6. INNER PRODUCT SPACES

To define lengths, angles want to generalize idea of dot product. FIS does real, complex cases together. I will do real case first, complex case later if time.

Definition: A *real inner product space* is a real vector space V with a *real inner product* giving $\langle x, y\rangle \in \mathbf{R}$ for all $x, y \in V$ that is

(RIP1) symmetric/commutative: $\langle x, y\rangle = \langle y, x\rangle \ \forall\ x, y \in V$;
(RIP2) linear in x: $\langle \alpha_1 x_1 + \alpha_2 x_2, y\rangle = \alpha_1\langle x_1, y\rangle + \alpha_2\langle x_2, y\rangle \ \forall\ \alpha_1, \alpha_2 \in \mathbf{R}$, $x_1, x_2, y \in V$; and
(RIP3) *positive definite*: $\langle x, x\rangle > 0 \ \forall\ x \in V$, $x \neq 0$.

Examples: (1) Usual dot product on $\mathbf{R}^2$ or $\mathbf{R}^3$ or $\mathbf{R}^n$: $\langle x, y\rangle = x \cdot y = x_1y_1 + \ldots + x_ny_n$.

(2) On $C[a, b] = \{$real-valued continuous functions on $[a, b]\}$ can take
$\langle f, g\rangle = \int_a^b f(x)g(x)\,dx$ or some multiple of this, e.g. $\frac{1}{b-a}$ times.
In general, positive multiple of inner product is inner product.

(3) On $\mathbf{R}^{n\times n}$ take $\langle A, B\rangle = \operatorname{tr} B^{\mathrm{T}}A = \sum_{i=1}^n \sum_{j=1}^n A_{ij}B_{ij}$. Like dot product, multiply corresponding elements and add.

Notes: (RIP4) By (RIP1) and (RIP2), $\langle x, y\rangle$ is also linear in y, so *bilinear*. 'Positive definite symmetric bilinear form' on real V = inner product.

(RIP5) Bilinearity implies $\langle 0, 0\rangle = \langle x, 0\rangle = \langle 0, y\rangle = 0 \ \forall\ x, y \in V$.

Definition: *Length* or *magnitude* of $x \in V$ is $\|x\| = \sqrt{\langle x, x\rangle}$. Generalizes length in $\mathbf{R}^n$.
Unit vector in direction of x is $\dfrac{x}{\|x\|}$.

Example: (3 ctd) If $A = \begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix}$ then $\|A\| = \sqrt{1 + 4 + 9 + 16} = \sqrt{30}$, and unit vector in direction of A is $\dfrac{1}{\sqrt{30}}\begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix}$.

Need following result to sensibly define angles.
(RIP6) **Cauchy-Schwartz Inequality:** $|\langle x, y\rangle| \le \|x\|\,\|y\|$.
Proof: True if $x = 0$ or $y = 0$. If $x \neq 0$ and $y \neq 0$,

$$0 \le \left\langle \frac{x}{\|x\|} - \frac{y}{\|y\|}, \frac{x}{\|x\|} - \frac{y}{\|y\|} \right\rangle = \left\langle \frac{x}{\|x\|}, \frac{x}{\|x\|} \right\rangle - 2\left\langle \frac{x}{\|x\|}, \frac{y}{\|y\|} \right\rangle + \left\langle \frac{y}{\|y\|}, \frac{y}{\|y\|} \right\rangle$$

since commutative

$$= \frac{\langle x, x\rangle}{\|x\|^2}(= 1) - 2\frac{\langle x, y\rangle}{\|x\|\,\|y\|} + \frac{\langle y, y\rangle}{\|y\|^2}(= 1) = 2 - 2\frac{\langle x, y\rangle}{\|x\|\|y\|},$$

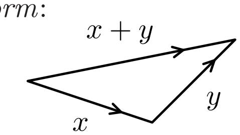
from which $\langle x, y\rangle \le \|x\|\,\|y\|$. Also $-\langle x, y\rangle = \langle -x, y\rangle \le \|-x\|\|y\| = \|x\|\,\|y\|$. Result follows. ∎

(RIP7) So can define angle θ between nonzero x and y: $\theta = \cos^{-1}\dfrac{\langle x, y\rangle}{\|x\|\,\|y\|}$. Generalizes result from $\mathbf{R}^n$.

Definition: x and y are *orthogonal* or *perpendicular*, $x \perp y$, if $\langle x, y\rangle = 0$. 0 is orthogonal to everything. Means angle is 90° if vectors nonzero.

(RIP8) Definition of length also has natural properties. $\|x\| = \sqrt{\langle x, x\rangle}$ is a *norm*:

(N1) $\|x\| \ge 0$, and if $x \neq 0$ then $\|x\| > 0$, $\forall\ x \in V$;
(N2) $\|\alpha x\| = |\alpha|\,\|x\| \ \forall\ \alpha \in \mathbf{R}$, $x \in V$; and
(N3) *triangle inequality*: $\|x + y\| \le \|x\| + \|y\| \ \forall\ x, y \in V$.



'Norm' is a general idea, need not come from inner product, e.g. in $\mathbf{R}^n$, $\|x\|_1 = |x_1| + \ldots + |x_n|$.

Proof of (N3): $\|x+y\|^2 = \langle x+y, x+y\rangle = \langle x,x\rangle + 2\langle x,y\rangle + \langle y,y\} = \|x\|^2 + 2\langle x,y\rangle + \|y\|^2$
$\le \|x\|^2 + 2\|x\|\,\|y\| + \|y\|^2$ by Cauchy-Schwartz
$= (\|x\| + \|y\|)^2$. ∎

Definition: *Orthogonal* set of vectors: pairwise orthogonal. *Orthonormal* set: orthogonal and all length 1 (unit vectors).

Lemma: If $S = \{v_1, v_2, \ldots, v_k\}$ is an orthogonal set of nonzero vectors, and $y \in \operatorname{span} S$, then y has a unique representation $y = \alpha_1 v_1 + \alpha_2 v_2 + \ldots + \alpha_k v_k$ given by $\alpha_i = \langle y, v_i\rangle / \|v_i\|^2 = \langle y, v_i\rangle / \langle v_i, v_i\rangle$.

Proof: We know $y = \alpha_1 v_1 + \ldots + \alpha_k v_k$ for some $\alpha_1, \ldots, \alpha_k$. Take any i, $1 \le i \le k$. We have

$$\begin{aligned}\langle y, v_i\rangle &= \langle \alpha_1 v_1 + \ldots + \alpha_i v_i + \ldots + \alpha_k v_k, v_i\rangle \\ &= \alpha_1\langle v_1, v_i\rangle + \ldots + \alpha_i\langle v_i, v_i\rangle + \ldots + \alpha_k\langle v_k, v_i\rangle \\ &= \alpha_i \langle v_i, v_i\rangle \quad \text{since all other terms are 0.}\end{aligned}$$

Thus $\alpha_i = \langle y, v_i\rangle / \langle v_i, v_i\rangle$. ∎

Consequences: (1) If S is an orthogonal set of nonzero vectors, then S is linearly independent. S can be infinite here, but linear independence just looks at finite subsets of S.

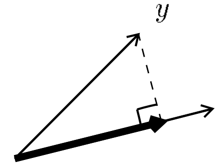
Proof: By the lemma, the unique way to represent 0 as a linear combination of (finitely many) $v_1, v_2, \ldots, v_k \in S$ is with zero coefficients. ∎

(2) If S is a finite orthogonal, or (even better) orthonormal set, then (a) S is a basis of $V = \operatorname{span} S$ and (b) it is easy to find coordinates in V relative to S (formula for α_i above).

So want to find orthogonal, or orthonormal bases.

Definition: Given $x, y \in V$ with $x \ne 0$, the *orthogonal projection of y onto x*, $\operatorname{proj}_x y$, is the unique vector parallel to (multiple of) x so that $y - \operatorname{proj}_x y$ is orthogonal to x.

$$\operatorname{proj}_x y = (\|y\| \cos\theta)\frac{x}{\|x\|} = \frac{\langle y, x\rangle}{\|x\|^2}\,x = \frac{\langle y, x\rangle}{\langle x, x\rangle}\,x.$$



Gram-Schmidt Orthogonalization Algorithm: Given linearly independent $w_1, w_2, \ldots, w_m$, constructs orthogonal $v_1, v_2, \ldots, v_m$ so that $\operatorname{span}\{v_1, v_2, \ldots, v_k\} = \operatorname{span}\{w_1, w_2, \ldots, w_k\}$ for all k, $0 \le k \le m$:

$v_1 = w_1$;
$v_2 = w_2 - \operatorname{proj}_{v_1} w_2$;
$v_3 = w_3 - \operatorname{proj}_{v_1} w_3 - \operatorname{proj}_{v_2} w_3$;
$\vdots$

$$v_k = w_k - \sum_{j=1}^{k-1} \operatorname{proj}_{v_j} w_k = w_k - \sum_{j=1}^{k-1} \frac{\langle w_k, v_j\rangle}{\langle v_j, v_j\rangle} v_j \quad \text{for } 1 \le k \le m.$$

Example: Find orthonormal basis for $V = \operatorname{span}\{(1,2,0,0), (4,3,1,0), (1,-3,-1,1)\} \subseteq \mathbf{R}^4$.

First apply Gram-Schmidt:

$v_1 = w_1 = (1,2,0,0)$; $\quad \langle v_1, v_1\rangle = v_1 \cdot v_1 = 5$.

$$v_2 = w_2 - \frac{\langle w_2, v_1\rangle}{\langle v_1, v_1\rangle} v_1 = (4,3,1,0) - \frac{(4,3,1,0)\cdot(1,2,0,0)}{5}(1,2,0,0) = (4,3,1,0) - \tfrac{10}{5}(1,2,0,0)$$
$$= (4,3,1,0) - (2,4,0,0) = (2,-1,1,0); \quad \langle v_2, v_2\rangle = v_2 \cdot v_2 = 6.$$

$$\begin{aligned} v_3 = w_3 - \frac{\langle w_3, v_1\rangle}{\langle v_1, v_1\rangle} v_1 - \frac{\langle w_3, v_2\rangle}{\langle v_2, v_2\rangle} v_2 &= (1,-3,-1,1) \\ &- \frac{(1,-3,-1,1)\cdot(1,2,0,0)}{5}(1,2,0,0) - \frac{(1,-3,-1,1)\cdot(2,-1,1,0)}{6}(2,-1,1,0) \\ &= (1,-3,-1,1) - \tfrac{-5}{5}(1,2,0,0) - \tfrac{4}{6}(2,-1,1,0) = (1,-3,-1,1) + (1,2,0,0) - (\tfrac{4}{3}, -\tfrac{2}{3}, \tfrac{2}{3}, 0) \\ &= (\tfrac{4}{3}, -\tfrac{1}{3}, -\tfrac{5}{3}, 1); \end{aligned}$$

can replace by $v_3' = 3v_3 = (2,-1,-5,3)$; $\langle v_3', v_3'\rangle = v_3' \cdot v_3' = 39$.

So $\{v_1, v_2, v_3'\}$ is orthogonal basis. For orthonormal basis, *normalize* each vector (divide by length):

$$\{u_1 = \frac{1}{\sqrt{5}}(1,2,0,0), u_2 = \frac{1}{\sqrt{6}}(2,-1,1,0), u_3 = \frac{1}{\sqrt{39}}(2,-1,-5,3)\}.$$

Terminology: (FIS) For ordered orthonormal basis $B = (u_1, u_2, \ldots, u_m)$, numbers $\langle y, u_i\rangle$ are *Fourier coefficients* of y with respect to B. Just coordinates, but computed very easily.

Definition: Given any $S \subseteq V$, the *orthogonal complement of* S, $S^\perp$ is $\{v \in V \mid \langle v, s\rangle = 0 \; \forall\, s \in S\}$.

Examples: If $S = \{(1,2,3)\} \subseteq \mathbf{R}^3$, $S^\perp$ is plane $x + 2y + 3z = 0$. Also true if S is whole line $\{t(1,2,3) \mid t \in \mathbf{R}\}$.

Properties: (OC1) $S^\perp$ is a subspace of V for any $S \subseteq V$.

(OC2) $\emptyset^\perp = \{0\}^\perp = V$.

(OC3) $S^\perp = (\text{span } S)^\perp \; \forall\, S \subseteq V$.

(OC4) Suppose W is a finite-dimensional subspace of V with orthonormal basis $\{v_1, v_2, \ldots, v_k\}$.

(a) Every $y \in V$ can be written uniquely as $y = w + z$ where $w \in W$, $z \in W^\perp$ and $w = \sum_{i=1}^k \langle y, v_i\rangle v_i$. We write $w = \text{proj}_W y$ (the *orthogonal projection of* y *onto* W). The function $\text{proj}_W : V \to W$ is a linear transformation.

(b) $\text{proj}_W y$ is the unique closest point (minimum distance) on W to y.

(OC5) Suppose V is finite-dimensional and W is a subspace of V.

(a) Suppose we take a basis $\{v_1, v_2, \ldots, v_k\}$ for W, extend it to a basis $\{v_1, v_2, \ldots, v_n\}$ for V, and apply Gram-Schmidt to get $\{w_1, w_2, \ldots, w_n\}$. Then $\{w_1, w_2, \ldots, w_k\}$ is an orthogonal basis for W, and $\{w_{k+1}, w_{k+2}, \ldots, w_n\}$ is an orthogonal basis for $W^\perp$.

Consequently $\dim W + \dim W^\perp = \dim V$.

(b) $(W^\perp)^\perp = W$ (duality result).

Example: W is the line $\{t(1,2,3) \mid t \in \mathbf{R}\}$, $W^\perp$ is $x + 2y + 3z = 0$, $(W^\perp)^\perp$ is just line again.

Definition: A *complex inner product space* is a complex vector space V with a *complex inner product* giving $\langle x, y\rangle \in \mathbf{C}$ for all $x, y \in V$ that is

(CIP1) conjugate symmetric: $\overline{\langle x, y\rangle} = \langle y, x\rangle \; \forall\, x, y \in V$;

(CIP2) linear in x: $\langle \alpha_1 x_1 + \alpha_2 x_2, y\rangle = \alpha_1 \langle x_1, y\rangle + \alpha_2 \langle x_2, y\rangle \; \forall\, \alpha_1, \alpha_2 \in \mathbf{R}$, $x_1, x_2, y \in V$; and

(CIP3) *positive definite*: $\langle x, x\rangle > 0$ (means it's real) $\forall\, x \in V$, $x \neq 0$.

Together (CIP1) and (CIP2) mean $\langle x, y\rangle$ is *conjugate linear* in y: $\langle x, \beta_1 y_1 + \beta_2 y_2\rangle = \overline{\beta_1}\langle x, y_1\rangle + \overline{\beta_2}\langle x, y_2\rangle$. Altogether $\langle x, y\rangle$ is *sesquilinear*.

Example: Standard inner product on $\mathbf{C}^n$ is $\langle x, y\rangle = x_1\overline{y_1} + x_2\overline{y_2} + \ldots + x_n\overline{y_n}$.

Complex inner products important for example in quantum mechanics.

Many properties (e.g., Cauchy-Schwartz, length is norm) same as for real inner products although proofs may need modifying.

Some results about real matrices proved most easily by using complex arguments, e.g. every real symmetric matrix is diagonalizable.