Linear Algebra (part 4) : Inner Products (by Evan Dummit, 2021, v. 3.00)

Contents

4 Inner Products **1**

4.1 Inner Products 1

4.1.1 Inner Products and Norms 1

4.1.2 The Cauchy-Schwarz Inequality and Applications 3

4.2 Orthogonality 5

4.2.1 Orthogonality and Orthonormal Sets 6

4.2.2 The Gram-Schmidt Procedure, QR Factorization 7

4.2.3 Orthogonal Complements 11

4.2.4 Orthogonal Projection 12

4.3 Applications of Inner Products 14

4.3.1 Least-Squares Estimates 14

4.3.2 Fourier Series 19

4 Inner Products

In this chapter we discuss vector spaces having an additional kind of structure called an "inner product" that generalizes the idea of the dot product of vectors in $\mathbb{R}^n$. We then use inner products to formulate notions of "length" and "angle" in more general vector spaces, and discuss in particular the central concept of orthogonality. We then discuss a few important applications of these ideas to least-squares problems and Fourier series.

4.1 Inner Products

- Recall that if $\mathbf{v}$ is a vector in $\mathbb{R}^n$, then the dot product $\mathbf{v} \cdot \mathbf{v} = ||\mathbf{v}||^2$ is the square of the length of $\mathbf{v}$, and that for any vectors $\mathbf{v}$ and $\mathbf{w}$, the angle θ between them satisfies the relation $\mathbf{v} \cdot \mathbf{w} = ||\mathbf{v}|| \, ||\mathbf{w}|| \cos\theta$.
  - Our goal is to define an abstract version of the dot product in a general vector space, and then use it to discuss the notions of length and angle.

4.1.1 Inner Products and Norms

- Definition: If V is a (real) vector space[1], an inner product on V is a pairing that assigns a real number to each ordered pair $(\mathbf{v}, \mathbf{w})$ of vectors in V. This pairing is denoted[2] $\langle \mathbf{v}, \mathbf{w} \rangle$ and must satisfy the following properties:

  [I1] Linearity in the first argument: $\langle \mathbf{v}_1 + c\mathbf{v}_2, \mathbf{w} \rangle = \langle \mathbf{v}_1, \mathbf{w} \rangle + c\langle \mathbf{v}_2, \mathbf{w} \rangle$ for any scalar c.

  [I2] Symmetry: $\langle \mathbf{v}, \mathbf{w} \rangle = \langle \mathbf{w}, \mathbf{v} \rangle$.

  [I3] Positive-definiteness: $\langle \mathbf{v}, \mathbf{v} \rangle \geq 0$ for all $\mathbf{v}$, and $\langle \mathbf{v}, \mathbf{v} \rangle = 0$ only when $\mathbf{v} = \mathbf{0}$.

[1] There is also a notion of an inner product on a complex vector space, whose scalars are the complex numbers. The definition is almost identical, except with condition [I2] replaced with "conjugate-symmetry": $\langle \mathbf{v}, \mathbf{w} \rangle = \overline{\langle \mathbf{w}, \mathbf{v} \rangle}$, rather than $\langle \mathbf{v}, \mathbf{w} \rangle = \langle \mathbf{w}, \mathbf{v} \rangle$.

[2] In this chapter, we will use angle brackets to denote inner products, and represent vectors in $\mathbb{R}^n$ using parentheses.

- The linearity and symmetry properties are fairly clear: if we fix the second component, the inner product behaves like a linear function in the first component, and we want both components to behave in the same way.
- The positive-definiteness property is intended to capture an idea about "length": namely, the length of a vector $\mathbf{v}$ should be the (inner) product of $\mathbf{v}$ with itself, and lengths are supposed to be nonnegative. Furthermore, the only vector of length zero should be the zero vector.

- Definition: A vector space V together with an inner product $\langle \cdot, \cdot \rangle$ on V is called an inner product space.
  - Any given vector space may have many different inner products.
  - When we say "suppose V is an inner product space", we intend this to mean that V is equipped with a particular (fixed) inner product.

- The entire purpose of defining an inner product is to generalize the notion of the dot product to more general vector spaces, so we first observe that the dot product on $\mathbb{R}^n$ is actually an inner product:

- Example: Show that the standard dot product on $\mathbb{R}^n$, defined as $(x_1, \ldots, x_n) \cdot (y_1, \ldots, y_n) = x_1 y_1 + \cdots + x_n y_n$ is an inner product.
  - [I1]-[I2]: It is an easy algebraic computation to verify the linearity and symmetry properties.
  - [I3]: If $\mathbf{v} = (x_1, \ldots, x_n)$ then $\mathbf{v} \cdot \mathbf{v} = x_1^2 + x_2^2 + \cdots + x_n^2$. Since each square is nonnegative, $\mathbf{v} \cdot \mathbf{v} \geq 0$, and $\mathbf{v} \cdot \mathbf{v} = 0$ only when all of the components of $\mathbf{v}$ are zero.

- There are other examples of inner products on $\mathbb{R}^n$ beyond the standard dot product.

- Example: Show that the pairing $\langle (x_1, y_1), (x_2, y_2) \rangle = 3x_1x_2 + 2x_1y_2 + 2x_2y_1 + 4y_1y_2$ on $\mathbb{R}^2$ is an inner product.
  - [I1]-[I2]: It is an easy algebraic computation to verify the linearity and symmetry properties.
  - [I3]: We have $\langle (x, y), (x, y) \rangle = 3x^2 + 4xy + 4y^2 = 2x^2 + (x + 2y)^2$, and since each square is nonnegative, the inner product is always nonnegative. Furthermore, it equals zero only when both squares are zero, and this clearly only occurs for $x = y = 0$.

- An important class of inner products are ones defined on function spaces:

- Example: Let V be the vector space of continuous (real-valued) functions on the interval $[a, b]$. Show that $\langle f, g \rangle = \int_a^b f(x)g(x)\,dx$ is an inner product on V.
  - [I1]: We have $\langle f_1 + cf_2, g \rangle = \int_a^b [f_1(x) + cf_2(x)]\, g(x)\,dx = \int_a^b f_1(x)g(x)\,dx + c\int_a^b f_2(x)g(x)\,dx = \langle f_1, g \rangle + c\,\langle f_2, g \rangle$.
  - [I2]: Observe that $\langle g, f \rangle = \int_a^b g(x)f(x)\,dx = \int_a^b f(x)g(x)\,dx = \langle f, g \rangle$.
  - [I3]: Notice that $\langle f, f \rangle = \int_a^b f(x)^2\,dx$ is the integral of a nonnegative function, so it is always nonnegative. Furthermore (since f is assumed to be continuous) the integral of f^2 cannot be zero unless f is identically zero.
  - Remark: More generally, if $w(x)$ is any fixed positive ("weight") function that is continuous on $[a, b]$, $\langle f, g \rangle = \int_a^b f(x)g(x) \cdot w(x)\,dx$ is an inner product on V.

- Here is another example of an inner product, on the space of matrices:

- Example: Let $V = M_{n \times n}$ be the vector space of $n \times n$ matrices. Show that $\langle A, B \rangle = \mathrm{tr}(AB^T)$ is an inner product on V.
  - [I1]: We have $\langle A + cC, B \rangle = \mathrm{tr}[(A + cC)B^T] = \mathrm{tr}[AB^T + cCB^T] = \mathrm{tr}(AB^T) + c\,\mathrm{tr}(CB^T) = \langle A, B \rangle + c\,\langle C, B \rangle$, where we used the facts that $\mathrm{tr}(M + N) = \mathrm{tr}(M) + \mathrm{tr}(M)$ and $\mathrm{tr}(cM) = c\,\mathrm{tr}(M)$.
  - [I2]: Observe that $\langle B, A \rangle = \mathrm{tr}(BA^T) = \mathrm{tr}(B^T A) = \mathrm{tr}(AB^T) = \langle A, B \rangle$, where we used the fact that $\mathrm{tr}(MN) = \mathrm{tr}(NM)$, which can be verified by expanding both sides explicitly.

- [I3]: We have $\langle A, A\rangle = \sum_{j=1}^{n}(AA^T)_{j,j} = \sum_{j=1}^{n}\sum_{k=1}^{n} A_{j,k}A^T_{k,j} = \sum_{j=1}^{n}\sum_{k=1}^{n} A^2_{j,k} \geq 0$, and equality can only occur when each element of A is zero, since squares are always nonnegative.
- Remark: This inner product is often called the Frobenius inner product.

- Our fundamental goal in studying inner products is to extend the notion of length in $\mathbb{R}^n$ to a more general setting. Using the positive-definiteness property [I3], we can define a notion of length in an inner product space.

- Definition: If V is an inner product space, we define the norm (or length) of a vector $\mathbf{v}$ to be $||\mathbf{v}|| = \sqrt{\langle \mathbf{v}, \mathbf{v}\rangle}$.
  - When $V = \mathbb{R}^n$ with the standard dot product, the norm on V reduces to the standard notion of "length" of a vector in $\mathbb{R}^n$.

- Here are a few basic properties of inner products and norms:

- Proposition (Properties of Inner Products): If V is an inner product space with inner product $\langle \cdot, \cdot\rangle$, then the following are true:

  1. For any vectors $\mathbf{v}$, $\mathbf{w}_1$, and $\mathbf{w}_2$, and any scalar c, $\langle \mathbf{v}, \mathbf{w}_1 + c\mathbf{w}_2\rangle = \langle \mathbf{v}, \mathbf{w}_1\rangle + c\langle \mathbf{v}, \mathbf{w}_2\rangle$
     - Proof: Apply [I1] and [I2]: $\langle \mathbf{v}, \mathbf{w}_1 + c\mathbf{w}_2\rangle = \langle \mathbf{w}_1 + c\mathbf{w}_2, \mathbf{v}\rangle = \langle \mathbf{w}_1, \mathbf{v}\rangle + c\langle \mathbf{w}_2, \mathbf{v}\rangle = \langle \mathbf{v}, \mathbf{w}_1\rangle + c\langle \mathbf{v}, \mathbf{w}_2\rangle$.
  2. For any vector $\mathbf{v}$, $\langle \mathbf{v}, \mathbf{0}\rangle = 0 = \langle \mathbf{0}, \mathbf{v}\rangle$.
     - Proof: Apply property (1) and [I2] with $c = 0$, using the fact that $0\mathbf{w} = \mathbf{0}$ for any $\mathbf{w}$.
  3. For any vector $\mathbf{v}$, $||\mathbf{v}||$ is a nonnegative real number, and $||\mathbf{v}|| = 0$ if and only if $\mathbf{v} = \mathbf{0}$.
     - Proof: Immediate from [I3].
  4. For any vector $\mathbf{v}$ and scalar c, $||c\mathbf{v}|| = |c| \cdot ||\mathbf{v}||$.
     - Proof: We have $||c\mathbf{v}|| = \sqrt{\langle c\mathbf{v}, c\mathbf{v}\rangle} = \sqrt{c^2\langle \mathbf{v}, \mathbf{v}\rangle} = |c| \cdot ||\mathbf{v}||$, using [I2] and property (1).

4.1.2 The Cauchy-Schwarz Inequality and Applications

- In $\mathbb{R}^n$, there are a number of fundamental inequalities about lengths, which generalize quite pleasantly to general inner product spaces.

- The following result, in particular, is one of the most fundamental inequalities in all of mathematics:

- Theorem (Cauchy-Schwarz Inequality): For any $\mathbf{v}$ and $\mathbf{w}$ in an inner product space V, we have $|\langle \mathbf{v}, \mathbf{w}\rangle| \leq ||\mathbf{v}||\,||\mathbf{w}||$, with equality if and only if the set $\{\mathbf{v}, \mathbf{w}\}$ is linearly dependent.
  - Proof: If $\mathbf{w} = \mathbf{0}$ then the result is trivial (since both sides are zero, and $\{\mathbf{v}, \mathbf{0}\}$ is always dependent), so now assume $\mathbf{w} \neq \mathbf{0}$.
  - Let $t = \dfrac{\langle \mathbf{v}, \mathbf{w}\rangle}{\langle \mathbf{w}, \mathbf{w}\rangle}$. By properties of inner products and norms, we can write

$$\begin{aligned} ||\mathbf{v} - t\mathbf{w}||^2 &= \langle \mathbf{v} - t\mathbf{w}, \mathbf{v} - t\mathbf{w}\rangle = \langle \mathbf{v}, \mathbf{v}\rangle - 2t\langle \mathbf{v}, \mathbf{w}\rangle + t^2\langle \mathbf{w}, \mathbf{w}\rangle \\ &= \langle \mathbf{v}, \mathbf{v}\rangle - \frac{\langle \mathbf{v}, \mathbf{w}\rangle^2}{\langle \mathbf{w}, \mathbf{w}\rangle}. \end{aligned}$$

  - Therefore, since $||\mathbf{v} - t\mathbf{w}||^2 \geq 0$ and $\langle \mathbf{w}, \mathbf{w}\rangle \geq 0$, clearing denominators and rearranging yields $\langle \mathbf{v}, \mathbf{w}\rangle^2 \leq \langle \mathbf{v}, \mathbf{v}\rangle\langle \mathbf{w}, \mathbf{w}\rangle$. Taking the square root yields the stated result.
  - Furthermore, we will have equality if and only if $||\mathbf{v} - t\mathbf{w}||^2 = 0$, which is in turn equivalent to $\mathbf{v} - t\mathbf{w} = \mathbf{0}$; namely, when $\mathbf{v}$ is a multiple of $\mathbf{w}$. Since we also get equality if $\mathbf{w} = \mathbf{0}$, equality occurs precisely when the set $\{\mathbf{v}, \mathbf{w}\}$ is linearly dependent.

- Remark: As written, this proof is completely mysterious: why does making that particular choice for t work? Here is some motivation: observe that $||\mathbf{v} - t\mathbf{w}||^2 = \langle \mathbf{v}, \mathbf{v} \rangle - 2t \langle \mathbf{v}, \mathbf{w} \rangle + t^2 \langle \mathbf{w}, \mathbf{w} \rangle$ is a quadratic function of t that is always nonnegative.
- To decide whether a quadratic function is always nonnegative, we can complete the square to see that $\langle \mathbf{v}, \mathbf{v} \rangle - 2t \langle \mathbf{v}, \mathbf{w} \rangle + t^2 \langle \mathbf{w}, \mathbf{w} \rangle = \langle \mathbf{w}, \mathbf{w} \rangle \left[t - \dfrac{\langle \mathbf{v}, \mathbf{w} \rangle}{\langle \mathbf{w}, \mathbf{w} \rangle} \right]^2 + \left[\langle \mathbf{v}, \mathbf{v} \rangle - \dfrac{\langle \mathbf{v}, \mathbf{w} \rangle^2}{\langle \mathbf{w}, \mathbf{w} \rangle} \right]$.
- Thus, the minimum value of the quadratic function is $\langle \mathbf{v}, \mathbf{v} \rangle - \dfrac{\langle \mathbf{v}, \mathbf{w} \rangle^2}{\langle \mathbf{w}, \mathbf{w} \rangle}$, and it occurs when $t = \dfrac{\langle \mathbf{v}, \mathbf{w} \rangle}{\langle \mathbf{w}, \mathbf{w} \rangle}$.

- The Cauchy-Schwarz inequality has many applications (most of which are, naturally, proving other inequalities). Here are a few such applications:

- Theorem (Triangle Inequality): For any $\mathbf{v}$ and $\mathbf{w}$ in an inner product space V, we have $||\mathbf{v} + \mathbf{w}|| \le ||\mathbf{v}|| + ||\mathbf{w}||$, with equality if and only if one vector is a positive-real scalar multiple of the other.
  - Proof: Using the Cauchy-Schwarz inequality and the fact that $\langle \mathbf{v}, \mathbf{w} \rangle \le |\langle \mathbf{v}, \mathbf{w} \rangle|$, we have

$$\begin{aligned} ||\mathbf{v} + \mathbf{w}||^2 = \langle \mathbf{v} + \mathbf{w}, \mathbf{v} + \mathbf{w} \rangle &= \langle \mathbf{v}, \mathbf{v} \rangle + 2 \langle \mathbf{v}, \mathbf{w} \rangle + \langle \mathbf{w}, \mathbf{w} \rangle \\ &\le \langle \mathbf{v}, \mathbf{v} \rangle + 2 |\langle \mathbf{v}, \mathbf{w} \rangle| + \langle \mathbf{w}, \mathbf{w} \rangle \\ &\le \langle \mathbf{v}, \mathbf{v} \rangle + 2 ||\mathbf{v}|| \, ||\mathbf{w}|| + \langle \mathbf{w}, \mathbf{w} \rangle \\ &= ||\mathbf{v}||^2 + 2 ||\mathbf{v}|| \, ||\mathbf{w}|| + ||\mathbf{w}||^2 . \end{aligned}$$

  Taking the square root of both sides yields the desired result.
  - Equality will hold if and only if $\{\mathbf{v}, \mathbf{w}\}$ is linearly dependent (for equality in the Cauchy-Schwarz inequality) and $\langle \mathbf{v}, \mathbf{w} \rangle$ is a nonnegative real number. If either vector is zero, equality always holds. Otherwise, we must have $\mathbf{v} = c\mathbf{w}$ for some nonzero constant c: then $\langle \mathbf{v}, \mathbf{w} \rangle = c \langle \mathbf{w}, \mathbf{w} \rangle$ will be a nonnegative real number if and only if c is a nonnegative real number.

- Example: Show that for any continuous function f on $[0, 3]$, it is true that $\int_0^3 x f(x)\, dx \le 3\sqrt{\int_0^3 f(x)^2\, dx}$.
  - Simply apply the Cauchy-Schwarz inequality to f and $g(x) = x$ in the inner product space of continuous functions on $[0, 3]$ with inner product $\langle f, g \rangle = \int_0^3 f(x)g(x)\, dx$.
  - We obtain $|\langle f, g \rangle| \le ||f|| \cdot ||g||$, or, explicitly, $\left| \int_0^3 x f(x)\, dx \right| \le \sqrt{\int_0^3 f(x)^2\, dx} \cdot \sqrt{\int_0^3 x^2\, dx} = 3\sqrt{\int_0^3 f(x)^2\, dx}$.
  - Since any real number is less than or equal to its absolute value, we immediately obtain the required inequality $\int_0^3 x f(x)\, dx \le 3\sqrt{\int_0^3 f(x)^2\, dx}$.

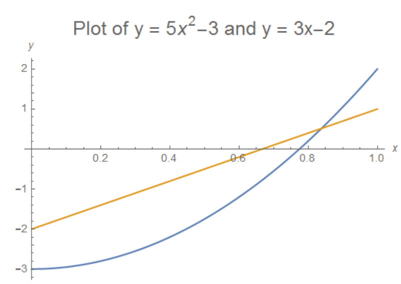
- Example: Show that for any positive reals a, b, c, it is true that $\sqrt{\dfrac{a + 2b}{a + b + c}} + \sqrt{\dfrac{b + 2c}{a + b + c}} + \sqrt{\dfrac{c + 2a}{a + b + c}} \le 3$.
  - Let $\mathbf{v} = (\sqrt{a + 2b}, \sqrt{b + 2c}, \sqrt{c + 2a})$ and $\mathbf{w} = (1, 1, 1)$ in $\mathbb{R}^3$. By the Cauchy-Schwarz inequality, $\mathbf{v} \cdot \mathbf{w} \le ||\mathbf{v}|| \, ||\mathbf{w}||$.
  - We compute $\mathbf{v} \cdot \mathbf{w} = \sqrt{a + 2b} + \sqrt{b + 2c} + \sqrt{c + 2a}$, along with $||\mathbf{v}||^2 = (a + 2b) + (b + 2c) + (c + 2a) = 3(a + b + c)$ and $||\mathbf{w}||^2 = 3$.
  - Thus, we see $\sqrt{a + 2b} + \sqrt{a + 2c} + \sqrt{b + 2c} \le \sqrt{3(a + b + c)} \cdot \sqrt{3}$, and upon dividing through by $\sqrt{a + b + c}$ we obtain the required inequality.

- Example (for those who like quantum mechanics): Prove the momentum-position formulation of Heisenberg's uncertainty principle: $\sigma_x \sigma_p \ge \hbar/2$. (In words: the product of uncertainties of position and momentum is greater than or equal to half of the reduced Planck constant.)
  - For this derivation we require the Cauchy-Schwarz inequality in an inner product space whose scalars are the complex numbers.

- It is a straightforward computation that, for two (complex-valued) observables X and Y, the pairing $\langle X, Y\rangle = E[X\overline{Y}]$, the expected value of $X\overline{Y}$, is an inner product on the space of observables.
- Assume (for simplicity) that x and p both have expected value 0.
- We assume as given the commutation relation $xp - px = i\overline{h}$.
- By definition, $(\sigma_x)^2 = E[x\overline{x}] = \langle x, x\rangle$ and $(\sigma_p)^2 = E[p\overline{p}] = E[\overline{p}p] = \langle \overline{p}, \overline{p}\rangle$ are the variances of x and p respectively (in the statistical sense).
- By the Cauchy-Schwarz inequality, we can therefore write $\sigma_x^2\sigma_p^2 = \langle x, x\rangle \langle \overline{p}, \overline{p}\rangle \geq |\langle x, \overline{p}\rangle|^2 = |E[xp]|^2$.
- We can write $xp = \frac{1}{2}(xp + px) + \frac{1}{2}(xp - px)$, where the first component is real and the second is imaginary, so taking expectations yields $E[xp] = \frac{1}{2}E[xp + px] + \frac{1}{2}E[xp - px]$, and therefore, $|E[xp]| \geq \frac{1}{2}|E[xp - px]| = \frac{1}{2}|i\overline{h}| = \frac{\overline{h}}{2}$.
- Combining with the inequality above yields $\sigma_x^2\sigma_p^2 \geq \overline{h}^2/4$, and taking square roots yields $\sigma_x\sigma_p \geq \overline{h}/2$.

4.2 Orthogonality

- Motivated by the Cauchy-Schwarz inequality, we can define a notion of angle between two nonzero vectors in an inner product space:
- Definition: If V is an inner product space, we define the angle between two nonzero vectors $\mathbf{v}$ and $\mathbf{w}$ to be the real number θ in $[0, \pi]$ satisfying $\cos\theta = \dfrac{\langle \mathbf{v}, \mathbf{w}\rangle}{||\mathbf{v}||\,||\mathbf{w}||}$.
  - By the Cauchy-Schwarz inequality, the quotient on the right is a real number in the interval $[-1, 1]$, so there is exactly one such angle θ.
- Example: Compute the angle between the vectors $\mathbf{v} = (3, -4, 5)$ and $\mathbf{w} = (1, 2, -2)$ under the standard dot product on $\mathbb{R}^3$.
  - We have $\mathbf{v}\cdot\mathbf{w} = -15$, $||\mathbf{v}|| = \sqrt{\mathbf{v}\cdot\mathbf{v}} = 5\sqrt{2}$, and $||\mathbf{w}|| = \sqrt{\mathbf{w}\cdot\mathbf{w}} = 3$.
  - Then the angle θ between the vectors satisfies $\cos(\theta) = \dfrac{-15}{15\sqrt{2}} = -\dfrac{1}{\sqrt{2}}$, so $\theta = \boxed{3\pi/4}$.
- Example: Compute the "angle" between $p = 5x^2 - 3$ and $q = 3x - 2$ in the inner product space of continuous functions with inner product $\langle f, g\rangle = \int_0^1 f(x)g(x)\,dx$.
  - We have $\langle p, q\rangle = \int_0^1 (5x^2-3)(3x-2)\,dx = 23/12$, $||p|| = \sqrt{\int_0^1 (5x^2-3)^2\,dx} = 2$, and $||q|| = \sqrt{\int_0^1 (3x-2)^2\,dx} = 1$.
  - Then the angle θ between the vectors satisfies $\cos(\theta) = \dfrac{23/12}{2} = \dfrac{23}{24}$, so $\theta = \boxed{\cos^{-1}(\frac{23}{24})}$.
  - The fact that this angle is so close to 0 suggests that these functions are nearly "parallel" in this inner product space. Indeed, the graphs of the two functions have very similar shapes:



- A particular case of interest is when the angle between two vectors is $\pi/2$, which we will discuss next.

4.2.1 Orthogonality and Orthonormal Sets

- Definition: We say two vectors in an inner product space are orthogonal if their inner product is zero. We say a set S of vectors is an orthogonal set if every pair of vectors in S is orthogonal.
  - By our basic properties, the zero vector is orthogonal to every vector. Two nonzero vectors will be orthogonal if and only if the angle between them is $\pi/2$. (This generalizes the idea of two vectors being "perpendicular".)
  - Example: In $\mathbb{R}^3$ with the standard dot product, the vectors $(1,0,0)$, $(0,1,0)$, and $(0,0,1)$ are orthogonal.
  - Example: In $\mathbb{R}^3$ with the standard dot product, the three vectors $(-1,1,2)$, $(2,0,1)$, and $(1,5,-2)$ form an orthogonal set, since the dot product of each pair is zero.
  - The first orthogonal set above seems more natural than the second. One reason for this is that the vectors in the first set each have length 1, while the vectors in the second set have various different lengths ($\sqrt{6}$, $\sqrt{5}$, and $\sqrt{30}$ respectively).
- Definition: We say a set S of vectors is an orthonormal set if every pair of vectors in S is orthogonal, and every vector in S has norm 1.
  - Example: In $\mathbb{R}^3$, $\{(1,0,0), (0,1,0), (0,0,1)\}$ is an orthonormal set, but $\{(-1,1,2), (2,0,1), (1,5,-2)\}$ is not.
- In both examples above, notice that the given orthogonal sets are also linearly independent. This feature is not an accident:
- Proposition (Orthogonality and Independence): In any inner product space, every orthogonal set of nonzero vectors is linearly independent.
  - Proof: Suppose we had a linear dependence $a_1\mathbf{v}_1 + \cdots + a_n\mathbf{v}_n = \mathbf{0}$ for an orthogonal set of nonzero vectors $\{\mathbf{v}_1, \ldots, \mathbf{v}_n\}$.
  - Then, for any j, $0 = \langle \mathbf{0}, \mathbf{v}_j \rangle = \langle a_1\mathbf{v}_1 + \cdots + a_n\mathbf{v}_n, \mathbf{v}_j \rangle = a_1\langle \mathbf{v}_1, \mathbf{v}_j \rangle + \cdots + a_n\langle \mathbf{v}_n, \mathbf{v}_j \rangle = a_j\langle \mathbf{v}_j, \mathbf{v}_j \rangle$, since each of the inner products $\langle \mathbf{v}_i, \mathbf{v}_j \rangle$ for $i \neq j$ is equal to zero.
  - But now, since $\mathbf{v}_j$ is not the zero vector, $\langle \mathbf{v}_j, \mathbf{v}_j \rangle$ is positive, so it must be the case that $a_j = 0$. This holds for every j, so all the coefficients of the linear dependence are zero. Hence there can be no nontrivial linear dependence, so any orthogonal set is linearly independent.
- Corollary: If V is an n-dimensional vector space and S is an orthogonal set of n nonzero vectors, then S is a basis for V. (We refer to such a basis as an orthogonal basis.)
  - Proof: By the proposition above, S is linearly independent, and by our earlier results, a linearly-independent set of n vectors in an n-dimensional vector space is necessarily a basis.
- If we have a basis of V, then (essentially by definition) every vector in V can be written as a unique linear combination of the basis vectors.
  - However, as we have seen, computing the coefficients of the linear combination can be quite cumbersome.
  - If, however, we have an orthogonal basis for V, then we can compute the coefficients for the linear combination much more conveniently.
- Theorem (Orthogonal Decomposition): If V is an n-dimensional vector space and $S = \{\mathbf{e}_1, \ldots, \mathbf{e}_n\}$ is an orthogonal basis, then for any $\mathbf{v}$ in S, we can write $\mathbf{v} = c_1\mathbf{e}_1 + \cdots + c_n\mathbf{e}_n$, where $c_k = \dfrac{\langle \mathbf{v}, \mathbf{e}_k \rangle}{\langle \mathbf{e}_k, \mathbf{e}_k \rangle}$ for each $1 \leq k \leq n$. In particular, if S is an orthonormal basis, then each $c_k = \langle \mathbf{v}, \mathbf{e}_k \rangle$.
  - Proof: Since S is a basis, there do exist such coefficients c_i and they are unique.
  - We then compute $\langle \mathbf{v}, \mathbf{e}_k \rangle = \langle c_1\mathbf{e}_1 + \cdots + c_n\mathbf{e}_n, \mathbf{e}_k \rangle = c_1 \langle \mathbf{e}_1, \mathbf{e}_k \rangle + \cdots + c_n \langle \mathbf{e}_n, \mathbf{e}_k \rangle = c_k \langle \mathbf{e}_k, \mathbf{e}_k \rangle$ since each of the inner products $\langle \mathbf{e}_j, \mathbf{e}_k \rangle$ for $j \neq k$ is equal to zero.

- Therefore, we must have $c_k = \dfrac{\langle \mathbf{v}, \mathbf{e}_k \rangle}{\langle \mathbf{e}_k, \mathbf{e}_k \rangle}$ for each $1 \le k \le n$.
- If S is an orthonormal basis, then $\langle \mathbf{e}_k, \mathbf{e}_k \rangle = 1$ for each k, so we get the simpler expression $c_k = \langle \mathbf{v}, \mathbf{e}_k \rangle$.

- Example: Write $\mathbf{v} = (7, 3, -4)$ as a linear combination of the basis vectors $\{(-1, 1, 2), (2, 0, 1), (1, 5, -2)\}$ of $\mathbb{R}^3$.
  - We saw above that this set is an orthogonal basis, so let $\mathbf{e}_1 = (-1, 1, 2)$, $\mathbf{e}_2 = (2, 0, 1)$, and $\mathbf{e}_3 = (1, 5, -2)$.
  - We compute $\mathbf{v} \cdot \mathbf{e}_1 = -12$, $\mathbf{v} \cdot \mathbf{e}_2 = 10$, $\mathbf{v} \cdot \mathbf{e}_3 = 30$, $\mathbf{e}_1 \cdot \mathbf{e}_1 = 6$, $\mathbf{e}_2 \cdot \mathbf{e}_2 = 5$, and $\mathbf{e}_3 \cdot \mathbf{e}_3 = 30$.
  - Thus, per the theorem, $\mathbf{v} = c_1 \mathbf{e}_1 + c_2 \mathbf{e}_2 + c_3 \mathbf{e}_3$ where $c_1 = \dfrac{-12}{6} = -2$, $c_2 = \dfrac{10}{5} = 2$, and $c_3 = \dfrac{30}{30} = 1$.
  - Indeed, we can verify that $\boxed{(7, 3, -4) = -2(-1, 1, 2) + 2(2, 0, 1) + 1(1, 5, -2)}$.

4.2.2 The Gram-Schmidt Procedure, QR Factorization

- Given a basis, there exists a way to write any vector as a linear combination of the basis elements: the advantage of having an orthogonal basis is that we can easily *compute* the coefficients. We now give an algorithm for constructing an orthogonal basis for any finite-dimensional inner product space:

- Theorem (Gram-Schmidt Procedure): Let $S = \{\mathbf{v}_1, \mathbf{v}_2, \dots\}$ be a basis of the inner product space V, and set $V_k = \text{span}(\mathbf{v}_1, \dots, \mathbf{v}_k)$. Then there exists an orthogonal set of vectors $\{\mathbf{w}_1, \mathbf{w}_2, \dots\}$ such that, for each $k \ge 1$, $\text{span}(\mathbf{w}_1, \dots, \mathbf{w}_k) = \text{span}(V_k)$ and $\mathbf{w}_k$ is orthogonal to every vector in V_{k-1}. Furthermore, this sequence is unique up to multiplying the elements by nonzero scalars.
  - Proof: We construct the sequence $\{\mathbf{w}_1, \mathbf{w}_2, \dots\}$ recursively: we start with the simple choice $\mathbf{w}_1 = \mathbf{v}_1$.
  - Now suppose we have constructed $\{\mathbf{w}_1, \mathbf{w}_2, \dots, \mathbf{w}_{k-1}\}$, where $\text{span}(\mathbf{w}_1, \dots, \mathbf{w}_{k-1}) = \text{span}(V_{k-1})$.
  - Define the next vector as $\mathbf{w}_k = \mathbf{v}_k - a_1 \mathbf{w}_1 - a_2 \mathbf{w}_2 - \cdots - a_{k-1} \mathbf{w}_{k-1}$, where $a_j = \langle \mathbf{v}_k, \mathbf{w}_j \rangle / \langle \mathbf{w}_j, \mathbf{w}_j \rangle$.
  - From the construction, we can see that each of $\mathbf{w}_1, \dots, \mathbf{w}_k$ is a linear combination of $\mathbf{v}_1, \dots, \mathbf{v}_k$, and vice versa. Thus, by properties of span, $\text{span}(\mathbf{w}_1, \dots, \mathbf{w}_k) = V_k$.
  - Furthermore, we can compute

$$\begin{aligned} \langle \mathbf{w}_k, \mathbf{w}_j \rangle &= \langle \mathbf{v}_k - a_1 \mathbf{w}_1 - a_2 \mathbf{w}_2 - \cdots - a_{k-1} \mathbf{w}_{k-1}, \mathbf{w}_j \rangle \\ &= \langle \mathbf{v}_k, \mathbf{w}_j \rangle - a_1 \langle \mathbf{w}_1, \mathbf{w}_j \rangle - \cdots - a_{k-1} \langle \mathbf{w}_{k-1}, \mathbf{w}_j \rangle \\ &= \langle \mathbf{v}_k, \mathbf{w}_j \rangle - a_j \langle \mathbf{w}_j, \mathbf{w}_j \rangle = 0 \end{aligned}$$

    because all of the inner products $\langle \mathbf{w}_i, \mathbf{w}_j \rangle$ are zero except for $\langle \mathbf{w}_j, \mathbf{w}_j \rangle$.
  - Therefore, $\mathbf{w}_k$ is orthogonal to each of $\mathbf{w}_1, \dots, \mathbf{w}_{k-1}$, and is therefore orthogonal to all linear combinations of these vectors.
  - The uniqueness follows from the observation that (upon appropriate rescaling) we must choose $\mathbf{w}_k = \mathbf{v}_k - a_1 \mathbf{w}_1 - a_2 \mathbf{w}_2 - \cdots - a_{k-1} \mathbf{w}_{k-1}$ for some scalars $a_1, \dots a_{k-1}$: orthogonality then forces the choice of the coefficients a_j that we used above, since $\langle \mathbf{w}_k, \mathbf{w}_j \rangle = \langle \mathbf{v}_k, \mathbf{w}_j \rangle - a_j \langle \mathbf{w}_j, \mathbf{w}_j \rangle$ must be zero.

- Corollary: Every finite-dimensional inner product space has an orthonormal basis.
  - Proof: Choose any basis $\{\mathbf{v}_1, \dots, \mathbf{v}_n\}$ for V and apply the Gram-Schmidt procedure: this yields an orthogonal basis $\{\mathbf{w}_1, \dots, \mathbf{w}_n\}$ for V.
  - Now simply normalize each vector in $\{\mathbf{w}_1, \dots, \mathbf{w}_n\}$ by dividing by its norm: this preserves orthogonality, but rescales each vector to have norm 1, thus yielding an orthonormal basis for V.

- The proof of the Gram-Schmidt procedure may seem involved, but applying it in practice is fairly straightforward (if somewhat cumbersome computationally).
  - We remark here that, although our algorithm above gives an orthogonal basis, it is also possible to perform the normalization at each step during the procedure, to construct an orthonormal basis one vector at a time.

- When performing computations by hand, it is generally disadvantageous to normalize at each step, because the norm of a vector will often involve square roots (which will then be carried into subsequent steps of the computation).
- When using a computer (with approximate arithmetic), however, normalizing at each step can avoid certain numerical instability issues. The particular description of the algorithm we have discussed turns out not to be especially numerically stable, but it is possible to modify the algorithm to avoid magnifying the error as substantially when iterating the procedure.

- Example: For $V = \mathbb{R}^2$ with the standard inner product, apply the Gram-Schmidt procedure to the vectors $\mathbf{v}_1 = (2,1)$, $\mathbf{v}_2 = (1,4)$. Use the result to find an orthonormal basis for $\mathbb{R}^2$.
  - We start with $\mathbf{w}_1 = \mathbf{v}_1 = \boxed{(2,1)}$.
  - Next, $\mathbf{w}_2 = \mathbf{v}_2 - a_1\mathbf{w}_1$, where $a_1 = \dfrac{\mathbf{v}_2 \cdot \mathbf{w}_1}{\mathbf{w}_1 \cdot \mathbf{w}_1} = \dfrac{(2,1)\cdot(1,4)}{(2,1)\cdot(2,1)} = \dfrac{6}{5}$. Thus, $\mathbf{w}_2 = \boxed{(-\frac{7}{5}, \frac{14}{5})}$.
  - For the orthonormal basis, we simply divide each vector by its length.
  - We get $\dfrac{\mathbf{w}_1}{||\mathbf{w}_1||} = \dfrac{1}{\sqrt{5}}(2,1)$, $\dfrac{\mathbf{w}_2}{||\mathbf{w}_2||} = \dfrac{1}{\sqrt{5}}(-1,2)$.

- Example: For $V = \mathbb{R}^3$ with the standard inner product, apply the Gram-Schmidt procedure to the vectors $\mathbf{v}_1 = (2,1,2)$, $\mathbf{v}_2 = (5,4,2)$, $\mathbf{v}_3 = (-1,2,1)$. Use the result to find an orthonormal basis for $\mathbb{R}^3$.
  - We start with $\mathbf{w}_1 = \mathbf{v}_1 = \boxed{(2,1,2)}$.
  - Next, $\mathbf{w}_2 = \mathbf{v}_2 - a_1\mathbf{w}_1$, where $a_1 = \dfrac{\mathbf{v}_2 \cdot \mathbf{w}_1}{\mathbf{w}_1 \cdot \mathbf{w}_1} = \dfrac{(5,4,2)\cdot(2,1,2)}{(2,1,2)\cdot(2,1,2)} = \dfrac{18}{9} = 2$. Thus, $\mathbf{w}_2 = \boxed{(1,2,-2)}$.
  - Finally, $\mathbf{w}_3 = \mathbf{v}_3 - b_1\mathbf{w}_1 - b_2\mathbf{w}_2$ where $b_1 = \dfrac{\mathbf{v}_3 \cdot \mathbf{w}_1}{\mathbf{w}_1 \cdot \mathbf{w}_1} = \dfrac{(-1,2,1)\cdot(2,1,2)}{(2,1,2)\cdot(2,1,2)} = \dfrac{2}{9}$, and $b_2 = \dfrac{\mathbf{v}_3 \cdot \mathbf{w}_2}{\mathbf{w}_2 \cdot \mathbf{w}_2} = \dfrac{(-1,2,1)\cdot(1,2,-2)}{(1,2,-2)\cdot(1,2,-2)} = \dfrac{1}{9}$. Thus, $\mathbf{w}_3 = \boxed{(-\frac{14}{9}, \frac{14}{9}, \frac{7}{9})}$.
  - For the orthonormal basis, we simply divide each vector by its length.
  - We get $\dfrac{\mathbf{w}_1}{||\mathbf{w}_1||} = (\frac{2}{3}, \frac{1}{3}, \frac{2}{3})$, $\dfrac{\mathbf{w}_2}{||\mathbf{w}_2||} = (\frac{1}{3}, \frac{2}{3}, -\frac{2}{3})$, and $\dfrac{\mathbf{w}_3}{||\mathbf{w}_3||} = (-\frac{2}{3}, \frac{2}{3}, \frac{1}{3})$.

- Example: For $V = P(\mathbb{R})$ with inner product $\langle f, g\rangle = \int_0^1 f(x)g(x)\,dx$, apply the Gram-Schmidt procedure to the polynomials $p_1 = 1$, $p_2 = x$, $p_3 = x^2$.
  - We start with $\mathbf{w}_1 = p_1 = \boxed{1}$.
  - Next, $\mathbf{w}_2 = p_2 - a_1\mathbf{w}_1$, where $a_1 = \dfrac{\langle p_2, \mathbf{w}_1\rangle}{\langle \mathbf{w}_1, \mathbf{w}_1\rangle} = \dfrac{\int_0^1 x\,dx}{\int_0^1 1\,dx} = \dfrac{1}{2}$. Thus, $\mathbf{w}_2 = \boxed{x - \frac{1}{2}}$.
  - Finally, $\mathbf{w}_3 = p_3 - b_1\mathbf{w}_1 - b_2\mathbf{w}_2$ where $b_1 = \dfrac{\langle p_3, \mathbf{w}_1\rangle}{\langle \mathbf{w}_1, \mathbf{w}_1\rangle} = \dfrac{\int_0^1 x^2\,dx}{\int_0^1 1\,dx} = \dfrac{1}{3}$, and $b_2 = \dfrac{\langle p_3, \mathbf{w}_2\rangle}{\langle \mathbf{w}_2, \mathbf{w}_2\rangle} = \dfrac{\int_0^1 x^2(x-1/2)\,dx}{\int_0^1 (x-1/2)^2\,dx} = \dfrac{1/12}{1/12} = 1$. Thus, $\mathbf{w}_3 = \boxed{x^2 - x + \frac{1}{6}}$.

- By applying the Gram-Schmidt procedure to the linearly independent columns of a matrix, we obtain a useful way to factor matrices:

- Theorem (QR Factorization): Suppose A is an $m \times n$ matrix whose columns are linearly independent. Then there exists a unique $m \times n$ matrix Q whose columns are orthonormal and a unique upper-triangular $n \times n$ matrix R with positive diagonal entries such that $A = QR$.
  - This factorization is known as the QR factorization of the matrix A. We remark that the $m \times n$ matrix Q has orthonormal columns if and only if $Q^T Q = I_n$.

- Proof: Suppose A has columns $\mathbf{v}_1, \dots, \mathbf{v}_m$, Q has orthonormal columns $\mathbf{e}_1, \dots, \mathbf{e}_m$, and $R = \begin{bmatrix} r_{1,1} & r_{1,2} & \cdots & r_{1,n} \\ 0 & r_{2,2} & \cdots & r_{2,n} \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & r_{n,n} \end{bmatrix}$
- Performing the matrix multiplication shows that QR has columns $r_{1,1}\mathbf{e}_1$, $r_{1,2}\mathbf{e}_1 + r_{2,2}\mathbf{e}_2$, $r_{1,3}\mathbf{e}_1 + r_{2,3}\mathbf{e}_2 + r_{3,3}\mathbf{e}_3$, ... , $r_{1,n}\mathbf{e}_1 + r_{2,n}\mathbf{e}_2 + \cdots + r_{n,n}\mathbf{e}_n$.
- Therefore, we have $\mathbf{v}_1 = r_{1,1}\mathbf{e}_1$, $\mathbf{v}_2 = r_{1,2}\mathbf{e}_1 + r_{2,2}\mathbf{e}_2$, ... , $\mathbf{v}_n = r_{1,n}\mathbf{e}_1 + r_{2,n}\mathbf{e}_2 + \cdots + r_{n,n}\mathbf{e}_n$.
- If we let $\mathbf{e}_i = \dfrac{\mathbf{w}_i}{||\mathbf{w}_i||}$, these relations are precisely the ones we get by applying the Gram-Schmidt procedure to the vectors $\mathbf{v}_1, \dots, \mathbf{v}_m$.
- Therefore, if we apply the Gram-Schmidt procedure to $\mathbf{v}_1, \dots, \mathbf{v}_m$ to obtain orthonormal vectors $\mathbf{e}_1, \dots, \mathbf{e}_m$, we obtain the matrix Q.
- The matrix R is embedded in the calculations for applying the Gram-Schmidt procedure, but we can also give a simple formula for the entries $r_{i,j}$ by observing that $\langle \mathbf{v}_i, \mathbf{e}_j \rangle = r_{i,j}$ is simply the coefficient of $\mathbf{e}_j$ in the expression for $\mathbf{v}_i$.
- To show uniqueness, suppose A is an $m \times n$ matrix with two QR factorizations $A = Q_1 R_1 = Q_2 R_2$. Then R_1 is invertible and its inverse is also upper-triangular with positive diagonal entries, so we may write $Q_1 = Q_2 R_2 R_1^{-1}$.
- Let $S = R_2 R_1^{-1}$: then S is the product of two upper-triangular matrices with positive diagonal entries, so it is also upper-triangular with positive diagonal entries.
- Then we have $I_m = Q_1^T Q_1 = (Q_2 S)^T (Q_2 S) = S^T Q_2^T S = S^T S$, meaning S also has orthogonal columns. But it is easy to see that the only upper-triangular matrix with positive diagonal entries and orthonormal columns is the identity matrix I_n. Thus, $S = I_n$, meaning that $R_1 = R_2$ and then $Q_1 = Q_2$. Thus, the QR factorization is unique, as claimed.

• Another way to think of the QR factorization is as a change of basis.

- Explicitly, the decomposition $A = QR$ is simply keeping track of the change of basis from the basis β given by the columns of A to the orthonormal basis γ obtained via Gram-Schmidt: the matrix R is simply the change-of-basis matrix $[I]_\beta^\gamma$.
- To calculate a QR factorization, we simply perform Gram-Schmidt on the columns and keep track of all of the coefficients used during the calculation.
- If we normalize each vector during the calculation, then we can also calculate the entries of the matrix R as $r_{i,j} = \langle \mathbf{e}_i, \mathbf{v}_j \rangle$.
- In terms of the notation $\mathbf{w}_j = \mathbf{v}_j - a_1\mathbf{w}_1 - a_2\mathbf{w}_2 - \cdots - a_{k-1}\mathbf{w}_{j-1}$, where $a_i = \langle \mathbf{v}_j, \mathbf{w}_i \rangle / \langle \mathbf{w}_i, \mathbf{w}_i \rangle$ from our Gram-Schmidt calculations earlier, we have $r_{i,i} = ||\mathbf{w}_i||$ for each i and $r_{i,j} = \dfrac{\mathbf{w}_i \cdot \mathbf{v}_j}{||\mathbf{w}_i||} = a_i \, ||\mathbf{w}_i||$ for each $i < j$.

• Example: Find the QR factorization of $A = \begin{bmatrix} 3 & 4 \\ 4 & 5 \end{bmatrix}$.

- We start with $\mathbf{w}_1 = \mathbf{v}_1 = (3, 4)$.
- Then $\mathbf{w}_2 = \mathbf{v}_2 - a_1\mathbf{w}_1$, where $a_1 = \dfrac{\mathbf{v}_2 \cdot \mathbf{w}_1}{\mathbf{w}_1 \cdot \mathbf{w}_1} = \dfrac{(4,5)\cdot(3,4)}{(3,4)\cdot(3,4)} = \dfrac{32}{25}$, so $\mathbf{w}_2 = (4/25, -3/25)$.
- For the orthonormal basis, we simply divide each vector by its length, yielding $\mathbf{e}_1 = \dfrac{\mathbf{w}_1}{||\mathbf{w}_1||} = (3/5, 4/5)$ and $\mathbf{e}_2 = \dfrac{\mathbf{w}_2}{||\mathbf{w}_2||} = (4/5, -3/5)$.
- We also compute the coefficients $r_{1,1} = \langle \mathbf{e}_1, \mathbf{v}_1 \rangle = ||\mathbf{w}_1|| = 5$, $r_{1,2} = \langle \mathbf{e}_1, \mathbf{v}_2 \rangle = \dfrac{\mathbf{w}_1 \cdot \mathbf{v}_2}{||\mathbf{w}_1||} = \dfrac{32}{5}$, and $r_{2,2} = \langle \mathbf{e}_2, \mathbf{v}_2 \rangle = ||\mathbf{w}_2|| = \dfrac{1}{5}$.

- Thus, $A = QR$ with $Q = \boxed{\begin{bmatrix} 3/5 & 4/5 \\ 4/5 & -3/5 \end{bmatrix}}$ and $R = \boxed{\begin{bmatrix} 5 & 32/5 \\ 0 & 1/5 \end{bmatrix}}$.

- <u>Example</u>: Find the QR factorization of $A = \begin{bmatrix} 1 & -3 \\ 2 & 5 \end{bmatrix}$.
  - We start with $\mathbf{w}_1 = \mathbf{v}_1 = (1, 2)$.
  - Then $\mathbf{w}_2 = \mathbf{v}_2 - a_1\mathbf{w}_1$, where $a_1 = \dfrac{\mathbf{v}_2 \cdot \mathbf{w}_1}{\mathbf{w}_1 \cdot \mathbf{w}_1} = \dfrac{(1,2) \cdot (-3,5)}{(1,2) \cdot (1,2)} = \dfrac{7}{5}$, so $\mathbf{w}_2 = (-22/5, 11/5)$.
  - For the orthonormal basis, we simply divide each vector by its length, yielding $\mathbf{e}_1 = \dfrac{\mathbf{w}_1}{||\mathbf{w}_1||} = \dfrac{1}{\sqrt{5}}(1,2)$ and $\mathbf{e}_2 = \dfrac{\mathbf{w}_2}{||\mathbf{w}_2||} = \dfrac{1}{\sqrt{5}}(-2,1)$.
  - We also compute the coefficients $r_{1,1} = \langle \mathbf{e}_1, \mathbf{v}_1 \rangle = ||\mathbf{w}_1|| = \sqrt{5}$, $r_{1,2} = \langle \mathbf{e}_1, \mathbf{v}_2 \rangle = \dfrac{\mathbf{w}_1 \cdot \mathbf{v}_2}{||\mathbf{w}_1||} = \dfrac{7}{\sqrt{5}}$, and $r_{2,2} = \langle \mathbf{e}_2, \mathbf{v}_2 \rangle = ||\mathbf{w}_2|| = \dfrac{11}{\sqrt{5}}$.
  - Thus, $A = QR$ with $Q = \boxed{\dfrac{1}{\sqrt{5}}\begin{bmatrix} 1 & -2 \\ 2 & 1 \end{bmatrix}}$ and $R = \boxed{\dfrac{1}{\sqrt{5}}\begin{bmatrix} 5 & 7 \\ 0 & 11 \end{bmatrix}}$.

- <u>Example</u>: Find the QR factorization of $A = \begin{bmatrix} 9 & 0 \\ 18 & 21 \\ 6 & -14 \end{bmatrix}$.
  - We start with $\mathbf{w}_1 = \mathbf{v}_1 = (9, 18, 6)$.
  - Then $\mathbf{w}_2 = \mathbf{v}_2 - a_1\mathbf{w}_1$, where $a_1 = \dfrac{\mathbf{v}_2 \cdot \mathbf{w}_1}{\mathbf{w}_1 \cdot \mathbf{w}_1} = \dfrac{(9,18,6) \cdot (0,21,-14)}{(9,18,6) \cdot (9,18,6)} = \dfrac{2}{3}$, so $\mathbf{w}_2 = (-6, 9, -18)$.
  - For the orthonormal basis, we simply divide each vector by its length, yielding $\mathbf{e}_1 = \dfrac{\mathbf{w}_1}{||\mathbf{w}_1||} = (3/7, 6/7, 2/7)$ and $\mathbf{e}_2 = \dfrac{\mathbf{w}_2}{||\mathbf{w}_2||} = (-2/7, 3/7, -6/7)$.
  - We also compute the coefficients $r_{1,1} = \langle \mathbf{e}_1, \mathbf{v}_1 \rangle = ||\mathbf{w}_1|| = 21$, $r_{1,2} = \langle \mathbf{e}_1, \mathbf{v}_2 \rangle = \dfrac{\mathbf{w}_1 \cdot \mathbf{v}_2}{||\mathbf{w}_1||} = 14$, and $r_{2,2} = \langle \mathbf{e}_2, \mathbf{v}_2 \rangle = ||\mathbf{w}_2|| = 21$.
  - Thus, $A = QR$ with $Q = \boxed{\begin{bmatrix} 3/7 & -2/7 \\ 6/7 & 3/7 \\ 2/7 & -6/7 \end{bmatrix}}$ and $R = \boxed{\begin{bmatrix} 21 & 14 \\ 0 & 21 \end{bmatrix}}$.

- <u>Example</u>: Find the QR factorization of $A = \begin{bmatrix} 2 & 5 & 2 \\ 1 & 7 & 10 \\ 2 & -4 & 2 \end{bmatrix}$.
  - We start with $\mathbf{w}_1 = \mathbf{v}_1 = (2, 1, 2)$.
  - Next, $\mathbf{w}_2 = \mathbf{v}_2 - a_1\mathbf{w}_1$, where $a_1 = \dfrac{\mathbf{v}_2 \cdot \mathbf{w}_1}{\mathbf{w}_1 \cdot \mathbf{w}_1} = \dfrac{(5,7,-4) \cdot (2,1,2)}{(2,1,2) \cdot (2,1,2)} = \dfrac{9}{9} = 1$. Thus, $\mathbf{w}_2 = (3, 6, -6)$.
  - Finally, $\mathbf{w}_3 = \mathbf{v}_3 - b_1\mathbf{w}_1 - b_2\mathbf{w}_2$ where $b_1 = \dfrac{\mathbf{v}_3 \cdot \mathbf{w}_1}{\mathbf{w}_1 \cdot \mathbf{w}_1} = \dfrac{(2,10,2) \cdot (2,1,2)}{(2,1,2) \cdot (2,1,2)} = 2$, and $b_2 = \dfrac{\mathbf{v}_3 \cdot \mathbf{w}_2}{\mathbf{w}_2 \cdot \mathbf{w}_2} = \dfrac{(2,10,2) \cdot (3,6,-6)}{(3,6,-6) \cdot (3,6,-6)} = \dfrac{2}{3}$. Thus, $\mathbf{w}_3 = (-4, 4, 2)$.
  - For the orthonormal basis we get $\mathbf{e}_1 = \dfrac{\mathbf{w}_1}{||\mathbf{w}_1||} = (\dfrac{2}{3}, \dfrac{1}{3}, \dfrac{2}{3})$, $\mathbf{e}_2 = \dfrac{\mathbf{w}_2}{||\mathbf{w}_2||} = (\dfrac{1}{3}, \dfrac{2}{3}, -\dfrac{2}{3})$, and $\mathbf{e}_3 = \dfrac{\mathbf{w}_3}{||\mathbf{w}_3||} = (-\dfrac{2}{3}, \dfrac{2}{3}, \dfrac{1}{3})$.

- We also compute the coefficients $r_{1,1} = ||\mathbf{w}_1|| = 3$, $r_{1,2} = a_1 ||\mathbf{w}_1|| = 3$, $r_{2,2} = ||\mathbf{w}_2|| = 9$, $r_{1,3} = b_1 ||\mathbf{w}_1|| = 6$, $r_{2,3} = b_2 ||\mathbf{w}_2|| = 6$, and $r_{3,3} = ||\mathbf{w}_3|| = 6$.
- Thus, $A = QR$ with $Q = \boxed{\frac{1}{3}\begin{bmatrix} 2 & 1 & -2 \\ 1 & 2 & 2 \\ 2 & -2 & 1 \end{bmatrix}}$ and $R = \boxed{\begin{bmatrix} 3 & 3 & 6 \\ 0 & 9 & 6 \\ 0 & 0 & 6 \end{bmatrix}}$.

4.2.3 Orthogonal Complements

- If V is an inner product space, W is a subspace, and $\mathbf{v}$ is some vector in V, we would like to study the problem of finding a "best approximation" of $\mathbf{v}$ in W.
  - For two vectors $\mathbf{v}$ and $\mathbf{w}$, the distance between $\mathbf{v}$ and $\mathbf{w}$ is $||\mathbf{v} - \mathbf{w}||$, so what we are seeking is a vector $\mathbf{w}$ in S that minimizes the quantity $||\mathbf{v} - \mathbf{w}||$.
  - As a particular example, suppose we are given a point P in $\mathbb{R}^2$ and wish to find the minimal distance from P to a particular line in $\mathbb{R}^2$. Geometrically, the minimal distance is achieved by the segment PQ, where Q is chosen so that PQ is perpendicular to the line.
  - In a similar way, the minimal distance between a point in $\mathbb{R}^3$ and a given plane will also be minimized by finding the segment perpendicular to the plane.
  - Both of these problems suggest that the solution to this optimization problem will involve some notion of "perpendicularity" to the subspace W.
- Definition: Let V be an inner product space. If S is a nonempty subset of V, we say a vector $\mathbf{v}$ in V is orthogonal to S if it is orthogonal to every vector in S. The set of all vectors orthogonal to S is denoted $S^\perp$ ("S-perpendicular", or often "S-perp" for short).
  - We will typically be interested in the case where S is a subspace of V. It is easy to see via the subspace criterion that $S^\perp$is always a subspace of V, even if S itself is not.
  - Example: In $\mathbb{R}^3$, if W is the xy-plane consisting of all vectors of the form $(x, y, 0)$, then $W^\perp$ is the z-axis, consisting of the vectors of the form $(0, 0, z)$.
  - Example: In $\mathbb{R}^3$, if W is the x-axis consisting of all vectors of the form $(x, 0, 0)$, then $W^\perp$ is the yz-plane, consisting of the vectors of the form $(0, y, z)$.
  - Example: In any inner product space V, $V^\perp = \{\mathbf{0}\}$ and $\{\mathbf{0}\}^\perp = V$.
- When V is finite-dimensional, we can use the Gram-Schmidt procedure to compute an explicit basis of $W^\perp$:
- Theorem (Basis for Orthogonal Complement): Suppose W is a subspace of the finite-dimensional inner product space V, and that $S = \{\mathbf{e}_1, \dots, \mathbf{e}_k\}$ is an orthonormal basis for W. If $\{\mathbf{e}_1, \dots, \mathbf{e}_k, \mathbf{e}_{k+1}, \dots, \mathbf{e}_n\}$ is any extension of S to an orthonormal basis for V, the set $\{\mathbf{e}_{k+1}, \dots, \mathbf{e}_n\}$ is an orthonormal basis for $W^\perp$. In particular, $\dim(V) = \dim(W) + \dim(W^\perp)$.
  - Remark: It is always possible to extend the orthonormal basis $S = \{\mathbf{e}_1, \dots, \mathbf{e}_k\}$ to an orthonormal basis for V: simply extend the linearly independent set S to a basis $\{\mathbf{e}_1, \dots, \mathbf{e}_k, \mathbf{x}_{k+1}, \dots, \mathbf{x}_n\}$ of V, and then apply Gram-Schmidt to generate an orthonormal basis $\{\mathbf{e}_1, \dots, \mathbf{e}_k, \mathbf{e}_{k+1}, \dots, \mathbf{e}_n\}$.
  - Proof: For the first statement, the set $\{\mathbf{e}_{k+1}, \dots, \mathbf{e}_n\}$ is orthonormal and hence linearly independent. Since each vector is orthogonal to every vector in S, each of $\mathbf{e}_{k+1}, \dots, \mathbf{e}_n$ is in $W^\perp$, and so it remains to show that $\{\mathbf{e}_{k+1}, \dots, \mathbf{e}_n\}$ spans $W^\perp$.
  - So let $\mathbf{v}$ be any vector in $W^\perp$. Since $\{\mathbf{e}_1, \dots, \mathbf{e}_k, \mathbf{e}_{k+1}, \dots, \mathbf{e}_n\}$ is an orthonormal basis of V, by the orthogonal decomposition we know that $\mathbf{v} = \langle \mathbf{v}, \mathbf{e}_1 \rangle \mathbf{e}_1 + \cdots + \langle \mathbf{v}, \mathbf{e}_k \rangle \mathbf{e}_k + \langle \mathbf{v}, \mathbf{e}_{k+1} \rangle \mathbf{e}_{k+1} + \cdots + \langle \mathbf{v}, \mathbf{e}_n \rangle \mathbf{e}_n$.
  - But since $\mathbf{v}$ is in $W^\perp$, $\langle \mathbf{v}, \mathbf{e}_1 \rangle = \cdots = \langle \mathbf{v}, \mathbf{e}_k \rangle = 0$: thus, $\mathbf{v} = \langle \mathbf{v}, \mathbf{e}_{k+1} \rangle \mathbf{e}_{k+1} + \cdots + \langle \mathbf{v}, \mathbf{e}_n \rangle \mathbf{e}_n$, and therefore $\mathbf{v}$ is contained in the span of $\{\mathbf{e}_{k+1}, \dots, \mathbf{e}_n\}$, as required.
  - The statement about dimensions follows immediately from our explicit construction of the basis of V as a union of the basis for W and the basis for $W^\perp$.
- Example: If $W = \text{span}[\frac{1}{3}(1, 2, -2), \frac{1}{3}(-2, 2, 1)]$ in $\mathbb{R}^3$ with the standard dot product, find a basis for $W^\perp$.

- Notice that the vectors $\mathbf{e}_1 = \frac{1}{3}(1,2,-2)$ and $\mathbf{e}_2 = \frac{1}{3}(-2,2,1)$ form an orthonormal basis for W.
- It is straightforward to verify that if $\mathbf{v}_3 = (1,0,0)$, then $\{\mathbf{e}_1, \mathbf{e}_2, \mathbf{v}_3\}$ is a linearly independent set and therefore a basis for $\mathbb{R}^3$.
- Applying Gram-Schmidt to the set $\{\mathbf{e}_1, \mathbf{e}_2, \mathbf{v}_3\}$ yields $\mathbf{w}_1 = \mathbf{e}_1$, $\mathbf{w}_2 = \mathbf{e}_2$, and $\mathbf{w}_3 = \mathbf{v}_3 - \langle \mathbf{v}_3, \mathbf{w}_1 \rangle \mathbf{w}_1 - \langle \mathbf{v}_3, \mathbf{w}_2 \rangle \mathbf{w}_2 = \frac{1}{9}(4,2,4)$.
- Normalizing $\mathbf{w}_3$ produces the orthonormal basis $\{\mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3\}$ for V, with $\mathbf{e}_3 = \frac{1}{3}(2,1,2)$.
- By the theorem above, we conclude that $\{\mathbf{e}_3\} = \boxed{\{\frac{1}{3}(2,1,2)\}}$ is an orthonormal basis of $W^\perp$.
- Alternatively, we could have computed a basis for $W^\perp$ by observing that $\dim(W^\perp) = \dim(V) - \dim(W) = 1$, and then simply finding one nonzero vector orthogonal to both $\frac{1}{3}(1,2,-2)$ and $\frac{1}{3}(-2,2,1)$. (For this, we could have either solved the system of equations explicitly, or computed the cross product of the two given vectors.)

- We can give a simpler (although ultimately equivalent) method for finding a basis for $W^\perp$ using matrices:
- <u>Theorem</u> (Orthogonal Complements and Matrices): If A is an $m \times n$ matrix, then the rowspace of A and the nullspace of A are orthogonal complements of one another in $\mathbb{R}^n$, with respect to the standard dot product.
  - <u>Proof</u>: Let A be an $m \times n$ matrix, so that the rowspace and nullspace are both subspaces of $\mathbb{R}^n$.
  - By definition, any vector in rowspace(A) is orthogonal to any vector in nullspace(A), so rowspace(A) $\subseteq$ nullspace$(A)^\perp$ and nullspace(A) $\subseteq$ rowspace$(A)^\perp$.
  - Furthermore, since $\dim(\text{rowspace}(A)) + \dim(\text{nullspace}(A)) = n$ from our results on the respective dimensions of these spaces, we see that $\dim(\text{rowspace}(A)) = \dim(\text{nullspace}(A)^\perp)$ and $\dim(\text{nullspace}(A)) = \dim(\text{rowspace}(A)^\perp)$.
  - Since all these spaces are finite-dimensional, we must therefore have equality everywhere, as claimed.
- From the theorem above, when W is a subspace of $\mathbb{R}^n$ with respect to the standard dot product, we can easily compute a basis for $W^\perp$ by computing the nullspace of the matrix whose rows are a spanning set for W.
  - Although this method is much faster, it will not produce an orthonormal basis of W. It can also be adapted for subspaces of an arbitrary finite-dimensional inner product space, but this requires having an orthonormal basis for the space computed ahead of time.
- <u>Example</u>: If $W = \text{span}[(1,1,-1,1),\ (1,2,0,-2)]$ in $\mathbb{R}^4$ with the standard dot product, find a basis for $W^\perp$.
  - We row-reduce the matrix whose rows are the given basis for W:

$$\begin{bmatrix} 1 & 1 & -1 & 1 \\ 1 & 2 & 0 & -2 \end{bmatrix} \xrightarrow{R_2 - R_1} \begin{bmatrix} 1 & 1 & -1 & 1 \\ 0 & 1 & 1 & -3 \end{bmatrix} \xrightarrow{R_1 - R_2} \begin{bmatrix} 1 & 0 & -2 & 4 \\ 0 & 1 & 1 & -3 \end{bmatrix}.$$

  - From the reduced row-echelon form, we see that $\boxed{\{(-4,3,0,1),\ (2,-1,1,0)\}}$ is a basis for the nullspace and hence of $W^\perp$.

4.2.4 Orthogonal Projection

- As we might expect from geometric intuition, if W is a subspace of the (finite-dimensional) inner product space V, we can decompose any vector uniquely as the sum of a component in W with a component in $W^\perp$:
- <u>Theorem</u> (Orthogonal Components): Let V be an inner product space and W be a finite-dimensional subspace. Then every vector $\mathbf{v} \in V$ can be uniquely written in the form $\mathbf{v} = \mathbf{w} + \mathbf{w}^\perp$ for some $\mathbf{w} \in W$ and $\mathbf{w}^\perp \in W^\perp$, and furthermore, we have the Pythagorean relation $||\mathbf{v}||^2 = ||\mathbf{w}||^2 + ||\mathbf{w}^\perp||^2$.

- Proof: First, we show that such a decomposition exists. Since W is finite-dimensional, it has some orthonormal basis $\{\mathbf{e}_1, \ldots, \mathbf{e}_k\}$.
- Now set $\mathbf{w} = \langle \mathbf{v}, \mathbf{e}_1 \rangle \mathbf{e}_1 + \langle \mathbf{v}, \mathbf{e}_2 \rangle \mathbf{e}_2 + \cdots + \langle \mathbf{v}, \mathbf{e}_k \rangle \mathbf{e}_k$, and then $\mathbf{w}^\perp = \mathbf{v} - \mathbf{w}$.
- Clearly $\mathbf{w} \in W$ and $\mathbf{v} = \mathbf{w} + \mathbf{w}^\perp$, so we need only check that $\mathbf{w}^\perp \in W^\perp$.
- To see this, first observe that $\langle \mathbf{w}, \mathbf{e}_i \rangle = \langle \mathbf{v}, \mathbf{e}_i \rangle$ since $\{\mathbf{e}_1, \ldots, \mathbf{e}_k\}$ is an orthonormal basis. Then, we see that $\langle \mathbf{w}^\perp, \mathbf{e}_i \rangle = \langle \mathbf{v} - \mathbf{w}, \mathbf{e}_i \rangle = \langle \mathbf{v}, \mathbf{e}_i \rangle - \langle \mathbf{w}, \mathbf{e}_i \rangle = 0$. Thus, $\mathbf{w}^\perp$ is orthogonal to each vector in the orthonormal basis of W, so it is in $W^\perp$.
- For the uniqueness, suppose we had two decompositions $\mathbf{v} = \mathbf{w}_1 + \mathbf{w}_1^\perp$ and $\mathbf{v} = \mathbf{w}_2 + \mathbf{w}_2^\perp$.
- By subtracting and rearranging, we see that $\mathbf{w}_1 - \mathbf{w}_2 = \mathbf{w}_2^\perp - \mathbf{w}_1^\perp$. Denoting this common vector by $\mathbf{x}$, we see that $\mathbf{x}$ is in both W and $W^\perp$: thus, $\mathbf{x}$ is orthogonal to itself, but the only such vector is the zero vector. Thus, $\mathbf{w}_1 = \mathbf{w}_2$ and $\mathbf{w}_1^\perp = \mathbf{w}_2^\perp$, so the decomposition is unique.
- For the last statement, since $\langle \mathbf{w}, \mathbf{w}^\perp \rangle = 0$, we have $||\mathbf{v}||^2 = \langle \mathbf{v}, \mathbf{v} \rangle = \langle \mathbf{w} + \mathbf{w}^\perp, \mathbf{w} + \mathbf{w}^\perp \rangle = \langle \mathbf{w}, \mathbf{w} \rangle + \langle \mathbf{w}^\perp, \mathbf{w}^\perp \rangle = ||\mathbf{w}||^2 + ||\mathbf{w}^\perp||^2$, as claimed.

- Definition: If V is an inner product space and W is a finite-dimensional subspace with orthonormal basis $\{\mathbf{e}_1, \ldots, \mathbf{e}_k\}$, the orthogonal projection of $\mathbf{v}$ into W is the vector $\text{proj}_W(\mathbf{v}) = \langle \mathbf{v}, \mathbf{e}_1 \rangle \mathbf{e}_1 + \langle \mathbf{v}, \mathbf{e}_2 \rangle \mathbf{e}_2 + \cdots + \langle \mathbf{v}, \mathbf{e}_k \rangle \mathbf{e}_k$.
  - If instead we only have an orthogonal basis $\{\mathbf{u}_1, \ldots, \mathbf{u}_k\}$ of W, the corresponding expression is instead $\text{proj}_W(\mathbf{v}) = \dfrac{\langle \mathbf{v}, \mathbf{u}_1 \rangle}{\langle \mathbf{u}_1, \mathbf{u}_1 \rangle}\mathbf{u}_1 + \dfrac{\langle \mathbf{v}, \mathbf{u}_2 \rangle}{\langle \mathbf{u}_2, \mathbf{u}_2 \rangle}\mathbf{u}_2 + \cdots + \dfrac{\langle \mathbf{v}, \mathbf{u}_2 \rangle}{\langle \mathbf{u}_k, \mathbf{u}_k \rangle}\mathbf{u}_k$.

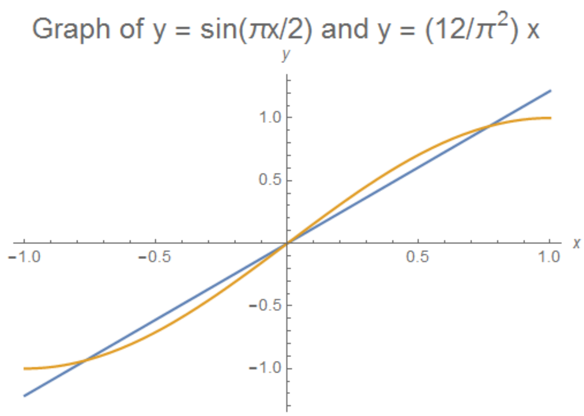
- Example: For $W = \text{span}[(1,0,0), \frac{1}{5}(0,3,4)]$ in $\mathbb{R}^3$ under the standard dot product, compute the orthogonal projection of $\mathbf{v} = (1,2,1)$ into W, and verify the relation $||\mathbf{v}||^2 = ||\mathbf{w}||^2 + ||\mathbf{w}^\perp||^2$.
  - Notice that the vectors $\mathbf{e}_1 = (1,0,0)$ and $\mathbf{e}_2 = \frac{1}{5}(0,3,4)$ form an orthonormal basis for W.
  - Thus, the orthogonal projection is $\mathbf{w} = \text{proj}_W(\mathbf{v}) = \langle \mathbf{v}, \mathbf{e}_1 \rangle \mathbf{e}_1 + \langle \mathbf{v}, \mathbf{e}_2 \rangle \mathbf{e}_2 = 1\,(1,0,0) + 2\,(0,3/5,4/5) = \boxed{(1,6/5,8/5)}$.
  - We see that $\mathbf{w}^\perp = \mathbf{v} - \mathbf{w} = (0,4/5,-3/5)$ is orthogonal to both $(1,0,0)$ and $(0,3/5,4/5)$, so it is indeed in $W^\perp$. Furthermore, $||\mathbf{v}||^2 = 6$, while $||\mathbf{w}||^2 = 5$ and $||\mathbf{w}^\perp||^2 = 1$, so indeed $||\mathbf{v}||^2 = ||\mathbf{w}||^2 + ||\mathbf{w}^\perp||^2$.

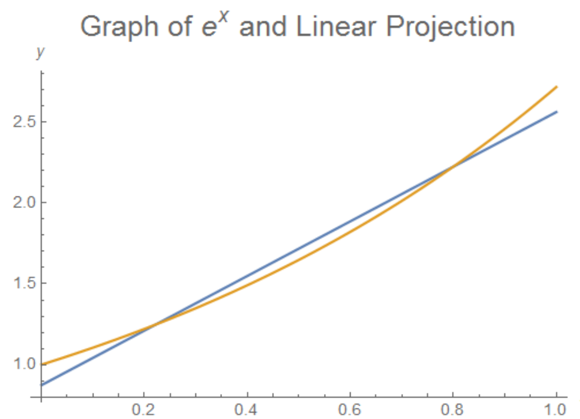
- The orthogonal projection gives the answer to the approximation problem we posed earlier:

- Corollary (Best Approximations): If W is a finite-dimensional subspace of the inner product space V, then for any vector $\mathbf{v}$ in V, the projection of $\mathbf{v}$ into W is closer to $\mathbf{v}$ than any other vector in W. Explicitly, if $\mathbf{w}$ is the projection, then for any other $\mathbf{w}' \in W$, we have $||\mathbf{v} - \mathbf{w}|| < ||\mathbf{v} - \mathbf{w}'||$.
  - Proof: By the theorem on orthogonal complements, we can write $\mathbf{v} = \mathbf{w} + \mathbf{w}^\perp$ where $\mathbf{w} \in W$ and $\mathbf{w}^\perp \in W^\perp$. Now, for any other vector $\mathbf{w}' \in W$, we can write $\mathbf{v} - \mathbf{w}' = (\mathbf{v} - \mathbf{w}) + (\mathbf{w} - \mathbf{w}')$, and observe that $\mathbf{v} - \mathbf{w} = \mathbf{w}^\perp$ is in $W^\perp$, and $\mathbf{w} - \mathbf{w}'$ is in W (since both $\mathbf{w}$ and $\mathbf{w}'$ are, and W is a subspace).
  - Thus, $\mathbf{v} - \mathbf{w}' = (\mathbf{v} - \mathbf{w}) + (\mathbf{w} - \mathbf{w}')$ is a decomposition of $\mathbf{v} - \mathbf{w}'$ into orthogonal vectors. Taking norms, we see that $||\mathbf{v} - \mathbf{w}'||^2 = ||\mathbf{v} - \mathbf{w}||^2 + ||\mathbf{w} - \mathbf{w}'||^2$.
  - Then, if $\mathbf{w}' \neq \mathbf{w}$, since the norm of $||\mathbf{w} - \mathbf{w}'||$ is positive, we conclude that $||\mathbf{v} - \mathbf{w}|| < ||\mathbf{v} - \mathbf{w}'||$.

- Example: Find the best approximation to $\mathbf{v} = (3,-3,3)$ lying in the subspace $W = \text{span}[\frac{1}{3}(1,2,-2), \frac{1}{3}(-2,2,1)]$, where distance is measured under the standard dot product.
  - Notice that the vectors $\mathbf{e}_1 = \frac{1}{3}(1,2,-2)$ and $\mathbf{e}_2 = \frac{1}{3}(-2,2,1)$ form an orthonormal basis for W.
  - Thus, the desired vector, the orthogonal projection, is $\text{proj}_W(\mathbf{v}) = \langle \mathbf{v}, \mathbf{e}_1 \rangle \mathbf{e}_1 + \langle \mathbf{v}, \mathbf{e}_2 \rangle \mathbf{e}_2 = -3\mathbf{e}_1 - 3\mathbf{e}_2 = \boxed{(1,-4,1)}$.

- Example: Find the best approximation to the function $f(x) = \sin(\pi x/2)$ that lies in the subspace $P_2(\mathbb{R})$, under the inner product $\langle f, g\rangle = \int_{-1}^{1} f(x)g(x)\,dx$.
  - First, by applying Gram-Schmidt to the basis $\{1, x, x^2\}$, we can generate an orthogonal basis of $P_2(\mathbb{R})$ under this inner product: the result (after rescaling to eliminate denominators) is $\{1,\ x,\ 3x^2 - 1\}$.
  - Now, with $p_1 = 1$, $p_2 = x$, $p_3 = 3x^2 - 1$ we can compute $\langle f, p_1\rangle = 0$, $\langle f, p_2\rangle = 8/\pi^2$, $\langle f, p_3\rangle = 0$, and also $\langle p_1, p_1\rangle = 2$, $\langle p_2, p_2\rangle = 2/3$, and $\langle p_3, p_3\rangle = 8/5$.
  - Thus, the desired orthogonal projection is $\text{proj}_{P_1(\mathbb{R})}(f) = \dfrac{\langle f, p_1\rangle}{\langle p_1, p_1\rangle}p_1 + \dfrac{\langle f, p_2\rangle}{\langle p_2, p_2\rangle}p_2 + \dfrac{\langle f, p_3\rangle}{\langle p_3, p_3\rangle}p_3 = \boxed{\dfrac{12}{\pi^2}x}$.
  - We can see from a plot of the orthogonal projection polynomial and f on $[-1, 1]$ that this line is a reasonably accurate approximation of $\sin(x)$ on the interval $[-1, 1]$:





- Example: Find the linear polynomial $p(x)$ that minimizes the expression $\int_0^1 (p(x) - e^x)^2\,dx$.
  - Observe that the minimization problem is asking us to find the orthogonal projection of e^x into $P_1(\mathbb{R})$ under the inner product $\langle f, g\rangle = \int_0^1 f(x)g(x)\,dx$.
  - First, by applying Gram-Schmidt to the basis $\{1, x\}$, we can generate an orthogonal basis of $P_1(\mathbb{R})$ under this inner product: the result (after rescaling to clear denominators) is $\{1,\ 2x - 1\}$.
  - Now, with $p_1 = 1$ and $p_2 = 2x - 1$, we can compute $\langle e^x, p_1\rangle = e - 1$, $\langle e^x, p_2\rangle = 3 - e$, $\langle p_1, p_1\rangle = 1$, and $\langle p_2, p_2\rangle = 1/3$.
  - Then $\text{proj}_{P_2(\mathbb{R})}(e^x) = \dfrac{\langle e^x, p_1\rangle}{\langle p_1, p_1\rangle}p_1 + \dfrac{\langle e^x, p_2\rangle}{\langle p_2, p_2\rangle}p_2 = \boxed{(10 - 4e) + (18 - 6e)x \approx 0.873 + 1.690x}$.
  - We can see from a plot (see above) of the orthogonal projection polynomial and e^x on $[0, 1]$ that this line is indeed a very accurate approximation of e^x on the interval $[0, 1]$.

4.3 Applications of Inner Products

- In this section we discuss several practical applications of inner products and orthogonality.

4.3.1 Least-Squares Estimates

- A fundamental problem in applied mathematics and statistics is data fitting: finding a model that well approximates some set of experimental data. Problems of this type are ubiquitous in the physical sciences, social sciences, life sciences, and engineering.
  - A common example is that of finding a linear regression: a line $y = mx + b$ that best fits a set of 2-dimensional data points $\{(x_1, y_1), \dots, (x_n, y_n)\}$ when plotted in the plane.
  - Of course, in many cases a linear model is not appropriate, and other types of models (polynomials, powers, exponential functions, logarithms, etc.) are needed instead.

- The most common approach to such regression analysis is the method of "least squares", which minimizes the sum of the squared errors (the error being the difference between the model and the actual data).

- As we will discuss, many of these questions ultimately reduce to the following: if A is an $m \times n$ matrix such that the matrix equation $A\mathbf{x} = \mathbf{c}$ has no solution, what vector $\hat{\mathbf{x}}$ is the closest approximation to a solution?
  - In other words, we are asking for the vector $\hat{\mathbf{x}}$ that minimizes the vector norm $||A\hat{\mathbf{x}} - \mathbf{c}||$.
  - Since the vectors of the form $A\hat{\mathbf{x}}$ are precisely those in the column space of A, from our analysis of best approximations in the previous section we see that the vector $\mathbf{w} = A\hat{\mathbf{x}}$ will be the projection of $\mathbf{c}$ into the column space of A.
  - Then, by orthogonal decomposition, we know that $\mathbf{w}^{\perp} = \mathbf{c} - A\hat{\mathbf{x}}$ is in the orthogonal complement of the column space of A.
  - Since the column space of A is the same as the row space of A^T, by our theorem on orthogonal complements we know that the orthogonal complement of the column space of A is the nullspace of A^T.
  - Therefore, $\mathbf{w}^{\perp}$ is in the nullspace of A^T, so $A^T\mathbf{w}^{\perp} = \mathbf{0}$.
  - Explicitly, this means $A^T(\mathbf{c} - A\hat{\mathbf{x}}) = \mathbf{0}$, or $A^TA\hat{\mathbf{x}} = A^T\mathbf{c}$: this is an explicit matrix system, called the normal equation, that we can solve for $\hat{\mathbf{x}}$.

- Definition: If A is an $m \times n$ matrix with $m > n$, a least-squares solution to the matrix equation $A\mathbf{x} = \mathbf{c}$ is a vector $\hat{\mathbf{x}}$ satisfying $A^TA\hat{\mathbf{x}} = A^T\mathbf{c}$.
  - The system $A^TA\hat{\mathbf{x}} = A^T\mathbf{c}$ for $\hat{\mathbf{x}}$ is always consistent for any matrix A, although it is possible for there to be infinitely many solutions (a trivial case would be when A is the zero matrix). Even in this case, the orthogonal projection $\mathbf{w} = A\hat{\mathbf{x}}$ onto the column space of A will always be unique.

- In typical cases, the rank of A is often equal to n. In this case, the matrix A^TA will always be invertible, and there is a unique least-squares solution:

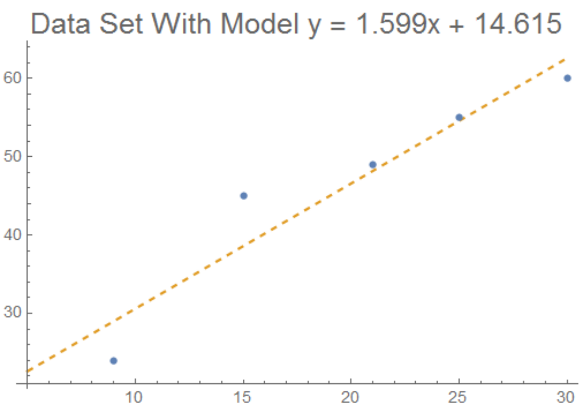
- Proposition (Least-Squares Solution): If A is an $m \times n$ matrix and $\text{rank}(A) = n$, then A^TA is invertible and the unique least-squares solution to $A\mathbf{x} = \mathbf{c}$ is $\hat{\mathbf{x}} = (A^TA)^{-1}A^T\mathbf{c}$.
  - Proof: Suppose that the rank of A is equal to n.
  - We claim that the nullspace of A^TA is the same as the nullspace of A. Clearly, if $A\mathbf{x} = \mathbf{0}$ then $(A^TA)\mathbf{x} = \mathbf{0}$, so it remains to show that if $(A^TA)\mathbf{x} = \mathbf{0}$ then $A\mathbf{x} = \mathbf{0}$.
  - Notice that the dot product $\mathbf{x} \cdot \mathbf{y}$ is the same as the matrix product $\mathbf{y}^T\mathbf{x}$, if $\mathbf{x}$ and $\mathbf{y}$ are column vectors.
  - Now suppose that $(A^TA)\mathbf{x} = \mathbf{0}$: then $\mathbf{x}^T(A^TA)\mathbf{x} = 0$, or equivalently, $(A\mathbf{x})^T(A\mathbf{x}) = 0$.
  - But this means $(A\mathbf{x}) \cdot (A\mathbf{x}) = \mathbf{0}$, so that the dot product of $A\mathbf{x}$ with itself is zero.
  - Since the dot product is an inner product, this means $A\mathbf{x}$ must itself be zero, as required.
  - Thus, the nullspace of A^TA is the same as the nullspace of A. Since the dimension of the nullspace is the number of columns minus the rank, and A^TA and A both have n columns, $\text{rank}(A^TA) = \text{rank}(A) = n$.
  - But since A^TA is an $n \times n$ matrix, this means A^TA is invertible. The second statement then follows immediately upon left-multiplying $A\mathbf{x} = \mathbf{c}$ by $(A^TA)^{-1}$.

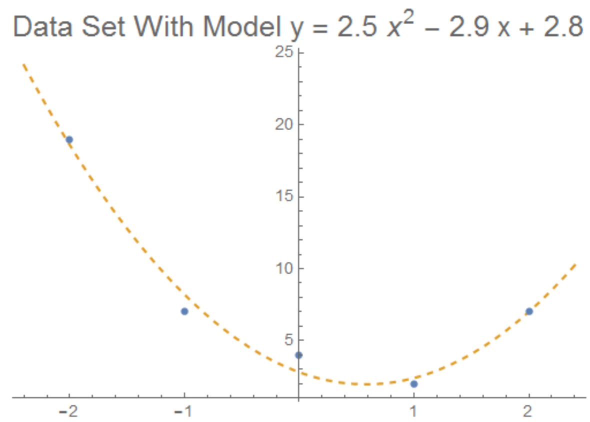
- Example: Find the least-squares solution to the inconsistent system $x + 2y = 3$, $2x + y = 4$, $x + y = 2$.
  - In this case, we have $A = \begin{bmatrix} 1 & 2 \\ 2 & 1 \\ 1 & 1 \end{bmatrix}$ and $\mathbf{c} = \begin{bmatrix} 3 \\ 4 \\ 2 \end{bmatrix}$. Since A clearly has rank 2, A^TA will be invertible and there will be a unique least-squares solution.
  - We compute $A^TA = \begin{bmatrix} 6 & 5 \\ 5 & 6 \end{bmatrix}$, which is indeed invertible and has inverse $(A^TA)^{-1} = \frac{1}{11}\begin{bmatrix} 6 & -5 \\ -5 & 6 \end{bmatrix}$.
  - The least-squares solution is therefore $\hat{\mathbf{x}} = (A^TA)^{-1}A^T\mathbf{c} = \boxed{\begin{bmatrix} 3 \\ -1 \end{bmatrix}}$.

- In this case, we see $A\hat{\mathbf{x}} = \begin{bmatrix} 2 \\ 5 \\ 2 \end{bmatrix}$, so the error vector is $\mathbf{c} - A\hat{\mathbf{x}} = \begin{bmatrix} 2 \\ 1 \\ 0 \end{bmatrix}$. Our analysis above indicates that this error vector has the smallest possible norm.

- We can apply these ideas to the problem of finding an optimal model for a set of data points.
  - For example, suppose that we wanted to find a linear model $y = mx + b$ that fits a set of data points $\{(x_1, y_1), \dots, (x_n, y_n)\}$, in such a way as to minimize the sum of the squared errors $(y_1 - mx_1 - b)^2 + \cdots + (y_n - mx_n - b)^2$.
  - If the data points happened to fit exactly on a line, then we would be seeking the solution to the system $y_1 = mx_1 + b$, $y_2 = mx_2 + b$, ... , $y_n = mx_n + b$.
  - In matrix form, this is the system $A\mathbf{x} = \mathbf{c}$ where $A = \begin{bmatrix} 1 & x_1 \\ 1 & x_2 \\ \vdots & \vdots \\ 1 & x_n \end{bmatrix}$, $\mathbf{x} = \begin{bmatrix} b \\ m \end{bmatrix}$, and $\mathbf{c} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix}$.
  - Of course, due to experimental errors and other random noise, it is unlikely for the data points to fit the model exactly. Instead, the least-squares estimate $\hat{\mathbf{x}}$ will provide the values of m and b that minimize the sum of the squared errors.
  - In a similar way, to find a quadratic model $y = ax^2 + bx + c$ for a data set $\{(x_1, y_1), \dots, (x_n, y_n)\}$, we would use the least-squares estimate for $A\mathbf{x} = \mathbf{c}$, with $A = \begin{bmatrix} 1 & x_1 & x_1^2 \\ 1 & x_2 & x_2^2 \\ \vdots & \vdots & \vdots \\ 1 & x_n & x_n^2 \end{bmatrix}$, $\mathbf{x} = \begin{bmatrix} c \\ b \\ a \end{bmatrix}$, and $\mathbf{c} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix}$.
  - In general, to find a least-squares model of the form $y = a_1 f_1(x) + \cdots + a_m f_m(x)$ for a data set $\{(x_1, y_1), \dots, (x_n, y_n)\}$, we would want the least-squares estimate for the system $A\mathbf{x} = \mathbf{c}$, with $A = \begin{bmatrix} f_1(x_1) & \cdots & f_m(x_1) \\ \vdots & \ddots & \vdots \\ f_1(x_n) & \cdots & f_m(x_n) \end{bmatrix}$, $\mathbf{x} = \begin{bmatrix} a_1 \\ \vdots \\ a_m \end{bmatrix}$, and $\mathbf{c} = \begin{bmatrix} y_1 \\ \vdots \\ y_n \end{bmatrix}$.

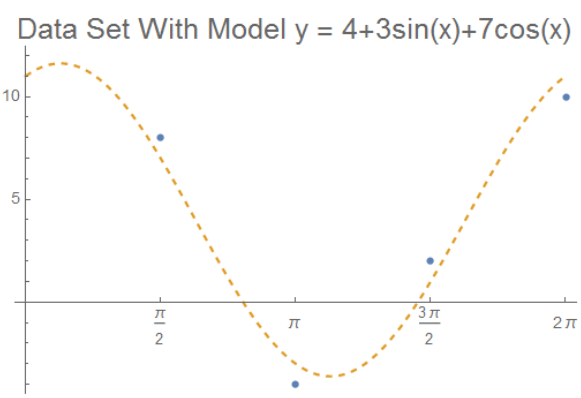
- Example: Use least-squares estimation to find the line $y = mx + b$ that is the best model for the data points $\{(9, 24), (15, 45), (21, 49), (25, 55), (30, 60)\}$.
  - We seek the least-squares solution for $A\mathbf{x} = \mathbf{c}$, where $A = \begin{bmatrix} 1 & 9 \\ 1 & 15 \\ 1 & 21 \\ 1 & 25 \\ 1 & 30 \end{bmatrix}$, $\mathbf{x} = \begin{bmatrix} b \\ m \end{bmatrix}$, and $\mathbf{c} = \begin{bmatrix} 24 \\ 45 \\ 49 \\ 55 \\ 60 \end{bmatrix}$.
  - We compute $A^T A = \begin{bmatrix} 5 & 100 \\ 100 & 2272 \end{bmatrix}$, so the least-squares solution is $\hat{\mathbf{x}} = (A^T A)^{-1} A^T \mathbf{c} \approx \begin{bmatrix} 14.615 \\ 1.599 \end{bmatrix}$.
  - Thus, to three decimal places, the desired line is $y = \boxed{1.599x + 14.615}$. From a plot, we can see that this line is fairly close to all of the data points:





- Example: Use least-squares estimation to find the quadratic function $y = ax^2 + bx + c$ best modeling the data points $\{(-2, 19), (-1, 7), (0, 4), (1, 2), (2, 7)\}$.
  - We seek the least-squares solution for $A\mathbf{x} = \mathbf{c}$, with $A = \begin{bmatrix} 1 & -2 & 4 \\ 1 & -1 & 1 \\ 1 & 0 & 0 \\ 1 & 1 & 1 \\ 1 & 2 & 4 \end{bmatrix}$, $\mathbf{x} = \begin{bmatrix} c \\ b \\ a \end{bmatrix}$, $\mathbf{c} = \begin{bmatrix} 19 \\ 7 \\ 4 \\ 2 \\ 7 \end{bmatrix}$.
  - We compute $A^T A = \begin{bmatrix} 5 & 0 & 10 \\ 0 & 10 & 0 \\ 10 & 0 & 34 \end{bmatrix}$, so the least-squares solution is $\hat{\mathbf{x}} = (A^T A)^{-1} A^T \mathbf{c} = \begin{bmatrix} 2.8 \\ -2.9 \\ 2.5 \end{bmatrix}$.
  - Thus, the desired quadratic polynomial is $y = \boxed{-2.5x^2 - 2.9x + 2.8}$. From a plot (see above), we can see that this quadratic function is fairly close to all of the data points.

- Example: Use least-squares estimation to find the trigonometric function $y = a + b\sin(x) + c\cos(x)$ best modeling the data points $\{(\pi/2, 8), (\pi, -4), (3\pi/2, 2), (2\pi, 10)\}$.
  - We seek the least-squares solution for $A\mathbf{x} = \mathbf{c}$, with $A = \begin{bmatrix} 1 & 1 & 0 \\ 1 & 0 & -1 \\ 1 & -1 & 0 \\ 1 & 0 & 1 \end{bmatrix}$, $\mathbf{x} = \begin{bmatrix} a \\ b \\ c \end{bmatrix}$, $\mathbf{c} = \begin{bmatrix} 8 \\ -4 \\ 2 \\ 10 \end{bmatrix}$.
  - We compute $A^T A = \begin{bmatrix} 4 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 2 \end{bmatrix}$, so the least-squares solution is $\hat{\mathbf{x}} = (A^T A)^{-1} A^T \mathbf{c} = \begin{bmatrix} 4 \\ 3 \\ 7 \end{bmatrix}$.
  - Thus, the desired function is $y = \boxed{4 + 3\sin(x) + 7\cos(x)}$. In this case, the model predicts the points $\{(\pi/2, 7), (\pi, -3), (3\pi/2, 1), (2\pi, 11)\}$, so it is a good fit to the original data:



- Our results on least squares from above also yield a method for writing down the matrix associated to orthogonal projection onto a subspace S of $\mathbb{R}^n$ with respect to the standard basis:

- Corollary (Associated Matrices for Projections): Suppose S is a subspace of $\mathbb{R}^n$ with basis $\{\mathbf{v}_1, \dots, \mathbf{v}_k\}$. If β is the standard basis for $\mathbb{R}^n$ and $T : \mathbb{R}^n \to \mathbb{R}^n$ represents orthogonal projection onto the subspace S, then the associated matrix $[T]_\beta^\beta = A(A^T A)^{-1} A^T$, where A is the $n \times k$ matrix whose columns are the vectors $\mathbf{v}_i$.
  - Proof: By our earlier results, for any column vector $\mathbf{c}$, the unique least-squares solution to $A\mathbf{x} = \mathbf{c}$ is $\hat{\mathbf{x}} = (A^T A)^{-1} A^T \mathbf{c}$, and the vector $A\hat{\mathbf{x}} = A(A^T A)^{-1} A^T \mathbf{c}$ represents the projection of $\mathbf{c}$ into the column space of A.
  - But the column space of the matrix A is precisely S, by definition, so the associated matrix $[T]_\beta^\beta$ is precisely $A(A^T A)^{-1} A^T$, as claimed.

- Example: Find the matrix M (with respect to the standard basis of $\mathbb{R}^2$) associated to orthogonal projection onto the subspace S spanned by $\{(1, 2)\}$ inside $\mathbb{R}^2$.
  - By the corollary above, the associated matrix is $A(A^T A)^{-1} A^T$ where $A = \begin{bmatrix} 1 \\ 2 \end{bmatrix}$.
  - We compute $(A^T A)^{-1} = [1/5]$ and so $M = A(A^T A)^{-1} A^T = \boxed{\frac{1}{5}\begin{bmatrix} 1 & 2 \\ 2 & 4 \end{bmatrix}}$.
  - Remark: Notice that this is simply orthogonal projection onto the line $y = 2x$ in the plane, whose associated matrix we also calculated in the previous chapter in a more geometric way.

- Example: Find the matrix M (with respect to the standard basis of $\mathbb{R}^4$) associated to orthogonal projection onto the subspace S spanned by $\{(1, 1, 0, 0), (-1, 1, 1, 1)\}$ inside $\mathbb{R}^4$.
  - By the corollary above, the associated matrix is $A(A^T A)^{-1} A^T$ where $A = \begin{bmatrix} 1 & -1 \\ 1 & 1 \\ 0 & 1 \\ 0 & 1 \end{bmatrix}$.
  - We compute $(A^T A)^{-1} = \begin{bmatrix} 1/2 & 0 \\ 0 & 1/4 \end{bmatrix}$ and so $M = A(A^T A)^{-1} A^T = \boxed{\frac{1}{4}\begin{bmatrix} 3 & 1 & -1 & -1 \\ 1 & 3 & 1 & 1 \\ -1 & 1 & 1 & 1 \\ -1 & 1 & 1 & 1 \end{bmatrix}}$.
  - To verify that this really is the correct matrix, we can check that the column space of M is equal to S (which it is), and that M sends each vector in S to itself (which it does).

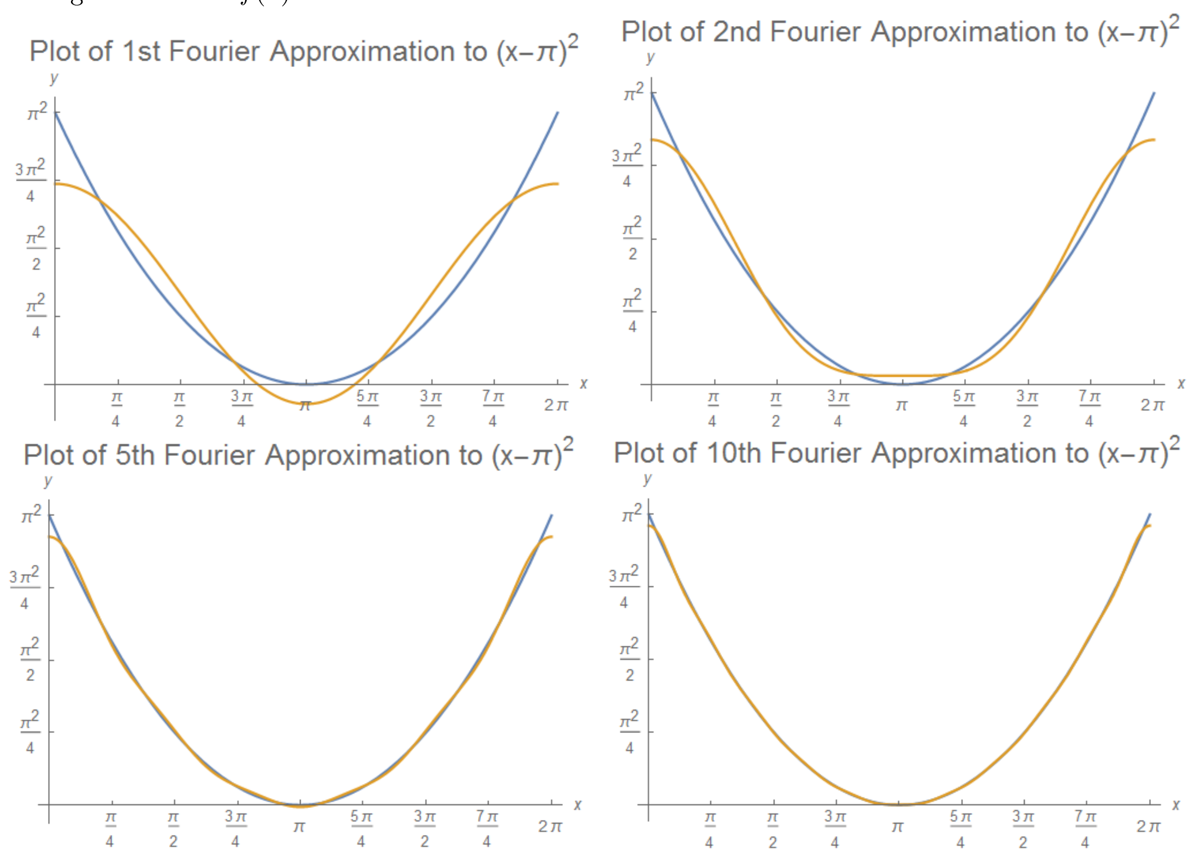
- Remark: From the formula for the associated matrix above, or even just from the definition, we can see that if P is any orthogonal projection, then $P^2 = P$.
  - In fact, this property also characterizes orthogonal projections: if $P : V \to V$ is any linear transformation such that $P^2 = P$, then P is an orthogonal projection onto its image $W = \text{im}(P)$.
  - To see this, suppose $\mathbf{v} \in V$ is any vector: we claim that the orthogonal decomposition $\mathbf{v} = \mathbf{w} + \mathbf{w}^\perp$ has $\mathbf{w} = P(\mathbf{v}) \in W$ and $\mathbf{w}^\perp = \mathbf{v} - P(\mathbf{v})$.
  - Clearly $P(\mathbf{v}) \in W$ by definition, and we also have $P(\mathbf{v} - P(\mathbf{v})) = P(\mathbf{v}) - P^2(\mathbf{v}) = 0$: this means $\mathbf{v} - P(\mathbf{v}) \in \ker(P) = W^\perp$, as required.

- In addition to least-squares methods (which are extremely important in statistics and the experimental sciences), orthogonal projections also have many applications in computer graphics, coding theory, and machine learning.
  - To illustrate the idea very generally, if we have a set of data that is high-dimensional (i.e., lies inside $\mathbb{R}^n$ where n is very large) that has a lot of underlying structure, it is often the case that projecting onto a much smaller-dimensional subspace will not lose very much information. Storing the projection of the data then requires much less information, which is the central idea of data compression.
  - To minimize the loss of information when compressing data, one may use tools such as principal-component analysis, which provide ways to calculate subspaces that carry as much of the information from the original data set as possible.

4.3.2 Fourier Series

- Another extremely useful application of the general theory of orthonormal bases is that of Fourier series.
  - Fourier analysis, broadly speaking, studies the problem of approximating a function on an interval by trigonometric functions. This problem is very similar to the question, studied in calculus, of approximating a function by a polynomial (the typical method is to use Taylor polynomials, although as we have already discussed, least-squares estimates provide another potential avenue).
  - Fourier series have a tremendously wide variety of applications, ranging from to solving partial differential equations (in particular, the famous wave equation and heat equation), studying acoustics and optics (decomposing an acoustic or optical waveform into simpler waves of particular frequencies), electrical engineering, and quantum mechanics.
- Although a full discussion of Fourier series belongs more properly to analysis, we can give some of the ideas.
  - A typical scenario in Fourier analysis is to approximate a continuous function on $[0, 2\pi]$ using a trigonometric polynomial: a function that is a polynomial in $\sin(x)$ and $\cos(x)$.
  - Using trigonometric identities, this question is equivalent to approximating a function $f(x)$ by a (finite) <u>Fourier series</u> of the form $s(x) = a_0 + b_1 \cos(x) + b_2 \cos(2x) + \cdots + b_k \cos(kx) + c_1 \sin(x) + c_2 \sin(2x) + \cdots + c_k \sin(kx)$.
  - Notice that, in the expression above, $s(0) = s(2\pi)$ since each function in the sum has period 2π. Thus, we can only realistically hope to get close approximations to functions satisfying $f(0) = f(2\pi)$.
- Let V be the vector space of continuous, real-valued functions on the interval $[0, 2\pi]$ having equal values at 0 and 2π, and define an inner product on V via $\langle f, g\rangle = \int_0^{2\pi} f(x)g(x)\,dx$.
- <u>Proposition</u>: The functions $\{\varphi_0, \varphi_1, \varphi_2, \dots\}$ are an orthonormal set on V, where $\varphi_0(x) = \frac{1}{\sqrt{2\pi}}$, and $\varphi_{2k-1}(x) = \frac{1}{\sqrt{\pi}}\cos(kx)$ and $\varphi_{2k}(x) = \frac{1}{\sqrt{\pi}}\sin(kx)$ for each $k \geq 1$.
  - <u>Proof</u>: Using the product-to-sum identities, such as $\sin(ax)\sin(bx) = \frac{1}{2}[\cos(a-b)x - \cos(a+b)x]$, it is a straightforward exercise in integration to verify that $\langle \varphi_i, \varphi_j\rangle = 0$ for each $i \neq j$.
  - Furthermore, we have $\langle \varphi_0, \varphi_0\rangle = \frac{1}{2\pi}\int_0^{2\pi} 1dx = 1$, $\langle \varphi_{2k-1}, \varphi_{2k-1}\rangle = \frac{1}{\pi}\int_0^{2\pi} \cos^2(kx)\,dx = 1$, and $\langle \varphi_{2k}, \varphi_{2k}\rangle = \frac{1}{\pi}\int_0^{2\pi} \sin^2(kx)\,dx = 1$. Thus, the set is orthonormal.
- If it were the case that $S = \{\varphi_0, \varphi_1, \varphi_2, \dots\}$ were an orthonormal basis for V, then, given any other function $f(x)$ in V, we could write f as a linear combination of functions in $\{\varphi_0, \varphi_1, \varphi_2, \dots\}$, where we can compute the appropriate coefficients using the inner product on V.
  - Unfortunately, S does not span V: we cannot, for example, write the function $g(x) = \sum_{n=1}^{\infty} \frac{1}{2^n}\sin(nx)$ as a finite linear combination of $\{\varphi_0, \varphi_1, \varphi_2, \dots\}$, since doing so would require each of the infinitely many terms in the sum.
  - Ultimately, the problem, as exemplified by the function $g(x)$ above, is that the definition of "basis" only allows us to write down finite linear combinations.
  - On the other hand, the finite sums $\sum_{j=0}^{k} a_j\varphi_j(x)$ for $k \geq 0$, where $a_j = \langle f, \varphi_j\rangle$, will represent the best approximation to $f(x)$ inside the subspace of V spanned by $\{\varphi_0, \varphi_1, \dots, \varphi_k\}$. Furthermore, as we increase k, we are taking approximations to f that lie inside larger and larger subspaces of V, so as we take $k \to \infty$, these partial sums will yield better and better approximations to f.
  - Provided that f is a sufficiently nice function, it can be proven that in the limit, our formulas for the coefficients do give a formula for $f(x)$ as an infinite sum:

- Theorem (Fourier Series): Let $f(x)$ be a twice-differentiable function on $[0, 2\pi]$ satisfying $f(0) = f(2\pi)$, and define the Fourier coefficients of f as $a_j = \langle f, \varphi_j \rangle = \int_0^{2\pi} f(x)\varphi_j(x)\,dx$, for the trigonometric functions $\varphi_j(x)$ defined above. Then $f(x)$ is equal to its Fourier series $\sum_{j=0}^{\infty} a_j \varphi_j(x)$ for every x in $[0, 2\pi]$.
  - This result can be interpreted as a "limiting version" of the theorem we stated earlier giving the coefficients for the linear combination of a vector in terms of an orthonormal basis: it gives an explicit way to write the function $f(x)$ as an "infinite linear combination" of the orthonormal basis elements $\{\varphi_0, \varphi_1, \varphi_2, \dots\}$.
- Example: Compute the Fourier coefficients and Fourier series for $f(x) = (x - \pi)^2$ on the interval $[0, 2\pi]$, and compare the partial sums of the Fourier series to the original function.
  - First, we have $a_0 = \int_0^{2\pi} f(x) \frac{1}{\sqrt{2\pi}}\,dx = \frac{1}{\sqrt{18}}\pi^{5/2}$.
  - For k odd, after integrating by parts twice, we have $a_{2k-1} = \int_0^{2\pi} f(x)\,\frac{1}{\sqrt{\pi}}\cos(kx)\,dx = \frac{4\sqrt{\pi}}{k^2}$.
  - For k even, in a similar manner we see $a_{2k} = \int_0^{2\pi} f(x)\,\frac{1}{\sqrt{\pi}}\sin(kx)\,dx = 0$.
  - Therefore, the Fourier series for $f(x)$ is $\boxed{\frac{1}{6}\pi^2 + \sum_{k=1}^{\infty} \frac{4}{k^2}\cos(kx)}$.
  - Here are some plots of the partial sums (up to the term involving $\cos(nx)$) of the Fourier series along with f. As is clearly visible from the graphs, the partial sums give increasingly close approximations to the original function $f(x)$ as we sum more terms:



Well, you're at the end of my handout. Hope it was helpful.