Math 164 – Optimization

University of California, Los Angeles

Duc Vu

Spring 2021

This is math 164 – Optimization taught by Professor Li. We meet weekly on MWF from 3:00 pm to 3:50 pm for lecture. The main textbook used for the class is *An Introduction to Optimization* 4^{th} by *Chong* and *Zak*. Other course notes can be found at my github. Please let me know through my email if you spot any typos in the note.

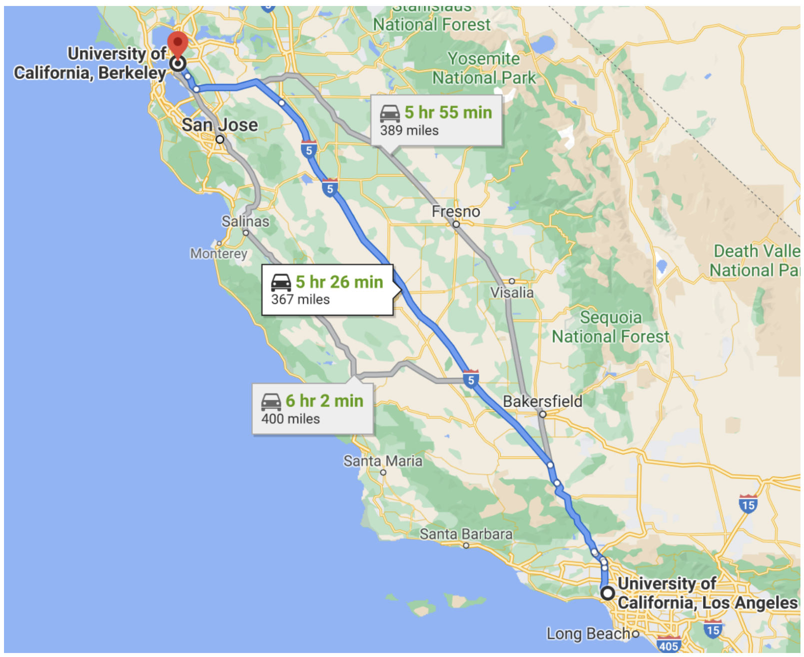
Contents

§1 Lec 1: Mar 29, 2021

§1.1 Introduction

Question 1.1. Why Optimization?

- Find the fastest route from A to B.
- Possible constraints: avoid tolls?



Question 1.2. So what is optimization?

- Optimization is an important tool in decision science and in the analysis of artificial or physical systems.
- An optimization problem involves
  - An objective, which is a scalar, quantitative measure of the performance of the system under study.
  - examples of objectives include profit, time, energy, error, loss, cost, etc.
  - The objective depends on certain characteristics of the system, called variables or unknowns or parameters.
  - Often the variables are restricted, or constrained in some way.
  - The goal of solving an optimization problem is to find values of the variables that satisfy the constraints and optimize/minimize/maximize the objective.

In general, an optimization problem can be summarized as

$$\boxed{\text{Optimized} \quad \text{Objective(Variables)} \qquad \text{Subject to} \quad \text{Constraints on variables}}$$

Applying the optimization framework to solve problems involves three steps:

1. Modeling: identifying objective, variables, and constraints for a given problem.
2. Solving: employing an optimization algorithm to find solutions, usually with the help of a computer.
3. Analyzing: recognizing whether the problem has been successfully solved using optimality conditions.

Mathematically speaking, optimization is the minimization or maximization of a (scalar valued) function subject to constraints on its variables.
We use the following notation

- x is a vector of variables/unknowns/parameters.
- $f(x)$ is the objective function, a scalar function of x that we want to maximize or minimize.
- $C_i(x)$ are constraint functions, which are scalar functions of x that define certain equations or inequalities that the unknown vector x must satisfy.

Using this notation, the optimization problem is

$$\boxed{\text{minimize} \quad \underbrace{f(x)}_{\text{objective}} \quad \text{with } \underbrace{x \in \mathbb{R}^n}_{\text{variables}} \text{ subject to } \begin{cases} c_i(x) = 0, & i \in \underbrace{\mathcal{E}}_{\text{equality}} \\ c_i(x) \geq 0, & i \in \underbrace{\mathcal{I}}_{\text{inequality}} \end{cases}}$$

The set of variables that satisfies all constraints, i.e.,

$$\Omega = \{x \in \mathbb{R}^n : c_i(x) = 0,\, i \in \mathcal{E},\, c_i(x) \geq 0,\, i \in I\}$$

is called the feasible region/set. So the optimization can also be written in an abstract manner as

$$\boxed{\text{minimize } f(x) \text{ with } x \in \mathbb{R}^n \text{ subject to } \underbrace{x \in \Omega}_{\text{feasible/constraint set}}}$$

§1.2 Some Examples

Example 1.1

Consider the problem

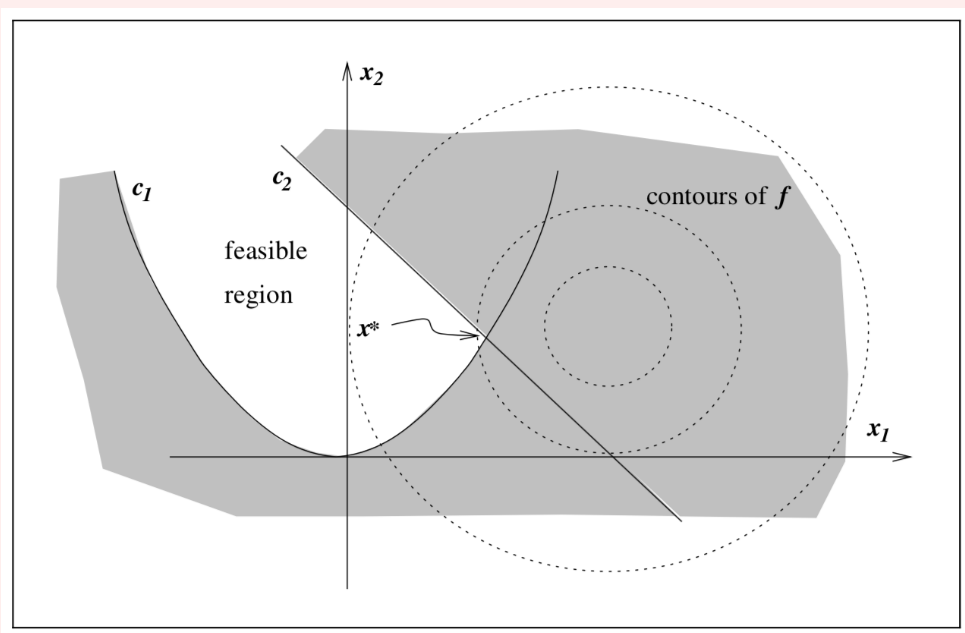
$$\text{minimize } (x_1-2)^2+(x_2-1)^2 \text{ subject to}$$
$$x_1^2 - x_2 \le 0$$
$$x_1 + x_2 \le 2$$

We identify

- the optimization variable $x = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$
- the objective(cost) function $f(x) = (x_1-2)^2+(x_2-1)^2$
- the constraint $c(x) = \begin{bmatrix} c_1(x) \\ c_2(x) \end{bmatrix} = \begin{bmatrix} x_2 - x_1^2 \\ 2 - x_1 - x_2 \end{bmatrix}$, $\mathcal{I} = \{1,2\}$, $\mathcal{E} = \emptyset$.

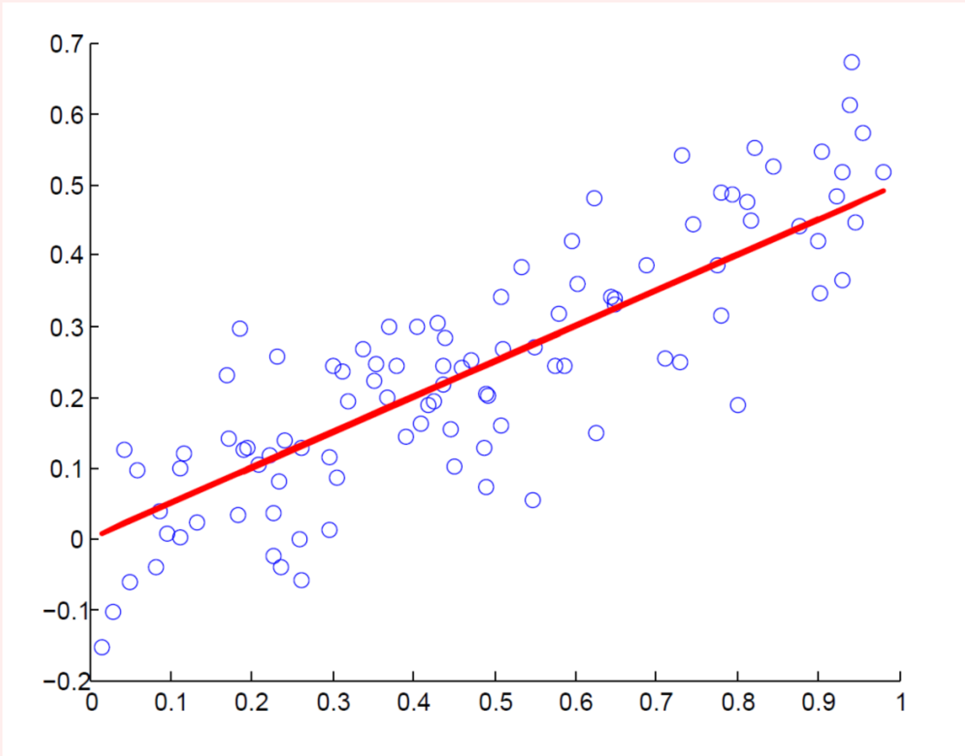
A lot of times we stack all equality constraints and/or inequality constraints into vector functions and write, e.g., $c(x) \ge 0$ meaning element-wise equality or inequality.

$$c(x) = \begin{bmatrix} c_1(x) \\ c_2(x) \\ c_3(x) \\ \vdots \end{bmatrix}$$



Example 1.2 (Linear Regression)

Given a set of feature vectors $a_i \in \mathbb{R}^n$ and outcomes y_i, $i = 1, \ldots, N$, find weights x that predict the outcome accuracy $x^T a_i \approx y_i$.

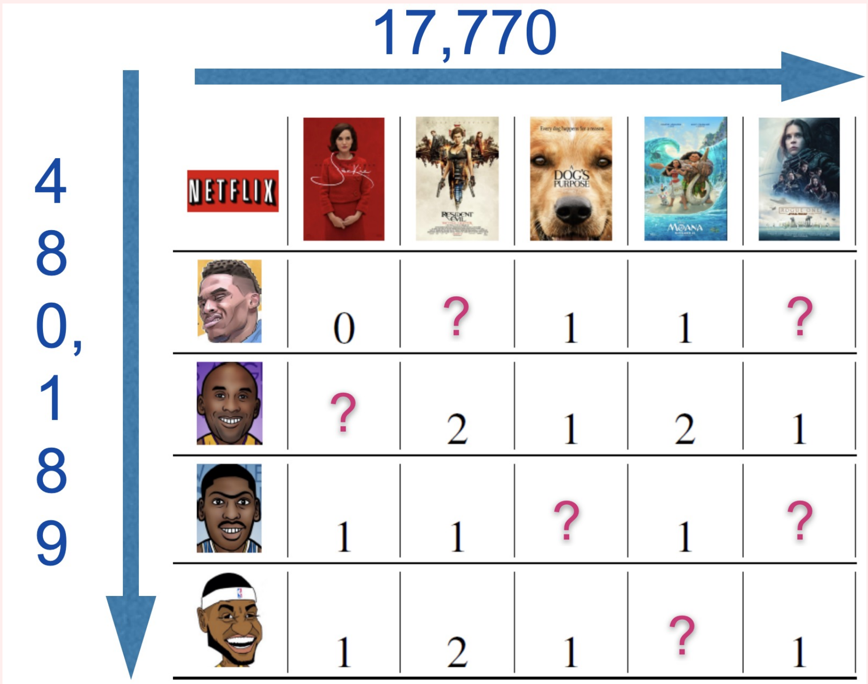


We can find the optimal x by solving the least squares problem.

$$\min_x \sum_{i=1}^{N} \left(y_i - x^T a_i\right)^2$$

Example 1.3

The Netflix prize: predict how a user will rate a movie.



- Some pattern exists: users do not assign ratings completely at random – if you like Godfather I, you'll probably like Godfather II.
- We have lots and lots of data: we know how a user has rated other movies, and we know how other users have rated this (and other) movies.
- Let y_{ij} denote the rate of user i for movie j.
- The Netflix price concerns finding a low-rank matrix X such that $x_{ij} = y_{ij}$ for observed (i, j), related to the following optimization problem

$$\min_X \text{(minimize)} \sum_{\text{observed}(i,j)} (x_{ij} - y_{ij})^2 \text{ subject to } \operatorname{rank}(X) \le r$$

or an alternative way is to

$$\min_X \operatorname{rank}(X) \text{ s.t. } x_{ij} = y_{ij} \text{ observed}(i, j)$$

§1.3 Classification of Optimizations

$$\begin{aligned} &\text{minimize } f(x) \\ &\text{subject to } c_i(x) = 0,\ i \in \mathcal{E} \\ &\qquad\qquad\quad c_i(x) \ge 0, i \in \mathcal{I} \end{aligned}$$

- Unconstrained Optimization: $\mathcal{E} = \mathcal{I} = \emptyset$.
  - Many practical problems are unconstrained or the constraints can be safely discarded.
  - Unconstrained problems arise also as reformulation of constrained ones by replacing the constraints with penalization.
- Contrained Optimization: when constraints are essential for the problem.
- Linear Programming: when the objective function and all the constraints are linear function of x.
  - widely used in management, financial, and economic applications.
- Nonlinear Programming: at least some of the constraints or the objective are nonlinear functions.
  - tend to arise naturally in physical sciences, engineering, signal processing, and machine learning.
  - become more widely used in management and economic sciences as well.
- Global Optimization: aim at fining the global optimal solution, which is generally very challenging.
- Local Optimization: focuses instead on the computation and characterization of local solutions.
- Convex Optimization: the objective function is convex, the equality constraint functions are linear, and the inequality constraint functions are concave.

§2 Lec 2: Mar 31, 2021

§2.1 An Overview of Linear Algebra

Vector Spaces:

In "linear algebra", we denote a vector as a list of numbers,

$$\vec{x} = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} \in \mathbb{R}^n \text{ or } \mathbb{C}^n$$

The general definition of a vector space is as follows.

- A vector space S is composed of a set of elements (called "vectors") and members of a field R (called scalars).
  - Roughly, vectors are objects that can be added together and multiplied by numbers (namely, the scalars).
  - A field is a set of numbers for which addition and multiplication are will defined. We will typically use $R = \mathbb{R}$ or $R = \mathbb{C}$.
- In a vector space, there must be two rules defined for combining vectors and scalars.
  - The first operation is *vector addition*, which associates with any two vectors $\vec{x}, \vec{y} \in S$ the sum $\vec{x} + \vec{y}$ which also must belong to S.
  - The second operation is *scalar multiplication*, which associates with any vector $\vec{x} \in S$ and any scalar $a \in R$ the scalar multiple of $\vec{x}$ by a, denoted by $a\vec{x}$ or $a \cdot \vec{x}$, which must belong to S.
- The vector addition operation must obey four rules:
  1. Commutativity: $\vec{x} + \vec{y} = \vec{y} + \vec{x}\ \forall \vec{x}, \vec{y} \in S$.
  2. Associativity: $\vec{x} + (\vec{y} + \vec{z}) = (\vec{x} + \vec{y}) + \vec{z}\ \forall \vec{x}, \vec{y}, \vec{z} \in S$.
  3. There is a unique "zero vector" $\vec{0} \in S$ such that $\vec{x} + \vec{0} = \vec{x}\ \forall \vec{x} \in S$.
  4. For each $\vec{x} \in S$, there is a vector $-\vec{x} \in S$ such that $\vec{x} + (-\vec{x}) = \vec{0}$.
- The scalar multiplication operation must also obey four rules:
  1. Distributivity: $a\,(\vec{x} + \vec{y}) = a\vec{x} + a\vec{y}$ and $(a+b)\vec{x} = a\vec{x} + b\vec{x}\ \forall \vec{x}, \vec{y} \in S$ and $a, b \in R$.
  2. Associativity: $(ab)\vec{x} = a(b\vec{x})\ \forall \vec{x} \in S$ and $a, b \in R$.
  3. For the multiplicative identity of R (denoted by the scalar $1 \in R$), we have $1 \cdot \vec{x} = \vec{x} \in S$.
  4. For the additive identity of R (denoted by the scalar $0 \in R$), we have $0 \cdot \vec{x} = \vec{0} \in S$.

Linear Subspaces:

The concept of a linear subspace is useful for modeling, approximating signals, discussing the concept of bases,etc.

Definition 2.1 (Linear Subspace) — A nonempty subset T of a vector space S is called a subspace (or linear subspace) of S if

$$a\vec{x} + b\vec{y} \in T$$

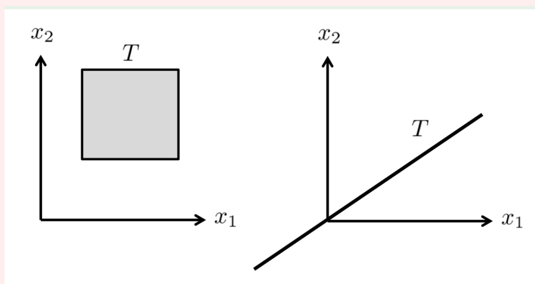
for all $\vec{x}, \vec{y} \in T$ and all $a, b \in R$.

Notes:

- Any linear subspace T must contain $\vec{0}$.
- Any vector space S is a linear subspace (of itself).
- Any linear subspace T meets all the properties of a vector space.

Example 2.2

Are either of these a subspace of $S = \mathbb{R}^2$



Left: No! Right: Yes!

Example 2.3

Which of the following are subspaces?

1. $S = \mathbb{R}^5$ and $T = \left\{\vec{x} \in \mathbb{R}^5 : x_4 = 0\right\}$ – Yes!
2. $S = \mathbb{R}^5$ and $T = \left\{\vec{x} \in \mathbb{R}^5 : x_4 = 1\right\}$ – No, add any two vectors in T and the sum will not belong to T.
3. $S = \mathbb{R}^5$ and T is the set of vectors in $\mathbb{R}^5$ with no more than 3 nonzero entries – No, can add certain vectors in T to get up to 5 nonzero elements.

Linear Combinations:

Linear combinations are used to build new vectors a weighted sum of other vectors.

Definition 2.4 (Linear Combination) — Let $M = \{\vec{v_1}, \vec{v_2}, \dots, \vec{v_n}\}$ be a collection of vectors in a vector space S. (we will stick with finite collections at the moment).
A linear combination of vectors in M is a sum of the form

$$a_1\vec{v_1} + a_2\vec{v_2} + \dots + a_n\vec{v_n}$$

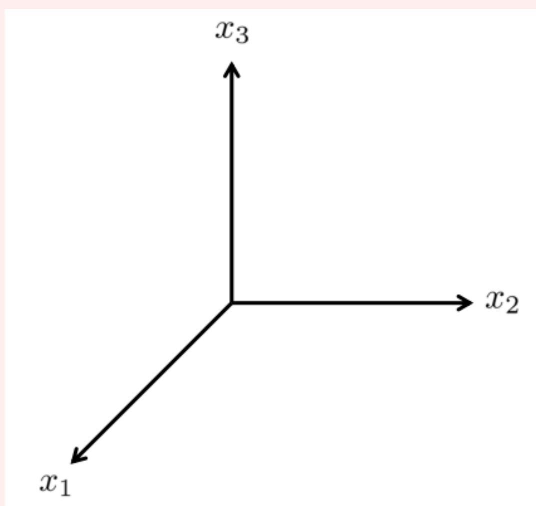
for some $a_1, a_2, \dots, a_n \in R$. Since S is a vector space, this sum must belong to S.

Mentally, you might find it useful to replace "linear combination" with "weighted sum", although this is not standard terminology.

Definition 2.5 (Span) — Let $M = \{\vec{v_1}, \dots, \vec{v_n}\}$ be a finite collection of vectors in a vector space S. The span of M, denoted by $\operatorname{span}(M)$ or $\operatorname{span}\{\vec{v_1}, \dots, \vec{v_n}\}$, is the set of all linear combinations of vectors in M.

Example 2.6

Consider the vectors $\vec{v_1} = \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix}$ and $\vec{v_2} = \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}$. What is $\operatorname{span}\{\vec{v_1}, \vec{v_2}\}$?



The entire (x_1, x_2)-plane.

When we are in $\mathbb{R}^n$ for some finite n, it is common to use matrix-vector notation as shorthand for linear combinations:

- Suppose $\vec{x} = c_1\vec{p_1} + \dots + c_k\vec{p_k}$
- Then we may define the $n \times k$ matrix

$$A = \begin{bmatrix} \vec{p_1} & \vec{p_2} & \dots & \vec{p_k} \end{bmatrix}$$

and the $k \times 1$ vector

$$\vec{c} = \begin{bmatrix} c_1 \\ c_2 \\ \vdots \\ c_k \end{bmatrix}$$

and this allows us to write the $n \times 1$ vector $\vec{x}$ as $\vec{x} = A\vec{c}$.

Definition 2.7 (Linear Dependence) — A finite set of vectors $\vec{v_1}, \ldots, \vec{v_n}$ in a vector space S is said to be linearly dependent if there exists scalars $a_1, \ldots, a_n \in R$, not all equal to zero, such that

$$a_1\vec{v_1} + \ldots + a_n\vec{v_n} = \vec{0}$$

Definition 2.8 (Linear Independence) — A finite set of vectors $\vec{v_1}, \ldots, \vec{v_n}$ in a vector space S is said to be linearly independent if

$$a_1\vec{v_1} + \ldots + a_n\vec{v_n} = \vec{0}$$

only when all $a_k = 0$.

Every vector in the span of a linearly independent set of vectors has a unique expansion in terms of those vectors. This is formalized in the following lemma.

Lemma 2.9

Suppose $\vec{v_1}, \ldots, \vec{v_n}$ are linearly indep. and suppose

$$a_1\vec{v_1} + \ldots + a_n\vec{v_n} = b_1\vec{v_1} + \ldots + b_n\vec{v_n}$$

for some scalars $a_1, \ldots, a_n \in R$ and $b_1, \ldots, b_n \in R$. Then $a_k = b_k$ for $k = 1, 2, \ldots, n$.

Proof. Note that $\sum_{k=1}^{n} (a_k - b_k)\, \vec{v_k} = \vec{0}$. Since $\vec{v_1}, \ldots, \vec{v_n}$ are linearly indep., it must follow that $a_k - b_k = 0$ for all $k = 1, 2, \ldots, n$. □

Bases and Dimension:

Now that we know how to combine vectors (via linear combinations), let's think about important sets of vectors we'd be interested in combining.

Definition 2.10 (Basis) — A finite set of vectors $\vec{v_1}, \ldots, \vec{v_n}$ in a vector space S is said to form a basis for S if the following two conditions are satisfied:

1. $\vec{v_1}, \ldots, \vec{v_n}$ are linearly independent.
2. $\text{span}\{\vec{v_1}, \ldots, \vec{v_n}\} = S$.

A vector space wit a finite basis, as in the above definition, is said to be finite-dimensional. Any two bases for a finite-dimensional vector space contain the same number of elements. This leads to a meaningful definition of dimension.

Definition 2.11 (Dimension) —

- For a vector space S that can be spanned using a finite set of basis vectors, the dimension of S is the number of vectors required in any basis for S.
- For a vector space S that cannot be spanned using a finite set of basis vectors, the dimension of S is said to be infinite.

Example 2.12 (Bases for $S = \mathbb{R}^n$)

- The standard, or canonical, basis for $\mathbb{R}^n$ is given by:

$$\{\vec{v_1}, \vec{v_2}, \dots, \vec{v_n}\} = \left\{ \begin{bmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \\ \vdots \\ 0 \end{bmatrix}, \dots, \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 1 \end{bmatrix} \right\}$$

- Any set of n linearly indep. vectors in $\mathbb{R}^n$ forms a basis for $\mathbb{R}^n$.

§3 Lec 3: Apr 2, 2021

§3.1 Lec 2 (Cont'd)

Normed Linear Spaces:

A norm is a function used to measure the size of vectors in a vector space.

Definition 3.1 (Norm) — A norm $\|\cdot\|$ on a vector space S is a mapping $\|\cdot\| : S \to \mathbb{R}$ with the following properties:

1. $\|\vec{x}\| \geq 0$ for all $\vec{x} \in S$, and $\|\vec{x}\| = 0 \iff \vec{x} = \vec{0}$.
2. Triangle inequality: $\|\vec{x} + \vec{y}\| \leq \|\vec{x}\| + \|\vec{y}\|$ for all $\vec{x}, \vec{y} \in S$.
3. $\|a\vec{x}\| = |a| \cdot \|\vec{x}\|$ for any $a \in R$ and $\vec{x} \in S$.

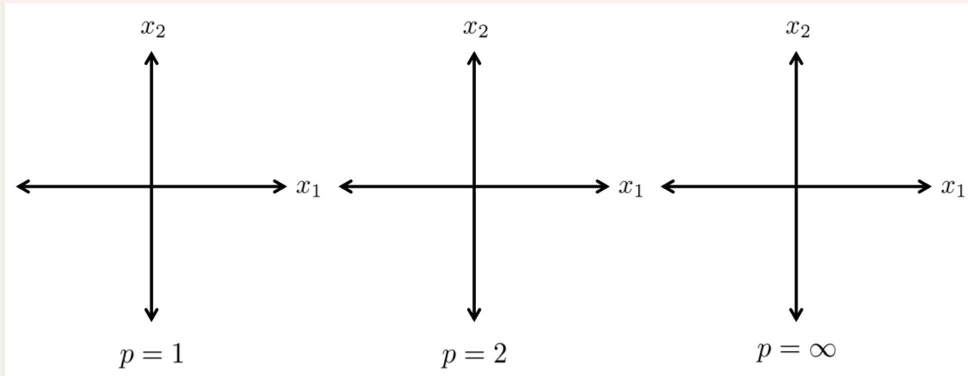
Definition 3.2 (Normed Linear Space) — A normed linear space is a vector space S together with a valid norm $\|\cdot\| : S \to \mathbb{R}$.

The l_p metrics for vectors in $\mathbb{R}^n$ extend naturally to l_p norms for these same spaces:

- l_1 norm: $\|\vec{x}\|_1 = \sum_{i=1}^n |x_i|$
- l_2 ("Euclidean") norm:
$$\|\vec{x}\|_2 = \left(\sum_{i=1}^n |x_i|^2\right)^{\frac{1}{2}}$$
- l_p norm for $1 \leq p < \infty$:
$$\|\vec{x}\|_p = \left(\sum_{i=1}^n |x_i|^p\right)^{\frac{1}{p}}$$
- l_∞ norm:
$$\|\vec{x}\|_\infty = \max_{i=1,\ldots,n} |x_i|$$
- Bonus – l_0 norm (not really a norm – not satisfied the definition of a norm)
$$\|\vec{x}\|_0 = \#\text{ of nonzeros in } \vec{x}$$

Example 3.3

The "unit ball" in a normed linear space is the set of all vectors in S having norm less than or equal to 1. Suppose $S = \mathbb{R}^2$ and draw the l_p balls for $p = 1, 2, \infty$.



- $p = 1$: $\|\vec{x}\|_1 = |x_1| + |x_2| \leq 1$ – diamond.
- $p = 2$: $\|\vec{x}\|_2 = x_1^2 + x_2^2 \leq 1$ – circle.
- $p = \infty$: $\|\vec{x}\|_\infty = \max\{|x_1|, |x_2|\}$ – square.

Inner Product Spaces:

An inner product is a function used to compare two vectors in a vector space. The concept of an inner product will give us additional geometric structure beyond what is available in general normed linear spaces. In particular, using an inner product we can define a meaningful measure of the angle between two vectors, discuss orthonormal bases and orthogonal projections, etc.

Definition 3.4 (Inner Product) — An inner product $\langle\cdot,\cdot\rangle$ on a vector space S is a mapping $\langle\cdot,\cdot\rangle : S \times S \to R$ with the following properties:

1. $\langle\vec{x},\vec{y}\rangle = \langle\vec{y},\vec{x}\rangle^*$ for all $\vec{x},\vec{y} \in S$.
2. For any $\vec{x},\vec{y},\vec{z} \in S$ and any $a, b \in R$, $\langle a\vec{x} + b\vec{y},\vec{z}\rangle = a\langle\vec{x},\vec{z}\rangle + b\langle\vec{y},\vec{z}\rangle$.
3. For any $\vec{x} \in S$, $\langle\vec{x},\vec{x}\rangle$ is real-valued and non-negative, and $\langle\vec{x},\vec{x}\rangle = 0$ iff $\vec{x} = \vec{0}$.

Definition 3.5 (Inner Product Space) — An inner product space is a vector space S together with a valid inner product $\langle\cdot,\cdot\rangle : S \times S \to R$.

Definition 3.6 (Orthogonality) — Two vectors $\vec{x}$ and $\vec{y}$ in an inner product space S are said to be orthogonal if $\langle\vec{x},\vec{y}\rangle = 0$.

Example 3.7

When $S = \mathbb{R}^n$, the standard inner product between two vectors $\vec{x}, \vec{y} \in S$ is given by

$$\langle \vec{x}, \vec{y} \rangle = \sum_{i=1}^{n} x_i y_i = \vec{y}^\top \vec{x}$$

The standard inner product on $\mathbb{R}^n$ is also known as the dot product.
$S = \mathbb{C}^n : \langle \vec{x}, \vec{y} \rangle = \vec{y}^H \vec{x} = \sum_{i=1}^{n} x_i y_i^*$ where H denotes conjugate transpose/hermitian.

Before we get to the connection between inner product and angles, it is worth noting that inner products can actually be used to measure the length of vectors (and thus distances between vectors as well).
In particular, any valid inner product induces a valid norm by

$$\|\vec{x}\| = \sqrt{\langle \vec{x}, \vec{x} \rangle}$$

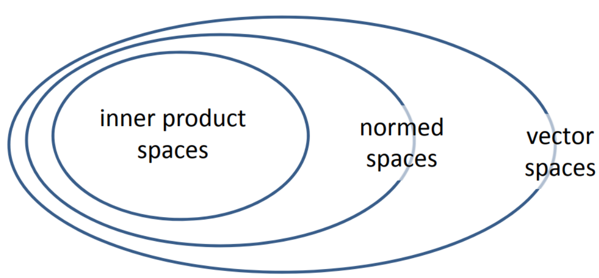
Example 3.8

When $S = \mathbb{R}^n$, the standard inner product induces the following norm:

$$\|\vec{x}\| = \sqrt{\langle \vec{x}, \vec{x} \rangle} = \sqrt{\sum_{i=1}^{n} x_i^2}$$

We can recognize this as the l_2 norm.

Other l_p norm (for $p \neq 2$)cannot be induced by inner products. Because every valid inner product induces a valid norm, every inner product space is also a normed linear space. But, not every normed linear space is also an inner product space:



Recall the definition of a basis in a generic, finite-dimensional vector space. In inner product spaces, a particularly useful class of bases are orthogonal bases.

Definition 3.9 (Orthogonal Basis) — A finite sets of non-zero vectors $\vec{v_1}, \ldots, \vec{v_n}$ in an inner product space S is said to form an orthogonal basis for S if the following two conditions are satisfied:

1. $\langle \vec{v_k}, \vec{v_l} \rangle = 0$ for all $k \neq l$ (note that this implies the vectors are linearly indep.)
2. $\operatorname{span}\{\vec{v_1}, \ldots, \vec{v_n}\} = S$

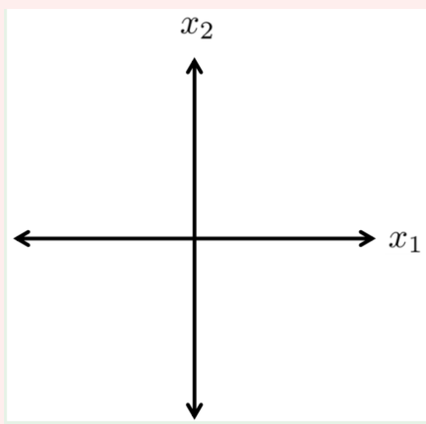
Definition 3.10 (Orthonormal Basis) — An orthogonal basis is called orthonormal basis or orthobasis if every basis vector $\vec{v_k}$ has unit norm (i.e., $\|\vec{v}_k = 1\|$) according to the induced norm in the inner product space.

Example 3.11

Using the standard inner product,

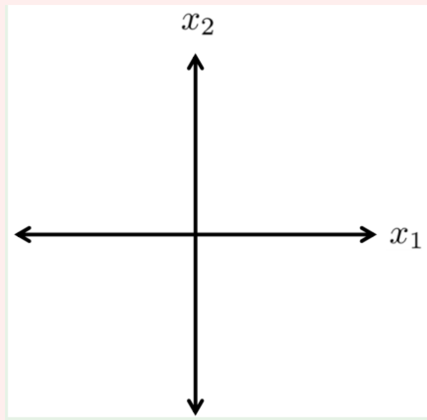
$$\vec{v_1} = \begin{bmatrix} 1 \\ 0 \end{bmatrix} \text{ and } \vec{v_2} = \begin{bmatrix} 0 \\ 1 \end{bmatrix}$$

form an orthobasis for $\mathbb{R}^2$.



Example 3.12 (Rotation of 45 degree from the last example)
Another possible orthobasis in $\mathbb{R}^2$ is given by

$$\vec{v_1} = \begin{bmatrix} \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} \end{bmatrix} \text{ and } \vec{v_2} = \begin{bmatrix} \frac{1}{\sqrt{2}} \\ -\frac{1}{\sqrt{2}} \end{bmatrix}$$



§3.2 Linear Operators

Operators are transformations that map vectors in some vector space to vectors in some other (possible different) vector space.

Definition 3.13 (Linear Operator) — Suppose X and Y are vector spaces. We say the operator $A : X \to Y$ is a linear operator if

$$A(\alpha_1 \vec{x}_1 + \alpha_2 \vec{x}_2) = \alpha_1 A\vec{x}_1 + \alpha_2 A\vec{x}_2$$

for all $\alpha_1, \alpha_2 \in R$ and $\vec{x}_1, \vec{x}_2 \in X$.

Fact 3.1. If $X = \mathbb{R}^n$ and $Y = \mathbb{R}^m$, any linear operator from X to Y can be represented as multiplication by an $m \times n$ matrix.

Therefore, a particularly interesting class of linear operators for us will simply be matrices.

§3.3 Operator Norms

Roughly speaking, operator norms help us talk about the "gain" of a system.

Definition 3.14 (Operator Norm) — Let X and Y be normed linear spaces with corresponding norms $\|\cdot\|_X$ and $\|\cdot\|_Y$ and suppose $A : X \to Y$ is linear operator. The operator norm $\|A\|$ is defined as

$$\|A\| := \sup_{\vec{0} \neq \vec{x} \in X} \frac{\|A\vec{x}\|_Y}{\|\vec{x}\|_X}$$

This is equivalent to

$$\|A\| := \sup_{\vec{x} \in X,\, \|\vec{x}\|_X = 1} \|A\vec{x}\|_Y$$

§4 Lec 4: Apr 5, 2021

§4.1 Operator Norms (Cont'd)

Any operator norm as defined in the last lecture will satisfy the following properties

1. $\|A\| \geq 0$ with equality iff $A = 0$
2. $\|\alpha A\| = |\alpha|\|A\|$ for all $\alpha \in R$
3. $\|A + B\| \leq \|A\| + \|B\|$ for all linear operators A and B between the vector spaces X and Y.
   - Examining these first three properties, we see that $\|A\|$ is a valid norm on the vector space of linear operators!
4. $\|A\vec{x}\|_Y \leq \|A\|\|\vec{x}\|_X$ for all $\vec{x} \in X$
   - Therefore, the operator norm helps us bound how much an operator can "amplify" a signal.
5. $\|AB\| \leq \|A\|\|B\|$
6. If $X = Y$ and if $\|A\| < 1$, then we can write

$$\sum_{i=0}^{\infty} A^i = (I - A)^{-1}$$

   just as for a scalar $a \in \mathbb{R}$ with $|a| < 1$, we can write $\sum_{i=0}^{\infty} a_i = \frac{1}{1-a}$

Let's restrict our attention to the special case where $X = \mathbb{R}^n$ and $Y = \mathbb{R}^m$. For these choices of X and Y, recall that any linear operator $A : X \to Y$ can be represented as multiplication by an $m \times n$ matrix. In such a case, the operator norm $\|A\|$ is also called a matrix norm.

When X and Y are both equipped with the l_p norm for $p \in [1, \infty]$, we can write

$$\|A\|_p := \sup_{\vec{x} \in X,\, \|\vec{x}\|_p = 1} \|A\vec{x}\|_p$$

We can relate $\|A\|_p$ to certain properties of the matrix A:

- In the case $p = \infty$, letting $\vec{y} = A\vec{x}$, we have

$$\|A\|_\infty := \sup_{\|\vec{x}\|_\infty = 1} \|\underbrace{A\vec{x}}_{\vec{y}}\|_\infty = \sup_{\|\vec{x}\|_\infty = 1} \left\| \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_m \end{bmatrix} \right\|_\infty$$

  Note that $|y_i| = \left|\sum_{j=1}^n a_{ij} x_j\right|$. Overall $\vec{x}$ with $\|\vec{x}\|_\infty = 1$, the largest $|y_i|$ for a given $i = 1, 2, \ldots, m$ is achieved by taking $x_j = \text{sign}(a_{ij})$ for $j = 1, 2, \ldots, n$, and for this choice of i and $\vec{x}$, we will have

$$|y_i| = \sum_{j=1}^n |a_{ij}| = \text{absolute sum of row i of } A$$

Thus,

$$\|A\|_\infty = \max_{i=1,2,\ldots,m} \sum_{j=1}^{n} |a_{ij}| = \text{maximum absolute row sum of } A$$

- Similarly, in the case $p = 1$, we have

$$\|A\|_1 = \max_{j=1,2,\ldots,n} = \sum_{i=1}^{m} |a_{ij}| = \text{maximum absolute column sum of } A$$

- When $p = 2$, $\|A\|_2$ is also referred to as the spectral norm of A. We can understand $\|A\|_2$ geometrically: the operator A maps the l_2 unit ball in $\mathbb{R}^n$ to an ellipsoid in $\mathbb{R}^m$.

  The length of the major axis of the ellipsoid is equal to $\|A\|_2$. We can also write

$$\|A\|_2 = \sqrt{\lambda_{\max}(A^\top A)} = \sigma_{\max}(A)$$

  where $\lambda_{\max}$ denotes the largest eigenvalue of a matrix and $\sigma_{\max}$ denotes the largest singular value of a matrix. If the matrix A happens to be symmetric (i.e. if $A = A^\top$) then we can also write

$$\|A\|_2 = \max_i |\lambda_i(A)|$$

There is also a special type of "matrix norm" that does not actually follow the definition of an operator norm,

$$\|A\| \coloneqq \sup_{\vec{x} \in X,\, \|\vec{x}\|_X = 1} \|A\vec{x}\|_Y$$

In particular, the Frobenius form of a matrix A is defined to be

$$\|A\|_F \coloneqq \sqrt{\sum_{i=1}^{m} \sum_{j=1}^{n} |a_{ij}|^2} = \sqrt{\text{trace}(A^H A)}$$

Note that Frobenius form $\|A\|_F$ is not an operator norm, because $A = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$.

$$\|A\|_F = \sqrt{2} > 1 = \sup_{\|x\|_X = 1} \|Ax\|_Y = \|x\|_X = 1$$

$$\|A\|_F = \sqrt{1+1} = \sqrt{2}$$

$$\|A\|_2 = \sup_{\|x\|_2 = 1} \|Ax\|_2 = 1$$

How about $A = \begin{bmatrix} 1 & 1 \\ 2 & 0 \end{bmatrix}$?

$$\|A\|_F = \sqrt{1+1+4+0} = \sqrt{6}$$

§4.2 Inverse Operator

Definition 4.1 (Invertibility) — A linear operator $A : X \to Y$ between two vector spaces X and Y is said to be invertible if there exists an operator $A^{-1} : Y \to X$ s.t.

- $AA^{-1} = I$, i.e., $AA^{-1}\vec{y} = \vec{y}$ for all $\vec{y} \in Y$ and
- $A^{-1}A = I$, i.e., $A^{-1}A\vec{x} = \vec{x}$ for all $\vec{x} \in X$.

In such a case, A^{-1} is referred to as the inverse of A.

Lemma 4.2

If A is an invertible linear operator, A^{-1} is itself a linear operator.

Invertibility is a topic of interest when we want to find an exact solution to a linear equation. Let us again restrict our attention to the special case where $X = \mathbb{R}^n$ and $Y = \mathbb{R}^m$. For these choices of X and Y, recall that any linear operator $A : X \to Y$ can be represented as multiplication by an $m \times n$ matrix.

Fact 4.1. An $m \times n$ matrix A cannot be invertible unless it is square.

Not all square matrices are invertible. An invertible matrix is also known as the nonsingular matrix.

Proposition 4.3

If A is a square matrix, the following statements are all equivalent

- A is invertible
- A is nonsingular
- $\det(A) \neq 0$
- $A\vec{x} = \vec{0} \iff \vec{x} = \vec{0}$
- The rows of A are linearly indep.
- The columns of A are linearly indep.
- $\dim(\mathcal{N}(A)) = 0$, i.e., $\mathcal{N} = \left\{\vec{0}\right\}$
- $\dim(\mathcal{R}(A)) = n$.
- A is full rank
- All eigenvalues of A are nonzero
- The matrix $A^\top A$ is positive definite.
- $A^\top$ is invertible.

§4.3 Adjoint Operators

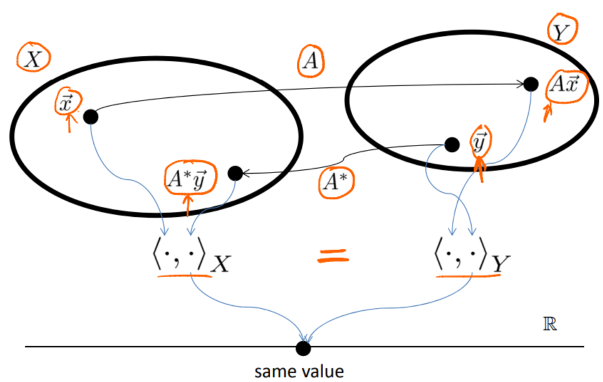
Definition 4.4 (Adjoint of an Operator) — Let $A : X \to Y$ be a bounded ($\|A\| < \infty$) linear operator between two inner product spaces X and Y. The adjoint of A, denoted $A^* : Y \to X$ is the unique operator such that

$$\langle A\vec{x}, \vec{y}\rangle_Y = \langle \vec{x}, A^*\vec{y}\rangle_X$$

for all $\vec{x} \in X$ and $\vec{y} \in Y$.

Definition 4.5 (Self-Adjoint Operator) — An operator $A : X \to X$ is said to be self-adjoint if $A = A^*$.

An illustration:



Lemma 4.6

If A is bounded linear operator with adjoint A^* then A^* is itself a bounded linear operator, and

$$\|A^*\| = \|A\|$$

Lemma 4.7

If A is bounded linear operator with adjoint A^* then $(A^*)^* = A$

Lemma 4.8

If A is an invertible bounded linear operator with adjoint A^* and bounded inverse A^{-1}, then

$$(A^{-1})^* = (A^*)^{-1}$$

In the special case where $X = \mathbb{R}^n$ and $Y = \mathbb{R}^m$, recall that any linear operator $A : X \to Y$ can be represented as multiplication by an $m \times n$ matrix.

Question 4.1. What is the adjoint of a matrix?

Using the standard inner product on X and Y, the adjoint of A is the unique operator s.t.

$$\vec{y}^H A\vec{x} = \langle A\vec{x}, \vec{y}\rangle = \langle \vec{x}, A^*\vec{y}\rangle = \vec{y}^H (A^*)^H \vec{x}$$

for all $\vec{x}$ and $\vec{y}$. This requires that $A = (A^*)^H$ which is satisfied by taking

$$A^* = A^H$$

Therefore, the adjoint of a matrix is simply its conjugate transpose (not its inverse). Self-adjoint matrices satisfy $A = A^H$. These are also known as symmetric (if real), conjugate symmetric or Hermitian (if complex).

Theorem 4.9

Let A be a real-valued $m \times n$ matrix. For a fixed $\vec{y} \in \mathbb{R}^m$, the vector $\vec{x} \in \mathbb{R}^n$ is a minimizer of $\|\vec{y} - A\vec{x}\|_2 \iff$

$$A^\top A\vec{x} = A^\top \vec{y}$$

If $A^\top A$ is invertible, then the unique minimizer of $\|\vec{y} - A\vec{x}\|_2$ is given by

$$\vec{x} = (A^\top A)^{-1} A^\top \vec{y}$$

The same theorem holds if $A, \vec{x}$ and $\vec{y}$ are all complex-valued.

§5 Lec 5: Apr 7, 2021

§5.1 Fundamental Subspaces of Linear Operators

Definition 5.1 (Range) — Let $A : X \to Y$ be a linear operator between two vector spaces X and Y. The range or range space of A, denoted by $\mathcal{R}(A)$, is defined to be

$$\mathcal{R}(A) \coloneqq \{\vec{y} \in Y : A\vec{x} = \vec{y} \text{ for some } \vec{x} \in X\}$$

The range space is a linear subspace of Y.

Definition 5.2 (Nullspace) — Let $A : X \to Y$ a linear operator between two vector spaces X and Y. The nullspace of A, denoted by $\mathcal{N}(A)$, is defined to be

$$\mathcal{N}(A) \coloneqq \left\{\vec{x} \in X : A\vec{x} = \vec{0}\right\}$$

The nullspace is a linear subspace of X.

Again, we consider the case where $X = \mathbb{R}^n$ and $Y = \mathbb{R}^m$ (or where $X = \mathbb{C}^n$ and $Y = \mathbb{C}^m$) and $A : X \to Y$ can be represented as multiplication by an $m \times n$ matrix.
When A is a matrix, $\mathcal{R}(A)$ is just the span of the columns of A:

$$\mathcal{R}(A) = \operatorname{colspan}(A)$$

For any $m \times n$ matrix A,

$$\operatorname{rank}(A) \le \min\{m, n\}$$

We say that A is full rank if $\operatorname{rank}(A) = \min\{m, n\}$, otherwise we call it rank deficient.
We can related the rank of A to the dimensions of the four fundamental subspaces of A:

- $\dim(\mathcal{R}(A)) = \dim(\operatorname{colspan}(A)) = \operatorname{rank}(A)$
- $\dim(\mathcal{N}(A)) = n - \operatorname{rank}(A)$
- $\dim(\mathcal{R}(A^*)) = \dim(\operatorname{rowspan}(A)) = \operatorname{rank}(A)$
- $\dim(\mathcal{N}(A^*)) = m - \operatorname{rank}(A)$
- $\dim(\mathcal{R}(A)) + \dim(\mathcal{N}(A)) = n = \#$ columns of A.

For two matrices A and B, we have

- $\operatorname{rank}(AB) \le \min\{\operatorname{rank}(A), \operatorname{rank}(B)\}$
- $\operatorname{rank}(A + B) \le \operatorname{rank}(A) + \operatorname{rank}(B)$

§5.2 Projection Operators

Definition 5.3 (Projection Operator) — A linear operator $P : X \to X$ from a vector space X into itself is called a projection or a projection operator if

$$P^2 = P$$

i.e., if $P\left(P(\vec{x})\right) = P(\vec{x})$ for all $\vec{x} \in X$.

For $P^2 = P$, P is called idempotent operator.

Definition 5.4 (Orthgonal Projection Operator) — A projection operator P is an inner product space X is called an orthogonal projection or an orthogonal projection operator if

$$\mathcal{R}(P) \perp \mathcal{N}(P)$$

i.e., if $\langle \vec{x}, \vec{y} \rangle = 0$ for all $\vec{x} \in \mathcal{R}(P)$ and $\vec{y} \in \mathcal{N}(P)$.

We notice

$$\vec{x} = \underbrace{P\vec{x}}_{\in\mathcal{R}(P)} + \underbrace{(I-P)\vec{x}}_{\in\mathcal{N}(P)}$$

If P is an orthogonal projection operator,

$$\langle P\vec{x}, (I-P)\vec{x} \rangle = 0$$

Lemma 5.5

A bounded linear operator $P : X \to X$ on an inner product space X is an orthogonal projection iff

1. $P^2 = P$ and
2. $P = P^*$

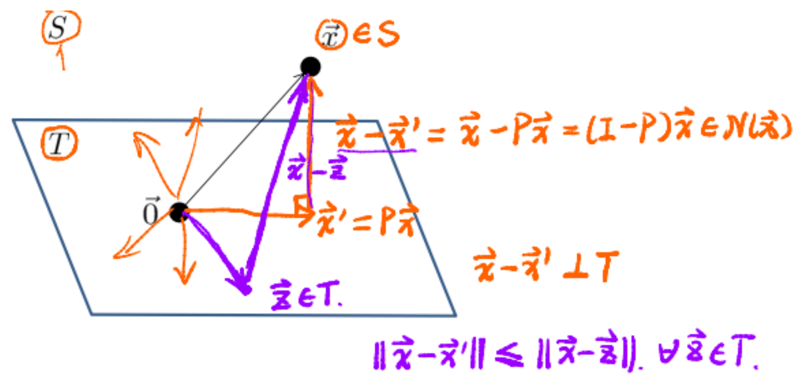
Theorem 5.6

Suppose S is an inner product space and suppose T is a linear subspace of S. For a given vector $\vec{x} \in S$, there is a unique vector $\vec{x}' \in T$ such that $\|\vec{x} - \vec{x}'\| \le \|\vec{x} - \vec{z}\|$ for all $\vec{z} \in T$. Furthermore, this minimizer has the property that

$$\vec{x} - \vec{x}' \perp T$$

i.e. $\langle \vec{x} - \vec{x}', \vec{y} \rangle = 0$ for all $\vec{y} \in T$.

The minimizing vector $\vec{x}'$ is referred to as the orthogonal projection of $\vec{x}$ onto T. In other words, $\vec{x}' = P\vec{x}$, where P is an orthogonal projection operator with $\mathcal{R}(P) = T$.



For an $n \times n$ matrix P,

- P is a projection if $P^2 = P$
- P is orthogonal projection if $P^2 = P$ and $P^\top = P$

Example 5.7

Consider the operator $P : \mathbb{R}^2 \to \mathbb{R}^2$ defined as

$$P\left(\begin{bmatrix} x_1 \\ x_2 \end{bmatrix}\right) = \begin{bmatrix} x_1 \\ x_1 \end{bmatrix}$$

We can express P as the 2×2 matrix $P = \begin{bmatrix} 1 & 0 \\ 1 & 0 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} x_1 \\ x_1 \end{bmatrix}$ Consider:

- Does $P^2 = P$? Yes!
- Does $P^\top = P$? No!
- What is $\mathcal{R}(P)$? the line $x_2 = x_1$
- What is $\mathcal{N}(P)$? x_2-axis
- Is $\mathcal{R}(P) \perp \mathcal{N}(P)$? No!

Example 5.8

Consider the operator $P : \mathbb{R}^3 \to \mathbb{R}^3$ defined as

$$P\left(\begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix}\right) = \begin{bmatrix} x_1 \\ x_2 0 \end{bmatrix}$$

where

$$P = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{bmatrix}$$

Consider:

- Does $P^2 = P$? Yes!
- Does $P^\top = P$? Yes!
- What is $\mathcal{R}(P)$? the plane $x_1 - x_2$
- What is $\mathcal{N}(P)$? x_3-axis
- Is $\mathcal{R}(P) \perp \mathcal{N}(P)$? Yes!

Consider a set of m linearly indep. vectors $\vec{v_1}, \vec{v_2}, \ldots, \vec{v_m} \in \mathbb{R}^n$ or $\mathbb{C}^n$. We can construct an orthogonal projection matrix P onto the subspace $T = \operatorname{span}\{\vec{v_1}, \vec{v_2}, \ldots, \vec{v_m}\}$ as follows:

1. Construct an $n \times m$ matrix

$$A = \begin{bmatrix} \vec{v_1} & \vec{v_2} & \ldots & \vec{v_m} \end{bmatrix}$$

   Note that colspan(A) = T.

2. Let

$$P = A\left(A^\top A\right)^{-1} A^\top = AA^\dagger$$

Example 5.9

Consider the vectors

$$\vec{v_1} = \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix} \text{ and } \vec{v_2} = \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}$$

The orthogonal projection can be constructed as follows

$$P = AA^\dagger = A(A^\top A)^{-1} A^\top = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{bmatrix}$$

§6 Lec 6: Apr 9, 2021

§6.1 Motivating Examples

Consider a 2×2 matrix

$$A = \begin{bmatrix} 1 & \frac{1}{2} \\ \frac{1}{2} & 1 \end{bmatrix}$$

For what nonzero vectors $\vec{x} \in S$ (eigenvectors) and scalar $\lambda \in \mathbb{C}$ (eigenvalue), do we have that $A\vec{x} = \lambda x$?

- We know that $\vec{x}$ and λ must satisfy $(A - \lambda I)\vec{x} = \vec{0}$.
- Thus, $\vec{x}$ must be in $\mathcal{N}(A - \lambda I)$.
- Thus, $A - \lambda I$ must have a nontrivial nullspace.
- Thus, $A - \lambda I$ must be singular.
- We can solve for λ s.t. $\det(A - \lambda I) = 0$:

$$\det(A - \lambda I) = \det\left(\begin{bmatrix} 1-\lambda & \frac{1}{2} \\ \frac{1}{2} & 1-\lambda \end{bmatrix}\right) = (1-\lambda)^2 - \frac{1}{4} = 0$$

  which equals 0 for $\lambda = 1.5$ or $\lambda = 0.5$

- Now we know the eigenvalues. What are the corresponding eigenvectors?
- For $\lambda = 1.5$, we need $A\vec{x} = 1.5\vec{x}$.

$$\begin{bmatrix} 1 & \frac{1}{2} \\ \frac{1}{2} & 1 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} 1.5x_1 \\ 1.5x_2 \end{bmatrix}$$

  which requires $x_1 = x_2$. To have unit norm, we can choose

$$\vec{x} = \begin{bmatrix} \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} \end{bmatrix}$$

  but notice that any rescaling of this $\vec{x}$ is also an eigenvector.

- For $\lambda = 0.5$, we need $A\vec{x} = 0.5\vec{x}$.

$$\begin{bmatrix} 1 & \frac{1}{2} \\ \frac{1}{2} & 1 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} 0.5x_1 \\ 0.5x_2 \end{bmatrix}$$

  This requires $x_1 = -x_2$. To have unit norm, we can choose

$$\vec{x} = \begin{bmatrix} \frac{1}{\sqrt{2}} \\ -\frac{1}{\sqrt{2}} \end{bmatrix}$$

  but notice that any rescaling of this $\vec{x}$ is also an eigenvector.

- We say that the eigenvectors of A are

$$\vec{v}_1 = \begin{bmatrix} \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} \end{bmatrix} \text{ and } \vec{v}_2 = \begin{bmatrix} \frac{1}{\sqrt{2}} \\ -\frac{1}{\sqrt{2}} \end{bmatrix}$$

  with the understanding that any rescaling of either one is also an eigenvector.

§6.2 Eigenvalues and Eigenvectors

Big picture:

- first find the eigenvalues
  - Suppose A is an $n \times n$ matrix.
  - We want to know: for what values of $\lambda \in \mathbb{C}$ does there exist a non-zero $\vec{x} \in \mathbb{C}^n$ s.t.
  $$A\vec{x} = \lambda\vec{x}?$$
  - We know that for such a $(\lambda, \vec{x})$ pair to exist, we must have
  $$(A - \lambda I)\,\vec{x} = \vec{0}$$
  and therefore, $\vec{x} \in \mathcal{N}(A - \lambda I)$.
  - For such a non-zero $\vec{x}$ to exist in the nullspace of $A - \lambda I$, we need
  $$\dim\left(\mathcal{N}(A - \lambda I)\right) > 0$$
  Since
  $$\dim\left(\mathcal{R}(A - \lambda I)\right) + \dim\left(\mathcal{N}(A - \lambda I)\right) = n = \#\text{ columns of } A - \lambda I$$
  which means we require
  $$\dim\left(\mathcal{R}(A - \lambda I)\right) < n$$
  so $A - \lambda I$ cannot be full rank. We also know this requires
  $$\det\left(A - \lambda I\right) = 0$$
  - Hence, by finding all λ s.t. $\det\left(A - \lambda I\right) = 0$, we get the eigenvalues of A.
- the find the eigenvectors
  - Suppose λ is an eigenvalue of A, which has size $n \times n$.
  - Then every $\vec{x} \in \mathcal{N}(A - \lambda I)$ is considered an eigenvector of A, corresponding to the eigenvalue λ.
  - Because $\mathcal{N}(A - \lambda I)$ is a linear subspace of $\mathbb{C}^n$, we conventionally just specify enough vectors to span this subspace, i.e., a basis for $\mathcal{N}(A - \lambda I)$.

Let's work through an example.

Example 6.1

Let

$$A = \begin{bmatrix} 2 & 0 & 0 \\ 0 & \frac{3}{2} & \frac{1}{2} \\ 0 & \frac{1}{2} & \frac{3}{2} \end{bmatrix}$$

Solving $\det(A - \lambda I) = 0$, we want

$$\det(A - \lambda I) = (2 - \lambda)\left(\left(\frac{3}{2} - \lambda\right)^2 - \frac{1}{4}\right) = 0$$

which is satisfied if $\lambda = 2, 1$.

- For $\lambda = 1$ (with multiplicity one)
  - We want to find all $\vec{x} \in \mathbb{C}^3$ s.t.

$$\begin{bmatrix} 1 & 0 & 0 \\ 0 & \frac{1}{2} & \frac{1}{2} \\ 0 & \frac{1}{2} & \frac{1}{2} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}$$

  - This requires that $x_1 = 0$ and that $\frac{x_2}{2} + \frac{x_3}{2} = 0$. We have two linear equations, which imply that $\mathcal{N}(A - \lambda I)$ is a line in $\mathbb{C}^3$.
  - So we can pick $\vec{v}_1 = \begin{bmatrix} 0 \\ \frac{1}{\sqrt{2}} \\ -\frac{1}{\sqrt{2}} \end{bmatrix}$ as our first eigenvector.
- For $\lambda = 2$ (with multiplicity two)
  - We want to find $\vec{x} \in \mathbb{C}^3$ s.t.

$$\begin{bmatrix} 0 & 0 & 0 \\ 0 & -\frac{1}{2} & \frac{1}{2} \\ 0 & \frac{1}{2} & -\frac{1}{2} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}$$

  - This requires that $\frac{-x_2}{2} + \frac{x_3}{2} = 0 \implies x_2 = x_3$. We have one linear equation, which implies that $\mathcal{N}(A - \lambda I)$ is a plane in $\mathbb{C}^3$.
  - So we can pick two linearly indep. vectors from this nullspace, e.g.,

$$\vec{v}_2 = \begin{bmatrix} 0 \\ \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} \end{bmatrix} \text{ and } \vec{v}_3 = \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}$$

  - Thus, for this particular matrix A, we have two distinct eigenvalues but three linearly indep. eigenvectors.

In the case where the eigenvalues of the matrix are distinct, we have an important result:

Theorem 6.2

Eigenvectors corresponding to distinct eigenvalues are linearly independent.

Proof. Left for readers to figure out on your own :) (just kidding, I guess I am just lazy). □

Eigenvectors corresponding to repeated eigenvalues could be linearly indep.

Example 6.3

Consider

$$A = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}, \lambda_1 = \lambda_2 = 1, \vec{v}_1 = \begin{bmatrix} 1 \\ 0 \end{bmatrix}, \vec{v}_2 = \begin{bmatrix} 0 \\ 1 \end{bmatrix}$$

in which case $A - \lambda I$ has a two-dimensional nullspace.

Example 6.4

Consider:

$$A = \begin{bmatrix} 4 & 2 \\ 0 & 4 \end{bmatrix}, \lambda_1 = \lambda_2 = 4, \vec{v}_1 = \vec{v}_2 = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$$

in which case $A - \lambda I$ has a one-dimensional nullspace.

§7 Dis 1: Mar 30, 2021

§7.1 Linear Algebra Review

Definition 7.1 (Linear Operator) — Suppose X and Y are vector spaces. A linear operator $A : X \to Y$ is a map $X \to Y$ such that for any $x_1, x_2 \in X$, $c_1, c_2 \in \mathbb{R}$

$$A(c_1x_1 + c_2x_2) = c_1Ax_1 + c_2Ax_2$$

The set of linear operators $X \to Y$ forms a vector space with the usual addition and scalar multiplication for functions: $(A + B)(x) = Ax + Bx$ and $(cA)(x) = c(Ax)$, denoted $L(X \to Y)$.

For finite-dimensional X and Y, linear operators $X \to Y$ can represented by matrices. In particular, we identify $L(\mathbb{R}^n \to \mathbb{R}^m)$ with the space $\mathbb{R}^{m\times n}$ of $m \times n$ matrices, associating a matrix A with the map $x \mapsto Ax$.
An operator $A : X \to Y$ is invertible if there is $B : Y \to X$ such that $AB = Id_Y$, i.e., $(AB)y = y$ for all $y \in Y$ and $BA = Id_X$, i.e. $(BA)x = x$ for all $x \in X$. Such a B is unique and is called the inverse of A, $B = A^{-1}$. If A is linear, so is A^{-1}. If $A : X \to Y$ is invertible, then $\dim X = \dim Y$. In particular, all invertible matrices are square, we always have $\left(A^{-1}\right)^{-1} = A$ and $(AB)^{-1} = B^{-1}A^{-1}$.
The followings are equivalent for $n \times n$ matrices A:

- A is invertible.
- A is nonsingular.
- $\text{rank}(A) = n$.
- $\det(A) \neq 0$
- The linear map A is bijective.
- $\dim(\ker(A)) = 0 \iff \ker A = \{0\}$
- $\dim(\text{im}(A)) = n$.
- Columns of A are linearly indep.
- Rows of A are linearly indep.
- All eigenvalues of A are non-zero.
- $A^\top$ is invertible ($\left(A^\top\right)^{-1} = \left(A^{-1}\right)^\top$)
- $A^\top A$ is invertible $\iff$ $A^\top A$ is positive definite, which means

  If $B = B^\top$, then B is positive definite if $\langle x, Bx\rangle > 0$ for all $x \neq 0$. Equivalently, all of $B's$ eigenvalues are positive.

Definition 7.2 (Adjoint) — Assume that X, Y are vector spaces with inner product $\langle , \rangle_X$ and $\langle , \rangle_Y$ and that $A : X \to Y$ is a linear operator. Then the adjoint of A is the unique linear operator $A^* : Y \to X$ such that

$$\langle Ax, y\rangle_Y = \langle x, A^* y\rangle_X$$

- An operator $A : X \to X$ is self-adjoint if $A^* = A$.
- We always have $(A^*)^* = A$, and if A is invertible then A^* is invertible with $(A^*)^{-1} = (A^{-1})^*$.
- If A is represented by a matrix, which we will also call A, then A^* is represented by the matrix, $A^\top (A^\dagger = \overline{A^\top}$ for complex matrices).
- $(A + B)^* = (A^* + B^*)$, $(cA)^* = \overline{c}A^*$, $(AB)^* = B^* A^*$

Approximate Solutions: Suppose $A : X \to Y$ is a linear operator between inner product spaces. Given $y \in Y$, we wish to minimize $\|y - Ax\|$, where $\|z\|^2 = \langle z, z\rangle$.
Need x to be s.t. for any $z \in X$, $\|y - A(x + z)\|^2 \geq \|y - Ax\|^2$. Have

$$\begin{aligned}\langle y - A(x + z), y - A(x + z)\rangle &= \langle (y - Ax) - Az, (y - Ax) - Az\rangle \\ &= \|y - Ax\|^2 + \|Az\|^2 - 2\langle y - Ax, Az\rangle\end{aligned}$$

If $\langle y - Ax, Az\rangle > 0$, then $\|y - A(x + \epsilon z)\|^2 < \|y - Ax\|^2$ for small $\epsilon > 0$. Similarly for $\langle y - Ax, Az\rangle < 0$. Thus, $\langle y - Ax, Az\rangle = 0$ for all $z \in X$, i.e., $\langle A^*(y - Ax), z >\rangle = 0$ for all $z \in X$. Hence $A^*(y - Ax) = 0$, i.e. $A^* Ax = A^* y$.