Finding The Best Approximation Of A Function Using A Basis Set

Treating Functions As A Vector, How Can A Representation Be Constructed Using A Set Of Basis Functions?

Mathematics Internal Assessment

April 14, 2021

Contents

1 Research Question

Treating Functions As A Vector, How Can A Representation Be Constructed Using A Set Of Basis Functions?

2 Introduction

For me, the idea of finding sets of basis functions that can be combined to approximate others, first came to me when we were learning about the Taylor Series representation of differentiable functions in class.

Using the Taylor Series, any infinitely differentiable function can be written as the sum of a polynomial series, using n-th degree derivatives as as coefficients on the respective power of x.

Afterwards, when learning linear algebra, I realized that functions shared many of the same crucial properties with vectors. In this light, the Taylor Series expansion can be seen as expressing the original function in terms of a set of basis functions, just like how any three dimensional vector can be written as a linear combination of the $\hat{\imath}, \hat{\jmath}$ and $\hat{k}$ basis vectors.

My goal in this investigation therefore became to determine what conditions were allowed for a function to be approximated using a linear combination of some other set of basis functions, using properties of vectors to analyze this problem.

3 Vector Spaces

Before any properties of functions as vectors can be discussed, a formal definition of a vector space must first be defined. From Taylor and Mann (1983), a vector space is defined the following manner:

Definition 3.1 (Vector Space)**.** A Vector Space $\mathbb{V}$ is defined along with a scalar field $\mathbb{S}$ to be any set that has the following properties:

1. Scalar Multiplication Operator: $\boldsymbol{x} \in \mathbb{V}, c \in \mathbb{S} \rightarrow c\boldsymbol{x} \in \mathbb{V}$
2. Commutative Vector Addition Operator: $\boldsymbol{x}, \boldsymbol{y} \in \mathbb{V} \rightarrow \boldsymbol{x} + \boldsymbol{y} \in \mathbb{V}$
3. Scalar Multiplication is Distributive Across Vectors: $c(\boldsymbol{x} + \boldsymbol{y}) = c\boldsymbol{x} + c\boldsymbol{y}$
4. Scalar Multiplication is Distributive Across Scalars: $(c_1 + c_2)\boldsymbol{x} = c_1\boldsymbol{x} + c_2\boldsymbol{x}$
5. $\vec{0}$ Vector Element: $\boldsymbol{x} + \vec{0} = \boldsymbol{x}\ \forall \boldsymbol{x} \in \mathbb{V}$
6. 0 Scalar Element: $0\boldsymbol{x} = \vec{0}\ \forall \boldsymbol{x} \in \mathbb{V}$
7. Identity Scalar Element: $1\boldsymbol{x} = \boldsymbol{x}\ \forall \boldsymbol{x} \in \mathbb{V}$

From this, we can rigorously show that the set of functions on x do indeed satisfy the conditions of a vector space - more specifically the set of functions defined over some common domain.

Theorem 3.1 (Function Space)**.** *Let $\mathcal{D} \subset \mathbb{R}$ be any domain over the real numbers. Let $\mathscr{F}(\mathcal{D})$ be defined as the set of all functions $f(x) : \mathbb{R} \rightarrow \mathbb{R}$ well defined over $\mathcal{D}$. In conjunction with the scalar field $\mathbb{R}$, this set defines a vector space satisfying definition (3.1).*

Proof. Let $f(x)$ be any element in the set $\mathscr{F}(\mathcal{D})$. For any scalar c, $cf(x)$ is also a well defined function defined over $\mathcal{D}$ and is therefore also in $\mathscr{F}(\mathcal{D})$, satisfying condition 1. Additionally, the sum of two functions $f(x) + g(x)$ is well defined over $\mathcal{D}$, satisfying condition 2. Conditions 3 and 4 are defined as part of the algebra of functions. The $\vec{0}$ function element can be defined to be $\vec{0}(x) = 0$, so that $f(x) + \vec{0}(x) = f(x)$, satisfying condition 5. Finally, conditions 6 and 7 are satisfied with the conventional scalars of 0 and 1, both of which, when multiplied into any function, produce results in correspondance to the definition, showing that the set $\mathscr{F}(\mathcal{D})$ is indeed a valid vector space.

□

4 Defining An Inner Product

One of the most powerful things about vector spaces is they allow any member element to be written as a linear combination of a much smaller set of standard basis vectors.

However, generally to determine these coefficients, an inner product is needed. In a normal 2 or 3 dimensional vector space, the inner product, or the dot product, can give a measure of the degree to which two vectors are geometrically 'close', determined by the angle between the two vectors. The Inner Product doesn't need to be restricted to this geometric definition though, and can be generalized to allow for a measurement of 'closeness' encompassing vector spaces in a broader sense.

Definition 4.1 (Inner Product)**.** Given a vector space $\mathbb{V}$ with a corresponding scalar field $\mathbb{S}$, it is said to posses and inner product if there exists a function $\langle \boldsymbol{x}|\boldsymbol{y}\rangle : \mathbb{V}\times\mathbb{V} \to \mathbb{S}$ defined over the entire vector space that satisfies the following conditions:

1. Scalar Associative in Second Argument: $\langle \boldsymbol{x}|c\boldsymbol{y}\rangle = c\,\langle \boldsymbol{x}|\boldsymbol{y}\rangle \ \forall \boldsymbol{x}, \boldsymbol{y} \in \mathbb{V}, c \in \mathbb{S}$
2. Distributive In Second Argument: $\langle \boldsymbol{x}|\ \boldsymbol{y}+\boldsymbol{z}\rangle = \langle \boldsymbol{x}|\boldsymbol{y}\rangle + \langle \boldsymbol{x}|\boldsymbol{z}\rangle \ \forall \boldsymbol{x}, \boldsymbol{y}, \boldsymbol{z} \in \mathbb{V}$
3. Symmetry: $\langle \boldsymbol{x}|\boldsymbol{y}\rangle = \langle \boldsymbol{y}|\boldsymbol{x}\rangle \ \forall \boldsymbol{x}, \boldsymbol{y} \in \mathbb{V}$
4. Positive Semi-Definite: $\langle \boldsymbol{x}|\boldsymbol{x}\rangle > 0 \leftrightarrow \boldsymbol{x} \neq \vec{0}, \langle \boldsymbol{x}|\boldsymbol{x}\rangle = 0 \leftrightarrow \boldsymbol{x} = \vec{0}$

4.1 Finding A Valid Inner Product in $\mathscr{F}$

In general, an inner product can be any mapping of two vectors into a scalar that satisfies the conditions outline above. However, taking into account the value of the function over its entire domain and mapping that into a scalar is most commonly associated with taking the definite integral of the function. As such, it is especially useful to define the inner product of two functions as follows, noting that this definition satisfies all the properties previous outlined on the inner product.

Theorem 4.1 (Inner Product In Function Space)**.** *Let the domain* $\mathcal{D}$ *be equal to some bounded interval* $[a, b] \in \mathbb{R}$*. Let* $\mathscr{F}_M(\mathcal{D})$ *be the subset of* $\mathscr{F}(\mathcal{D})$ *containing every measurable function over the domain, or for which*

$$\int_a^b f(x)dx$$

is defined. Then define the function

$$\langle f(x)|g(x)\rangle = \int_a^b f(x)g(x)dx \tag{1}$$

This function is defined and constitutes a valid inner product for all $f(x), g(x) \in \mathscr{F}_M(\mathcal{D})$*.*

5 Orthogonal Functions

Now that an inner product has been established, we still need to outline some properties on the set of basis vectors we are using to attempt to reconstruct every other function. Now, theoretically any set of vectors can be chosen - however, when choosing basis vectors, two properties we find in 2 and 3 dimensional basis vectors are especially useful and can be generalized to our current case.

Firstly, the set of vectors $\hat{\imath}, \hat{\jmath}, \hat{k}$ are all *orthogonal* - meaning geometrically that any two different basis vectors are perpendicular.

Secondly, each of the basis vectors is *normalized*, meaning in the 3-dimensional case that they all have a length of 1.

To rigorously define these properties in the case of function space, the following definition for an Orthonormal Set is used:

Definition 5.1 (Orthonormal Set)**.** A set of vectors $U = \{\boldsymbol{u}_1, ..., \boldsymbol{u}_n\}$ is defined to be an orthonormal set iff

$$\langle \boldsymbol{u}_i | \boldsymbol{u}_j \rangle = \delta_{ij}, \ \forall \boldsymbol{u}_i, \boldsymbol{u}_j \in U \tag{2}$$

With the Kronecker Delta δ_{ij} being defined as

$$\delta_{ij} = \begin{cases} 0, & i \neq j \\ 1, & i = j \end{cases} \tag{3}$$

This provides a meaning to an orthogonal set of basis functions within function space. For some set of functions $U = \boldsymbol{u}_n(x)$ to be orthogonal, combining the inner product from equation (1) with definition (5.1) gives:

$$(\boldsymbol{u}_i(x), \boldsymbol{u}_j(x)) = \int_a^b \boldsymbol{u}_i(x)\boldsymbol{u}_j(x)dx = \delta_{ij} \tag{4}$$

5.1 Defining The Closest Representation in an Orthonormal Basis

Given a set of orthonormal basis vectors $\boldsymbol{u}_n(x)$, my next goal was figuring out the process by which to determine coefficients c_n such that the linear combination $\sum_{k=1}^{n} c_k \boldsymbol{u}_k(x)$ either equaled or best approximated the original arbitrary function $f(x)$. Speaking in terms of a traditional 3 dimensional vector space with a vector $\vec{v}$, this amounts to determining its x, y and z coordinates, allowing the original vector to be written as a linear combination of the basis vectors $\hat{i}, \hat{j}, \hat{k}$.

As established earlier, establishing a vector representation in terms of a set of basis vectors is tremendously useful, because many important operations possessing the property of linearity can first be computed individually on the set of basis vectors, then rescaled by the given coefficients to determine the effect of applying the function on the original vector itself.

However, unlike vector spaces of finite dimension, the uncountable amount of ways in which any two functions can differ makes it quite likely that no set of orthonormal vectors $\{\boldsymbol{u}_n\}$ can span our entire vector space. Additionally, even in the case that there does exist a valid set of basis vectors, it is reasonable to believe that such a set would have an infinite number of elements. Because of this, it is immensely useful to construct a measure of how *close* a given linear approximation is to the original function - giving a formal meaning to our intuitive sense of 'bad', 'good' or 'best' approximation.

Let $\{\boldsymbol{u}_n\}$ be an orthonormal set, and the $f(x)$ be any arbitrary function in our vector space. For a given set of scalars c_n, define the function

$$F(x) = c_1\boldsymbol{u}_1(x) + c_2\boldsymbol{u}_2(x)... + c_n\boldsymbol{u}_n(x) \tag{5}$$

to be the representation of $f(x)$ in the basis set. The difference between the representation and original function is another function of x given by $\delta(x) = f(x) - F(x)$. Because $\delta(x)$ is defined as the linear combination of $f(x)$, an element of $\mathscr{F}_M(\mathcal{D})$, and $\{\boldsymbol{u}_n\}$, all of which are elements of $\mathscr{F}_M(\mathcal{D})$, $\delta(x)$ must then also be a member of $\mathscr{F}_M(\mathcal{D})$ - which in turn means that it has a valid inner product.

Again using intuition from geometry, classically the 'distance' between two vectors is the norm of the difference between those vectors - written in terms of the inner product as $\|\vec{v}-\vec{u}\|^2 = (\vec{v}-\vec{u}, \vec{v}-\vec{u})$. This can be applied to say that the measure of how 'close' an approximation function is from the original is given by

$$\|\delta(x)\|^2 = \langle f(x) - F(x) \mid f(x) - F(x) \rangle \tag{6}$$

Therefore, given the orthonormal basis $\{\boldsymbol{u}_n\}$, the best approximation is defined to be the set of coefficients $\{c_n\}$ that minimize 'distance' from equation (6).

5.2 Determining The Linear Representation of a Given Function

Finally, to answer my research question, the last part of this essay will be dedicated to determining how these coefficients relate to the given set of orthonormal basis functions, and attempting to find (if it exists) a basis such that the best linear approximation can reproduce exactly any arbitrary function in the entire space.

Firstly, for a given $f(x)$ and basis $\{\boldsymbol{u}_n\}$, define the set $\{c_n\}$ and corresponding function $F_c(x) = \sum_{k=1}^{n} c_k \boldsymbol{u}_k$ such that, for any other linear combination $\{d_n\}$,

$$\langle f - F_c \mid f - F_c \rangle \leq \langle f - F_d \mid f - F_d \rangle \tag{7}$$

with the two sides being equal if and only if the set $\{d_n\}$ is exactly equal to $\{c_n\}$. (Note: the (x)'s have been dropped from here forwards for convenience, but are implied whenever not explicitly written). Following from equation (7) using the properties of linearity,

$$\begin{aligned}
\langle f - F_c \mid f - F_c \rangle &\leq \langle f - F_d \mid f - F_d \rangle \\
\langle f|f \rangle - 2 \langle f|F_c \rangle + \langle F_c|F_c \rangle &\leq \langle f|f \rangle - 2 \langle f|F_d \rangle + \langle F_d|F_d \rangle \\
-2 \langle f|F_c \rangle + \langle F_c|F_c \rangle &\leq -2 \langle f|F_d \rangle + \langle F_d|F_d \rangle \qquad (8)
\end{aligned}$$

Expanding $\langle F_c|F_c \rangle$ using $\boldsymbol{u}_i(x)\boldsymbol{u}_j(x) = 0 \leftrightarrow i \neq j$ and $\langle \boldsymbol{u}_i(x)|\boldsymbol{u}_i(x) \rangle = 1$, we get:

$$\begin{aligned}
\langle F_c|F_c \rangle &= \int F_c(x)F_c(x)dx \\
&= \int (\sum_k^n c_k \boldsymbol{u}_k(x) \sum_k^n c_k \boldsymbol{u}_k(x))dx \\
&= \int \sum_k^n (c_k \boldsymbol{u}_k(x))^2 dx \\
&= \sum_k^n {c_k}^2 \int (\boldsymbol{u}_k(x))^2 dx \\
&= \sum_k^n {c_k}^2 \langle \boldsymbol{u}_k|\boldsymbol{u}_k \rangle \\
&= \sum_k^n {c_k}^2 \qquad (9)
\end{aligned}$$

Similarly expanding $\langle f|F_c \rangle$, we get:

$$\begin{aligned}
\langle f|F_c \rangle &= \int f(x)F_c(x)dx \\
&= \int \sum_k^n c_k (f(x)\boldsymbol{u}_k(x))dx \\
&= \sum_k^n c_k \langle f(x)|\boldsymbol{u}_k \rangle \qquad (10)
\end{aligned}$$

And plugging equations (9) and (10) into (8) to get:

$$\begin{aligned}
2\langle f|F_c\rangle - \langle F_c|F_c\rangle &\geq 2\langle f|F_d\rangle - \langle F_d|F_d\rangle \\
2\sum_k^n c_k \langle f(x)|\boldsymbol{u}_k\rangle - \sum_k^n {c_k}^2 &\geq 2\sum_k^n d_k \langle f(x)|\boldsymbol{u}_k\rangle - \sum_k^n {d_k}^2 \\
\sum_k^n (2c_k \langle f(x)|\boldsymbol{u}_k\rangle - {c_k}^2) &\geq \sum_k^n (2d_k \langle f(x)|\boldsymbol{u}_k\rangle - {d_k}^2) \qquad (11)
\end{aligned}$$

Because equation (11) is true for any set $\{d_k\}$, even sets that only differ in one single coefficient, equation (11) must be true term for term - that is to say:

$$\begin{aligned}
(2c_k \langle f(x)|\boldsymbol{u}_k\rangle - {c_k}^2) &\geq (2d_k \langle f(x)|\boldsymbol{u}_k\rangle - {d_k}^2)\ \forall k \in n, \qquad (12) \\
(2c_k \langle f(x)|\boldsymbol{u}_k\rangle - {c_k}^2) - \langle f(x)|\boldsymbol{u}_k\rangle^2 &\geq (2d_k \langle f(x)|\boldsymbol{u}_k\rangle - {d_k}^2) - \langle f(x)|\boldsymbol{u}_k\rangle^2 \\
-(c_k - \langle f(x)|\boldsymbol{u}_k\rangle)^2 &\geq -(d_k - \langle f(x)|\boldsymbol{u}_k\rangle)^2 \\
(c_k - \langle f(x)|\boldsymbol{u}_k\rangle)^2 &\leq (d_k - \langle f(x)|\boldsymbol{u}_k\rangle)^2\ \forall d_k \in \mathbb{R} \\
\therefore\ c_k &= \langle f(x)|\boldsymbol{u}_k\rangle,\ \forall k \in n \qquad (13)
\end{aligned}$$

Therefore, from equation (13), the representation of a function in an orthonormal basis is obtained by taking the inner product between that function and each basis function.

References

Taylor, A. E. and Mann, W. R. (1983). *Advanced Calculus.* John Wiley and Sons, Inc., 3rd edition.