Lecture 9

Econ 2001

2015 August 20

Lecture 9 Outline

1. Linear Functions
2. Linear Representation of Matrices
3. Analytic Geometry in $\mathbf{R}^n$: Lines and Hyperplanes
4. Separating Hyperplane Theorems

- Back to vector spaces so that we can illustrate the formal connection between linear functions and matrices.
- This matters for a lot calculus since derivatives are linear functions.
- Then some useful geometry in $\mathbf{R}^n$.

Announcements:
- Test 2 will be tomorrow at 3pm, in WWPH 4716, and there will be recitation at 1pm. The exam will last an hour.

Linear Transformations

Definition

Let X and Y be two vector spaces. We say $T : X \rightarrow Y$ is a linear transformation if

$$T(\alpha_1 \mathbf{x}_1 + \alpha_2 \mathbf{x}_2) = \alpha_1 T(\mathbf{x}_1) + \alpha_2 T(\mathbf{x}_2) \quad \forall \mathbf{x}_1, \mathbf{x}_2 \in X,\ \alpha_1, \alpha_2 \in \mathbf{R}$$

Let $L(X, Y)$ denote the set of all linear transformations from X to Y.

- $L(X, Y)$ is a vector space.

Example

The function $T : \mathbb{R}^n \longrightarrow \mathbb{R}^m$ defined by $T(\mathbf{x}) = \mathbf{A}\mathbf{x}$, with $\underset{m \times n}{\mathbf{A}}$ is linear:

$$T(a\mathbf{x} + b\mathbf{y}) = \mathbf{A}(a\mathbf{x} + b\mathbf{y}) = a\mathbf{A}\mathbf{x} + b\mathbf{A}\mathbf{y} = aT(\mathbf{x}) + bT(\mathbf{y})$$

Linear Functions

An equivalent characterization of linearity

$l\colon \mathbb{R}^n \longrightarrow \mathbb{R}^m$ is linear if and only if

1. for all $\mathbf{x}$ and $\mathbf{y}$ in $\mathbb{R}^n$,
$$l(\mathbf{x}+\mathbf{y}) = l(\mathbf{x}) + l(\mathbf{y})$$
and
2. for all scalars λ,
$$l(\lambda\mathbf{x}) = \lambda l(\mathbf{x})$$

- Linear functions are additive and exhibit constant returns to scale (econ jargon).
- If l is linear, then $l(\mathbf{0}) = \mathbf{0}$ and, more generally, $l(\mathbf{x}) = -l(-\mathbf{x})$.

Compositions of Linear Transformations

The compositions of two linear functions is a linear function

Given $R \in L(X, Y)$ and $S \in L(Y, Z)$, define $S \circ R : X \to Z$.

- $$\begin{aligned}(S \circ R)(\alpha \mathbf{x}_1 + \beta \mathbf{x}_2) &= S(R(\alpha \mathbf{x}_1 + \beta \mathbf{x}_2)) \\ &= S(\alpha R(\mathbf{x}_1) + \beta R(\mathbf{x}_2)) \\ &= \alpha S(R(\mathbf{x}_1)) + \beta S(R(\mathbf{x}_2)) \\ &= \alpha (S \circ R)(\mathbf{x}_1) + \beta (S \circ R)(\mathbf{x}_2)\end{aligned}$$

  so $S \circ R \in L(X, Z)$.

- Think about statement and proof for the equivalent characterization of linearity.

Kernel and Rank

Definition

Let $T \in L(X, Y)$.

- The image of T is $\mathrm{Im}\, T = T(X) = \{\mathbf{y} \in Y : \mathbf{y} = T(\mathbf{x}) \text{ for some } \mathbf{x} \in X\}$
- The kernel of T is $\ker T = \{\mathbf{x} \in X : T(\mathbf{x}) = 0\}$
- The rank of T is $\mathrm{Rank}\, T = \dim(\mathrm{Im}\, T)$

$\dim X$, is the cardinality of any basis of X.

- The kernel is the set of solutions to $T(\mathbf{x}) = 0$
- The image is the set of vectors in Y for which $T(\mathbf{x}) = \mathbf{y}$ has at least one solution.
- Exercise: prove that $T \in L(X, Y)$ is one-to-one if and only if $\ker T = \{0\}$.

Theorem (The Rank-Nullity Theorem)

Let X be a finite-dimensional vector space, $T \in L(X, Y)$. Then $\mathrm{Im}\, T$ and $\ker T$ are vector subspaces of Y and X respectively, and

$$\dim X = \dim \ker T + \mathrm{Rank}\, T$$

Invertible Linear Functions

Definition

$T \in L(X, Y)$ is invertible if there exists a function $S : Y \to X$ such that

$$
\begin{aligned}
S(T(\mathbf{x})) &= \mathbf{x} \qquad \forall \mathbf{x} \in X \\
T(S(\mathbf{y})) &= y \qquad \forall \mathbf{y} \in Y
\end{aligned}
$$

Denote S by T^{-1}.

In other words, T is invertible if and only if it is one-to-one and onto.

- This is the usual condition for the existence of an inverse function.
- The linearity of the inverse function follows from the linearity of T.

Theorem

If $T \in L(X, Y)$ is invertible, then $T^{-1} \in L(Y, X)$; that is, T^{-1} is linear.

Proof.

Problem Set 9. □

From Linear Functions to Matrices

- $l : \mathbb{R}^n \longrightarrow \mathbb{R}^m$ is linear if and only if (i) $l(\mathbf{x}+\mathbf{y}) = l(\mathbf{x}) + l(\mathbf{y})$ and $l(\lambda \mathbf{x}) = \lambda l(\mathbf{x})$.

Any linear function is described by a matrix.

- *Compute $l(\mathbf{e}_i)$, where $\mathbf{e}_i$ is the ith component of the standard basis in $\mathbb{R}^n$.*
- *For each $\mathbf{e}_i$, $l(\mathbf{e}_i)$ is a vector in $\mathbb{R}^m$, call it $\mathbf{a}_i$.*
- *Form $\mathbf{A}$ as the square matrix with ith column equal to $\mathbf{a}_i$.*
- *By construction, $\mathbf{A}$ has n columns and m rows and therefore $f(\mathbf{x}) = \mathbf{A}\mathbf{x}$ for all $\mathbf{x}$.*

- Next, apply this procedure to any linear basis of a vector space.
- Using this observation, we derive a conection between composition of linear functions and multiplication of matrices.

Linear Transformations and Bases

If X and Y are vector spaces, to define a linear transformation T from X to Y, it is sufficient to define T for every element of a basis for X.

- Why?
  - Let V be a basis for X. Then every vector $\mathbf{x} \in X$ has a unique representation as a linear combination of a finite number of elements of V.

Theorem

Let X and Y be two vector spaces, and let $V = \{\mathbf{v}_\lambda : \lambda \in \Lambda\}$ be a basis for X. Then a linear function $T \in L(X, Y)$ is completely determined by its values on V, that is:

1. *Given any set $\{\mathbf{y}_\lambda : \lambda \in \Lambda\} \subset Y$, $\exists T \in L(X, Y)$ such that*
$$T(\mathbf{v}_\lambda) = \mathbf{y}_\lambda \;\; \forall \lambda \in \Lambda$$
2. *If $S, T \in L(X, Y)$ and $S(\mathbf{v}_\lambda) = T(\mathbf{v}_\lambda)$ for all $\lambda \in \Lambda$, then $S = T$.*

Linear Transformations and Bases

Proof.

Proof of 1. If $\mathbf{x} \in X$, $\mathbf{x}$ has a unique representation of the form

$$\mathbf{x} = \sum_{i=1}^{n} \alpha_i \mathbf{v}_{\lambda_i} \qquad \alpha_i \neq 0 \; i = 1, \ldots, n$$

- Define $T(\mathbf{x}) = \sum_{i=1}^{n} \alpha_i \mathbf{y}_{\lambda_i}$
- Then $T(\mathbf{x}) \in Y$ and you will show that T is linear as an exercise.

Proof of 2. Suppose $S(\mathbf{v}_\lambda) = T(\mathbf{v}_\lambda)$ for all $\lambda \in \Lambda$. We need to show that $S = T$.

- Given $\mathbf{x} \in X$,

$$\begin{aligned} S(\mathbf{x}) &= S\left(\sum_{i=1}^{n} \alpha_i \mathbf{v}_{\lambda_i}\right) = \sum_{i=1}^{n} \alpha_i S\left(\mathbf{v}_{\lambda_i}\right) = \sum_{i=1}^{n} \alpha_i T\left(\mathbf{v}_{\lambda_i}\right) = T\left(\sum_{i=1}^{n} \alpha_i \mathbf{v}_{\lambda_i}\right) \\ &= T(\mathbf{x}) \end{aligned}$$

  so $S = T$. □

Isomorphisms

Definitions

Two vector spaces X and Y are isomorphic if there is an invertible $T \in L(X, Y)$
T is then called an isomorphism.

Isomorphic vector spaces are essentially indistinguishable

If T is an isomorphism it is one-to-one and onto:

- $\forall \mathbf{x}_1, \mathbf{x}_2 \in X$, there exist unique $\mathbf{y}_1, \mathbf{y}_2 \in Y$ s.t. $\mathbf{y}_1 = T(\mathbf{x}_1)$ and $\mathbf{y}_2 = T(\mathbf{x}_2)$
- $\forall \mathbf{y}_1, \mathbf{y}_2 \in Y$, there exist unique $\mathbf{x}_1, \mathbf{x}_2 \in X$ s.t. $\mathbf{x}_1 = T^{-1}(\mathbf{y}_1)$ and $\mathbf{x}_2 = T^{-1}(\mathbf{y}_2)$
- moreover, by linearity

$$T(\alpha\mathbf{x}_1 + \beta\mathbf{x}_2) = \alpha\mathbf{y}_1 + \beta\mathbf{y}_2 \quad \text{and} \quad T^{-1}(\alpha\mathbf{y}_1 + \beta\mathbf{y}_2) = \alpha\mathbf{x}_1 + \beta\mathbf{x}_2$$

Theorem

Two vector spaces X and Y are isomorphic if and only if $\dim X = \dim Y$.

Coordinate Representation of Vectors

By the previous theorem, any vector space of dimension n is isomorphic to $\mathbf{R}^n$. What's the isomorphism?

- Let X be a finite-dimensional vector space over $\mathbf{R}$ with $\dim X = n$.
- Fix a basis $V = \{\mathbf{v}_1, \ldots, \mathbf{v}_n\}$ of X, then any $\mathbf{x} \in X$ has a unique representation $\mathbf{x} = \sum_{j=1}^{n} \beta_j \mathbf{v}_j$ (so the β_j are unique real numbers).
- Define the coordinate representation of $\mathbf{x}$ with respect to the basis V as

$$crd_V(\mathbf{x}) = \begin{pmatrix} \beta_1 \\ \vdots \\ \beta_n \end{pmatrix} \in \mathbf{R}^n$$

- By construction

$$crd_V(\mathbf{v}_1) = \begin{pmatrix} 1 \\ 0 \\ \vdots \\ 0 \\ 0 \end{pmatrix} \quad crd_V(\mathbf{v}_2) = \begin{pmatrix} 0 \\ 1 \\ \vdots \\ 0 \\ 0 \end{pmatrix} \quad \ldots \quad crd_V(\mathbf{v}_n) = \begin{pmatrix} 0 \\ 0 \\ \vdots \\ 0 \\ 1 \end{pmatrix}$$

- crd_V is an isomorphism from X to $\mathbf{R}^n$ (verify this).

Matrix Representation

- Suppose $T \in L(X, Y)$, where X and Y have dimension n and m respecively.
- Fix bases $V = \{\mathbf{v}_1, \ldots, \mathbf{v}_n\}$ of X and $W = \{\mathbf{w}_1, \ldots, \mathbf{w}_m\}$ of Y.
- Since $T(\mathbf{v}_j) \in Y$, we can write

$$T(\mathbf{v}_j) = \sum_{i=1}^{m} \alpha_{ij} \mathbf{w}_i$$

- Define

$$Mtx_{W,V}(T) = \begin{pmatrix} \alpha_{11} & \cdots & \alpha_{1n} \\ \vdots & \ddots & \vdots \\ \alpha_{m1} & \cdots & \alpha_{mn} \end{pmatrix}$$

  - Each column is given by $crd_W(T(\mathbf{v}_j))$, the coordinates of $T(\mathbf{v}_j)$ with respect to W.
- Thus any linear transformation from X to Y is equivalent to a matrix once one fixes the two bases.

Matrix Representation

Example

$X = Y = \mathbf{R}^2,\ V = \{(1,0),(0,1)\},\ W = \{(1,1),(-1,1)\};$

- Let T be the identity: $T(\mathbf{x}) = \mathbf{x}$ for each $\mathbf{x}$.
- Notice that $Mtx_{W,V}(T) \neq \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$. $Mtx_{W,V}(T)$ is the matrix that changes basis from V to W.
- How do we compute it?

$$\mathbf{v}_1 = (1,0) = \alpha_{11}(1,1) + \alpha_{21}(-1,1)$$
$$\alpha_{11} - \alpha_{21} = 1 \quad \text{and} \quad \alpha_{11} + \alpha_{21} = 0$$
$$2\alpha_{11} = 1,\ \alpha_{11} = \tfrac{1}{2} \quad \text{hence } \alpha_{21} = -\tfrac{1}{2}$$
$$\mathbf{v}_2 = (0,1) = \alpha_{12}(1,1) + \alpha_{22}(-1,1)$$
$$\alpha_{12} - \alpha_{22} = 0 \quad \text{and} \quad \alpha_{12} + \alpha_{22} = 1$$
$$2\alpha_{12} = 1,\ \alpha_{12} = \tfrac{1}{2} \quad \text{hence } \alpha_{22} = \tfrac{1}{2}$$

- So $Mtx_{W,V}(id) = \begin{pmatrix} 1/2 & 1/2 \\ -1/2 & 1/2 \end{pmatrix}$

Matrices and Vectors

- Given $T \in L(X, Y)$, where $\dim X = n$ and $\dim Y = m$, and $V = \{\mathbf{v}_1, \ldots, \mathbf{v}_n\}$ and $W = \{\mathbf{w}_1, \ldots, \mathbf{w}_m\}$ are bases of X and Y respectively.
- We built: $Mtx_{W,V}(T) = \begin{pmatrix} \alpha_{11} & \cdots & \alpha_{1n} \\ \vdots & \ddots & \vdots \\ \alpha_{m1} & \cdots & \alpha_{mn} \end{pmatrix}$ and $crd_V(\mathbf{v}_1) = \begin{pmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix}$.
- Clearly:

$$\begin{pmatrix} \alpha_{11} & \cdots & \alpha_{1n} \\ \vdots & \ddots & \vdots \\ \alpha_{m1} & \cdots & \alpha_{mn} \end{pmatrix} \begin{pmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix} = \begin{pmatrix} \alpha_{11} \\ \vdots \\ \alpha_{m1} \end{pmatrix}$$

- Thus

$$Mtx_{W,V}(T) \cdot crd_V(\mathbf{v}_j) = crd_W(T(\mathbf{v}_j))$$

- and more generally

$$Mtx_{W,V}(T) \cdot crd_V(\mathbf{x}) = crd_W(T(\mathbf{x}))\ \forall \mathbf{x} \in X$$

REMARK: Multiplying a vector by a matrix does two things

1. Computes the action of the linear function T
2. Accounts for the change in basis.

Matrices and Linear Transofrmations are Isomorphic

Theorem

Let X and Y be vector spaces over $\mathbf{R}$, with $\dim X = n$, $\dim Y = m$. Then $L(X, Y)$, the space of linear transformations from X to Y, is isomorphic to $\mathcal{M}_{m\times n}$, the vector space of $m \times n$ matrices over $\mathbf{R}$.
If $V = \{\mathbf{v}_1, \ldots, \mathbf{v}_n\}$ is a basis for X and $W = \{\mathbf{w}_1, \ldots, \mathbf{w}_m\}$ is a basis for Y, then

$$Mtx_{W,V} \in L(L(X, Y), \mathcal{M}_{m\times n})$$

and $Mtx_{W,V}$ is an isomorphism from $L(X, Y)$ to $\mathcal{M}_{m\times n}$.

- This is mostly the consequence of things we already know about dimensionality once you realize the relationship between matrices and linear transformations.
- Note that X and Y are general vector spaces, but they are isomorphic to matrices of real numbers.

Summary and Illustration

- Let $T \in L(V, W)$, where V and W are vectors spaces over $\mathbf{R}$ with $\dim V = N$ and $\dim W = M$, and fix bases $\{\mathbf{v}_1, \ldots, \mathbf{v}_N\}$ of V and $\{\mathbf{w}_1, \ldots, \mathbf{w}_M\}$ of W.
- Any $\mathbf{v} \in V$ can be uniquely represented as $\mathbf{v} = \sum_{n=1}^{N} x_n \mathbf{v}_n$ (the x_ns are unique).
- Any $T(\mathbf{v}) \in W$ is uniquely represented as $T(\mathbf{v}) = \sum_{m=1}^{M} y_m \mathbf{w}_m$ (the y_ms are unique).

- Since $T(\mathbf{v}_n) \in W$ for each n, $T(\mathbf{v}_n) = \sum_{m=1}^{M} a_{nm} \mathbf{w}_m$ (where the a_{nm}s are unique).
- Thus:

$$\sum_{m=1}^{M} y_m \mathbf{w}_m = T(\mathbf{v}) = T(\sum_{n=1}^{N} x_n \mathbf{v}_n) = \sum_{n=1}^{N} x_n T(\mathbf{v}_n) = \sum_{n=1}^{N} x_n \sum_{m=1}^{M} a_{nm} \mathbf{w}_m = \sum_{m=1}^{M} (\sum_{n=1}^{N} x_n a_{mn}) \mathbf{w}_m$$

- Therefore,

$$y_m = \sum_{n=1}^{N} x_n a_{nm} \text{ for all } m$$

- Let $\mathbf{y} = (y_1, .., y_M) \in \mathbb{R}^M$, $\mathbf{x} = (x_1, .., x_N) \in \mathbb{R}^N$, and $\mathbf{A} = [a_{mn}]$ (the $m \times n$ matrix with m, n entry given by the real number a_{mn}); then, we have

$$\mathbf{y} = \mathbf{A}\mathbf{x}$$

- In words: the matrix of real numbers $\mathbf{A}$ represents the linear transformation T given the bases $\{\mathbf{v}_1, \ldots, \mathbf{v}_N\}$ and $\{\mathbf{w}_1, \ldots, \mathbf{w}_M\}$.
  - there is one and only one linear transformation T corresponding to $\mathbf{A}$.
  - there is one and only one matrix $\mathbf{A}$ corresponding to T.

Matrix Product as Composite Transformation

- Suppose we have another linear transformation $S \in L(W, Q)$ where Q is a vector space with basis $\{\mathbf{q}_1, \ldots, \mathbf{q}_J\}$.
- Let $\mathbf{B} = [b_{jm}]$ be the $J \times M$ matrix representing S with respect to the bases $\{\mathbf{w}_1, \ldots, \mathbf{w}_M\}$ and $\{\mathbf{q}_1, \ldots, \mathbf{q}_J\}$ for W and Q, respectively.
- Then, for all m, $S(\mathbf{w}_m) = \sum_{j=1}^{J} b_{jm}\mathbf{q}_j$.

- $S \circ T : V \to Q$ is the linear transformation $S \circ T(\mathbf{v}) = S(T(\mathbf{v}))$.
- Using the same logic of the previous slide:

$$\begin{aligned} S \circ T(\mathbf{v}_n) &= S(T(\mathbf{v}_n)) = S(\sum_{m=1}^{M} a_{nm}\mathbf{w}_m) \\ &= \sum_{m=1}^{M} a_{nm} S(\mathbf{w}_m) = \sum_{m=1}^{M} a_{nm} \sum_{j=1}^{J} b_{jm}\mathbf{q}_j = \sum_{j=1}^{J} (\sum_{m=1}^{M} b_{jm} a_{mn})\mathbf{q}_j \end{aligned}$$

- Therefore

$$S \circ T(\mathbf{v}_n) = \sum_{j=1}^{J} c_{jn} q_j$$

where

$$c_{jn} = \sum_{m=1}^{M} b_{jm} a_{mn} \text{ for } j = 1, ..., J$$

- Thus the $J \times N$ matrix $\mathbf{C} = [c_{jn}]$ represents $S \circ T$.
- What is $\mathbf{C}$? The product $\mathbf{BA}$ (check the definition of matrix product).

Analytic Geomtetry: Lines

- How do we talk about points, lines, planes,... in $\mathbb{R}^n$?

Definition

A line in $\mathbb{R}^n$ is described by a point $\mathbf{x}$ and a direction v. It can be represented as

$$\{\mathbf{z} \in \mathbb{R}^n : \text{there exists } t \in \mathbb{R} \text{ such that } \mathbf{z} = \mathbf{x} + t\mathbf{v}\}$$

- If $t \in [0, 1]$, this is the line segment connecting $\mathbf{x}$ to $\mathbf{x} + \mathbf{v}$.

REMARK

Even in $\mathbb{R}^n$ two points still determine a line: the line connecting $\mathbf{x}$ to $\mathbf{y}$ is the line containing $\mathbf{x}$ in the direction $\mathbf{v}$.

- Check that this is the same as the line through $\mathbf{y}$ in the direction $\mathbf{v}$.

Analytic Geomtetry: Hyperplanes

Definition

A hyperplane is a set described by a point $\mathbf{x}_0 \in \mathbb{R}^n$ and a normal direction of the plane $\mathbf{p} \in \mathbb{R}^n$, $\mathbf{p} \neq 0$. It can be represented as

$$\{\mathbf{z} \in \mathbb{R}^n : \mathbf{p} \cdot (\mathbf{z} - \mathbf{x}_0) = 0\}.$$

- A hyperplane consists of all the $\mathbf{z}$ such that the direction $\mathbf{z} - \mathbf{x}_0$ is orthogonal to $\mathbf{p}$.
- Hyperplanes can equivalently be written as $\mathbf{p} \cdot \mathbf{z} = \mathbf{p} \cdot \mathbf{x}_0$ (and if $\mathbf{x}_0 = \mathbf{0}$ this is $\mathbf{p} \cdot \mathbf{z} = 0$)
- In $\mathbb{R}^2$ lines are also hyperplanes. In $\mathbb{R}^3$ hyperplanes are “ordinary” planes.

REMARK

Lines and hyperplanes are two kinds of “flat” subsets of $\mathbb{R}^n$.

- Lines are subsets of dimension one.
- Hyperplanes are subsets of dimension $n - 1$ or co-dimension one.
- One can have flat subsets of any dimension less than n.

Linear Manifolds

- Lines and hyperplanes are not subspaces because they do not contain the origin. (what is a subspace anyhow?)
- They are obtained by "translating" a subspace: adding the same constant to all of its elements.

Definition

A linear manifold of $\mathbb{R}^n$ is a set S such that there is a subspace V on $\mathbb{R}^n$ and $\mathbf{x}_0 \in \mathbb{R}^n$ with

$$S = V + \{\mathbf{x}_0\} \equiv \{\mathbf{y} : \mathbf{y} = \mathbf{v} + \mathbf{x}_0 \text{ for some } \mathbf{v} \in V\}$$

- Lines and hyperplanes are linear manifolds (not linear subspaces).

Lines

Here is another way to describe a line.

Given two points $\mathbf{x}$ *and* $\mathbf{y}$ *in* $\mathbb{R}^n$*, the line that passes through these points is:*

$$\{\mathbf{z} \in \mathbb{R}^n : \mathbf{z} = \mathbf{x} + t(\mathbf{y} - \mathbf{x}) \text{ for some } t\}.$$

This is called "parametric" representation (it defines n equations).

Example

Given $\mathbf{y}, \mathbf{x} \in \mathbb{R}^2$, we find a line through them by solving:

$$z_1 = x_1 + t(y_1 - x_1) \qquad \text{and} \qquad z_2 = x_2 + t(y_2 - x_2).$$

Solve the first equation for t and substite out:

$$z_2 = x_2 + \frac{(y_2 - x_2)(z_1 - x_1)}{y_1 - x_1} \qquad \text{or} \qquad z_2 - x_2 = \frac{y_2 - x_2}{y_1 - x_1}(z_1 - x_1).$$

This is the standard way to represent the equation of a line (in the plane) through (x_1, x_2) with slope $(y_2 - x_2)(y_1 - x_1)$.

Lines

Given two points $\mathbf{x}$ *and* $\mathbf{y}$ *in* $\mathbb{R}^n$*, the line that passes through these points is:*

$$\{\mathbf{z} \in \mathbb{R}^n : \mathbf{z} = \mathbf{x} + t(\mathbf{y} - \mathbf{x}) \text{ for some } t\}.$$

This is called "parametric" representation (it defines n equations).

- The parametric representation is almost equivalent to the standard representation in $\mathbb{R}^2$.
- Why almost? It is more general since it allows for lines parallel to the axes.
- One needs two pieces of information to describe a line:
  - Point and direction, or
  - Two points (you get the direction by subtracting the points).

Hyperplanes

- How does one describe an hyperplane?
- One way is to use a point and a (normal) direction

$$\{\mathbf{z} \in \mathbb{R}^n : \mathbf{p} \cdot (\mathbf{z} - \mathbf{x}_0) = 0\}$$

- Another way is to use n points in $\mathbb{R}^n$, provided these are in "general position."
- You can go from points to the normal solving a linear system of equations.

How to Find an Hyperplane when n=3

In $\mathbb{R}^3$ an hyperplane is a 'plane' and one can describe it using three points.

Example

Given $(1, 2, -3), (0, 1, 1), (2, 1, 1)$ find A, B, C, D such that

$$Ax_1 + Bx_2 + Cx_3 = D$$

- Solve the following system of equations: $\begin{cases} A + 2B - 3C = D \\ B + C = D \\ 2A + B + C = D \end{cases}$
- This yields $(A, B, C, D) = (0, .8D, .2D, D)$.
- Hence, if we find one set of coefficients, any non-zero multiple will also work.
- Hence an equation for the plane is: $4x_2 + x_3 = 5$ (check that the three points actually satisfy this equation).

Describing Hyperplanes with Normals

- Given some points, look for a normal direction.
  - A normal direction is a direction that is orthogonal to **all** directions in the plane.
  - A direction in the plane is a direction of a line in the plane.
- Hence, we can get such a direction by subtracting any two points in the plane.

Given $(1, 2, -3), (0, 1, 1), (2, 1, 1)$ find a two dimensional hyperplane

- It has two independent directions.
  - One direction comes from the difference between the first two points:
    $$(1, 2, -3) - (0, 1, 1) = (1, 1, -4).$$
  - The other can come from the difference between the second and third points
    $$(0, 1, 1) - (2, 1, 1) = (-2, 0, 0).$$
  - We can now find a normal to both of them.
- That is, a $\mathbf{p}$ such that $\mathbf{p} \neq 0$ and
  $$\mathbf{p} \cdot (1, 1, -4) = \mathbf{p} \cdot (-2, 0, 0) = 0$$
  - any multiple of $(0, 4, 1)$ solves this system of two equations and three unknowns.
- Hence, the equation for the hyperplane is
  $$(0, 4, 1) \cdot (x_1 - 1, x_2 - 2, x_3 + 3) = 0$$

Separating Hyperplanes and Convexity

- Consider two sets X and Y in $\mathbb{R}^n$ which do not interserct.
- A hyperplane in $\mathbb{R}^n$ is expressed as $\mathbf{p} \cdot \mathbf{x} = c$ where $\mathbf{x} \in \mathbb{R}^n$, $\mathbf{p} \neq \mathbf{0}$ is in $\mathbb{R}^n$, and $c \in \mathbb{R}$.
- Under what conditions can we find an hyperplane that divides the space in two, each side containing only one of those sets?
- Draw a few pictures.
- The following is a minimal condition sets have to satisfy to make separation possible.

Definition

$X \subset \mathbb{R}^n$ is convex if $\forall\, \mathbf{x}, \mathbf{y} \in X$ and $\forall\, \alpha \in [0, 1]$ we have

$$\alpha \mathbf{x} + (1 - \alpha)\mathbf{y} \in X$$

Separating Hyperplane Theorem

- Start with the simpler case of separate a point from a set.

Theorem (Separating Hyperplane)

Given a nonempty, closed, convex set $X \subset \mathbb{R}^n$ and $\mathbf{x} \in \mathbb{R}^n$, with $\mathbf{x} \notin X$, There exists a $\mathbf{p} \in \mathbb{R}^n, \mathbf{p} \neq \mathbf{0}$, and a $c \in \mathbb{R}$ such that

$$X \subset \{\mathbf{y} \in \mathbb{R}^n : \mathbf{p} \cdot \mathbf{y} \geq c\}$$

and

$$\mathbf{p} \cdot \mathbf{x} < c$$

- That is,

$$\mathbf{p} \cdot \mathbf{y} = c$$

  defines a separating hyperplane for X: it leaves all of X on one side and $\mathbf{x}$ on the other.
- Without loss of generality, one can normalize the normal to the separating hyperplane.
    - That is, we can assume that $\|\mathbf{p}\| = 1$.

Proof of the separating hyperplane theorem in 9(+1 for you) steps

Consider the problem of minimizing the distance between $\mathbf{x}$ and X. That is: find a vector that solves $\min_{y \in X} \|\mathbf{x} - \mathbf{y}\|$.

1. X is nonempty, so we can find some element $\mathbf{z} \in X$. While X is not necessarily bounded, without loss of generality we can replace X by $\{y \in X : \|\mathbf{y} - \mathbf{x}\| \leq \|\mathbf{z} - \mathbf{x}\|\}$. This set is compact because X is closed.
2. The norm is a continuous function. Hence there is a solution $\mathbf{y}^*$.
3. Let $\mathbf{p} = \mathbf{y}^* - \mathbf{x}$. Since $\mathbf{x} \notin X$, $\mathbf{p} \neq \mathbf{0}$.
4. Let $c = \mathbf{p} \cdot \mathbf{y}^*$. Since $c - \mathbf{p} \cdot \mathbf{x} = \mathbf{p} \cdot (\mathbf{y}^* - \mathbf{x}) = \|\mathbf{p}\|^2$, therefore $c > \mathbf{p} \cdot \mathbf{x}$.
5. We need to show
$$\text{if } \mathbf{y} \in X, \text{ then } \mathbf{p} \cdot \mathbf{y} \geq c.$$
6. Notice how this inequality is equivalent to
$$(\mathbf{y}^* - \mathbf{x}) \cdot (\mathbf{y} - \mathbf{y}^*) \geq 0.$$
7. Since X is convex and $\mathbf{y}^*$ solves the minimzation problem, it must be that
$$\|t\mathbf{y} + (1-t)\mathbf{y}^* - \mathbf{x}\|^2$$
is minimized when $t = 0$.
8. Since the derivative of $\|t\mathbf{y} + (1-t)\mathbf{y}^* - \mathbf{x}\|^2$ is non-negative at $t = 0$, the first order condition will do it.
9. Check that differentiating $\|t\mathbf{y} + (1-t)\mathbf{y}^* - \mathbf{x}\|^2$ (as a function of t) and simplifying yields the desired inequality.

Extensions

- If $\mathbf{x}$ is in the boundary of X, then you can approximate $\mathbf{x}$ by a sequence $\mathbf{x}_k$ such that each $\mathbf{x}_k \notin X$.
  - This yields a sequence of $\mathbf{p}_k$, which can be taken to be unit vectors, that satisfy the conclusion of the theorem.
  - A subsequence of the $\mathbf{p}_k$ must converge.
  - The limit point $\mathbf{p}^*$ will satisfy the conclusion of the theorem (except we can only guarantee that $c \geq \mathbf{p}^* \cdot \mathbf{x}$ rather than the strict equality).
- The closure of any convex set is convex.
  - Given a convex set X and a point $\mathbf{x}$ not in the interior of the set, we can separate $\mathbf{x}$ from the closure of X.
- Putting these things together...

A Stronger Separating Hyperplane Theorem

Theorem (Supporting Hyperplane)

Given a convex set $X \subset \mathbb{R}^n$ and $\mathbf{x} \in \mathbb{R}^n$, $\mathbf{x}$. If $\mathbf{x}$ is not in the interior of X, then there exists $\mathbf{p} \in \mathbb{R}^n, \mathbf{p} \neq \mathbf{0}$, and $c \in \mathbb{R}$ such that

$$X \subset \{\mathbf{y} \in \mathbb{R}^n \mid \mathbf{p} \cdot \mathbf{y} \geq c\}$$

and

$$\mathbf{p} \cdot \mathbf{x} \leq c$$

- Other general versions separate two convex sets. The easiest case is when the sets do not have any point in common, and the slightly harder is the one in which no point of one set is in the interior of the other set.
  - Let $A, B \subseteq \mathbf{R}^n$ be nonempty, disjoint convex sets. Then there exists a nonzero vector $p \in \mathbf{R}^n$ such that
  $$p \cdot a \leq p \cdot b \quad \forall a \in A, b \in B$$
- The separating hyperplane theorem can be used to provide intuition for the way we solve constrained optimization problems.
- In the typical economic application, the separating hyperplane's normal is a price vector, and the separation property states that a particular vector costs more than vectors in a consumption set.

Tomorrow

Calculus

1. Level Sets
2. Derivatives and Partial Derivatives
3. Differentiability
4. Tangents to Level Sets