

BHextractor / PCA Update & Notes

James A. Clark

Things we have going on / I know about

- Continuing BHextractor / Sant-Cugant work
 - 'September paper' (Kimbrell)
- NR catalogue studies
 - PCAT-style clustering (Magana)
 - Waveform / parameter Interpolation (London, Clark)
- Analytic Catalogue Studies
 - PCA Characterization & Exploration (Day)

BHextractor: previous work

- Small (20 waveforms) catalogues (Q, HR, RO3)
- Used all available harmonics, constructed optimally oriented (face-on) waveforms
- PCs computed from truncated, aligned & SNR-scaled waveforms
- Evidence computed in old SMEE matlab script for each catalogue
- No sky-location search, no time-search, no mass-search
- Single IFO
- Reasonable success in distinguishing the more 'distinct' waveforms
- No attempt at parameter estimation; goal was *only* separation of catalogues

BHextractor: Recent / On-going

- N. Mangini (now left): started looking at PCs with complex ($h_+ - ih_\times$) waveforms and handling inclination
- D. Leininger (Glasgow summer): investigated sky-localisation and multi-IFO application
- A. Lombardi (UMass): understand / incorporate search over total mass (amplitude scaling & \sim resampling)
- S. Kimbrell: tentative plan is to repeat & extend Sant-Cugat study using LALInference implementation of SMEE. J. Powell working to publish a branch of LAL with the SMEE edits.

Note: we also need

- 1 **careful treatment of orientation & harmonics**
- 2 a consistent and meaningful way to define catalogues (do we even *want* catalogues?)

Waveform Clustering

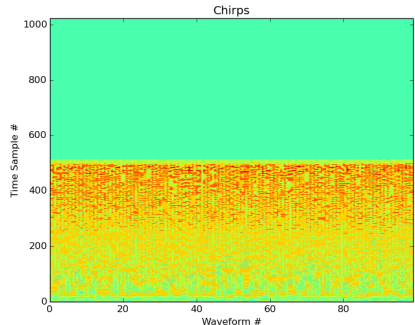
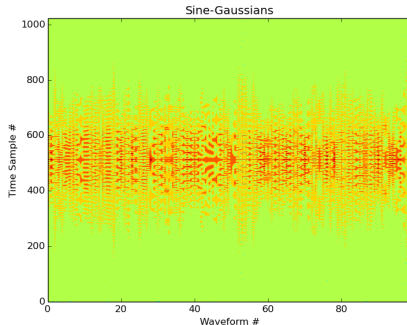
‘PCAT’: algorithm for glitch classification in detchar:

- 1 Construct a *single* catalogue of all waveforms
- 2 Perform PCA
- 3 Use GMM to identify clusters of principle component ‘scores’
- 4 Clusters in PC-score space represent morphologically similar waveforms

So far: prototype script to generate a catalogue comprised of sine-Gaussians & chirps (randomised params), perform PCA / clustering and identify the different families

- *Potential* use in constructing morphologically similar catalogues
- *Potential* use in grouping together similar waveforms for constructing sensible interpolants in PCA-based waveform interpolation ...

Waveform Clustering Demo / Prototype



500 waveforms of each type: algorithm identifies 2 distinct waveform morphologies from clustering the PCA scores with 48/52% membership, using just the first 50 PCs.

PCA & Parameter Estimation / Waveform Interpolation

Finally understand the fundamental idea behind waveform interpolation and connection with parameter estimation!

- ① Build catalogue of waveforms with varying (say) $q = m_1/m_2$; catalogue: $\{h(t|q_1), h(t|q_2), \dots\}$
- ② Compute PCs U_i and coefficients β_i where waveform is

$$h(t|q) = \sum_{i=1} \beta_i(q) U_i(t) \quad (1)$$

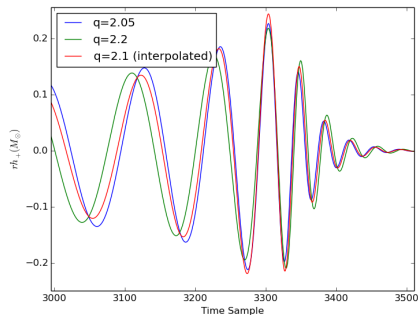
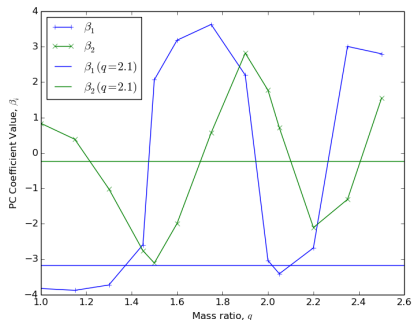
- ③ Realise we can interpolate *each* β_i as a function of q
- ④ E.g., have discretely sampled $\{\beta_1(q)\} = \{\beta_1(q_1), \beta_1(q_2), \dots\}$ (and similarly for β_2, \dots)
- ⑤ So we can get $h(t|q')$ for arbitrary q' by interpolating $\beta(q)$

PCA & Parameter Estimation / Waveform Interpolation

Example: Q-series has mass ratios

$q = \{1., 1.15, 1.3, 1.45, 1.5, 1.6, 1.75, 1.9, 2., 2.05, 2.2, 2.35, 2.5\}$;

interpolate to find $h(t|q = 2.1)$:



PCA & Parameter Estimation / Waveform Interpolation

- 1-D interpolation between (e.g.) mass ratios \sim trivial
- To be useful, we need multivariate interpolation to handle spins and internal orientation angles
- http://en.wikipedia.org/wiki/Multivariate_interpolation
- See also §8 of arXiv:1402.4146 - seems to be supported in Mathematica, Matlab but still casting around for python solutions
- Any local experience / expertise with *tensor product spline interpolation*???
- Note: it may well be useful / necessary to group like-waveforms together, essentially into distinct catalogues, each with their own interpolants; this is the *potential*¹ use of the PCAT-style clustering

¹no idea if this will be sensitive to different BBH waveforms!

PCA & Parameter Estimation / Waveform Interpolation

Other thoughts / concerns:

- Do we want to be stuck in the time-domain? (The SEOBNR ROMs are F-domain). F-domain is *strongly* preferred for any kind of parameter estimation.
- How can we use PCA & interpolation to identify under-sampled regions of parameter space in NR simulations?
- *What do we do with higher modes & unknown source orientation?*
 - Include orientation in PCA? Seems unfeasible / unnecessary / ugly
 - PCA & interpolants for *each* mode?

Where Next / Plans

- Sant-Cugant study using LALInference implementation of SMEE
 - will get evidence for ‘catalogues’ and posterior samples for β_i ; can we say something useful using these without going as far as waveform interpolation?
- ‘September paper’: overhaul of Sant-Cugant study
- Clustering: tidy up demo script, understand and characterise clustering behaviour, apply to BBH waveforms
- Interpolation: understand implementation of tensor spline interpolation
 - Call to be scheduled with J. Veitch next week to talk about what we could do
 - Particularly interested in identifying under-sampled regions of parameter space to guide NR simulations
 - B. Day’s studies with EOBNR & PCA could be a good test-bed for interpolation studies

Wavelets?

Maybe something useful to do with wavelet scaleograms (a la 'eigenfaces')?

Left side: Q-series, face-on; Right side: RO3-series, edge-on

