Greedy algorithm for building a reduced basis of gravitational wave templates

Scott Field[1] Chad Galley[2] Frank Herrmann[3]
Jan Hesthaven (Advisor)[4] Evan Ochsner[5] Manuel Tiglio[3]

[1]Brown University, Department of Physics
[2]California Institute of Technology, Jet Propulsion Laboratory
[3]University of Maryland, Department of Physics
[4]Brown University, Division of Applied Mathematics
[5]University of Wisconsin–Milwaukee, Center for Gravitation and Cosmology

January 14, 2011

Outline

- Gravitational wave templates and matched filtering
- Greedy construction of a reduced basis catalog
- Results: 2 parameter stationary phase approximation (SPA) waveforms at the 2nd post–Newtonian (PN) order
- Future work and conclusion

Gravitational wave templates and matched filtering

Physical Motivation: Gravitational Wave Sources

- We focus on astrophysical gravitational wave sources where two compact objects inspiral, merge, and ringdown
  - Binary neutron star (1-3 solar masses) and stellar sized binary black holes (3-30 solar masses)
  - Initial and advanced LIGO, Virgo
  - Detection rates of 1 to 10^3 per year (LIGO collaboration)

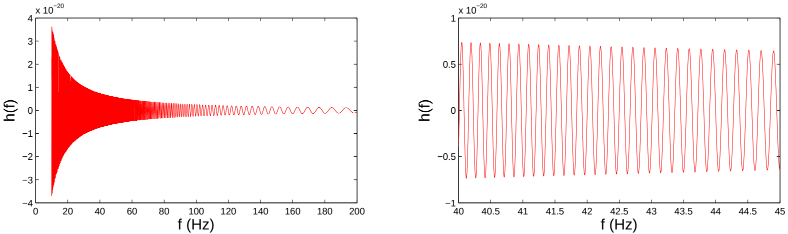


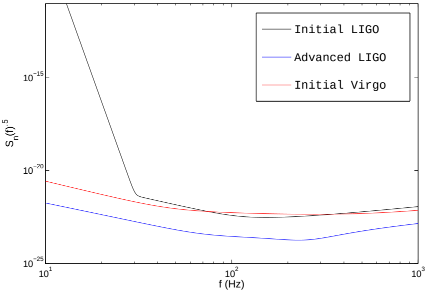
Figure: Chirp waveform for two 3 solar mass objects

Physical Motivation: Gravitational Wave Sources

- Let P parameters $\vec{\mu} = \{\mu_1, \ldots, \mu_P\}$ be associated with the sources (compact objects)
- Any relevant intrinsic or extrinsic parameters could be considered
  - Will specialize to compact objects' mass (2)
- Let $\mathcal{H}$ be the space of all normalized waveforms for the considered sources
  - Solutions to Einstein's equation, reduced models, analytic waveforms

Physical Motivation: Gravitational Wave detectors

- A passing gravitational wave causes a path length change in the interferometer's arm
  - $\Delta L/L \sim 10^{-20}$ and smaller
- A time series is recorded at some sample rate
- Data segments are Fourier transformed and analyzed
- Detector noise in frequency domain given by $S_n(f)$



Physical Motivation: Matched filtering

- Weak signals are buried in noise
- Gravitational waveform *templates* h from a *catalog* C are correlated with data s through a process known as matched filtering

$$< h, s >_{MF} = 4\mathrm{Re} \int_a^b \frac{hs^*}{S_n(f)} df$$

- When h normalized, signal–to–noise $\rho_C =< h, s >_{MF}$
- *Minimal match* measures the closeness of a catalog C w.r.t. the continuum $\mathcal{H}$

$$\min_{s \in \mathcal{H}} \max_{h \in C} < h, s >_{MF} \quad = \quad MM \leq 1$$

- $\rho_C \geq \rho_{\mathcal{H}} \times MM$
- "Detection" when $\rho_C \geq 8$ in multiple detectors

Physical Motivation: Matched filtering

Matched filtering bottlenecks...

- How do we populate the catalog? Metric based approach
  - Taylor expand analytic waveform expression in parameters
  - Analytic expressions and best coordinates to use
  - Metric must be worked out on a case by case basis
  - Large number of required templates goes like $(1 - MM)^{-P/2}$
- With 2 mass parameters, many thousands of templates in a catalog for $MM = .97$
- Performing $< h, s >_{MF}$ over many templates can be computationally expensive
- As the noise curve changes we must recompute the catalog
- What if waveforms are solutions to ODEs or PDEs?

Goal of talk

Introduce a generic reduced basis method for...

- Generation of an accurate and compact reduced basis space of waveforms
- Selection of nearly optimal parameter points

Will also show...

- Non–linear "space of waveforms" can be represented as a linear space with arbitrarily high accuracy
- $MM \sim 1 - 10^{-10}$ for *all* parameter values (and at low cost)
- Significantly fewer templates (basis) for a given MM
- Reduced basis space is robust against changes to noise curve

Greedy construction of a reduced basis catalog

Problem statement

- Given: P parameters $\vec{\mu} = \{\mu_1, \ldots, \mu_P\}$ and the space of all normalized waveforms $\mathcal{H}$
- Each waveform is denoted $h_{\vec{\mu}} \in \mathcal{H}$
- Seek an N dimensional linear space W_N to accurately represent $\mathcal{H}$
  - Ansatz: W_N is span of N waveforms chosen from $\mathcal{H}$
- W_N is called the *reduced basis space*

Notation: inner product, projection, and errors

- Weighted complex inner product (not to be confused with $<>_{MF}$)

$$< g, f > \equiv \int_a^b \frac{g f^*}{S_n(f)} df \qquad ||g|| = \sqrt{< g, g >}$$

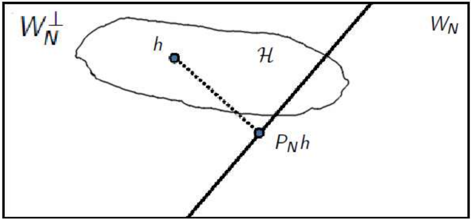
- Projection operator $P_N : \mathcal{H} \to W_N$ defined wrt to its action on the orthonormal basis e_i of W_N

$$P_N(h) \equiv \sum_{i=1}^{N} < e_i, h > e_i$$

- Orthogonal projection in sense of ...

P_N operator

- $\mathcal{H}^+$ such that $\mathcal{H} \subset \mathcal{H}^+$ and $\mathcal{H}^+$ is a Hilbert space
- Existence of $W_N^\perp = \{\delta h \in \mathcal{H}^+ :< \delta h, w >= 0 \ \forall w \in W_N\}$
- Unique decomposition $h = P_N h + \delta h$, $< P_N h, h - P_N h >= 0$
- In $\mathcal{H}$, $P_N h$ closest element to h: $||h - w||^2 = ||P_N h + \delta h - w||^2 = ||P_N h - w||^2 + ||\delta h||^2 \geq ||\delta h||^2 = ||h - P_N h||^2$



When and how should we look for W_N?

- Kolmogorov N–width

$$d_N(\mathcal{H}) = \min_{W_N} \max_{\vec{\mu}} \min_{u \in W_N} ||u - h_{\vec{\mu}}||$$

- Measurements in L^∞ over the parameter domain, L^2 over frequency
- There exists a "best" N dimensional space W_N^{Kol}
- Challenges to finding W_N^{Kol}
  - Minimization over W_N is computationally challenging
  - Sampling $\mathcal{H}$ at an infinite number of parameters
- If solutions depend smoothly on the parameters

$$d_N(\mathcal{H}) \leq A e^{-bN}$$

A practical approach to finding W_N

- Sample $\mathcal{H}$ at a finite set of points
  - *Training space* of points Ξ, not necessarily tied to a pre-existing (metric based) selection
  - Now we seek to approximate $\mathcal{H}_\Xi = \{h(\vec{\mu}) \in \mathcal{H} : \vec{\mu} \in \Xi\}$ by W_N
  - Show convergence with finer sampling
  - If this were the only approximation, we expect convergence to W_N^{Kol}
- Build W_N by solving N easier problems
  - Sequence of spaces are constructed $W_1 \subset W_2 \subset \ldots \subset W_N$
  - (reminder) Ansatz: W_N is span of N waveforms chosen from $\mathcal{H}$
  - Greedy algorithm outputs a collection of points $\{\vec{\mu}_i\}_{i=1}^N$ and corresponding waveforms $\{h_i\}_{i=1}^N$ such that $W_N = \mathrm{span}\left(\{h_i\}_{i=1}^N\right)$

The greedy algorithm

- Definition: A greedy algorithm is any algorithm that solves the problem by making the locally optimal choice at each stage with the hope of finding the global optimum. (Wikipedia)
- Suppose we have W_i. The algorithm optimally chooses W_{i+1} and continues to W_N
- Expect that W_N is sub–optimal compared to W_N^{Kol}
- Nearly optimal, expected convergence [1]

$$\text{If} \quad d_N(\mathcal{H}) \leq Ae^{-bN} \quad \text{then} \quad \max_{\vec{\mu}} \min_{u \in W_N} ||u - h_{\vec{\mu}}|| \leq \tilde{A}e^{-\tilde{b}N}$$

[1] P. Binev et al., Convergence rates for greedy algorithms in reduced basis methods, in Preprint, 2010.

Greedy error

- The *greedy error*

$$\varepsilon_N \equiv \max_{\vec{\mu} \in \Xi} ||h_{\vec{\mu}} - P_N(h_{\vec{\mu}})||$$

- Recall that $P_N h$ realizes the minimum over $u \in W_N$
- $\varepsilon_N^2 = 1 - \max_{\vec{\mu} \in \Xi} < h_{\vec{\mu}}, P_N(h_{\vec{\mu}}) >_{MF}$
- Recall the minimal match

$$\min_{s \in \mathcal{H}} \max_{\vec{\mu} \in \Xi} < s, h_{\vec{\mu}} >_{MF} \quad = \quad MM$$

- In limit $\dim(\Xi) \to \infty$, $1 - MM \sim \varepsilon_N^2$

Greedy algorithm

- Setup: choose a parameter and waveform space (continuous and discrete)
- Initialize reduced basis with choice of $\vec{\mu}_1$ and thus $W_1 = \mathrm{span}(\{h_1\})$

While $\varepsilon_N \geq \mathrm{Tol}$

$i \rightarrow i + 1$

Greedy algorithm

- Setup: choose a parameter and waveform space (continuous and discrete)
- Initialize reduced basis with choice of $\vec{\mu}_1$ and thus $W_1 = \mathrm{span}\left(\{h_1\}\right)$

While $\varepsilon_N \geq \mathrm{Tol}$

$i \rightarrow i+1$

1. For all $\vec{\mu} \in \Xi$ compute $||h_{\vec{\mu}} - P_i(h_{\vec{\mu}})||$

Greedy algorithm

- Setup: choose a parameter and waveform space (continuous and discrete)
- Initialize reduced basis with choice of $\vec{\mu}_1$ and thus $W_1 = \mathrm{span}(\{h_1\})$

While $\varepsilon_N \geq \mathrm{Tol}$

$i \to i + 1$

1. For all $\vec{\mu} \in \Xi$ compute $||h_{\vec{\mu}} - P_i(h_{\vec{\mu}})||$
2. Find the parameter which maximizes the error of step 1

Greedy algorithm

- Setup: choose a parameter and waveform space (continuous and discrete)
- Initialize reduced basis with choice of $\vec{\mu}_1$ and thus $W_1 = \mathrm{span}\left(\{h_1\}\right)$

While $\varepsilon_N \geq \mathrm{Tol}$

$i \rightarrow i+1$

1. For all $\vec{\mu} \in \Xi$ compute $||h_{\vec{\mu}} - P_i(h_{\vec{\mu}})||$
2. Find the parameter which maximizes the error of step 1
3. $W_{i+1} = \mathrm{span}\left(\{h_1, ..., h_{i+1}\}\right)$

- W_N approximates $\mathcal{H}_\Xi$ with an error of better than Tol
- To show this for $\mathcal{H}$ we can increase the size of Ξ

The greedy algorithm: Implementation details

- Basis vectors are kept orthonormal by Gram–Schmidt
  - Otherwise some linear degeneracy, projection will become numerically unstable
- Hierarchical basis structure exploited: $P_{i+1} = P_i + (P_{i+1} - P_i)$
  - Extending the basis is independent of $\dim(W_i)$
- Loop over $||h_{\vec{\mu}} - P_i(h_{\vec{\mu}})||$ is embarrassingly parallel
  - Computational time weakly scales with finer sampling of $\mathcal{H}$
- Smart choice of training space Ξ
  - If chosen points cluster, should insert more training space points there

The greedy algorithm: Features

- Straightforward to implement
- Error measured in the L^∞ norm, ensuring a strict upper bound over the entire parameter space
- Globally selects points of interest
  - Not restricted to choosing point in a small neighborhood
  - *hp* adaptive greedy can be implemented
  - Points are good for interpolation
- Constructing W_N is $O(N)$
- Resulting reduced basis
  - Captures most dominant solution features
  - Application–specific spectral basis

Results: 2 parameter stationary phase approximation (SPA) waveforms at the 2nd post–Newtonian (PN) order

Analytic waveforms

- 2 dimensional parameter space described by the compact objects' masses, m_1 and m_2
- Ignore extrinsic parameters for now
- Stationary phase approximation of restricted 2nd PN inspiral waveforms

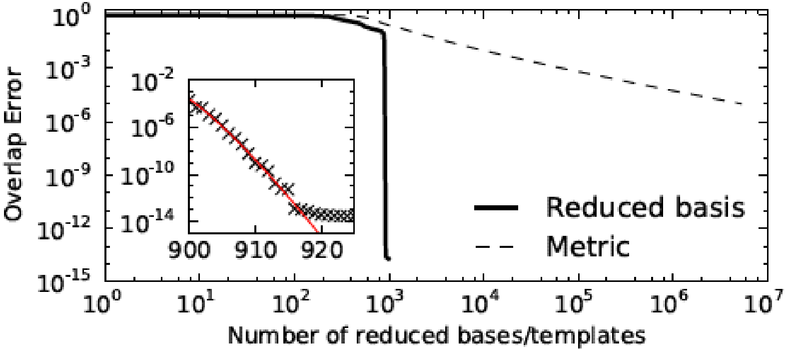
$$h(f) = \mathcal{A} f^{-7/6} \exp\left(-\mathrm{i}\pi/4 + \frac{3\mathrm{i}}{128\eta} \left(\frac{GM}{c^3} \pi f \right)^{-5/3} + \dots \right)$$

- $\mathcal{A}$ Depends on the distance and orientation of the source
- Total mass $M = m_1 + m_2$, and symmetric mass ratio $\eta = m_1 m_2 / M^2$

Analytic waveforms

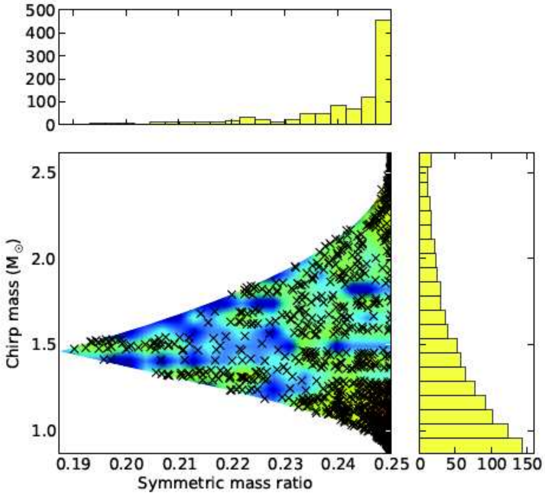
- Waveforms terminated at the greatest innermost stable circular orbit (ISCO) frequency in parameter space
  - With a parameter dependent ISCO cut, waveforms are nonsmooth function of parameters, spoiling exponential decay
  - Not a problem for inspiral, merger, and ringdown waveforms
  - Our method continues to work for ISCO cut waveforms (will address in Future Work)
- Waveforms are normalized w.r.t. the detectors noise curve

Results for [1-3]$M_\odot$ with Initial LIGO

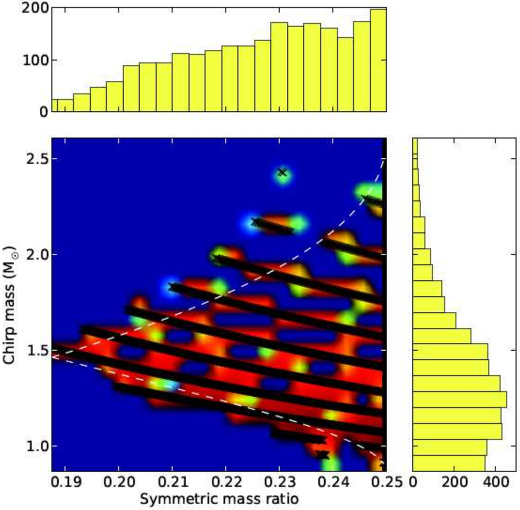


Plot of 1-MM (or ε_N^2). Same generic feature for all mass ranges considered.

Results for [1-3]$M_\odot$ with Initial LIGO (ChirpM $= \eta^{3/5} M$)



Metric Template with $MM = .97$



Results

Table: Accuracy measures 1-MM. TM = template metric (number of templates) and RB = reduced basis (dimension of space)

Detector	Overlap Error	BBH		BNS	
		RB	TM	RB	TM

Results

Table: Accuracy measures 1-MM. TM = template metric (number of templates) and RB = reduced basis (dimension of space)

Detector	Overlap Error	BBH		BNS	
		RB	TM	RB	TM
InitLIGO	10^{-2}	165	$2,450$	898	$10,028$
	10^{-5}	170	6.6×10^5	904	2.2×10^6
	2.5×10^{-13}	182	8.5×10^{11}	917	1.5×10^{12}

Results

Table: Accuracy measures 1-MM. TM = template metric (number of templates) and RB = reduced basis (dimension of space)

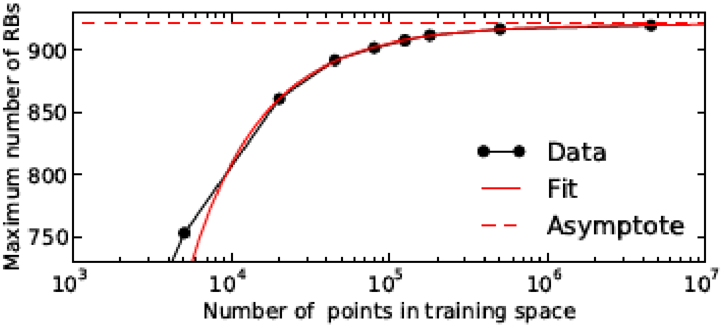
Detector	Overlap Error	BBH		BNS	
		RB	TM	RB	TM
InitLIGO	10^{-2}	165	$2,450$	898	$10,028$
	10^{-5}	170	6.6×10^5	904	2.2×10^6
	2.5×10^{-13}	182	8.5×10^{11}	917	1.5×10^{12}
AdvLIGO	10^{-2}	$1,058$	$19,336$	$5,395$	$72,790$
	10^{-5}	$1,687$	1.1×10^7	$8,958$	3.2×10^7
	2.5×10^{-13}	$1,700$	8.0×10^{13}	$8,976$	1.4×10^{14}

Results

Table: Accuracy measures 1-MM. TM = template metric (number of templates) and RB = reduced basis (dimension of space)

Detector	Overlap Error	BBH		BNS	
		RB	TM	RB	TM
InitLIGO	10^{-2}	165	$2,450$	898	$10,028$
	10^{-5}	170	6.6×10^5	904	2.2×10^6
	2.5×10^{-13}	182	8.5×10^{11}	917	1.5×10^{12}
AdvLIGO	10^{-2}	$1,058$	$19,336$	$5,395$	$72,790$
	10^{-5}	$1,687$	1.1×10^7	$8,958$	3.2×10^7
	2.5×10^{-13}	$1,700$	8.0×10^{13}	$8,976$	1.4×10^{14}
AdvVirgo	10^{-2}	$1,395$	$42,496$	$7,482$	$156,127$
	10^{-5}	$1,690$	3.2×10^7	$8,960$	2.6×10^7
	2.5×10^{-13}	$1,703$	1.4×10^{14}	$8,977$	2.9×10^{14}

Results for [1-3]$M_\odot$ with Initial LIGO



Confirms expectation that we are converging to a "unique" space (choice of initial pick could be important). $\dim(W_N) = 921$ for error of 10^{-8}.

Analytic results: PSD drifts

- The PSD will fluctuate in time
- Since the PSD weights the inner products, will a new RB need to be constructed? One can show

$$\delta_1 \equiv ||h_{\vec{\mu},1} - P_N(h_{\vec{\mu},1})||_1 \leq \varepsilon_N \frac{||h_{\vec{\mu}}||_0}{||h_{\vec{\mu}}||_1} \sqrt{\max_{f \in [f_L, f_U]} \left(\frac{S_0}{S_1} \right)}$$

- Where 1 and 0 subscripts refer to the old and new PSD
- RB space is robust against large perturbations in the PSD, no need to recompute
    - Order 1 change in accuracy

What have we gained?

- To compute the matched filter integral, integrate against the basis functions e_j
  - Recover this integral for any $h_i \in \mathcal{H}$ through $P_N h_i = \alpha_{ij} e_j$
  - Off–line computation of α_{ij} is done once and stored
  - For a given minimal match, significantly fewer integrations needed
  - Factor of 10 fewer integrations for $MM = .99$, a few additional RB resolves $\mathcal{H}$ to machine precision
- It appears the set of (inspiral) gravitational waves can be represented by a finite linear one with arbitrarily high accuracy

Future work and conclusion

Future work

Interpolation with selected points (Empirical Interpolation)

- Not yet implemented, but expect error on the order of ε_N

False alarms

- How do glitches project into the space? Will they be easier or harder to identify?

Solutions to PDE/ODE

- Can the method be extended when we must solve for the waveform at each point in parameter space? Must estimate the error

Future work

ISCO cut–off moved from b to c

- Parameter dependent chopping leads to a nonsmooth solution
- With more realistic waveforms this is not a problem
- One can smoothly transition between inspiral and 0
- Note that $||h_{\vec{\mu}} - P_i(h_{\vec{\mu}})||_{\Omega(a,b)}$ is positive definite, so we can bound

$$\frac{||h_{\vec{\mu}} - P_N(h_{\vec{\mu}})||_{\Omega(a,c)}}{||h_{\vec{\mu}}||_{\Omega(a,c)}} \leq \varepsilon_N X(c)$$

- Where c is new cut, $c \leq b$
- $X(c) = 1 + (\int_c^b f^{-7/3}/S_n)/(\int_a^c f^{-7/3}/S_n)$
- Assumes all error in range $[c, b]$, in practice we find 1 order of magnitude worst

Conclusion

- Motivation for the conjecture that the space of waveforms can be **accurately represented** by a compact set of basis function
- Proposed an efficient algorithm for finding this space
- Applied the algorithm to SPA waveforms at the 2 PN order and compared results to standard methods
  - Compact, accurate, and selects points of interest
- Many questions remain (Interpolation, false alarms, etc.)