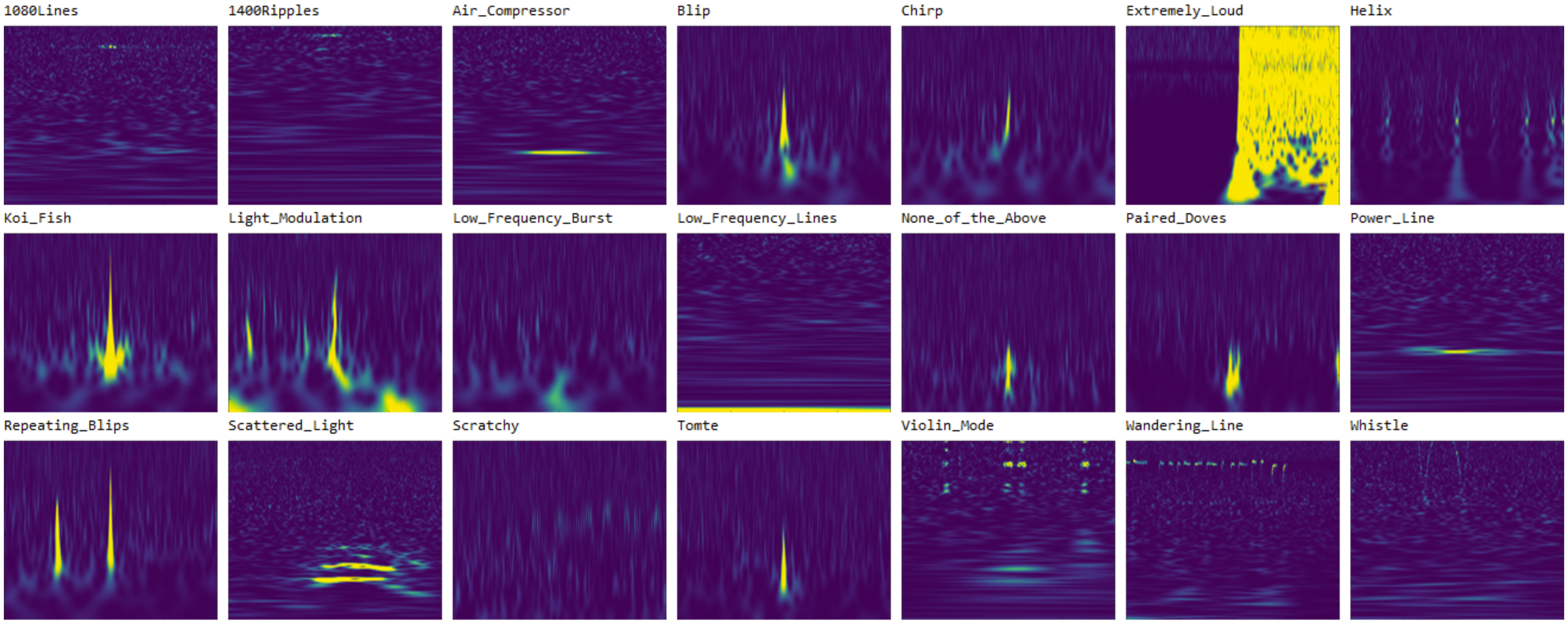
→ Hundreds of them…



Acernese et al. 2014

Noise transients - glitches

Non-Gaussian transients of noise. Large variety of morphologies.



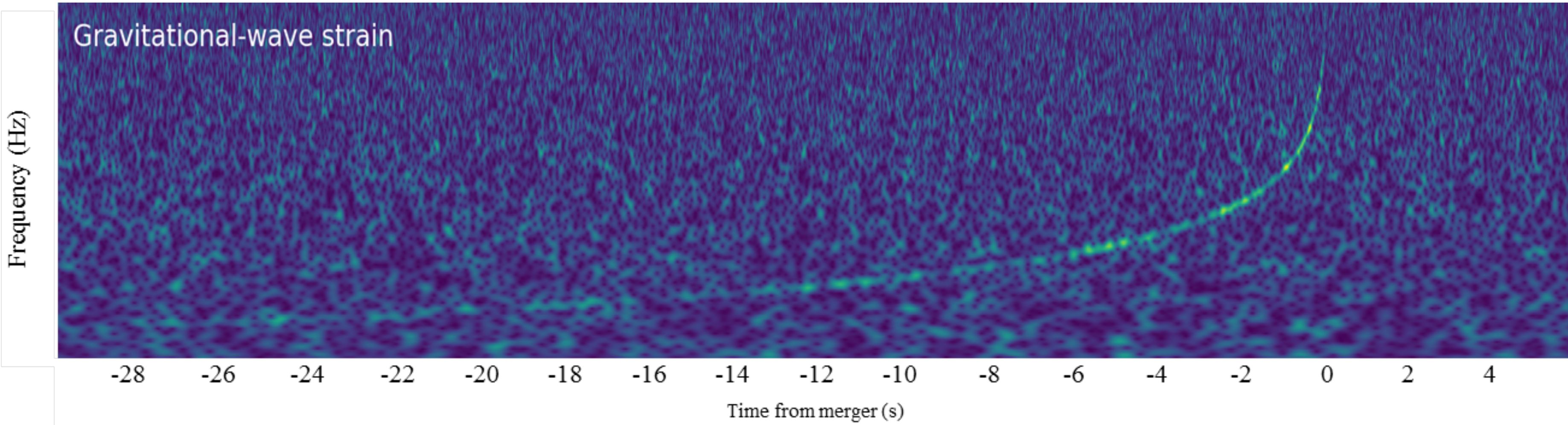
Gravity Spy, Zevin et al (2017) www.zooniverse.org/projects/zooniverse/gravity-spy

Effect on detectors

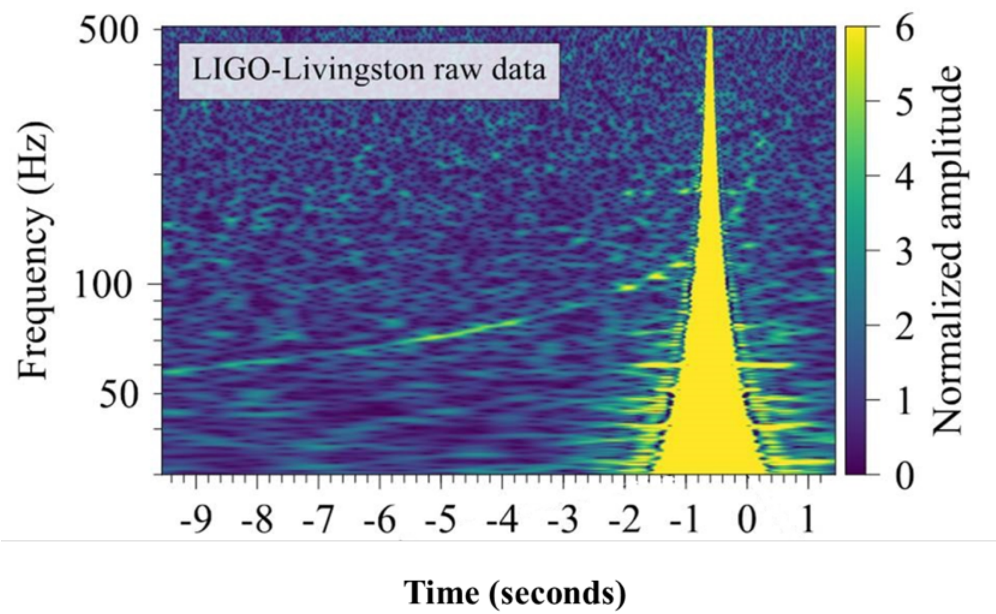
1. Reduce significance of candidate GW events.
2. Affect estimation of physical parameters.
3. Reduce amount of usable data.

Prompt characterization of noise critical for improving sensitivity.
Fast methods for glitch classification are needed.

GW170817 glitch



Omega Scan of LIGO's First Neutron Star Merger Detection



Abbott et al (2017)

Gravitational-wave data analysis: steps

Proof of concept
- Test with non-white Gaussian noise.
- Proof that the algorithm works with GW signals.

Application to real data
- Test with non-white non-Gaussian noise.
- Proof that the algorithm works with GW real data.

Test with standard pipelines
- Use in combination with standard pipelines.
- Proof that the algorithm improves the results.

Total-variation methods for GW denoising

- Data analysis technique to remove noise.

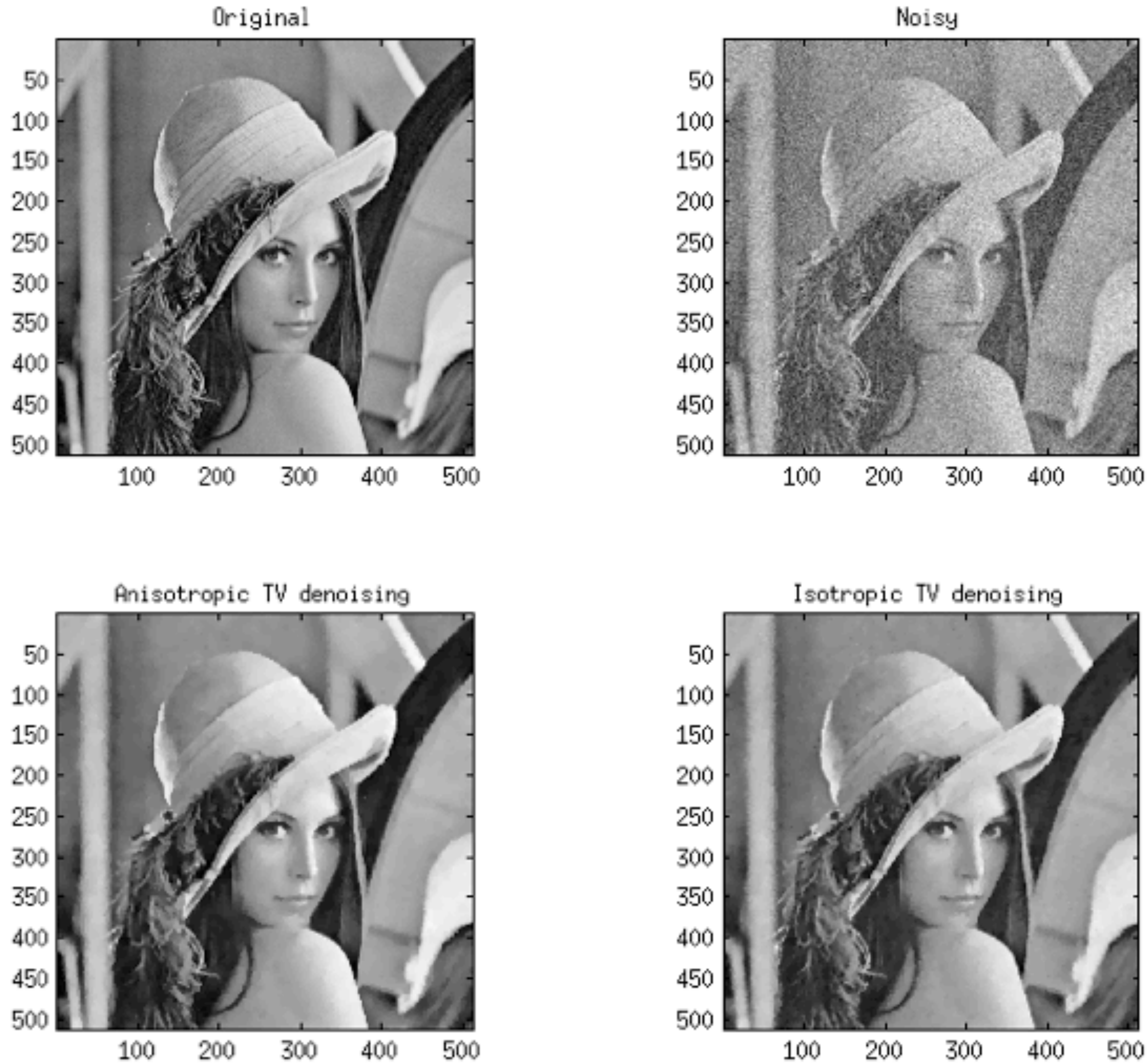
- TV denoising is one of the best denoising models but also one of the hardest to compute as it is based on the L1 norm.

- Developed and tested (mainly) in the context of image restoration.

- At UV we have adapted TV techniques to GW signals.

- Interesting features:

  No a priori information of the source needed.
  Can be applied in both time and frequency domains.
  Can be easily extended to higher dimensionality.
  Complementary to existing data analysis methods.



Example image TV denoising using the split Bregman method for L1 regularized problems (Goldstein & Osher 2009).

(as implemented by B. Tremoulheac in MathWorks)

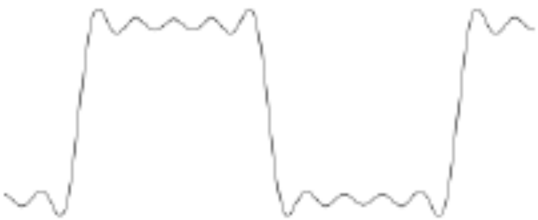
Introduction to variational problems

Linear degradation model: $f = u + n$

f measured signal
n noise
u signal to recover

Solution: find a functional u whose L2-norm distance to f is the standard deviation of the noise

$$||f - u||_{L_2}^2 = \sigma^2$$ (e.g. least squares or Fourier series)



Gibbs phenomena. Non unique solution.

Issues overcome using an auxiliary energy prior $\mathcal{R}(u)$ to regularize the least-squares problem, solving a constrained variational problem:

$$\min_u \mathcal{R}(u) \quad \text{subject to} \quad ||f - u||_{L_2}^2 = \sigma^2$$

Unique solution if $\mathcal{R}(u)$ is convex.

Variational problem can be formulated as an unconstrained problem (Tikhonov regularization):

$$u = \underset{u}{\operatorname{argmin}} \left\{ \mathcal{R}(u) + \frac{\lambda}{2} \mathcal{F}(u) \right\}$$

$\mathcal{F}$: Fidelity term. Measures the similarity of the solution to the data.

$\mathcal{R}$: Regularization term. Constraint we impose on the data.

λ : Regularization parameter. Controls relative weight of both terms.

$$u = \arg\min_{u} \left\{ \mathcal{R}(u) + \frac{\lambda}{2} ||f - u||_{L_2}^2 \right\}, \ \lambda > 0$$

Unique solution for a given value of λ .

λ becomes the scale parameter. Larger values allow to recover finer scales.

If $\mathcal{R}(u) \equiv \int |\nabla u|_{L_2}^2$ Wiener filter

Associated Euler-Lagrange eq.

$$\Delta u + \lambda (f - u) = 0$$

Elliptic PDE. Easy to solve due to differentiability and strict convexity.

Issues in the presence of noise: a) amplification of high frequencies; b) the recovered smooth solution shows spurious oscillations near steep gradients or edges.

Rudin-Osher-Fatemi model (ROF)

$$u = \underset{u}{\operatorname{argmin}} \left\{ \mathrm{TV}(u) + \frac{\lambda}{2} ||u - f||_2^2 \right\} \qquad \mathrm{TV}(u) = \int |\nabla u|_1$$

Convex problem.
Preserves steep gradients or edges and avoids spurious oscillations.
Unique solution for a given value of the regularization parameter.

Fine scales are unresolved by the effect of the TV norm.
A good estimation of λ results in an ill-conditioned Euler-Lagrange eq:

$$\nabla \cdot \frac{\nabla u}{|\nabla u|} + \lambda(f - u) = 0$$

E-L eq. ill-defined for $\nabla u = 0$

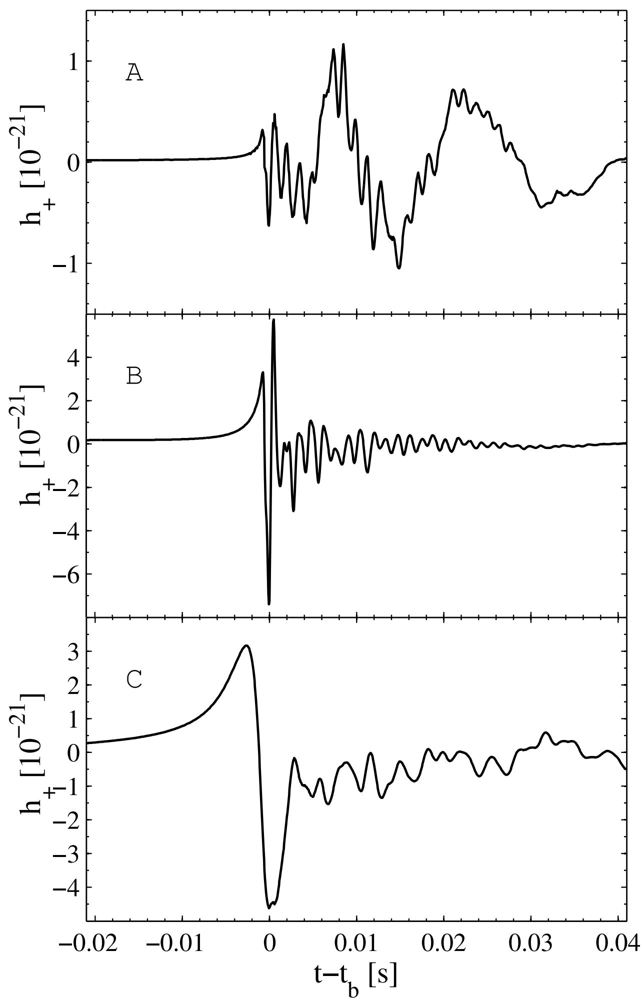
Solution: to slightly perturb the TV functional (regularized ROF model)

$$u = \arg\min_u \left\{ \mathrm{TV}_\beta(u) + \frac{\lambda}{2} ||f - u||_{L_2}^2 \right\}, \ \mathrm{TV}_\beta(u) \equiv \int \sqrt{|\nabla u|^2 + \beta}$$

Associated Euler-Lagrange equation is elliptic and non-degenerate.
Approximate solution can be obtained by e.g. a nonlinear Gauss-Seidel iterative procedure.

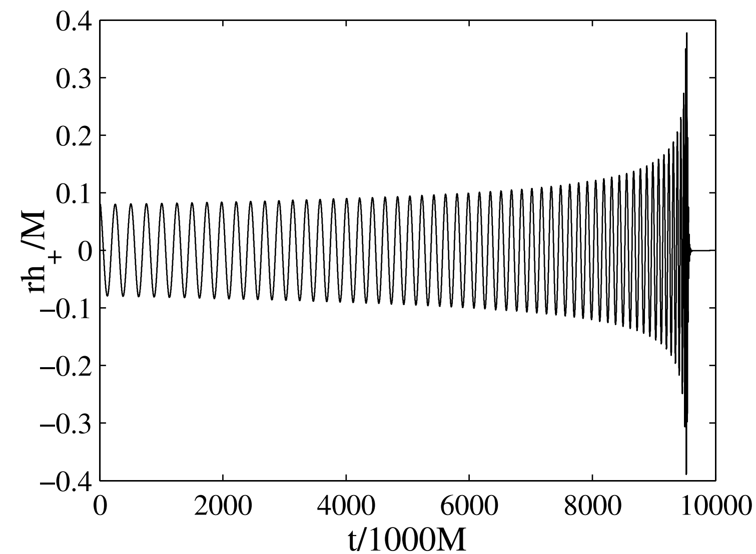
CCSN catalog
(Dimmelmeier+ 2008)

Magneto-rotational mechanism
128 waveforms
Short duration



BBH catalog
(SXS Collaboration Mroué+ 2013)

174 waveforms
Frequency and amplitude increase with time
Long duration



Signals injected in Gaussain noise provided by the LAL library from LSC.

Advanced LIGO proposed broadband configuration.

Torres-Forné+ 2014

Regularization parameter estimation

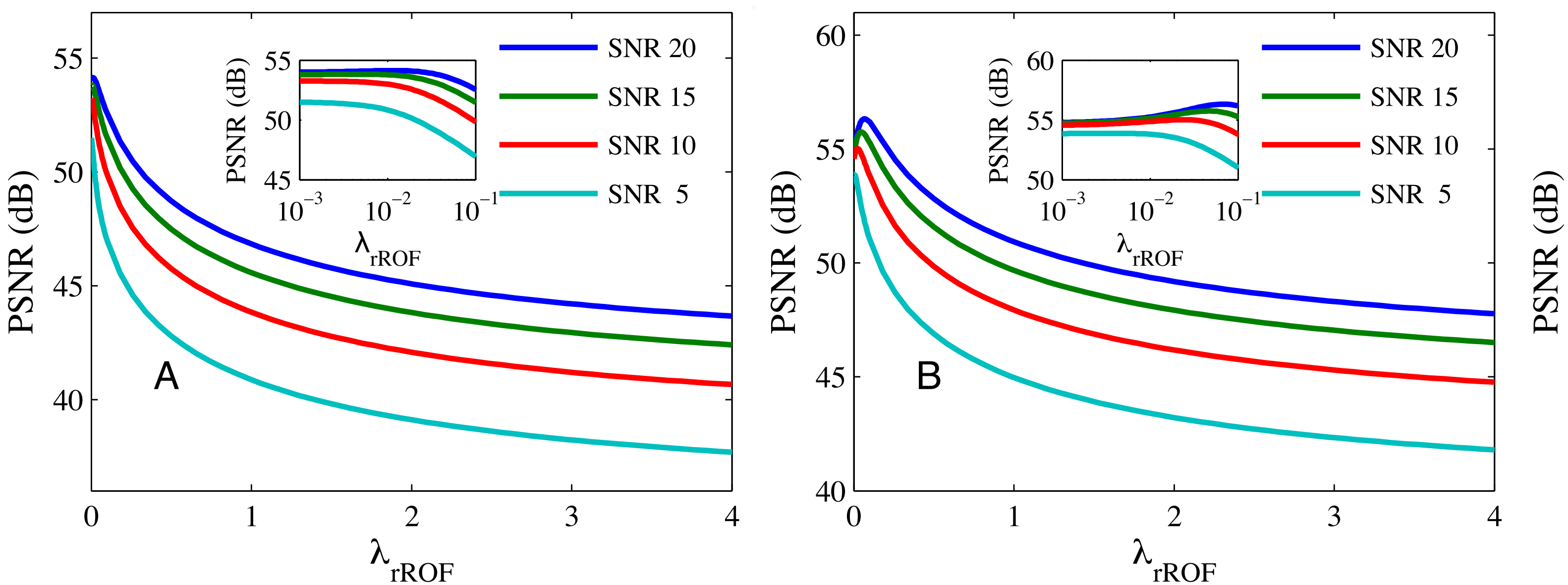
We must determine the **optimal value** of the regularization parameter that produces the best results. Heuristic search when the standard deviation of the noise is not known.

λ_{opt} used to denoise GW signals under different SNR conditions.

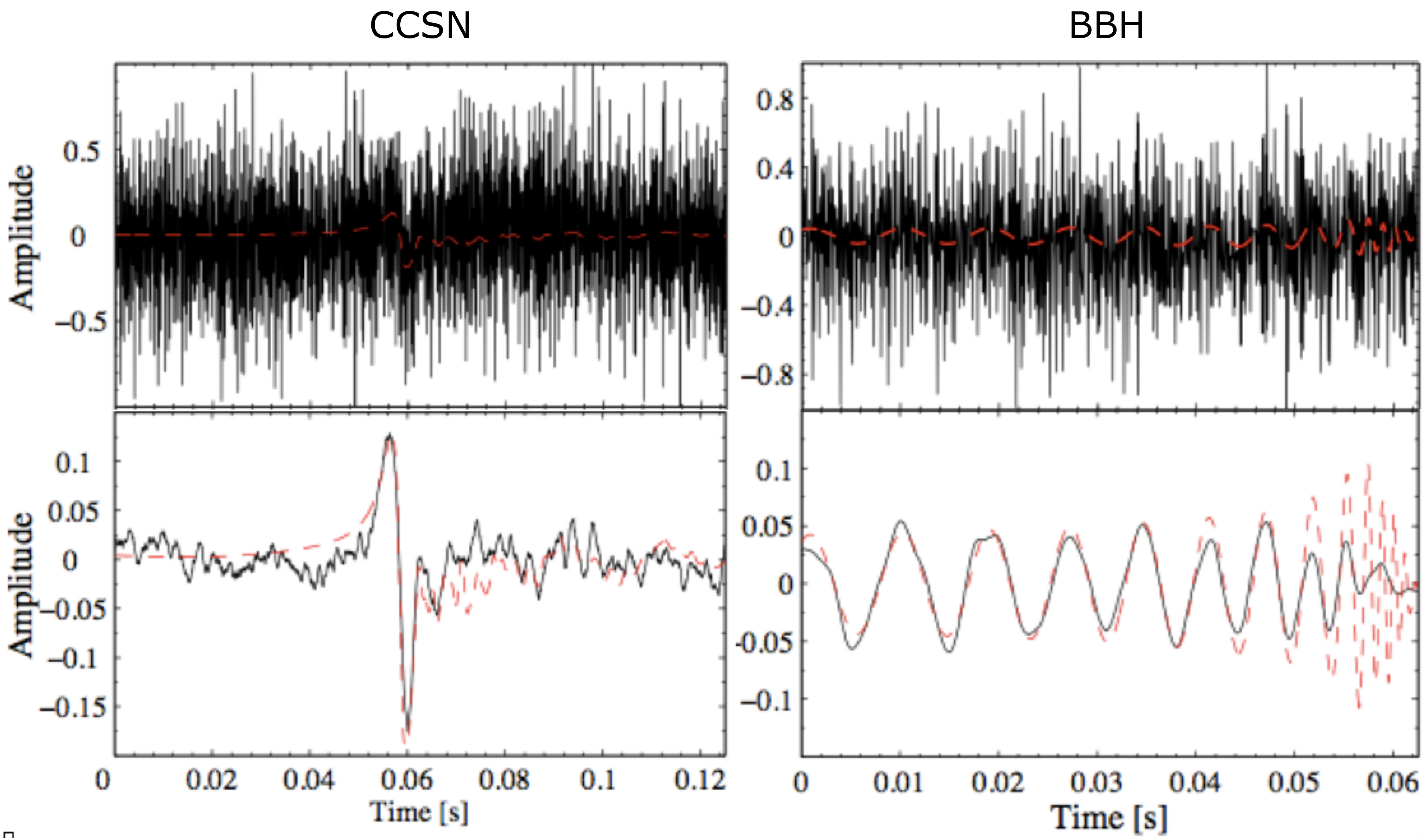
Quality estimators:

$$\mathrm{MSE} \equiv \frac{1}{n}\sum_{i}^{n}(\hat{y}_i - y_i)^2 \longrightarrow \mathrm{PSNR(dB)} = 10\log_{10}\left(\frac{n}{\mathrm{MSE}}\right)$$

$$\mathrm{SSIM}(x,y) \equiv \frac{(2\mu_x\mu_y + c_1)(2\sigma_{xy} + c_2)}{(\mu_x^2 + \mu_y^2 + c_1)(\sigma_x^2 + \sigma_y^2 + c_2)} \qquad \mathrm{DSSIM}(x,y) \equiv \frac{1-\mathrm{SSIM(x,y)}}{2}$$



Results for Gaussian noise



SNR=20

Torres-Forné+ 2014

TV-denoising with Advanced LIGO data

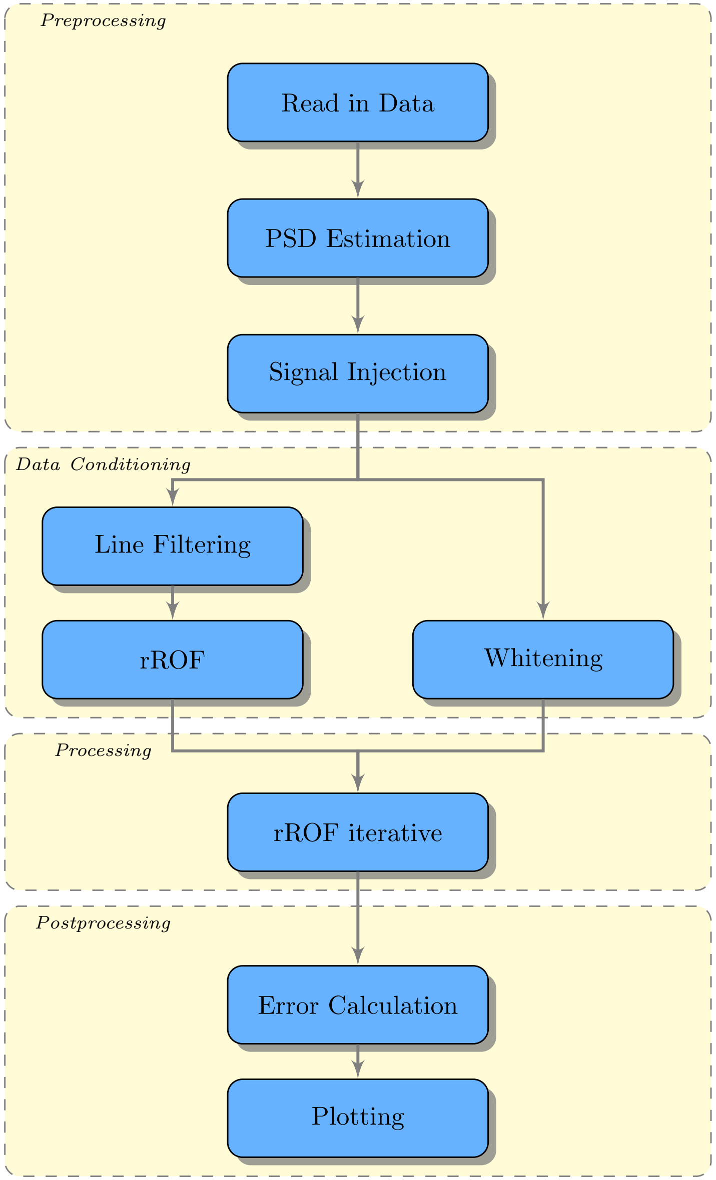
Detector noise is not Gaussian (e.g. sources of narrow-band noise such as the electric power, mirror suspension resonances, calibration lines, etc).

Our data: 10 chunks of data of 5 s each randomly chosen from the entire O1 data from LLO.

Data must be first pre-conditioned to make the noise flat in frequency (**whitening**).

Simple approach: remove spectral lines and filter low frequencies.

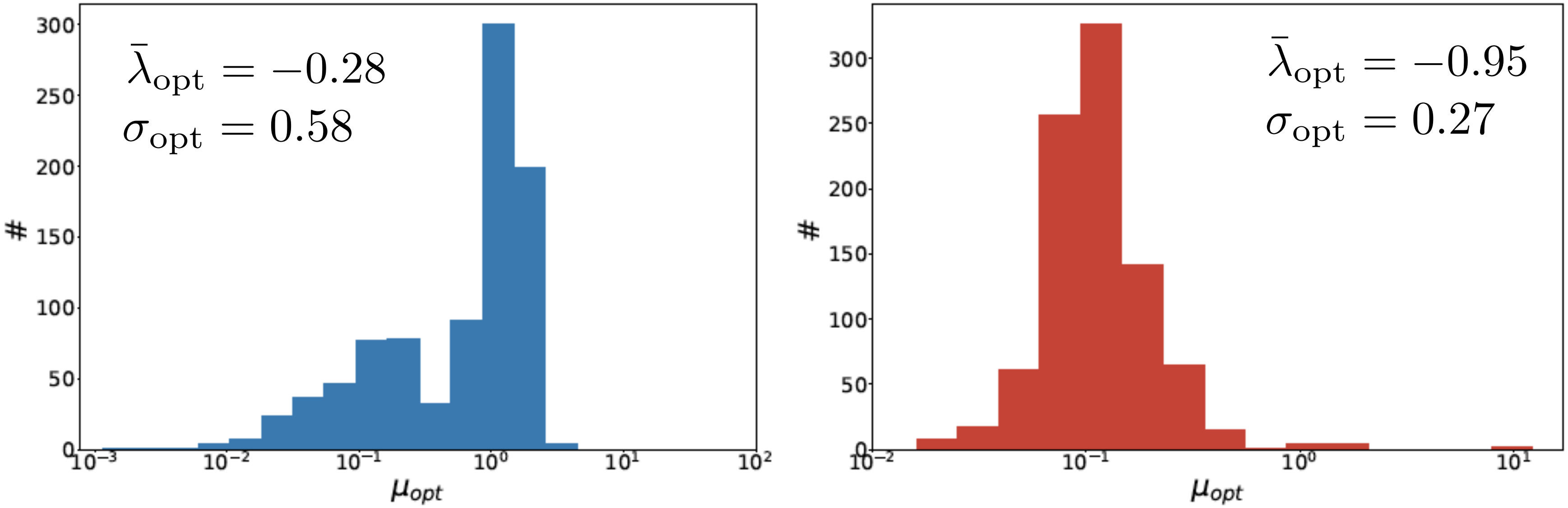
Sophisticated approach: transform the colored noise into white noise using an autoregressive (AR) model (Cuoco et al 2001).



Regularization parameter estimation

NR signals from CCSN and BBH catalogs injected into O1 data (10 different random GPS times).

Histograms of optimal values of λ (maximum SSIM)



30 CCSN signals
@ 5, 10, 20 kpc

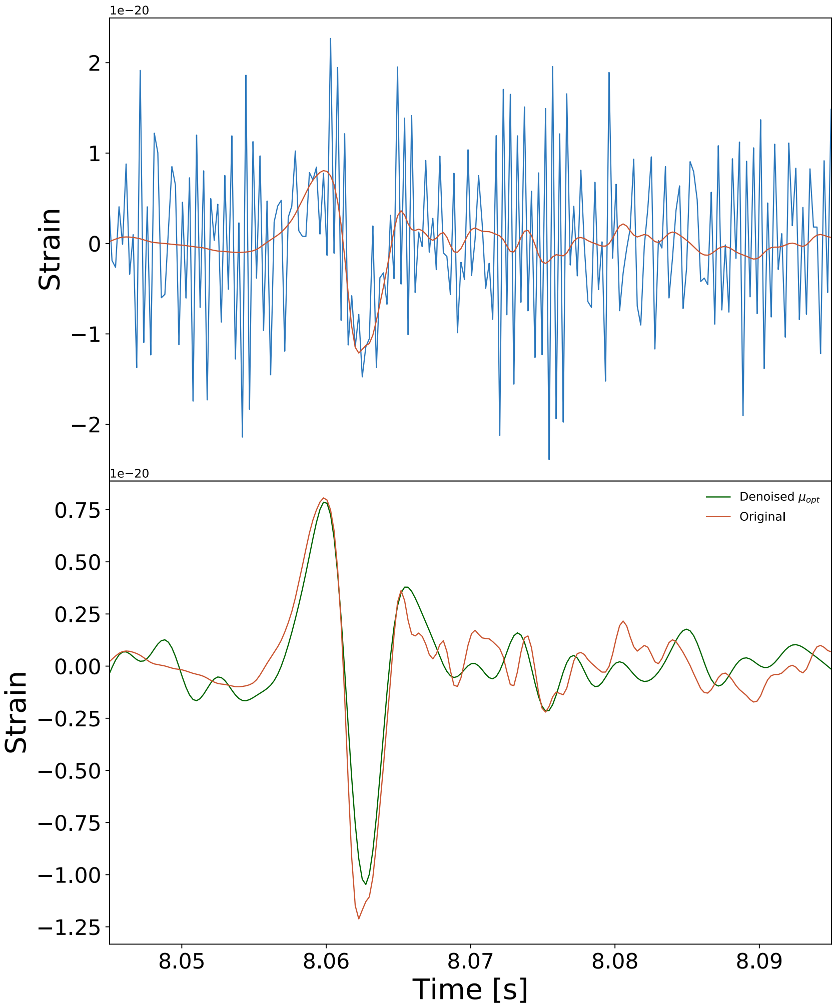
100 BBH signals
@ 400, 800, 1000 Mpc

We follow two approaches:

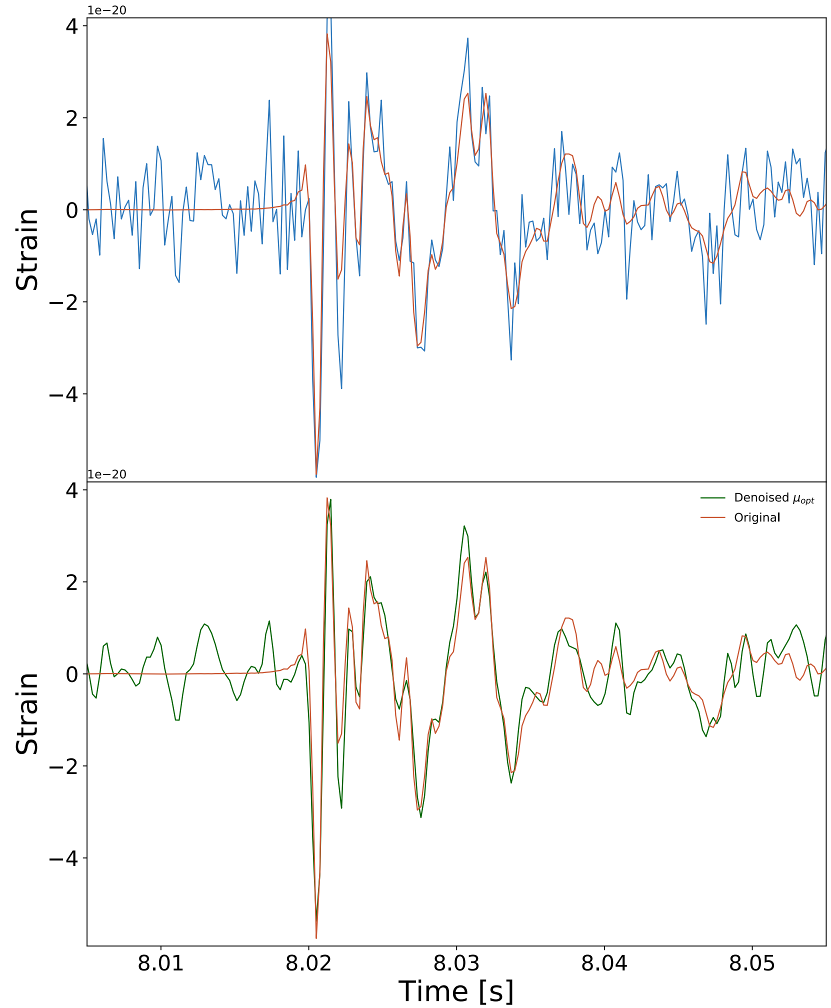
1. Use the mean value of regularization parameters for all waveforms.
2. Use the average of 20 different values sampled from a Gaussian distribution with same mean and variance than the histograms).

Results for real noise: CCSN waveforms

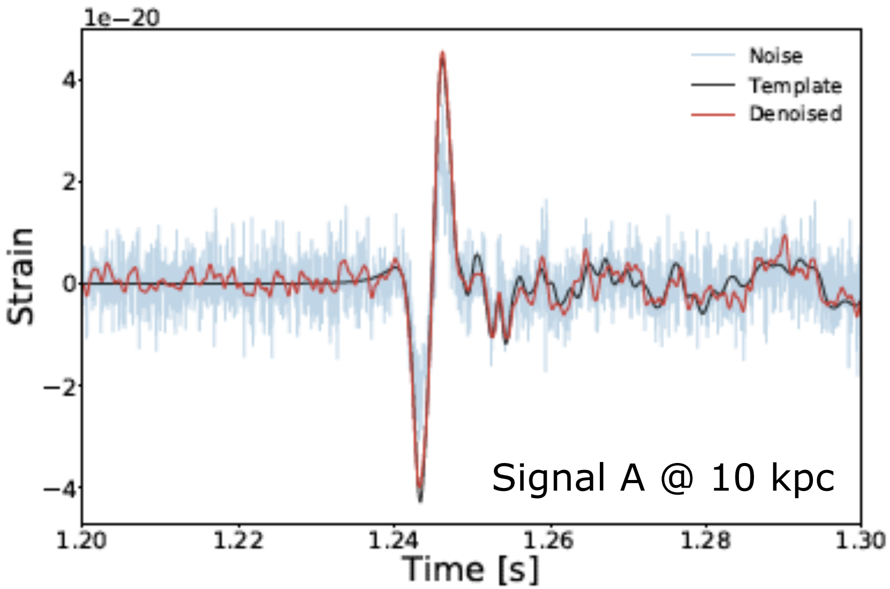
Filtering



AR-whitening



Results for real noise: CCSN waveform



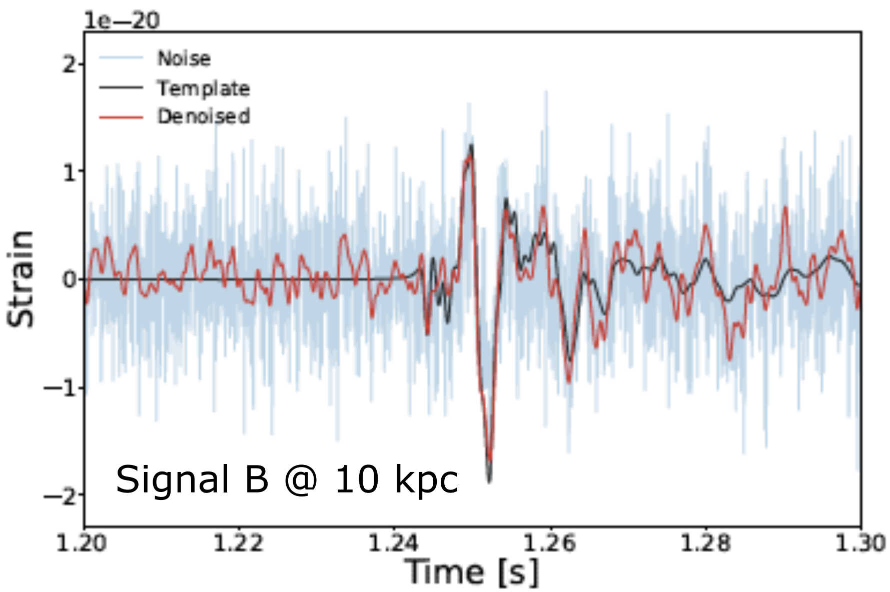


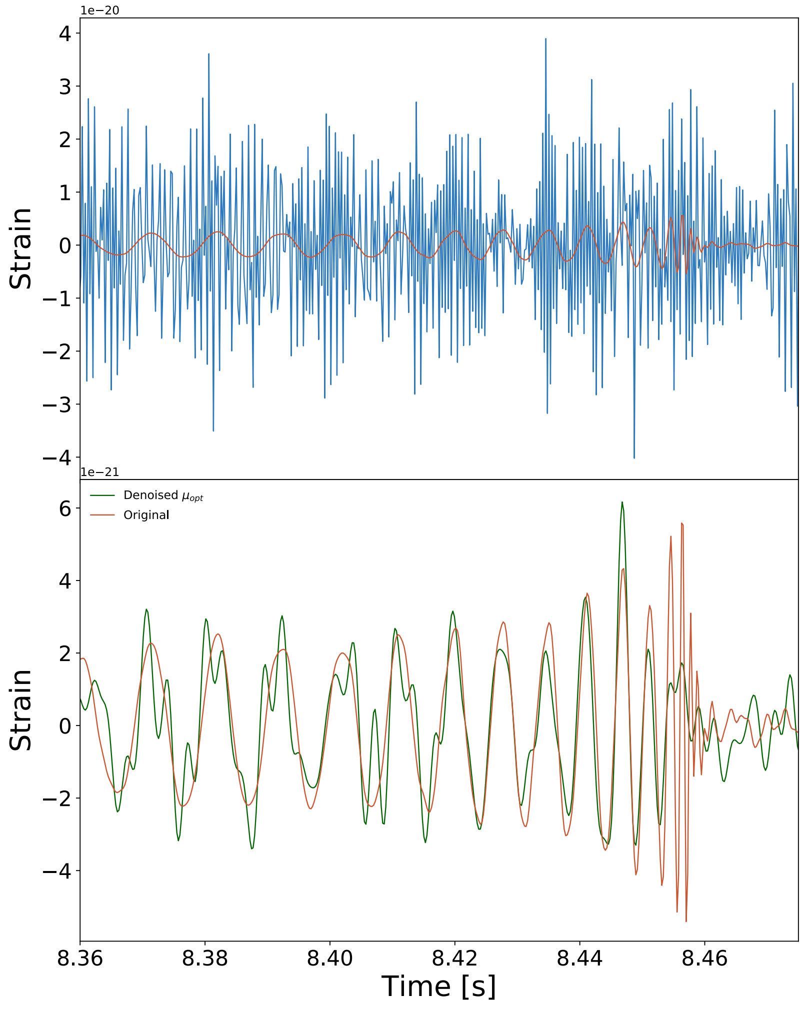
TABLE I. Values of the SSIM index for CCSN waveforms when using the optimal value of the regularization parameter for each signal, $\mu_{\rm opt}$; the mean value for all signals, $\bar{\mu}$; and multiple values, $\mu_{\rm m}$. The final column, Ref, indicates the SSIM index computed for the signal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

		SSIM index			
Signal	Distance (kpc)	$[\mu_{\rm opt}]$	$[\bar{\mu}]$	$[\mu_{\rm m}]$	[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

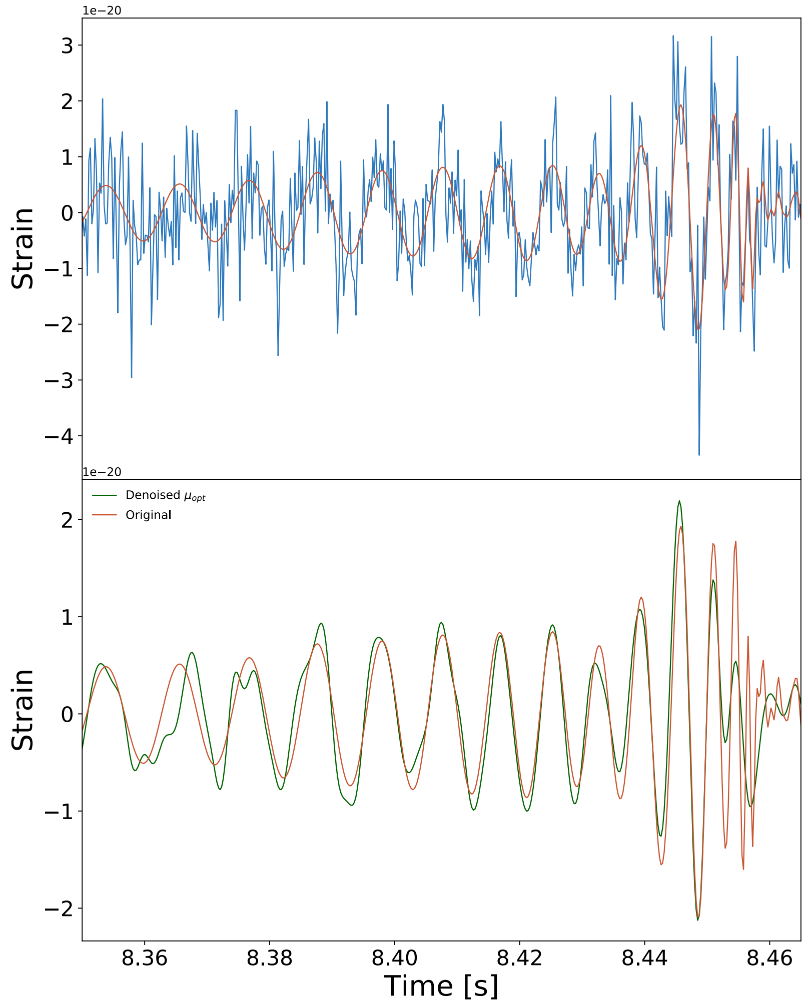
Not very strong dependence on the regularization parameter.

Results for real noise: BBH waveform

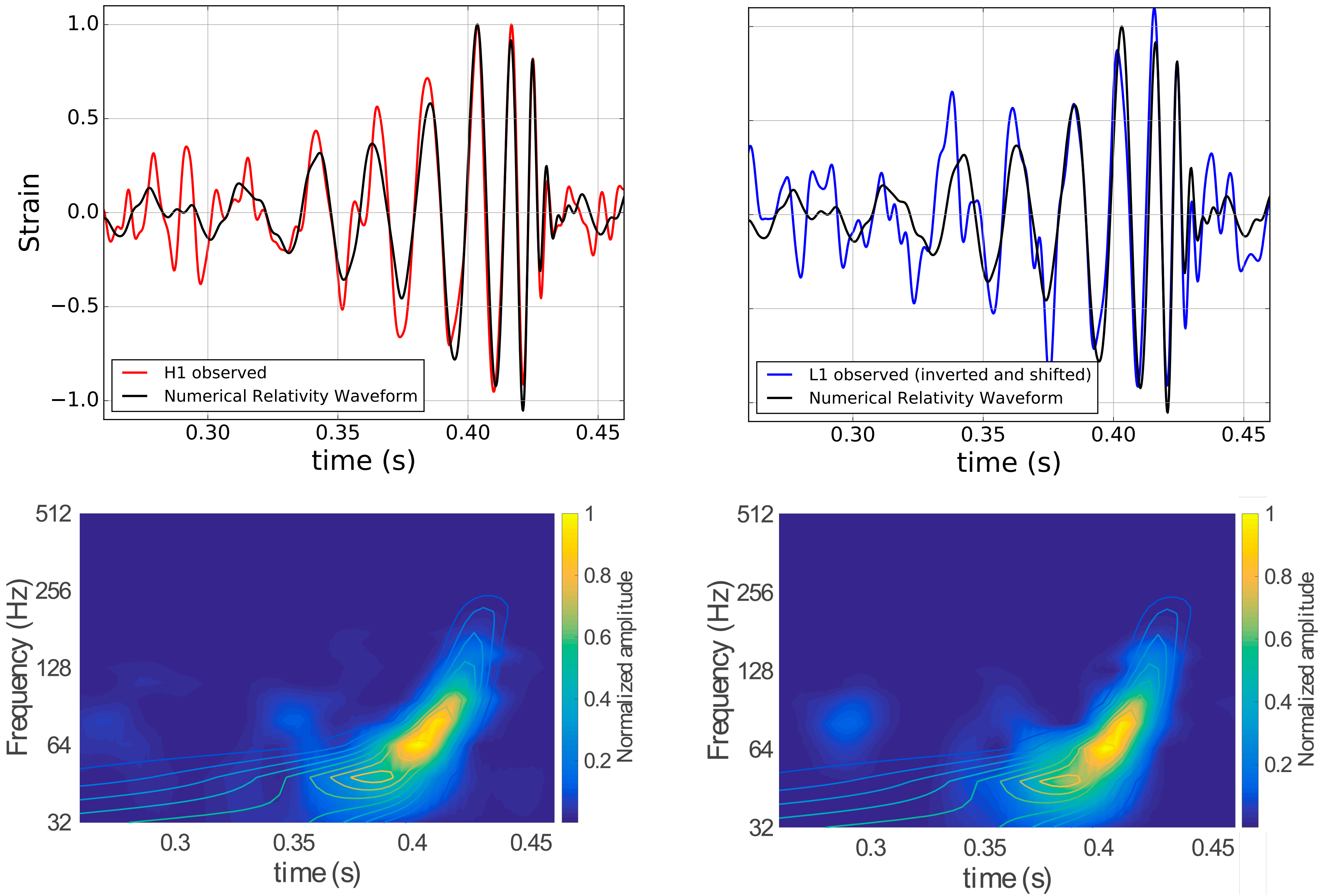
Filtering



AR-whitening

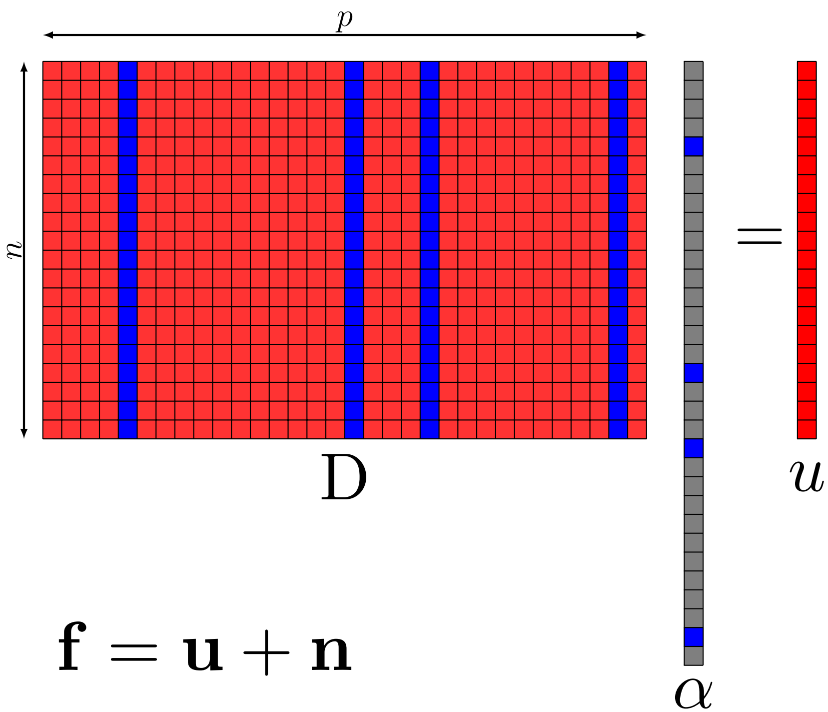


Results for GW150914



Denoising of GWs with Dictionary Learning

Growing interest in the problem of the sparse representation of signals. The use of sparse representations via **learned dictionaries** has proved to be very effective for signal denoising problems.



$\mathbf{D} \in \mathbb{R}^{n \times p}$

$p \equiv$ number of atoms

$n \equiv$ number of samples

Dictionary: Matrix of prototype signals (atoms).

Signals are described by sparse linear combinations of the atoms using a few coefficients.

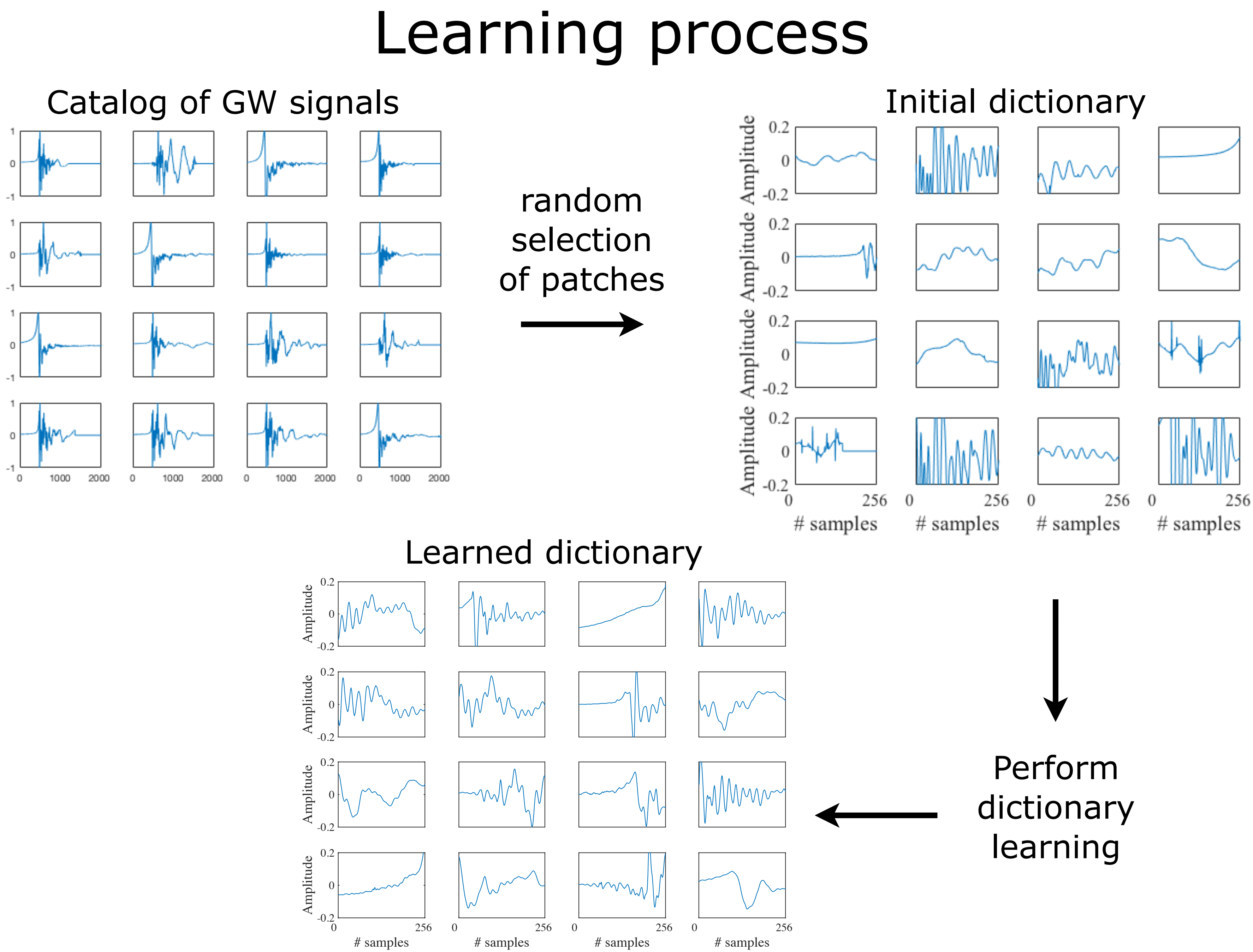
$$\mathbf{f} = \mathbf{u} + \mathbf{n}$$

$$\alpha = \arg\min_{\alpha} \left\{ ||\mathbf{D}\alpha - \mathbf{f}||_2^2 + \lambda ||\alpha||_1 \right\} \qquad \mathbf{u} = \mathbf{D}\alpha$$

LASSO (Tibsibirani R. JSOR, B, 58. 1996)

[Least Absolute Shrinkage and Selection Operator]

LASSO problem solved using the Split-Bregman algorithm


Learning process
Catalog of GW signals
random selection of patches
Initial dictionary
Amplitude
samples
Perform dictionary learning
Learned dictionary
Amplitude
samples

Dictionary learning problem

1. We start by considering a finite number of **training signals**: *m* patches of length *n*. $\mathbf{U} = [\mathbf{u}_1, \cdots, \mathbf{u}_m] \in \mathbb{R}^{n \times m}$ $\quad n \ll m$

2. To obtain the **trained dictionary**, we add the dictionary matrix **D** as a variable in the minimization problem:

$$\alpha = \arg\min_{\alpha, \mathbf{D}} \frac{1}{n} \sum_{i=1}^{m} ||\mathbf{D}\alpha_i - \mathbf{u}_i||_2^2 + \lambda ||\alpha_i||_1$$

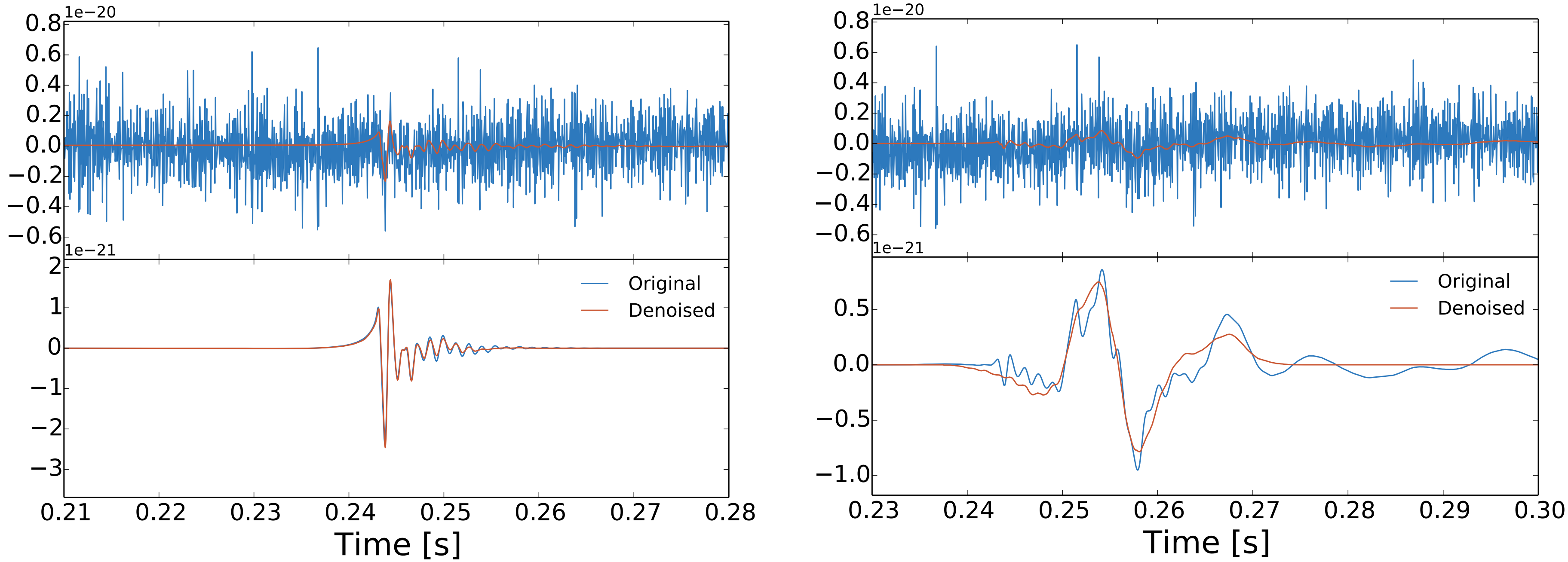
α_i *i*-th row of $\alpha \in \mathbb{R}^{p \times n}$ Contains the coefficients of the sparse representation of each atom in the dictionary.

3. Dictionary updated using a block-coordinate descent method (Mairal et al 2009, Tseng 2001).

CCSN and BBH gravitational wave catalogs. 80% of waveforms to train the dictionary, 15% for method validation, and 5% to test algorithm.

Signals shifted to be aligned with minimum peak (CCSN) or maximum peak in the merger (BBH). 2048 samples to train the dictionary.

CCSN signals (Dimmelmeier+ 2008)

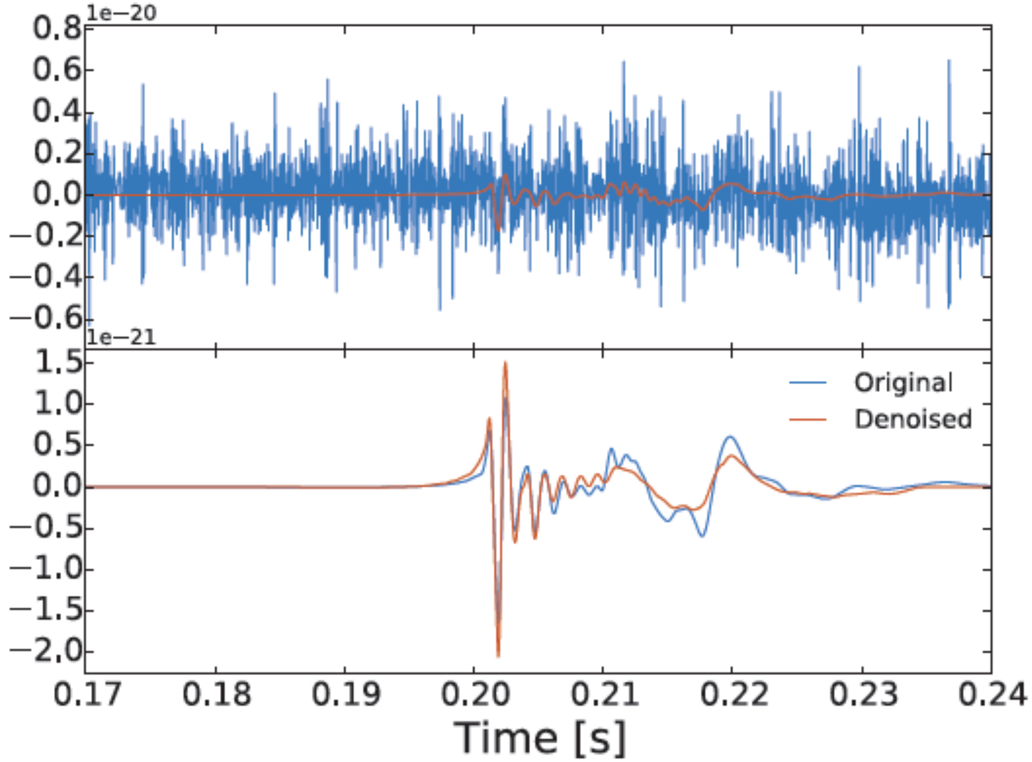


SNR 20, Gaussian noise Torres-Forné+ (2017)

Peaks well recovered PNS ringdown smoothed out	Small oscillations lost Broad morphology captured
$\mathrm{SSIM} = 0.98$	$\mathrm{SSIM} = 0.67$

Best (left) and worst (right) denosing results for a given λ and noise realisation.

CCSN signal from a different catalog



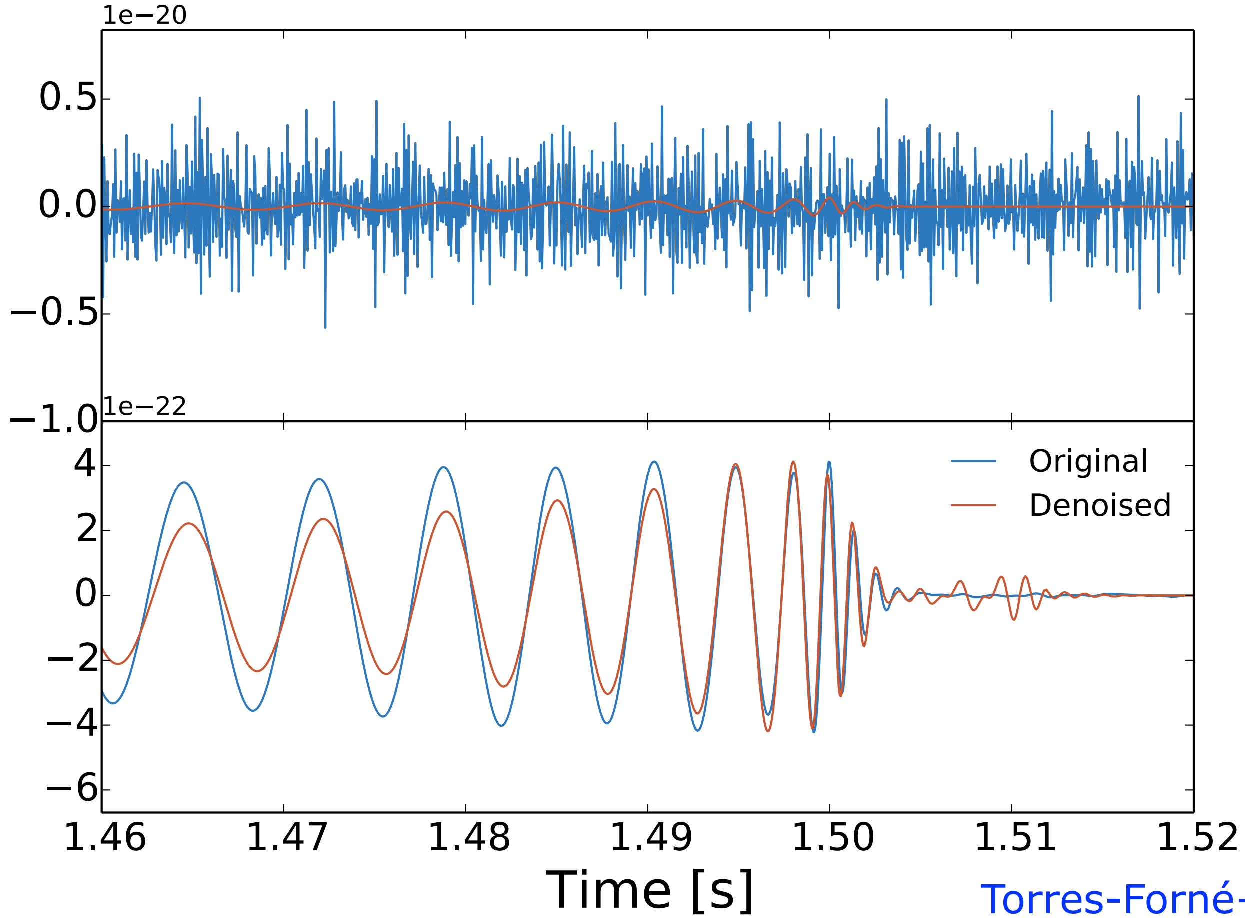
Torres-Forné+ (2017)

CCSN signal from the catalog of Abdikamalov et al (2014).

Accurate results:

Similar structure (burst)
Some peaks missing (PNS damped oscillations)

BBH signals (SXS Collaboration Mroué+ 2013)



Torres-Forné+ (2017)

SNR 20
Gaussian noise
Random signal from test set
Random time of arrival

SSIM index = 0.86

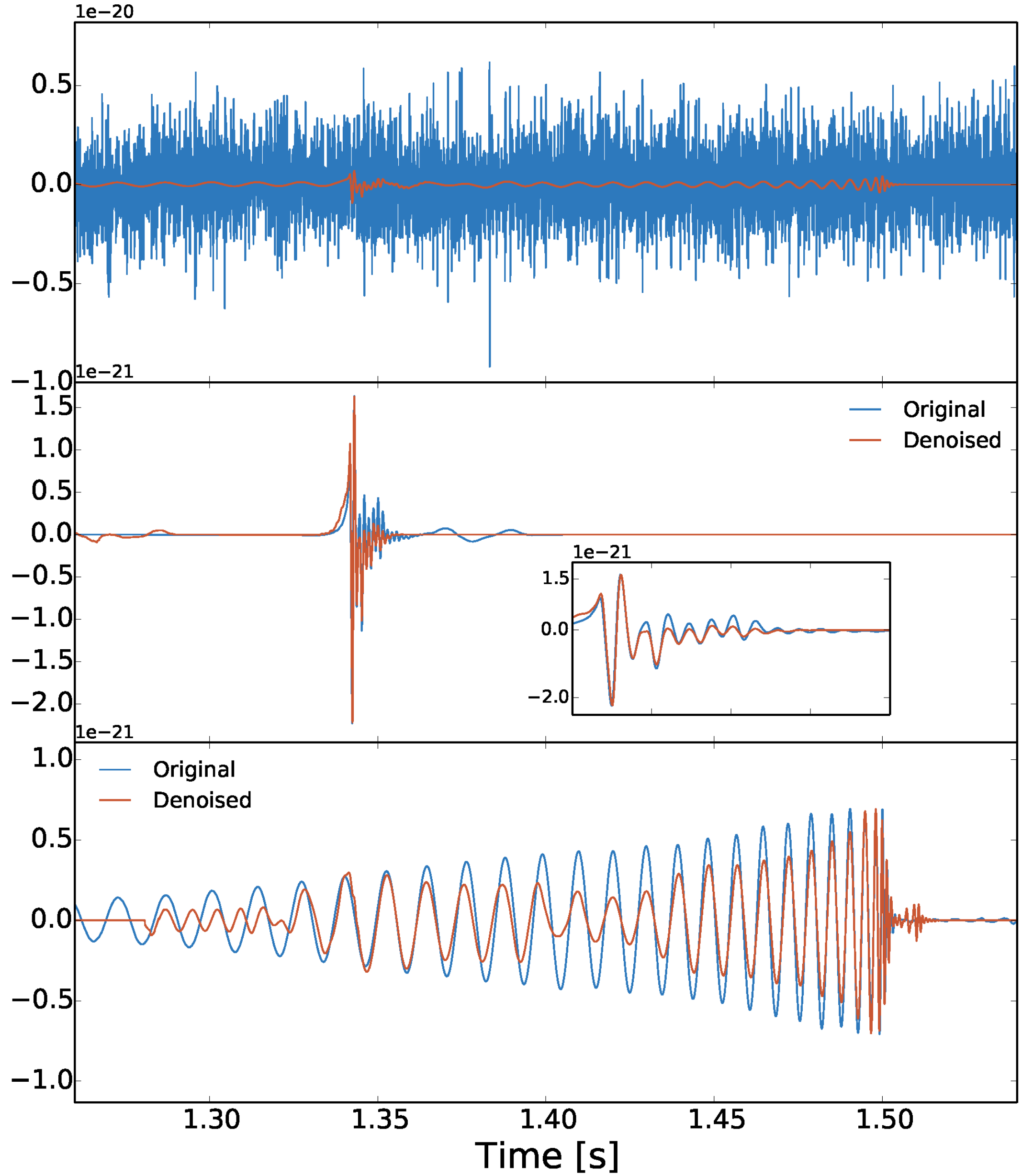
Dictionary specifically designed to recover the merger part.
Phase very well captured.

Signal discrimination

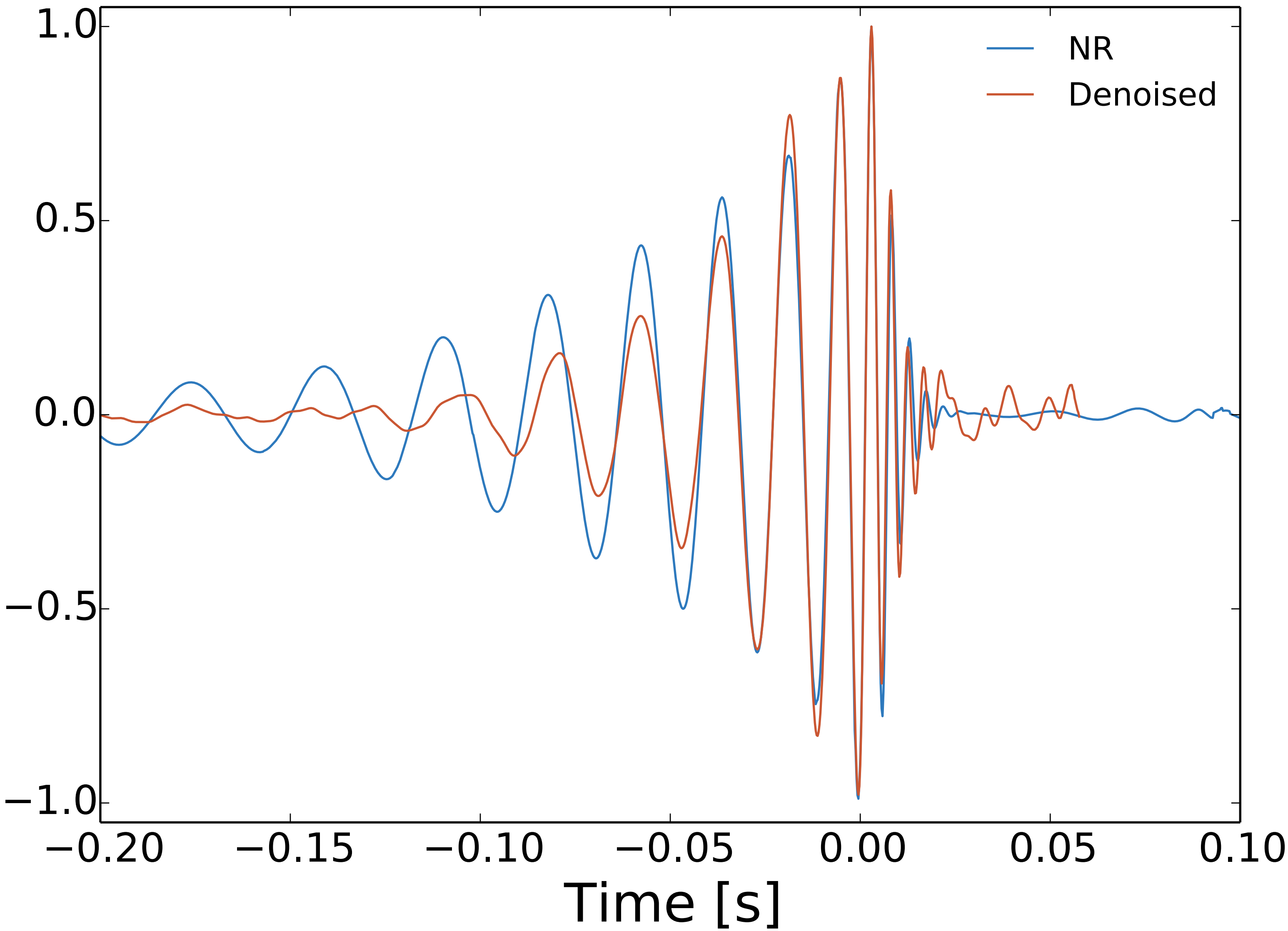
Test the dictionary when dealing with signals different from the type they are designed for.

Use both dictionaries independently.

Each dictionary discriminates well between the type of signal.



Results for GW150914



Glitch classification

In the LVC there exist diverse strategies to classify glitches in the detectors:

Powell et al (2015, 2017):

- PCAT: Principal Component Analysis for Transients. Uses PC coefficients to classify glitches using a Gaussian Mixture Model.
- PC-LIB: Based on LAL-Inference. Computes Bayes factor for glitch selection. Supervised classification.
- WDF-ML: Wavelet Detection Filter + (unsupervised) ML algorithm (GMM).

Zevin et al (2017): Gravity Spy, Zooniverse Platform. Citizen science + ML.

Mukund et al (2017): Difference Boosting Neural Network (supervised Bayesian classifier).

George et al (2018): Deep Learning + Transfer Learning.

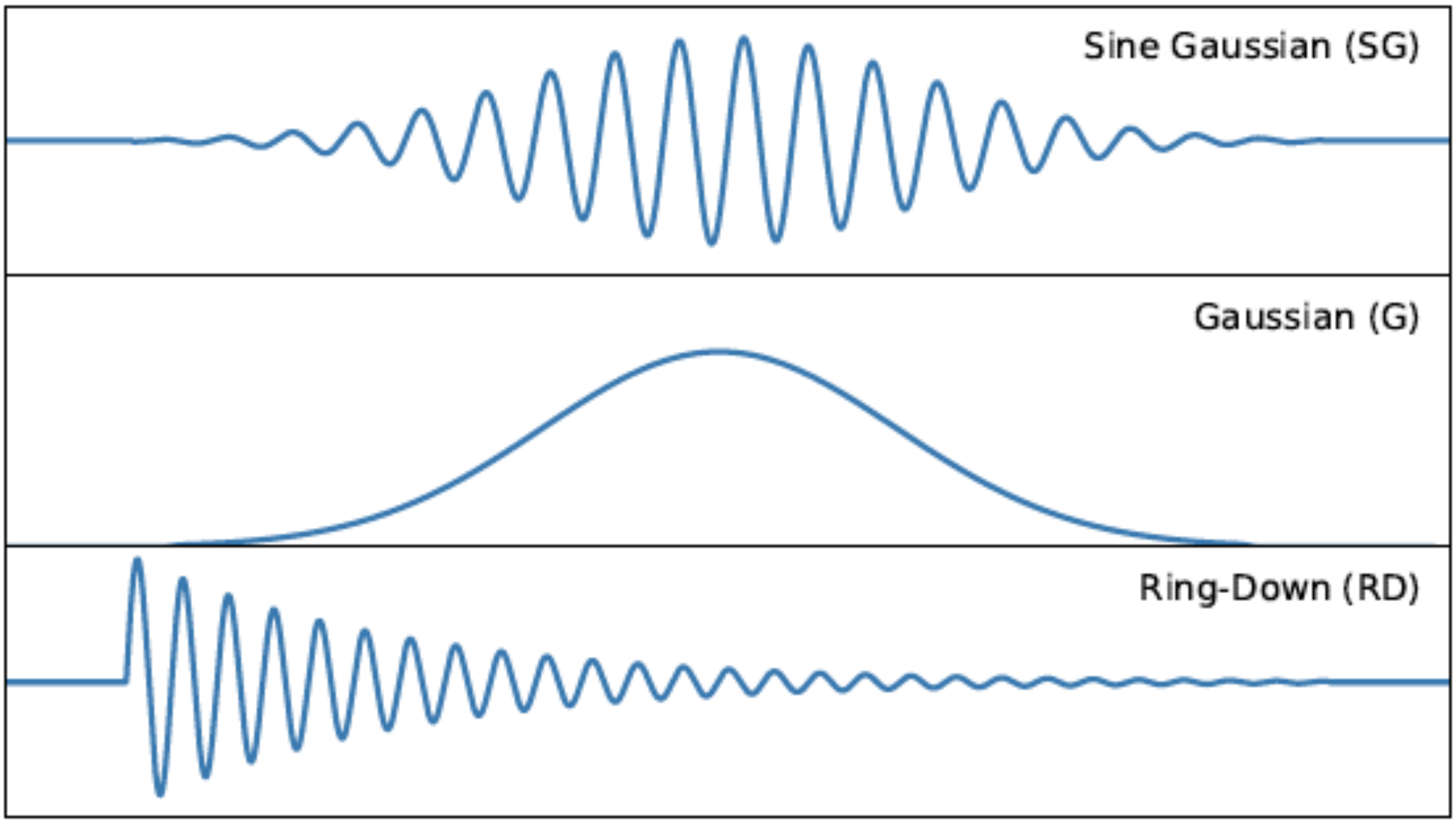
Razzano & Cuoco (2018): Convolutional Neural Networks to classify glitches from their spectrograms (time-frequency evolution)

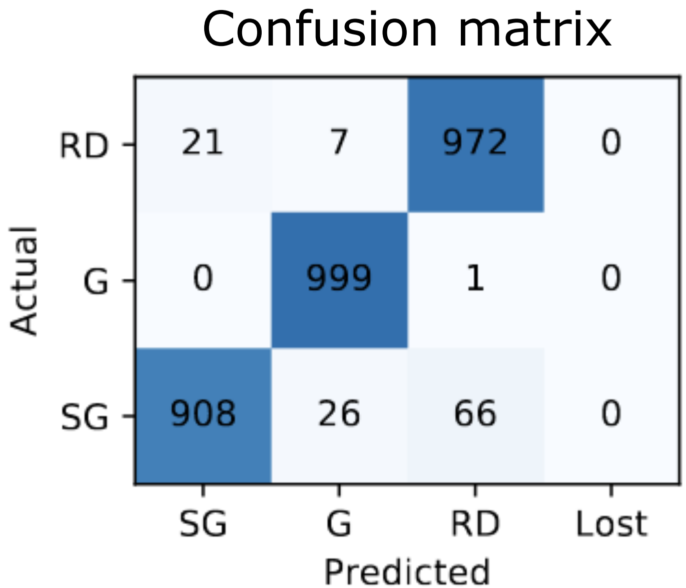
Llorens-Monteagudo et al (2018): Dictionary learning.

Simulated glitches embedded in Gaussian noise to simulate the background noise of advanced LIGO in its broadband configuration.

Data set of 3000 simulated glitches of three different waveform morphologies, comprising 1000 glitches per morphology.

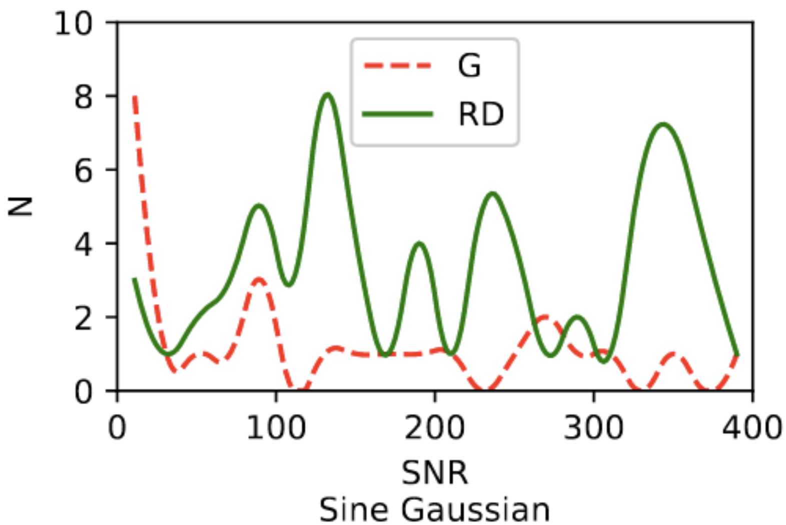
3 simple types of glitch morphologies (following Powell et al, 2015)

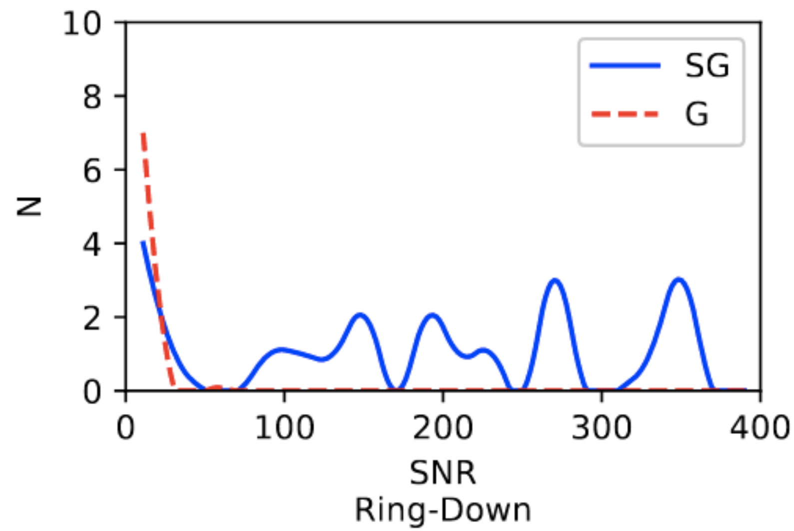




Llorens-Monteagudo+ (2018)

Out of 3000 glitches, 2879 (96%) are correctly classified.





Performance barely decreases with SNR (down to SNR~10).

Most misclassified glitches have highest and lowest frequencies. More affected by noise than intermediate frequencies.

Summary

- We have discussed **variational methods** for minimization problems based on the **TV-norm** in the context of gravitational-wave signals.

- **Novel strategy** in the field.

- We have shown that TV algorithms can be useful in the field of Gravitational-Wave Astronomy as a tool to **remove noise**.

- ROF model has been tested both for Gaussian noise and with Advanced LIGO data.

- We have discussed a **machine-learning algorithm** based on **dictionaries**.

- We have shown that they can be successfully applied for both, gravitational-wave denoising and **glitch classification**.

Thanks for your attention!

Upcoming GR22/Amaldi13 conference in Valencia

www.gr22amaldi13.com

Topics from the G2NET Cost Action will be discussed at the GR22/Amaldi13 conference