EECS 275 Matrix Computation

Ming-Hsuan Yang

Electrical Engineering and Computer Science
University of California at Merced
Merced, CA 95344
http://faculty.ucmerced.edu/mhyang

Lecture 2
Overview

- Basic definition: vector space, norm, subspace, linear independence, convexity, normed linear spaces, range, rank, null space, matrix inverse
- Elementary analytical and topological properties
Reading

- Chapter 1-3 of *Numerical Linear Algebra* by Trefethen and Bau
- Chapter 5 of *Matrix Analysis and Applied Linear Algebra* by Carl Meyer
- Chapter 2 of *Matrix Computations* by Gene Golub and Charles Van Loan
- Chapter 3 and Chapter 4 of *Matrix Algebra From a Statistician’s Perspective* by David Harville
- Chapter 2 of *Optimization by Vector Space Methods* by David Luenberger
Vector space

- A vector space $X$ is a set of elements called vectors together with two operations.
  - Vector addition: let $x, y \in X$, then $x + y \in X$
  - Scalar multiplication: let $x \in X$ and $\alpha$ be any scalar, then $\alpha x \in X$

- Axioms:
  - $x + y = y + x$ (commutative law)
  - $(x + y) + z = x + (y + z)$ (associative law)
  - $(\alpha \beta)x = \alpha(\beta x)$ (associative law)
  - There is a null (zero) vector $0 \in X$ such that $x + 0 = x$, $\forall x \in X$.
  - $\alpha(x + y) = \alpha x + \alpha y$ (distributive law)
  - $(\alpha + \beta)x = \alpha x + \beta x$ (distributive law)
  - $0x = 0$, $1x = x$
Let $X$ and $Y$ be vector spaces over the same field of scalars. Then the Cartesian product of $X$, and $Y$, denoted $X \times Y$, consists of the collection of ordered pairs $(x, y)$ with $x \in X$, $y \in Y$. Addition and scalar multiplication are defined on $X \times Y$ by 

$$(x_1, y_1) + (x_2, y_2) = (x_1 + x_2, y_1 + y_2) \text{ and } \alpha(x, y) = (\alpha x, \alpha y).$$

Generalized to product of $n$ vector spaces, $X_1, X_2, \ldots, X_n$. Denote it as $X^n$ for the product of a vector space of with itself $n$ times.
Subspace

- A nonempty subset $M$ of a vector space $X$ is a subspace of $X$ if every vector of the form $\alpha x + \beta y$ is in $M$ whenever $x$ and $y$ are both in $M$.
- $M$ is a subspace if and only if
  - The null vector $0 \in M$.
  - If $x, y \in M$, then $x + y \in M$.
  - If $\alpha$ is a scalar and $x \in M$, then $\alpha x \in M$.
- The simplest subspace is the set consisting of $0$ alone.
- In three-dimensional space, a plane passing through the origin is a subspace.
- Let $M$ and $N$ be subspaces of a vector space $X$. Then the intersection $M \cap N$, of $M$ and $N$ is a subspace of $X$.
- The sum of two subsets $S$ and $T$ in a vector space, denoted $S + T$, consists of all vectors of the forms $s + t$ where $s \in S$ and $t \in T$.
- Let $M$ and $N$ be subspaces of a vector space $X$, then their sum $M + N$ is a subspace of $X$. 
Linear subspace

- A linear combination of the vectors $\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_n$ in a vector space is a sum of the form $\alpha_1 \mathbf{x}_1 + \alpha_2 \mathbf{x}_2 + \cdots + \alpha_n \mathbf{x}_n$.
- Suppose $S$ is a subset of a vector space $X$. The set $[S]$, called the subspace generated by $S$, consists of all vectors in $X$ which are linear combinations of vectors in $S$.
- The translation of a subspace is a linear variety (affine subspace).
- An affine subspace of $\mathbb{R}^3$ is a point $P(x, y)$ or a line whose points are solution of a linear system
  \[
  a_1 x + a_2 y + a_3 z = a_4 \\
  b_1 x + b_2 y + b_3 z = b_4
  \]
  or a plane, formed by the solutions of a linear equation
  \[
  ax + by + cz = d
  \]
  These are not necessarily subspaces of vector space $\mathbb{R}^3$, unless $x$ is the origin or the equations are homogeneous.
- Affine subspace is obtained from a vector space by translation, and in this sense a generalization of linear.
Convexity and cones

A set $K$ in a linear vector space is convex if, given $x_1, x_2 \in K$, all points of the form $\alpha x_1 + (1 - \alpha)x_2$ with $0 \leq \alpha \leq 1$ are in $K$.

Let $K$ and $G$ be convex sets in a vector space. Then

- $\alpha K = \{x : x = \alpha k, k \in K\}$
- $K + G$ is convex

Let $S$ be an arbitrary set in a linear vector space. The convex hull, denoted $co(S)$ is the smallest convex set containing $S$. In other words, $co(S)$ is the intersection of all convex sets containing $S$.

A set $C$ in a linear vector space is a cone with vertex at the origin if $x \in C$ implies $\alpha x \in C \ \forall \alpha \geq 0$.

In $\mathbb{R}^n$, the set

$$P = \{x : x = \{\xi_1, \xi_2, \ldots, \xi_n\}, \xi_i \geq 0 \ \forall i\}$$

defining the positive orthant, is a convex cone.
Linear independence and dimension

- A set of \(\{x_1, x_2, \ldots, x_n\}\) is linearly independent if \(\sum_{j=1}^{n} \alpha_j x_j = 0\) implies \(\alpha_j = 0\) for all \(j = 1, \ldots, n\).

- Span: \(\text{span}\ \{x_1, \ldots, x_n\} = \{\sum_{j=1}^{n} \beta_j x_i : \beta_j \in \mathbb{R}\}\).

- If \(\{x_1, \ldots, x_n\}\) is independent and \(y \in \text{span} \ \{x_1, \ldots, x_n\}\), then \(y\) is a unique linear combination of \(x_i\).

- A finite set \(S\) of linearly independent vectors is a basis for the space \(X\) if \(S\) generates \(X\). A vector space having a finite basis is said to be finite dimensional.

- Usually we characterize a finite-dimensional space by the number of elements in a basis. Thus, a space with a basis consisting of \(n\) elements is referred to as \(n\)-dimensional space.

- Any two bases for a finite-dimensional vector space contain the same number of elements.
Normed linear spaces

- A normed linear vector space is a vector space $X$ on which there is defined a real-valued function which maps each element $x$ in $X$ into a real number $\|x\|$ called the norm of $x$. The norm satisfies the following axioms:
  - $\|x\| \geq 0$ for all $x \in X$, $\|x\| = 0$ iff $x = 0$.
  - $\|x + y\| \leq \|x\| + \|y\|$ for each $x, y \in X$.
  - $\|\alpha x\| = |\alpha| \cdot \|x\|$ for all scalar $\alpha$ and each $x \in X$.

- An abstraction of our usual concept of length.

- In a normed linear space,

\[ \|x\| - \|y\| \leq \|x - y\| \]

for any two vectors $x, y$.

- A normed linear space is a vector space having a measure of distance defined on it.

- Recall a metric on a set $X$ is a function, $d : X \times X \rightarrow \mathbb{R}$ that satisfies non-negativity, identity, symmetry and triangle inequality conditions.
Normed linear spaces: examples

- Normed linear space $C[a, b]$ consists of continuous functions on the real interval $[a, b]$ together with the norm $\|x\| = \max_{a \leq t \leq b} |x(t)|$.
  $$\max |x(t) + y(t)| \leq \max(|x(t)| + |y(t)|) \leq \max |x(t)| + \max |y(t)|$$
  $$\max |\alpha x(t)| = \max |\alpha| |x(t)| = |\alpha| \max |x(t)|.$$  

- Euclidean $n$-space, denoted $E^n$, consists of $n$-tuples with the norm of an element $x = \{\xi_1, \xi_2, \ldots, \xi_n\}$ defined as $\|x\| = (\sum_{i=1}^{n} |\xi_i|^2)^{1/2}$.

- Consider the space $BV[a, b]$ consisting of functions of bounded variation on the interval $[a, b]$. By a partition of the interval $[a, b]$, we mean a finite set of points $t_i \in [a, b]$, $i = 0, 1, \ldots, n$, such that $a = t_0 \leq t_1 \cdots \leq t_n = b$.

- A function $x$ defined on $[a, b]$ is of bounded variation if there is a constant $K$ so that for any partition of $[a, b]$
  $$\sum_{i=1}^{n} |x(t_i) - x(t_{i-1})| \leq K.$$
Normed linear spaces: examples

- The total variation of $x$, denoted $TV(x)$ is defined as

$$TV(x) = \sup \sum_{i=1}^{n} |x(t_i) - x(t_{i-1})|$$

where the supremum is taken w.r.t. all partitions of $[a, b]$.

- A convenient notation for the total variation

$$TV(x) = \int_{a}^{b} |dx(t)|.$$

- The $TV$ of a monotonic function is the absolute value of the difference between function values at the end points $a$ and $b$.

- The space $BV[a, b]$ is the space of all functions of bounded variation on $[a, b]$ together with the norm defined as

$$\|x\| = |x(a)| + TV(x).$$
Open and close sets

- An element \( x \in C \subseteq \mathbb{R}^n \) is called an interior point of \( C \) if there exists an \( \varepsilon > 0 \) such that \( \{ y \mid \| y - x \|_2 < \varepsilon \} \subseteq C \), i.e., there exists a ball centered at \( x \) that lies entirely in \( C \).
- The set of all points interior to \( C \) is called interior of \( C \), denoted by \( \text{int}(C) \).
- A set \( C \) is open if \( \text{int}(C) = C \), i.e., every point in \( C \) is an interior point. For example, \((0, 2)\) is open, and \((0, 2]\) is not open.
- A set is closed if its complement \( \mathbb{R}^n \setminus C = \{ x \in \mathbb{R}^n \mid x \notin C \} \) is open. For example, \([0, 2]\) is closed.
- The closure of a set \( C \) is \( \text{cl}(C) = \mathbb{R}^n \setminus \text{int}(\mathbb{R}^n \setminus C) \), i.e., the complement of the interior of the complement of \( C \).
- A point is in the closure of \( C \) if for all \( \varepsilon > 0 \), there is a \( y \in C \) with \( \| x - y \| < \varepsilon \).

Blue points: \( x^2 + y^2 = r^2 \), and red points: \( x^2 + y^2 < r^2 \). The blue points form a closure set. The red points form an open set. The union of the red and blue points is a boundary set.
Closed sets and convergent sequences

- Can describe closed sets in terms of convergent sequences and limit points.
- A set $C$ is closed if and only if it contains the limit point of every convergent sequences in it.
- In other words, if $x_1, x_2, \ldots$ converges to $x$ and $x_i \in C$, then $x \in C$. The closure of $C$ is the set of all limit points of convergent sequences in $C$.
- The boundary of $C$ is defined as $\text{bd}(C) = \text{cl}(C) \setminus \text{int}(C)$.
- A boundary point $x$ satisfies the following property: For all $\varepsilon > 0$, there exist $y \in C$ and $z \notin C$ with
  \[ \|y - x\|_2 \leq \varepsilon, \quad \|z - x\|_2 \leq \varepsilon, \]
  i.e., there exist arbitrarily close points in $C$, and also arbitrarily close points not in $C$.
- $C$ is closed if it contains its boundary, $\text{bd}(C) \subseteq C$. It is open if it contains no boundary points, $C \cap \text{bd}(C) = \emptyset$. 
Vector norm

- Vector norm: a function that assigns a strictly positive length or size to all vectors $\mathbf{x}$ in a vector space $X$, other than the zero vector, $\mathbf{0}$, i.e., $f : X \to \mathbb{R}; \mathbf{x} \mapsto f(\mathbf{x})$ that satisfies the following properties:

  $$f(\mathbf{x}) \geq 0 \quad \mathbf{x} \in \mathbb{R}^n, \ (f(\mathbf{x}) = 0 \text{ iff } \mathbf{x} = \mathbf{0})$$
  $$f(\mathbf{x} + \mathbf{y}) \leq f(\mathbf{x}) + f(\mathbf{y}) \quad \mathbf{x}, \mathbf{y} \in \mathbb{R}^n$$
  $$f(\alpha \mathbf{x}) = |\alpha|f(\mathbf{x}) \quad \alpha \in \mathbb{R}, \mathbf{x} \in \mathbb{R}^n$$

- A simple example is the 2-dimensional space $\mathbb{R}^2$ with Euclidean norm, e.g., a point (2,5) is drawn as an arrow from the origin. As such, Euclidean norm is often known as magnitude.

- Euclidean norm: $\|\mathbf{x}\| = \sqrt{x_1^2 + \cdots + x_n^2}$ for $\mathbf{x} \in \mathbb{R}^n$.

- Manhattan norm: $\|\mathbf{x}\|_1 = \sum_{i=1}^n |x_i|$. 
Vector norm (cont’d)

- $\ell_p$-norm: $\|x\|_p = (\sum_{i=1}^{n} |x_i|^p)^{\frac{1}{p}}$ for $p \geq 1$. Note that for $p = 1$, we get the $\ell_1$ norm or Manhattan norm, and for $p = 2$, we get the Euclidean norm.

- $\ell_\infty$-norm (maximum norm): $\|x\|_\infty = \max(|x_1|, \ldots, |x_n|)$.

- When $0 < p < 1$, $\ell_p$-norm does not define a norm as it violates the triangle inequality. However, the function $d_p(x, y) = \sum_{i=1}^{n} |x_i - y_i|^p$ defines a metric.

- When $p = 0$, the zero norm of $x$ is defined as $\lim_{p \to 0} \|x\|_p^p$. Define $0^0 = 0$, then we can write the zero norm as $\sum_{i=1}^{n} x_i^0$, which is simply the number of non-zero elements.

- If $x \in \mathbb{R}^n$, then $\|x\|_p \leq \|x\|_q$ if $p > q$, and $p > 0, q > 0$, e.g., $\|x\|_\infty \leq \|x\|_2 \leq \|x\|_1$. 
Vector norm (cont’d)

- Norms play an important role in solving optimization problems.

\[ p \geq 1 \quad p < 1 \quad \text{Lasso (left) and ridge (right) regression} \]

- The residual sum of squares has elliptical contours, centered at the full least squares estimate.

- The constraint region for Lasso is \( \|x_1\| + \|x_2\| \leq t \) while the constraint region for ridge regression is \( x_1^2 + x_2^2 \leq t^2 \).

- The diamond has corners and if the solution occurs at a corner, then it has one parameter \( x_i \) equal to zero.
Vector norm (cont’d)

- For \( x \in \mathbb{R}^n \), then
  \[
  \|x\|_2 \leq \|x\|_1 \leq \sqrt{n}\|x\|_2 \\
  \|x\|_{\infty} \leq \|x\|_2 \leq \sqrt{n}\|x\|_{\infty} \\
  \|x\|_{\infty} \leq \|x\|_1 \leq n\|x\|_{\infty}
  \]

- Holder inequality: For the \( \ell_p \)-norm, we have
  \[
  |x^\top y| \leq \|x\|_p \|y\|_q, \quad \frac{1}{p} + \frac{1}{q} = 1
  \]

- A special case is the Cauchy-Schwartz inequality,
  \[
  |x^\top y| \leq \|x\|_2 \|y\|_2
  \]

- The triangle inequality for inner product is often shown using Cauchy-Schwartz inequality
  \[
  \|x + y\|^2 \leq (\|x\| + \|y\|)^2
  \]