GW data analysis – A primer (1)



Eric Chassande-Mottin

AstroParticule et Cosmologie (APC) CNRS Université de Paris

Pulsar timing array



LISA space mission



Ground-based detectors



Known

Transient Short duration

Permanent Long duration

waveform waveform Coalescing 'Bursts' Binary asymmetric core Systems collapse supernovae Neutron Stars, cosmic strings Black Holes ??? Credit: Chandra X-ray Observatory Credit AEL CCT LSU Astrophysical or Cosmic GW Continuous background Sources stochastic, Spinning neutron stars incoherent crustal deformations. background accretion NASA/WMAP Science Tear Casey Reed, Penn State

Unknown

Outline - Part 1

- General statement and main concepts
 - Observation model (signal and noise) Likelihood
 - Detection theory **Optimal statistics**
- From general principles to transient searches
 - Compact binary coalescences Part 1 (Detection)
 - Expected gravitational-wave signal Waveform models
 - Search implementation : matched filtering
 - Observing with **multiple detectors**
 - Background estimation and **significance** assessment
 - Pipeline and software packages
 - Exercises !



Jun 7 2021

Needle in the haystack

- We are looking for weak signals in noisy data
- Few numbers for compact binaries
 - The observing run O3a lasted 6 months, ~ 1.5×10^7 seconds.
 - 40 events were found, each lasting < ~5 seconds in the detector band
 - All but ~ 10⁻⁶ of the data contained nothing but noise...
 - ... and probably loads of undetectable signals from the distant universe.
- Search pipelines must find the ~ 10⁻⁶ of the data containing a detectable signal
 - Can't skimp on the detection software! It must be close to "optimal" the best that can be achieved in the presence of unavoidable noise.

Observation equation?

data = function(signal, noise)

Detector transfer function and noise coupling

Recap – Gravitational waves

Outward propagating wave in z-direction

$$h(t,z) = h_{\mu\nu}e^{i(\omega t - kz)}$$

$$h_{\mu\nu} = \begin{pmatrix} 0 & 0 & 0 & 0\\ 0 & h_{+} & h_{\times} & 0\\ 0 & h_{\times} & -h_{+} & 0\\ 0 & 0 & 0 & 0 \end{pmatrix}$$

$$\mathbf{h}(t) = h_{+}(t)\boldsymbol{\epsilon}^{+} + h_{\times}(t)\boldsymbol{\epsilon}^{\times}$$

Polarization tensors

$$\epsilon^{+} = \hat{p} \otimes \hat{p} - \hat{q} \otimes \hat{q}$$
$$\epsilon^{\times} = \hat{p} \otimes \hat{q} + \hat{q} \otimes \hat{p}$$

 \hat{p},\hat{q} Preferred basis in source frame

Recap – Reference frames

$$\mathbf{h}(t) = h_{+}(t)\boldsymbol{\epsilon}_{+} + h_{\times}(t)\boldsymbol{\epsilon}_{\times}$$



$$\mathbf{e}^{+} = \hat{u} \otimes \hat{u} - \hat{v} \otimes \hat{v}$$
$$\mathbf{e}^{\times} = \hat{u} \otimes \hat{v} + \hat{v} \otimes \hat{u}$$

$$\begin{aligned} \boldsymbol{\epsilon}^+ &= \hat{p} \otimes \hat{p} - \hat{q} \otimes \hat{q} \\ &= \cos 2\psi \, \mathbf{e}^+ - \sin 2\psi \, \mathbf{e}^\times \end{aligned}$$

$$\begin{aligned} \boldsymbol{\epsilon}^{\times} &= \hat{p} \otimes \hat{q} + \hat{q} \otimes \hat{p} \\ &= \sin 2\psi \, \mathbf{e}^{+} + \cos 2\psi \, \mathbf{e}^{\times} \end{aligned}$$

8 Cornish, arXiv: 0910.4372

Detector response



Antenna patterns



Antenna patterns































































5













7

Propagation time delays



Time delays allows triangulation





Credit: LIGO/Virgo/NASA/Leo Singer (Milky Way image: Axel Mellinger)

Observation equation (signal part)

data = function(signal, noise)

$$h_{\text{signal}}(t) = F_{+}h_{+}(t-\tau) + F_{\times}h_{\times}(t-\tau)$$

 $F_+, F_ imes, au$ depends on sky position $heta, \phi$ and polarization angle ψ



data = function(signal, noise)

The noise that affects the measurement combines linearly

data = signal + noise

Noise components that can be modelled are Gaussian

$$p(n_i) \propto \mathcal{N}(0, \sigma) = \frac{1}{\sqrt{2\pi\sigma}} \exp{-\frac{n_i^2}{2\sigma^2}}$$

$$p(n) = \frac{1}{(2\pi)^{N/2}\sigma^N} \exp{-\sum_i \frac{n_i^2}{2\sigma^2}}$$

 $\mathbf{\Omega}$

• Gaussian and correlated

$$p(n) \propto \mathcal{N}(0, C) = \frac{\exp{-\frac{1}{2}n_i C_{ij}^{-1} n_j}}{\sqrt{|2\pi C|}}$$

-1

Multi-variate Gaussian variable

- Correlated and stationary
 - Can be described by a **spectrum**
 - Results in noise "color"

Time
$$\leftrightarrow$$
 Frequency
Correlation \leftrightarrow Power spectral density
 C $S_{noise}(f)$



Observation equation

signal d = h + nn



 $d = h + n \implies d - h = n$



26

Statistical modelling (1 detector)



Likelihood
$$p(d|h) \propto \mathcal{N}(h, C) = \frac{\exp{-\frac{1}{2}(d_i - h_i)C_{ij}^{-1}(d_j - h_j)}}{\sqrt{|2\pi C|}}$$

Jun 7 2021 Encapsulates everything: Gaussian, correlated, additive

Statistical modelling (multiple detectors)

Detector k response : $h^{(k)}$ (includes antenna pattern and time delay)

Likelihood (single det): $p(d^{(k)}|h^{(k)}) = \frac{\exp{-\frac{1}{2}(d_i^{(k)} - h_i^{(k)})C_{(k)}^{-1}ij(d_j^{(k)} - h_j^{(k)})}}{\sqrt{|2\pi C_{(k)}|}}$

Likelihood (multi det) : $p(\lbrace d^{(k)} \rbrace | h) = \prod_{k} p(d^{(k)} | h^{(k)})$

Assuming detector noises are independent

Allows the data from all detectors to be analyzed jointly (coherently) Proper "triangulation" for source sky localization

Detection – Optimal statistics

Hypothesis testing - (H1) d = h + n vs (H0) d = n (h=0)

Likelihood ratio statistic

$$\Lambda = \frac{p(d|h)}{p(d|h=0)} \leq \text{threshold}$$



Maximize detection prob at a given false alarm prob

Estimation – Maximum likelihood

Find h that maximizes:
$$p(d|h) = \frac{\exp{-\frac{1}{2}(d_i - h_i)C_{ij}^{-1}(d_j - h_j)}}{\sqrt{|2\pi C|}}$$

Maximize the likelihood amounts to minimize residual norm

$$\chi^{2} = (d_{i} - h_{i})C_{ij}^{-1}(d_{j} - h_{j}) = \langle d - h | d - h \rangle$$

where we used a freq-weighted scalar product

$$\langle a|b\rangle = a_i \ C_{ij}^{-1} \ b_j = \int \frac{\hat{a}(f)\hat{b}^*(f)}{S_{\text{noise}}(f)} \mathrm{d}f$$

Jun 7 2021

Key question: Do we know anything about the signal?

- Completely known
 - Compact binary coalescence \rightarrow matched filtering
- Partially known
 - "Unmodelled" transients \rightarrow time-frequency searches
- Unknown or completely random
 - Stochastic background \rightarrow cross-correlation

Other important questions

- Is the stationary assumption for the noise really valid?
- Is the Gaussian assumption for the noise really valid?

Compact binary coalescences – Detection



Compact stars : Neutron star or black holes or more exotic objects (boson stars, gravastars, ...)



Orbital dynamics and gravitational radiation

- In the inspiral phase, the orbits are quasi-Keplerian
 - The energy loss due to GW emission is small enough that the orbits nearly close on themselves
 - Adiabatic sequence of quasi-circular orbits

which ends at the "Innermost Stable Circular Orbit" (ISCO), at twice the Schwarzchild radius.

• GW frequency is twice the orbital frequency

Inspiral-merger-ringdown waveform

- The exact resolution of the twobody problem is hard in general relativity
 - Analytic approximations
 - Numerical relativity
- Major families
 - Phenomenological ansatz
 - Effective one-body approximation
 - Surrogate numerical relativity (for large masses and large q)

Note: very active developments on calculation techniques borrowed from quantum physics "scattering amplitudes"



36

Inspiral-merger-ringdown waveform

orbita

separatic

Orbital energy

$$E = -\frac{\nu M v^2}{2} \left\{ 1 + \left(-\frac{9+\nu}{12} \right) v^2 + \left(\frac{-81+57\nu-\nu^2}{24} \right) v^4 + \left(-\frac{675}{64} + \left[\frac{34445}{576} - \frac{205\pi^2}{96} \right] \nu - \frac{155}{96} \nu^2 - \frac{35}{5184} \nu^3 \right) v^6 + \mathcal{O}(v^8) \right\}$$

Energy loss due to GW radiation $\mathcal{F} = \frac{32\nu^2 v^{10}}{5} \left\{ 1 - \left(\frac{1247}{336} + \frac{35}{12}\nu\right)v^2 + 4\pi v^3 + \left(-\frac{44711}{9072} + \frac{9271}{504}\nu + \frac{65}{18}\nu^2\right)v^5 \right\}$

Resolve eqs of motion:

$$\frac{dv}{dt} = \frac{-\mathcal{F}(v)}{E'(v)}$$

Deduce orbital phase and polar waveform

Jun 7 2021

$$\frac{dv}{dt} = \frac{-\mathcal{F}(v)}{E'(v)}$$

 $\pi()$

 $\frac{d\varphi(t)}{dt} = \frac{v^3}{M} \qquad h_+ = \frac{2\nu M}{D} v^2 (1 + \cos^2 \iota) \, \cos[2\varphi(t)]$ $h_{\times} = \frac{4\nu M}{D} v^2 \cos \iota \, \sin[2\varphi(t)]$

path to merger

Numerical

Relativity

(NR)

perturbatic

post-Newtonian theory

(PN)

Differences between waveform families

- There are small differences between waveform approximants
 - Can help to estimate the "systematic error" wrt the "true" GR waveform



Waveform parameters

- Component masses: m₁, m₂
- Component spins: S_1 , S_2 (2 x 3 params)
- Sky location: right ascension α , declination δ
- Luminosity distance: *d*_L
- Orbital plane orientation: inclination $\iota \cong \theta_{\scriptscriptstyle LN}$, polarization angle ψ
 - If component spins are not aligned with the orbital angular momentum L, the orbital plane will *precess* so the inclination angle is $\theta_{_{JN}}$
- At coalescence: phase φ_c , time t_c
- That's 15 parameters, for (adiabatically) quasi-circular orbits



Waveform parameters

- Component masses: m₁, m₂
- Component spins: S_1 , S_2 (2 x 3 params)
- Sky location: right ascension α , declination δ
- Luminosity distance: d_L
- Orbital plane orientation: inclination $\iota \cong \theta_{LN}$, polarization angle ψ
 - If component spins are not aligned with the orbital angular momentum L, the orbital plane will *precess* so the inclination angle is $\theta_{_{JN}}$
- At coalescence: phase φ_c , time t_c
- That's 15 parameters, for (adiabatically) quasi-circular orbits
 - 8 intrinsic and 7 extrinsic

Waveform variability





Varying the mass parameters

Effect: overall phase evolution, and final ringdown frequency – Chirp mass and mass ratio

Varying the spin amplitude

Effect: "orbital hang-up" - Effective spin



Varying the spin orientation



Effect: relativistic spin precession – Precessing spin

Post-Newtonian Expansion $u = (\pi \mathcal{M} f)^{1/3} \sim v$

$$\begin{array}{lll} \text{OPN} & \frac{3}{128} \, u^{-5} & \text{Measure chirp mass} & \text{Chirp mass:} & \mathcal{M} = \frac{(m_1 m_2)^{3/5}}{(m_1 + m_2)^{1/5}} \\ \text{IPN} & \left(\frac{3715}{32256} + \eta \frac{55}{384}\right) \eta^{-2/5} u^{-3} & \text{Measure individual masses} & \text{Mass ratio:} & q = \frac{m_1}{m_2} \\ \text{I.SPN} & -\left(\frac{3\pi}{8} - \frac{1}{32} \left[113(1 \pm \sqrt{1 - 4\eta}) - 76\eta\right] \hat{L} \cdot \vec{\chi}_{1,2}\right) \eta^{-3/5} u^{-2} & \text{Measure spin combination} & \text{Effective spin:} \\ \chi_{\text{eff}} = \frac{m_1 \chi_1 \cos \theta_1 + m_2 \chi_2 \cos \theta_2}{m_1 + m_2} \\ \text{2PN} & \left(\frac{1529305}{21567022} + \frac{27145}{21504} \eta + \frac{3085}{3072} \eta^2 + \sigma(\hat{L} \cdot \vec{\chi}_{1,2}, \vec{\chi}_{1} \cdot \vec{\chi}_{2}, \vec{\chi}_{1,2})\right) \eta^{-4/6} u^{-1} & \text{Measure individual spins} \end{array}$$

Jun 7 2021

Matched filtering

Likelihood:
$$p(d|\boldsymbol{\lambda}) \propto \exp{-\frac{1}{2}\langle d - h(\boldsymbol{\lambda})|d - h(\boldsymbol{\lambda})\rangle}$$

Parameters: $\boldsymbol{\lambda} = \{m_1, m_2, \mathbf{S}_1, \mathbf{S}_2, \ldots\}$
Likelihood ratio: $\Lambda(\boldsymbol{\lambda}) = \frac{p(d|h(\boldsymbol{\lambda}))}{p(d|h=0)} = e^{-\langle d|h \rangle + \frac{1}{2}\langle h|h|}$

Maximize over λ Let's do over amplitude only for now

$$h(\boldsymbol{\lambda}) \equiv \rho \ h_{\mathrm{norm}}(\boldsymbol{\lambda})$$

Compute ρ such that $\; \partial \Lambda/d\rho = 0 \;$ and replace in Λ

Matched filtering

Likelihood:
$$p(d|\boldsymbol{\lambda}) \propto \exp{-\frac{1}{2}\langle d - h(\boldsymbol{\lambda})|d - h(\boldsymbol{\lambda}) \rangle}$$

Parameters: $\boldsymbol{\lambda} = \{m_1, m_2, \mathbf{S}_1, \mathbf{S}_2, \ldots\}$
Likelihood ratio: $\Lambda(\boldsymbol{\lambda}) = \frac{p(d|h(\boldsymbol{\lambda}))}{p(d|h=0)} = e^{-\langle d|h \rangle + \frac{1}{2}\langle h|h \rangle}$
Generalized
likelihood ratio: $\rho(\boldsymbol{\lambda}) = \langle d|h_{\text{norm}}(\boldsymbol{\lambda}) \rangle = \int{\frac{\hat{d}(f)\hat{h}_{\text{norm}}^*(f;\boldsymbol{\lambda})}{S_{\text{noise}}(f)}}df$



Signal-to-noise ratio SNR



Constrast between the two hypotheses

$$SNR^{2} = \int \frac{|h(f)|^{2}}{S_{\text{noise}}(f)} df$$

48

Figure-of-merit used to predict detectability – Typ. SNR > 7

$$SNR = \frac{E[\rho]}{\operatorname{var}[\rho_0]^{1/2}}$$
$$= \langle h | h_{\operatorname{norm}} \rangle$$
$$= \sqrt{\langle h | h \rangle}$$

Jun 7 2021

Matched filtering (cont'd)

Maximize over λ : analytically over extrinsic params (d_1, φ_c, t_c)

$$z(\tau; \boldsymbol{\lambda}) = \int \frac{\hat{d}(f)\hat{h}_{\text{norm}}^*(f; \boldsymbol{\lambda})}{S_{\text{noise}}(f)} e^{2\pi i f \tau} df$$
$$\rho(\boldsymbol{\lambda}) = \max_{\tau} |z(\tau; \boldsymbol{\lambda})| \qquad \text{Can be computed with FFT}$$

Maximize over λ : for the rest, use brute force "grid search"

Matched filtering (cont'd) – Grid search

Grid spacing ensures < 3 % of SNR loss (2 % in BBH region)

Correlation or match <h1lh2> between neighboring templates is > 97 %



Jun 7 2021

Dal Canton, Harry – arXiv:1705.01845 **50**

Matched filtering (cont'd) – Grid search



https://github.com/gwastro/pycbc-config/tree/master/O1/bank https://github.com/gwastro/pycbc-config/tree/master/O2/bank https://git.ligo.org/lscsoft/gstlal/-/tree/master/gstlal-inspiral/share/O3

Jun 7 2021

Dal Canton, Harry – arXiv:1705.01845 **51**

Non-Gaussian/non-stationary noise artifacts



Noise mitigation and glitch rejection (1)

- Use multiple detector data
 - **Coincidence** in time and template parameters
 - Consistency in phase and SNR
 - Note: there are no fully coherent searches in production so far
- Use info from environmental and detector monitoring channels



Noise mitigation and glitch rejection (2)

- X² signal consistency test
 - Companion statistics that tests the spectral profile of triggers

$$\chi^2 = \sum_{k \in \text{bins}} (\rho_k - \rho/N_{\text{bins}})^2 \quad \rho = \sum_{k \in \text{bins}} \rho_k$$

- Combine p and X² statistics

$$\hat{\rho} = \begin{cases} \frac{\rho}{[(1+(\chi_r^2)^{p/2})/2]^{1/p}} & \text{for } \chi_r^2 > 1\\ \rho & \text{for } \chi_r^2 \le 1 \end{cases}$$





Matched filtering (cont'd) – Whitening

$$z(\tau; \boldsymbol{\lambda}) = \int \frac{\hat{d}(f)\hat{h}_{\text{norm}}^*(f; \boldsymbol{\lambda})}{S_{\text{noise}}(f)} e^{2\pi i f \tau} df$$

- Down-weighting noisy frequency regions (low-, high- and "lines")
- Filtering with $S_{\text{noise}}^{-1/2}$ is known as "whitening"
- Requires regular estimates of noise PSD to adapt slow drift in the noise level



Workflows and pipelines



Workflows and pipelines

- GstLAL lscsoft.docs.ligo.org/gstlal
- PyCBC pycbc.org
- MBTA T. Adams et al. (2016)
- SPIIR Q. Chu (2017)
- IAS Venumadhav et al. (2020)
- Bulk of computing is **embarrassingly parallel**
- Run on dedicated **computer clusters** of CPU or GPU or on distributed computing infrastructures
- **Complex workflows** orchestrated using schedulers such as Condor or Pegasus



Jun 7 2021

Event significance assessment



- Background rate of accidental noise (glitch) coincidence in two or more detectors
- Impossible to shield against GW There are no "noise-only" data
- Empirical estimate using surrogate "noise only" data obtained by applying non-physical time shifts
 - Allows to generate > 100,000 equivalent yr!
- Coincidences in time shifted triggers = background
- Deduce event significance
 - False alarm rate = rate of background triggers ranked higher than the foreground

Background estimate

O2 science run



Jun 7 2021

Conclusions

- Fundamentals of GW data analysis
 - Observation model (detector response and noise model)
 - Deduced the **likelihood**
- Compact binary coalescence searches
 - Application of detection statistics
 - Matched filtering and its practical implementation
 - Dealing with noise imperfections
- You will experiment with those concepts with the tutorials (kick-off at the project session today)

Jun 7 2021

This presentation borrowed materials from other excellent lectures. Thank you to Neil Cornish, Alan Weinstein and others!

gw-openscience.org

٠



Download data **A**

-

- \sim Join the email list
- **Open Data Workshops** R

Jun 7 2021

- Public release of GW data and event catalogs
 - O3a data published on Apr 30 2021! 🔤 ٠
- Documentation, usage recommendations
- Online training: video tutorials and Jupyter notebooks
 - https://gw-odw.thinkific.com ٠





References

- Papers/Reviews
 - LIGO/Virgo, "A guide to LIGO-Virgo detector noise and extraction of transient gravitational-wave signals", Class Quantum Grav 37, 055002 (2020)
- Books
 - Maggiore, "Gravitational Waves: Volume 1: Theory and Experiments"
 - Creighton & Anderson, "Gravitational-Wave Physics and Astronomy: An Introduction to Theory, Experiment and Data Analysis"