# Reduced Basis Catalogs for Gravitational Wave Templates

Scott E. Field,[1] Chad R. Galley,[2,3] Frank Herrmann,[3] Jan S. Hesthaven,[4] Evan Ochsner,[5,6] and Manuel Tiglio[3]

[1]*Department of Physics, Brown University, Providence, Rhode Island 02912, USA*
[2]*Jet Propulsion Laboratory and TAPIR, California Institute of Technology, Pasadena, California 91125, USA*
[3]*Department of Physics, Maryland Center for Fundamental Physics, Joint Space Sciences Institute, Center for Scientific Computation
and Mathematical Modeling, University of Maryland, College Park, Maryland 20742, USA*
[4]*Division of Applied Mathematics, Brown University, Providence, Rhode Island 02912, USA*
[5]*Center for Gravitation and Cosmology, University of Wisconsin-Milwaukee, Milwaukee, Wisconsin 53201, USA*
[6]*Department of Physics, Maryland Center for Fundamental Physics, University of Maryland, College Park, Maryland 20742, USA*
(Received 19 January 2011; published 2 June 2011)

We introduce a reduced basis approach as a new paradigm for modeling, representing and searching for
gravitational waves. We construct waveform catalogs for nonspinning compact binary coalescences, and
we find that for accuracies of 99% and 99.999% the method generates a factor of about $10$–$10^5$ fewer
templates than standard placement methods. The continuum of gravitational waves can be represented by
a finite and comparatively compact basis. The method is robust under variations in the noise of detectors,
implying that only a single catalog needs to be generated.

*Introduction.*—The second generation of earth-based gravitational wave detectors, such as Advanced LIGO and Virgo, will become operational in 2014–2015. These detectors are expected to directly measure gravitational waves (GWs), with likely event rates of 0.4–400 per year for binary neutron stars (NS) and 0.4–1000 for binary black holes (BH) [1]. Direct detections would allow tests of general relativity in the nonlinear regime as well as access to portions of the universe otherwise unobservable.

Compact binary coalescences (CBCs), which consist of a pair of NSs and/or BHs inspiraling and merging, are considered to be one of the most promising sources of gravitational waves. The preferred method to search for GWs from CBCs is to employ matched filtering, which compares data from a detector to a bank of possible template waveforms and checks for a strong correlation between them. In low mass searches, the inspiral regime dominates the observable signal but as the mass increases the merger regime becomes increasingly relevant. The merger regime requires numerical simulations—even if used only for calibration of semianalytic models. Given the number and cost of these simulations, knowing an optimal choice of parameters is critical in order to limit the number of large simulations needed to accurately represent the variation over the parameter space. For this reason, it is desirable to seek a method that builds a template bank of waveforms by sequentially selecting only the most relevant points in the parameter space.

In addition, once a waveform catalog is constructed, there is a significant computational cost in performing an actual search for GWs due to the size of the catalogs. Real-time analysis of the data is critical to generate alerts to search for electromagnetic counterparts and enable multimessenger astronomy [2–4]. With the standard catalog placement method the number of templates needed grows rapidly with the dimension $P$ of the parameter space (as $(1 - MM)^{-P/2}$, with $MM$ the minimal match [5]) and such an approach could become impractical for searches of spinning binaries or other complex physical systems.

*Reduced basis method.*—The RB framework [6] constructs a global basis rather than using local methods and can be seen as an application-specific spectral expansion. In such an approach one seeks to enable a rapid online evaluation of the reduced model at the expense of having to build the basis prior to the application. While such an approach is not suitable for certain types of applications, it is particularly well suited for those requiring near real-time or many inquiry response as is the case in the present application. It has the following advantageous features over standard model reduction techniques such as proper orthogonal decompositions, singular value decompositions, or principal component analysis (see [7] for a general review of these methods and [8,9] for applications to GWs), see also [10]: (i) It is applicable to situations in which one must choose the most relevant parameters on the fly. (ii) It yields nested, hierarchically constructed catalogs which can be easily extended. If $C_N = \{h_1, \ldots, h_N\}$ is a catalog from the RB method then adding additional waveforms for higher accuracy implies that the resulting catalogs contain the previous ones, $C_N \subset C_{N+1} \subset C_{N+2} \cdots$. (iii) It is highly computationally efficient. The cost of adding a new member to an existing catalog of size $N$ is independent of $N$. Hence, the total cost of generating a catalog of size $N$ scales linearly with $N$, in contrast to many other approaches. (iv) It yields catalogs that are nearly optimal in terms of the error in approximating the whole spectrum of GWs by a compact set of basis

elements. This error is measured in the $L^\infty$ norm, ensuring a strict upper bound over the entire parameter space.

A gravitational wave is a function of time (or frequency in Fourier space) and of the $P$ parameters $\vec{\mu} = \{\mu_1, \ldots, \mu_P\}$ associated with the source. We denote each of them simply by $h_{\vec{\mu}}$ and do not explicitly write the time or frequency dependence. Let $\mathcal{H}$ be the space of all normalized GWs for the considered source(s). Although $\mathcal{H}$ is a not a linear space (the sum of two waveforms is not a waveform), we show that it can be represented by a linear one with arbitrarily high accuracy. We start with a theoretical description of our approach, followed by a description of an actual implementation.

We are interested in approximating $\mathcal{H}$ by the best linear combinations of members $\Psi_i \equiv h_{\vec{\mu}=\vec{\mu}_i}$ of a catalog $C_N = \{\Psi_i\}_{i=1}^N$. All such linear combinations form the *reduced basis space* $W_N = \text{span}(C_N)$. The waveforms that make up this catalog could be optimally chosen so that the error in representing $\mathcal{H}$ with $W_N$ is minimized over the choice of $N$ catalog members. Such an optimal error is given by the Kolmogorov $N$ width [11],

$$d_N(\mathcal{H}) = \min_{C_N} \max_{\vec{\mu}} \min_{u \in W_N} \|u - h_{\vec{\mu}}\|. \tag{1}$$

That is, one computes the error in the best approximation of $h_{\vec{\mu}}$ by a member of $W_N$, then finds the parameter $\vec{\mu}$ yielding the largest error, and lastly finds the smallest such error for all possible $N$-member catalogs. Here, the norm in Eq. (1) is calculated from the complex inner product $\langle \cdot, \cdot \rangle$, which is related to the standard overlap of Wiener filtering by $4\mathfrak{R}[\langle \cdot, \cdot \rangle]$, such that for two waveforms $F$ and $G$ in Fourier space,

$$\langle F, G \rangle \equiv \int_{f_L}^{f_U} \frac{F^*(f)G(f)}{S_n(f)} df, \tag{2}$$

where $S_n(f)$ is the power spectral density (PSD) of the detector.

Finding a catalog that exactly achieves the $N$ width is a computationally demanding optimization problem. Instead, we use a greedy approach, which is an inexpensive and practical procedure for hierarchically generating catalogs that *nearly* satisfy the $N$ width [12].

One constructs a catalog by first choosing a waveform for an arbitrary parameter value. A basis vector $e_1$ is then identified with this waveform, $e_1 = h_{\vec{\mu}_1}$, and the catalog is $C_1 = \{\Psi_1 = h_{\vec{\mu}_1}\}$. To add another waveform to the catalog, one seeks the parameter value $\vec{\mu}_2$ that maximizes $\|h_{\vec{\mu}} - P_1(h_{\vec{\mu}})\|$ where $P_1(h_{\vec{\mu}}) = e_1 \langle e_1, h_{\vec{\mu}} \rangle$ is the (orthogonal) projection of $h_{\vec{\mu}}$ onto $W_1$. We call this step a *greedy sweep*. The waveform corresponding to $\vec{\mu}_2$ is added to the catalog so that $C_2 = \{\Psi_1, \Psi_2\}$. The new basis vector $e_2$ is then constructed via Gram-Schmidt orthonormalization. Notice that $C_1 \subset C_2$, which demonstrates the hierarchical nature of the catalogs generated. Additional members of the reduced basis catalog are generated by

mathematical induction. At each step one picks up the parameter value $\vec{\mu}_j$ that maximizes $\|h(\vec{\mu}) - P_{j-1}(h(\vec{\mu}))\|$.

It can be shown [12] that if the decay of the $N$ width with $N$ can be bounded by an exponential,

$$d_N(\mathcal{H}) \leq Ae^{-cN^\alpha},$$

for some real $c$ and $\alpha$, then the decay of the maximum error for a catalog $C_N$ generated by this approach, which we call the *greedy error* $\varepsilon_N$, is also exponential,

$$\varepsilon_N \equiv \max_{\vec{\mu}} \|h_{\vec{\mu}} - P_N(h_{\vec{\mu}})\| \leq \tilde{A}e^{-dN^\beta}, \tag{3}$$

where $P_N(h_{\vec{\mu}}) = \sum_{i=1}^N \langle e_i, h_{\vec{\mu}} \rangle e_i$ and $d, \beta$ depend on $c, \alpha$ (see [12] for more details). Similar results hold in the case of power-law fall-off, i.e. $d_N(\mathcal{H}) \leq BN^{-s}$ implies $\varepsilon_N \leq \tilde{B}N^{-s}$. Note that $\varepsilon_N$ is a bound on the error between a waveform and its representation, and that $\varepsilon_N^2 = \max_{\vec{\mu}}(1 - \mathfrak{R}[\langle h_{\vec{\mu}}, P_N(h_{\vec{\mu}}) \rangle])$, so that $\varepsilon_N^2$ is an error on the overlap directly comparable to $(1 - MM)$. Given that GWs appear to depend smoothly on the parameters $\vec{\mu}$, we expect $d_N(\mathcal{H})$, and hence the greedy error $\varepsilon_N$, to decay rapidly (in fact exponentially) with $N$, which is a key feature of this method. Notice that (3) implies that any waveform can be represented as $h_{\vec{\mu}} = P_N(h_{\vec{\mu}}) + \delta h_{\vec{\mu}}(f)$ where $\|\delta h_{\vec{\mu}}(f)\| \leq \varepsilon_N$. Therefore, if $\varepsilon_N$ is of the order of numerical round-off then, in practice, the projection of $h_{\vec{\mu}}$ onto $W_N$ equals the waveform itself. In addition, the number of RBs needed to represent any $h_{\vec{\mu}}$ is comparatively small (see below).

In the case in which one is numerically solving equations on the fly while building the RB, the error (3) is replaced by an inexpensive error estimate evaluation (such as a residual), which is referred to as the *weak greedy approach*. Once the next parameter value is chosen one solves the full problem for it (referred to as the *offline stage*) and proceeds to the next greedy sweep. In any greedy approach, the maximum over $\vec{\mu}$ is searched for, in practice, using a *training space* of samples $\vec{\mu}$. However, since this is done as part of the offline process, the training space can be finely sampled and one can take advantage of the observation that evaluations for different parameters values are decoupled and, hence, embarrassingly parallel.

If one attempted a matched filter search with a RB catalog $C_N$ by filtering each basis function against the data and maximizing over arbitrary linear combinations of these filter outputs, one would of course get a very high false alarm rate. Instead, it is important to allow only linear combinations that correspond to physical waveforms. To be more specific, one could easily store the matrix of overlaps between waveforms in the original template bank (the training space) and the reduced basis, i.e., $\alpha_{ij} = \langle e_i, h_{\vec{\mu}_j} \rangle$. In fact, our algorithm provides this reconstruction matrix as output. A matched filtering computation may then be performed by integrating the incoming signal $s$ against each member of the basis $\langle s, e_i \rangle$ and using the reconstruction matrix $\alpha_{ij}$ to recover the matched filtering

integral of the signal with any template $h_{\tilde{\mu}_j}$ in the original bank. Explicitly, $\langle s, h_{\tilde{\mu}_j}\rangle = \sum_i \langle s, e_i\rangle \alpha_{ij}$. In this way, using the reduced bases is equivalent to using the original waveform space but with many fewer matched filtering integrals to compute for a given signal. Hence, using RB yields no increase in the false alarm rate.

*Catalogs for compact binary inspirals.*—We discuss our results for constructing reduced bases for "chirp" gravitational waveforms for binary inspirals without spins [13,14]. We use the 2nd order post-Newtonian accurate waveforms in the stationary phase approximation, which are known in closed form, so that the parameter space is two-dimensional (the binary's masses). For simplicity, we take the coalescence time and phase to be constant for each waveform.

Figure 1 shows results for the greedy error using a reduced basis model for inspirals of binary neutron stars (BNS) with mass components in the range $[1\text{–}3]M_\odot$ (for Initial LIGO with a lower frequency cutoff at 40 Hz) compared with the standard metric template placement method [5]. After a slowly decaying region, the reduced basis model gives very fast exponential convergence decay, which can be fitted by $\varepsilon_N^2 = a e^{-bN^p}$ with $a = 9.65 \times 10^{-4}$, $b = 0.598$, $p = 1.25$. The metric method yields approximately linear decay for a two-dimensional parameter space. As already mentioned, this decay becomes slower as the dimensionality $P$ of the parameter space increases. The fast decay of the reduced basis model allows a representation of the whole set of gravitational waves for these sources and mass ranges to within machine precision. We have found the

same feature in all mass ranges that we have explored. This leads to the rather remarkable finding that for all practical purposes the set of relevant gravitational waveforms in compact parameter regions appears to be *finite* dimensional. When increasing the number of samples $x$ in the training set we find the following fit for the number of RB for machine precision error, $N = a + bx^{-1/2} + cx^{-1}$ with $a = 921$, $b = -2090$, $c = -9.18 \times 10^5$ for the case of Fig. 1. In particular, in the limit $x \to \infty$ only 921 bases are needed to represent, within numerical accuracy, the full space of waveforms $\mathcal{H}$ for this range of masses for BNS inspirals.

Figure 2 shows the chosen parameter values in the chirp mass vs symmetric mass ratio plane and a density plot of the number of RBs. The histograms highlight that most values are picked for (nearly) equal mass systems of low chirp mass.

Table I shows the number of RB that we need to represent, for different overlap error tolerances, inspirals of BNS and stellar size binary black holes (BBH, with mass components in the range $[3\text{–}30]M_\odot$). The limit $x \to \infty$ is not taken here for simplicity so the RB values listed in Table I are slightly underestimated.

*Sensitivity to nonstationary noise.*—The PSD of any ground-based interferometer will fluctuate in time due to changes in environmental noises and other factors. Since the PSD weights the inner products used to construct the reduced basis, one might worry that a new RB needs to be constructed for any variation in the PSD.

Remarkably, we find indications that the RB constructed assuming a fiducial PSD is highly robust against rather large perturbations. From a histogram of the sensitivity of the LIGO interferometers during a portion of LIGO's fifth science reported in [15], we conclude that a 20% increase
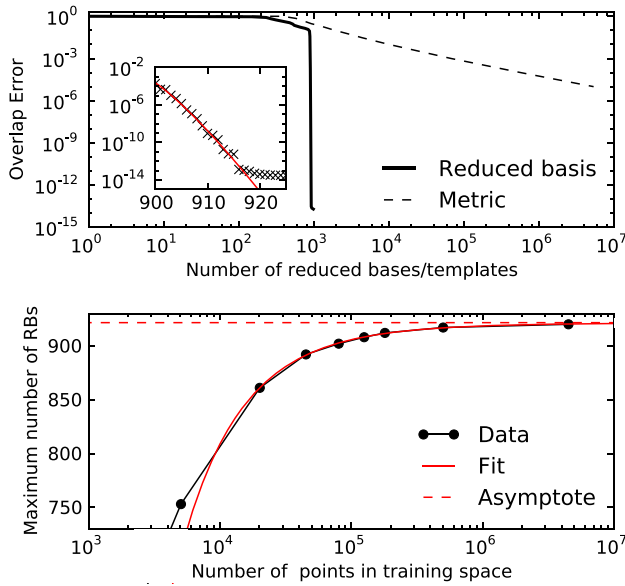


FIG. 1 (color online). Error in approximating the space of waveforms by a discrete catalog for BNS inspirals with Initial LIGO. For reduced basis, the error is the square of the greedy error (3) while for metric placement the error is $(1 - MM)$ with $MM$ the minimal match. The lower panel shows the extrapolation of the maximum number of RBs generated for an infinitely large training space. The fit shown (red) excludes the two points with largest $x$, which change the asymptotic value by 0.2.
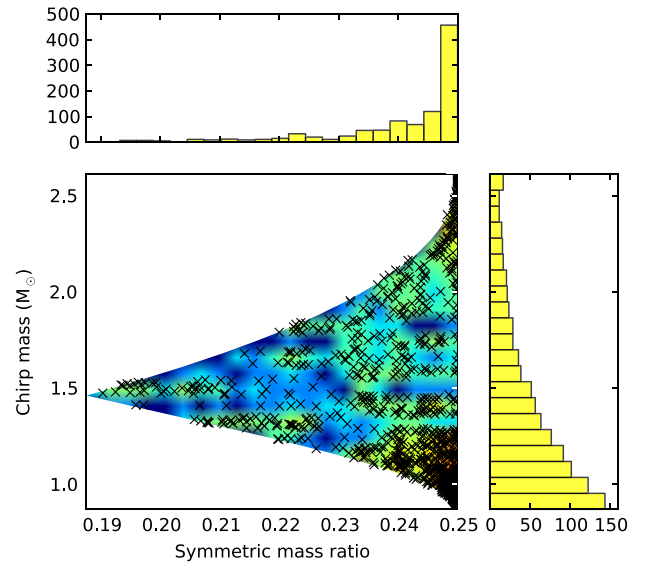


FIG. 2 (color online). The points show the parameter values chosen for the catalog of BNS and Initial LIGO. The density of parameter values is shown using a coloramp as well as histograms.

TABLE I.    Number of reduced bases and templates for different target accuracies with the
reduced basis (RB) and template metric (TM) approaches for binary neutron stars (BNS) and
binary black holes (BBH), using spinless chirp waveforms. We assume a lower frequency cutoff
of 40 Hz for Initial LIGO and 10 Hz for advanced LIGO and Virgo. The error is given by $\varepsilon_N^2$ for
RB and $(1 - MM)$ for TM.

| Detector | Overlap Error | BBH | | BNS | |
|---|---|---|---|---|---|
| | | RB | TM | RB | TM |
| InitLIGO | $10^{-2}$ | 165 | 2450 | 898 | 10, 028 |
| | $10^{-5}$ | 170 | $1.2 \times 10^6$ | 904 | $4.3 \times 10^6$ |
| | $2.5 \times 10^{-13}$ | 182 | $5.9 \times 10^{12}$ | 917 | $1.4 \times 10^{13}$ |
| AdvLIGO | $10^{-2}$ | 1, 058 | 19, 336 | 5, 395 | 72, 790 |
| | $10^{-5}$ | 1, 687 | $1.5 \times 10^7$ | 8, 958 | $4.9 \times 10^7$ |
| | $2.5 \times 10^{-13}$ | 1, 700 | $2.3 \times 10^{14}$ | 8, 976 | $5.6 \times 10^{14}$ |
| AdvVirgo | $10^{-2}$ | 1, 395 | 42, 496 | 7, 482 | 156, 127 |
| | $10^{-5}$ | 1, 690 | $3.1 \times 10^7$ | 8, 960 | $8.3 \times 10^7$ |
| | $2.5 \times 10^{-13}$ | 1, 703 | $4.8 \times 10^{14}$ | 8, 977 | $6.0 \times 10^{14}$ |

or decrease in the sensitivity to BNS signals is a large but
realistic fluctuation to the sensitivity of the Initial LIGO
interferometers. Therefore, we constructed smooth defor-
mations of the Initial LIGO design PSD meant to simulate
variations in the seismic, thermal and shot noise levels
which yield roughly a 20% increase or decrease in sensi-
tivity to BNS signals. We find that the RB generated for the
Initial LIGO design PSD with a greedy error tolerance of
$\varepsilon_N$ can represent any inspiral waveform within the per-
turbed PSDs considered here with an error of no more than
$1.3\varepsilon_N$. This result implies that one needs to compute only a
*single reduced basis* for a given source for a particular
detector. This is unlike the current operating procedure in
which a new template bank is generated every $\sim 2048$
seconds because of the drifts in the nonstationary noise.
We will provide further details in a forthcoming paper.

*Conclusions.*—We have considered the development and
use of a reduced basis method to template bank construc-
tion and found rapid exponential convergence of the wave-
form catalog over the full parameter space. The catalog is
computationally cheap to derive, hierarchical (i.e., if a
more accurate catalog is required, elements can be added),
can be extended for a computational cost that is indepen-
dent of $N$, and is robust under changes in a detector's noise.
We have found that the space of gravitational waveforms
considered in this paper is essentially finite dimensional for
any finite range of physical parameters, and conjecture that
it is in general the case. Elsewhere, we will present a more
detailed description of these results and further applica-
tions of the RB framework.

We thank P. Ajith, E. Barausse, A. Buonanno, C. Cutler,
A. Le Tiec, S. Lau, T. Littenberg, J. McFarland, R.
Nochetto, L. Price, M. Vallisneri, and B. Zheng for helpful
discussions and suggestions, and especially B. Stamm for
introducing several of us to the use of Reduced Basis. This
work has been supported by NSF Grants PHY0801213 and
PHY0908457 to UMD, NSF DMS 0554377 and OSD/

[1] J. Abadie *et al.* (LIGO Scientific Collaboration), Classical
Quantum Gravity **27**, 173001 (2010).
[2] J. Kanner *et al.*, Classical Quantum Gravity **25**, 184034
(2008).
[3] D. Buskulic (LIGO and Virgo Scientific Collaborations),
Classical Quantum Gravity **27**, 194013 (2010).
[4] P. S. Shawhan (LIGO Scientific Collaboration) and (Virgo
Collaboration), in *Bulletin of the American Astronomical
Society* (American Astronomical Society, Washington,
DC, 2010), Vol. 42, p. 230.
[5] B. J. Owen, Phys. Rev. D **53**, 6749 (1996).
[6] C. Prud'homme *et al.*, J. Fluids Eng. **124**, 70 (2002).
[7] R. Pinnau, in *Model Order Reduction: Theory, Research
Aspects and Applications*, edited by Wilhelmus H. A.
Schilders, Henk A. van der Vorst, and Joost Rommes,
Mathematics in Industry Vol. 13 (Springer, Berlin,
Heidelberg, 2008), p. 95.
[8] I. S. Heng, Classical Quantum Gravity **26**, 105005
(2009).
[9] K. Cannon *et al.*, Phys. Rev. D **82**, 044025 (2010).
[10] P. R. Brady and S. Ray-Majumder, Classical Quantum
Gravity **21**, S1839 (2004).
[11] A. Pinkus, *N-Widths in Approximation Theory* (Springer,
Amsterdam, 1985).
[12] "Convergence Rates for Greedy Algorithms in Reduced
Basis Methods," P. Binev *et al.*, SIAM J. Math. Anal. (to
be published).
[13] L. Blanchet, T. Damour, and B. R. Iyer, Phys. Rev. D **51**,
5360 (1995).
[14] C. M. Will and A. G. Wiseman, Phys. Rev. D **54**, 4813
(1996).
[15] B. Abbott *et al.* (LIGO Scientific Collaboration), Rep.
Prog. Phys. **72**, 076901 (2009).