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 $\mathbf{x}, \mathbf{y} \in X$, then $\mathbf{x} + \mathbf{y} \in X$
  - Scalar multiplication: let $\mathbf{x} \in X$ and α be any scalar, then $\alpha\mathbf{x} \in X$
- Axioms:
  - $\mathbf{x} + \mathbf{y} = \mathbf{y} + \mathbf{x}$ (commutative law)
  - $(\mathbf{x} + \mathbf{y}) + \mathbf{z} = \mathbf{x} + (\mathbf{y} + \mathbf{z})$ (associative law)
  - $(\alpha\beta)\mathbf{x} = \alpha(\beta\mathbf{x})$ (associative law)
  - There is a null (zero) vector $\mathbf{0} \in \mathbf{X}$ such that $\mathbf{x} + \mathbf{0} = \mathbf{x}$, $\forall \mathbf{x} \in X$.
  - $\alpha(\mathbf{x} + \mathbf{y}) = \alpha\mathbf{x} + \alpha\mathbf{y}$ (distributive law)
  - $(\alpha + \beta)\mathbf{x} = \alpha\mathbf{x} + \beta\mathbf{x}$ (distributive law)
  - $0\mathbf{x} = \mathbf{0}$, $1\mathbf{x} = \mathbf{x}$

Cartesian product

- 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 $(\mathbf{x}, \mathbf{y})$ with $\mathbf{x} \in X$, $\mathbf{y} \in Y$. Addition and scalar multiplication are defined on $X \times Y$ by $(\mathbf{x}_1, \mathbf{y}_1) + (\mathbf{x}_2, \mathbf{y}_2) = (\mathbf{x}_1 + \mathbf{x}_2, \mathbf{y}_1 + \mathbf{y}_2)$ and $\alpha(\mathbf{x}, \mathbf{y}) = (\alpha\mathbf{x}, \alpha\mathbf{y})$.
- 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 $\alpha\mathbf{x} + \beta\mathbf{y}$ is in M whenever $\mathbf{x}$ and $\mathbf{y}$ are both in M.
- M is a subspace if and only if
  - The null vector $\mathbf{0} \in M$.
  - If $\mathbf{x}, \mathbf{y} \in M$, then $\mathbf{x} + \mathbf{y} \in M$.
  - If α is a scalar and $\mathbf{x} \in M$, then $\alpha\mathbf{x} \in M$.
- The simplest subspace is the set consisting of $\mathbf{0}$ 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 $\mathbf{s} + \mathbf{t}$ where $\mathbf{s} \in S$ and $\mathbf{t} \in T$.
- 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

$$\begin{aligned} a_1x + a_2y + a_3z &= a_4 \\ b_1x + b_2y + b_3z &= b_4 \end{aligned}$$

  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 $\mathbf{x}$ 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 $\mathbf{x}_1$, $\mathbf{x}_2 \in K$, all points of the form $\alpha\mathbf{x}_1 + (1-\alpha)\mathbf{x}_2$ with $0 \le \alpha \le 1$ are in K.
- Let K and G be convex sets in a vector space. Then
  - $\alpha K = \{\mathbf{x} : \mathbf{x} = \alpha\mathbf{k}, \mathbf{k} \in K\}$
  - $K + G$ 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 $\mathbf{x} \in C$ implies $\alpha\mathbf{x} \in C \ \forall\alpha \ge 0$.
- In $\mathbb{R}^n$, the set

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

defining the positive orthant, is a convex cone.

Linear independence and dimension

- A set of $\{\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_n\}$ is linearly independent if $\sum_{j=1}^{n} \alpha_j \mathbf{x}_j = \mathbf{0}$ implies $\alpha_j = 0$ for all $j = 1, \ldots, n$.
- Span: span $\{\mathbf{x}_1, \ldots, \mathbf{x}_n\} = \{\sum_{j=1}^{n} \beta_i \mathbf{x}_i : \beta_j \in \mathbb{R}\}$.
- If $\{\mathbf{x}_1, \ldots, \mathbf{x}_n\}$ is independent and $\mathbf{y} \in$ span $\{\mathbf{x}_1, \ldots, \mathbf{x}_n\}$, then $\mathbf{y}$ is a unique linear combination of $\mathbf{x}_i$.
- 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 $\mathbf{x}$ in X into a real number $\|\mathbf{x}\|$ called the norm of $\mathbf{x}$. The norm satisfies the following axioms:
  - $\|\mathbf{x}\| \geq 0$ for all $\mathbf{x} \in X$, $\|\mathbf{x}\| = 0$ iff $\mathbf{x} = \mathbf{0}$.
  - $\|\mathbf{x} + \mathbf{y}\| \leq \|\mathbf{x}\| + \|\mathbf{y}\|$ for each $\mathbf{x}$, $\mathbf{y} \in X$.
  - $\|\alpha \mathbf{x}\| = |\alpha| \cdot \|\mathbf{x}\|$ for all scalar α and each $\mathbf{x} \in X$.
- An abstraction of our usual concept of length.
- In a normed linear space,

$$\|\mathbf{x}\| - \|\mathbf{y}\| \leq \|\mathbf{x} - \mathbf{y}\|$$

  for any two vectors $\mathbf{x}$, $\mathbf{y}$.
- 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, $d : X \times X \to \mathbb{R}$ 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 $\|\mathbf{x}\| = \max_{a \le t \le b} |\mathbf{x}(t)|$.

$$\max |\mathbf{x}(t) + \mathbf{y}(t)| \le \max(|\mathbf{x}(t)| + |\mathbf{y}(t)|) \le \max |\mathbf{x}(t)| + \max |\mathbf{y}(t)|$$

$$\max |\alpha \mathbf{x}(t)| = \max |\alpha||\mathbf{x}(t)| = |\alpha| \max |\mathbf{x}(t)|.$$

- Euclidean n-space, denoted E^n, consists of n-tuples with the norm of an element $\mathbf{x} = \{\xi_1, \xi_2, \dots, \xi_n\}$ defined as $\|\mathbf{x}\| = (\sum_{i=1}^n |\xi_n|^2)^{1/2}$.
- 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 $t_i \in [a, b]$, $i = 0, 1, \dots, n$, such that $a = t_0 \le t_1 \cdots \le t_n = b$.
- A function $\mathbf{x}$ defined on $[a, b]$ is of bounded variation if there is a constant K so that for any partition of $[a, b]$

$$\sum_{i=1}^{n} |\mathbf{x}(t_i) - \mathbf{x}(t_{i-1})| \le K.$$

Normed linear spaces: examples

- The total variation of $\mathbf{x}$, denoted $TV(\mathbf{x})$ is defined as

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

  where the supremum is taken w.r.t. all partitions of $[a, b]$.
- A convenient notation for the total variation

$$TV(\mathbf{x}) = \int_a^b |d\mathbf{x}(t)|.$$

- 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

$$\|\mathbf{x}\| = |\mathbf{x}(a)| + TV(\mathbf{x}).$$

Open and close sets

- An element $\mathbf{x} \in C \subseteq \mathbb{R}^n$ is called an interior point of C if there exists an $\varepsilon > 0$ such that $\{\mathbf{y} | \ \|\mathbf{y} - \mathbf{x}\|_2 < \varepsilon\} \subseteq C$, i.e., there exists a ball centered at $\mathbf{x}$ 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 $\mathbb{R}^n \backslash C = \{\mathbf{x} \in \mathbb{R}^n | \mathbf{x} \notin C\}$ is open. For example, $[0, 2]$ is closed.
- The closure of a set C is $\text{cl}(C) = \mathbb{R}^n \backslash \text{int}(\mathbb{R}^n \backslash C)$, 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 $\mathbf{y} \in C$ with $\|\mathbf{x} - \mathbf{y}\| < \varepsilon$.

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 $\mathbf{x}_1, \mathbf{x}_2, \ldots$ converges to $\mathbf{x}$ and $\mathbf{x}_i \in C$, then $\mathbf{x} \in C$. The closure of C is the set of all limit points of convergent sequences in C.
- The boundary of C is defined as $\mathrm{bd}(C) = \mathrm{cl}(C) \backslash \mathrm{int}(C)$.
- A boundary point $\mathbf{x}$ satisfies the following property: For all $\varepsilon > 0$, there exist $\mathbf{y} \in C$ and $\mathbf{z} \notin C$ with

  $$\|\mathbf{y} - \mathbf{x}\|_2 \leq \varepsilon, \quad \|\mathbf{z} - \mathbf{x}\|_2 \leq \varepsilon,$$

  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, $\mathrm{bd}(C) \subseteq C$. It is open if it contains no boundary points, $C \cap \mathrm{bd}(C) = \emptyset$.

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:

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

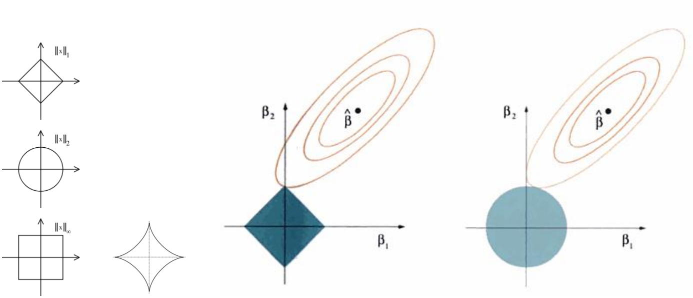
- A simple example is the 2-dimensional space $\mathbb{R}^2$ with Euclidean norm, e.g., a point (2,5) is drawn as an arrow from the origin. As such, Euclidean norm is often known as magnitude.
- Euclidean norm: $\|\mathbf{x}\| = \sqrt{x_1^2 + \cdots + x_n^2}$ for $\mathbf{x} \in \mathbb{R}^n$.
- Manhattan norm: $\|\mathbf{x}\|_1 = \sum_{i=1}^{n} |x_i|$.

Vector norm (cont'd)

- ℓ_p-norm: $\|\mathbf{x}\|_p = (\sum_{i=1}^n |x_i|^p)^{\frac{1}{p}}$ for $p \geq 1$. Note that for $p = 1$, we get the ℓ_1 norm or Manhattan norm, and for $p = 2$, we get the Euclidean norm.
- ℓ_∞-norm (maximum norm): $\|\mathbf{x}\|_\infty = \max(|x_1|, \dots, |x_n|)$.
- When $0 < p < 1$, ℓ_p-norm does not define a norm as it violates the triangle inequality. However, the function $d_p(\mathbf{x}, \mathbf{y}) = \sum_{i=1}^n |x_i - y_i|^p$ defines a metric.
- When $p = 0$, the zero norm of $\mathbf{x}$ is defined as $\lim_{p \to 0} \|\mathbf{x}\|_p^p$. 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 $\mathbf{x} \in \mathbb{R}^n$, then $\|\mathbf{x}\|_p \leq \|\mathbf{x}\|_q$ if $p > q$, and $p > 0, q > 0$, e.g., $\|\mathbf{x}\|_\infty \leq \|\mathbf{x}\|_2 \leq \|\mathbf{x}\|_1$.

Vector norm (cont'd)

- Norms play an important role in solving optimization problems.



$p \geq 1$ $p < 1$ 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 $\mathbf{x} \in \mathbb{R}^n$, then

$$\begin{array}{ccccc} \|\mathbf{x}\|_2 & \leq & \|\mathbf{x}\|_1 & \leq & \sqrt{n}\|\mathbf{x}\|_2 \\ \|\mathbf{x}\|_\infty & \leq & \|\mathbf{x}\|_2 & \leq & \sqrt{n}\|\mathbf{x}\|_\infty \\ \|\mathbf{x}\|_\infty & \leq & \|\mathbf{x}\|_1 & \leq & n\|\mathbf{x}\|_\infty \end{array}$$

- Holder inequality: For the ℓ_p-norm, we have

$$|\mathbf{x}^\top \mathbf{y}| \leq \|\mathbf{x}\|_p \|\mathbf{y}\|_q, \quad \frac{1}{p} + \frac{1}{q} = 1$$

- A special case is the Cauchy-Schwartz inequality,

$$|\mathbf{x}^\top \mathbf{y}| \leq \|\mathbf{x}\|_2 \|\mathbf{y}\|_2$$

- The triangle inequality for inner product is often shown using Cauchy-Schwartz inequality

$$\|\mathbf{x} + \mathbf{y}\|^2 \leq (\|\mathbf{x}\| + \|\mathbf{y}\|)^2$$