# EECS 127/227AT Discussion 1 Slides

Druv Pai

September 2, 2020

# About Me

- 3rd year undergrad, CS/Stats major
- Interested in:
  - Statistical learning theory, specifically:
  - Robust estimation
  - High dimensional statistics
  - *All of these require optimization!*
- Other: Playing basketball, running, reading
- Email: druvpaiberkeley.edu
- **Office hours: 5-6 PM PST on Monday – please stop by!**

# Q1

## Definition (Vector)

Element of vector space (has scalar multiplication and vector addition defined in ways you would expect)

## Example

Element of $\mathbb{R}^n$ ($n$-tuple $(x_1, \ldots, x_n)$); matrices; more exotic objects

## Definition (Inner Product)

Inner product of $x$ and $y$, $\langle x, y \rangle$, is any function which has properties:

1. (Conjugate) symmetry: $\langle x, y \rangle = \overline{\langle y, x \rangle}$ (in $\mathbb{R}^n$ $\langle x, y \rangle = \langle y, x \rangle$)
2. (Bi)linearity: $\langle \alpha x, y \rangle = \alpha \langle x, y \rangle$; $\langle x + y, z \rangle = \langle x, z \rangle + \langle y, z \rangle$
3. Positiveness: $\langle x, x \rangle > 0$ for $x > 0$

## Example

If $x = (x_1, \ldots, x_n)$, $y = (y_1, \ldots, y_n)$ then $\langle x, y \rangle = x^\mathsf{T} y = \sum_{i=1}^n x_i y_n$

### Definition (Norm)

Norm of vector $x$, $\|x\|$, is any function which has properties:

- Homogeneity: $\|\alpha x\| = |\alpha|\|x\|$ for $\alpha \in \mathbb{R}$
- Positiveness: $\|x\| > 0$ for $x \neq 0$
- Triangle Inequality: $\|x + y\| \leq \|x\| + \|y\|$

### Example

$\|(x_1, \ldots, x_m)\|_p = \left(\sum_{i=1}^{n} |x_i|^p\right)^{1/p}$; matrix norms (covered later).

Norm *induced* from inner product: $\|x\| = \sqrt{\langle x, x\rangle}$. (Though we can consider non-induced norms as well, for $p \neq 2$ then $\|x\|_p$ isn't induced by an inner product).

### Theorem (Cauchy-Schwarz)

$|\langle x, y\rangle| \leq \|x\|_2\|y\|_2$.

### Proof.

Problem 1! □

# Q2

- Partial derivatives – "like regular derivatives, but hold everything except the variable you want constant"
- Matrix calculus: Homework 0 Problem 6 for basics
- Small trick to find the gradient: Suppose $f\colon \mathbb{R}^n \to \mathbb{R}$, then $f(x+\varepsilon) = f(x) + \varepsilon^\top(\nabla_x f(x)) + \frac{1}{2}\varepsilon^\top\left(\nabla_x^2 f(x)\right)\varepsilon + \cdots$
- Get gradient and Hessian by pattern matching! Very fast.
- Can be applied to $f\colon \mathbb{R}^n \to \mathbb{R}^m$ Jacobian if careful

# Q3

## Definition

Angle $\theta$ between two vectors: $\langle x, y \rangle = \|x\| \|y\| \cos(\theta)$

## Definition (Orthogonality)

$x, y$ orthogonal if $\langle x, y \rangle = 0 \to \theta = \frac{\pi}{2}$ (radians)

## Definition (Linear Independence)

Set $S$ of vectors is *linearly independent* if $\sum_{x \in S} \alpha_x x = 0 \to \alpha_x = 0$ for all $x$; if $S$ is finite of size $n$ then $\sum_{i=1}^{n} \alpha_i x_i = 0 \to \alpha_i = 0$

## Definition (Span, Linear Combination)

If $S$ is set of vectors, then span($S$) $= \{\sum_{x \in S} \alpha_x x \mid \forall x, \alpha_x \in \mathbb{R}\}$ = set of all *linear combinations* of vectors in $S$

## Definition (Basis, Dimension)

Set of vectors $S$ is basis for $X$ if span($S$) = $X$ and $S$ is linearly independent; dimension: dim($X$) = number of vectors in $S$

### Definition (Orthogonal Basis)

$S$ is orthogonal basis for $X$ if $S$ is basis for $X$ and $\langle x, y \rangle = 0$ for $x \neq y$, $x, y \in S$ (pairwise orthogonality)

### Definition (Normalized Vector)

Vector where $\|x\| = 1$.

### Definition (Orthonormal Basis)

Orthogonal basis where every vector in the basis is normalized.

### Definition (Projection)

Suppose $X$ is vector space and $V \subseteq X$ is a subspace (subset that is itself a vector space) has orthonormal basis $\{v_1, \ldots, v_n\}$. Then $\text{proj}_V(x) = \sum_{i=1}^{n} \langle v_i, x \rangle v_i$. "Closest point in $V$ to $X$."

### Definition (Gram-Schmidt Process)

Suppose we have basis $U = \{u_1, \ldots, u_n\}$. We want to find an orthonormal basis $V = \{v_1, \ldots, v_n\}$.

1. Set $v_1 \leftarrow \frac{u_1}{||u_1||}$.
2. For $i \in \{2, \ldots, n\}$:
   2.1 Set $s_i \leftarrow u_i - \text{proj}_{\text{span}(v_1, \ldots, v_{i-1})}(u_i)$ – "subtract non-orthogonal part from $u_i$"
   2.2 Set $v_i \leftarrow \frac{s_i}{||s_i||}$ – "normalize"

NB: projection defined on previous slide.