An F-statistic based multi-detector veto for detector artifacts in continuous-wave gravitational wave data

D. Keitel, R. Prix, M.A. Papa, M. Siddiqi

Albert-Einstein-Institut, Hannover and Golm, Germany

Abstract

Continuous gravitational waves (CW) are expected from spinning neutron stars with non-axisymmetric deformations. A network of interferometric detectors (LIGO, Virgo and GEO600) is looking for these signals. They are predicted to be very weak and retrievable only by integration over long observation times.
One of the standard methods of CW data analysis is the multi-detector F-statistic. In a typical search, the F-statistic is computed over a range in frequency, spindown and sky position, and the candidates with highest F values are kept for further analysis. However, this detection statistic is susceptible to a class of noise artifacts, strong monochromatic lines in a single detector.
By assuming an extended noise model - standard Gaussian noise plus single-detector lines - we can use a Bayesian odds ratio to derive a generalized detection statistic, the line veto (LV-) statistic. In the absence of lines, it behaves similarly to the F-statistic, but it is much more robust against line artifacts.
In the past, ad-hoc post-processing vetoes have been implemented in searches to remove these artifacts. Here we provide a systematic framework to develop and benchmark this class of vetoes. We present our results from testing this LV-statistic on simulated data. Furthermore, we plan to use it on the current Einstein@Home distributed computing analysis of LIGO S6 data.

Standard detection statistic

We want to detect a signal at given frequency, spindown and sky position.

- Assuming Gaussian noise only, we have two hypotheses:

  $\mathcal{H}_G$: pure noise: $\quad \boldsymbol{x}(t) = \boldsymbol{n}(t)$

  $\mathcal{H}_S$: noise plus signal (with amplitude parameters $\mathcal{A}$) : $\quad \boldsymbol{x}(t) = \boldsymbol{n}(t) + \boldsymbol{h}(t;\mathcal{A})$
- The Bayesian approach allows hypothesis testing by computing the odds ratio, and marginalizing over the unknown parameters $\mathcal{A}$:

$$O_{SG}(\boldsymbol{x}) \equiv \frac{P(\mathcal{H}_S|\boldsymbol{x})}{P(\mathcal{H}_G|\boldsymbol{x})} \propto \int \mathcal{L}(\boldsymbol{x};\mathcal{A})\, P(\mathcal{A}|\mathcal{H}_S)\, d\mathcal{A} \qquad (1)$$

- Here, $\mathcal{L}(\boldsymbol{x};\mathcal{A})$ the likelihood ratio:

$$\mathcal{L}(\boldsymbol{x};\mathcal{A}) \equiv \frac{P(\boldsymbol{x}|\mathcal{H}_S\,\mathcal{A})}{P(\boldsymbol{x}|\mathcal{H}_G)}$$

- The marginalization integral in Eq. (1) can be carried out analytically (if we assume specific priors on $\mathcal{A}$, see [3]). We obtain

$$O_{SG}(\boldsymbol{x}) \propto e^{\mathcal{F}(\boldsymbol{x})} \qquad (2)$$

- The detection statistic $\mathcal{F}(\boldsymbol{x})$ is calculated directly from multi-detector interferometer strain and detector position time-series. [1, 2]

Derivation of the Line Veto

Problem with $O_{SG}(\boldsymbol{x}) \propto e^{\mathcal{F}(\boldsymbol{x})}$: quasi-monochromatic, stationary detector artifacts ("lines") look more like $\mathcal{H}_S$ than $\mathcal{H}_G$ and will result in large values for O_{SG}.

- So we add an *alternative noise hypothesis* $\mathcal{H}_L$ that fits lines in single detectors better than the multi-detector coherent $\mathcal{H}_S$:

$$\mathcal{H}_L^X : x^X(t) = n^X(t) + h^X(t;\mathcal{A}) \quad \text{for detector } X$$

- For two detectors $X = 1, 2$:

$$\mathcal{H}_L : \left(\mathcal{H}_L^1 \text{ and } \mathcal{H}_G^2\right) \text{ or } \left(\mathcal{H}_G^1 \text{ and } \mathcal{H}_L^2\right)$$

$$\Longrightarrow P(\mathcal{H}_L) = P(\mathcal{H}_L^1)\, P(\mathcal{H}_G^2) + P(\mathcal{H}_G^1)\, P(\mathcal{H}_L^2) \qquad (3)$$

- Again using the "$\mathcal{F}$-statistic priors" and analytically maximizing over $\mathcal{A}$, we can compute

$$P(\mathcal{H}_L|\boldsymbol{x}) = P(\mathcal{H}_G|\boldsymbol{x})\, c_0 \left[l^1 e^{\mathcal{F}^1(x^1)} + l^2 e^{\mathcal{F}^2(x^2)}\right]$$

  with a prior $l^X \equiv P(\mathcal{H}_L^X)/P(\mathcal{H}_G^X)$ for the line density in detector X.

Now we can consider two possible applications:

- Assuming we already decided on $\mathcal{H}_S$ over $\mathcal{H}_G$ with the standard detection statistic $\mathcal{F}$, we use a new statistic to veto lines:

$$O_{SL}(\boldsymbol{x}) \equiv \frac{P(\mathcal{H}_S|\boldsymbol{x})}{P(\mathcal{H}_L|\boldsymbol{x})} \propto \frac{e^{\mathcal{F}(\boldsymbol{x})}}{l^1 e^{\mathcal{F}^1(x^1)} + l^2 e^{\mathcal{F}^2(x^2)}} \qquad (4)$$

- In a single step, we replace the standard $\mathcal{F}$-statistic by a new detection statistic, directly deciding between signal vs. extended noise hypotheses:

$$O_{SN}(\boldsymbol{x}) \equiv \frac{P(\mathcal{H}_S|\boldsymbol{x})}{P(\mathcal{H}_L|\boldsymbol{x}) + P(\mathcal{H}_G|\boldsymbol{x})} = \left[O_{SG}^{-1}(\boldsymbol{x}) + O_{SL}^{-1}(\boldsymbol{x})\right]^{-1}$$

$$\propto \frac{e^{\mathcal{F}(\boldsymbol{x})}}{\rho_{\max}^4/70 + l^1 e^{\mathcal{F}^1(x^1)} + l^2 e^{\mathcal{F}^2(x^2)}} \qquad (5)$$

  Here, $\rho_{\max}$ is the cutoff in an uniform prior on signal strength.

Discussion and features

- The new detection statistic, Eq. (5), downweights candidates which have higher single-detector than multi-detector $\mathcal{F}$-statistics, thereby penalizing lines.
- The prior cutoff $\rho_{\max}$ allows us to tune the detection statistic, determining how much discrepancy between detectors is attributed to Gaussian noise and how soon vetoing sets in. Further work on simulated data is necessary to choose this prior optimally.
- Another interesting result is that, in the special case $l^1 = l^2$, the line veto O_{SL} of Eq. (4) can be expressed as

$$\ln O_{SL}(\boldsymbol{x}) - \ln O_{SL}^{(0)} \approx \mathcal{F}(\boldsymbol{x}) - \max\{\mathcal{F}^1(x^1), \mathcal{F}^2(x^2)\} \qquad (6)$$

  We therefore recover an *ad-hoc* veto criterion, which has been used in post-processing of LIGO S5 data [4, 5], as a special case, namely

$$\text{veto if } \max\{\mathcal{F}^1, \mathcal{F}^2\} > \mathcal{F}(\boldsymbol{x}) \iff \left[\ln O_{SL}(\boldsymbol{x}) - \ln O_{SL}^{(0)}\right] < 0 \qquad (7)$$

- Similar ideas have been explored previously in the context of binary coalescences [6] and burst searches [7, 8]. But those are not readily applicable to CW searches.

Extension to hierarchical searches

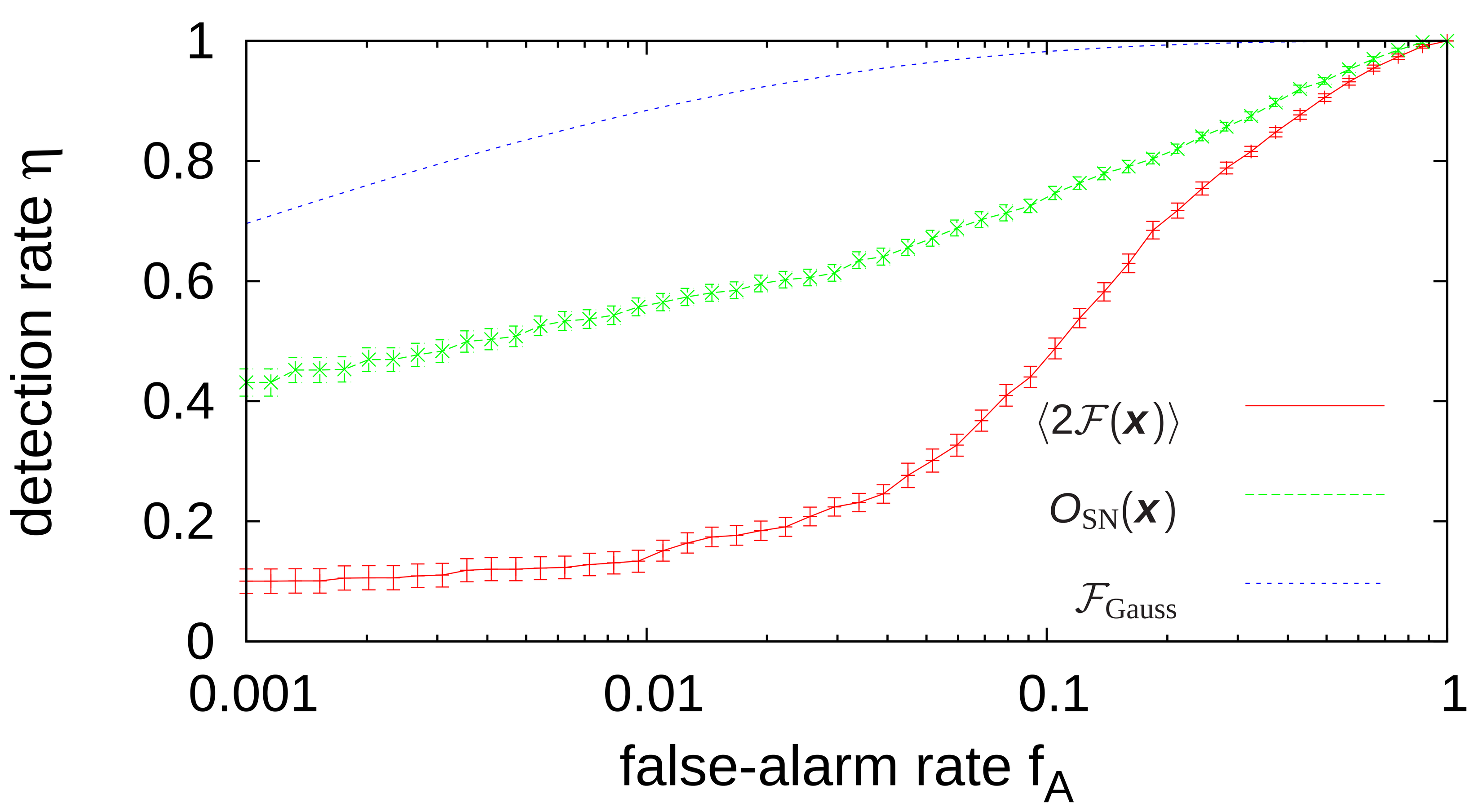
To detect weak CW signals, long integration times are necessary. For computational efficiency, these are done *hierarchically*: the detection statistic is calculated coherently only for segments $\sim$ 1 day and then summed up incoherently.
We can easily generalize our results for such hierarchical searches, e.g. Eq. (4) becomes

$$O_{SL}(\boldsymbol{x}) \propto \frac{\exp\left(\sum_k \mathcal{F}_k\right)}{\sum_X l^X \exp\left(\sum_k \mathcal{F}_k^X\right)} \qquad (8)$$

where k runs over the data segments and X over detectors. (And similarly for Eq. (5)).
For hierarchical blind searches, care needs to be taken to optimally choose the the *doppler parameters* (frequency, spindown, sky location) in each segment, e.g. by the GCT method [9].

Preliminary results

We simulated interferometer data for 2 detectors, 14 segments with 25 hours of data each. 2500 noise draws contain either pure Gaussian noise or a strong line in one detector (20% probability per detector). 2500 signal draws have a pulsar CW signal injected coherently in both detectors (average SNR $\rho = \sqrt{(x|x)} = 4.46$).
This data is fed into the same hierarchical search algorithm running on Einstein@Home. From the noise draws, we obtain the false-alarm rate f_A, while the signal draws yield the detection rate η.



The new detection statistic (semi-coherent version of (5)) is found to be much more effective than the standard semi-coherent $\mathcal{F}$-statistic. Especially at low false-alarm rates, which are desirable for GW searches, the new statistic allows for more detections. Using the line veto, the multi-detector $\mathcal{F}$-statistic threshold can be lowered, accepting more real signals, while still vetoing the lines which are suspiciously strong in a single detector.

References

- [1] Jaranowski, Królak, Schutz, *PRD* **58**, 063001 (1998)
- [2] Cutler, Schutz, *PRD* **72**, 063006 (2005)
- [3] Prix, Krishnan, *CQG* **26**, 204013 (2009)
- [4] Abbott et al., *PRD* **76**, 082001 (2007)
- [5] S5 E@H search, in preparation
- [6] Veitch, Vecchio *PRD* **81**, 062003 (2010)
- [7] Clark, Heng, Pitkin, Woan, *PRD* **76**, 043003 (2007)
- [8] Littenberg, Cornish, *PRD* **82**, 103007 (2010)
- [9] Pletsch, Allen, *PRD* **103**, 181102 (2009)
- [10] Prix, Keitel, Papa, Leaci, Siddiqi, in preparation

Outlook

More detailed results from representative studies of simulated data will be published soon [10]. We will also use Eq. (8) as a follow-up veto for strong candidates from the LIGO S5 run. The current S6 analysis, running on the distributed computing project Einstein@Home (`http://www.einsteinathome.org`) since May 2011, will presumably be analyzed using the semi-coherent $O_{SN}(\boldsymbol{x})$ as a detection statistic.