Math E-23c - Mathematics for Computation and Data Science

Proofs

Jimmy Tran[1]

Spring 2024

[1]Originally scribed by Chris Watson. Adapted with permission.

Preface

Math E–23c is the second half of a moderately rigorous sequence in linear algebra, real analysis, and multivariable (integral) calculus. It is aimed at students who want to *prove* the theorems they later *use*, yet care just as much about applying those results in computer science, statistics, data science, and economics. Topics span discrete mathematics, vector spaces, infinite series, the axiomatic foundations of probability, and integration, with hands-on work in the R programming language.
These notes serve as a reference for proofs that students in the course are expected to know how to recreate and teach to others.

Historical Note. Senior Lecturer Paul Bamberg designed *Math 23c* as an “alternative second half” to *Math 23b*. The course deliberately traded most of 23b’s differential-forms material for probability, statistics, and computation, thereby giving social-science, data-science, and industry-bound students a direct path from a proof-based 23a foundation to the mathematics they will actually use. For many years the Extension and College versions were cross-listed: the two registrations shared lectures, problem sets, and exams—the catalog number was the only difference. As of Spring 2024, Math E–23b’s differential-forms content (determinants, exterior derivative, Stokes’s theorem, etc.) was folded into a revamped *E–23c* syllabus. This course therefore preserves the full, proof-oriented curriculum while offering a single, streamlined pathway after Math E–23a.

Contents

Proof 1 (Orthogonal Eigenbasis Implies Symmetric Matrix)

Let A be an $n \times n$ matrix of real numbers, with a basis of real orthogonal eigenvectors. Then A is symmetric.

Proof

Since we have a basis of real eigenvectors, then we can diagonalize A as PDP^{-1}, where

$$P = \begin{bmatrix} | & | & & | \\ \vec{v}_1 & \vec{v}_2 & \cdots & \vec{v}_n \\ | & | & & | \end{bmatrix} \quad \text{and} \quad D = \begin{bmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & \lambda_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \lambda_n \end{bmatrix}$$

We can normalize each $\vec{v}_i$ by dividing by their lengths so that they are *unit vectors*. So we will work with *orthonormal* eigenvectors.

Consider P^TP:

$$P^TP = \begin{bmatrix} — & \vec{v}_1 & — \\ — & \vec{v}_2 & — \\ & \vdots & \\ — & \vec{v}_n & — \end{bmatrix} \begin{bmatrix} | & | & & | \\ \vec{v}_1 & \vec{v}_2 & \cdots & \vec{v}_n \\ | & | & & | \end{bmatrix} = \begin{bmatrix} \vec{v}_1 \cdot \vec{v}_1 & \vec{v}_1 \cdot \vec{v}_2 & \cdots & \vec{v}_1 \cdot \vec{v}_n \\ \vec{v}_2 \cdot \vec{v}_1 & \vec{v}_2 \cdot \vec{v}_2 & \cdots & \vec{v}_2 \cdot \vec{v}_n \\ \vdots & \vdots & \ddots & \vdots \\ \vec{v}_n \cdot \vec{v}_1 & \vec{v}_n \cdot \vec{v}_2 & \cdots & \vec{v}_n \cdot \vec{v}_n \end{bmatrix}$$

Since the vectors are orthonormal, $\vec{v}_i \cdot \vec{v}_i = 1$ and $\vec{v}_i \cdot \vec{v}_j = 0$ for $i \neq j$. So $P^TP = I_n$, which means $P^{-1} = P^T$.

So $A = PDP^{-1} = PDP^T$. We take the tranpose of both sides:

$$\begin{aligned} A^T &= \left(PDP^T\right)^T \\ &= \left(P^T\right)^T D^T P^T && \text{(Use } (AB)^T = B^TA^T\text{)} \\ &= PDP^T && \text{(Since } D^T = D\text{)} \\ &= A \end{aligned}$$

Therefore, A must be symmetric. ■

Proof 2 (Sum of Integrals)

Let f and g be two functions that are both integrable on $[a, b]$. Prove that their sum $f + g$ is integrable on $[a, b]$, and that the integral of the sum equals the sum of the integrals:

$$\int_a^b f + g = \int_a^b f + \int_a^b g$$

Proof

Part (a)

We first look in any subinterval S, $[t_{k-1}, t_k]$, of some partition P. In any S, the following two statements are true for any f and g for all $x \in S$:

$$\begin{aligned} f(x) &\geq m(f, S) && (m \text{ is the } \textit{infimum}) \\ g(x) &\geq m(g, S) \end{aligned}$$

We can add these inequalities so that the following is true for all $x \in S$.

$$\begin{aligned} f(x) + g(x) &\geq m(f, S) + m(g, S) \\ \iff (f + g)(x) &\geq m(f, S) + m(g, S) \\ \implies m(f + g, S) &\geq m(f, S) + m(g, S) \\ \Rightarrow m(f + g, S)(t_k - t_{k-1}) &\geq m(f, S)(t_k - t_{k-1}) + m(g, S)(t_k - t_{k-1}) \end{aligned}$$

When we sum over all subintervals S of partition P, the above inequality will be

$$L(f + g, P) \geq L(f, P) + L(g, P)$$

Similarly, with supremums we will get

$$U(f + g, P) \leq U(f, P) + U(g, P)$$

For any function and fixed partition, $L \leq U$, so after combining the above inequalities,

$$L(f, P) + L(g, P) \leq L(f + g, P) \leq U(f + g, P) \leq U(f, P) + U(g, P) \tag{1}$$

Consider any $\varepsilon > 0$. Since f is integrable, $\exists$ a partition P_1 such that $U(f, P_1) - L(f, P_1) < \frac{\varepsilon}{2}$. Since g is integrable, $\exists$ a partition P_2 such that $U(g, P_2) - L(g, P_2) < \frac{\varepsilon}{2}$.

Let $P = P_1 \cup P_2$. For a finer partition, upper sums decrease and lower sums increase. So $U(f, P) - L(f, P) < \frac{\varepsilon}{2}$ and $U(g, P) - L(g, P) < \frac{\varepsilon}{2}$. We combine these inequalities:

$$(U(f, P) + U(g, P)) - (L(f, P) + L(g, P)) < \varepsilon$$

We see from Equation (1) that the rightmost term minus the leftmost term can be made less than ε, so we then see that (because the two "inner terms" of the inequality are "squeezed")

$$U(f + g, P) - L(f + g, P) < \varepsilon$$

Therefore, $f + g$ is integrable.

Part (b)

Since $f+g$ is integrable, we insert the integral into the middle of Equation (1):

$$L(f,P)+L(g,P) \leq L(f+g,P) \leq \int f+g \leq U(f+g,P) \leq U(f,P)+U(g,P)$$

We can also assert that

$$L(f,P)+L(g,P) \leq \int f + \int g \leq U(f,P)+U(g,P)$$

Since the difference between the leftmost and rightmost terms can be made less than ε, $\forall \varepsilon > 0$, both $\int f+g$ and $\int f + \int g$ are squeezed between that difference. Therefore,

$$\int f+g = \int f + \int g$$

$\blacksquare$

Proof 3 (Improper Integral of e^{-x^2})

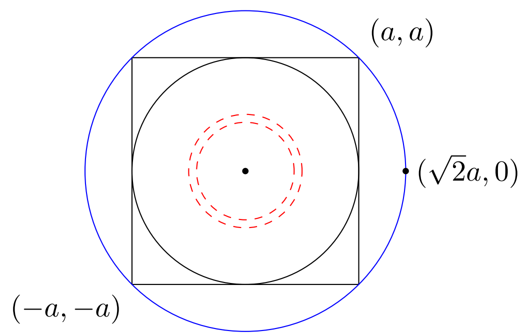
Define $I(a) = \int_{-a}^{a} e^{-x^2}\,\mathrm{d}x = \int_{-a}^{a} e^{-y^2}\,\mathrm{d}y$.
By considering integrals over circular and square regions, prove that

$$\pi\left(1 - e^{-a^2}\right) < I(a)^2 < \pi\left(1 - e^{-2a^2}\right)$$

and use a squeeze argument to show that

$$\int_{-\infty}^{\infty} e^{-x^2}\,\mathrm{d}x = \lim_{a\to\infty} I(a) = \sqrt{\pi}$$

Proof



We will consider integrals over circular and square regions. In the figure to the left, the point at the center is the origin. The black circle has radius a, the blue circle has radius $\sqrt{2}a$. We will call the square region S, the black circle C_1, and the blue circle C_2.

We start with the number $(I(a))^2$, which is an integral over the square S:

$$(I(a))^2 = \int_{-a}^{a} e^{-x^2}\,\mathrm{d}x \cdot \int_{-a}^{a} e^{-y^2}\,\mathrm{d}y = \iint_S e^{-x^2-y^2}\,\mathrm{d}x\,\mathrm{d}y$$

Instead, we will calculate the integral of the same function over the circle with radius a (C_1). We choose a partition of C_1 which will consist of concentric rings; for each ring, we take some function value multiplied by the area of each ring and add them. An arbitrary ring will have radius r and width Δr. For the function values, we look only at the inside of the ring: $f\begin{pmatrix} x \\ y \end{pmatrix} = e^{-(x^2+y^2)} = e^{-r^2}$. The area of each ring is approximately $2\pi r\Delta r$. When adding the rings, we calculate the following integral (substituting $u = r^2$):

$$\iint_{C_1} e^{-x^2-y^2}\,\mathrm{d}x\,\mathrm{d}y = \int_{r=0}^{a} e^{-r^2} 2\pi r\,\mathrm{d}r = \int_0^{a^2} e^{-u}\pi\,\mathrm{d}u = \pi\left(1 - e^{-a^2}\right)$$

Next we look at the larger circle C_2, which has radius $\sqrt{2}a$. By the same procedure as above (but with different bounds), we see that

$$\iint_{C_2} e^{-x^2-y^2}\,\mathrm{d}x\,\mathrm{d}y = \pi\left(1 - e^{-2a^2}\right)$$

We now have, for all a, an inequality relating the areas of the circles and the square:

$$\pi\left(1 - e^{-a^2}\right) \le (I(a))^2 \le \pi\left(1 - e^{-2a^2}\right)$$

In the limit as $a \to \infty$, both the LHS and RHS approach π, so $\lim_{a\to\infty} (I(a))^2 = \pi$ by the *squeeze lemma*, and $\lim_{a\to\infty} I(a) = \sqrt{\pi}$. This is equivalent to $\int_{-\infty}^{\infty} e^{-x^2}\,\mathrm{d}x = \sqrt{\pi}$. ∎

Proof 4 (Determinant of a Product)

Let A and B be $n \times n$ matrices, and $C = AB$. Prove that $\det C = \det A \det B$.

Proof

Consider C to be a matrix of column vectors: $C = \begin{bmatrix} | & | & & | \\ \vec{c}_1 & \vec{c}_2 & \cdots & \vec{c}_n \\ | & | & & | \end{bmatrix}$. An arbitrary column of C, by matrix multiplication, can be viewed as a linear combination of the columns of A: $\vec{c}_k = A\vec{b}_k = \displaystyle\sum_{i=1}^{n} \vec{a}_i b_{ik}$. The determinant of C is a function of all the columns:

$$\begin{aligned} \det C &= \det \begin{bmatrix} \vec{c}_1 & \vec{c}_2 & \cdots & \vec{c}_n \end{bmatrix} \\ &= \det \begin{bmatrix} \sum_{i_1=1}^{n} \vec{a}_{i_1} b_{i_1,1} & \sum_{i_2=1}^{n} \vec{a}_{i_2} b_{i_2,2} & \cdots & \sum_{i_n=1}^{n} \vec{a}_{i_n} b_{i_n,n} \end{bmatrix} \\ &= \sum_{i_1=1}^{n} b_{i_1,1} \det \begin{bmatrix} \vec{a}_{i_1} & \sum_{i_2=1}^{n} \vec{a}_{i_2} b_{i_2,2} & \cdots & \sum_{i_n=1}^{n} \vec{a}_{i_n} b_{i_n,n} \end{bmatrix} && \text{(Multilinearity)} \\ &= \sum_{i_1=1}^{n} \sum_{i_2=1}^{n} \cdots \sum_{i_n=1}^{n} b_{i_1,1} b_{i_2,2} \cdots b_{i_n,n} \det \begin{bmatrix} \vec{a}_{i_1} & \vec{a}_{i_2} & \cdots & \vec{a}_{i_n} \end{bmatrix} && \text{(Multilinearity } n-1 \text{ times)} \end{aligned}$$

There are n^n terms in the above expression, but if any indices are repeated there will be repeated columns, so the determinant for those terms equals 0 (by *anti-symmetry*). The only nonzero terms will be when the indices are a permutation of n (of which there are $n!$), so we write this as

$$\det C = \sum_{\substack{i_1, i_2, \ldots, i_n \\ \in \mathrm{Perm}_n}} b_{i_1,1} b_{i_2,2} \cdots b_{i_n,n} \det \begin{bmatrix} \vec{a}_{i_1} & \vec{a}_{i_2} & \cdots & \vec{a}_{i_n} \end{bmatrix}$$

We can swap the columns $\vec{a}_{i_j}$ to place them in order, so for some permutation σ, we write the above as

$$\det C = \sum_{\substack{i_1, i_2, \ldots, i_n \\ \in \mathrm{Perm}_n}} b_{i_1,1} b_{i_2,2} \cdots b_{i_n,n} \operatorname{sgn}(\sigma) \det \begin{bmatrix} \vec{a}_1 & \vec{a}_2 & \cdots & \vec{a}_n \end{bmatrix}$$

The latter term is the determinant of A, which does not depend on the indices, so we can take it out of the sum:

$$\det C = \det A \cdot \sum_{\substack{i_1, i_2, \ldots, i_n \\ \in \mathrm{Perm}_n}} b_{i_1,1} b_{i_2,2} \cdots b_{i_n,n} \operatorname{sgn}(\sigma) \tag{2}$$

This formula works for any matrix A; in particular, for $A = I_n$. So $\det A = \det I_n = 1$ (by the *normalization* property). Then $C = AB = B$, so $\det B = \det C$, and substituting into Equation (2) results in

$$\det B = \sum_{\substack{i_1, i_2, \ldots, i_n \\ \in \mathrm{Perm}_n}} b_{i_1,1} b_{i_2,2} \cdots b_{i_n,n} \operatorname{sgn}(\sigma)$$

Therefore, from Equation (2) again we have $\det C = \det A \det B$. ∎

Proof 5 (Linear Change of Variables)

Suppose $T : \mathbb{R}^n \to \mathbb{R}^n$ is an invertible linear transformation, and $f : \mathbb{R}^n \to \mathbb{R}$ is integrable. Prove that $f \circ T$ is also integrable, and that

$$\int_{\mathbb{R}^n} f(\vec{y})|\mathrm{d}^n\vec{y}| = |\det T| \int_{\mathbb{R}^n} f(T(\vec{x}))|\mathrm{d}^n\vec{x}|$$

Proof

In parameter space, consider a dyadic partition D_N made up of dyadic n-cubes C. These cubes map by T to n-parallelograms $T(C)$ that form a partition of original space $T(D_N)$.

The integral (if it exists) is a limit of upper Riemann sums in parameter space:

$$\begin{aligned}
\int_{\mathbb{R}^n} f(T(\vec{x}))|\det T||\mathrm{d}^n\vec{x}| &= \lim_{N\to\infty} U\left(f \circ T, D_N\right)|\det T| \\
&= \lim_{N\to\infty} \sum_{C \in D_N} M_C\left(f \circ T\right)|\det T| \operatorname{vol}_n C \\
&= \lim_{N\to\infty} \sum_{C \in D_N} M_C\left(f \circ T\right) \operatorname{vol}_n T(C)
\end{aligned}$$

(The third line comes from Module 4, in which we showed that T volume-stretches an n-cube by a factor of $|\det T|$: $\mathrm{vol}_n T(C) = |\det T| \operatorname{vol}_n C$.)

We can express the supremum in the above equation differently:

$$\begin{aligned}
M_C(f \circ T) &= \sup\{f(T(\vec{x})) \mid \vec{x} \in C\} \\
&= \sup\{f(T(\vec{x})) \mid T(\vec{x}) \in T(C)\} \\
&= \sup\{f(\vec{y}) \mid \vec{y} \in T(C)\} \\
&= M_{T(C)} f
\end{aligned}$$

Then, continuing from the first set of equalities,

$$\begin{aligned}
&= \lim_{N\to\infty} \sum_{C \in D_N} M_{T(C)}(f) \operatorname{vol}_n T(C) \\
&= \lim_{N\to\infty} \sum_{T(C) \in T(D_N)} M_{T(C)}(f) \operatorname{vol}_n T(C) \\
&= \lim_{N\to\infty} U(f, T(D_N)) \\
&= \int_{\mathbb{R}^n} f(\vec{y})|\mathrm{d}^n\vec{y}| \qquad \text{(Since } f \text{ is integrable)}
\end{aligned}$$

■

Proof 6 (The Beta Distribution)

The *beta distribution* is defined as having a probability density function

$$\mu(x) = \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} x^{\alpha-1}(1-x)^{\beta-1} \qquad 0 \le x \le 1$$

Use the "taxi-cab" change of variables formula to prove that this distribution is properly normalized:

$$\int_0^1 \mu(x)\,\mathrm{d}x = 1$$

You may use the definition of the Gamma function as

$$\Gamma(\alpha) = \int_0^\infty x^{\alpha-1}e^{-x}\,\mathrm{d}x$$

Proof

Taxi-cab coordinates are the following, which we will invert:

$$\begin{cases} u = x + y \\ v = \frac{y}{x+y} \end{cases}$$

Multiplying both together gives $uv = y$, so $x = u - y = u(1-v)$. Then

$$\begin{pmatrix} x \\ y \end{pmatrix} = \Phi \begin{pmatrix} u \\ v \end{pmatrix} = \begin{pmatrix} u(1-v) \\ uv \end{pmatrix}$$

The Jacobian is simply $\left[D\Phi\right] = \begin{bmatrix} 1-v & -u \\ v & u \end{bmatrix}$, and $\det\left[D\Phi\right] = u$. By our definition, we can see that $0 \le u \le \infty$ and $0 \le v \le 1$, which we will use to change the bounds of integration.

Consider

$$\begin{aligned}
\Gamma(\alpha)\Gamma(\beta) &= \int_0^\infty y^{\alpha-1}e^{-y}\,\mathrm{d}y \int_0^\infty x^{\beta-1}e^{-x}\,\mathrm{d}x \\
&= \int_0^\infty \int_0^\infty y^{\alpha-1}x^{\beta-1}e^{-(x+y)}\,\mathrm{d}x\,\mathrm{d}y \\
&= \int_{u=0}^\infty \int_{v=0}^1 u^{\alpha-1}v^{\alpha-1}u^{\beta-1}(1-v)^{\beta-1}e^{-u}u\,\mathrm{d}v\,\mathrm{d}u && (\det\left[D\Phi\right] = u) \\
&= \int_0^\infty u^{\alpha+\beta-1}e^{-u}\,\mathrm{d}u \int_0^1 v^{\alpha-1}(1-v)^{\beta-1}\,\mathrm{d}v \\
&= \Gamma(\alpha+\beta)\int_0^1 v^{\alpha-1}(1-v)^{\beta-1}\,\mathrm{d}v \\
\implies 1 &= \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)}\int_0^1 x^{\alpha-1}(1-x)^{\beta-1}\,\mathrm{d}x && (\text{Dividing both sides by } \Gamma(\alpha)\Gamma(\beta)) \\
\iff 1 &= \int_0^1 \mu(x)\,\mathrm{d}x && \blacksquare
\end{aligned}$$

Proof 7 (Different Bases Have the Same Number of Elements)

Prove that any two bases of an (abstract) vector space V have the same number of elements (using the fact that if a matrix is invertible, then it is square).

Proof

Consider a basis $\mathbf{v} = \{\vec{v}_1, \dots, \vec{v}_k\}$ with k elements, and $\mathbf{w} = \{\vec{w}_1, \dots, \vec{w}_p\}$ with p elements.
There are "concrete-to-abstract" functions such that

$$\mathbb{R}^k \xrightarrow{\Phi_v} V \xleftarrow{\Phi_w} \mathbb{R}^p$$

The following are true for Φ_w as well as Φ_v:

First, Φ_v is one-to-one (because basis vectors are linearly independent): for $\vec{x} \in V$, if $\vec{x} = \Phi_v(\vec{a}) = \Phi_v(\vec{b})$, then $\vec{x} = \sum_{i=1}^k a_i\vec{v}_i = \sum_{i=1}^k b_i\vec{v}_i$, and $\sum_{i=1}^k (a_i - b_i)\vec{v}_i = \vec{0}$. So $\forall i,\ a_i = b_i$, and $\vec{a} = \vec{b}$.

Second, Φ_v is onto (because basis vectors are spanning): for all $\vec{x} \in V,\ \exists a_1, \dots, a_k : \vec{x} = \sum_{i=1}^k a_i\vec{v}_i$. This is the same as writing $\vec{x} = \Phi_v(\vec{a})$.

So, Φ_v is one-to-one and onto, and thus invertible; that is, Φ_v^{-1} exists.

So there are functions such that

$$\mathbb{R}^k \xleftarrow{\Phi_v^{-1}} V \xleftarrow{\Phi_w} \mathbb{R}^p$$

The composition $\Phi_v^{-1} \circ \Phi_w$ takes a concrete vector in $\mathbb{R}^p$ to a concrete vector in $\mathbb{R}^k$. Therefore, it has a matrix, which will have size $k \times p$.

Also, the composition of two invertible functions is invertible, so the $k \times p$ matrix is invertible. Since only square matrices can be invertible, then $k = p$. Therefore, the bases have the same number of elements. ■

Proof 8 (Lebesgue Integral)

Define the Lebesgue integral and illustrate the definition $\int_0^\infty e^{-x}\,\mathrm{d}x$ as a Lebesgue integral.

Proof

First write $f(t) = e^{-t},\ t \in [0, \infty)$ as a series of functions:

$$\sum_{k=0}^{\infty} f_k(t) \quad \text{where} \quad f_k(t) = \begin{cases} e^{-t} & k \le t \le k+1 \\ 0 & \text{otherwise} \end{cases}$$

First, test for Lebesgue integrability: calculate

$$I_k = \int_0^\infty |f_k(t)|\,\mathrm{d}t = \int_k^{k+1} e^{-t}\,\mathrm{d}t$$

Does $\sum_{k=0}^\infty I_k$ converge? Each $I_k \le e^{-k}$ (i.e., the area under each $f_k(t)$ is less than the area of a rectangle), so we can use the *comparison test* for series:

$$\sum_{k=0}^{\infty} e^{-k} = \sum_{k=0}^{\infty} \left(\frac{1}{e}\right)^k \quad \text{converges}$$

By the comparison test, then $\sum_{k=0}^\infty I_k$ converges, so the Lebesgue integral exists, and

$$\begin{aligned}
\int_0^\infty f(t)\,\mathrm{d}t &= \sum_{k=0}^{\infty} \int_0^\infty f_k(t)\,\mathrm{d}t \\
&= \sum_{k=0}^{\infty} \int_k^{k+1} e^{-t}\,\mathrm{d}t \\
&= \sum_{k=0}^{\infty} -e^{-t}\Big|_k^{k+1} \\
&= \sum_{k=0}^{\infty} e^{-k} - e^{-(k+1)} \\
&= \sum_{k=0}^{\infty} (1 - e^{-1}) e^{-k} \\
&= (1 - e^{-1}) \sum_{k=0}^{\infty} \left(\frac{1}{e}\right)^k \\
&= (1 - e^{-1}) \frac{1}{1 - \frac{1}{e}} \\
&= 1
\end{aligned}$$

■

Proof 9 (Projection Matrices)

Let M be a k-dimensional subspace of $\mathbb{R}^n$, and let A be a $n \times k$ matrix whose columns are a basis for M. Prove that $P = A(A^TA)^{-1}A^T$ is the projection matrix by proving that

(a) $\forall \vec{m} \in M,\ P\vec{m} = \vec{m}$

(b) $\forall \vec{w} \in \mathbb{R}^n,\ P\vec{w} \in M$

(c) $\forall \vec{w} \in \mathbb{R}^n,\ \vec{w} - P\vec{w} \perp M$

Proof

Part (a)

Since the columns of A span M, then $\forall \vec{m} \in M,\ \exists \vec{x} : \vec{m} = A\vec{x}$. Then $P\vec{m} = A(A^TA)^{-1}A^TA\vec{x} = AI_k\vec{x} = A\vec{x} = \vec{m}$.

Part (b)

$\forall \vec{w} \in \mathbb{R}^n,\ P\vec{w} = A(A^TA)^{-1}A^T\vec{w} = A\vec{x}$ for some $\vec{x}$; this latter product is a linear combination of the columns of A. Since the columns of A span M, then the product will be in M, which proves that $P\vec{w} \in M$.

Part (c)

Consider any $\vec{w} \in \mathbb{R}^n$, and let $P\vec{w} = \vec{m}_0 \in M$ (by part (b)). Let $\vec{v} = \vec{w} - P\vec{w} = \vec{w} - \vec{m}_0$.
We want to show that $\vec{v}$ is orthogonal to all $\vec{m} \in M$, which is equivalent to showing that $\vec{v}$ is orthogonal to all the columns of A. This is also equivalent to showing that $\vec{v} \in \ker A^T$. Then

$$\begin{aligned}
A^T\vec{v} &= A^T\vec{w} - A^TP\vec{w} \\
&= A^T\vec{w} - A^TA(A^TA)^{-1}A^T\vec{w} \\
&= A^T\vec{w} - A^T\vec{w} \\
&= \vec{0}
\end{aligned}$$

So $\vec{v} \in \ker A^T$, which means $\vec{v} \perp M$. ■

Proof 10 (SVD)

Any $n \times m$ matrix A can be written as $A = PDQ^T$ where

- P is a $n \times n$ orthonormal matrix
- D is a $n \times m$ matrix ("nonnegative rectangular diagonal") with $d_{i,j} = \begin{cases} 0 & i \neq j \\ \geq 0 & i = j \end{cases}$
- Q is a $m \times m$ orthonormal matrix

Proof

Part (a)

We want to show that $A^T A$ has nonnegative eigenvalues:

$$\begin{aligned} A^T A\vec{v} &= \lambda\vec{v} \\ \vec{v}^T A^T A\vec{v} &= \lambda\vec{v}^T\vec{v} \\ (A\vec{v})^T A\vec{v} &= \lambda\vec{v}^T\vec{v} \\ (A\vec{v}) \cdot (A\vec{v}) &= \lambda\vec{v} \cdot \vec{v} \\ \|A\vec{v}\|^2 &= \lambda\,\|\vec{v}\|^2 \end{aligned} \qquad (\lambda \text{ must be nonnegative})$$

Part (b)

Let $\vec{v}$ be an eigenvector of $A^T A$ (with $\lambda > 0$); then $A^T A\vec{v} = \lambda\vec{v}$. We want to show that $\vec{w} = \frac{1}{\sqrt{\lambda}}A\vec{v}$ is an eigenvector of AA^T.

$$\begin{aligned} AA^T\vec{w} &= AA^T\frac{1}{\sqrt{\lambda}}A\vec{v} \\ &= \frac{1}{\sqrt{\lambda}}AA^TA\vec{v} \\ &= \frac{1}{\sqrt{\lambda}}A\lambda\vec{v} \\ &= \lambda \cdot \frac{1}{\sqrt{\lambda}}A\vec{v} \\ &= \lambda\vec{w} \end{aligned}$$

Part (c)

(Q): Since $A^T A$ is a symmetric $m \times m$ matrix, it has m eigenvectors, which form the columns of Q. We know by the *Spectral Theorem* that the eigenvectors are orthogonal, and we normalize them so Q is orthonormal ($Q^T = Q^{-1}$).

(P): Consider each of the vectors $\vec{w}$ from the previous step; these are also orthonormal. If there are less than n of them, we create additional orthonormal vectors to complete a basis for $\mathbb{R}^n$. These vectors will form the columns of P.

(D): Take any non-zero eigenvalues λ_i of $A^T A$ and set $d_{i,i} = \sqrt{\lambda_i} = \sigma_i$. We set all other $d_{i,j} = 0$.

Part (d)

We want to show that $A = PDQ^T$, so we will show $AQ = PD$.

$$A\begin{bmatrix}\underbrace{\vec{v}_1 \ \cdots \ \vec{v}_k}_{\lambda>0} & \underbrace{\vec{v}_{k+1} \ \cdots \ \vec{v}_m}_{\lambda=0}\end{bmatrix} = \begin{bmatrix}\underbrace{\vec{w}_1 \ \cdots \ \vec{w}_k}_{\frac{1}{\sqrt{\lambda}}A\vec{v}} & \vec{w}_{k+1} & \cdots & \vec{w}_n\end{bmatrix}\begin{bmatrix}\sqrt{\lambda_1} & 0 & \cdots & & 0\\ 0 & \sqrt{\lambda_2} & & & 0\\ \vdots & & \ddots & & \vdots\\ 0 & 0 & \cdots & \sqrt{\lambda_k} & 0\\ 0 & 0 & \cdots & 0 & 0\end{bmatrix}$$

On the LHS, for the first k columns the product will be $A\vec{v}_i$. On the RHS, the first k columns will be of the form $P \cdot \begin{bmatrix}0\\ \vdots\\ \sqrt{\lambda_i}\\ 0\\ \vdots\\ 0\end{bmatrix} = \vec{w}_i\sqrt{\lambda_i}$.

For columns $k+1$ through m, on the LHS the product will be of the form $A\vec{v}_i$.
On the RHS, they will all equal the 0 vector, so we want to show that $A\vec{v}_i = \vec{0}$.

We know that $A^TA\vec{v}_i = 0 \cdot \vec{v}_i = \vec{0}$, because they are eigenvectors with eigenvalue 0. Let $A\vec{v}_i = \vec{y}$ be a linear combination of the columns of A. Then $A^T\vec{y} = \vec{0}$, which means $\vec{y}$ is orthogonal to the columns of A. The only vector that is both a linear combination of and orthogonal to a set of vectors is the 0 vector; i.e., $\vec{y} = A\vec{v}_i = \vec{0}$. ■

Proof 11 (Elementary 2-forms)

Suppose that ϕ is an arbitrary element of $A^2(\mathbb{R}^3)$. All we know about it is that it is multilinear and alternating. Prove that

$$\phi = \sum_{1 \le i_1 < i_2 \le 3} a_{i_1,i_2} dx_{i_1} \wedge dx_{i_2}$$

where $a_{i_1,i_2} = \phi(\vec{e}_{i_1}, \vec{e}_{i_2})$.

Proof

Let $\vec{v}_1 = \begin{bmatrix} a \\ b \\ c \end{bmatrix} = a\vec{e}_1 + b\vec{e}_2 + c\vec{e}_3$ and $\vec{v}_2 = \begin{bmatrix} u \\ v \\ w \end{bmatrix} = u\vec{e}_1 + v\vec{e}_2 + w\vec{e}_3$.
We have

$$\begin{aligned} \varphi(\vec{v}_1, \vec{v}_2) &= \varphi(a\vec{e}_1 + b\vec{e}_2 + c\vec{e}_3, \vec{v}_2) \\ &= a\varphi(\vec{e}_1, \vec{v}_2) + b\varphi(\vec{e}_2, \vec{v}_2) + c\varphi(\vec{e}_3, \vec{v}_2) && \text{(Linearity)} \end{aligned}$$

φ is also linear in its second input; we look at the first term above:

$$\begin{aligned} \varphi(\vec{e}_1, \vec{v}_2) &= \varphi(\vec{e}_1, u\vec{e}_1 + v\vec{e}_2 + w\vec{e}_3) \\ &= u\varphi(\vec{e}_1, \vec{e}_1) + v\varphi(\vec{e}_1, \vec{e}_2) + w\varphi(\vec{e}_1, \vec{e}_3) && \text{(Linearity)} \end{aligned}$$

φ is also alternating, so $\varphi(\vec{e}_1, \vec{e}_2) = -\varphi(\vec{e}_2, \vec{e}_1)$ and $\varphi(\vec{e}_1, \vec{e}_1) = 0$.
We expand and regroup from above:

$$\begin{aligned} \varphi(\vec{v}_1, \vec{v}_2) &= av\varphi(\vec{e}_1, \vec{e}_2) + bu\varphi(\vec{e}_2, \vec{e}_1) \\ &\quad + aw\varphi(\vec{e}_1, \vec{e}_3) + cu\varphi(\vec{e}_3, \vec{e}_1) \\ &\quad + bw\varphi(\vec{e}_2, \vec{e}_3) + cv\varphi(\vec{e}_3, \vec{e}_2) \\ &= (av - bu)\varphi(\vec{e}_1, \vec{e}_2) + (aw - cu)\varphi(\vec{e}_1, \vec{e}_3) + (bw - cv)\varphi(\vec{e}_2, \vec{e}_3) \\ &= \varphi(\vec{e}_1, \vec{e}_2) \det \begin{bmatrix} a & u \\ b & v \end{bmatrix} + \varphi(\vec{e}_1, \vec{e}_3) \det \begin{bmatrix} a & u \\ c & w \end{bmatrix} + \varphi(\vec{e}_2, \vec{e}_3) \det \begin{bmatrix} b & v \\ c & w \end{bmatrix} \\ &= \varphi(\vec{e}_1, \vec{e}_2)\, dx \wedge dy(\vec{v}_1, \vec{v}_2) + \varphi(\vec{e}_1, \vec{e}_3)\, dx \wedge dz(\vec{v}_1, \vec{v}_2) + \varphi(\vec{e}_2, \vec{e}_3)\, dy \wedge dz(\vec{v}_1, \vec{v}_2) \end{aligned}$$

So we have written the 2-form φ as a linear combination of what φ does on the basis vectors times the basis 2-forms. The general differential form is

$$\varphi = \varphi(\vec{e}_1, \vec{e}_2)\, dx \wedge dy + \varphi(\vec{e}_1, \vec{e}_3)\, dx \wedge dz + \varphi(\vec{e}_2, \vec{e}_3)\, dy \wedge dz$$

Proof 12 (Exterior Derivative)

Let $\varphi = f\begin{pmatrix} x \\ y \end{pmatrix} dx + g\begin{pmatrix} x \\ y \end{pmatrix} dy$ be a smooth differential 1-form on $\mathbb{R}^2$.
State the definition of

$$\mathbf{d}\varphi(P_{\mathbf{x}}(\vec{e}_1, \vec{e}_2))$$

and use it to show that

$$\mathbf{d}\varphi = (D_x g - D_y f) dx \wedge dy$$

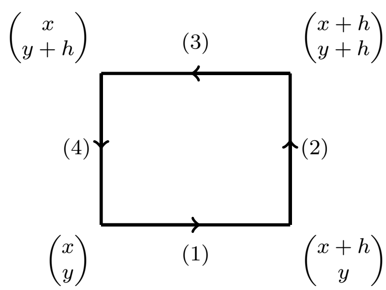
Proof

Since φ is a 1-form, $\mathbf{d}\varphi$ will be a 2-form. It is fully defined if we know what it does on $\vec{e}_1, \vec{e}_2$:

$$\mathbf{d}\varphi = a_{1,2} dx \wedge dy$$

To find $a_{1,2}$, we calculate the derivative on some parallelogram anchored at a point $\mathbf{x}$:

$$\mathbf{d}\varphi\left(P_{\mathbf{x}}(\vec{e}_1, \vec{e}_2)\right) = \lim_{h \to 0} \frac{1}{h^2} \int_{\partial P_{\mathbf{x}}(h\vec{e}_1, h\vec{e}_2)} \varphi$$



To integrate over the boundary, we have to integrate over the four segments of a parallelogram with proper orientation.
Here, positive orientation will be counterclockwise (see figure to the left).
We number each of the segments of the boundary over which we will parametrize.

For the first segment, we parametrize as $\gamma(t) = \begin{pmatrix} x + ht \\ y \end{pmatrix}$ for $0 \le t \le 1$. Then its Jacobian is $[D\gamma] = \begin{bmatrix} h \\ 0 \end{bmatrix}$. We let the differential form act on the Jacobian: $dx(D\gamma) = h$ and $dy(D\gamma) = 0$.
So, since $\varphi = f\begin{pmatrix} x \\ y \end{pmatrix} dx + g\begin{pmatrix} x \\ y \end{pmatrix} dy$, the contribution to the exterior derivative will be

$$\frac{1}{h^2} \int_{t=0}^{1} f\begin{pmatrix} x + ht \\ y \end{pmatrix} h \, \mathrm{d}t \tag{S1}$$

For side 3, $\gamma(t) = \begin{pmatrix} x + ht \\ y + h \end{pmatrix}$ for t from 1 to 0. The Jacobian will be $[D\gamma] = \begin{bmatrix} h \\ 0 \end{bmatrix}$, and side 3's contribution will be

$$\frac{1}{h^2} \int_{t=1}^{0} f\begin{pmatrix} x + ht \\ y + h \end{pmatrix} h \, \mathrm{d}t \tag{S3}$$

To combine the contributions from sides 1 and 3, we flip the sign for side 3 and combine (factoring out h):

$$-\frac{1}{h} \int_0^1 \left[f\begin{pmatrix} x + ht \\ y + h \end{pmatrix} - f\begin{pmatrix} x + ht \\ y \end{pmatrix} \right] \mathrm{d}t$$

By the *Mean Value Theorem*, $\exists s \in [0,1]$ such that the above can be written as

$$-\frac{1}{h}\int_0^1 hD_y f\begin{pmatrix} x+ht \\ y+hs \end{pmatrix} \mathrm{d}t = -\int_0^1 D_y f\begin{pmatrix} x+ht \\ y+hs \end{pmatrix} \mathrm{d}t$$

Taking the limit as $h \to 0$, we get $-D_y f\begin{pmatrix} x \\ y \end{pmatrix}$ as the contribution from sides 1 and 3.

For side 2, we have $\gamma(t) = \begin{pmatrix} x+h \\ y+ht \end{pmatrix}$ for $0 \le t \le 1$. Its Jacobian is $[D\gamma] = \begin{bmatrix} 0 \\ h \end{bmatrix}$, and side 2's contribution will be

$$\frac{1}{h^2}\int_{t=0}^1 g\begin{pmatrix} x+h \\ y+ht \end{pmatrix} h\,\mathrm{d}t \tag{S2}$$

For side 4, we have $\gamma(t) = \begin{pmatrix} x \\ y+ht \end{pmatrix}$ with t going from 1 to 0. The Jacobian is again $[D\gamma] = \begin{bmatrix} 0 \\ h \end{bmatrix}$, and side 4's contribution will be

$$\frac{1}{h^2}\int_{t=1}^0 g\begin{pmatrix} x \\ y+ht \end{pmatrix} h\,\mathrm{d}t \tag{S4}$$

We combine and factor again, and as above use the *Mean Value Theorem* to get:

$$\frac{1}{h}\int_{t=0}^1 \left[g\begin{pmatrix} x+h \\ y+ht \end{pmatrix} - g\begin{pmatrix} x \\ y+ht \end{pmatrix}\right] \mathrm{d}t = \int_0^1 D_x g\begin{pmatrix} x+hs \\ y+ht \end{pmatrix} \mathrm{d}t$$

Taking the limit as $h \to 0$, we are left with $D_x g\begin{pmatrix} x \\ y \end{pmatrix}$.

Overall, we have that

$$\mathbf{d}\varphi\left(P_{\mathbf{x}}(\vec{e}_1, \vec{e}_2)\right) = D_x g\begin{pmatrix} x \\ y \end{pmatrix} - D_y f\begin{pmatrix} x \\ y \end{pmatrix}$$

So

$$\mathbf{d}\varphi = (D_x g - D_y f)dx \wedge dy$$

∎