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.

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. Why can’t we consider all the functions of the form to live in a linear space?

We have the feeling this is related to the fact that this series goes to infinity and therefore it would not fulfil the null-vector property required in linear spaces.

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?

1. **Theory Question:**

We will try to explain this by first reviewing the logic used in the Transformed Feature Space, its main parallelism with the best approximation in the Hilbert Space (Function Sapces) and will finally highlight a few of the main properties that make this parallelism possible.

As we have seen in class, having a finite number of feature variables (for the sack of simplicity lets make this 1 and denote it by ), we can use these to predict another target variable (1 dimensional) through a linear combination as follows:

We might not be able to be able to capture all the information in through this simple transformation (when projecting on the 1-dimensional space of we might loose too much information). For this reason, we might want to think of a more complex set of transformation we can apply on our feature in order to capture more details about the target variable. This will be done through a function such that:

This is basically the same principle used in the series expansion in which we try to expand a function m(x) (possibly complex) by adding up less complex and therefore easier to handle functions together:

It is important to highlight that one of the basic reasons for which these two principles work (in the vector and function space) is because in both can exploit the dot product property between the elements living in their respective spaces.

Finally, we would also like to highlight that very nice properties pop up in both scenarios if the transformations () form an orthonormal set of functions/vectors:

* 1. Adding additional components to the sum would not change the previous values.
  2. can be estimated using the Generalized Fourier Expansion, which in the limit tends to an interpolation function of or .