Likelihood and parameter estimation acceleration

Sylvain Marsat (L2IT, Toulouse)



GdR DA

Lyon, IP2I - 2022-11-15

Introduction

- Bayesian parameter estimation: $p(\theta|d) = \frac{p(d|\theta)p(\theta)}{p(d)}$
- Performance of likelihood function is crucial (millions of evals) performance of PE algorithm is also crucial, time to converge can be unreasonable
- Standard likelihood for stationary Gaussian noise: $\ln \mathcal{L} = -\frac{1}{2}(h d|h d)$

$$(a|b) = 4\operatorname{Re} \int df \; \frac{\tilde{a}(f)\tilde{b}^*(f)}{S_n(f)} \qquad (a|b) = 4\operatorname{Re}\Delta f \sum_i \frac{\tilde{a}(f_i)\tilde{b}^*(f_i)}{S_n(f_i)} \qquad f_i = i\Delta f$$

- Ingredients: fast waveforms, fast likelihoods, fast PE algorithms
- Accelerating waveforms: Reduced Order Models (ROMs), Surrogates crucial for NR and EOB
- Acceleration of likelihoods go beyond simply using a fast waveform important for data interface

- How long are GW signals ? Ground-based detectors and LISA
- Accelerating the likelihood: multibanding, heterodyning, Reduced Order Quadratures
- Accelerating PE: burn-in vs sampling, marginalization/ extremization, fast/slow parameters
- Dealing with degeneracies: example of MBHBs for LISA

How long are GW signals ? Ground-based detectors and LISA

- Accelerating the likelihood: multibanding, heterodyning, Reduced Order Quadratures
- Accelerating PE: burn-in vs sampling, marginalization/ extremization, fast/slow parameters
- Dealing with degeneracies: example of MBHBs for LISA

How long are GW signals ?





Time-frequency relation

With $M_s = GM/c^3$ the total mass in seconds $(M_{\odot} = 5 \times 10^{-6}s)$:

$$f = \frac{1}{\pi M_s} \left[\frac{256\nu}{5} \frac{\Delta T}{M_s} \right]^{-3/8}$$

$$\Delta T = M_s \frac{5}{256\nu} \left[\pi M_s f \right]^{-8/3}$$

at leading order, for a circularized binary

Nyquist sampling

 $\Delta t = 1/f_s$ $f_s = 2f_{\text{Nyquist}}$

 $f_{\rm Nyquist}$ highest frequency present in the signal

How long are signals for ground-based detectors ?



Frequency	/	Hz
-----------	---	----

$\begin{array}{c} \text{BBH} \\ M = 65 M_{\odot} \end{array}$	ΔT	$N \\ f_s = 4096 \text{Hz}$	$\Delta \phi_{ m orb}/(2\pi)$
$f_{\rm min} = 20 {\rm Hz}$	0.8 s	3.5e3	14
$f_{\rm min} = 10 {\rm Hz}$	5.5 s	2.3e4	44
$f_{\rm min} = 5{ m Hz}$	35 s	1.4e5	140

BNS $M = 2.8 M_{\odot}$	ΔT	$N f_s = 4096 \mathrm{Hz}$	$\Delta \phi_{ m orb}/(2\pi)$
$f_{\rm min} = 20 {\rm Hz}$	163 s	6.7e5	2600
$f_{\rm min} = 10 {\rm Hz}$	1030 s	4.2e6	8300
$f_{\rm min} = 5 \mathrm{Hz}$	6560 s	2.7e7	26000

LISA signals

- **MBHBs**: very loud, mergerdominated (mostly short)
- **SBHBs**: early inspiral, some chirping during LISA obs. (multiband ?)
- **GBs**: quasi-monochromatic, superposed
- **EMRIs**: long-lived, many harmonics
- Stochastic backgrounds



How long are LISA signals ? MBHBs

- How long before merger can we detect the signal ?
- SNR=10 to claim detection

Astrophysical models [Barausse 2012]:

- Heavy seeds delay
- Heavy seeds no delay
- PopIII seeds delay



SBHB signals

- Can last for years in the LISA band and chirp to exit the band at high frequencies
- Data with Nyquist sampling 0.2Hz for 10yrs: N=6.3e6

GB signals

- Last for the whole mission
- Quasi mono-chromatic, very compact in Fourier domain
- Millions superposed, ~20000 resolvable !

EMRI signals

- Can last for months
- Waveform generation and representation very challenging, complex signal with many harmonics

- How long are GW signals ? Ground-based detectors and LISA
- Accelerating the likelihood: multibanding, heterodyning, Reduced Order Quadratures
- Accelerating PE: burn-in vs sampling, marginalization/ extremization, fast/slow parameters
- Dealing with degeneracies: example of MBHBs for LISA

Accelerating likelihoods: 'multibanding'



$$N = f_{\rm max} / \Delta f = f_{\rm max} \ast \Delta T$$

Max frequency set by merger, duration dominated by inspiral

Nyquist sampling is far from optimal for chirping signals

Solution: separate data in signal-adapted bands

Interpolation acceleration

If on each band, phase is linearly interpolated, replace exp. with mult.

 $e^{i\lambda(j+1)\Delta f} = e^{i\lambda j\Delta f} e^{i\lambda\Delta f}$

Generalizes to higher-order polynomials



Accelerating likelihoods: quasi-monochromatic heterodyning

GB signals in LISA: quasi-monochromatic

$$\Phi = \Phi_0 + 2\pi \left(f_0 t + \frac{1}{2} \dot{f}_0 t^2 + \dots \right)$$
$$\frac{\dot{f}_0 T}{f_0} \ll 1$$

Sidebands: created by chirp and modulation of the LISA response

 Idea: factor out the fast-varying part, work with the slow-varying part [Conish-Littenberg 2007]

 $s(t) = \overline{s}(t) \exp[2i\pi f_0 t]$

- FFT of the slow part, with a very reduced Nyquist frequency (bandwidth: from 32 to 512)
- Go back to original signal with a simple shift in frequency



Accelerating likelihoods: heterodyning

Overview

- Structure of the likelihood $\ln \mathcal{L} = -\frac{1}{2}(h - d|h - d)$ $(a|b) = 4 \operatorname{Re} \int df \ \frac{\tilde{a}(f)\tilde{b}^*(f)}{S_n(f)}$ $h = Ae^{i\Phi}$ smooth amp/phase d numerical data
- Introduce a reference waveform $\overline{h}(f)$ $\zeta(f) \equiv h(f)/\overline{h}(f)$ now slowly variable in the vicinity of reference parameters
- Separate integrand in slowly and rapidly variable parts

$$(h|d) \sim \int df \, \frac{\overline{h}d^*}{S_n} \times \zeta \qquad (h|h) \sim \int df \, \frac{\overline{hh}*}{S_n} \times \zeta \zeta^*$$

- Interpolate and precompute
 - ζ interpolated on a coarse, reduced grid

$$\begin{split} (h|d) &\sim \sum_{i} \int_{f_{i}}^{f_{i+1}} df \, \frac{\overline{h}d^{*}}{S_{n}} \times (a_{i} + b_{i}f) \\ (h|h) &\sim \sum_{i} \int_{f_{i}}^{f_{i+1}} df \, \frac{\overline{h}h^{*}}{S_{n}} \times (a_{i} + b_{i}f + c_{i}f^{2}) \end{split}$$

- Evaluate
 - h on coarse grid, then sum weights and coeffs

[Cornish 2010, Cornish 2021]

[Zackay+ 2018] (relative binning)

Usage in practice

- Small reduced grid (N~100)
- Different interpolation methods (linear, polynomial)
- Requires reference waveform (first guess for signal parameters) — can be updated on the way
- Distinguish burn-in from actual sampling, the latter happens close to the true signal



Accelerating likelihoods: heterodyning example for MBHB

Decomposing the likelihood:

$$\ln \mathcal{L} = -\frac{1}{2}(s - d|s - d)$$

= $-\frac{1}{2}(s - s_0|s - s_0) + (s - s_0|d - s_0) - \frac{1}{2}(s_0 - d|s_0 - d)$

Residuals from reference waveform:

 $s_{\ell m} - s_{\ell m}^0 = r_{\ell m} e^{i\Phi_{\ell m}^0}$

Implementation:

$$(s - s_0 | s - s_0) = \sum_{\ell m} \sum_{\ell' m'} (r_{\ell m} r^*_{\ell' m'} | e^{i(\Phi^0_{\ell' m'} - \Phi^0_{\ell m})})$$
$$(s - s_0 | d - s_0) = \sum_{\ell m} (r_{\ell m} | e^{-i\Phi^0_{\ell m}} (d - s_0))$$

- Fix a sparse frequency grid (~128)
- Linear interpolation of the residuals, mode-by-mode
- Precompute 0-th and 1st polynomial inner products against phase and data terms, with a fine resolution



Accelerating likelihoods: ROQs

Linear ROM for GW signal

Building a reduced basis and empirical interpolant directly for h_+, h_\times and products $h_A h_B (A = +, \times)$:

$$h[\theta](f) = \sum_{j=0}^{p} B_j(f)h[\theta](f_j^{\text{node}}) \qquad h_A[\theta]h_B[\theta](f) = \sum_{k=0}^{q} C_k(f)h_A[\theta]h_B[\theta](f_k^{\text{node}})$$

p,q size of the reduced basis (linear and quadratic)

- + efficient representation for likelihood
- larger basis for longer signals, challenging to build

Likelihood evaluation

- Precompute all inner products $(B_j|d) \ (C_k|1)$
- Evaluate waveform model on interpolant nodes

 $\lambda_j = h[\theta](f_j^{\text{node}}) \qquad \mu_k = h_A[\theta]h_B[\theta](f_k^{\text{node}})$

• Use linear/quad structure to compute likelihood

$$\ln \mathcal{L} = -\frac{1}{2}(h - d|h - d)$$

$$\ln \mathcal{L} = \sum_{j=0}^{p} \lambda_j (B_j|d) + \sum_{k=0}^{q} \mu_k (C_k|1) + \text{const}$$



- How long are GW signals ? Ground-based detectors and LISA
- Accelerating the likelihood: multibanding, heterodyning, Reduced Order Quadratures
- Accelerating PE: burn-in vs sampling, marginalization/extremization, fast/slow parameters
- Dealing with degeneracies: example of MBHBs for LISA

Accelerating PE: burn-in vs sampling



- Simulating realistic PE: start from prior
- Prospective parameter estimation, only interested in final result: cheat with initialization

Techniques for burn-in (search) or sampling can differ !

Accelerating PE: marginalization, optimization

Separate parameters to reduce dimensionality:

 θ 'interesting' parameters (intrinsic) λ parameters to eliminate (extrinsic)

Marginalization

$$p(\theta|d) = \int d\lambda \; p(\theta, \lambda|d)$$

review: [Talbot-Thrane 2018]

- Approximate marginalization on time: IFFT of integrand mimics time shifts
- Phase marginalization for 22-only signals (likelihood becomes modified Bessel function)
- Distance marginalization

Optimization

Useful at search stage, for faster burn-in

- F-stat to build a proposal
- Directly search 'sampling' F-stat as a pseudo-likelihood

$$\mathcal{F} = \max_{\lambda} \ln \mathcal{L}(\theta, \lambda)$$

- F-stat analytical over distance, inclination, phase, polarization
- LISA GB in low-frequency approximation: +sky
- LISA MBHB low-f and short signal approximation: +sky

Accelerating PE: fast and slow parameters

Separate parameters for computational efficiency:

- θ 'costly' parameters (intrinsic)
- λ 'cheap' parameters (extrinsic)

Intrinsic parameters: requires to solve GR (analytical models, numerical relativity)

Extrinsic parameters: geometry and signal propagation, simple and completely universal

Gibbs sampling

Successive sampling steps by blocks

$$\theta_{i+1} \sim p(\theta | \lambda_i)$$

$$\lambda_{i+1} \sim p(\lambda | \theta_{i+1})$$



- resolving correlations might be costly

Likelihood pre-interpolation

Pre-interpolate intrinsic likelihood (e.g. Gaussian Process Regression)

- LIGO/Virgo: RIFT [Lange+ 2017]
- LISA GB: [Strub+ 2022]



Accelerating PE: dealing with degeneracies

Parallel tempering

- Introduce parallel chains with temperatures, posterior: $p(\theta)^{\beta_i}$ $\beta_i = 1/T_i$
- Propose swaps with acceptance:

$$p_{\text{swap}} = \min\left[1, \left(\frac{p(\theta_i)}{p(\theta_j)}\right)^{\beta_j - \beta_i}\right]$$



• Crucial for robustness, avoids being stuck in a local maximum and ensures exploration of the parameter space

Tailored proposals

- In presence of known degeneracies, include jumps in proposal
- Very efficient for very disconnected multimodal posteriors



- How long are GW signals ? Ground-based detectors and LISA
- Accelerating the likelihood: multibanding, heterodyning, Reduced Order Quadratures
- Accelerating PE: burn-in vs sampling, marginalization/ extremization, fast/slow parameters

• Dealing with degeneracies: example of MBHBs for LISA

Dealing with degeneracies: MBHB example

Toy problem, completely degenerate extrinsic 22 likelihood without motion and high-f effects



Dealing with degeneracies: parameter map



MBHB example: I) F-statistic search on small data segments



MBHB example: II) initial PE with low frequencies



MBHB example: III) sampling with all frequencies



LISA data - band-passing, whitening



- **Band-passing**: select frequencies below 2mHz
- Whitening: work with signal/noise, so that all frequencies/times contribute equally

LISA data - band-passed, whitened in time domain



LISA data - band-passed, whitened in time domain

Whitened, band-passed data 2010TDI A 0 -10-201.11 1.12 1.13 1.15 1.14 1.101.16t (s) $\times 10^7$ sigma-thresholding for detection 100 -104.7854.7904.7954.8054.800 $\times 10^{6}$ 5.02.50.0 -2.5-5.08.742 8.744 8.746 8.7488.750

Noise properties

Noise PSD

- Noise autocorrelation function: (stationarity: depends only on τ)
- Noise PSD formal definition:
- Stationarity: independance in FD $\langle \tilde{n}(f)\tilde{n}^*(f')\rangle = \frac{1}{2}S_n(f)\delta(f-f')$
- Gaussianity: noise in freq. bins is Gaussian



$$K(\tau) = \langle n(t)n(t+\tau) \rangle$$

$$S_n(f) = 2 \int d\tau e^{2i\pi f\tau} K(\tau)$$

$$\sum_{n=1}^{\infty} \langle \tilde{x}(f) \tilde{x}^*(f') \rangle = \frac{1}{2} \sum_{n=1}^{\infty} \langle f \rangle \tilde{x}(f)$$

$$\tilde{n}(f) \sim \mathcal{N}(0, \frac{1}{2\Delta f}S_n(f))$$



Less-than ideal assumptions for LISA ! Non-stationarity, glitches...

Likelihood

- Likelihood: $\mathcal{L} = p(\text{data}|\text{signal params})$
- PDF of the noise: collection of independent $\ln p(n = n_i) = \text{const} \frac{1}{2} \sum_i \Delta f \frac{2}{S_n(f)} |\tilde{n}_i|^2$ Gaussian noise variables in each bin
- Likelihood is the probability that the noise makes up for the difference between observed data and theoretical signal: $d = h(\theta) + n$ $\ln p(d|\theta) = \ln p(n = d - h(\theta)) = -\frac{1}{2}(d - h(\theta)|d - h(\theta))$

Bayesian formalism

• Matched-filtering overlap:
$$(h_1|h_2) = 4 \operatorname{Re} \int df \, \frac{\tilde{h}_1(f)\tilde{h}_2^*(f)}{S_n(f)}$$

• For Gaussian, stationary noise, for independent channels:

$$\ln \mathcal{L}(d|\theta) = -\sum_{\text{channels}} \frac{1}{2} (h(\theta) - d|h(\theta) - d)$$

$$d = h(\theta_0) + n_0$$

• Bayes theorem defines the posterior:

$$\theta|d) = \frac{\mathcal{L}(d|\theta)p_0(\theta)}{p(d)}$$

- h GW signal
- θ parameters
- d data stream
- θ_0 signal params.
- n_0 noise real.
- S_n noise PSD

 $p_0(\theta)$ prior p(d) evidence

p(

Signal-to-noise (SNR)

Measures loudness of signal:

$$\mathrm{SNR}^2 = (h|h) = 4 \int \frac{df}{S_n} |h|^2$$

Simple detection statistics: SNR>8-10 (true detection statistics LIGO/Virgo more complicated)

Fisher matrix analysis

• Quadratic expansion of log-likelihood around true signal, approx. likelihood as a Gaussian

$$h(\theta) = h(\theta_0) + \Delta \theta_i \partial_i h + \dots$$

$$\ln \mathcal{L} = -\frac{1}{2} \Delta \theta_i F_{ij} \Delta \theta_j + \mathcal{O}(\Delta \theta^3)$$

$$F_{ij} = (\partial_i h | \partial_j h)$$

- Matrix inversion to get to the covariance of the Gaussian $C = F^{-1}$
- Valid at high SNR, and misses degeneracies

Bayesian sampling tools

- MCMC methods, nested sampling
- MCMC proposals: ensemble samplers (emcee), differential evolution, ...
- Parallel tempering: explore full parameter space
- Informed proposals to deal with degeneracies

Levels of approximation

- Fisher: for high SNR limit (depends on signal !)
- Set noise realization to 0
- Initialize MCMC from Fisher
- Full run with initialization from priors
- Full run with noise
- Superposition of sources, unknown noise, noise artifacts...

GW signals seen by LISA - the basics



LISA: different BHB signals

- **MBHBs**: very loud, mergerdominated (mostly short)
- **SBHBs**: early inspiral, some chirping during LISA obs. (multiband ?)
- **GBs**: quasi-monochromatic, superposed
- **EMRIs**: long-lived, many harmonics
- Stochastic backgrounds
- TDEs !

Phases of the signal:

- **Inspiral**: covered by post-Newtonian (PN) perturbative series
- **Merger**: covered only by numerical relativity (NR)
- **Ringdown**: NR, superposition of Quasi-Normal Modes (QNM)



Contrasting LIGO/Virgo and LISA responses: LISA

LISA-frame



Low-f approximation: **two LIGO-type detectors** in motion [Cutler 1997]



High-f: **three channels** with complicated frequency-dependence

Sky localisation from the modulations induced by the orbits for long-lived signals

Sky localization can also come from high-f effects.

Degeneracies - multimodality in the sky possible !

Massive black holes: signals and challenges

- Very loud sources, SNRs of several thousands !
- Detection of merger easy, but detection as early as possible ?
- Advance localization for multimessenger observations ?
- Signals can be short (< I day) and degenerate
- Waveform model systematics for such loud signals ? Biases, residuals for other sources ?
- Subdominant features in the signal are important



Galactic binaries: signals and challenges

- Mostly WD-WD, some other compact objects
- Full galaxy: ~20 million systems !
- About ~20000 individually resolvable
- Form a (non-stationary) background
- Verification binaries
- Quasi-monochromatic GW emitters
- Modulation by LISA motion (sidebands in Fourier-domain)
- Superposition of signals in Fourier-domain





38

Stellar-mass BHs

- Quiet signals: a few detections in the LISA band
- Inspiral regime far from merger, very large number of cycles $(10^5-10^6)\,$
- Challenge of detection: template banks impossible
- Multiband analysis, archival searches ?



LISA data - LDC-2 Sangria



- **MBHBs**: chirping signals, emerging from low-f noise
- **GBs**: quasi-monochromatic, horizontal lines

LISA data - LDC-2 Sangria Time-Domain



- **MBHBs**: loudest ones clearly visible by eye above the noise
- **GBs**: superposed signals, annual modulation due to the LISA motion

LISA data - LDC-2 Sangria Frequency-Domain



- **MBHBs**: loudest ones visible in the spectrum, subdominant
- **GBs**: signals local in frequency, both individually resolvable and building up a background

LISA Fourier-domain response

Response

Laser frequency shift, spacecrafts s to r through link I: $y = \Delta \nu / \nu$

$$y_{slr} = \frac{1}{2} \frac{1}{1 - \hat{k} \cdot n_l} n_l \cdot (h(t_s) - h(t_r)) \cdot n_l$$



Transfer function for modulated and delayed signal $FT[F(t)h(t + d(t))] = \mathcal{T}(f)\tilde{h}(f)$

Fourier-domain for **chirping signals** (separation of timescales):

$$\mathcal{T}_{slr} = \frac{i\pi fL}{2} \operatorname{sinc} \left[\pi fL \left(1 - k \cdot n_l\right)\right] \exp\left[i\pi f \left(L + k \cdot \left(p_r + p_s\right)\right)\right] n_l \cdot P \cdot n_l(t_f)$$

Time and frequency-dependency Time: motion of LISA on its orbit Frequency: departure from long-wavelength

+ Time-delay interferometry (TDI)

linear combinations of y_{slr} with more delays



Orbits











LISA sources

Terminology:

- Massive black holes binaries (MBHBs)
- Stellar-mass black hole binaries (SBHBs): masses observable by ground-based detectors [Sesana 2016]
- Galactic Binaries (GBs): mostly WD-WD
- Extreme Mass Ratio Inspirals (EMRIs)
- Sochastic backgrounds (GBs, cosmo.)
- TDEs !



Contrasting LIGO/Virgo and LISA responses: LIGO/Virgo

Pattern functions

Simple multiplicative response

 $s = F_+h_+ + F_\times h_\times$

Angular dependence:

$$F_{+} = \frac{1}{2} \left(1 + \cos^{2} \theta \right) \cos \left(2\phi \right) ,$$

$$F_{\times} = \cos \theta \sin \left(2\phi \right)$$

Time-of-arrival triangulation

- Two detectors: ~ring on the sky
- Better localization for 3 or more detectors (even low SNR!)



One-arm frequency observables

From spacecraft s to spacecraft r through link s: $y = \Delta \nu / \nu$

$$y_{slr} = \frac{1}{2} \frac{1}{1 - \hat{k} \cdot n_l} n_l \cdot (h(t_s) - h(t_r)) \cdot n_l$$

$$t_s = t - L - \hat{k} \cdot p_s, \quad t_r = t - \hat{k} \cdot p_r$$

$$h = h_+ P_+(\hat{k}) + h_\times P_\times(\hat{k}) \quad \text{GW at SSB}$$

Time-delay interferometry (TDI)

- Crucial to cancel laser noise
- First generation: unequal arms
- Second generation: propagation and flexing
- Michelson X,Y,Z Uncorrelated noises A,E,T

Approximations

- Long-wavelength approximation: two moving LIGOs rotated by $\,\pi/4\,$ + orbital delay
- Rigid approximation (order of the delays does not matter, delay=L simple in Fourier domain)

46



$$\begin{split} X_{1}^{\mathrm{GW}} &= \underbrace{\left[(y_{31}^{\mathrm{GW}} + y_{13,2}^{\mathrm{GW}}) + (y_{21}^{\mathrm{GW}} + y_{12,3}^{\mathrm{GW}})_{,22} - (y_{21}^{\mathrm{GW}} + y_{12,3}^{\mathrm{GW}}) - (y_{31}^{\mathrm{GW}} + y_{13,2}^{\mathrm{GW}})_{,33}\right]}_{X^{\mathrm{GW}}(t)} \\ &- \underbrace{\left[(y_{31}^{\mathrm{GW}} + y_{13,2}^{\mathrm{GW}}) + (y_{21}^{\mathrm{GW}} + y_{12,3}^{\mathrm{GW}})_{,22} - (y_{21}^{\mathrm{GW}} + y_{12,3}^{\mathrm{GW}}) - (y_{31}^{\mathrm{GW}} + y_{13,2}^{\mathrm{GW}})_{,33}\right]_{,2233}}_{X^{\mathrm{GW}}(t-2L_2-2L_3) \simeq X^{\mathrm{GW}}(t-4L)} \end{split}$$