From Polynomials to Hilbert Spaces: Completeness, Orthogonal Projection, and the Geometry of Least Squares

Ryan Hausner

Abstract

Two ideas that appear to belong to separate disciplines—the convergence of infinite polynomial series and the minimization of squared error in regression—are instances of a single geometric principle: orthogonal projection onto a subspace. This note develops the argument from first principles. Beginning with finite-dimensional polynomial spaces, we show that attempting to represent transcendental functions such as e^x as limits of polynomial combinations exposes a structural failure: the space of all polynomials is not closed under its own limiting operations. The Hilbert space $L^2([-1, 1])$ is the natural completion that repairs this failure. We then show that ordinary least squares regression is orthogonal projection of a response vector onto the column space of the design matrix, a finite-dimensional subspace of $\mathbb{R}^n$. The normal equations, hat matrix, and sum-of-squares decomposition all follow directly from the geometry of inner product spaces, with no calculus required. In both settings, the best approximation is characterized by a single condition: the residual is orthogonal to the approximating subspace. The Hilbert space framework unifies these instances and provides the completeness necessary to guarantee that projections exist and infinite series converge.

Contents

1 Introduction **2**

2 Finite-Dimensional Polynomial Spaces **2**

3 The Taylor Series and the Failure of Finite Spaces **3**

4 Hilbert Spaces as the Necessary Extension **3**

5 Orthogonal Projection in Inner Product Spaces **4**

6 Ordinary Least Squares as Orthogonal Projection in $\mathbb{R}^n$ **5**

7 The Pythagorean Decomposition and ANOVA **6**

8 The Infinite-Dimensional Analogue: OLS in Function Space **6**

9 Residual Orthogonality as a Universal Principle **7**

10 Conclusion **8**

1 Introduction

A first course in linear algebra works entirely in $\mathbb{R}^n$: vectors are finite tuples, subspaces are finite-dimensional, and every object of interest is reachable by a finite linear combination. A first course in statistics introduces the ordinary least squares estimator as the minimizer of a sum of squared residuals, solved by setting a gradient to zero. Neither course typically reveals that these two topics are the same topic viewed from different angles.

The connecting thread is orthogonal projection. Given a vector $\mathbf{y}$ and a subspace V, there is a unique element of V closest to $\mathbf{y}$, and it is characterized by the condition that the error $\mathbf{y} - \hat{\mathbf{y}}$ is orthogonal to V. This is the geometric content of least squares. It is also, when V is a space of polynomials and $\mathbf{y}$ is replaced by a function f, the geometric content of function approximation by series expansion.

This note traces the argument from its most concrete form to its most general. Section 2 establishes finite-dimensional polynomial spaces and their limitations. Section 3 shows precisely where the polynomial framework breaks down via the Taylor series of e^x. Section 4 introduces Hilbert spaces as the necessary extension. Sections 5 and 6 develop the geometry of orthogonal projection in $\mathbb{R}^n$ and apply it to OLS. Section 7 connects the hat matrix and sum-of-squares decomposition to Pythagorean geometry. Section 8 shows that the same projection formula, extended to countably infinite bases, recovers function approximation in L^2. Section 9 summarizes the universal principle unifying all of these settings.

2 Finite-Dimensional Polynomial Spaces

Definition 1. *Let $\mathcal{P}_N$ denote the vector space of all real polynomials of degree at most N, with pointwise addition and scalar multiplication over $\mathbb{R}$.*

The set $\{1, x, x^2, \ldots, x^N\}$ is a basis for $\mathcal{P}_N$, so $\dim(\mathcal{P}_N) = N + 1$. Every element $p \in \mathcal{P}_N$ takes the form

$$p(x) = \sum_{k=0}^{N} a_k x^k, \qquad a_k \in \mathbb{R}.$$

On the interval $[-1, 1]$, define the inner product

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

which induces the norm $\|f\| = \langle f, f \rangle^{1/2}$. Under this inner product, $\mathcal{P}_N$ is a genuine inner product space. Because it is finite-dimensional, it is automatically complete: every Cauchy sequence in $\mathcal{P}_N$ converges to an element of $\mathcal{P}_N$.

Proposition 1. *Every finite-dimensional inner product space is complete.*

This follows from the equivalence of all norms in finite dimensions. Within $\mathcal{P}_N$, limits are well-behaved: the space is closed under all limiting operations one can perform on its elements. The difficulty arises the moment we attempt to represent functions that require infinitely many basis terms.

3 The Taylor Series and the Failure of Finite Spaces

Consider the exponential function e^x on $[-1, 1]$. Its Taylor expansion around $x = 0$ is

$$e^x = \sum_{n=0}^{\infty} \frac{x^n}{n!} = 1 + x + \frac{x^2}{2!} + \frac{x^3}{3!} + \cdots$$

The N-th partial sum

$$S_N(x) = \sum_{n=0}^{N} \frac{x^n}{n!}$$

belongs to $\mathcal{P}_N$ for each N. The sequence $(S_N)_{N\geq 0}$ converges pointwise to e^x on all of $\mathbb{R}$, uniformly on compact sets. In the L^2 norm,

$$\|e^x - S_N\|^2 = \int_{-1}^{1} \left(e^x - \sum_{n=0}^{N} \frac{x^n}{n!} \right)^2 dx \;\to\; 0 \quad \text{as } N \to \infty.$$

So (S_N) is a Cauchy sequence, yet its limit e^x is transcendental.

Remark 1. *The function e^x is not a polynomial. Suppose for contradiction that $e^x = \sum_{k=0}^{N} a_k x^k$ for some N and coefficients a_k. Differentiating both sides $N + 1$ times yields $e^x = 0$, a contradiction. More directly, $\frac{d}{dx} e^x = e^x$, but differentiation strictly reduces the degree of any nonzero polynomial. Thus $e^x \notin \mathcal{P}_N$ for any finite N.*

The natural ambient space for all polynomials is

$$\mathcal{P} = \bigcup_{N=0}^{\infty} \mathcal{P}_N.$$

But $\mathcal{P}$ is not closed under limits: (S_N) is Cauchy in $\mathcal{P}$ yet its limit lies outside $\mathcal{P}$. The same failure occurs for $\sin(x)$, $\cos(x)$, $\ln(1 + x)$, and every function whose Taylor series is non-terminating. Completeness is not a luxury. Without it, the standard tools of analysis—taking limits, solving operator equations, defining projections—cannot be guaranteed to produce elements that remain in the space.

4 Hilbert Spaces as the Necessary Extension

The resolution is to pass to the *completion* of $\mathcal{P}$ under the L^2 inner product.

Definition 2. *A **Hilbert space** $\mathcal{H}$ is a vector space over $\mathbb{R}$ (or $\mathbb{C}$) equipped with an inner product $\langle \cdot, \cdot \rangle$ such that the induced norm $\| \cdot \| = \langle \cdot, \cdot \rangle^{1/2}$ makes $\mathcal{H}$ a complete metric space: every Cauchy sequence converges to an element of $\mathcal{H}$.*

The space $L^2([-1,1])$, consisting of all measurable functions $f : [-1,1] \to \mathbb{R}$ satisfying $\int_{-1}^{1} f(x)^2\,dx < \infty$, with the inner product above, is a Hilbert space. Every continuous function on $[-1,1]$ belongs to $L^2([-1,1])$, including e^x. The polynomial subspace $\mathcal{P}$ is dense in $L^2([-1,1])$ by the Weierstrass approximation theorem: every L^2 function can be approximated arbitrarily closely by polynomials in the L^2 norm. Within this Hilbert space, the sequence (S_N) converges to e^x, and the limit is a genuine element of the space.

The inner product structure supports a theory of orthonormal bases generalizing the finite-dimensional case. In $L^2([-1,1])$, the Legendre polynomials $\{P_0, P_1, P_2, \ldots\}$, suitably normalized, form a countable orthonormal basis. Any $f \in L^2([-1,1])$ admits the representation

$$f = \sum_{n=0}^{\infty} \langle f, P_n \rangle\, P_n,$$

where the series converges in the L^2 norm and the coefficients $\langle f, P_n \rangle$ are the generalized Fourier coefficients of f. This is the infinite-dimensional analogue of expanding a vector in an orthonormal basis of $\mathbb{R}^n$, and it is valid precisely because L^2 is complete. The meaning of these coefficients, and why they are optimal, is the subject of the next sections.

5 Orthogonal Projection in Inner Product Spaces

The geometric engine behind all of the above is a single theorem. Let $\mathcal{H}$ be a Hilbert space and $V \subset \mathcal{H}$ a closed subspace.

Definition 3. *The **orthogonal projection** of $f \in \mathcal{H}$ onto V is the unique element $\hat{f} \in V$ satisfying*

$$f - \hat{f} \perp V, \qquad \textit{i.e.,} \quad \langle f - \hat{f},\, v \rangle = 0 \quad \textit{for all } v \in V.$$

Theorem 1 (Projection Theorem)**.** *Let V be a closed subspace of a Hilbert space $\mathcal{H}$. For every $f \in \mathcal{H}$, the orthogonal projection $\hat{f}$ exists, is unique, and satisfies*

$$\|f - \hat{f}\| = \min_{v \in V} \|f - v\|.$$

Proof. For any $v \in V$, write $f - v = (f - \hat{f}) + (\hat{f} - v)$. Since $f - \hat{f} \perp V$ and $\hat{f} - v \in V$, the Pythagorean theorem gives

$$\|f - v\|^2 = \|f - \hat{f}\|^2 + \|\hat{f} - v\|^2 \geq \|f - \hat{f}\|^2,$$

with equality if and only if $v = \hat{f}$. Uniqueness follows immediately. □

This equivalence is the central fact: *the projection is the least-squares approximation, and the least-squares approximation is the projection.* The orthogonality condition and the minimization condition are two descriptions of the same object. All subsequent results in both function approximation and regression are consequences of this theorem.

6 Ordinary Least Squares as Orthogonal Projection in $\mathbb{R}^n$

The standard OLS derivation minimizes

$$\mathcal{L}(\boldsymbol{\beta}) = \|\mathbf{y} - \mathbf{X}\boldsymbol{\beta}\|^2$$

by setting the gradient to zero, arriving at the normal equations $\mathbf{X}^\top\mathbf{X}\hat{\boldsymbol{\beta}} = \mathbf{X}^\top\mathbf{y}$. The algebra is correct but obscures the geometry. We now rederive OLS purely from the Projection Theorem.

Let $\mathbf{X} \in \mathbb{R}^{n\times p}$ have full column rank, and let $\mathcal{C}(\mathbf{X}) \subset \mathbb{R}^n$ denote its column space, a p-dimensional subspace. The space $\mathbb{R}^n$ equipped with the standard inner product $\langle \mathbf{u}, \mathbf{v}\rangle = \mathbf{u}^\top\mathbf{v}$ is a Hilbert space (finite-dimensional, hence complete). By the Projection Theorem, the unique minimizer of $\|\mathbf{y} - \mathbf{v}\|$ over $\mathbf{v} \in \mathcal{C}(\mathbf{X})$ is the orthogonal projection $\hat{\mathbf{y}}$, characterized by

$$\mathbf{y} - \hat{\mathbf{y}} \perp \mathcal{C}(\mathbf{X}).$$

Since any element of $\mathcal{C}(\mathbf{X})$ takes the form $\mathbf{X}\boldsymbol{\beta}$, the orthogonality condition is

$$\langle \mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}},\, \mathbf{X}\boldsymbol{\beta}\rangle = 0 \quad \text{for all } \boldsymbol{\beta} \in \mathbb{R}^p.$$

Writing this as $\boldsymbol{\beta}^\top\mathbf{X}^\top(\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}}) = 0$ for all $\boldsymbol{\beta}$ yields

$$\mathbf{X}^\top(\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}}) = \mathbf{0},$$

which are the normal equations. They do not arise from calculus; they are the finite-dimensional expression of residual orthogonality. Solving gives

$$\hat{\boldsymbol{\beta}} = (\mathbf{X}^\top\mathbf{X})^{-1}\mathbf{X}^\top\mathbf{y}, \qquad \hat{\mathbf{y}} = \underbrace{\mathbf{X}(\mathbf{X}^\top\mathbf{X})^{-1}\mathbf{X}^\top}_{\mathbf{H}}\,\mathbf{y},$$

where $\mathbf{H}$ is the **hat matrix** (projection matrix onto $\mathcal{C}(\mathbf{X})$).

Proposition 2. *The hat matrix $\mathbf{H} = \mathbf{X}(\mathbf{X}^\top\mathbf{X})^{-1}\mathbf{X}^\top$ satisfies:*

1. $\mathbf{H}^2 = \mathbf{H}$ *(idempotency: projecting twice equals projecting once),*
2. $\mathbf{H}^\top = \mathbf{H}$ *(self-adjointness: orthogonal projections are symmetric),*
3. $\operatorname{rank}(\mathbf{H}) = p$*,*
4. $\mathbf{I} - \mathbf{H}$ *is the orthogonal projection onto $\mathcal{C}(\mathbf{X})^\perp$.*

All four properties follow immediately from the geometric definition of projection; none requires matrix algebra beyond substitution.

7 The Pythagorean Decomposition and ANOVA

The residual vector $\mathbf{e} = (\mathbf{I} - \mathbf{H})\mathbf{y}$ is the component of $\mathbf{y}$ orthogonal to $\mathcal{C}(\mathbf{X})$. The full orthogonal decomposition of $\mathbf{y}$ is

$$\mathbf{y} = \underbrace{\mathbf{H}\mathbf{y}}_{\hat{\mathbf{y}} \in \mathcal{C}(\mathbf{X})} + \underbrace{(\mathbf{I} - \mathbf{H})\mathbf{y}}_{\mathbf{e} \perp \mathcal{C}(\mathbf{X})}.$$

Since $\hat{\mathbf{y}} \perp \mathbf{e}$, the Pythagorean theorem applies directly:

$$\|\mathbf{y}\|^2 = \|\hat{\mathbf{y}}\|^2 + \|\mathbf{e}\|^2.$$

When $\mathbf{y}$ is mean-centered (so the intercept column of $\mathbf{X}$ is absorbed), this is

$$\underbrace{\sum_{i=1}^{n}(y_i - \bar{y})^2}_{\text{SS}_{\text{Total}}} = \underbrace{\sum_{i=1}^{n}(\hat{y}_i - \bar{y})^2}_{\text{SS}_{\text{Model}}} + \underbrace{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}_{\text{SS}_{\text{Residual}}}.$$

ANOVA is Pythagoras. The F-test in regression is a comparison of the lengths of two orthogonal components of a vector. The coefficient of determination is

$$R^2 = \frac{\|\hat{\mathbf{y}}\|^2}{\|\mathbf{y}\|^2} = \cos^2\theta,$$

where θ is the angle between $\mathbf{y}$ and its projection $\hat{\mathbf{y}}$ onto the column space. An R^2 of 1 means $\mathbf{y}$ lies exactly in $\mathcal{C}(\mathbf{X})$; an R^2 of 0 means $\mathbf{y}$ is orthogonal to every predictor.

Orthonormal Predictors and Fourier Coefficients

When the columns of $\mathbf{X}$ are orthonormal ($\mathbf{X}^\top\mathbf{X} = \mathbf{I}_p$), the projection formula simplifies to

$$\hat{\mathbf{y}} = \mathbf{X}\mathbf{X}^\top\mathbf{y} = \sum_{j=1}^{p}\langle \mathbf{y}, \mathbf{x}_j\rangle\, \mathbf{x}_j,$$

and the OLS coefficients become $\hat{\beta}_j = \langle \mathbf{y}, \mathbf{x}_j\rangle$. These are the generalized Fourier coefficients of $\mathbf{y}$ with respect to the basis $\{\mathbf{x}_j\}$: the fitted value is the sum of the one-dimensional projections of $\mathbf{y}$ onto each basis vector. When predictors are not orthonormal, Gram–Schmidt orthogonalization produces an orthonormal basis $\{\mathbf{q}_1, \ldots, \mathbf{q}_p\}$ spanning the same column space, reducing OLS to p independent one-dimensional projections. The QR decomposition $\mathbf{X} = \mathbf{Q}\mathbf{R}$ formalizes this: $\hat{\boldsymbol{\beta}} = \mathbf{R}^{-1}\mathbf{Q}^\top\mathbf{y}$, numerically more stable than inverting $\mathbf{X}^\top\mathbf{X}$.

8 The Infinite-Dimensional Analogue: OLS in Function Space

The projection formula

$$\hat{\mathbf{y}} = \sum_{j=1}^{p}\langle \mathbf{y}, \mathbf{x}_j\rangle\, \mathbf{x}_j$$

does not depend on the finiteness of p. In a Hilbert space $\mathcal{H}$ with a countable orthonormal basis $\{\phi_j\}_{j=1}^{\infty}$, every $f \in \mathcal{H}$ admits the representation

$$f = \sum_{j=1}^{\infty} \langle f, \phi_j \rangle \, \phi_j,$$

where convergence holds in the norm of $\mathcal{H}$. By Parseval's identity,

$$\|f\|^2 = \sum_{j=1}^{\infty} |\langle f, \phi_j \rangle|^2 < \infty,$$

so the partial sums

$$\hat{f}_N = \sum_{j=1}^{N} \langle f, \phi_j \rangle \, \phi_j$$

form a Cauchy sequence converging in norm to f. Each $\hat{f}_N$ is the orthogonal projection of f onto $\operatorname{span}\{\phi_1, \ldots, \phi_N\}$, and by the Projection Theorem it minimizes $\|f - g\|$ over all g in that subspace. This is OLS in function space: fitting f with an N-term expansion is the least-squares problem of approximating f from an N-dimensional subspace, solved by orthogonal projection.

For $\mathcal{H} = L^2([-1, 1])$ with the Legendre polynomial basis $\{P_j\}$, the projection $\hat{f}_N$ is the best L^2 polynomial approximation to f of degree at most N. The Taylor series partial sums S_N constitute a special case, but in a non-orthogonal basis $\{1, x, x^2, \ldots\}$: the Taylor coefficients are not inner products with an orthonormal set, which is why Legendre expansions converge faster in L^2 for generic smooth functions.

Completeness of $\mathcal{H}$ is essential throughout. The Cauchy sequence $(\hat{f}_N)$ must converge to an element of the space. The polynomial space $\mathcal{P}$ fails this requirement: (S_N) is Cauchy in $\mathcal{P}$ but $e^x \notin \mathcal{P}$. In $L^2([-1, 1])$, the limit exists and belongs to the space. The extension from finite- to infinite-dimensional projection is valid precisely because L^2 is complete.

9 Residual Orthogonality as a Universal Principle

The following table summarizes the three settings developed in this note. In each case, the best approximation from a subspace is characterized by the same geometric condition: the residual is orthogonal to the subspace.

Setting	**Space**	**Subspace**	**Orthogonality Condition**
OLS regression	$\mathbb{R}^n$	$\mathcal{C}(\mathbf{X})$, $\dim = p$	$\mathbf{X}^\top(\mathbf{y} - \hat{\mathbf{y}}) = \mathbf{0}$
Polynomial approximation	$L^2([-1, 1])$	$\operatorname{span}\{P_0, \ldots, P_N\}$	$\langle f - \hat{f}_N, P_j \rangle = 0, \ j \leq N$
Fourier series	$L^2([0, 2\pi])$	$\operatorname{span}\{e^{ij\,\cdot}\}_{\lvert j \rvert \leq N}$	$\langle f - \hat{f}_N, e^{ij\,\cdot} \rangle = 0, \ \lvert j \rvert \leq N$

The normal equations of OLS are not an algebraic artifact of calculus. They are the finite-dimensional instance of a universal optimality condition. The hat matrix is not a

computational convenience; it is the matrix representation of the projection operator onto a subspace. The sum-of-squares decomposition is not a statistical convention; it is the Pythagorean theorem applied to an orthogonal decomposition of $\mathbf{y}$.

10 Conclusion

The argument in this note follows a single line from beginning to end. Finite-dimensional polynomial spaces are complete and geometrically well-behaved, but they cannot contain the limits of infinite polynomial combinations. The Taylor series of e^x forces this limitation into view: the partial sums are polynomials, but their limit is transcendental, and the space of all polynomials is incomplete. The Hilbert space $L^2([-1,1])$ is the natural completion, and within it, the projection formula $\hat{f}_N = \sum_{j=1}^{N} \langle f, \phi_j \rangle \phi_j$ provides the best N-term approximation to any function.

The same projection formula, restricted to $\mathbb{R}^n$ with $\mathbf{y}$ in place of f and the columns of $\mathbf{X}$ in place of the orthonormal basis, is ordinary least squares. The normal equations are residual orthogonality. The hat matrix is the projection operator. The sum-of-squares identity is the Pythagorean theorem. R^2 is a squared cosine. ANOVA is a comparison of orthogonal lengths.

These are not analogies. They are the same theorem, instantiated in different Hilbert spaces. The completeness of those spaces is what guarantees the projection exists, the series converges, and the least-squares problem has a unique solution.