Inner product spaces  Let $V$ be a real vector space.

- An inner product on $V$ is a symmetric, positive-definite [bilinear] 2-form on $V$,
  i.e., a function that maps each pair of vectors $\vec{v}, \vec{w} \in V$ to their inner product $\langle \vec{v}, \vec{w} \rangle \in \mathbb{R}$, with the following properties:

  - Bilinear: $\langle \alpha \vec{v}, \vec{w} \rangle = \alpha \langle \vec{v}, \vec{w} \rangle = \langle \vec{v}, \alpha \vec{w} \rangle$  
    ($\forall \vec{v}, \vec{w} \in V$ and $\alpha \in \mathbb{R}$)

  $\langle \vec{v}_1 + \vec{v}_2, \vec{w} \rangle = \langle \vec{v}_1, \vec{w} \rangle + \langle \vec{v}_2, \vec{w} \rangle$  
    ($\forall \vec{v}_1, \vec{v}_2, \vec{w} \in V$)

  $\langle \vec{v}, \vec{v}_1 + \vec{v}_2 \rangle = \langle \vec{v}, \vec{v}_1 \rangle + \langle \vec{v}, \vec{v}_2 \rangle$  
    ($\forall \vec{v}, \vec{v}_1, \vec{v}_2 \in V$)

  - Symmetric: $\forall \vec{v}, \vec{w} \in V$, $\langle \vec{v}, \vec{w} \rangle = \langle \vec{w}, \vec{v} \rangle$

  - Positive-definite: $\langle \vec{v}, \vec{v} \rangle \geq 0$  
    ($\forall \vec{v} \in V$)

  $\langle \vec{v}, \vec{v} \rangle = 0 \Rightarrow \vec{v} = \vec{0}$

  - $\langle \vec{v}, \vec{w} \rangle$ can be intuitively interpreted as the degree of “agreement” of the vectors $\vec{v}$ and $\vec{w}$.

- Given an inner product $\langle \cdot, \cdot \rangle$ on $V$, we define the norm of a vector $\vec{v} \in V$ by $\| \vec{v} \| \overset{\text{def}}{=} \sqrt{\langle \vec{v}, \vec{v} \rangle}$.

  This norm has the following properties:

  - Absolutely homogeneous: $\| \alpha \vec{v} \| = | \alpha | \| \vec{v} \|$  
    ($\forall \vec{v} \in V$ and $\alpha \in \mathbb{R}$)

  - Positive-definite: $\| \vec{v} \| \geq 0$  
    ($\forall \vec{v} \in V$)

  $\| \vec{v} \| = 0 \Rightarrow \vec{v} = \vec{0}$

  - $\| \vec{v} \|$ is geometrically interpreted as the “length” of $\vec{v}$, with $\| \vec{v} - \vec{w} \|$ representing the distance between $\vec{v}$ and $\vec{w}$.

- An inner product space is a real vector space $V$ equipped with an inner product $\langle \cdot, \cdot \rangle$.

  - In an inner product space, we have not only the linear algebra concepts arising from linear combinations, but also notions such as distance, length, angles, projection, and approximation.

  - Two common inner product spaces:

    - $\mathbb{R}^m$, equipped with the dot product $\begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_m \end{bmatrix} \cdot \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_m \end{bmatrix} \overset{\text{def}}{=} x_1 y_1 + x_2 y_2 + \cdots + x_m y_m$

      and the resulting norm $\left\| \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_m \end{bmatrix} \right\| = \sqrt{x_1^2 + x_2^2 + \cdots + x_m^2}$.

    - $C([a, b])$, equipped with the inner product $\langle f, g \rangle \overset{\text{def}}{=} \int_a^b f g$  
      (or some scaling thereof)  
      and the resulting norm $\| f \| = \sqrt{\int_a^b f^2}$.

  - Two important inequalities hold in any inner product space:

    - The Cauchy-Schwarz Inequality: $| \langle \vec{v}, \vec{w} \rangle | \leq \| \vec{v} \| \| \vec{w} \|$.  

    - The Triangle Inequality: $\| \vec{v} + \vec{w} \| \leq \| \vec{v} \| + \| \vec{w} \|$, and its counterpart, $\| \vec{v} + \vec{w} \| \geq \| \vec{v} \| - \| \vec{w} \|$.  

The transpose  In $\mathbb{R}^m$, we can transpose matrices and vectors by turning their columns into rows.

  - The transpose is linear: $(\alpha A)^T = \alpha (A^T)$  
    and $(A + B)^T = A^T + B^T$.

  - The transpose reverses the order of compositions: $(AB)^T = B^T A^T$.

... and the dot product

Suppose that:  
- $\vec{x}, \vec{y} \in \mathbb{R}^m$.  
- $A = \begin{bmatrix} \vec{a}_1 & \cdots & \vec{a}_n \end{bmatrix}$ is an $m \times n$ matrix.  
- $B = \begin{bmatrix} \vec{b}_1 & \cdots & \vec{b}_p \end{bmatrix}$ is an $m \times p$ matrix.

The transpose relates to the dot product in the following ways:

<table>
<thead>
<tr>
<th>Vector-Vector</th>
<th>Matrix-Vector</th>
<th>Matrix-Matrix</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\vec{x}^T \vec{y} = \vec{x} \cdot \vec{y}$</td>
<td>$A^T \vec{x} = \begin{bmatrix} \vec{a}_1 \cdot \vec{x} \ \vdots \ \vec{a}_n \cdot \vec{x} \end{bmatrix}$</td>
<td>$A^T B = \begin{bmatrix} \vec{a}_1 \cdot \vec{b}_1 &amp; \cdots &amp; \vec{a}_1 \cdot \vec{b}_p \ \vdots &amp; \ddots &amp; \vdots \ \vec{a}_n \cdot \vec{b}_1 &amp; \cdots &amp; \vec{a}_n \cdot \vec{b}_p \end{bmatrix}$</td>
</tr>
</tbody>
</table>
Orthogonality and vector projection  Suppose that $V$ is an inner product space.

- **Orthogonality:**
  - vector-vector: if $\vec{v}, \vec{w} \in V$, then $\vec{v} \perp \vec{w}$ means $\langle \vec{v}, \vec{w} \rangle = 0$, read “$\vec{v}$ and $\vec{w}$ are orthogonal”.
  - vector-subspace: if $\vec{v} \in V$ and $W$ is a subspace of $V$, then $\vec{v} \perp W$ means $\forall \vec{w} \in W$, $\vec{v} \perp \vec{w}$.
  - subspace-subspace: if $W, W'$ are subspaces of $V$, then $W \perp W'$ means $\forall \vec{w} \in W$ and $\vec{u}' \in W'$, $\vec{w} \perp \vec{u}'$.

- $\vec{v} \in V$ is a **unit vector** means $\|\vec{v}\| = 1$, i.e., $\langle \vec{v}, \vec{v} \rangle = 1$.
- Any nonzero vector $\vec{v} \in V$ can be **normalized**, i.e., converted into a unit vector in the same direction, via $\vec{v} \rightsquigarrow \frac{1}{\|\vec{v}\|} \vec{v}$.
- The Pythagorean Theorem holds for orthogonal vectors in any inner product space: $\vec{v} \perp \vec{w} \iff \|\vec{v} + \vec{w}\|^2 = \|\vec{v}\|^2 + \|\vec{w}\|^2$.
  - Important consequence: if $\vec{v} \perp W$, then $\forall \vec{w} \in W$, $\|\vec{v} + \vec{w}\| \geq \|\vec{v}\|$.  

- The **vector projection** of $\vec{v} \in V$ onto a nonzero vector $\vec{w} \in V$ is defined by $\text{proj}_{\vec{w}} \vec{v} = \frac{\langle \vec{v}, \vec{w} \rangle}{\langle \vec{w}, \vec{w} \rangle} \vec{w}$.
  - $\text{proj}_{\vec{w}} \vec{v}$ is the multiple of $\vec{w}$ that best approximates $\vec{v}$.
  - Proven by minimizing the function $f(\alpha) = \|\vec{v} - \alpha \vec{w}\|^2$.
  - Defining $\text{orth}_{\vec{w}} \vec{v} = \vec{v} - \text{proj}_{\vec{w}} \vec{v}$, we find that $\text{orth}_{\vec{w}} \vec{v} \perp \vec{w}$; thus, $\vec{v} = \text{proj}_{\vec{w}} \vec{v} + \text{orth}_{\vec{w}} \vec{v}$ expresses any vector $\vec{v}$ as a multiple of $\vec{w}$ plus a vector orthogonal to $\vec{w}$.
  - In the case that $\vec{w}$ is a unit vector, the projection formula $\text{proj}_{\vec{w}} \vec{v}$ simplifies to $\langle \vec{v}, \vec{w} \rangle \vec{w}$.

**Orthonormal collections**

A collection $\mathcal{U} = \{\vec{u}_1, \vec{u}_2, \ldots, \vec{u}_k\}$ is **orthonormal** means that

$$\forall i, j, \langle \vec{u}_i, \vec{u}_j \rangle = \delta_{ij} \overset{\text{def}}{=} \begin{cases} 1, & \text{if } i = j \\ 0, & \text{if } i \neq j \end{cases}$$

in other words, $\mathcal{U}$ is orthonormal if the vectors of $\mathcal{U}$ are mutually orthogonal unit vectors.

- If $\mathcal{C}$ is an orthonormal collection, then $\mathcal{C}$ is linearly independent.
- Any finite collection $\mathcal{C}$ in $V$ can be converted into an orthonormal collection $\mathcal{U} = \{\vec{u}_1, \vec{u}_2, \ldots, \vec{u}_k\}$ having the same span (i.e., an orthonormal basis for span $\mathcal{C}$) by the Gram-Schmidt Process:
  
  For each vector $\vec{v}$ of $\mathcal{C}$ in turn:
  - Subtract its projections onto the vectors of $\mathcal{U}$ so far obtained: $\vec{v} \rightsquigarrow \vec{v}' = \vec{v} - \langle \vec{v}, \vec{u}_1 \rangle \vec{u}_1 - \langle \vec{v}, \vec{u}_2 \rangle \vec{u}_2 - \ldots$
  - If $\vec{v}' = \vec{0}$, toss it out; otherwise, normalize it and append it to the orthonormal collection: $\frac{1}{\|\vec{v}'\|} \vec{v}' \perp \mathcal{U}$.
  
  Tip: Keep any messy scalars resulting from normalization outside the vectors, rather than distributing them to the entries.
- Some useful orthonormal collections:
  - The standard basis $\{\vec{e}_1, \vec{e}_2, \ldots, \vec{e}_n\}$ for $\mathbb{R}^n$ is orthonormal (with respect to the dot product).
  - The collection $\{\frac{1}{\sqrt{2}}, \cos x, \sin x, \cos 2x, \sin 2x, \ldots\}$ is orthonormal in $C([-\pi, \pi])$
    (with respect to the inner product $\langle f, g \rangle = \frac{1}{\pi} \int_{-\pi}^{\pi} f \, g$).
Projection and orthonormal collections

Suppose that \( W = \text{span} \{ \vec{w}_1, \vec{w}_2, \ldots, \vec{w}_n \} \) is the subspace spanned by an orthonormal collection in an inner product space \( V \).

- The orthogonal projection of \( \vec{v} \in V \) onto \( W \) is given by \( \text{proj}_W \vec{v} = \frac{\langle \vec{v}, \vec{w}_1 \rangle}{\langle \vec{w}_1, \vec{w}_1 \rangle} \vec{w}_1 + \frac{\langle \vec{v}, \vec{w}_2 \rangle}{\langle \vec{w}_2, \vec{w}_2 \rangle} \vec{w}_2 + \cdots + \frac{\langle \vec{v}, \vec{w}_n \rangle}{\langle \vec{w}_n, \vec{w}_n \rangle} \vec{w}_n \).
- Defining \( \text{orth}_W \vec{v} = \vec{v} - \text{proj}_W \vec{v} \), we find that \( \text{orth}_W \vec{v} \perp W \); thus, \( \vec{v} = \text{proj}_W \vec{v} + \text{orth}_W \vec{v} \) expresses any vector \( \vec{v} \) as a vector of \( W \) plus a vector orthogonal to \( W \).
- \( \text{proj}_W \vec{v} \) is the vector of \( W \) that best approximates \( \vec{v} \).
  · Proven via the corollary to the Pythagorean Theorem and the fact that \( \text{orth}_W \vec{v} \perp W \).

- Example: Fourier polynomials for continuous functions on \( [-\pi, \pi] \).
  - The collection \( \mathcal{F}_n = \{ \frac{1}{\sqrt{2}}, \cos x, \sin x, \cos 2x, \sin 2x, \ldots, \cos nx, \sin nx \} \) is orthonormal in \( C([-\pi, \pi]) \) (with respect to the inner product \( \langle f, g \rangle = \frac{1}{\pi} \int_{-\pi}^{\pi} f(x)g(x) \, dx \)).
  - Given \( f(x) \in C([-\pi, \pi]) \), the \( n \text{th}-\)order Fourier polynomial for \( f(x) \) is:

\[
\text{proj}_{\text{span} \mathcal{F}_n} f(x) = \left( f(x), \frac{1}{\sqrt{2}} \right) \frac{1}{\sqrt{2}} + \langle f(x), \cos x \rangle \cos x + \langle f(x), \sin x \rangle \sin x + \cdots + \langle f(x), \cos nx \rangle \cos nx + \langle f(x), \sin nx \rangle \sin nx
\]

This function is the linear combination of such sines and cosines that best approximates \( f(x) \) on \( [-\pi, \pi] \).

Projection matrices and approximation for linear systems

Suppose that \( A \) is an \( m \times n \) matrix with linearly independent columns.

- \( A^T A \) is invertible, so we can define the projection matrix \( P = A(A^T A)^{-1} A^T : \mathbb{R}^m \rightarrow \mathbb{R}^m \).
  - \( P \) maps vectors orthogonal to \( C(A) \) to zero: \( \forall \vec{y} \perp C(A), \quad P\vec{y} = \vec{0} \).
  - \( P \) maps vectors in \( C(A) \) to themselves: \( \forall \vec{y} \in C(A), \quad P\vec{y} = \vec{y} \).
  - Applying these results to any vector \( \vec{v} = \text{proj}_{C(A)} \vec{v} + \text{orth}_{C(A)} \vec{v} \) shows us that \( P\vec{v} = \text{proj}_{C(A)} \vec{v} \).
  - Thus, given any basis for a subspace of \( \mathbb{R}^m \) (not necessarily orthonormal), \( P \) allows us to project any vector of \( \mathbb{R}^m \) onto their span (by placing the basis vectors into the matrix \( A \) and computing \( P \)).

- This allows us find the best approximate solution to an inconsistent system \([ A \mid \vec{b} ]\), as follows:
  - Even if the linear system \( A\vec{x} = \vec{b} \) is inconsistent, the system \( A\vec{x} = P\vec{b} \) is consistent because \( P\vec{b} \in C(A) \).
  - From this, we can find the best approximate solution to our original system by transforming it as follows:

\[
[A \mid \vec{b}] \sim [A^T A \mid A^T \vec{b}]
\]

- Application: Curve-Fitting

We can find the best-fitting curve (line, polynomial, exponential, etc.) to some given set of data, simply by using the data points to set up a [possibly inconsistent] system for the coefficients involved, then using the above method to find the best approximate solution.