

An automated parameter domain decomposition approach for gravitational wave surrogates using hp-greedy refinement

Franco Cerino and Manuel Tiglio
*Facultad de Matemática, Astronomía, Física y Computación,
Universidad Nacional de Córdoba, (5000) Córdoba, Argentina and
CONICET*

J. Andrés Díaz-Pace
*ISISTAN-CONICET Research Institute,
UNICEN University, (7000) Tandil, Buenos Aires, Argentina*

We introduce hp-greedy, a refinement approach for building gravitational wave surrogates as an extension of the standard reduced basis framework. Our proposal is data-driven, with a domain decomposition of the parameter space, local reduced basis, and a binary tree as the resulting structure, which are obtained in an automated way. When compared to the standard global reduced basis approach, the numerical simulations of our proposal show three salient features: i) representations of lower dimension with no loss of accuracy, ii) a significantly higher accuracy for a fixed maximum dimensionality of the basis, in some cases by orders of magnitude, and iii) results that depend on the reduced basis seed choice used by the refinement algorithm. We first illustrate the key parts of our approach with a toy model and then present a more realistic use case of gravitational waves emitted by the collision of two spinning, non-precessing black holes. We discuss performance aspects of hp-greedy, such as overfitting with respect to the depth of the tree structure, and other hyperparameter dependences. As two direct applications of the proposed hp-greedy refinement, we envision: i) a further acceleration of statistical inference, which might be complementary to focused reduced-order quadratures, and ii) the search of gravitational waves through clustering and nearest neighbors.

I. INTRODUCTION

Gravitational waves (GWs) are perturbations of space-time produced by massive accelerating objects, predicted by Einstein's equations for General Relativity (GR). Since 2015, when they were directly measured for the first time [1], they have become a new window to the universe, in addition to electromagnetic radiation.

When studying GWs, the ones produced by coalescing compact binary systems are widely considered. These are the strongest ones and hence the easiest to measure. Also, they provide information of the strong field regime of gravity. To model the waves emitted when the merger in a binary coalescence occurs, Numerical Relativity (NR) is needed, a task that requires expensive computational power. For example, a single simulation of a binary black hole system can cost $10^4 - 10^5$ CPU hours [2]. Furthermore, some studies are only feasible if the GWs can be computed fast enough, in real or quasi-real time, as in parameter estimation [3–5] and matched filtering [6–8]. Therefore, developments that enable model evaluation and analysis in a fast and accurate way are necessary.

The use of the reduced basis method [9–13] has been largely adopted in GW science [14–20], both for building surrogate models and for statistical inference, significantly reducing the associated computational costs while retaining high accuracy –see [21] for a review. Existing approaches in this application domain have normally used a global basis, similar in spirit to spectral methods. In fact, the reduced basis approach is sometimes referred to as a *domain-specific spectral expansion*. However, there are cases in which being able to partially localize the basis is useful and might have significant performance effects. In this paper we do so by proposing an hp-greedy refinement approach [22]. From a numerical relativity perspective, hp-greedy is similar to spectral elements [23] but partitioning the parameter domain instead of the physical one (space-time). As we discuss, this strategy has several advantages when compared to a global approach, most prominently: a higher accuracy basis for the same number of elements, faster surrogate evaluations, and faster statistical inference.

The organization of this paper is as follows. In Section II we describe in detail the hp-greedy reduced basis framework, its algorithms and supporting notation. In Section III we illustrate its application to a toy model with an intended strong discontinuity in the parameter space. In Section IV we apply the framework to the case study of the gravitational waves emitted by the collision of two spinning, non-precessing black holes in an initial quasi-circular orbit. We close in Section V with comments and possible future directions of research.

II. HP-GREEDY REDUCED BASIS

The hp-greedy approach leverages the standard reduced basis method, building a partitioning (h-refinement) of the parameter space and a reduced basis (p-refinement, here denoted as RB) for each partition.

A global reduced basis is initially built. If this basis is not as accurate and compact as wanted, the domain is partitioned and new local bases are built in each subdomain. The idea is to find reduced bases for spaces with lower complexity by means of a divide-and-conquer strategy.

The partitioning is adaptive and recursive. It depends on how the structure of the solutions vary in parameter space, allowing to focus the partitioning where it is needed, stopping when the local basis is sufficiently accurate or after a maximum number of partitions.

This section explains the h and p refinement procedures (subsections II A and II B) and how they work in synergy to obtain the hp-greedy approximation (subsection II C).

A. p-refinement

This is the standard reduced basis method (RBM). It is referred to as *p-refinement* borrowing language from spectral methods [23, 24], where the basis are polynomials (thus, the “p”) and the *refinement* part refers to the property that the representation error decreases as the degree of the polynomial increases. By analogy, then *p-refinement* in this context means that the error of the reduced basis representation decreases as the dimensionality n of the basis increases.

The RBM is traditionally targeted to computationally-intensive parametrized problems which require multiple queries. It is an alternative to repeatedly solving the full problem, which might not be feasible or realistic in practice, allowing for compact and accurate representations of the elements under study. In the case of GW modeling, the golden standard for solving the full problem is numerical solutions of the Einstein equations on supercomputers, which tends to be a remarkably difficult and expensive task. The RBM has allowed the construction of surrogate predictive models which are essentially indistinguishable from numerical relativity supercomputer simulations but can be evaluated in less than a second on a standard laptop.

The RBM starts with a solution space of functions $\mathcal{F} := \{h_\lambda := h_\lambda(t) := h(\lambda, t)\}$, where λ is a parameter, in general multi-dimensional, in a compact domain D . In our application, $h(t)$ is a complex time series: a gravitational wave. More specifically, λ can denote, for example, the masses and spins of each black hole in a binary collision.

A sampling of \mathcal{F} is used to form a training set $\mathcal{T} := \{h_i = h_{\lambda_i}, i = 1 \dots m\}$ for a number m of parameter values $\{\lambda_i\}_{i=1}^m$. The training set is used to build a compact reduced basis $\{e_i\}_{i=1}^n$, with $n \ll m$ in general, the compression ratio being n/m , which represents \mathcal{T} through its linear span of the form

$$h(t) \approx \sum_{i=1}^n c_i e_i(t),$$

and furthermore, through validation, the original space \mathcal{F} within a prescribed accuracy ϵ or a maximum dimension n_{max} . The choice of the coefficients $\{c_i\}$ should be such that the approximation is optimal, or quasi-optimal, in a precise mathematical sense.

1. Searching for an optimal reduced basis

A basis of dimension n , being optimal with respect to the maximum error in the parameter space, is characterized by the Kolmogorov n -width [25],

$$d_n := \min_{\{e_i\}_{i=1}^n} \max_{\lambda \in D} \min_{c_{i,\lambda} \in \mathbb{C}} \|h_\lambda(\cdot) - \sum_{i=1}^n c_{i,\lambda} e_i(\cdot)\|^2. \quad (1)$$

It measures the maximum representation error in parameter space, given by an optimal basis and the optimal coefficients $\{c_{i,\lambda}\}$, in the norm $\|\cdot\|$. In our case, the latter is given by

$$\|h_\lambda(\cdot)\|^2 := \int_{t_i}^{t_f} dt |h_\lambda(t)|^2, \quad (2)$$

inherited from the scalar product

$$\langle h_1, h_2 \rangle := \int_{t_i}^{t_f} \bar{h}_i(t) h_f(t) dt, \quad (3)$$

where the bar indicates complex conjugation.

If a basis is fixed in Equation (1), the first minimization problem, i.e., the minimum over the coefficients $\{c_{i,\lambda}\}$, turns to be a least squares one for those coefficients. There is a unique solution to this problem: it is the orthogonal projection \mathcal{P}_n with respect to the scalar product (3) to the span of the reduced basis [21]. The n-width then takes the form

$$d_n = \min_{\{e_i\}_{i=1}^n} \max_{\lambda \in D} \|h_\lambda(\cdot) - \mathcal{P}_n h_\lambda(\cdot)\|^2. \quad (4)$$

In some cases, the n-width can be calculated theoretically [26]. More generally, for functions with its first $(r - 1)$ derivatives continuous with respect to parameter variation, and for functions with C^∞ dependence, it can be proven that the n-widths are given by $d_n \sim n^{-r}$ and $d_n \sim e^{-an^b}$, respectively [30]. These can be thought of as theoretical bounds to the approximation error of a reduced basis. In the case of GWs, they do depend smoothly with respect to parameter variation, thus an optimal basis has *asymptotic* exponential convergence with n [20, 27]; this explains the existence of compact reduced bases of high accuracy (typically machine precision in GW science).

The task of finding an optimal reduced basis is of combinatorial complexity and unfeasible in practice: all combinations of basis elements must be evaluated in order to find one achieving the optimal n-width. Therefore, computationally cheaper approaches become attractive. An effective one is through a greedy algorithm, which is quasi-optimal in a precise mathematical sense, highly parallelizable, and of linear complexity; for more details see [21].

2. Greedy algorithm to build a quasi-optimal reduced basis

This approach consists of an iterative procedure, in which the basis is built from a training set \mathcal{T} , and at each iteration, a new basis function is added to the basis set so that the overall precision is improved. The procedure is said to be greedy in the usual optimization sense: at each iteration, the algorithm chooses the worst element represented from \mathcal{T} as a new basis function to be added. The training error of a basis of dimension n (we also refer to it as the *greedy error*) is defined as

$$\sigma_n := \max_{\lambda} \|h_\lambda - \mathcal{P}_n h_\lambda\|^2. \quad (5)$$

In our validation tests, we use an independent validation set of test functions and compute the equivalent of this error; we refer to it as the *maximum validation error*.

From a computational point of view, the cost of enriching a reduced basis with a new element is independent of the dimension of the basis already built, and linear with respect to the size of the training set. In terms of accuracy, the algorithm finds a nearly-optimal basis with respect to the Kolmogorov measure: if d_n decays as a power law, so does the greedy error σ_n , and if d_n decays exponentially with n , the same applies to σ_n [28–30].

A pseudocode of the greedy procedure is presented in Algorithm 1, whose main points are explained next:

- As input, a training set \mathcal{T} of size m is given, along with its associated parameters $\{\lambda_i\}_{i=1}^m$, the target maximum training representation error ϵ , and the maximum dimension n_{max} of the basis.
- First, a function of \mathcal{T} is chosen and defined as the first element or seed of the reduced basis (Step 1). Note that for a global basis, this choice is not relevant [31]; as we will see, this is very different in hp-greedy. Next, the basis is enriched iteratively with the function of \mathcal{T} that is worst represented by an orthogonal projection onto the span of the basis (the corresponding parameter is found in Step 4).
- From a practical viewpoint, the different solutions might be almost linearly dependent, resulting in a large conditioning number of the Gram matrix [32] used to calculate the projections. Therefore, it is convenient to orthonormalize the solutions to obtain the basis functions. Here, a Gram-Schmidt orthonormalization algorithm [33] is applied (Steps 5 and 6).
- The representation error σ_n is computed at each iteration (Step 8).
- The algorithm ends when $\sigma_n \leq \epsilon$ or $n = n_{max}$, with a reduced basis of dimensionality n , built with $h_\Lambda = \{h_{\Lambda_i}\}_{i=1}^n$, where $\Lambda = \{\Lambda_i\}_{i=1}^n$ are referred to as the *greedy parameters*.

- The outputs are the reduced basis RB , Λ , and σ .

Algorithm 1 GreedyRB($\lambda, h_\lambda, \epsilon, n_{max}$)

input: $\lambda, h_\lambda, \epsilon, n_{max}$

```

1:  $i = 1, \sigma = 1, \Lambda_1 = \lambda_1, RB = \{h_{\Lambda_1}/\|h_{\Lambda_1}\|\}$ 
2: while  $\sigma > \epsilon$  and  $i < n_{max}$  do
3:    $i = i + 1$ 
4:    $\Lambda_i = \operatorname{argmax}_\lambda \|h_\lambda - \mathcal{P}_{i-1} h_\lambda\|^2$  (selection of greedy parameter)
5:    $e_i = h_{\Lambda_i} - \mathcal{P}_{i-1} h_{\Lambda_i}$  (Gram-Schmidt)
6:    $e_i = e_i/\|e_i\|$  (normalization)
7:    $RB = RB \cup \{e_i\}$ 
8:    $\sigma = \max_\lambda \|h_\lambda - \mathcal{P}_i h_\lambda\|^2$  (representation error)
9: end while

```

output: $RB, \Lambda = \{\Lambda_i\}_{i=1}^n, \sigma$

B. h-refinement

The terminology *h-refinement* is borrowed from finite differences/elements, where the size of each cell on the mesh is often denoted by h . In the context of differential equations, *h-refinement* then refers to improving the accuracy of the quantity of interest by decreasing h , either by adding more points per domain or by decreasing the size of the latter. We will not elaborate much on this analogy here, it suffices to say that in this context we deal with a domain decomposition in parameter space, which is recursively partitioned and results in a binary tree structure.

We introduce some notation:

$$\begin{aligned}
 V &= \text{parameter space for a given subdomain} \\
 D &= \cup V, \text{entire parameter space} \\
 V_1, V_2 &= \text{partitions of } V \\
 \Lambda_V &= \text{greedy parameters for } V \\
 \hat{\Lambda}_V &= \Lambda_V[1], \text{anchor point for } V \\
 \hat{\Lambda}_{V_1}, \hat{\Lambda}_{V_2} &= \Lambda_V[1] \text{ or } \Lambda_V[2]
 \end{aligned}$$

Each (sub)domain $V \subseteq D$ has an anchor point, which we denote by $\hat{\Lambda}_V$. For the domain decomposition or partitioning of V , we assume that the anchor points $\hat{\Lambda}_{V_1}, \hat{\Lambda}_{V_2}$ are known and a sampling of V , λ_V , is given. A *proximity function* $d = d(\lambda_1, \lambda_2)$,

$$d(\lambda_1, \lambda_2) = \|\lambda_1 - \lambda_2\|_2,$$

is used to find the anchor point being closest to each parameter of λ_V . Then, two sets of parameters $\lambda_{V_1}, \lambda_{V_2}$ are created, each one with the parameters nearest to one of the two anchor points. In case a point is at the same distance from the two anchor points, it can be arbitrarily assigned to any of the sets, or both. Finally, λ_{V_1} and λ_{V_2} are returned, representing a sampling of the partitions V_1 and V_2 , respectively. A pseudocode for domain decomposition through a binary partitioning is described in Algorithm 2.

Algorithm 2 Partition($\lambda_V, \hat{\Lambda}_{V_1}, \hat{\Lambda}_{V_2}$)

input: $\lambda_V, \hat{\Lambda}_{V_1}, \hat{\Lambda}_{V_2}$

```

1:  $\lambda_{V_1} = \emptyset = \lambda_{V_2}$ 
2: for each  $\lambda_i \in \lambda_V$  do
3:   if  $d(\lambda_i, \hat{\Lambda}_{V_1}) < d(\lambda_i, \hat{\Lambda}_{V_2})$  then
4:      $\lambda_{V_1} = \lambda_{V_1} \cup \{\lambda_i\}$ 
5:   else if  $d(\lambda_i, \hat{\Lambda}_{V_1}) > d(\lambda_i, \hat{\Lambda}_{V_2})$  then
6:      $\lambda_{V_2} = \lambda_{V_2} \cup \{\lambda_i\}$ 
7:   else
8:      $\lambda_{V'} = \text{random choice}([\lambda_{V_1}, \lambda_{V_2}])$ 
9:      $\lambda_{V'} = \lambda_{V'} \cup \{\lambda_i\}$ 
10:  end if
11: end for

```

output: $\lambda_{V_1}, \lambda_{V_2}$

As the initial parameter domain D is decomposed, the recursive partitions are structured in a binary tree, in which each node corresponds to a subspace obtained with Algorithm 2. The maximum number of levels ℓ among all branches of the tree (i.e., its depth) is denoted by ℓ_{max} , $0 \leq \ell \leq \ell_{max}$, where $\ell = 0$ represents the case with no partitioning at all (i.e., the standard reduced basis approach). Each node at level ℓ is labeled by a series of indices B_ℓ as

$$B_\ell = (0, i_2, \dots, i_\ell), \text{ with } i_j = \{0, 1\},$$

where, by convention, $i = 0$ for the left leaf and $i = 1$ for the right one. For example, the root ($\ell = 0$) comprises the whole parameter domain and is labeled by

$$B_0 = (0,),$$

and its two children ($\ell = 1$) by

$$B_1 = (0, 0) \text{ or } (0, 1).$$

Figure 1 gives an example of the notation for a tree with $\ell_{max} = 2$, where all leaves reach the maximum allowed depth .

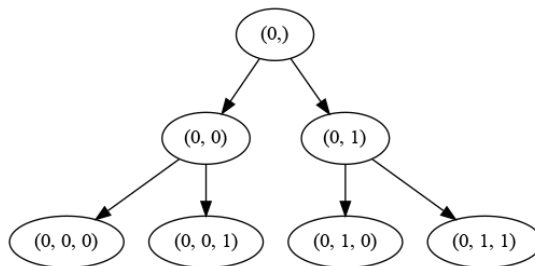


FIG. 1: Tree representing a partitioning of the parameter space. The index notation for the nodes of a tree of depth $\ell_{max} = 2$ is shown.

C. hp-greedy refinement

This approach combines *h-refinement* and *p-refinement* within the greedy reduced basis framework. In order to decide whether to partition a given parameter domain, a reduced basis is built, with an accuracy threshold ϵ and stopping criteria of n_{max} basis elements *per subdomain*, as in Algorithm 1. If the accuracy threshold is not reached, then the anchor point of the domain to be split is used as the anchor point of its left children, and the second greedy parameter of the reduced basis of the (parent) domain as the anchor point for the right children. Then, a binary

domain decomposition is performed, as described in Algorithm 2. Next, a reduced basis for each child is built. Note that the reduced basis from a parent node is *not* reused when building the bases for each of its children. The children domains where ϵ is not reached within n_{max} greedy iterations are further split into two more children according to Algorithm 2. This adaptive process continues until the accuracy threshold ϵ is achieved, the depth of the partition l reaches l_{max} , or the number of training set parameters is exhausted. If the latter happens, it means that the accuracy threshold cannot be reached with the prescribed n_{max} stopping criteria. The maximum allowed depth of the tree, l_{max} , is referred to as *early stopping* in machine learning.

A pseudocode for this hp-greedy refinement [22] approach is presented in Algorithm 3.

Algorithm 3 hpGreedy($\lambda_V, h_{\lambda_V}, \epsilon, n_{max}, \ell, \ell_{max}, B_\ell$)

input: $\lambda_V, h_{\lambda_V}, \epsilon, n_{max}, \ell, \ell_{max}, B_\ell$

```

1:  $RB, \Lambda_V, \sigma = \text{GreedyRB}(\lambda_V, h_{\lambda_V}, \epsilon, n_{max})$ 
2: if  $\sigma > \epsilon$  and  $\ell < \ell_{max}$  then
3:    $\hat{\Lambda}_1 = \Lambda_V[1]$ 
4:    $\hat{\Lambda}_2 = \Lambda_V[2]$ 
5:    $\lambda_{V_1}, \lambda_{V_2} = \text{Partition}(\lambda_V, \hat{\Lambda}_1, \hat{\Lambda}_2)$ 
6:    $out_1 = \text{hpGreedy}(\lambda_{V_1}, h_{\lambda_{V_1}}, \epsilon, n_{max}, \ell + 1, \ell_{max}, (B_\ell, 0))$ 
7:    $out_2 = \text{hpGreedy}(\lambda_{V_2}, h_{\lambda_{V_2}}, \epsilon, n_{max}, \ell + 1, \ell_{max}, (B_\ell, 1))$ 
8:    $out = out_1 \cup out_2$ 
9: else
10:   $out = \{(RB, \Lambda_V, h_{\Lambda_V}, B_\ell)\}$ 
11: end if

```

output: out

The first greedy parameter of the global basis (the seed of the algorithm), built for the first partition of the entire domain D , can *in principle* be chosen arbitrarily (we will see that it does have an impact on the accuracy of the resulting bases). Taking into account that the partitioning is carried out with the first two greedy parameters of the reduced basis (Step 5 in Algorithm 3), we notice that the seed of the algorithm determines the first partition, and thus, it also conditions the successive partitions and the reduced bases associated with those partitions. As we discuss in Section IV, the seed is a relevant hyperparameter of the algorithm.

There is no rigorous rule for choosing n_{max} , l_{max} and the seed. From a machine learning perspective, a possible approach to find them is through hyperparameter optimization, as they can be seen as hyperparameters of the algorithm in the sense of being parameters whose values are set before the learning process begins.

The rationale of hp-greedy is that if the greedy error is decaying slowly, the number of greedy iterations in the domain to be split gets too large, then the domain is partitioned (i.e., refined). The notion of a slowly decaying error is problem-dependent.

III. A TOY MODEL APPLICATION

We illustrate how hp-greedy reduced basis works for a toy model of functions that we artificially constructed with an intended strong discontinuity in the parameter space D . The hp-greedy procedure was run several times for a grid of hyperparameters (ℓ_{max}, n_{max}) , a fixed seed and a greedy tolerance $\epsilon = 10^{-10}$. For visualization purposes, we chose D to be two-dimensional:

$$D := [0, 1] \times [0, 1],$$

labeled by a tuple (α, β) . Our chosen parametrized functions are of the form $f_{\alpha, \beta}(x) : \mathbb{R} \rightarrow \mathbb{R}$ with $x \in [0, 10]$ and

$$f_{\alpha, \beta}(x) := \begin{cases} x^{1+\beta+\alpha}, & \beta < 1/2 \\ \sin(\beta\alpha + x), & 1/2 \leq \beta \end{cases} \quad (6)$$

After this choice, we normalized the functions so that

$$\|f_{\alpha,\beta}(\cdot)\|^2 := \int_0^{10} f_{\alpha,\beta}^2(x) dx = 1 \quad \forall (\alpha, \beta),$$

to place emphasis on their structure rather than on their size. Examples of $f_{\alpha,\beta}$ for different parameter values are shown in Figure 2.

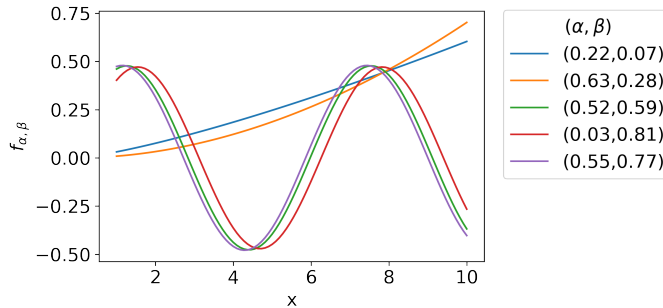


FIG. 2: A random sample of normalized functions $f_{\alpha,\beta}$ for our toy model. As it can be seen, the functions have a very different structure and there is a clear discontinuity in the parameter space.

We sampled a training set with 100 equispaced points per parameter dimension. That is, 100×100 numerical values of (α, β) were chosen and used by hp-greedy to subdivide each domain. For the validation set, we used 101 different (from those of the training set) equispaced points per parameter dimension.

To build hp-greedy models we used the open-source Python package Arby [34] to obtain a reduced basis for a given training set of waveforms (Algorithm 1). In addition, special-purpose code was written to deal with the partitioning of the domain (Algorithm 2) using hp-greedy (Algorithm 3).

A. Domain partitioning

By design, the algorithm is expected to automatically identify regions in the parameter space where functions have dissimilar structures, and then recursively divide them into subdomains. Each subdomain has its own reduced basis, with at most n_{max} elements, and is partitioned until a given threshold ϵ is reached or the depth of the tree ℓ is equal to ℓ_{max} .

To illustrate hp-greedy in a specific example, we analyzed the successive partitions obtained at each step of the algorithm, with $\ell_{max} = n_{max} = 5$ and seed $(\alpha, \beta) = (0, 0)$, as exemplified in Figure 3. A darker color of the partition indicates that the algorithm needed more steps to describe that subdomain; that is, the corresponding leaf in the tree is deeper (has a larger value of ℓ) than those of the partitions with a lighter color.

The upper left plot corresponds to the first iteration of hp-greedy, in which the domain is divided into two parts, using the first two greedy parameters of the global RB as anchor points. The next plot (on the right) shows the second domain decomposition, in which each subspace (from the first iteration) is partitioned using the first two greedy parameters of its associated RB. The plots below show the partitionings of the following iterations. In this case the algorithm had to do more partitions in the region around the discontinuity, because the change of the functions with parameter variation is larger than the places where there is no discontinuity and the variation of the functions is smooth.

Figure 4 shows the resulting tree structure for this toy model. The two deepest leaves, with $\ell = 5$, $(0, 1, 0, 1, 0, 0)$ and $(0, 1, 0, 1, 0, 1)$, contain the region of the discontinuity, which is harder to represent when compared to a domain with no discontinuity. These two nodes are associated with the subspaces in the center of the parameter space, which are black-colored in the fifth plot of Figure 3.

To assess the behavior of the partitioning in a more extreme case, we let the algorithm keep partitioning with a larger ℓ_{max} : we set $\ell_{max} = 8$ and $n_{max} = 4$. The result is shown in Figure 5. In this case, it can also be seen that the algorithm performed more partitions in the region of the discontinuity, almost “detecting” the discontinuity. Furthermore, the partitioning stopped earlier in those subspaces with no discontinuity.

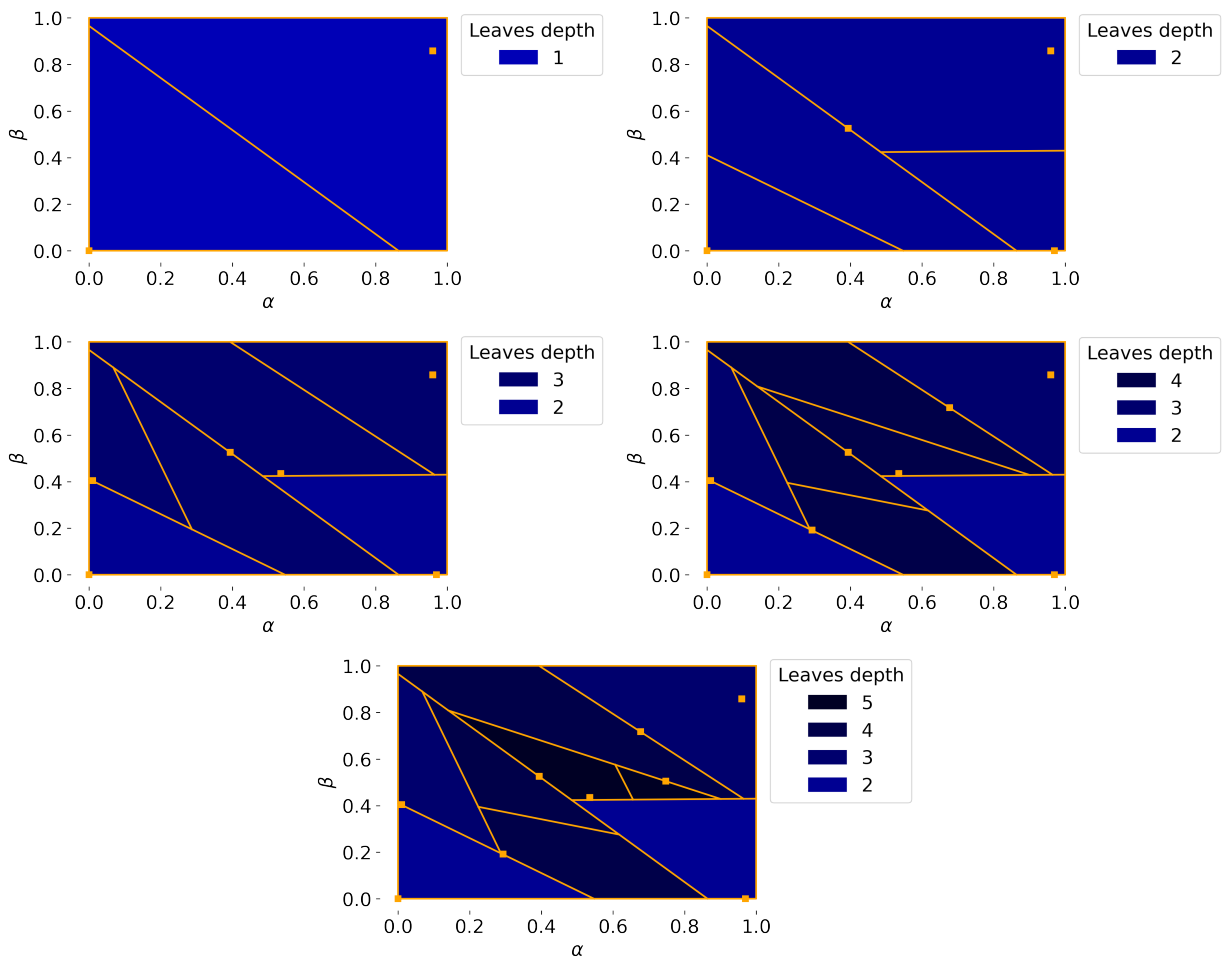


FIG. 3: Successive partitions performed by hp-greedy for the toy model, with $\ell_{max} = n_{max} = 5$. The *leaves depth* is the value of ℓ of that subdomain. The yellow squares denote the anchor points obtained by the hp-greedy algorithm to partition the parameter space.

B. Convergence

Figure 6 presents some results for different combinations of hyperparameters (n_{max}, ℓ_{max}) using validation data. The first aspect to emphasize here is the observed spectral convergence: for a fixed ℓ_{max} , the validation errors decay exponentially as a function of n_{max} , even with a discontinuity in the parameter space. It is unclear to us why this is so even for $\ell_{max} = 0$, (i.e. no partitioning) and without evidence of Gibb’s phenomenon [9, 24]; we speculate that this effect should appear at very high resolutions, below our chosen threshold $\epsilon = 10^{-10}$.

We now focus on the differences between a global basis, $\ell_{max} = 0$, and partitioning, $\ell_{max} > 0$, by comparing them for a given maximum dimensionality n_{max} for each basis¹. This comparison can be qualitatively seen by fixing n_{max} in the left panel of Figure 6. Except for cases of very low dimensionality and poor accuracy ($\epsilon < 10^{-2}$, $n_{max} < 3$), we notice that increasing ℓ_{max} significantly improves the maximum validation error. As an example, for $n_{max} = 5$, from a global basis $\ell_{max} = 0$ to, say, $\ell_{max} = 5$, there are around four orders of magnitude improvements in the error.

We finally focus on the value of n_{max} needed to achieve a given representation error for different values of ℓ_{max} . This analysis is appealing because a model with a lower n_{max} and the same or better accuracy can yield faster representations for the same precision. This happens because the dimensionality of the basis is a key point when evaluating a representation or surrogate, and it could be a way to accelerate statistical inference, as discussed in Section V. In our study, it can be seen that a model with partitioning can have a lower n_{max} than a model with

¹ We point out that the *total* number of basis elements is, in general, larger if there is a partition of the parameter space ($\ell_{max} > 0$) because there are more basis with the same constraint: a dimension less or equal than n_{max} .

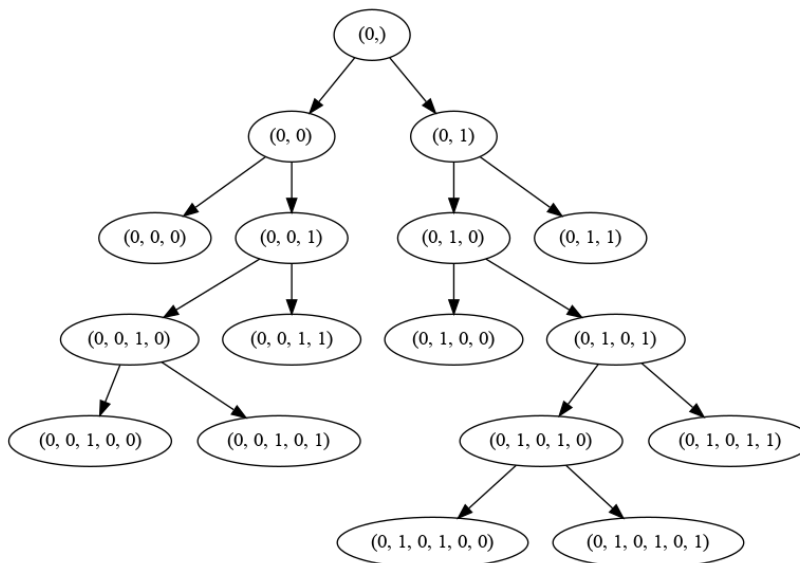


FIG. 4: Tree produced by the hp-greedy algorithm for the toy model with $\ell_{max} = n_{max} = 5$. For leaves with depth $\ell < \ell_{max}$, hp-greedy stopped when reaching the threshold training error $\epsilon = 10^{-10}$.

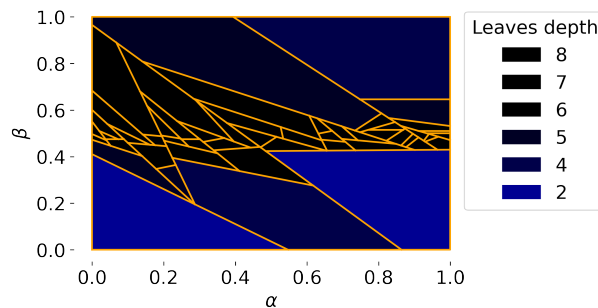


FIG. 5: Partitions performed by hp-greedy for the toy model, with $\ell_{max} = 8$ and $n_{max} = 4$. Notice that hp-greedy needed to refine notably more close to where the discontinuity is located, at $\beta = 0.5$.

no partitioning and the same or lower error. For example, for an error of 10^{-6} , models with $\ell_{max} = 2, 3, 4, 5$ have a comparable or lower error than the case without partitioning (see Figure 6).

C. Overfitting

Overfitting is a well-known behavior in machine learning, which entails that the learning process does not improve indefinitely when using more complex models; in fact, the errors might become worse [35]. This effect becomes evident when the training error decreases, but the opposite happens with the validation errors.

As the successive partitions of hp-greedy can be structured in a tree, overfitting in our approach can be related to the standard overfitting pattern of decision trees. If the maximum depth ℓ_{max} is large enough, the training data can be very well represented, but using validation data will likely show overfitting. In other words, there is a tradeoff between accuracy and tree depth when training while avoiding overfitting. Hyperparameter optimization can be approached in a number of ways to deal with this tradeoff, this aspect is left to future work.

In our numerical experiments for the toy model, we found that –as expected– for certain, but not all, values of n_{max} , larger values of ℓ_{max} lead to models with higher accuracy models for the training data; however, after a certain value, the maximum *validation* errors start increasing, up to orders of magnitude. In the left panel of Figure 7 we show how overfitting takes place for $n_{max} = 3$: it starts at $\ell_{max} = 8$ and the difference between maximum training and validation error gets larger than three orders of magnitude. Nonetheless, this pattern of overfitting did not appear in our experiments for all values of n_{max} . The right panel of Figure 7 shows the behavior of the validation error for

multiple values of (n_{max}, ℓ_{max}) . It can be seen that for $n_{max} > 4$ an increase in ℓ_{max} does not necessarily result in overfitting.

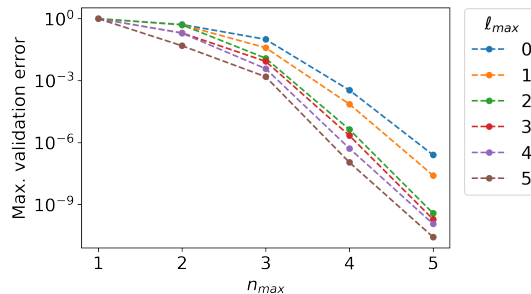


FIG. 6: Maximum validation errors for different values of n_{max} and ℓ_{max} for the toy model. The dashed blue curve shows the results using a global basis (no refinement), while the other curves are based on hp-refinement.

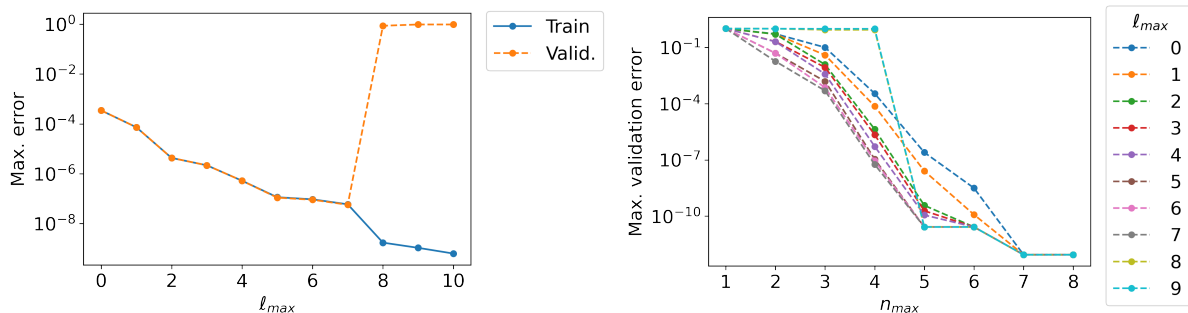


FIG. 7: **Left:** Training and validation errors for the toy model of Section III with $n_{max} = 3$. Overfitting is present, starting at $\ell_{max} = 8$. **Right:** Maximum validation errors for different values of n_{max} and ℓ_{max} for the toy model. Overfitting is present for $n_{max} \leq 4$ when increasing ℓ_{max} .

IV. GRAVITATIONAL WAVES FROM SPINNING BLACK HOLE COLLISIONS

We discuss the results of hp-greedy refinement using the physical setup of the GWs emitted by the collision of two spinning, non-precessing binary black holes initially in quasi-circular orbit. We use the hybrid numerical relativity (NR) and post-Newtonian (PN) surrogate NRHybSur3dq8 [36] as starting point.

Each waveform h is represented by a complex time series, $h = h_+ - ih_\times$, where h_+ and h_\times are the two polarizations of the gravitational wave. The time domain used in Ref. [36] is $[-5.4 \times 10^8, 135]M$, the long time interval explained by the use of PN approximations at early times, where $t = 0$ represents the peak of amplitude of the waveforms and M is the total mass of the binary system. A flat noise curve, $l \leq 4$ and $(5, 5)$ angular modes except for $(4, 1)$ or $(4, 0)$ were used in the construction of the surrogate.

In the late inspiral part of the waveforms, starting at $-3500M$ before the peak, the surrogate reproduces waveforms with mismatches $\lesssim 3 \times 10^{-4}$; where the latter are evaluated computing out-of-sample errors, randomly dividing the 104 training waveforms into groups of ~ 5 waveforms each and doing a cross-validation study. The errors are well within the truncation error of the NR simulations, which are computed calculating the mismatch between the two highest available resolutions of each NR waveform.

A. Datasets

In Ref. [36] the surrogate NRHybSur3dq8 was built in the parameter range of mass ratios $1 \leq q \leq 8$ and dimensionless spins $-0.8 \leq \chi_{1z}, \chi_{2z} \leq 0.8$. In this paper, we use those same ranges but only the dominant angular mode $\ell = m = 2$ for simplicity and the sake of illustrating hp-greedy. Furthermore, in order to speed up our numerical

experiments, we sampled NRHybSur3dq8 waveforms in the late inspiral part and merger regimes $t \in [-3000, 130]M$, with $\Delta t = 0.1M$. We also normalized the waveforms with respect to the L_2 norm (Equation 2), to put emphasis on structure/shape, instead of size/amplitude.

We studied three different cases, namely:

- 1D: No spin, the only free parameter is $q = m_1/m_2$, due to the scale invariance of GR.
- 2D: Two aligned spins with the same magnitude are added to the 1D case, meaning that $\chi_{1z} = \chi_{2z}$.
- 3D: The two spins are allowed to vary, but independently: in general, $\chi_{1z} \neq \chi_{2z}$.

Regarding training and validation sets, in the 1D case we generated two different sets of 500 waves to train and validate. For the 2D and 3D cases, we used 3,000 waves to train and 1,000 waves to validate.

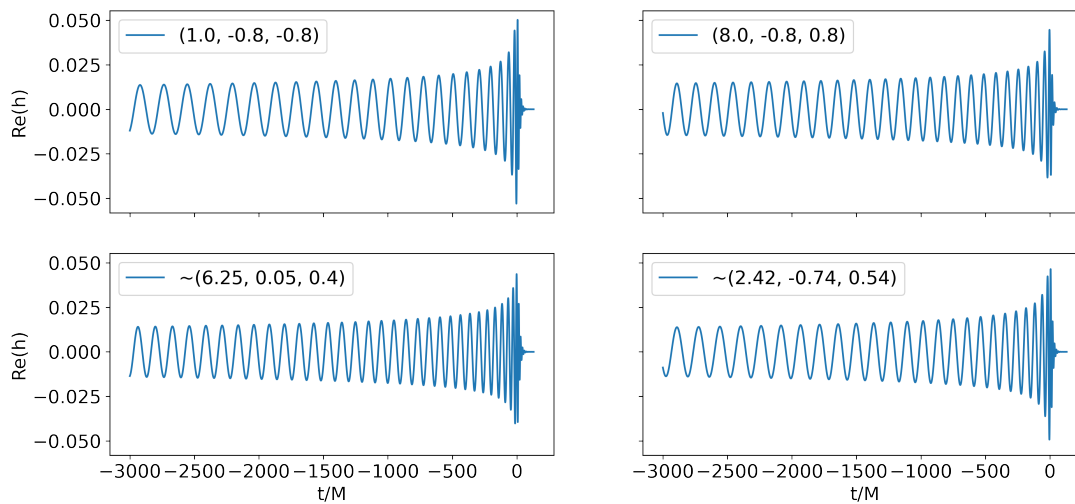


FIG. 8: Visualization of the real part of training waveforms. Each one has a different associated parameter value for the tuple $(q, \chi_{1z}, \chi_{2z})$.

B. hp-greedy refinement, dependence on seed

In our numerical experiments we set the greedy tolerance at double precision, $\epsilon = 10^{-16}$. As explained in Section II C, a seed is used as the anchor point for hp-greedy to initialize the whole algorithm and build local reduced bases. For a fixed seed, we built a number of hp-greedy bases with different values of (n_{max}, l_{max}) and manually chose, for each n_{max} , the l_{max} value leading to the highest accuracy multidomain representation. Since overfitting appears for large values of l_{max} , as discussed in Section III C, our 1D explorations were restricted to $0 \leq l_{max} \leq 9$, and $0 \leq l_{max} \leq 5$ in 2D and 3D.

With respect to the algorithm seed, it has been consistently found [31, 37] through numerical experiments that for global reduced bases, its choice is irrelevant because the greedy algorithm performs a global optimization –see for example Figure 1 of [31]. Interestingly, for hp-greedy and local bases, we found that the seed choice is highly relevant, and the accuracy of the resulting bases does depend on its choice. This situation is exemplified in Figure 9; for example, in the 2D case there are differences of up to three orders of magnitude in the error when varying the seed. In fact, it is possible that extensive seed searches might reveal larger differences. Therefore, for hp-greedy the seed should be taken as another hyperparameter of relevance, which is one of the main findings of this work.

In Figures 10 and 11 we show the domain decomposition obtained by hp-greedy for the parameter space in 1D and 2D, respectively, for different seed choices. It can be qualitatively seen that there is a significant impact on the partitioning of the domain.

We observed that almost all the trees of the resulting models are balanced and with depth $\ell = l_{max}$, unlike the toy model of Section III. This means that in our numerical experiments hp-greedy did not find specific regions where more refinement was needed. Examples of this behavior are shown in Figures 10 and 11.

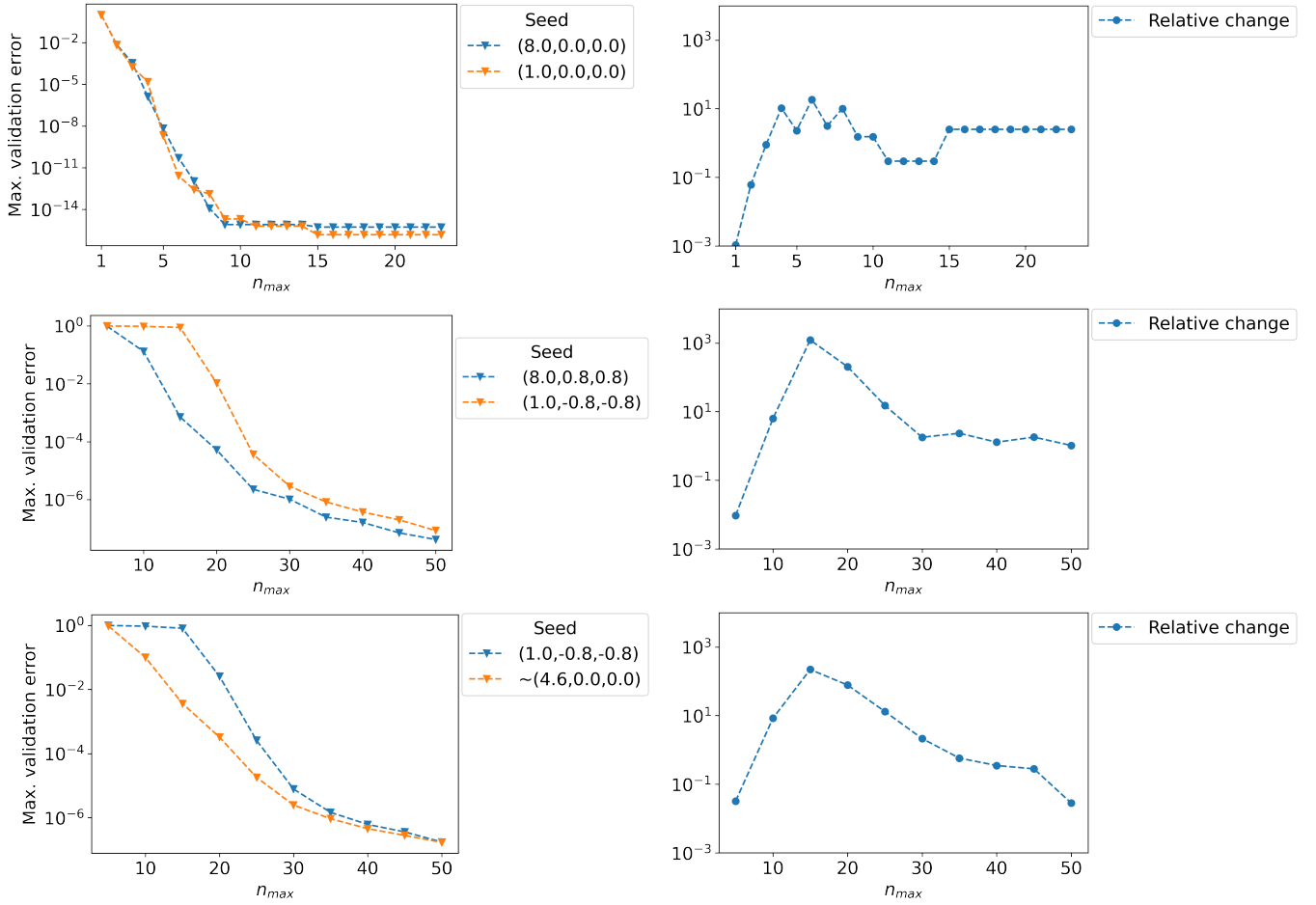


FIG. 9: From top to bottom, hp-greedy results are shown for 1D, 2D and 3D cases, using different seeds. Each curve in the left panels shows the highest accuracy model for each n_{max} , choosing the “optimal” value of ℓ_{max} . Unlike a global, standard reduced basis approach, in hp-greedy there is a strong dependence of the achieved errors on the seeds used. The right panels show the ratios between the errors of the two models of each left panel; we can observe up to three orders of magnitude of difference in the errors when varying the seed.



FIG. 10: Domain decompositions obtained with hp-greedy in the 1D gravitational wave case, using mass ratios $q = 1$ and $q = 8$ as seeds (left and right panel, respectively), $n_{max} = 5$ and $\ell_{max} = 4$. The dependence on the seed can be qualitatively noticed.

C. Convergence

We assessed the convergence of hp-greedy with respect to n_{max} , choosing for each value the highest accuracy model when varying the seed and ℓ_{max} , and comparing against a global basis (i.e., the standard approach). The results for the 1D, 2D, and 3D cases are shown in Figure 12. On one hand, there are always accuracy improvements at fixed n_{max} , in some cases by several orders of magnitude. On the other hand, for all cases, bases with lower dimension n_{max} are obtained for a fixed accuracy. As we discuss in Section V, smaller values of n_{max} are related to shorter evaluation times and hence faster statistical inference. For hp-greedy, we notice an exponential convergence from the

onset, $n_{max} \geq 1$. In contrast, for a global basis, this spectral convergence appears asymptotically, i.e., for sufficiently large values of n_{max} .

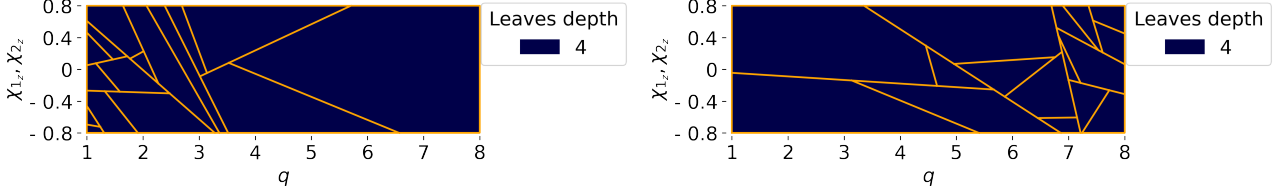


FIG. 11: Domain decompositions obtained with hp-greedy using different seeds for the 2D case. In order, from left to right, the figures have seeds located at $(1, -0.8, -0.8)$ and $(8, 0.8, 0.8)$. For pure visualization purposes and definiteness, hyperparameter values $l_{max} = 4$ and $n_{max} = 20$ were arbitrarily used. In these cases the leaves have the same depth, and also the highest possible. This means that lower depths were not enough for every subspace and the algorithm stopped in each leaf because of the value of l_{max} . The impact of the seed choice on the domain decomposition is apparent.

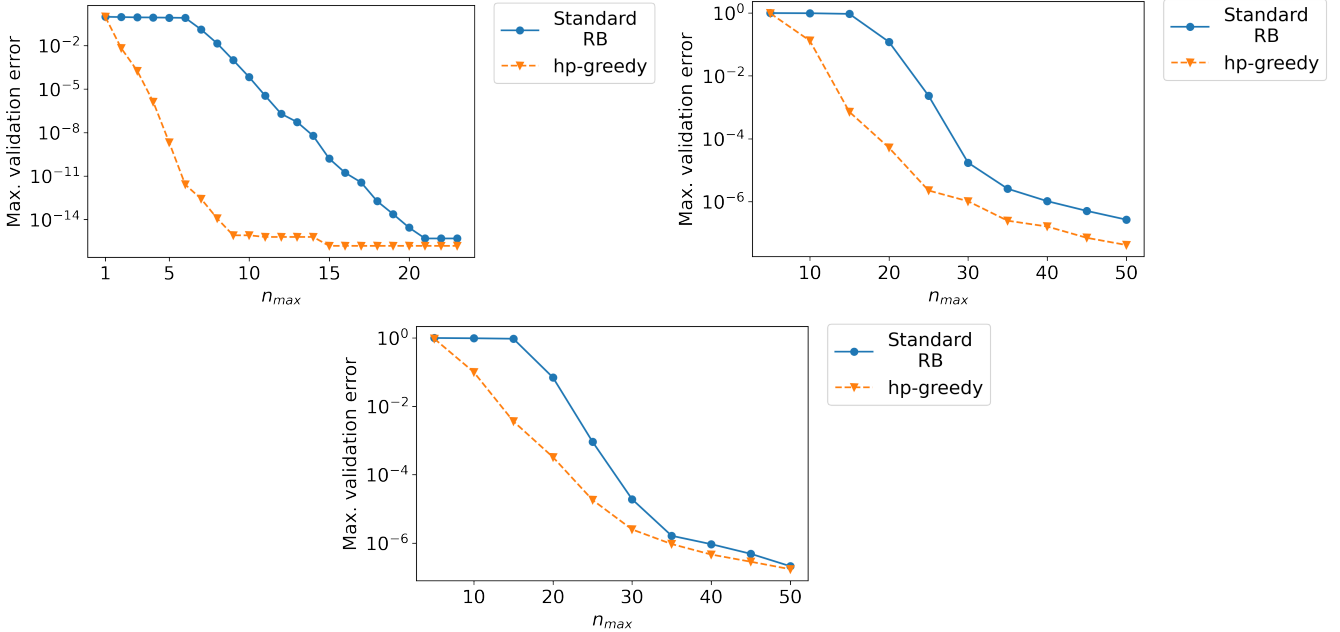


FIG. 12: Convergence of best performing hp-greedy models compared to the one of a global reduced basis (labeled as “Standard RB”) for the 1D, 2D and 3D cases, from left to right and top to bottom.

V. DISCUSSION AND FUTURE WORK

In this work, the usage of hp-greedy refinement in reduced basis within GW science is proposed, as a framework for partitioning the parameter space in an unstructured way and building a set of local reduced basis. This framework is a generalization of the standard reduced basis approach (which can be obtained as a special case by setting the hyperparameter $l_{max} = 0$), and aims at improving the accuracy and compactness of a reduced basis representation.

One main conclusion from our numerical experiments with hp-greedy is that the seed of the algorithm should be treated as an hyperparameter, in addition to n_{max} and l_{max} , since in general it does affect the accuracy of the resulting reduced bases. This behavior differs from the standard reduced basis framework, in which there is no partitioning of the parameter space, and the accuracy of the basis is insensitive to the seed choice. In addition, in the applications studied here, the seed choice does show a qualitatively noticeable impact on the structure of the domain partition.

It is known that a global reduced basis obtained with a greedy approach is a quasi-optimal approximation. This

means that it is difficult (or theoretically almost impossible) to reduce its dimensionality without accuracy loss: hp-greedy overcomes this challenge by partitioning the parameter space. From a simple two-dimensional model with a discontinuity in the parameter space, we concluded that the approach works with a reduction of the dimensionality needed for the bases to represent the function space for almost every precision used. Moreover, we have found that this improvement holds when applied to the more realistic case of two colliding spinning black holes, up to three dimensions, which is the highest dimensionality considered in this work. We have also found improvements in the accuracy for a fixed maximum dimensionality of the bases, n_{max} .

We envision several applications of our approach. One potential use case is to further accelerate statistical inference, for example in parameter estimation of the source of a detected gravitational wave, using reduced order quadratures (ROQ) [5, 38–41]. In more detail, parameter estimation serves to compute likelihoods, which involve integrals being often computationally expensive and requiring multiple on-demand sequential evaluations (e.g., via Markov chain Monte Carlo simulations). ROQ accelerates this process by constructing an application-specific quadrature rule using a reduced basis and the Empirical Interpolation Method [5], with a cost proportional to the dimensionality of the basis. Along this line, ROQ likelihood computations can be accelerated by using a multi-domain hp-greedy basis, since, for any given precision, the dimensionality of each local basis is expected to be smaller than that one of a global approach. This perspective would be similar in spirit to focused ROQ [39], where a reduced basis is constructed in a region close to the parameters found in the trigger part of the detection pipeline.

In any application of hp-greedy, the domain partition corresponding to any value of a parameter λ must be found in order to use the sub-space associated with a leaf that contains that value. Having a tree structure to divide the parameter space allows us to perform a fast search. This entails a binary search in which each node contains a pair of anchor points $\hat{\Lambda}_{V_1}$ and $\hat{\Lambda}_{V_2}$. Then, for a given parameter value λ , we can compute its distance to a pair of anchor points to descend one level in the tree, choosing the node that contains the closest anchor point with respect to that distance. The distance comparisons can begin with the anchor points of the root of the tree until reaching a leaf, being associated with a local reduced basis, that represents the sub-space where λ is located. For example, if the tree is balanced and there are n sub-spaces as leaves, the computational cost of the search becomes $\mathcal{O}(\log n)$, which improves the $\mathcal{O}(n)$ cost required if one looks at each subspace one by one.

Besides surrogate modeling, a second use case for hp-greedy is the search for gravitational waves using a nearest neighbors strategy, instead of the standard direct approach of comparing each candidate signal with a bank of templates one by one.

Another natural application can be one in which a physical discontinuity in the parameter space is present, and thus, a global reduced basis is likely to show slow convergence due to Gibb’s phenomena [24]. Some examples might include: i) two compact objects with a non-vanishing impact parameter, for which there can be fly-off, or collision [42], ii) the merger of two neutron stars or a mixed pair of a black hole and a neutron star, for which there can be a merger into a larger neutron star, or to a black hole [43].

In future work, we plan to perform a systematic hyperparameter optimization to devise rules for choosing appropriate seeds, n_{max} and ℓ_{max} , allowing us to build hp-greedy models with high accuracy in a faster way.

VI. ACKNOWLEDGMENTS

This work was partially supported by CONICET-Argentina. We thank Marcelo Rubio and Atuel Villegas for their feedback on a previous version of this manuscript.

-
- [1] Observation of gravitational waves from a binary black hole merger, *Phys. Rev. Lett.* **116**, 061102 (2016).
 - [2] L. Lehner and F. Pretorius, *Numerical relativity and astrophysics*, *Annual Review of Astronomy and Astrophysics* **52**, 661 (2014).
 - [3] J. Veitch, V. Raymond, B. Farr, W. Farr, P. Graff, S. Vitale, B. Aylott, K. Blackburn, N. Christensen, M. Coughlin, W. D. Pozzo, F. Feroz, J. Gair, C.-J. Haster, V. Kalogera, T. Littenberg, I. Mandel, R. O’Shaughnessy, M. Pitkin, C. Rodriguez, C. Röver, T. Sidery, R. Smith, M. V. D. Sluys, A. Vecchio, W. Vusden, and L. Wade, *Parameter estimation for compact binaries with ground-based gravitational-wave observations using the LALInference software library*, *Physical Review D* **91** (2015).
 - [4] N. Christensen and R. Meyer, *Parameter estimation with gravitational waves*, *Reviews of Modern Physics* **94** (2022).
 - [5] P. Canizares, S. E. Field, J. R. Gair, and M. Tiglio, *Gravitational wave parameter estimation with compressed likelihood evaluations*, *Phys. Rev.* **D87**, 124005 (2013).
 - [6] B. J. Owen and B. Sathyaprakash, *Matched filtering of gravitational waves from inspiraling compact binaries: Computational cost and template placement*, *Phys. Rev.* **D60**, 022002 (1999).

- [7] K. Cannon, R. Cariou, A. Chapman, M. Crispin-Ortuzar, N. Fotopoulos, M. Frei, C. Hanna, E. Kara, D. Keppel, L. Liao, S. Privitera, A. Searle, L. Singer, and A. Weinstein, Toward early-warning detection of gravitational waves from compact binary coalescence, *The Astrophysical Journal* **748**, 136 (2012).
- [8] S. Babak, R. Biswas, P. R. Brady, D. A. Brown, K. Cannon, C. D. Capano, J. H. Clayton, T. Cokelaer, J. D. E. Creighton, T. Dent, A. Dietz, S. Fairhurst, N. Fotopoulos, G. González, C. Hanna, I. W. Harry, G. Jones, D. Keppel, D. J. A. McKechnan, L. Pekowsky, S. Privitera, C. Robinson, A. C. Rodriguez, B. S. Sathyaprakash, A. S. Sengupta, M. Vallisneri, R. Vaulin, and A. J. Weinstein, Searching for gravitational waves from binary coalescence, *Physical Review D* **87** (2013).
- [9] J. S. Hesthaven, G. Rozza, and B. Stamm, *Certified Reduced Basis Methods for Parametrized Partial Differential Equations*, 1st ed., Springer Briefs in Mathematics (Springer, Switzerland, 2015) p. 135.
- [10] Y. Chen, J. S. Hesthaven, Y. Maday, and J. Rodríguez, Certified reduced basis methods and output bounds for the harmonic maxwell's equations, *SIAM J. Sci. Comput.* **32**, 970 (2010).
- [11] S. E. Field, C. R. Galley, F. Herrmann, J. S. Hesthaven, E. Ochsner, and M. Tiglio, Reduced basis catalogs for gravitational wave templates, *Phys. Rev. Lett.* **106**, 221102 (2011).
- [12] C. Prud'homme, D. V. Rovas, K. Veroy, L. Machiels, Y. Maday, A. T. Patera, and G. Turinici, Reliable real-time solution of parametrized partial differential equations: Reduced-basis output bound methods, *J. Fluids Eng.* **124**, 70 (2002).
- [13] A. Quarteroni, A. Manzoni, and F. Negri, *Reduced Basis Methods for Partial Differential Equations: An Introduction*, UNITEXT (Springer International Publishing, 2015).
- [14] V. Varma, S. E. Field, M. A. Scheel, J. Blackman, D. Gerosa, L. C. Stein, L. E. Kidder, and H. P. Pfeiffer, Surrogate models for precessing binary black hole simulations with unequal masses, *Phys. Rev. Research.* **1**, 033015 (2019).
- [15] GWSurrogate, GWSurrogate (2014–2020), accessed 31 May 2021.
- [16] S. E. Field, C. R. Galley, J. S. Hesthaven, J. Kaye, M. Tiglio, J. Blackman, B. Szilágyi, M. A. Scheel, D. A. Hemberger, P. Schmidt, R. Smith, C. D. Ott, M. Boyle, L. E. Kidder, H. P. Pfeiffer, and V. Varma, Binary black-hole surrogate waveform catalog (2019).
- [17] J. Blackman, S. E. Field, M. A. Scheel, C. R. Galley, D. A. Hemberger, P. Schmidt, and R. Smith, A surrogate model of gravitational waveforms from numerical relativity simulations of precessing binary black hole mergers, *Phys. Rev. D* **95**, 104023 (2017).
- [18] J. Blackman, S. E. Field, C. R. Galley, B. Szilágyi, M. A. Scheel, M. Tiglio, and D. A. Hemberger, Fast and accurate prediction of numerical relativity waveforms from binary black hole coalescences using surrogate models, *Phys. Rev. Lett.* **115**, 121102 (2015).
- [19] J. Blackman, S. E. Field, M. A. Scheel, C. R. Galley, C. D. Ott, M. Boyle, L. E. Kidder, H. P. Pfeiffer, and B. Szilágyi, Numerical relativity waveform surrogate model for generically precessing binary black hole mergers, *Phys. Rev. D* **96**, 024058 (2017).
- [20] S. E. Field, C. R. Galley, J. S. Hesthaven, J. Kaye, and M. Tiglio, Fast prediction and evaluation of gravitational waveforms using surrogate models, *Phys. Rev. X* **4**, 031006 (2014).
- [21] M. Tiglio and A. Villanueva, Reduced order and surrogate models for gravitational waves, *Living Reviews in Relativity* **25** (2022).
- [22] J. L. Eftang, A. T. Patera, and E. M. Ronquist, An hp certified reduced basis method for parametrized elliptic partial differential equations, *SIAM J. Sci. Comput.* **32**, 3170 (2010).
- [23] O. Sarbach and M. Tiglio, Continuum and discrete initial-boundary value problems and einstein's field equations, *Living Rev. Relativ.* **15**, 9 (2012).
- [24] J. Hesthaven, S. Gottlieb, and D. Gottlieb, *Spectral Methods for Time-Dependent Problems*, Cambridge Monographs on Applied and Computational Mathematics (Cambridge University Press, 2007).
- [25] A. Pinkus, *N-widths in approximation theory* (Springer, Amsterdam, 1985).
- [26] G. G. Magaril-Il'yaev, K. Y. Osipenko, and V. M. Tikhomirov, On exact values of n-widths in a hilbert space, *J. Approxim. Theory* **108**, 97 (2001).
- [27] S. E. Field, C. R. Galley, and E. Ochsner, Towards beating the curse of dimensionality for gravitational waves using reduced basis, *Phys. Rev. D* **86** (2012).
- [28] R. DeVore, G. Petrova, and P. Wojtaszczyk, Greedy algorithms for reduced bases in banach spaces, *Constructive Approximation* **37**, 455 (2013).
- [29] A. Buffa, Y. Maday, A. Patera, C. Prud'homme, and G. Turinici, A priori convergence of the greedy algorithm for the parametrized reduced basis method, *ESAIM: Mathematical Modelling and Numerical Analysis* **46** (2012).
- [30] P. Binev, A. Cohen, W. Dahmen, R. A. DeVore, G. Petrova, and P. Wojtaszczyk, Convergence rates for greedy algorithms in reduced basis methods, *SIAM J. Math. Analysis* **43**, 1457 (2011).
- [31] S. Caudill, S. E. Field, C. R. Galley, F. Herrmann, and M. Tiglio, Reduced Basis representations of multi-mode black hole ringdown gravitational waves, *Class. Quant. Grav.* **29**, 095016 (2012).
- [32] J. M. Taylor, The condition of gram matrices and related problems, *Proc. R. Soc. Edinburgh: Sect. A Math.* **80**, 45–56 (1978).
- [33] W. Hoffmann, Iterative algorithms for gram-schmidt orthogonalization, *Computing* **41**, 335 (1989).
- [34] A. Villanueva, M. Beroiz, J. Cabral, M. Chalela, and M. Dominguez, Arby - fast data-driven surrogates (2021).
- [35] G. James, D. Witten, T. Hastie, and R. Tibshirani, *An Introduction to Statistical Learning: with Applications in R* (Springer, 2013).
- [36] V. Varma, S. E. Field, M. A. Scheel, J. Blackman, L. E. Kidder, and H. P. Pfeiffer, Surrogate model of hybridized numerical relativity binary black hole waveforms, *Phys. Rev. D* **99**, 064045 (2019).

- [37] S. E. Field, C. R. Galley, and E. Ochsner, Towards beating the curse of dimensionality for gravitational waves using Reduced Basis, *Phys. Rev.* **D86**, 084046 (2012).
- [38] H. Antil, S. E. Field, F. Herrmann, R. H. Nochetto, and M. Tiglio, Two-step greedy algorithm for reduced order quadratures, *J. Sci. Comput.* **57**, 604 (2013).
- [39] S. Morisaki and V. Raymond, Rapid Parameter Estimation of Gravitational Waves from Binary Neutron Star Coalescence using Focused Reduced Order Quadrature, *Phys. Rev. D* **102**, 104020 (2020).
- [40] P. Canizares, S. E. Field, J. Gair, V. Raymond, R. Smith, and M. Tiglio, Accelerated gravitational wave parameter estimation with reduced order modeling, *Phys. Rev. Lett.* **114**, 071104 (2015).
- [41] R. Smith, S. E. Field, K. Blackburn, C.-J. Haster, M. Pürrer, V. Raymond, and P. Schmidt, Fast and accurate inference on gravitational waves from precessing compact binaries, *Phys. Rev. D* **94**, 044031 (2016).
- [42] F. Pretorius and D. Khurana, Black hole mergers and unstable circular orbits, *Classical and Quantum Gravity* **24**, S83 (2007).
- [43] M. Shibata, *Numerical Relativity*, 100 Years of General Relativity (World Scientific Publishing Company, 2015).