SGWB characterization with LISA

Mauro Pieroni

Imperial College London

mpieroni@ic.ac.uk

GW primordial cosmology workshop

May 17, 2021

Outline

SGWBs at GW detectors

The data $\tilde{d}$ (in frequency space) can be expressed as

$$\tilde{d} = \tilde{s} + \tilde{n}$$

For an isotropic SGWB $\longrightarrow$ $\langle h_\lambda(\vec{k})\, h^*_{\lambda'}(\vec{k}')\rangle = P_h^\lambda(k)(2\pi)^3\delta_{\lambda\lambda'}\delta(\vec{k}-\vec{k}')$

Assuming $\langle \tilde{s}\tilde{n}\rangle = 0$ and Gaussian signal

$$\boxed{\left\langle \tilde{d}^2 \right\rangle = \left\langle \tilde{s}^2 \right\rangle + \left\langle \tilde{n}^2 \right\rangle = \mathcal{R}\, P_h^\lambda + N \equiv \mathcal{R}\left[P_h^\lambda + S_n\right]}$$

where we have introduced

- The response function of the instrument $\mathcal{R}$
- The noise power spectrum N
- The (square of the) Strain sensitivity S_n (in 1/Hz)

SGWBs at GW detectors

The data $\tilde{d}$ (in frequency space) can be expressed as

$$\tilde{d} = \tilde{s} + \tilde{n}$$

For an isotropic SGWB $\longrightarrow$ $\langle h_\lambda(\vec{k})\, h^*_{\lambda'}(\vec{k}')\rangle = P_h^\lambda(k)(2\pi)^3\delta_{\lambda\lambda'}\delta(\vec{k}-\vec{k}')$

Assuming $\langle\tilde{s}\tilde{n}\rangle = 0$ and Gaussian signal

$$\boxed{\left\langle \tilde{d}^2 \right\rangle = \left\langle \tilde{s}^2 \right\rangle + \left\langle \tilde{n}^2 \right\rangle = \mathcal{R}\, P_h^\lambda + N \equiv \mathcal{R}\left[P_h^\lambda + S_n\right]}$$

where we have introduced

- The response function of the instrument $\mathcal{R}$
- The noise power spectrum N
- The (square of the) Strain sensitivity S_n (in 1/Hz)

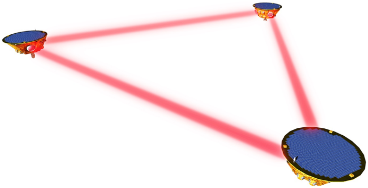
In order to compare with cosmological predictions it's customary to introduce

$$\Omega_n(f) = \frac{4\pi^2}{3H_0^2} f^3 S_n(f)\,, \qquad \text{and} \qquad \Omega_{\rm GW} \equiv \frac{1}{3H_0^2 M_p^2}\frac{\partial\rho_{\rm GW}}{\partial \ln f} = \frac{4\pi^2}{3H_0^2} f^3 \sum_\lambda P_h^\lambda$$

where $H_0 \simeq 3.24\times10^{-18}\, h_0\, \mathrm{Hz}$ is the Hubble constant today.

Laser Interferometer Space Antenna

Few details on LISA:



- First direct GW detector in space
- Constellation of three satellites
- 2.5 million km arm lengths
- Peak sensitivity $10^{-2} \div 10^{-3}$Hz
- Three correlated interferometers (XYZ basis)
- $\sim$ two independent detectors (AET basis)
- Expected launch in 2034
- Operating for 4yrs (nominal)

Very interesting for cosmology since we can:

- Measure H_0 (see 1601.07112)
- Test modified gravity (see1906.01593)
- (Hopefully) detect and characterize SGWBs! (This talk!)

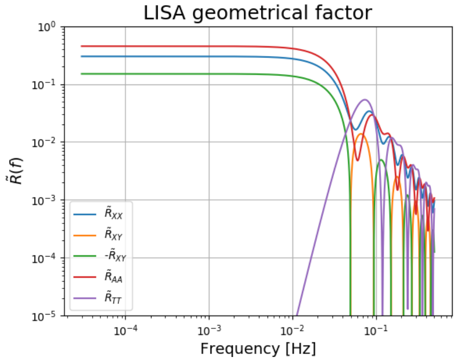
LISA response function

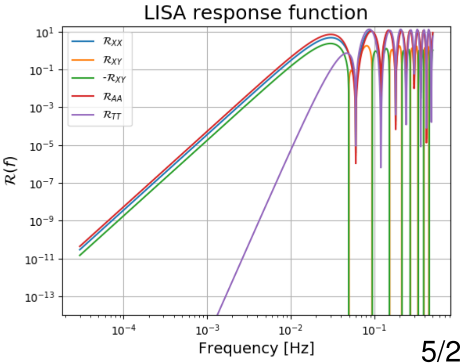
For an isotropic and non-chiral spectrum we get (see 2009.11845):

$$\left\langle \Delta F^{TDI}_{i(jk)} \Delta F^{TDI}_{l(mn)} \right\rangle = \int \mathrm{d}k \, P_h(k) \mathcal{R}_{ij}(k) \, , \qquad \mathcal{R}_{ij}(k) \equiv 4 \, (2\pi k L)^2 \left| W(kL) \right|^2 \tilde{R}_{il(jk)(mn)}(k) \, .$$

where $\mathcal{R}_{ij}(k)$ is the LISA response function.

For XYZ/AET (AET is noise diagonal) combinations we get:





The noise model I

Two analytical approximations for acceleration and interferometric noise:

$$P_{acc}(f,A) = A^2 \cdot 10^{-30} \cdot \left[1 + \left(\frac{4\cdot 10^{-4}}{f}\right)^2\right]\left[1 + \left(\frac{f}{8\cdot 10^{-3}}\right)^4\right]\left(\frac{1}{2\pi f}\right)^4\left(\frac{2\pi f}{c}\right)^2 ,$$

$$P_{IMS}(f,P) = P^2 \cdot 10^{-24} \cdot \left[1 + \left(\frac{2\cdot 10^{-3}}{f}\right)^4\right]\left(\frac{2\pi f}{c}\right)^2 .$$

The power spectral densities are ($L = 2.5 \times 10^9$m is the arm length):

$$P_{PSD}^{XX}(f) = 16\sin^2\left(\frac{2\pi f L}{c}\right)\left\{P_{IMS}(f,P) + \left[3 + \cos\left(\frac{4\pi f L}{c}\right)\right]P_{acc}(f,A)\right\} ,$$

$$P_{PSD}^{XY}(f) = -8\sin^2\left(\frac{2\pi f L}{c}\right)\cos\left(\frac{2\pi f L}{c}\right)\left\{P_{IMS}(f,P) + 4P_{acc}(f,A)\right\} ,$$

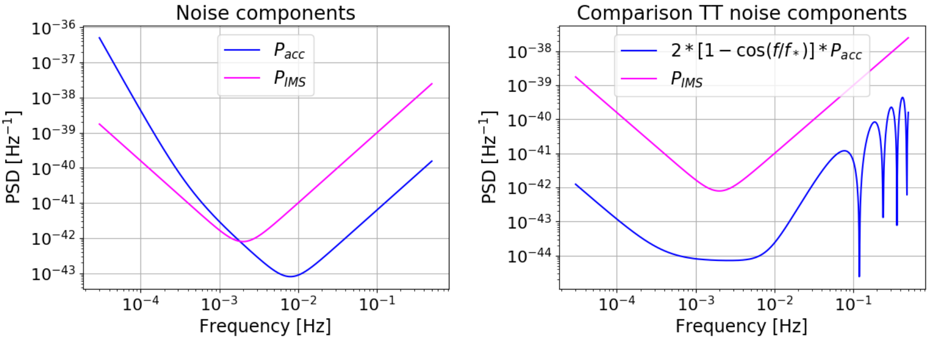
which for the TT combination gives:

$$P_{PSD}^{TT}(f,A,P) = 16\sin^2\left(\frac{2\pi f L}{c}\right)\left\{2\left[1 - \cos\left(\frac{2\pi f L}{c}\right)\right]^2 P_{acc}(f,A) + \right.$$
$$\left. + \left[1 - \cos\left(\frac{2\pi f L}{c}\right)\right]P_{IMS}(f,P)\right\} .$$

The noise model II

At low frequencies this becomes ($f_* \equiv (2\pi L/c)^{-1} \simeq 0.019\,\mathrm{Hz}$):

$$P^{TT}_{PSD}(f,A,P) \simeq 8\left(\frac{f}{f_*}\right)^2 \sin^2\left(\frac{f}{f_*}\right)\left[\left(\frac{f}{f_*}\right)^2 P_{acc}(f,A) + P_{IMS}(f,P)\right],$$

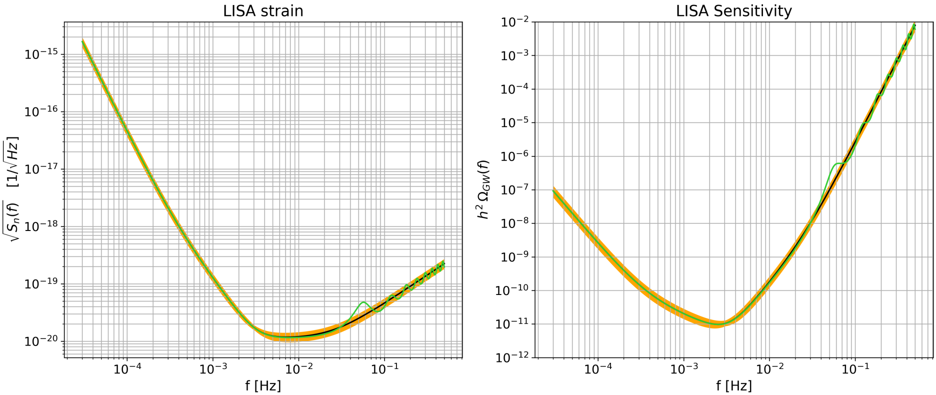


The interferometric noise dominates TT at all frequencies!

Analytical vs numerical

An approximation of the response function and the strain sensitivity are:

$$\tilde{R}(f) = \frac{0.3}{1 + 0.6\left(\frac{2\pi f L}{c}\right)^2}\,, \qquad S_n(f,P,A) = \frac{P_{PSD}(f,P,A)}{\tilde{R}(f)\times 16\sin^2\left(\frac{2\pi f L}{c}\right)\times\left(\frac{2\pi f L}{c}\right)^2}\,.$$



Central values in black (analytical) / green (numerical) ±20% in orange

Data generation

Assume signal and noise (Ω units) to be Gaussian distributed

The spectra ($\Omega_{\rm GW}$ and Ω_n) quantify the variance of fluctuations

$$\tilde{s}_c(f_i) = \left| \frac{G(0, \sqrt{\Omega_{\rm GW}(f_i)}) + i\, G(0, \sqrt{\Omega_{\rm GW}(f_i)})}{\sqrt{2}} \right|$$

$$\tilde{n}_c(f_i) = \left| \frac{G(0, \sqrt{\Omega_n(f_i)}) + i\, G(0, \sqrt{\Omega_n(f_i)})}{\sqrt{2}} \right|$$

For each data segment and frequency we generate a gaussian realization.

Data generation

Assume signal and noise (Ω units) to be Gaussian distributed

The spectra ($\Omega_{\rm GW}$ and Ω_n) quantify the variance of fluctuations

$$\tilde{s}_c(f_i) = \left| \frac{G(0, \sqrt{\Omega_{\rm GW}(f_i)}) + i\, G(0, \sqrt{\Omega_{\rm GW}(f_i)})}{\sqrt{2}} \right|$$

$$\tilde{n}_c(f_i) = \left| \frac{G(0, \sqrt{\Omega_n(f_i)}) + i\, G(0, \sqrt{\Omega_n(f_i)})}{\sqrt{2}} \right|$$

For each data segment and frequency we generate a gaussian realization.

Given that:

- LISA will be operating for 4yrs (75% efficiency)
- We choose data segments of roughly 12 days

we conclude that:

- Roughly 95 independent measurements at each frequency.
- The resolution of the detector is roughly 10^{-6}Hz

Data pre-processing and likelihood

Starting from $D_c(f_i)$ (our data), defined as:

$$D_c(f_i) \equiv \langle \tilde{d}_c^2(f_i) \rangle = \langle (\tilde{s}_c(f_i) + \tilde{n}_c(f_i))^2 \rangle = \langle \tilde{s}_c^2(f_i) \rangle + \langle \tilde{n}_c^2(f_i) \rangle \ .$$

we can reduce the complexity of the problem by performing two operations:

Data pre-processing and likelihood

Starting from $D_c(f_i)$ (our data), defined as:

$$D_c(f_i) \equiv \langle \tilde{d}_c^2(f_i) \rangle = \langle (\tilde{s}_c(f_i) + \tilde{n}_c(f_i))^2 \rangle = \langle \tilde{s}_c^2(f_i) \rangle + \langle \tilde{n}_c^2(f_i) \rangle \ .$$

we can reduce the complexity of the problem by performing two operations:

- We average over the (95) data segments:
  This leaves us with some $D(f_i)$ (the averaged data) and
  an estimate of the error $\sigma(f_i)$ (the standard deviation or the data).

Data pre-processing and likelihood

Starting from $D_c(f_i)$ (our data), defined as:

$$D_c(f_i) \equiv \langle \tilde{d}_c^2(f_i) \rangle = \langle (\tilde{s}_c(f_i) + \tilde{n}_c(f_i))^2 \rangle = \langle \tilde{s}_c^2(f_i) \rangle + \langle \tilde{n}_c^2(f_i) \rangle \ .$$

we can reduce the complexity of the problem by performing two operations:

- We average over the (95) data segments:
  This leaves us with some $D(f_i)$ (the averaged data) and
  an estimate of the error $\sigma(f_i)$ (the standard deviation or the data).
- We coarse grain the data:
  i.e. from the initial linear 10^{-6}Hz spacing ($\sim 5 \times 10^5$ points)
  $\longrightarrow$ we go to some final (and less dense) set of frequencies f_i
  This leaves us with the final data set D_i and errors σ_i.

Data pre-processing and likelihood

Starting from $D_c(f_i)$ (our data), defined as:

$$D_c(f_i) \equiv \langle \tilde{d}_c^2(f_i) \rangle = \langle (\tilde{s}_c(f_i) + \tilde{n}_c(f_i))^2 \rangle = \langle \tilde{s}_c^2(f_i) \rangle + \langle \tilde{n}_c^2(f_i) \rangle \ .$$

we can reduce the complexity of the problem by performing two operations:

- We average over the (95) data segments:
  This leaves us with some $D(f_i)$ (the averaged data) and
  an estimate of the error $\sigma(f_i)$ (the standard deviation or the data).
- We coarse grain the data:
  i.e. from the initial linear 10^{-6}Hz spacing ($\sim 5 \times 10^5$ points)
  $\longrightarrow$ we go to some final (and less dense) set of frequencies f_i
  This leaves us with the final data set D_i and errors σ_i.

Finally we assume the data to be described by the likelihood:

$$\mathcal{L}\left(\vec{\theta}, \vec{n}\right) \propto \exp\left[-\frac{N_{chunks}}{2} \sum_i \left(\frac{D_i - h^2\Omega_{\rm GW}\left(f_i, \vec{\theta}\right) - h^2\Omega_n\left(f_i, \vec{n}\right)}{\sigma_i}\right)^2\right]$$

with i labeling the data points and $\Omega_{\rm GW}$, Ω_n models for signal and noise.

SGWBinner algorithm

Based on two LISA COSWG projects:

- 1906.09244: C. Caprini, D. Figueroa, R. Flauger, M.P., G. Nardini, M. Peloso, A. Ricciardone, G. Tasinato
- 2009.11845: R. Flauger, N. Karnesis, G. Nardini, M. P., A. Ricciardone, J. Torrado

We look for best approximation of the signal with a multi-PL

$$h^2\Omega_{\rm GW}\left(f,\,\vec{\theta}\right)=\sum_i 10^{\alpha_i}\left(\frac{f}{\sqrt{f_{\min,i}\,f_{\max,i}}}\right)^{p_i}\Theta\left(f-f_{\min,i}\right)\Theta\left(f_{\max,i}-f\right)\,.$$

where Θ is the Heaviside step function.

N bins $\rightarrow$ $4N$ $(f_{\min,i}, f_{\max,i}, \alpha_i, p_i)$ $+2N$ (noise) parameters.

SGWBinner algorithm

Based on two LISA COSWG projects:

- 1906.09244: C. Caprini, D. Figueroa, R. Flauger, M.P., G. Nardini, M. Peloso, A. Ricciardone, G. Tasinato
- 2009.11845: R. Flauger, N. Karnesis, G. Nardini, M. P., A. Ricciardone, J. Torrado

We look for best approximation of the signal with a multi-PL

$$h^2\Omega_{\rm GW}\left(f,\vec{\theta}\right)=\sum_i 10^{\alpha_i}\left(\frac{f}{\sqrt{f_{{\rm min},i}\,f_{{\rm max},i}}}\right)^{p_i}\Theta\left(f-f_{{\rm min},i}\right)\Theta\left(f_{{\rm max},i}-f\right)\,.$$

where Θ is the Heaviside step function.

N bins $\rightarrow$ $4N$ $(f_{{\rm min},i}, f_{{\rm max},i}, \alpha_i, p_i)$ $+2N$ (noise) parameters.

The basic procedure is composed of four steps

1. Build a robust prior for the noise model (to force bin-by-bin measurements)
2. Divide the frequency range in a set of bins and reconstruct the signal
3. Merge as many bins as possible (to avoid overfitting)
4. Define a procedure to compute the error on the reconstruction
5. Final MCMC run with common noise parameters

Few more detail on steps 1 and 3 ...

Characterizing the noise

Bad noise reconstruction $\longrightarrow$

- No detection/false detections
- Bad parameter reconstruction

Some useful observations:

- Noise parameters are correlated over the full frequency range!
- Noise is expected to dominate at small and at large frequencies
- Noise dominates over signal in TT

Characterizing the noise

Bad noise reconstruction $\longrightarrow$

- No detection/false detections
- Bad parameter reconstruction

Some useful observations:

- Noise parameters are correlated over the full frequency range!
- Noise is expected to dominate at small and at large frequencies
- Noise dominates over signal in TT

As a consequence (single channel version):

1. We divide the range into three bands (small and large frequencies + a central band)
2. Estimate signal and noise parameters in the external bands
3. Use this as a prior for the measurements in the bins in the central part

As a consequence (three channels version):

1. Estimate signal and noise parameters in TT
2. Use this as a prior for the AA/EE.

Criterion for merging

Two motivations for merging:

- Larger N means smaller bins which implies larger errors
- For large values of N, the reconstruction with N bins ($i.e.$ $2N$ parameters) may overfit the signal

  Typically reducing the number N of bins may improve the analysis!

Criterion for merging

Two motivations for merging:

- Larger N means smaller bins which implies larger errors
- For large values of N, the reconstruction with N bins (*i.e.* $2N$ parameters) may overfit the signal

Typically reducing the number N of bins may improve the analysis!

AIC for model comparison we use:

$$\mathrm{AIC} = -2 \ln \mathcal{L} + 2k = \chi^2 + 2k$$

For a couple of consecutive bins $(i, i+1)$ we can compute

$$\Delta\,\mathrm{AIC} = \mathrm{AIC}_{\mathrm{after\ merging}} - \mathrm{AIC}_{\mathrm{before\ merging}} = \chi^2_{\mathrm{after\ merging}} - \chi^2_{\mathrm{before\ merging}} - 2\,k_{\mathrm{1\text{-}bin}}$$

According to the AIC definition:

$\Delta\mathrm{AIC} < 0$ $\longrightarrow$ It is convenient to merge the two bins

A more accurate likelihood

A Gaussian likelihood would give a systematic low bias!
(astro-ph/9808264, astro-ph/0205387, astro-ph/0302218, 0801.0554)

Consider the Gaussian likelihood:

$$\ln \mathcal{L}_G\left(\vec{\theta}, \vec{n}\right) \propto -\frac{N_{chunks}}{2} \sum_{i,j} \sum_k w_{ij}^{(k)} \left(\frac{D_{ij}^{(k)} - h^2\Omega_{\rm GW}\left(f_{ij}^{(k)}, \vec{\theta}\right) - h^2\Omega_n\left(f_{ij}^{(k)}, \vec{n}\right)}{h^2\Omega_{\rm GW}\left(f_{ij}^{(k)}, \vec{\theta}\right) + h^2\Omega_n\left(f_{ij}^{(k)}, \vec{n}\right)} \right)^2$$

and the Lognormal likelihood:

$$\ln \mathcal{L}_{LN}\left(\vec{\theta}, \vec{n}\right) \propto -\frac{N_{chunks}}{2} \sum_{i,j} \sum_k w_{ij}^{(k)} \ln^2 \left(\frac{h^2\Omega_{\rm GW}\left(f_{ij}^{(k)}, \vec{\theta}\right) + h^2\Omega_n\left(f_{ij}^{(k)}, \vec{n}\right)}{D_{ij}^{(k)}} \right)$$

Then we define our likelihood as (astro-ph/0302218, 2009.11845)

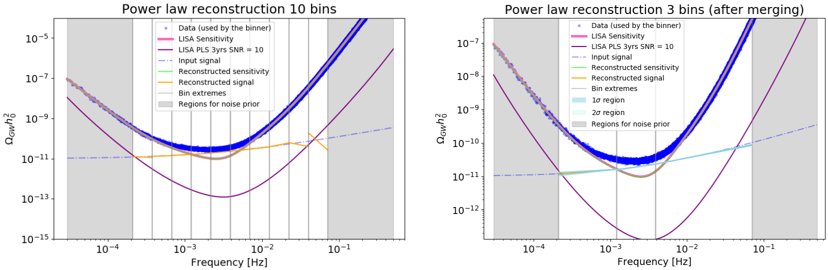
$$\ln \mathcal{L} = \frac{1}{3} \ln \mathcal{L}_G + \frac{2}{3} \ln \mathcal{L}_{LN}$$

which removes the skewness contributions and thus is more accurate.

Linear signal + “LIGO binaries”

As a first example let us consider

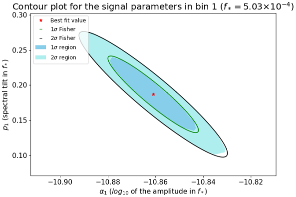
$$h^2\Omega_{\rm GW}(f) = h^2\Omega_{\rm GW,const}(f)+h^2\Omega_{\rm GW,BHB+NSB}(f) = 10^{-11}+5.4\times10^{-12}\left(\frac{f}{0.001}\right)^{2/3}$$

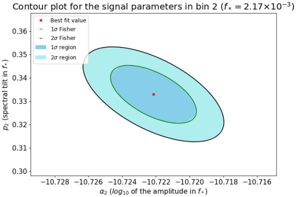


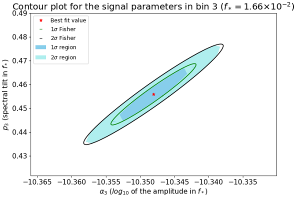
After the merging procedure only 3 bins with small error bands are left.

Linear signal + “LIGO binaries” contour plots

Let us have a closer look at the contour plots in each bin



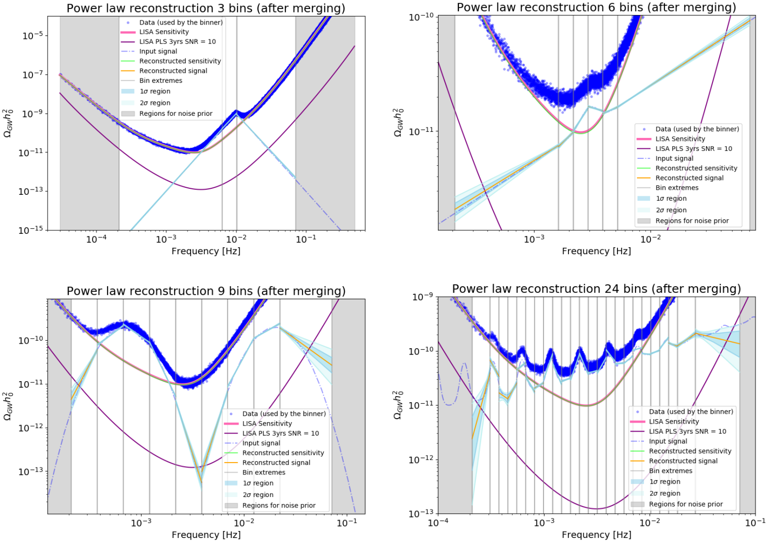




The three contour plots clearly show a progressive increase in the slope

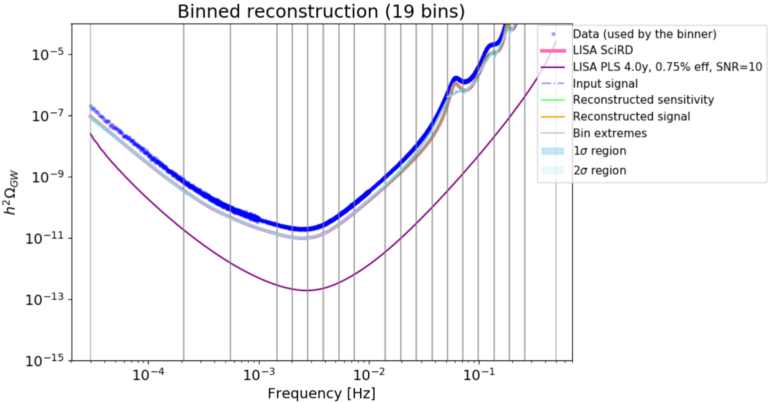
Consistent with the values of the ratio $\Omega_{\rm GW,const}/\Omega_{\rm GW,binaries}$ in the three bins!

More cases


Power law reconstruction 3 bins (after merging)
Power law reconstruction 6 bins (after merging)
Power law reconstruction 9 bins (after merging)
Power law reconstruction 24 bins (after merging)
$\Omega_{GW}h_0^2$
Frequency [Hz]
Data (used by the binner)
LISA Sensitivity
LISA PLS 3yrs SNR = 10
Input signal
Reconstructed sensitivity
Reconstructed signal
Bin extremes
1σ region
2σ region
Regions for noise prior

Degenerate cases

Since the noise in TT has a different shape
this may help in breaking degeneracies!

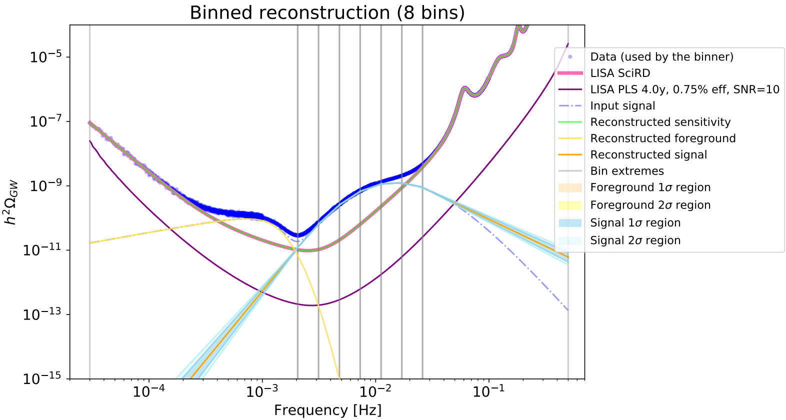


Foreground removal

The code can now account for perform component separation too!

Consider for example the signal due to Galactic binaries:

$$h^2\Omega_{\rm GW} = 10^{\alpha_{FG}} f^{2/3} \mathrm{e}^{-a_1 f + a_2 f \sin(a_3 f)} \left\{1 + \tanh\left[a_4(f_k - f)\right]\right\} \ .$$



An exact solution for the parameters

Based on 2004.01135, in collaboration with Enrico Barausse

If the model (for both signal and noise) is linear in the $\vec{\theta}$:

- The log likelihood is quadratic in the parameters
- The Fisher matrix does not depend on the parameters
- Finding the best fit reduces to solving a linear equation

An exact solution for the parameters

Based on 2004.01135, in collaboration with Enrico Barausse

If the model (for both signal and noise) is linear in the $\vec{\theta}$:

- The log likelihood is quadratic in the parameters
- The Fisher matrix does not depend on the parameters
- Finding the best fit reduces to solving a linear equation

Starting from:

$$-\ln \mathcal{L}\left(\vec{\theta}, \vec{n}\right) \propto \frac{1}{2} \sum_i \left(\frac{D_i - M\left(f_i, \vec{\theta}\right)}{\sigma_i} \right)^2 ,$$

for a linear model we get:

$$F_{lk} \equiv \sum_i \frac{1}{\sigma_i^2} \frac{\partial M\left(f_i, \vec{\theta}\right)}{\partial \theta_l} \frac{\partial M\left(f_i, \vec{\theta}\right)}{\partial \theta_k} , \qquad \bar{\theta}_l = F_{lk}^{-1} \sum_i \frac{1}{\sigma_i^2} D_i \frac{\partial M\left(f_i, \vec{\theta}\right)}{\partial \theta_k}$$

where $\bar{\theta}_l$ is the MLE for the parameters.

A simple model for the signal

Assume the signal can be expressed as:

$$S(f) = \sum_{j=1}^{n} a_j \, \delta_w(f - f_j) \, ,$$

where:

- a_j are the parameters
- w is some correlation length
- $\delta_w(f - f_j)$ are some functions

The choice of $\delta_w(f - f_j)$ defines a basis to express the signal.

A simple model for the signal

Assume the signal can be expressed as:

$$S(f) = \sum_{j=1}^{n} a_j \, \delta_w(f - f_j) \,,$$

where:

- a_j are the parameters
- w is some correlation length
- $\delta_w(f - f_j)$ are some functions

The choice of $\delta_w(f - f_j)$ defines a basis to express the signal.

Depending on the choice of w we have two regimes:

- Small w: the measurements in f_j are not correlated
- Large w: the measurements in f_j are correlated

Properly choosing w we can smooth the signal!

Principal component analysis

It is interesting to notice that:

- In general the parameters a_j are correlated
- Eigenvectors $e_j^{(i)}$ of F_{lk} are uncorrelated combinations of a_j.
- Eigenvalues $\lambda^{(i)}$ of F_{lk} give the information on the $e_j^{(i)}$.

Principal component analysis

It is interesting to notice that:

- In general the parameters a_j are correlated
- Eigenvectors $e_j^{(i)}$ of F_{lk} are uncorrelated combinations of a_j.
- Eigenvalues $\lambda^{(i)}$ of F_{lk} give the information on the $e_j^{(i)}$.

Principal Component Analysis (PCA):

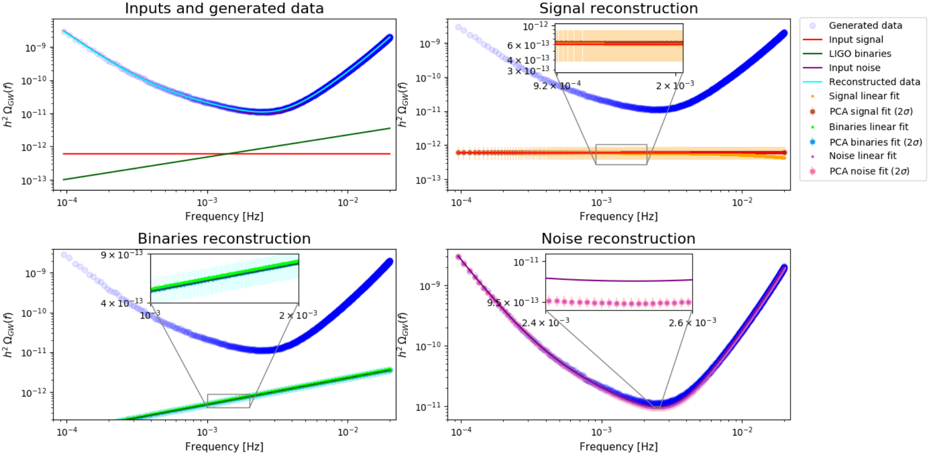
1. Compute the eigensystem of F_{lk}
2. Cut $e_j^{(i)}$ corresponding to $\lambda^{(i)}$ smaller than some threshold
3. Project $\delta_w(f - f_j)$ and a_j on this subset of $e_j^{(i)}$ (say $\eta_k(f)$, b_k)
4. Reconstruct the signal as: $S(f) = \sum_k b_k \eta_k(f)$

Corresponds to reconstructing the signal in terms of the components which can be well determined!

In the following plots all parameters are normalized to 1!

Subtracting the foreground 1

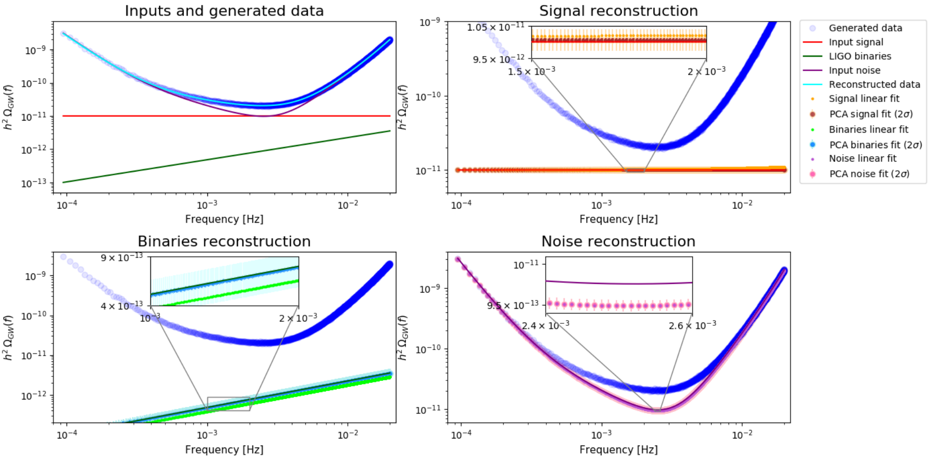
Flat signal (SNR $\sim$ 30) + LIGO binaries (gaussian prior $\sigma = 0.5$)



$L = 1.011 \pm 0.111$, $A \simeq 0.972 \pm 0.005$ and $O \simeq 0.973 \pm 0.001$

Subtracting the foreground 2

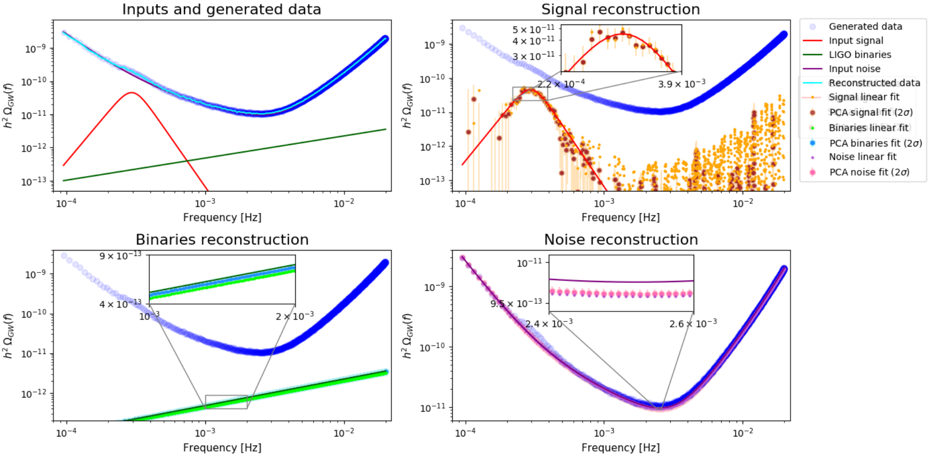
Flat signal (SNR $\sim$ 520)+ LIGO binaries (gaussian prior $\sigma = 0.5$)



$$L = 0.989 \pm 0.141,\ A = 0.971 \pm 0.005,\ O = 0.977 \pm 0.001$$

No degeneracy 1

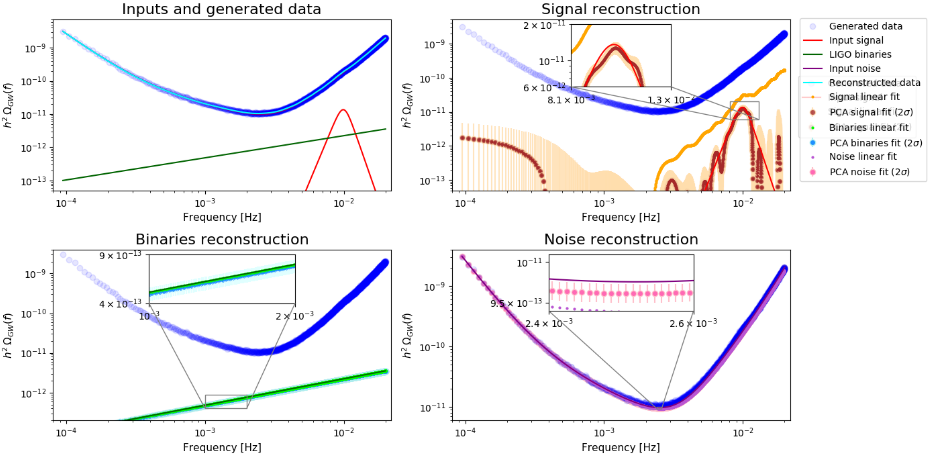
Broken PL (SNR $\sim$ 30) + LIGO binaries (gaussian prior $\sigma = 0.5$)



$L \simeq 0.960 \pm 0.033, A \simeq 0.990 \pm 0.004, O \simeq 0.976 \pm 0.001$

No degeneracy 2

Another broken PL (SNR $\sim$ 30) + LIGO binaries (gaussian prior $\sigma = 0.5$)



$L \simeq 0.972 \pm 0.054$, $A \simeq 0.987 \pm 0.008$, $O \simeq 0.982 \pm 0.002$

Conclusions and future perspective

Conclusions

- PLS is (qualitatively) useful but not the end of the story
- SGWBinner: a flexible algorithm to reconstruct general signal (multiple power lows, bumps,...)
- PCA reconstruction: an alternative approach for SGWB reconstruction
- In both cases it's possible to perform component separation

Future perspectives

- Keep improving on detector modeling
- Application to concrete case (inflation, phase transitions, ...)
- New techniques?

Last Slide

The End

Thank you

Bayes theorem and data analysis

Given some events D and $\vec{\theta}$ we define:

$$P(D|\theta_j) \equiv \frac{P(D \cap \theta_j)}{P(\theta_j)}$$

the conditional probability of D occurring given $\vec{\theta}$.

Since $P(D \cap \vec{\theta}) = P(\vec{\theta} \cap D)$ we get the Bayes theorem:

$$P(\vec{\theta}|D) = \frac{P(D|\vec{\theta}) \cdot P(\vec{\theta})}{P(D)} \propto P(D|\vec{\theta}) \cdot P(\vec{\theta})$$

D is the set of data and $\vec{\theta}$ is the vector of parameters of the theory.

- $P(\vec{\theta}|D)$ are the *Posterior probabilities*
- $P(\vec{\theta})$ are the *Prior probabilities*
- $P(D|\vec{\theta})$ is the *Likelihood* which from now on is denoted $\mathcal{L}(\vec{\theta})$
- $P(D)$ is the *Model Evidence*

Some statistical tools

- The maximum likelihood estimate of best fit parameters $\vec{\theta}_0$ is

$$\partial_{\vec{\theta}} \ln \mathcal{L}(\vec{\theta}_0) = 0$$

- The Fisher matrix (*i.e.* the inverse of the covariance matrix C_{ij}) is

$$F \equiv C_{ij}^{-1} = -\langle \partial_i \partial_j \ln \mathcal{L}|_{\vec{\theta}_0} \rangle$$

- Gaussian approximation of $\mathcal{L}$ around $\vec{\theta}_0$

$$\mathcal{L}(\vec{\theta}) \simeq \frac{1}{\sqrt{\det(2\pi C)}} \exp\left\{ -\frac{1}{2} (\vec{\theta} - \vec{\theta}_0)^T C^{-1} (\vec{\theta} - \vec{\theta}_0) \right\}$$

- Confidence intervals are obtained by solving

$$-2 \left[\ln \mathcal{L}(\vec{\theta}) - \ln \mathcal{L}(\vec{\theta}_0) \right] = v_{n\sigma}(k)$$

- Akaike Information Criterion (AIC) for a fit with k parameters:

$$\mathrm{AIC} = -2 \ln \mathcal{L} + 2k = \chi^2 + 2k$$