1. **Take-away message:**

* One of the most interesting ideas is the ability of interpreting/handling functions in a bounded interval as a vector in an infinite dimensional space (which later turns out to be the Hilber Space using -norms). This is possible because functions have the same properties as vectors belonging to a specific vector space. This also implies that we can compute the norm between two functions living in the same space. Functions living in the space might have a norm of zero despite not being a *null* function. To overcome this, we will no longer have a space of functions, but rather a space of *classes* of functions.
* Functions defined in an interval [a; b] (like those of linear/polynomial regression) can be interpreted as vectors in the R1nf space. Hence, thanks to the knowledge coming from the linear algebra, these functions can be approximated through the notion of orthonormal spaces, i.e. if we can construct a space of non-zero orthogonal vectors, we can find a basis of vectors that approximates the target function as much as desired.
* Going further, in practice point (2) results in handling dot products between vectors phi\_j and coefficients beta\_j ; vectors phi\_j come from a chosen basis, for example from the cosine basis or the Fourier basis that are straightforward to compute; the beta\_j can be computed through the generalized Fourier expansion, i.e. computing, or approximating, the dot product between the vectors of the basis and the target function; dot products are already easy to compute, but, thanks to the Parceval's theorem, we can further simplify them: the target function m(.) can be approximated directly from the beta\_j Fourier coefficients. We can control the precision of the approximation choosing the number J of coefficients. Powerful point: no need to recompute the previously computed coefficients as we change J. Even from a finite sample, we can estimate the overall trend of a function.
* This analysis can be easily extended to functions m(\_; \_; :::) of higher dimensions.

1. **Questions:**

We have two questions (we have also included what we think could be the solution to it):

1. How can we compute the n Fourier coefficients from n observed data-points?

In the theory document we have seen that we can approximate (which is a well specified function) to by computing the Fourier coefficients of the first j terms through an orthogonal function basis (one that ideally would minimize j). In this well specified scenario, it is “easy” to estimate these coefficients by taking the inner product between the well-defined function and each component of the basis up to J. How would this work in the case in which we observe the data generated by a function but do not know the shape of the function; would it suffice to convert the integral used to compute the inner product with a summation over the observed data?

where denotes the i-th observed data point.

1. How do we choose which orthonormal set of functions minimizes the number of GFC required?

Depending on the function we want to approximate, using different orthonormal sets of functions can lead to using more or less GFC’s. Is there any way we can find a “best” basis for this?