GES-207-4 — Mathematical Methods of Physics

National Institute for Space Research (INPE)
Graduate Program in Space Geophysics (PPGGES)

Lecture Notes 03

Gram–Schmidt Process, Orthonormal Bases, and Orthogonal Projections

Tuesday, September 29, 2026 · Prof. Dr. José Paulo Marchezi · DICEP · 3rd Term 2026

Learning goals and study scope. This self-contained lecture develops Gram–Schmidt rather than merely stating its formula. We begin with bases and orthogonality, construct the method in $\mathbb{R}^2$, extend it to $\mathbb{R}^3$ and arbitrary inner-product spaces, and then apply it to polynomials. We subsequently study orthogonal projection, best approximation, projectors, least squares, QR, Bessel's inequality, and Parseval's identity. Every use of projection assumes a specified inner product; changing the inner product changes orthogonality.

Contents

1 Review: Bases, Orthogonality, and Notation **1**
1.1 What a basis does 2

2 Constructing Gram–Schmidt in $\mathbb{R}^2$ **2**

3 From $\mathbb{R}^3$ to an Arbitrary Dimension **3**
3.1 Extracting the QR factorization from the same calculation 4
3.2 Complex spaces and a convention caveat 4

4 Function Spaces and Orthogonal Polynomials **5**

5 Orthogonal Projection onto a Subspace **5**

6 Least Squares and the QR Factorization **6**

7 Bessel's Inequality and Parseval's Identity **8**

8 Summary and Checks for Understanding **8**

1. Review: Bases, Orthogonality, and Notation

1.1. What a basis does

A finite set $B = \{\mathbf{v}_1, \dots, \mathbf{v}_n\}$ is a basis of V when it is linearly independent and spans V. Consequently every $\mathbf{v} \in V$ has a unique representation $\mathbf{v} = \sum_j c_j \mathbf{v}_j$. A basis need not be orthogonal. In an inner-product space, $\mathbf{u} \perp \mathbf{v}$ means $\langle \mathbf{u}, \mathbf{v} \rangle = 0$. A set is *orthogonal* if distinct members are orthogonal, and *orthonormal* if in addition every member has norm one.

How to read span

The expression $\operatorname{span}\{\mathbf{v}_1, \mathbf{v}_2\}$ is read "the span of $\mathbf{v}_1$ and $\mathbf{v}_2$," or "the space generated by them," and means

$$\operatorname{span}\{\mathbf{v}_1, \mathbf{v}_2\} = \{a\mathbf{v}_1 + b\mathbf{v}_2 : a, b \in \mathbb{R}\}.$$

Thus $\operatorname{span}\{\mathbf{w}_1, \mathbf{w}_2\} = \operatorname{span}\{\mathbf{v}_1, \mathbf{v}_2\}$ says that the vectors have changed but the space they describe has not.

For $B = \{(2, -1), (3, 4)\}$, the determinant $\det\left(\begin{smallmatrix} 2 & 3 \\ -1 & 4 \end{smallmatrix}\right) = 11 \neq 0$, so B is a basis of $\mathbb{R}^2$. Yet $\langle (2, -1), (3, 4) \rangle = 6 - 4 = 2 \neq 0$; it is not orthogonal. Can we replace it by an orthogonal basis without changing its span? Gram–Schmidt answers yes.

2. Constructing Gram–Schmidt in $\mathbb{R}^2$

Choose $\mathbf{w}_1 = \mathbf{v}_1$. Seek $\mathbf{w}_2 = \mathbf{v}_2 + \alpha \mathbf{w}_1$ with $\langle \mathbf{w}_1, \mathbf{w}_2 \rangle = 0$. Linearity gives

$$0 = \langle \mathbf{w}_1, \mathbf{v}_2 \rangle + \alpha \langle \mathbf{w}_1, \mathbf{w}_1 \rangle, \qquad \alpha = -\frac{\langle \mathbf{w}_1, \mathbf{v}_2 \rangle}{\langle \mathbf{w}_1, \mathbf{w}_1 \rangle}.$$

Hence

$$\boxed{\mathbf{w}_2 = \mathbf{v}_2 - \frac{\langle \mathbf{w}_1, \mathbf{v}_2 \rangle}{\langle \mathbf{w}_1, \mathbf{w}_1 \rangle} \mathbf{w}_1.} \tag{1}$$

How to read and interpret proj

$$\operatorname{proj}_{\mathbf{w}_1}(\mathbf{v}_2) = \frac{\langle \mathbf{w}_1, \mathbf{v}_2 \rangle}{\langle \mathbf{w}_1, \mathbf{w}_1 \rangle} \mathbf{w}_1$$

is read "the projection of $\mathbf{v}_2$ onto $\mathbf{w}_1$." It is the part of $\mathbf{v}_2$ pointing along $\mathbf{w}_1$: the result is a scalar multiple of $\mathbf{w}_1$, and the scalar measures how much of $\mathbf{v}_2$ lies in that direction. Equation (1) says: subtract the parallel part; the remainder is perpendicular.

Example 3.1: the basis $\{(2, -1), (3, 4)\}$

Let $\mathbf{v}_1 = (2, -1)$ and $\mathbf{v}_2 = (3, 4)$. Then

$$\mathbf{w}_1 = (2, -1), \quad \langle \mathbf{w}_1, \mathbf{v}_2 \rangle = 2, \quad \langle \mathbf{w}_1, \mathbf{w}_1 \rangle = 5,$$

and

$$\mathbf{w}_2 = (3,4) - \frac{2}{5}(2,-1) = \left(\frac{11}{5}, \frac{22}{5}\right) = \frac{11}{5}(1,2).$$

The check is explicit: $\langle (2,-1), (11/5, 22/5)\rangle = 22/5 - 22/5 = 0$. Scaling a nonzero basis vector does not change its direction or span, so the simpler orthogonal basis $\{(2,-1),(1,2)\}$ is equivalent. Normalizing gives

$$\boxed{\mathbf{u}_1 = \frac{1}{\sqrt{5}}(2,-1), \qquad \mathbf{u}_2 = \frac{1}{\sqrt{5}}(1,2).}$$

Both vectors have norm one and their inner product is zero.

Orthogonalization is not normalization. Gram–Schmidt determines mutually perpendicular directions. Dividing each nonzero result by its norm is a separate operation that makes the basis orthonormal.

3. From $\mathbb{R}^3$ to an Arbitrary Dimension

For three independent vectors,

$$\begin{aligned}
\mathbf{w}_1 &= \mathbf{v}_1, \\
\mathbf{w}_2 &= \mathbf{v}_2 - \mathrm{proj}_{\mathbf{w}_1}(\mathbf{v}_2), \\
\mathbf{w}_3 &= \mathbf{v}_3 - \mathrm{proj}_{\mathbf{w}_1}(\mathbf{v}_3) - \mathrm{proj}_{\mathbf{w}_2}(\mathbf{v}_3).
\end{aligned}$$

Removing only the first component would not generally make $\mathbf{w}_3$ orthogonal to $\mathbf{w}_2$; every previously constructed direction must be removed.

Gram–Schmidt theorem

For a linearly independent set $\{\mathbf{v}_1, \ldots, \mathbf{v}_n\}$, define

$$\mathbf{w}_1 = \mathbf{v}_1, \qquad \mathbf{w}_k = \mathbf{v}_k - \sum_{j=1}^{k-1} \frac{\langle \mathbf{w}_j, \mathbf{v}_k\rangle}{\langle \mathbf{w}_j, \mathbf{w}_j\rangle}\mathbf{w}_j, \qquad \mathbf{u}_k = \frac{\mathbf{w}_k}{\|\mathbf{w}_k\|}.$$

Then $\{\mathbf{u}_1, \ldots, \mathbf{u}_n\}$ is orthonormal and, for every k,

$$\mathrm{span}\{\mathbf{u}_1, \ldots, \mathbf{u}_k\} = \mathrm{span}\{\mathbf{v}_1, \ldots, \mathbf{v}_k\}.$$

To verify orthogonality, take $j < k$ and compute

$$\langle \mathbf{w}_j, \mathbf{w}_k\rangle = \langle \mathbf{w}_j, \mathbf{v}_k\rangle - \sum_{i=1}^{k-1} \frac{\langle \mathbf{w}_i, \mathbf{v}_k\rangle}{\langle \mathbf{w}_i, \mathbf{w}_i\rangle}\langle \mathbf{w}_j, \mathbf{w}_i\rangle = 0,$$

because all terms vanish except $i = j$, which cancels the first term. If some $\mathbf{w}_k$ were zero, $\mathbf{v}_k$ would be a combination of earlier vectors, contradicting independence. This also

explains why the hypothesis is essential.

Example 3.2: a complete computation in $\mathbb{R}^3$

For $\mathbf{v}_1 = (1,1,0)^T$, $\mathbf{v}_2 = (1,0,1)^T$, $\mathbf{v}_3 = (0,1,1)^T$, one obtains

$$\mathbf{u}_1 = \frac{1}{\sqrt{2}}(1,1,0)^T.$$

Next,

$$\mathbf{w}_2 = \mathbf{v}_2 - \frac{1}{2}(1,1,0)^T = (1/2,-1/2,1)^T, \quad \mathbf{u}_2 = \frac{1}{\sqrt{6}}(1,-1,2)^T.$$

Since $\langle \mathbf{u}_1, \mathbf{v}_3 \rangle = 1/\sqrt{2}$ and $\langle \mathbf{u}_2, \mathbf{v}_3 \rangle = 1/\sqrt{6}$,

$$\mathbf{w}_3 = (0,1,1)^T - \frac{1}{2}(1,1,0)^T - \frac{1}{6}(1,-1,2)^T = (-2/3,2/3,2/3)^T,$$

so $\mathbf{u}_3 = (-1,1,1)^T/\sqrt{3}$. Direct dot products verify all three pairs are orthogonal.

3.1. Extracting the QR factorization from the same calculation

Let $A = [\mathbf{v}_1|\mathbf{v}_2|\mathbf{v}_3]$ for Example 3.2 and $Q = [\mathbf{u}_1|\mathbf{u}_2|\mathbf{u}_3]$. At each Gram–Schmidt step,

$$\mathbf{v}_k = \sum_{j=1}^{k} r_{jk}\mathbf{u}_j, \qquad r_{jk} = \langle \mathbf{u}_j, \mathbf{v}_k \rangle \ (j<k), \qquad r_{kk} = \|\mathbf{w}_k\|.$$

The coefficients below the diagonal are zero because $\mathbf{v}_k$ uses only the first k orthonormal directions. In this example,

$$Q = \begin{pmatrix} 1/\sqrt{2} & 1/\sqrt{6} & -1/\sqrt{3} \\ 1/\sqrt{2} & -1/\sqrt{6} & 1/\sqrt{3} \\ 0 & 2/\sqrt{6} & 1/\sqrt{3} \end{pmatrix}, \quad R = \begin{pmatrix} \sqrt{2} & 1/\sqrt{2} & 1/\sqrt{2} \\ 0 & \sqrt{6}/2 & 1/\sqrt{6} \\ 0 & 0 & 2/\sqrt{3} \end{pmatrix}.$$

Thus $A = QR$. This equality is not a new theorem added after Gram–Schmidt: it is the same sequence of projection coefficients written in matrix form.

3.2. Complex spaces and a convention caveat

In $\mathbb{C}^n$ the Euclidean inner product becomes Hermitian: $\langle \mathbf{u}, \mathbf{v} \rangle = \mathbf{u}^H\mathbf{v}$. Complex conjugation is indispensable. This text takes the inner product conjugate-linear in the first argument and linear in the second. Some books reverse the convention; their projection coefficients must be written consistently.

4. Function Spaces and Orthogonal Polynomials

Gram–Schmidt needs an inner product, not a list of coordinates. For functions, a standard choice is

$$\langle f, g\rangle = \int_a^b f(x)g(x)\,dx.$$

Thus orthogonality means a zero integral, not pointwise perpendicular graphs.

Example 3.3: $\{1, x, x^2\}$ ***on*** $[-1, 1]$

Use $\langle f, g\rangle = \int_{-1}^{1} f(x)g(x)\,dx$. Start with $w_1 = 1$. Because x is odd,

$$w_2 = x - \frac{\langle 1, x\rangle}{\langle 1, 1\rangle} = x.$$

For the third polynomial,

$$w_3 = x^2 - \frac{\langle 1, x^2\rangle}{\langle 1, 1\rangle} - \frac{\langle x, x^2\rangle}{\langle x, x\rangle}x.$$

Now

$$\langle 1, x^2\rangle = \int_{-1}^{1} x^2 dx = \frac{2}{3}, \quad \langle 1, 1\rangle = 2, \quad \langle x, x^2\rangle = \int_{-1}^{1} x^3 dx = 0.$$

Therefore

$$\boxed{w_3 = x^2 - \frac{1}{3}}, \qquad \boxed{\{1, x, x^2 - 1/3\} \text{ is an orthogonal basis of } P_2[-1, 1].}$$

For an orthonormal basis, the norms are $\|1\|^2 = 2$, $\|x\|^2 = 2/3$, and $\|x^2 - 1/3\|^2 = 8/45$. Hence

$$u_0 = \frac{1}{\sqrt{2}}, \qquad u_1 = \sqrt{\frac{3}{2}}x, \qquad u_2 = \sqrt{\frac{45}{8}}\left(x^2 - \frac{1}{3}\right).$$

The first Legendre polynomials are $P_0 = 1$, $P_1 = x$, and $P_2 = (3x^2 - 1)/2$. Since $x^2 - 1/3 = (2/3)P_2$, Gram–Schmidt applied to $1, x, x^2, \ldots$ on $[-1, 1]$ produces the Legendre family up to scaling. In physics these polynomials appear in Laplace's equation in spherical coordinates, multipole expansions, gravitational and electrostatic potentials, and angular dependence of fields.

5. Orthogonal Projection onto a Subspace

Projection and best approximation

Let $W = \text{span}\{\mathbf{u}_1, \ldots, \mathbf{u}_k\}$ with an orthonormal basis. Then

$$P_W\mathbf{v} = \sum_{j=1}^{k} \langle \mathbf{u}_j, \mathbf{v}\rangle \mathbf{u}_j, \qquad \mathbf{v} - P_W\mathbf{v} \in W^{\perp}.$$

Moreover, $P_W\mathbf{v}$ is the unique point of W closest to $\mathbf{v}$.

Indeed, for any $\mathbf{w} \in W$,

$$\mathbf{v} - \mathbf{w} = (\mathbf{v} - P_W\mathbf{v}) + (P_W\mathbf{v} - \mathbf{w}),$$

where the two terms are orthogonal. Pythagoras gives

$$\|\mathbf{v} - \mathbf{w}\|^2 = \|\mathbf{v} - P_W\mathbf{v}\|^2 + \|P_W\mathbf{v} - \mathbf{w}\|^2 \geq \|\mathbf{v} - P_W\mathbf{v}\|^2.$$

Equality occurs only for $\mathbf{w} = P_W\mathbf{v}$.

If $Q = [\mathbf{u}_1|\cdots|\mathbf{u}_k]$, then

$$\boxed{P_W = QQ^T \quad \text{(real)}, \qquad P_W = QQ^H \quad \text{(complex)}.}$$

The projector satisfies $P_W^2 = P_W$, $P_W^* = P_W$, $\operatorname{im} P_W = W$, $\ker P_W = W^\perp$, and $P_{W^\perp} = I - P_W$. Idempotence says that projecting an already projected vector changes nothing; self-adjointness distinguishes an orthogonal projector from an oblique one.

Example 3.4: projection and distance to a plane

Let

$$W = \operatorname{span}\left\{\frac{1}{\sqrt{2}}(1,1,0)^T, (0,0,1)^T\right\}, \qquad \mathbf{v} = (2,0,3)^T.$$

The displayed spanning vectors are orthonormal. Their coefficients are

$$\left\langle \frac{1}{\sqrt{2}}(1,1,0)^T, \mathbf{v} \right\rangle = \sqrt{2}, \qquad \langle (0,0,1)^T, \mathbf{v}\rangle = 3.$$

Consequently

$$P_W\mathbf{v} = (1,1,0)^T + (0,0,3)^T = (1,1,3)^T, \qquad \mathbf{r} = \mathbf{v} - P_W\mathbf{v} = (1,-1,0)^T.$$

The residual is orthogonal to both basis vectors, so the distance from $\mathbf{v}$ to W is $\|\mathbf{r}\| = \sqrt{2}$. For any $\mathbf{w} \in W$,

$$\|\mathbf{v} - \mathbf{w}\|^2 = 2 + \|P_W\mathbf{v} - \mathbf{w}\|^2 \geq 2,$$

which verifies the best-approximation statement numerically and symbolically.

If the spanning columns of A are independent but not orthonormal, the projector is $P = A(A^TA)^{-1}A^T$ in the real case. Writing merely AA^T is correct only when $A^TA = I$.

6. Least Squares and the QR Factorization

An inconsistent system $A\mathbf{x} = \mathbf{b}$ cannot be solved exactly. Least squares chooses $\widehat{\mathbf{x}}$ minimizing $\|A\mathbf{x} - \mathbf{b}\|$. The fitted vector $A\widehat{\mathbf{x}}$ is the projection of $\mathbf{b}$ onto the column space

of A, so the residual $\mathbf{r} = \mathbf{b} - A\hat{\mathbf{x}}$ is orthogonal to every column:

$$A^T\mathbf{r} = 0 \quad \Longrightarrow \quad A^T A\hat{\mathbf{x}} = A^T\mathbf{b}.$$

These are the normal equations. They are mathematically transparent but can be numerically less stable because forming $A^T A$ squares the condition number.

Gram–Schmidt applied to the columns of a full-rank matrix produces $A = QR$, where $Q^T Q = I$ and R is upper triangular. Then

$$\|A\mathbf{x} - \mathbf{b}\|^2 = \|R\mathbf{x} - Q^T\mathbf{b}\|^2 + \|(I - QQ^T)\mathbf{b}\|^2,$$

and the minimizer solves the triangular system

$$\boxed{R\hat{\mathbf{x}} = Q^T\mathbf{b}.}$$

In serious numerical work, modified Gram–Schmidt or Householder reflections are preferred to classical Gram–Schmidt because roundoff can destroy orthogonality.

Example 3.5: fitting a linear trend

For observations (t_i, b_i), write

$$A = \begin{pmatrix} 1 & t_1 \\ \vdots & \vdots \\ 1 & t_N \end{pmatrix}, \quad \mathbf{x} = \begin{pmatrix} c_0 \\ c_1 \end{pmatrix}, \quad \mathbf{b} = (b_1, \ldots, b_N)^T.$$

The first normalized column is $\mathbf{u}_1 = \mathbf{1}/\sqrt{N}$. Removing its projection from the time column gives $\tilde{\mathbf{u}}_2 = \mathbf{t} - \bar{t}\mathbf{1}$. Therefore the two orthogonal directions represent a constant level and a centered linear trend. Projecting $\mathbf{b}$ onto their span yields the least-squares trend; the residual is orthogonal to both and may represent higher-frequency variation after instrumental drift removal.

Example 3.6: a complete three-point least-squares fit

Fit $b(t) = c_0 + c_1 t$ to the measurements $(-1, 1)$, $(0, 2)$, and $(1, 2)$. Here

$$A = \begin{pmatrix} 1 & -1 \\ 1 & 0 \\ 1 & 1 \end{pmatrix}, \qquad \mathbf{b} = \begin{pmatrix} 1 \\ 2 \\ 2 \end{pmatrix}.$$

The columns are already orthogonal: their squared norms are 3 and 2. Thus the orthonormal columns are

$$\mathbf{u}_1 = \frac{1}{\sqrt{3}}(1, 1, 1)^T, \qquad \mathbf{u}_2 = \frac{1}{\sqrt{2}}(-1, 0, 1)^T.$$

Equivalently, the normal equations are diagonal:

$$A^T A \begin{pmatrix} c_0 \\ c_1 \end{pmatrix} = \begin{pmatrix} 3 & 0 \\ 0 & 2 \end{pmatrix} \begin{pmatrix} c_0 \\ c_1 \end{pmatrix} = A^T \mathbf{b} = \begin{pmatrix} 5 \\ 1 \end{pmatrix}.$$

Hence $c_0 = 5/3$ and $c_1 = 1/2$. The fitted data and residual are

$$\widehat{\mathbf{b}} = A\widehat{\mathbf{x}} = \begin{pmatrix} 7/6 \\ 5/3 \\ 13/6 \end{pmatrix}, \qquad \mathbf{r} = \mathbf{b} - \widehat{\mathbf{b}} = \begin{pmatrix} -1/6 \\ 1/3 \\ -1/6 \end{pmatrix}.$$

Finally $A^T \mathbf{r} = \mathbf{0}$, confirming that no infinitesimal change of intercept or slope can reduce the squared residual.

7. Bessel's Inequality and Parseval's Identity

For an orthonormal family $\{\mathbf{u}_j\}$ and $S_N \mathbf{v} = \sum_{j=1}^{N} \langle \mathbf{u}_j, \mathbf{v} \rangle \mathbf{u}_j$, orthogonality yields

$$\|\mathbf{v}\|^2 = \sum_{j=1}^{N} |\langle \mathbf{u}_j, \mathbf{v} \rangle|^2 + \|\mathbf{v} - S_N \mathbf{v}\|^2.$$

Hence

$$\boxed{\sum_{j=1}^{N} |\langle \mathbf{u}_j, \mathbf{v} \rangle|^2 \leq \|\mathbf{v}\|^2} \qquad \text{(Bessel)}.$$

For a complete orthonormal basis, the remainder tends to zero and

$$\boxed{\|\mathbf{v}\|^2 = \sum_{j} |\langle \mathbf{u}_j, \mathbf{v} \rangle|^2} \qquad \text{(Parseval)}.$$

The coefficients are components of the vector or function, and Parseval says that total squared norm (energy, in many physical interpretations) equals the sum of squared modal amplitudes. In Fourier analysis this becomes conservation of signal energy between physical and frequency representations.

8. Summary and Checks for Understanding

- Gram–Schmidt removes projections sequentially and preserves every intermediate span.
- Normalization converts an orthogonal basis into an orthonormal one.
- Projection is an orthogonal decomposition and the solution of a unique best-approximation problem.
- Least squares is projection onto a column space; QR computes it without forming the normal equations.

- Bessel bounds the captured modal energy; Parseval gives equality for a complete basis.

Self-study exercises

1. Recompute Example 3.1 without simplifying $\mathbf{w}_2$ before normalization and verify that the same $\mathbf{u}_2$ results.
2. Apply Gram–Schmidt to $(1,1,0)^T, (1,0,1)^T, (0,1,1)^T$ and construct the complete matrices Q and R; verify $Q^TQ = I$ and $QR = A$.
3. Project $(2,3,4)^T$ onto the xy-plane, identify the residual, and prove directly that no other point in that plane is closer.
4. Continue the polynomial calculation to x^3 and compare the result with $P_3(x) = \frac{1}{2}(5x^3 - 3x)$.