Hierarchcical search EMRI signal via matched filtering: insight and progress

Xiaobo Zou, HIAS,UCAS

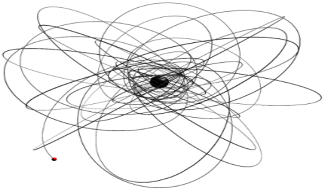
2025/10/19 Qiandaohu

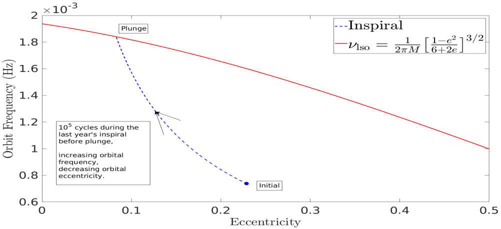
Collabrator: Prof.Soumya.Mohanty and Prof.Wenbiao Han

Outline I

1. Overview of EMRI data analysis
2. Methods
3. Results
4. Unsolved issues

Overview I

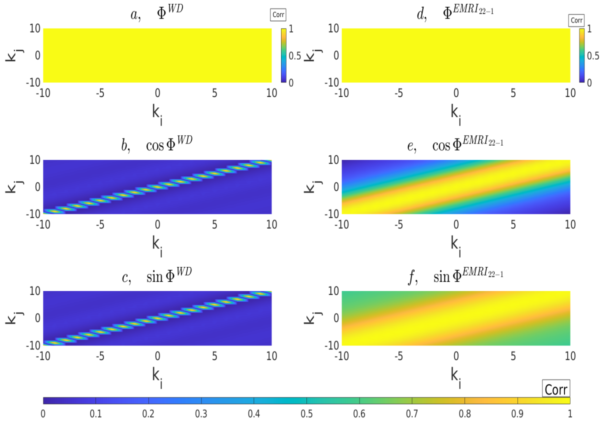




1. inspiral signal over several years
2. orbital motion + pericenter precessing + orbital plane precessing
3. 14D(more e.g., 15D) parameter space, $\mu, M, \lambda, S/M^2$, $e_0, \underline{\nu_0}$, $(\delta\tilde{Q})$ phase-coupled θ_s, ϕ_s, detector motion $\theta_k, \phi_k, \phi_0, \tilde{\gamma}_0, \alpha_0, \underline{D}$ constant modulation

Waveform modeling is challengeable and computationally expensive, currently Analytical kluge (AK) is used in LISA Data Challenge.

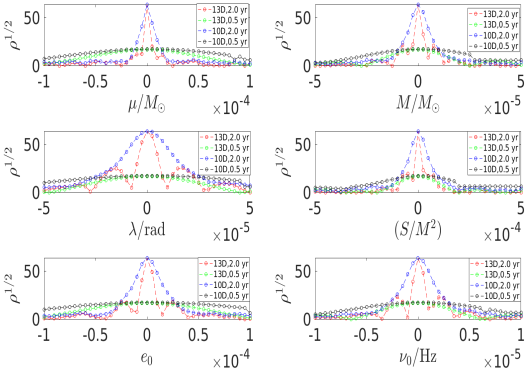
Overview II



1. 1st column,
   $\Phi_{\rm GB} = 2\pi ft$,
   $f = f_0 + k * 10^{-5}$,
   $k = -10 : 10$
2. 2nd column,
   $\Phi^{22-1}_{\rm EMRI} = 2\phi + 2\tilde{\gamma} - \alpha$,
   $\mu = \mu_0 + k * 4 \times 10^{-7}$,
   $k = -10 : 10$
3. The correlation coefficients (Corr) among varied Φ, $\sin\Phi$ and $\cos\Phi$ reveal the sensitivity.

The tiny change of phase-coupled parameters significantly change the GW waveform.

Overview III



1. The phase-coupled parameters contribute a extremly 6D sharp peak as GW phase sensitivity.
2. The longer the signal is, the sharper the primary peak is.

The signal morphology is a 6D (phase-coupled parameters) modulated sharp primary peak.

Overview IV

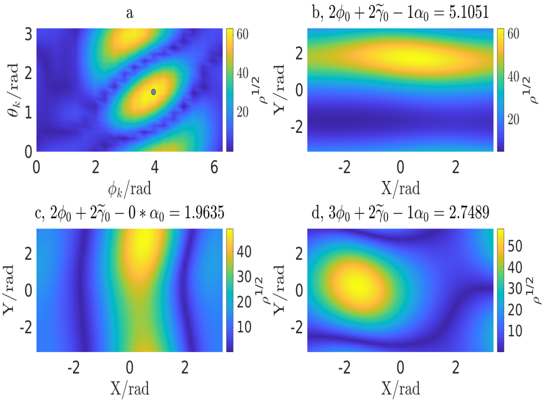


Figure: The 5D (θ_k,ϕ_k,ϕ_0,$\widetilde{\gamma}_0$,α_0 (constant modulations)) subspace has simple landscape with few peaks and no extrem sharpness.

Overview V

The primary peak is surrounded by multiple secondary peaks.

1. $h = \sum_{n2m} A_{nm} e^{i\Phi^{n2m}}$. The primary peak corresponds to the match of all the harmonics, while the partial match result in secondary peaks.
2. The interaction among the three subset of parameters produce secondary peaks.

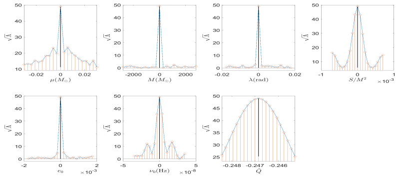
Overview VI

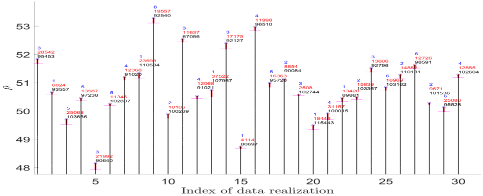
Table: The peak width is $\sim 1\sigma$, range 1 is currently what we can handle, range MLDC/LDC correspond to the general search space.

Peak params \ ranges width in σ	peak width	range 1 width	MLDC width	LDC width
$\mu(M_\odot)$	1	~ 20	~ 20	~ 400
$M(M_\odot)$	1	~ 20	~ 20	~ 600
λ	1	~ 20	~ 320	~ 350
S/M^2	1	~ 20	~ 40	~ 160
e_0	1	~ 20	~ 400	~ 1100
ν_0	1	~ 100	$\sim 10^5$	10^5 (FD) 10^6 lags (TD)
Relative volumn of the $5D$ search range (no ν_0)	1	$\sim 3.2 \times 10^6$	$\sim 2 \times 10^9$	$\sim 1.5 \times 10^{13}$
CPU cores needed	$\sim$	1000 **(DONE)**	10^4 or 10^5 **TODO**	10^5 **TODO**

MLDC (2005-2013 yr), LDC (2018-now yr)

The astrophysically motivated search range is too wide so that it is a Needle-in-the-Haystack for EMRI data analysis.





1. when considering bumpy-AK, $\tilde{Q} = \tilde{Q}_{\text{kerr}} + \delta\tilde{Q}$,
2. the dimensionality turns to 15D, the primary peak turnd to 7D.
3. The degernate peaks have higher SNR than injection, bias the parameter estimation.

The $\delta\tilde{Q}$ contribute extra degeneracy to likelihood surface of bumpy-AK besides the intrinsic nonlocal degeneracy of AK itself.

Methods I

1. **Time frequency method**
   fitting excess power beyond the noise background.
   Ref: Wenlin Qing&Gair series works during 2005∼2008.
2. **Machine learning method**
   Use AI for signal detection and partial parameter estimation.
   Ref: Xueting Zhang 2020, Tianyu Zhao 2023, Qianyun Yun 2023, Bo Liang 2024 and 2025
3. **Matched filtering**

$$SNR = \frac{(d|h)}{\sqrt{(h|h)}}, \quad (a|b) = \int_{-f}^{f} \frac{\tilde{a}\tilde{b}^* + \tilde{a}^*\tilde{b}}{S_n} df.$$

maximal match the phase between data and template model.
Ref: Babak 2008, Cornish 2008, Ali 2013, Changqin Ye 2023, ZXB 2024, Cole 2025, Strub 2025

Methods: Dimension reduction I

The EMRI's 14 parameters Θ could be split as following,
$\Theta = \Theta_1 \cup \Theta_2 \cup \Theta_3$,

1. $\Theta_1 = \{\mu, M, \lambda, S/M^2, e_0, \nu_0, \}$, those 6 phase-coupled parameters, contribute to the GW phase evolution, producing a 6D sharp primary peak. In which, e_0,ν_0 are more sensitive as being initials.
2. $\Theta_2 = \{\theta_s, \phi_s\}$, the sky location bring detector motion to GW time dependent quantities.
3. $\Theta_3 = \{\theta_k, \phi_k, \phi_0, \tilde{\gamma}_0, \alpha_0, D\}$, those extrinsic parameters, contribute time independent modulation to the GW amplitude and phase.

Methods: Dimension reduction II

We proposed dimension reduced likelihood via variable separation in likelihood function via waveform decomposition and nested optimization,

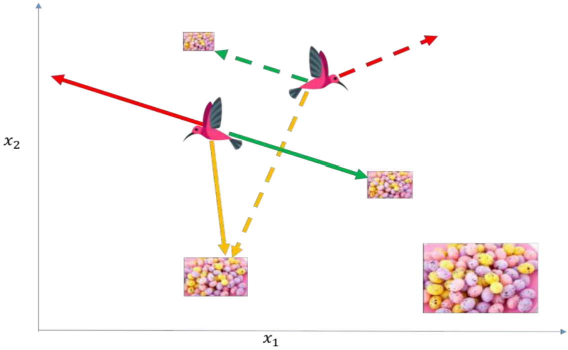
$$\Lambda(\Theta) = \sum_{I\in\{A,E\}} \left[-(\bar{h}^I(\Theta)|\bar{h}^I(\Theta)) + 2(\bar{d}^I|\bar{h}^I(\Theta))\right] , \quad (1)$$

$$\bar{h}_I^i(\Theta) = \sum_{p=1}^{4} a_p^i(\Theta_3)\bar{x}_{I,p}^i(\Theta_1,\Theta_2) , \quad (2)$$

$$(\bar{d}^I|\bar{h}^I(\Theta)) = \sum_{i=1}^{N}\sum_{p=1}^{4} a_p^i(\Theta_3)(\bar{d}^I\left|\bar{x}_p^{I,i}(\Theta_1,\Theta_2)\right) , \quad (3)$$

$$(\bar{h}^I(\Theta)|\bar{h}^I(\Theta)) = \sum_{i=1}^{N}\sum_{j=1}^{N}\sum_{p=1}^{4}\sum_{q=1}^{4} a_p^i(\Theta_3)a_q^j(\Theta_3)(\bar{x}_p^{I,i}\left|\bar{x}_q^{I,j}\right) , \quad (4)$$

8D/**7D** likelihood: analystical maximization over D **+ lag shift for** ν_0+ global optimizer for Θ_1, Θ_2 (8D/**7D**)+ local optimizer for Θ_3 (5D), they are used in a nested way. (*ZXB 2024*)



1. the figure is from Mohanty's book in 2018.
2. PSO is a point estimation method, no posterior.
3. PSO beongs to genetic algorithm family, it can provides the better initial start location for MCMC.

The PSO is suitable for peak morphology of EMRI signal.

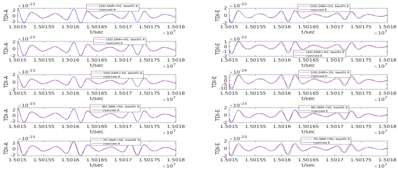


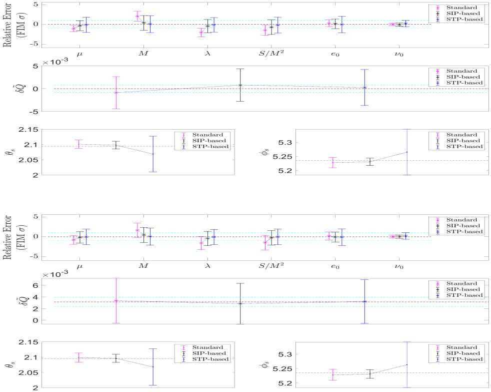
Table 3. PSO outputs of 8-dimensional searches. The square root of the fitness value at the true 8-dimensional location is **47.879594**. Further details about the table are discussed in Section 5.

	1st PSO	2nd PSO	3rd PSO	4th PSO
Square root of fitness values				
Best location from PSO	47.546001	46.381273	47.069351	**47.988164**
Parameter estimation errors				
$\mu(M_\odot)$	−3.1	−2.3	0.21	2.4
$M(M_\odot)$	1.9	2.1	−1.1	−2.6
λ(rad)	−2.1	−2.1	0.96	2.5
S/M^2	−2.2	−2.2	0.91	2.5
e_0	7.8	2.9	3.6	−1.2
ν_0(mHz)	−6.8	−4.5	−8.2	−1.9
D(Gpc)	−0.03	0.00011	−0.12521	0.015
θ_s(rad)	0.068	−0.078970 *	0.13	−0.012
ϕ_s(rad)	0.015	−0.167177 *	−0.0062	0.046
Overlap between the estimated and true signals				
ff_A	−0.970817	0.972518	0.964058	−0.990312
ff_E	−0.965563	0.940148	0.939171	−0.982537
ff_{AE}	−0.968851	0.959972	0.954244	−0.987405

1. Using narrow prior, or search range, $\pm 20\sigma$ around the injection, we can locate it.
2. the olverlap betweem the recovered signal and the inection is good, and the parameter estimation error is small within $\sim 1\sigma$ for the most successful PSO run.

We need hierarchical search to cover the Needle-in-the-Haystack.

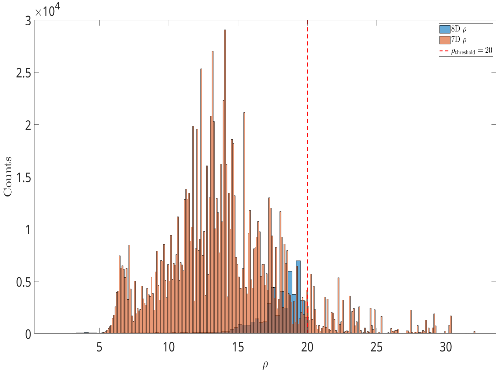
Results: Search bumpy-AK signal using narrow prior (Zou 2025) I



1. The narrow prior is ± 10 FIM σ for phase-coupled parameters.
2. We use PSO to collect the degenerate peaks, superinjection (SIP) and superthreshold (STP), their average statistics can mitigate the bias of parameter estimation.

We need hierarchical search to cover the Needle-in-the-Haystack and dual layer degeneracy.

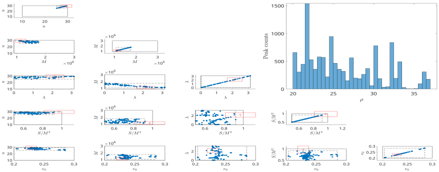
Results: Hierarchical search EMRI signal using wide prior I

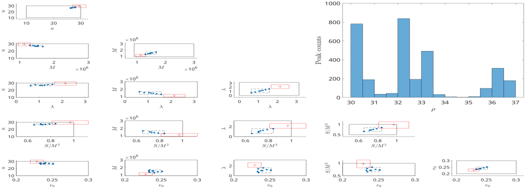


Using the population model predicted wide prior, only the $7D$ likelihood can locate the secondary peaks aring from GW, e.g., SNR threshold exceeds 20.

So it may be potential to use the secondary peaks to estimate the narrow prior in a hierarchichal way

Results: Hierarchical search EMRI signal using wide prior II

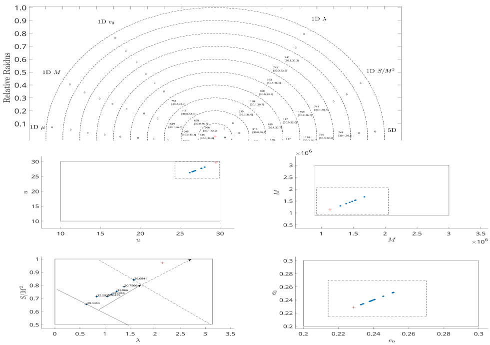




1. The wide prior is given by the population model.
2. we observe clustering of secondary peaks for u, M and e_0 with threshold 20, for λ, S/M^2 with threshold 30.

We are testing PSO to collect the secondary peaks and extract the narrow search range from them in a hierarchcal way.

Results: Hierarchical search EMRI signal using wide prior III



1. The distribution of secondary peaks in 5D is a band, which resulting from λ and S/M^2 where u, M, e_0 are always clucstering around the signal.
2. We can estimate the narrow prior based on the clustering.

We have find an inspiring clue to design our hierarchical search strategy.

Unsolved issues I

1. For Vacuum-GR EMRI, we need hierarchical search to overcome the Needle-in-the-Haystack issue within population model predicted prior.

2. For Beyond Vacuum-GR EMRI, e.g., the bumpy-AK, how to break the correlation and degeneracy between EMRI's parameters, environment parameters or other parameters related with no-GR effect?
The hierarchical search strategy need to be well designed to cover the dual layer degeneracy.

Acknowledgement

Thanks for the HIAS HPC.

Thanks for your attention.