Vector Spaces

A vector space is defined as a set V over a (scalar) field F, together with two binary operations, i.e., vector addition $(+)$ and scalar multiplication $(\cdot)$, satisfying the following axioms:

- Commutativity of $+$: $u + v = v + u$, $\forall u, v, \in V$;
- Associativity of $+$: $u + (v + w) = (u + v) + w$, $\forall u, v, w \in V$;
- Identity element for $+$: $\exists 0 \in V : v + 0 = 0 + v = v$, $\forall v \in V$;
- Inverse element for $+$: $\forall v \in V \exists (-v) \in V : v + (-v) = (-v) + v = 0$;
- Associativity of $\cdot$: $a(bv) = (ab)v$, $\forall a, b \in F, v \in V$;
- Distributivity of $\cdot$ w.r.t. vector $+$: $a(v + w) = av + aw$, $\forall a \in F, v, w \in V$;
- Distributivity of $\cdot$ w.r.t. scalar $+$: $(a + b)v = av + bv$, $\forall a, b \in F, v \in V$;
- Normalization: $1v = v$, $\forall v \in V$.

Vector space examples (or not?)

- $\mathbb{R}^n$, $\mathbb{C}^n$;
- Real continuous functions $f : \mathbb{R} \to \mathbb{R}$
- The set of $m \times n$ matrices;
- The set of solutions $y(t)$ of the LTI ODE $dy(t)/dt + 3y(t) = 0$;
- The set of points $(x_1, x_2, x_3) \in \mathbb{R}^3$ satisfying $x_1^2 + x_2^2 + x_3^2 = 1$.
- The set of solutions $y(t)$ of the LTI ODE $dy(t)/dt + 3y\hat{}2(t) = 0$.

Subspaces

- A **subspace** of a vector space is a subset of vectors that itself forms a vector space.
- A necessary and sufficient condition for a subset of vectors to form a subspace is that this subset be closed with respect to vector addition and scalar multiplication.

Subspace examples (or not?)

- The range on any real $n \times m$ matrix, and the nullspace of any $m \times n$ matrix.
- The set of all linear combinations of a given set of vectors.
- The intersection of two subspaces.
- The union of two subspaces.
- The Minkowski (or direct) sum of two subspaces.

Linear (in)dependence, bases

- n vectors $v_1, v_2, \ldots, v_n \in V$ are (linearly) independent if

$$c_1 v_1 + c_2 v_2 + \ldots + c_n v_n = 0 \quad \Leftrightarrow \quad c_1, c_2, \ldots, c_n = 0.$$

- A space is n-dimensional if every set of more than n vectors is dependent, but there is some set of n vectors that are independent.
- Any set of n independent vectors is also called a basis for the space.
- if a space contains a set of n independent vectors for any $n \in \mathbb{N}$, then the space is infinite-dimensional.

Norms

Norms measure the ‘length” of a vector. A norm maps all vectors in a vector space to a non-negative scalar, with the following properties:

- Positivity: $\|x\| > 0$ for $x \neq 0$.
- Homogeneity: $\|ax\| = |a| \, \|x\|, \quad \forall a \in \mathbb{R}, x \in V$.
- Triangle inequality: $\|x + y\| \leq \|x\| + \|y\|$.

Norm examples (or not?)

- Usual Euclidean norm in $\mathbb{R}^n$, $\|x\| = \sqrt{x'x}$;
  (where x' is the conjugate transpose of x, i.e., as in Matlab).
- A matrix Q is Hermitian if $Q' = Q$, and positive definite if $x'Qx > 0$ for $x \neq 0$. Then $\|x\| = \sqrt{x'Qx}$ is a norm.
- For $x \in \mathbb{R}^n$, $\|x\|_1 = \sum_1^n |x_i|$, and $\|x\|_\infty = \max_i |x_i|$.
- For a continuous function $f : [0,1] \to \mathbb{R}$:
  $\|f\|_\infty = \sup_{t \in [0,1]} |f(t)|$, and $\|f\|_2 = \left(\int_0^1 |f(t)|^2 dt \right)^{1/2}$.

Inner product

- An inner product on a vector space V (with scalar field F) is a binary operation $\langle \cdot, \cdot \rangle : V \times V \to F$, with the following properties:
  - Symmetry: $\langle x, y \rangle = \langle y, x \rangle'$, $\forall x, y \in V$;
  - Linearity: $\langle x, ay + bz \rangle = a\langle x, y \rangle + b\langle x, z \rangle$;
  - Positivity: $\langle x, x \rangle > 0$ for $x \neq 0$.
- The inner product gives a geometric structure to the space; e.g., it allows to reason about angles, and in particular, it defines orthogonality. Two vectors x and y are orthogonal if $\langle x, y \rangle = 0$.
- Let $S \subseteq V$ be a subspace of V. The set of vectors orthogonal to all vectors in S is called $S^{\perp}$, the orthogonal complement of S, and is itself a subspace.

Inner product and norms

- An inner product induces a norm $\|x\| = \sqrt{\langle x, x\rangle}$.
- For example, define $\langle x, y\rangle = x'Qy$ with Q Hermitian positive definite.
- For f, g continuous functions on $[0, 1]$, let $\langle f, g\rangle = \int_0^1 f(t)g(t)\, dt$
- Cauchy-Schwartz inequality: $|\langle x, y\rangle| \leq \|x\|\,\|y\|$, $\forall x, y \in V$,
  with equality only if $y = \alpha x$ for some $\alpha \in F$.
  (assuming that the norm is that induced by the inner product)

  Proof

$$0 \leq \langle x + \alpha y, x + \alpha y\rangle = x'x + \alpha' y' x + \alpha x' y + |\alpha|^2 y' y$$

$$\text{Choose } \alpha = -x'y/\langle y, y\rangle :$$

$$0 \leq \langle x, x\rangle\langle y, y\rangle - \langle x, y\rangle^2.$$

The Projection Theorem

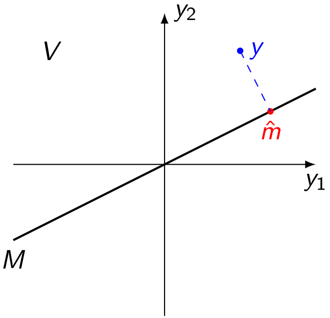
- Let M be a subspace of an inner product space V. Given some $y \in V$, consider the following minimization problem:

$$\min_{m \in M} \|y - m\|,$$

where the norm is that induced by the inner product in V.

Projection theorem

The optimal solution $\hat{m}$ is such that $(y - \hat{m}) \perp M$



Proof of the projection theorem

- By contradiction: assume that $y - \hat{m}$ is not orthogonal to M, i.e., there is some m_0, $\|m_0\| = 1$, such that $\langle y - \hat{m}, m_0 \rangle = \delta \neq 0$.

- Then argue that $(\hat{m} + \delta' m_0) \in M$ achieves a better solution than $\hat{m}$. In fact:

$$\begin{aligned}\|y - \hat{m} - \delta' m_0\|^2 &= \|y - \hat{m}\|^2 - \delta' \langle y - \hat{m}, m_0 \rangle - \delta \langle m_0, y - \hat{m} \rangle + |\delta|^2 \|m_0\|^2 \\ &= \|y - \hat{m}\|^2 - |\delta|^2 - |\delta|^2 + |\delta|^2 \|m_0\|^2 = \|y - \hat{m}\|^2 - |\delta|^2.\end{aligned}$$

Linear Systems of equations

- Consider the following system of (real or complex) linear equations:

$$Ax = y, \qquad A \in \mathbb{R}^{m \times n}, x \in \mathbb{R}^n, y \in \mathbb{R}^m.$$

- Given A and y, is there a solution x?

$$\exists \text{ a solution } x \quad \Leftrightarrow \quad y \in \mathcal{A} \quad \Leftrightarrow \quad \mathcal{R}([A|y]) = \mathcal{R}(A).$$

- There are three cases:
  1. $n = m$: if $\det(A) \neq 0) \Rightarrow x = A^{-1}y$ is the unique solution.
  2. $m > n$: more equations than unknowns, the system is overconstrained. Happens in, e.g., estimation problems, where one tries to estimate a small number of parameters from a lot of experimental measurements. In such cases the problem is typically inconsistent, i.e., $y \notin \mathcal{R}(A)$. So one is interested in finding the solution that minimizes some error criterion.
  3. $m < n$: more unknown than equations, the system is overconstrained. Happens in, e.g., control problems, where there may be more than one way to complete a desired task. If there is a solution x_p (i.e., $Ax_p = y$), then typically there are many other solutions of the form $x = x_p + x_h$, where $x_h \in \mathcal{N}(A)$ (i.e., $Ax_h = 0$). In this case it is desired to find the solution than minimizes some cost criterion.