Finding a Whisper in a Storm: An Intuitive Walkthrough of LIGO-Style Matched Filtering

September 4, 2025

Abstract

Gravitational waves are faint ripples in spacetime, so faint that even the most sensitive instruments see mostly noise. This document tells the story of how data analysis—especially matched filtering—turns noisy strain measurements into confident detections. We build a minimal, reproducible "toy detector": a synthetic chirp hidden inside random noise. We then recover it using the core idea of LIGO's searches, with clear intuition about "Gaussian" noise and why correlation is the right question to ask of the data.

1 Generative Model and Hypotheses

We model the observed strain $d(t)$ as either noise alone or noise plus a signal:

$$\mathcal{H}_0 : \; d(t) = n(t), \tag{1}$$

$$\mathcal{H}_1 : \; d(t) = h(t) + n(t), \tag{2}$$

where $n(t)$ is a (roughly) stationary random process and $h(t)$ is a deterministic template.

If the noise is (approximately) Gaussian, the likelihood ratio can be expressed with the noise-weighted inner product:

$$(a|b) \equiv 4 \; \mathrm{Re} \int_0^\infty \frac{\tilde{a}(f)\, \tilde{b}^*(f)}{S_n(f)} \, \mathrm{d}f, \tag{3}$$

where tildes denote Fourier transforms and $S_n(f)$ is the one-sided PSD.

2 Matched Filtering and SNR

For a template h with relative time shift τ, the SNR time series is

$$\rho(\tau) \equiv \frac{(d|h_\tau)}{\sqrt{(h|h)}}, \qquad h_\tau(t) \equiv h(t - \tau). \tag{4}$$

Peaks of $\rho(\tau)$ indicate times where d contains h most strongly, with (3) downweighting noisy bands.

3 Whitening and Spectrograms

To visualize features and approximate the weighting in (3), we whiten the data:

$$\tilde{x}_{\mathrm{w}}(f) \equiv \frac{\tilde{x}(f)}{\sqrt{S_n(f)}}, \qquad x_{\mathrm{w}}(t) = \mathcal{F}^{-1}\big[\tilde{x}_{\mathrm{w}}(f)\big]. \tag{5}$$

We then compute an STFT with a long window and high overlap:

$$Z(t_k, f_m) \equiv \sum_t x_{\mathrm{w}}(t)\, w(t - t_k)\, e^{-i2\pi f_m t}, \tag{6}$$

and plot $10 \log_{10} |Z|^2$ after subtracting, for each f_m, the median over t_k:

$$S_{\mathrm{dB}}^{\mathrm{ms}}(t_k, f_m) \equiv 10 \log_{10}\big(|Z|^2\big) \;-\; \mathrm{median}_{t_k}\big[\, 10 \log_{10}\big(|Z|^2\big)\, \big]\,. \tag{7}$$

Finally, we clip color limits to the 95th–99.5th percentiles of $S_{\mathrm{dB}}^{\mathrm{ms}}$ to enhance contrast.

4 Figures

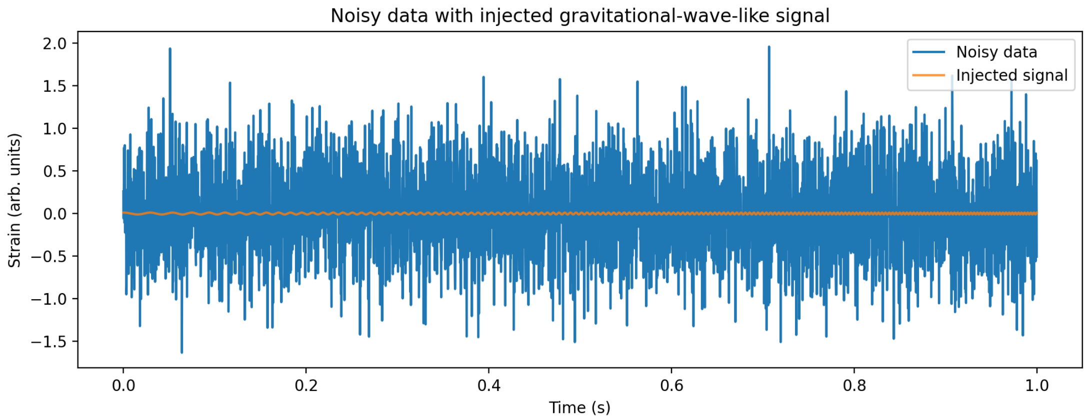

Figure 1: Noisy data with injected chirp (unit-energy template scaled by an amplitude) overlaid.

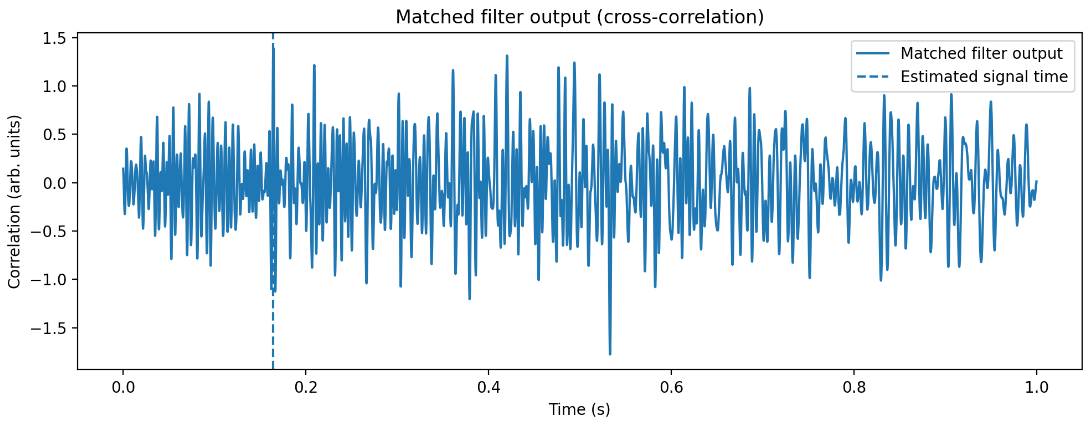

Figure 2: Matched-filter output (toy correlation) vs. time. The dashed line marks the peak (estimated arrival time).

Parameters

Table 1 lists the parameters used in the experiment.

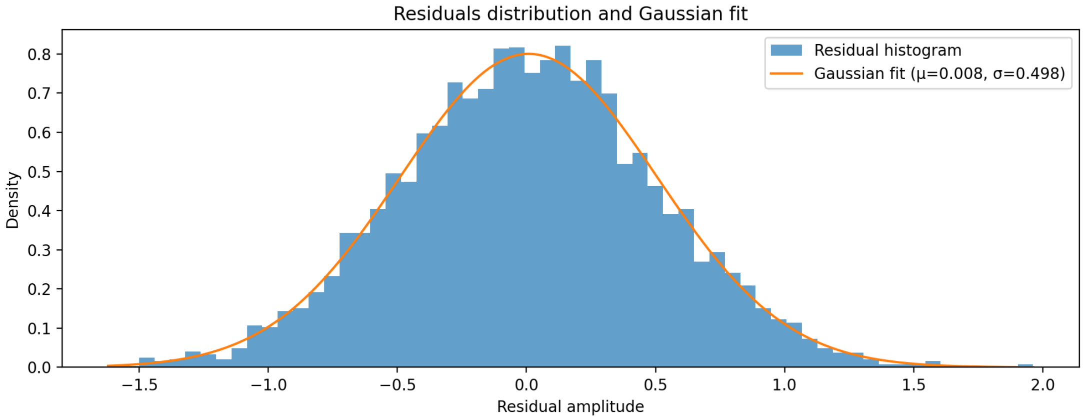

Figure 3: Residual histogram with Gaussian fit. This is the operational meaning of “Gaussian”: a bell-shaped distribution with predictable tails.

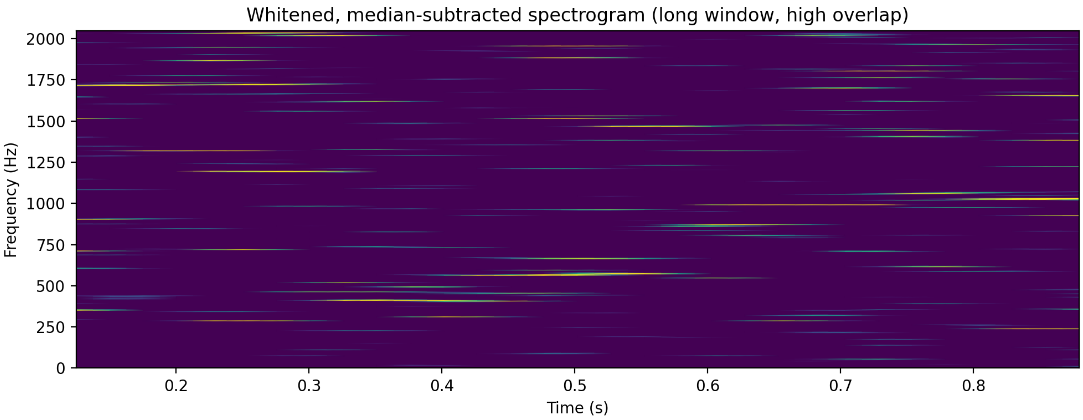

Figure 4: Whitened, median-subtracted spectrogram using a long Hann window with 90% overlap; color limits clipped to the 95th–99.5th percentiles.

Table 1: Parameters of the toy experiment.

Parameter	Value
Sampling rate	4096 Hz
Duration	1.00 s
Chirp start frequency	30.0 Hz
Chirp end frequency	300.0 Hz
Signal amplitude	0.50
Noise std. dev.	0.50
Estimated arrival time	0.1643 s
Approx. SNR (toy)	3.38
Residual ,	0.008, 0.498