Using Vectors to Approximate Functions

An Exploration of Linear Algebra's Applications Outside of Geometry

Benjamin Jones — April 13, 2026

When most people think of linear algebra, they picture vectors in a geometric sense as two- or three-dimensional objects which can be acted on by certain geometric transformations encoded within matrices. This is certainly a good image to have in terms of visualising what's going on, but it betrays the breadth of linear algebra's applicability, especially in algebraic problems. In this essay, we shall explore how geometric ideas about vectors can inform more abstract definitions, and how these can be used to solve problems in completely separate fields.

1 Geometric Foundations

Geometric vectors (as we shall refer to them here) are two- or three-component objects corresponding to positions in or translations through a physical plane or volume. They are used frequently in problems regarding polygons and polyhedra, especially for finding intersections of shapes and reflections in given lines/planes.

1.1 Basis

Vectors are often written out as columns of two or three numbers ("column vectors"). When we do this, we are implicitly writing the vector as a sum of several *basis vectors*. These are special vectors which essentially form the "base ingredients" for the space we're operating in. Each one introduces a new dimension that our vector can inhabit; for example, a basis set in three dimensions could be the three unit vectors in the x-, y-, and z-directions. The first of these allows our vector to have any x-coordinate it wants, the second extends our range to the entire xy-plane, and the third allows us to make up any vector in all of xyz-space. If we denote these basis vectors by $\vec{i}, \vec{j}, \vec{k}$, then any three-dimensional vector may be written as

$$\begin{pmatrix} x \\ y \\ z \end{pmatrix} = x\vec{i} + y\vec{j} + z\vec{k}$$

and such a representation is unique.

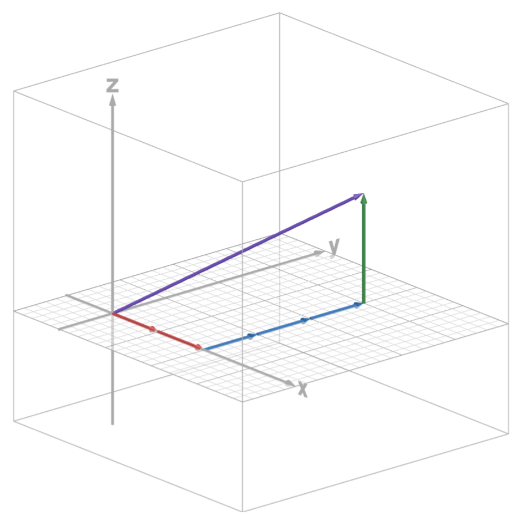

Figure 1: Example decomposition of a 3D vector into its basis components $(2\vec{i} + 3\vec{j} + \vec{k})$

If we wish to write a vector in n dimensions, we will clearly need one basis vector for each dimension and so our basis must consist of n different vectors where none can be written in terms of each other. The specific set of basis vectors chosen is not unique, but the representation of a given vector in any one set is unique.

1.2 Matrices

At its core, a matrix encodes a linear transformation by recording how the basis vectors are transformed. To understand why this is a sufficient amount of information to uniquely determine the transformation, consider a linear transformation with matrix M acting on a vector $\vec{v} = a\vec{i} + b\vec{j}$. We may use the linearity of the transformation to write

$$M\vec{v} = M(a\vec{i} + b\vec{j}) = aM\vec{i} + bM\vec{j}.$$

It is therefore clear that an entire linear transformation can be reduced to a transformation of the basis vectors, all other vectors transforming suitably as a result.

1.3 Dot products

The dot product is a way of producing a scalar value from two vectors. It is denoted by $\vec{u} \cdot \vec{v}$ and equals both the sum of the products of corresponding components (i.e. $u_1v_1 + u_2v_2 + u_3v_3$) and the value $|\vec{u}||\vec{v}|\cos(\theta)$, where $|\vec{u}|$ and $|\vec{v}|$ are vector lengths and θ is the angle between $\vec{u}$ and $\vec{v}$. Notably, the latter definition implies that $|\vec{u}| = \sqrt{\vec{u} \cdot \vec{u}}$ and that, for perpendicular vectors $\vec{u}$ and $\vec{v}$, $\vec{v} \cdot \vec{w} = 0$.

2 Introducing Abstraction

By this point, we have seen that the geometric properties of standard $\mathbb{R}^2$ and $\mathbb{R}^3$ vectors make them convenient for use in the analysis of linear transformations and projection problems. However, these same ideas arise in many areas outside of geometry. For example, differentiation follows the defining properties of a linear transformation and so its not too outlandish to imagine that there should be some way to apply the methods of linear algebra to analyse things like differential equations.

As we've briefly explored above, all of linear algebra can be derived from just three core operations:

- Vector addition (i.e. $\vec{u} + \vec{v}$ is well-defined)
- Vector scaling (i.e. $c\vec{u}$ is well-defined)
- Scalar product (i.e. $\vec{u} \cdot \vec{v}$ is well-defined)

An interesting question then naturally arises: are there any other sets of objects which support these operations? If so, it should then be possible to apply all methods of linear algebra to problems regarding these sets. Before investigating this question, it will help to give a more detailed explanation of "well-defined" for the above operations.

2.1 Vector addition and scaling

To give the most concise explanation, for vector addition and multiplication by scalars to be "well-defined" means that the vectors form an abelian group under addition and they form a ring homomorphism from the scalar field onto the endomorphism ring of the addition group. These are rather technical-sounding requirements, but they correspond to a list of very reasonable properties.

Vector addition and multiplication by scalars are well-defined if, for any vectors $\vec{u}, \vec{v}, \vec{w}$ and any scalars a, b, c, the following rules hold:

- Adding any two vectors yields another vector
- Order of addition does not matter at all ($\vec{u} + \vec{v} = \vec{v} + \vec{u}$ and $[\vec{u} + \vec{v}] + \vec{w} = \vec{u} + [\vec{v} + \vec{w}]$)
- There is an identity element under addition (there exists a vector $\vec{0}$ for which $\vec{u} + \vec{0} = \vec{0} + \vec{u} = \vec{u}$)
- Each vector $\vec{u}$ has an inverse under addition ($-\vec{u}$) such that $\vec{u} + [-\vec{u}] = \vec{0}$
- Multiplying a vector by a scalar yields another vector
- Order of multiplication is irrelevant ($a[b\vec{u}] = [ab]\vec{u}$)
- Multiplication by 1 does not change the vector ($1\vec{u} = \vec{u}$)
- The distributive law applies over vector sums ($c[\vec{u} + \vec{v}] = c\vec{u} + c\vec{v}$)
- The distributive law applies over scalar sums ($[a + b]\vec{u} = a\vec{u} + b\vec{u}$)

Any set of objects which satisfy the above rules is called a ***vector space***, the objects themselves being called ***vectors*** (Riley et al. 2006, p. 242). This definition alone is sufficient to define matrix algebra, and indeed there are some problems for which this implementation suffices (e.g. recording balances in various savings accounts, modelling the distribution of genotypes in controlled populations, and working with tokens in large language models). However, what this definition lacks is any analogue for the geometric ideas of "length" and "angle". These properties are what the dot product provides to our usual geometric vectors, so we shall define an abstraction for this next.

2.2 Inner products

An inner product, denoted by $(\vec{u}, \vec{v})$, is at its core a generalised version of a dot product. Just as was the case with dot products, the inner product takes two vectors and produces a scalar from them.

Similarly to how we went about defining vector addition and multiplication by scalars above, we will not prescribe any specific form to the inner product. Instead, we define it by its properties, allowing the specific form to be chosen to fit our problem.

For any vectors $\vec{u}, \vec{v}, \vec{w}$ and any scalars a, b, the following rules hold regarding the (real[1]) inner product:

- It is linear in the first argument, so $(a\vec{u} + b\vec{v}, \vec{w}) = a(\vec{u}, \vec{w}) + b(\vec{v}, \vec{w})$
- It is symmetrical on argument interchange, so $(\vec{u}, \vec{v}) = (\vec{v}, \vec{u})$
- It is "positive-definite", i.e. $(\vec{u}, \vec{u}) \geq 0$ with the inequality only saturated for $\vec{u} = \vec{0}$

With an inner product defined, we can then define a "norm" (a generalised length) identically to how we defined length in terms of the dot product. We write $||\vec{u}|| = \sqrt{(\vec{u}, \vec{u})}$ to be the norm of $\vec{u}$. The positive-definitive nature of the inner product then implies that this will always be defined, with only the zero vector having a length of 0.

Defining angles is the more interesting point. Having moved away from geometry, there is an argument that "angles" no longer mean anything; indeed, what possible use could there be in finding the angle between two tokens in an LLM or the angle between two different savings portfolios? Despite this, there is one specific angle which still remains incredibly important: 90°. Knowing that two vectors are perpendicular remains important even after leaving geometry, the interpretation then becoming that the vectors being compared have *no common content* (e.g. the vectors for "chalk" and "cheese" may be perpendicular in an LLM's token space).

Perpendicular vectors were a special case when considering the dot product, with this being the only situation where two non-zero vectors give a dot product of zero. We then take this as justification to define the idea of "orthogonal" vectors as being vectors whose inner product is zero. This is our generalised notion of a 90° angle.

If we take a standard vector space and equip it with an inner product, we define a so-called ***inner product space***. This is a set of objects for which *all of standard linear algebra is applicable.* The benefit of all this trouble is that now we can make a rather shocking claim: **suitable sets of functions constitute such spaces**.

3 Function Spaces

Consider the set of all (real) functions which are continuous on $[-1; 1]$, which we shall denote by $\mathcal{C}[-1; 1]$. If we use the standard definitions for addition and multiplication, then it is easily verifiable that both requirements for a vector space are satisfied on $\mathcal{C}[-1; 1]$. Furthermore, defining

$$(f, g) = \int_{-1}^{1} f(x)g(x)\, dx,$$

we can readily prove that this expression satisfies all the conditions for an inner product on $\mathcal{C}[-1; 1]$.

[1] If we allow our inner product to take on complex values, the second rule is altered such that interchange of arguments corresponds to complex conjugation.

- $(af+bg,h) = \int_{-1}^{1} [af(x)+bg(x)]h(x)\,dx = a\int_{-1}^{1} f(x)h(x)\,dx + b\int_{-1}^{1} g(x)h(x)\,dx = a(f,h)+b(g,h)$
- $(f,g) = \int_{-1}^{1} f(x)g(x)\,dx = \int_{-1}^{1} g(x)f(x)\,dx = (g,f)$
- $(f,f) = \int_{-1}^{1} f(x)f(x)\,dx = \int_{-1}^{1} [f(x)]^2\,dx \geq 0$

This is an example of a ***function space***, an inner product space whose "vectors" are functions. The form of inner product used here is typical for function spaces.

3.1 Basis sets

An introductory exercise regarding function spaces is to find a suitable basis set. Ideally, we would like to find a set of functions which are both mutually orthogonal and of unit length (an *orthonormal* basis set). Continuing with the $\mathcal{C}[-1;1]$ function space defined above, this corresponds to finding a set of functions $\{e_1(x), e_2(x), \cdots\}$ for which

$$(e_i, e_j) = \int_{-1}^{1} e_i(x)e_j(x)\,dx = \delta_{ij} = \begin{cases} 1 & i=j, \\ 0 & i \neq j. \end{cases}$$

It is a known result from Sturm-Liouville theory that a set of functions *almost* obeying the above condition is the *Legendre polynomials*, $P_n(x)$ $(n=0,1,\cdots)$ (Kreyszig. 1988, p. 231). Each $P_n(x)$ is a polynomial of degree n and the first few are shown below.

$$\begin{aligned} P_0(x) &= 1 \\ P_1(x) &= x \\ P_2(x) &= \frac{1}{2}(3x^2-1) \\ P_3(x) &= \frac{1}{2}(5x^3-3x) \\ &\vdots \end{aligned}$$

These functions are all orthogonal on the interval being considered, but they are not normalised (each having a norm of $2/(2n+1)$). This is easily rectified by dividing each polynomial by the root of its norm, yielding

$$e_n(x) = \sqrt{\frac{2n+1}{2}}P_n(x).$$

As a result of this, the full collection of (normalised) Legendre polynomials forms an orthonormal basis on $\mathcal{C}[-1;1]$. Furthermore, it can be shown that for the function space of n^{th} degree polynomials on the same interval ($\mathcal{P}_n[-1;1]$), the set $\{e_0, e_1, \cdots, e_n\}$ is a valid orthonormal basis.

4 Curve Fitting

It is often necessary to find approximate values for certain analytic functions, for example in calculators or computers. Taylor series can be used to do this if we only need accuracy over a small range of inputs, but the sheer number of terms needed to get accurate values further from the centre generally counteracts any benefits of this method. What would be preferable would be to find a polynomial approximation which is "accurate enough" throughout an entire interval. This is a common application of continuous curve fitting.

An important first step is to define our goal more quantitatively. We would like to find a polynomial $p(x)$ which fits our given continuous function $f(x)$ "fairly well" on some interval $[a;b]$. Another way of phrasing this is that we would like to reduce the error of our approximation across the interval. The issue with this statement is that error is a signed quantity: it can be negative or positive. As such, we could end up finding an approximation whose total error is very small but which massively overestimates and underestimates the given curve to roughly equal degrees. Instead, we choose to minimise the total *square* error, i.e. we wish to minimise the quantity

$$\int_a^b [f(x) - p(x)]^2 \, dx.$$

Assuming that the function we're approximating is a member of $\mathcal{C}[a;b]$, the total square error defined above is actually just the norm of the quantity $f(x) - p(x)$ (since $\mathcal{P}_n[a;b] \subset \mathcal{C}[a;b]$). In essence, we are therefore attempting to minimise the length of some "resultant vector" in our space of continuous functions.

4.1 Application to a period of sine

We shall demonstrate the method by finding the best cubic approximation $p(x)$ for $\sin(x)$ across one whole period: $[-\pi;\pi]$.

We begin by noting that $p(x) \in \mathcal{P}_3[-\pi;\pi]$ and $\sin(x) \in \mathcal{C}[-\pi;\pi]$. Let $B = \{e_0, e_1, e_2, e_3\}$ be an orthonormal basis set for $\mathcal{P}_3[-\pi;\pi]$ and $B' = \{e_0, e_1, \cdots\}$ be an orthonormal basis set for $\mathcal{C}[-\pi;\pi]$ (where we have chosen $B \subset B'$). We may therefore write

$$\begin{aligned} p(x) &= p_0 e_0(x) + p_1 e_1(x) + p_2 e_2(x) + p_3 e_3(x), \\ \sin(x) &= a_0 e_0(x) + a_1 e_1(x) + a_2 e_2(x) + a_3 e_3(x) + a_4 e_4(x) + \cdots . \end{aligned}$$

With these definitions, we are then at liberty to expand our resultant vector $\sin(x) - p(x)$ in components as

$$\begin{aligned} \sin(x) - p(x) &= (a_0 e_0 + a_1 e_1 + a_2 e_2 + a_3 e_3 + a_4 e_4 + \cdots) - (p_0 e_0 + p_1 e_1 + e_2 e_2 + p_3 e_3) \\ &= (a_0 - p_0) e_0 + (a_1 - p_1) e_1 + (a_2 - p_2) e_2 + (a_3 - p_3) e_3 + a_4 e_4 + \cdots \\ &= \sum_{i=0}^{\infty} r_i e_i(x). \end{aligned}$$

The resultant's norm, $||\sin(x) - p(x)||$, is then

$$\begin{aligned} ||\sin(x) - p(x)|| &= \int_{-\pi}^{\pi} [\sin(x) - p(x)]^2 \, dx \\ &= \int_{-\pi}^{\pi} \left(\sum_{i=0}^{\infty} r_i e_i(x) \right) \left(\sum_{i=0}^{\infty} r_i e_i(x) \right) \, dx \\ &= \int_{-\pi}^{\pi} \left[\sum_{i=0}^{\infty} \sum_{j=0}^{\infty} r_i r_j e_i(x) e_j(x) \right] \, dx \\ &= \sum_{i=0}^{\infty} \sum_{j=0}^{\infty} r_i r_j \left[\int_{-\pi}^{\pi} e_i(x) e_j(x) \, dx \right] = \sum_{i=0}^{\infty} \sum_{j=0}^{\infty} r_i r_j (e_i, e_j) \\ &= \sum_{i=0}^{\infty} r_i^2 \qquad \text{(By the orthonormality of the basis set)} \\ &= (a_0 - p_0)^2 + (a_1 - p_1)^2 + (a_2 - p_2)^2 + (a_3 - p_3)^3 + a_4^2 + \cdots . \end{aligned}$$

We now wish to minimise the above quantity. Since it is the polynomial $p(x)$ that we are choosing, we have no control over any of the a_n and hence we are only able to impact the first four terms of this expansion. Since these are all squared terms, the lowest value they can each take is zero. This is achieved by setting $p_i = a_i$ $(i = 0, 1, 2, 3)$. To draw a geometrical analogy, this essentially corresponds to "projecting" the $\sin(x)$ vector into the space $\mathcal{P}_3[-\pi;\pi]$.

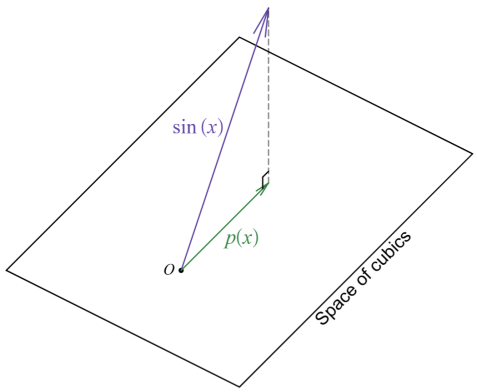

Figure 2: Geometric intuition for projecting $\sin(x)$ into $\mathcal{P}_3[-\pi;\pi]$

Now all that's left to do is to determine the coefficients $a_0, \cdots, a_3$. These can be obtained by taking the inner product of $\sin(x)$ with each of the basis vectors $e_0(x), \cdots, e_3(x)$ thanks to the orthonormality of this set. Referring to the results of section 3.1 and utilising the variable substitution $x = \pi u$, we find

$$\begin{aligned}
(e_i, e_j) &= \int_{-\pi}^{\pi} e_i(x) e_j(x)\, dx, \\
&= \int_{-1}^{1} e_i(\pi u) e_j(\pi u)\, (\pi\, du), \\
&= \int_{-1}^{1} \left[\sqrt{\pi} e_i(\pi u)\right]\left[\sqrt{\pi} e_j(\pi u)\right] du = \delta_{ij}, \\
\therefore \sqrt{\pi} e_n(\pi u) &= \sqrt{\frac{2n+1}{2}} P_n(u), \\
\implies e_n(x) &= \sqrt{\frac{2n+1}{2\pi}} P_n\left(\frac{x}{\pi}\right).
\end{aligned}$$

We may now carry out the required integrations to determine the value of each inner product, finding

$$a_0 = (\sin x, e_0) = 0, \qquad a_1 = (\sin x, e_1) = \sqrt{\frac{6}{\pi}}, \qquad a_2 = (\sin x, e_3) = 0, \qquad a_3 = (\sin x, e_3) = \sqrt{\frac{14}{\pi}}\left(1 - \frac{15}{\pi^2}\right).$$

This then finally gives our best fitting cubic of

$$\boxed{p(x) = \frac{35}{2\pi^4}\left(1 - \frac{15}{\pi^2}\right) x^3 - \frac{15}{2\pi^2}\left(1 - \frac{21}{\pi^2}\right) x}$$

with a maximum error of around 0.09059.

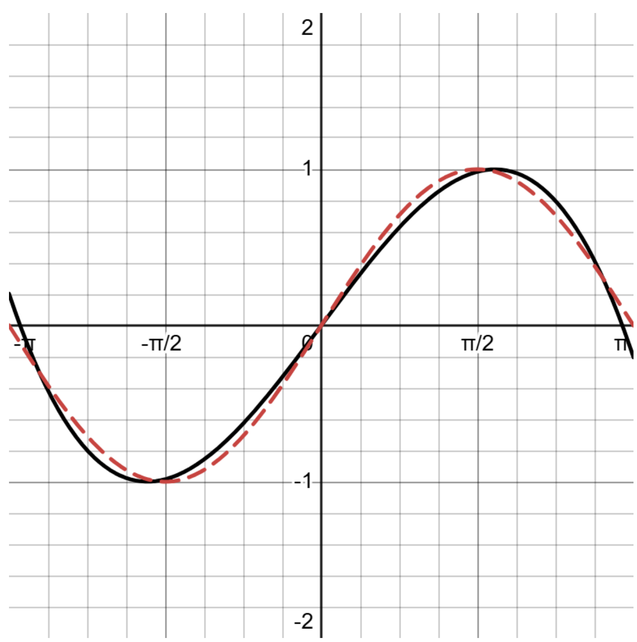

Figure 3: Cubic of best fit to $\sin(x)$

References

RILEY, K., HOBSON, M., and BENCE, S. (2006) *Mathematical Methods for Physics and Engineering.* 3rd ed. Cambridge: Cambridge University Press.

KREYSZIG, E. (1988) *Advanced Engineering Mathematics*. 6th ed. New Jersey: Wiley.