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: $\langle \alpha\vec{v}, \vec{w} \rangle = \alpha \langle \vec{v}, \vec{w} \rangle = \langle \vec{v}, \alpha\vec{w} \rangle$ $\quad (\forall\, \vec{v}, \vec{w} \in V \text{ and } \alpha \in \mathbb{R})$
    $\langle \vec{v}_1 + \vec{v}_2, \vec{w} \rangle = \langle \vec{v}_1, \vec{w} \rangle + \langle \vec{v}_2, \vec{w} \rangle$ $\quad (\forall\, \vec{v}_1, \vec{v}_2, \vec{w} \in V)$
    $\langle \vec{v}, \vec{w}_1 + \vec{w}_2 \rangle = \langle \vec{v}, \vec{w}_1 \rangle + \langle \vec{v}, \vec{w}_2 \rangle$ $\quad (\forall\, \vec{v}, \vec{w}_1, \vec{w}_2 \in V)$
  - Symmetric: $\forall\, \vec{v}, \vec{w} \in V,\ \langle \vec{v}, \vec{w} \rangle = \langle \vec{w}, \vec{v} \rangle$
  - Positive-definite: $\langle \vec{v}, \vec{v} \rangle \geq 0$ $\quad (\forall\, \vec{v} \in V)$
    $\langle \vec{v}, \vec{v} \rangle = 0 \Rightarrow \vec{v} = \vec{0}$
  - $\langle \vec{v}, \vec{w} \rangle$ 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: $\|\alpha\vec{v}\| = |\alpha|\, \|\vec{v}\|$ $\quad (\forall\, \vec{v} \in V \text{ and } \alpha \in \mathbb{R})$
  - Positive-definite: $\|\vec{v}\| \geq 0$ $\quad (\forall\, \vec{v} \in V)$
    $\|\vec{v}\| = 0 \Rightarrow \vec{v} = \vec{0}$
  - $\|\vec{v}\|$ 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* $\begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_m \end{bmatrix} \bullet \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_m \end{bmatrix} \overset{\text{def}}{=} x_1y_1 + x_2y_2 + \cdots + x_my_m$
      and the resulting norm $\left\| \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_m \end{bmatrix} \right\| = \sqrt{x_1^2 + x_2^2 + \cdots + x_m^2}$.
    - $\mathrm{C}([a,b])$, equipped with the inner product $\langle f, g \rangle \overset{\text{def}}{=} \int_a^b fg$ (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, $\|\vec{v} + \vec{w}\| \geq \big|\, \|\vec{v}\| - \|\vec{w}\| \,\big|$.

The transpose

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

- The transpose is linear: $(\alpha A)^{\mathrm{T}} = \alpha(A^{\mathrm{T}})$ and $(A+B)^{\mathrm{T}} = A^{\mathrm{T}} + B^{\mathrm{T}}$.
- The transpose *reverses* the order of compositions: $(AB)^{\mathrm{T}} = B^{\mathrm{T}}A^{\mathrm{T}}$.

...and the dot product

Suppose that: ▪ $\vec{x}, \vec{y} \in \mathbb{R}^m$. ▪ $A = \left[\vec{a}_1 \middle| \cdots \middle| \vec{a}_n \right]$ is an $m \times n$ matrix. ▪ $B = \left[\vec{b}_1 \middle| \cdots \middle| \vec{b}_p \right]$ is an $m \times p$ matrix.

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

Vector−Vector

$$\vec{x}^{\mathrm{T}}\vec{y} = \vec{x} \bullet \vec{y}$$

Matrix−Vector

$$A^{\mathrm{T}}\vec{x} = \begin{bmatrix} \vec{a}_1 \bullet \vec{x} \\ \vdots \\ \vec{a}_n \bullet \vec{x} \end{bmatrix}$$

Matrix−Matrix

$$A^{\mathrm{T}}B = \left[\begin{matrix} \vec{a}_1 \bullet \vec{b}_1 \\ \vdots \\ \vec{a}_n \bullet \vec{b}_1 \end{matrix} \middle| \cdots \middle| \begin{matrix} \vec{a}_1 \bullet \vec{b}_p \\ \vdots \\ \vec{a}_n \bullet \vec{b}_p \end{matrix} \right]$$

Orthogonality

Math 200
Swenton

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 $\boxed{\langle \vec{v}, \vec{w} \rangle = 0}$, 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 $\boxed{\forall\, \vec{w} \in W,\ \vec{v} \perp \vec{w}}$.
  - subspace-subspace: if W, W' are subspaces of V, then $W \perp W'$ means $\boxed{\forall\, \vec{w} \in W \text{ and } \vec{w}\,' \in W',\ \vec{w} \perp \vec{w}\,'}$.
- $\vec{v} \in V$ is a *unit vector* means $\boxed{\|\vec{v}\| = 1}$, 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: $\vec{v} \perp \vec{w} \ \Leftrightarrow\ \|\vec{v} + \vec{w}\|^2 = \|\vec{v}\|^2 + \|\vec{w}\|^2$.
  - Important consequence: if $\vec{n} \perp W$, then $\forall\, \vec{w} \in W$, $\|\vec{n} + \vec{w}\| \geq \|\vec{n}\|$.
- The *vector projection* of $\vec{v} \in V$ onto a nonzero vector $\vec{w} \in V$ is defined by $\boxed{\operatorname{proj}_{\vec{w}} \vec{v} = \dfrac{\langle \vec{v}, \vec{w} \rangle}{\langle \vec{w}, \vec{w} \rangle} \vec{w}}$.
  - $\operatorname{proj}_{\vec{w}} \vec{v}$ 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 $\boxed{\operatorname{orth}_{\vec{w}} \vec{v} = \vec{v} - \operatorname{proj}_{\vec{w}} \vec{v}}$, we find that $\operatorname{orth}_{\vec{w}} \vec{v} \perp \vec{w}$;
    thus, $\vec{v} = \operatorname{proj}_{\vec{w}} \vec{v} + \operatorname{orth}_{\vec{w}} \vec{v}$ 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 $\operatorname{proj}_{\vec{w}} \vec{v}$ simplifies to $\langle \vec{v}, \vec{w} \rangle\, \vec{w}$.

Orthonormal collections

A collection $\mathscr{U} = \{\, \vec{u}_1,\, \vec{u}_2,\, \ldots,\, \vec{u}_k \,\}$ is *orthonormal* means that $\boxed{\forall\, i, j,\ \langle \vec{u}_i, \vec{u}_j \rangle = \delta_{ij} \overset{\text{def}}{=} \begin{cases} 1, & \text{if } i = j \\ 0, & \text{if } i \neq j \end{cases}}$;

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

- If $\mathscr{C}$ is an orthonormal collection, then $\mathscr{C}$ is linearly independent.
- Any finite collection $\mathscr{C}$ in V can be converted into an *orthonormal* collection $\mathscr{U} = \{\, \vec{u}_1,\, \vec{u}_2,\, \ldots,\, \vec{u}_k \,\}$
  having the same span (i.e., an orthonormal basis for $\operatorname{span}\mathscr{C}$) by the *Gram-Schmidt Process*:
  For each vector $\vec{v}$ of $\mathscr{C}$ in turn:
  - Subtract its projections onto the vectors of $\mathscr{U}$ 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: $\frac{1}{\|\vec{v}\,'\|}\vec{v}\,' \curvearrowright \mathscr{U}$.

  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 $\mathrm{C}([-\pi, \pi])$
    (with respect to the inner product $\langle f, g \rangle = \frac{1}{\pi}\int_{-\pi}^{\pi} fg$).

Projection and orthonormal collections

Suppose that $W = \operatorname{span}\{\vec{w}_1, \vec{w}_2, \ldots, \vec{w}_n\}$ 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 $\boxed{\operatorname{proj}_W \vec{v} = \langle \vec{v}, \vec{w}_1 \rangle \vec{w}_1 + \langle \vec{v}, \vec{w}_2 \rangle \vec{w}_2 + \cdots + \langle \vec{v}, \vec{w}_n \rangle \vec{w}_n}$.
  - Defining $\boxed{\operatorname{orth}_W \vec{v} = \vec{v} - \operatorname{proj}_W \vec{v}}$, we find that $\operatorname{orth}_W \vec{v} \perp W$;
    thus, $\vec{v} = \operatorname{proj}_W \vec{v} + \operatorname{orth}_W \vec{v}$ expresses any vector $\vec{v}$ as a vector of W plus a vector orthogonal to W.
  - $\operatorname{proj}_W \vec{v}$ is the vector of W that best approximates $\vec{v}$.
    - Proven via the corollary to the Pythagorean Theorem and the fact that $\operatorname{orth}_W \vec{v} \perp W$.
- Example: Fourier polynomials for continuous functions on $[-\pi, \pi]$.
  - The collection $\mathscr{F}_n = \{ \frac{1}{\sqrt{2}}, \cos x, \sin x, \cos 2x, \sin 2x, \ldots, \cos nx, \sin nx \}$ is orthonormal in $\mathrm{C}([-\pi,\pi])$ (with respect to the inner product $\langle f, g \rangle = \frac{1}{\pi} \int_{-\pi}^{\pi} fg$).
  - Given $f(x) \in \mathrm{C}([-\pi,\pi])$, the n^{th}*-order Fourier polynomial* for $f(x)$ is:

$$\boxed{\begin{aligned} \operatorname{proj}_{\operatorname{span} \mathscr{F}_n} f(x) = \left\langle f(x), \tfrac{1}{\sqrt{2}} \right\rangle \tfrac{1}{\sqrt{2}} + \langle f(x), \cos x \rangle \cos x + \langle f(x), \sin x \rangle \sin x \\ + \ldots + \langle f(x), \cos nx \rangle \cos nx + \langle f(x), \sin nx \rangle \sin nx \end{aligned}}$$

    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 <u>linearly independent columns</u>.

- $A^{\mathrm{T}}A$ is invertible, so we can define the *projection matrix* $\boxed{P = A(A^{\mathrm{T}}A)^{-1}A^{\mathrm{T}} : \mathbb{R}^m \to \mathbb{R}^m}$.
  - P maps vectors orthogonal to $C(A)$ to zero: $\forall\, \vec{y} \perp C(A),\ P\vec{y} = \vec{0}$.
  - P maps vectors in $C(A)$ to themselves: $\forall\, \vec{y} \in C(A),\ P\vec{y} = \vec{y}$.
  - Applying these results to any vector $\vec{v} = \operatorname{proj}_{C(A)} \vec{v} + \operatorname{orth}_{C(A)} \vec{v}$ shows us that $\boxed{P\vec{v} = \operatorname{proj}_{C(A)} \vec{v}}$.
  - 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:

$$\boxed{[\, A \mid \vec{b}\,] \rightsquigarrow \overset{\text{consistent!}}{[\, A^{\mathrm{T}}A \mid A^{\mathrm{T}}\vec{b}\,]}}$$

- 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.