Mathematical preliminaries

Pramesh Kumar

IIT Delhi

August 7, 2025

Definitions

Definition (Vector). An array of scalars.

Definition (Vector Space). A vector space is a set V equipped with two operations - *addition* and *multiplication*:

1. (Addition) For $\mathbf{u}, \mathbf{v} \in V$, we have $\mathbf{u} + \mathbf{v} \in V$
2. (Scalar multiplication) For any scalar $c \in \mathbb{R}$ and $\mathbf{u} \in V$, we have $c\mathbf{u} \in V$

Example(s). $\mathbb{R}^n$, $\mathbb{M} = \mathbb{R}^{m \times n}$, $\mathbf{0}$, etc.

Definition (Subspace). A non-empty subset $S \subset V$ of a vector space is a subspace iff for every $\mathbf{x}, \mathbf{y} \in S$ and $c, d \in \mathbb{R}$, we have $c\mathbf{x} + d\mathbf{y} \in S$.

1. Geometric interpretation: If $\mathbf{x}, \mathbf{y} \in S$, then plane passing through $0, \mathbf{x},$ and $\mathbf{y}$ lies in S.
2. Intersection of finite number of subspaces is a subspace.
3. If S is a linear subspace, then there exists $A \in \mathbb{R}^{m \times n}$ such that $S = \{\mathbf{x} \in \mathbb{R}^n \mid A\mathbf{x} = 0\}$

Fundamental subspaces

Definition (Column space or range or image). Column space of matrix $A \in \mathbb{R}^{m \times n}$ denoted by $\mathcal{C}(A)$ or $\mathcal{R}(A)$ or $\text{img}(A)$ is defined as $\mathcal{C}(A) = \{Ax \mid x \in \mathbb{R}^n\}$, i.e., collection of all linear combinations of columns of A.

Definition (Null space or kernel). Null space of a matrix $A \in \mathbb{R}^{m \times n}$ denoted by $\mathcal{N}(A)$ or $\ker(A)$ is defined as $\mathcal{N}(A) = \{\mathbf{x} \in \mathbb{R}^n \mid A\mathbf{x} = 0\}$.

Example $\mathcal{C}\left(\begin{bmatrix}1 & 0\\ 0 & 1\end{bmatrix}\right)$ is $\mathbb{R}^2$ and $\mathcal{N}\left(\begin{bmatrix}1 & 2\\ 3 & 6\end{bmatrix}\right) = c\begin{bmatrix}-2\\ 1\end{bmatrix}$, where c is a scalar.

Remark. The other two fundamental subspaces are rowspace or coimage and left nullspace or cokernel defined as $\mathcal{C}(A^T)$ and $\mathcal{N}(A^T)$ respectively.

Matrices

Definition (Matrix). A rectangular array of scalars $A = \{a_{ij}\}_{i=1,\cdots,m,j=1,\cdots,n}$, $a_{ij} \in \mathbb{R}$.

Definition (Transpose). The transpose of a matrix A is matrix A^T produced by interchanging the rows with columns.

Definition (Identity matrix). A matrix $A \in \mathbb{R}^{n \times n}$ with $a_{ii} = 1, \forall i$ and $a_{ij} = 0, \forall i \neq j$

Definition (Symmetric matrix). A square matrix $A = \{a_{ij}\}$ with $a_{ij} = a_{ji}, \forall i, j$, i.e., transpose $A = A^T$ is a symmetric matrix. The set of symmetric matrices of size $n \times n$ is denoted by $\mathbb{S}^n$.

Definition (Positive (semi) definite matrix). A symmetric matrix with all positive (non-negative) eigen values. A matrix $A \in \mathbb{S}^n$. is positive (semi) definite (p.s.d.) if $\mathbf{x}^T A \mathbf{x} > 0$ ($\mathbf{x}^T A \mathbf{x} \geq 0$) for any nonzero vector $\mathbf{x}$. The set pf (semi) positive definite matrices of size $n \times n$ are denoted as ($\mathbb{S}^n_+$) $\mathbb{S}^n_{++}$.

Inner products and norms

Definition (Inner product). An inner product on real vector space V is a pairing that takes two vectors $\mathbf{x}, \mathbf{y} \in V$ and outputs a real number $\langle \mathbf{x}, \mathbf{y} \rangle = \mathbf{x}^T \mathbf{y} \in \mathbb{R}$. The inner product should satisfy three axioms with $\mathbf{x}, \mathbf{y}, \mathbf{z} \in V$ and scalars $\lambda_1, \lambda_2 \in \mathbb{R}$.

1. *Bilinearity*: $\langle \lambda_1 \mathbf{x} + \lambda_2 \mathbf{y}, \mathbf{z} \rangle = \lambda_1 \langle \mathbf{x}, \mathbf{z} \rangle + \lambda_2 \langle \mathbf{y}, \mathbf{z} \rangle$
   $\langle \mathbf{z}, \lambda_1 \mathbf{x} + \lambda_2 \mathbf{y} \rangle = \lambda_1 \langle \mathbf{z}, \mathbf{x} \rangle + \lambda_2 \langle \mathbf{z}, \mathbf{y} \rangle$
2. *Symmetry*: $\langle \mathbf{x}, \mathbf{y} \rangle = \langle \mathbf{y}, \mathbf{x} \rangle$
3. *Positivity*: $\langle \mathbf{x}, \mathbf{x} \rangle > 0$ whenever $\mathbf{x} \neq 0$, while $\langle \mathbf{0}, \mathbf{0} \rangle = 0$.

Remark. A vector space equipped with inner product is called an inner product space. Given an inner product, the associated norm of a vector $\mathbf{x} \in V$ is defined as

$$\boxed{\|x\| = \sqrt{\langle \mathbf{x}, \mathbf{x} \rangle}} \tag{1}$$

Remark. The standard inner product of two real matrices $X, Y \in \mathbb{R}^{m \times n}$ can be defined as $\langle X, Y \rangle = \textbf{trace}(X^T Y) = \sum_{j=1}^{n} \sum_{i=1}^{n} X_{ij}$

Cauchy-Schwarz inequality

Theorem

$$\boxed{|\langle \mathbf{x}, \mathbf{y} \rangle| \leq \|\mathbf{x}\| \|\mathbf{y}\|, \text{ for every } \mathbf{x}, \mathbf{y} \in V} \tag{2}$$

Equality holds iff $\mathbf{x}, \mathbf{y}$ are parallel vectors.

Proof.

One can prove it geometrically using the fact that $\mathbf{x}^T\mathbf{y} = \|\mathbf{x}\|\|\mathbf{y}\| cos\theta$ and $|cos\theta| \leq 1$.
Other way: The case when $\mathbf{y} = 0$ trivial. For $\mathbf{y} \neq \mathbf{0}$, let $\lambda \in \mathbb{R}$. We have,

$$0 \leq \|\mathbf{x}+\lambda\mathbf{y}\|^2 = \langle \mathbf{x}+\lambda\mathbf{y}, \mathbf{x}+\lambda\mathbf{y} \rangle = \langle \mathbf{x}, \mathbf{x} \rangle + 2\lambda\langle \mathbf{x}, \mathbf{y} \rangle + \lambda^2 \langle \mathbf{y}, \mathbf{y} \rangle = \|\mathbf{x}\|^2 + 2\lambda\langle \mathbf{x}, \mathbf{y} \rangle + \lambda^2\|\mathbf{y}\|^2 \tag{3}$$

with inequality holding only if $\mathbf{x} = -\lambda\mathbf{y}$, which requires $\mathbf{x}$ and $\mathbf{y}$ to be parallel vectors. Considering (3) to be quadratic function of λ, let's substitute minimum value of $\lambda = -\frac{\langle \mathbf{x}, \mathbf{y} \rangle}{\|\mathbf{y}\|^2}$ in (3).

$$0 \leq \|\mathbf{x}\|^2 - 2\frac{\langle \mathbf{x}, \mathbf{y} \rangle^2}{\|\mathbf{y}\|^2} + \frac{\langle \mathbf{x}, \mathbf{y} \rangle^2}{\|\mathbf{y}\|^2} = \|\mathbf{x}\|^2 - \frac{\langle \mathbf{x}, \mathbf{y} \rangle^2}{\|\mathbf{y}\|^2} \tag{4}$$

Rearranging this inequality, we have $\langle \mathbf{x}, \mathbf{y} \rangle^2 \leq \|\mathbf{x}\|^2\|\mathbf{y}\|^2$. The equality holds iff $\mathbf{x}, \mathbf{y}$ are parallel or $\mathbf{y} = \mathbf{0}$, which is of course parallel to every $\mathbf{x}$. Taking (positive) square root proves the result. $\square$

The triangle inequality

Theorem

The norm associated with inner product satisfies triangle inequality

$$\|\mathbf{x}+\mathbf{y}\| \leq \|\mathbf{x}\| + \|\mathbf{y}\|, \textit{ for all } \mathbf{x}, \mathbf{y} \in V \tag{5}$$

Equality holds iff $\mathbf{x}$ *and* $\mathbf{y}$ *are parallel vectors.*

Proof.

Other way: The case when $\mathbf{y} = 0$ trivial. For $\mathbf{y} \neq \mathbf{0}$, let $\lambda \in \mathbb{R}$. We have,

$$\|\mathbf{x}+\mathbf{y}\|^2 = \|\mathbf{x}\|^2 + 2\langle\mathbf{x},\mathbf{y}\rangle + \|\mathbf{y}\|^2 \leq \|\mathbf{x}\|^2 + 2\|\mathbf{x}\|\|\mathbf{y}\| + \|\mathbf{y}\|^2 = (\|\mathbf{x}\| + \|\mathbf{y}\|)^2. \tag{6}$$

□

Definition (Orthogonal vectors). Two vectors $\mathbf{x}, \mathbf{y} \in V$ of inner product space V are called orthogonal is their inner product vanishes, i.e., $\langle\mathbf{x},\mathbf{y}\rangle = 0$.

Norms

Definition (Norm). A function $f : \mathbb{R}^n \mapsto \mathbb{R}$ is called a norm if f is

1. *Non-negative*: $f(\mathbf{x}) \geq 0, \forall \mathbf{x} \in \mathbb{R}^n$
2. *Definite*: $f(\mathbf{x}) = 0$ iff $\mathbf{x} = 0$
3. *Homogeneous*: $f(t\mathbf{x}) = |t| f(\mathbf{x}), \forall \mathbf{x} \in \mathbb{R}^n, \forall t \in R$
4. satisfies *Triangle inequality*: $f(\mathbf{x} + \mathbf{y}) \leq f(\mathbf{x}) + f(\mathbf{y}), \forall \mathbf{x}, \mathbf{y} \in \mathbb{R}^n$.

Examples:

1. l_p norm, $\|\mathbf{x}\|_p = \left(\sum_{i=1}^n |x_i|^p\right)^{\frac{1}{p}}, 1 \leq p \leq \infty$. Triangular inequality for general p is known as Minkoswski's inequality. $\left(\sum_{i=1}^n |x_i + y_i|^p\right)^{\frac{1}{p}} \leq \left(\sum_{i=1}^n |x_i|^p\right)^{\frac{1}{p}} + \left(\sum_{i=1}^n |y_i|^p\right)^{\frac{1}{p}}$.
2. $l_0$norm is not a norm. Why?

Sets

Definition (Set). A collection of objects satisfying some conditions.

Definition (Interior point). An element $\mathbf{x} \in C \subseteq \mathbb{R}^n$ is called an interior point of C if $\exists \epsilon > 0$ for which $\{\mathbf{y} \mid \|\mathbf{y} - \mathbf{x}\| \leq \epsilon\} \subseteq C$, i.e., a ball centered at $\mathbf{x}$ of radius ϵ lies inside C.

Definition (Interior of a set). The set of all interior points of C is called interior of C, denoted by $\textbf{int}(C)$. A set is solid if it has nonempty interior.

Definition (Open set). A set C is open if all of its elements are interior points, i.e., $\textbf{int}(C) = C$.

Definition (Closed set). A set $C \subseteq \mathbb{R}^n$ is closed if $\mathbb{R}^n \backslash C$ is open. Alternatively, a set C is closed iff for any convergent sequence $\{\mathbf{x}_k\} \in S$ with limit point $\bar{\mathbf{x}}$, we also have $\bar{\mathbf{x}} \in C$[1].
Definition (Closure of a set). The closure of a set $C \subseteq \mathbb{R}^n$ is defined as $\textbf{cl}(C) = \mathbb{R}^n \backslash \textbf{int}(\mathbb{R}^n \backslash C)$.

[1] A limit point $\bar{\mathbf{x}}$ of any convergent sequence should lie in the interior or on the boundary of the set, otherwise $\exists \epsilon > 0$ s.t. $\{\mathbf{x} \mid \|\mathbf{x} - \bar{\mathbf{x}}\| < \epsilon\}\} \cap C = \phi$

Compact sets and projections

Definition (Boundary of a set). The boundary of a set $C \subseteq \mathbb{R}^n$ is defined as $\mathbf{bd}(C) = \mathbf{cl}(C) \backslash \mathbf{int} C$.

Remark. A set C is closed if it contains its boundary, i.e., $\mathbf{bd}(C) \subseteq C$. It is open if it contains no boundary points, i.e., $\mathbf{bd}(C) \cap C = \phi$.

Definition (Bounded set).: A set $C \subseteq \mathbb{R}$ is a bounded if $\|\mathbf{x} - \mathbf{y}\| \leq \epsilon, \forall \mathbf{x}, \mathbf{y} \in C$ for some finite $\epsilon > 0$.

Definition (Compact set). A set C is compact it is both closed as well as bounded.

Definition (Projection of a point onto a set). The projection of a point $\mathbf{x} \in \mathbb{R}^n$ onto a set $C \subseteq$ is point in C which is closest to $\mathbf{x}$, i.e., $\mathbf{proj}_x(C) = \text{argmin}_{y \in C}\{\|y - x\|\}$.

Definition (Projection of a set onto a space).Let $C \subseteq \mathbb{R}^n \times \mathbb{R}^p$ whose feasible points are denoted by $(\mathbf{x}, \mathbf{y})$ with $\mathbf{x} \in \mathbb{R}^n$ and $\mathbf{y} \in \mathbb{R}^p$. We define the projection of set C onto the space of variables $\mathbf{x}$ as the set

$$\mathbf{proj}_x(C) = \left\{\mathbf{x} \in \mathbb{R}^n \;\middle|\; \exists \mathbf{y} \in \mathbb{R}^p \text{ with } (\mathbf{x}, \mathbf{y}) \in C\right\}$$

Max, min, inf, sup

Definition (Maximum). Let $S \subseteq \mathbb{R}$. We say that x is a maximum of S iff $x \in S$ and $x \geq y, \forall y \in S$.

Definition (Minimum). Let $S \subseteq \mathbb{R}$. We say that x is a minimum of S iff $x \in S$ and $x \leq y, \forall y \in S$.

Definition (Bounds). Let $S \subseteq \mathbb{R}$. We say that u is an upper bound of S iff $u \geq x, \forall x \in S$. Similarly, l is a lower bound of S iff $l \leq x, \forall x \in S$.

Definition (Supremum). Let $S \subseteq \mathbb{R}$. We define the supremum of S denoted by $\mathbf{sup}(S)$ to be the smallest upper bound of S. If no such upper bound exists, then we set $\mathbf{sup}(S) = +\infty$.

Definition (Infimum). Let $S \subseteq \mathbb{R}$. We define the infimum of S denoted by $\mathbf{inf}(S)$ to be the largest lower bound of S. If no such lower bound exists, then we set $\mathbf{inf}(S) = -\infty$

Remark. If $x \in S$ such that $x = \mathbf{sup}(S)$, we say that supremum of S is achieved (which means that there is a maximum to the problem). Similar definition for whether infimum is achieved.

Weierstrass Extreme Value Theorem

Theorem

Let $X \subseteq \mathbb{R}^n$. A continuous function $f : X \mapsto \mathbb{R}$ defined on a closed and bounded set X attain a maximum and minimum value.

Proof (Bazaraa et al. (2006)).

We present the proof for minimum. A similar proof can be constructed for maximum. Since f is continuous on X (which is both bounded and closed), f is bounded below on X. Since $S \neq \phi$, there exists a greatest lower bound $l = \inf\{f(\mathbf{x}) \bigm| \mathbf{x} \in X\}$. Let $0 < \epsilon < 1$, and consider the sets $X_k = \{\mathbf{x} \in X \bigm| l \leq f(\mathbf{x}) \leq l + \epsilon^k\}$ for each $k = 1, 2, \cdots$. By the definition of infimum $X_k \neq \phi$ for each k, so we may construct a sequence of points $\{\mathbf{x}_k\} \in X$ by selecting a point $\mathbf{x}_k \in X_k$ for each $k = 1, 2, \cdots$. Since X is bounded, there exists a convergent sequence $\{\mathbf{x}_k\} \mapsto \bar{\mathbf{x}}$. By closedness of X, we have $\bar{\mathbf{x}} \in X$ and by continuity of f, since $\alpha \leq f(\mathbf{x}_k) \leq \epsilon^k, \forall k$, we have $\alpha = \lim_{k \mapsto \infty} f(\mathbf{x}_k) = f(\bar{\mathbf{x}})$. We have shown that infimum is achieved at $\bar{\mathbf{x}}$. □

Linear subspaces, affine sets, cones, convex sets

A set $C \subseteq \mathbb{R}^n$ is said to be

1. linear subspace iff $\forall \mathbf{x}, \mathbf{y} \in C$ and $\lambda_1, \lambda_2 \in \mathbb{R}$, we have $\lambda_1 \mathbf{x} + \lambda_2 \mathbf{y} \in C$.
2. cone iff $\forall \mathbf{x} \in C$ and $\lambda \in \mathbb{R}$ such that $\lambda \geq 0$, we have $\lambda \mathbf{x} \in C$.
3. affine set iff $\forall \mathbf{x}, \mathbf{y} \in C$ and $\lambda_1, \lambda_2 \in \mathbb{R}$ such that $\lambda_1 + \lambda_2 = 1$, we have $\lambda_1 \mathbf{x} + \lambda_2 \mathbf{y} \in C$ (line passing through any two points in C lies in C).
4. convex set iff $\forall \mathbf{x}, \mathbf{y} \in C$ and $\lambda_1, \lambda_2 \in \mathbb{R}$ such that $\lambda_1, \lambda_2 \geq 0$ and $\lambda_1 + \lambda_2 = 1$, we have $\lambda_1 \mathbf{x} + \lambda_2 \mathbf{y} \in C$ (line segment between any two points in C lies in C).

Linear, conic, affine, and convex combination of vectors

For a given set of vectors $\mathbf{x}^1, \mathbf{x}^2, \cdots, \mathbf{x}^k$ and scalars $\lambda_1, \lambda_2, \cdots, \lambda_k$, the weighted combination $\boxed{\mathbf{x} = \sum_{i=1}^{k} \lambda_i \mathbf{x}^i}$ is said to be

1. linear combination of vectors $\mathbf{x}^1, \mathbf{x}^2, \cdots, \mathbf{x}^k$ if $\lambda_1, \cdots, \lambda_k \in \mathbb{R}$
2. conic combination of vectors $\mathbf{x}^1, \mathbf{x}^2, \cdots, \mathbf{x}^k$ if $\lambda_1, \cdots, \lambda_k \in \mathbb{R}$ and $\lambda_1, \cdots, \lambda_k \geq 0$.
3. affine combination of vectors $\mathbf{x}^1, \mathbf{x}^2, \cdots, \mathbf{x}^k$ if $\lambda_1, \cdots, \lambda_k \in \mathbb{R}$ and $\sum_{i=1}^{k} \lambda_i = 1$.
4. convex combination of vectors $\mathbf{x}^1, \mathbf{x}^2, \cdots, \mathbf{x}^k$ if $\lambda_1, \cdots, \lambda_k \in \mathbb{R}$ such $\lambda_1, \cdots, \lambda_k \geq 0$ and $\sum_{i=1}^{k} \lambda_i = 1$.

Hulls

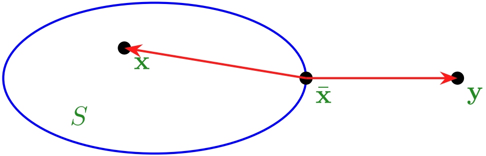
Accordingly, we can define

1. Linear hull of set C denoted by **lin**(C) is minimal inclusion-wise linear subspace containing C.
2. Conic hull of set C denoted by **cone**(C) is minimal inclusion-wise cone containing C.
3. Affine hull of set C denoted by **aff**(C) is minimal inclusion-wise affine set containing C.
4. Convex hull of set C denoted by **conv**(C) is minimal inclusion-wise convex set containing C.

Theorem

Let X *be nonempty, closed convex set in* $\mathbb{R}^n$ *and* $\mathbf{y} \notin S$*. Then, there exists a unique point* $\bar{\mathbf{x}} \in X$ *with minimum distance to* $\mathbf{y}$*. Furthermore,* $\bar{\mathbf{x}}$ *is also a minimizing point if and only if*

$$(\mathbf{y} - \bar{\mathbf{x}})^T(\mathbf{x} - \bar{\mathbf{x}}) \leq 0, \forall \mathbf{x} \in S$$



Theorem

Let X be nonempty, closed convex set in $\mathbb{R}^n$ and $\mathbf{y} \notin S$. Then, there exists a unique point $\bar{\mathbf{x}} \in X$ with minimum distance to $\mathbf{y}$. Furthermore, $\bar{\mathbf{x}}$ is also a minimizing point if and only if

$$(\mathbf{y} - \bar{\mathbf{x}})^T(\mathbf{x} - \bar{\mathbf{x}}) \leq 0, \forall \mathbf{x} \in S$$

Proof (Bazaraa et al. (2006)).

Let us establish the first result. Since $X \neq \phi, \exists \tilde{\mathbf{x}} \in X$. Consider the set $\tilde{X} = X \cap \{\mathbf{x} \in X \mid \|\mathbf{y} - \mathbf{x}\| \leq \|\mathbf{y} - \tilde{\mathbf{x}}\|\}$. The task of finding the closest point $\inf\{\|\mathbf{y} - \mathbf{x}\| \mid \mathbf{x} \in X\}$ is equivalent to $\inf\{\|\mathbf{y} - \mathbf{x}\| \mid \mathbf{x} \in \tilde{X}\}$. But the latter involves finding a minimum of a continuous function over a compact set, so by Weierstrass theorem, we have a minimum point $\bar{\mathbf{x}} \in X$ which is closest to $\mathbf{y}$.
To show uniqueness, suppose there exists another $\bar{\mathbf{x}}' \in X$ such that $\|\mathbf{y} - \bar{\mathbf{x}}\| = \|\mathbf{y} - \bar{\mathbf{x}}'\| = \alpha$. Due to convexity of X, the point $\frac{\bar{\mathbf{x}} + \bar{\mathbf{x}}'}{2} \in X$ and using triangle inequality, we have

$$\left\|\mathbf{y} - \frac{\bar{\mathbf{x}} + \bar{\mathbf{x}}'}{2}\right\| \leq \frac{1}{2}\|\mathbf{y} - \bar{\mathbf{x}}\| + \frac{1}{2}\|\mathbf{y} - \bar{\mathbf{x}}'\| = \alpha$$

Proof contd.

The strict inequality cannot hold because it will contradict the fact that $\bar{\mathbf{x}}$ is the closest point. Therefore, equality holds. Therefore, $\mathbf{y} - \bar{\mathbf{x}} = \lambda(\mathbf{y} - \bar{\mathbf{x}}^{'})$ for some λ. Since $\|\mathbf{y} - \bar{\mathbf{x}}\| = \|\mathbf{y} - \bar{\mathbf{x}}^{'}\| = \alpha$, we have $|\lambda| = 1$. $\lambda \neq -1$ because otherwise $\mathbf{y} \notin X$. So, $\lambda = 1$, proving that $\bar{\mathbf{x}} = \bar{\mathbf{x}}^{'}$.

" $\Longleftarrow$ " Let $\mathbf{x} \in X$. Then,

$$\|\mathbf{y} - \mathbf{x}\|^2 = \|\mathbf{y} - \bar{\mathbf{x}} + \bar{\mathbf{x}} - \mathbf{x}\|^2 = \|\mathbf{y} - \bar{\mathbf{x}}\|^2 + \|\bar{\mathbf{x}} - \mathbf{x}\|^2 + 2(\bar{\mathbf{x}} - \mathbf{x})^T(\mathbf{y} - \bar{\mathbf{x}})$$

Since $\|\bar{\mathbf{x}} - \mathbf{x}\|^2 \geq 0$ and $(\bar{\mathbf{x}} - \mathbf{x})^T(\mathbf{y} - \bar{\mathbf{x}}) \geq 0$ by assumption, we have $\|\mathbf{y} - \mathbf{x}\|^2 \geq \|\mathbf{y} - \bar{\mathbf{x}}\|^2$ showing that $\bar{\mathbf{x}}$ is the minimizing point.

" $\Longrightarrow$ " Assume that $\bar{\mathbf{x}}$ is the minimizing point, i.e., $\|\mathbf{y} - \mathbf{x}\|^2 \geq \|\mathbf{y} - \bar{\mathbf{x}}\|^2, \forall \mathbf{x} \in X$. Let $\mathbf{x} \in X$ and note that $\bar{\mathbf{x}} + \lambda(\mathbf{x} - \bar{\mathbf{x}}) \in X$ for $\lambda \in [0, 1]$ by the convexity of X. Therefore,

$$\begin{aligned}
\|\mathbf{y} - (\bar{\mathbf{x}} + \lambda(\mathbf{x} - \bar{\mathbf{x}}))\|^2 &\geq \|\mathbf{y} - \bar{\mathbf{x}}\|^2 \\
\|\mathbf{y} - \bar{\mathbf{x}}\|^2 + \lambda^2\|\mathbf{x} - \bar{\mathbf{x}}\|^2 - 2\lambda(\mathbf{y} - \bar{\mathbf{x}})^T(\mathbf{x} - \bar{\mathbf{x}}) &\geq \|\mathbf{y} - \bar{\mathbf{x}}\|^2 \\
\|\mathbf{x} - \bar{\mathbf{x}}\|^2 - 2\lambda(\mathbf{y} - \bar{\mathbf{x}})^T(\mathbf{x} - \bar{\mathbf{x}}) &\geq 0 \\
2(\mathbf{y} - \bar{\mathbf{x}})^T(\mathbf{x} - \bar{\mathbf{x}}) \leq \lambda\|\mathbf{x} - \bar{\mathbf{x}}\|^2
\end{aligned}$$

due to dividing by $\lambda \in [0, 1]$. Let $\lambda \mapsto 0^+$, the result follows. $\square$

Separating hyperplane theorem

Theorem

Suppose C and D are two disjoint convex sets i.e., $C \cap D = \phi$. Then, there exists $\mathbf{a} \neq 0$ and b such that

$$\boxed{\mathbf{a}^T\mathbf{x} \leq b, \forall \mathbf{x} \in C \quad \textit{and} \quad \mathbf{a}^T\mathbf{x} \geq b, \forall \mathbf{x} \in D}$$

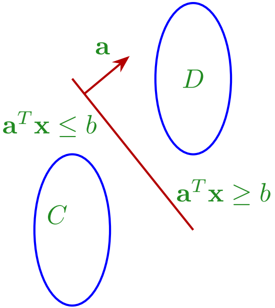


Figure: Separating Hyperplane Theorem

Separation of a convex set and a point

Theorem

Let C be a nonempty convex set in $\mathbb{R}^n$ and $\mathbf{y} \notin S$. Then there exists a nonzero vector $\mathbf{a}$ and a scalar b such that $\mathbf{a}^T\mathbf{y} > b$ and $\mathbf{a}^T\mathbf{x} \leq b, \forall \mathbf{x} \in S$.

Proof (Bazaraa et al. (2006)).

Using previous theorem, there is a unique minimizing point $\bar{\mathbf{x}} \in S$ such that $(\mathbf{x} - \bar{\mathbf{x}})^T(\mathbf{y} - \bar{\mathbf{x}}) \leq 0, \forall \mathbf{x} \in S$. Letting $\mathbf{a} = (\mathbf{y} - \bar{\mathbf{x}}) \neq 0$ and $b = \bar{\mathbf{x}}^T(\mathbf{y} - \bar{\mathbf{x}})$, we get $\mathbf{a}^T\mathbf{x} \leq b, \forall \mathbf{x} \in S$ while $\mathbf{a}^T\mathbf{y} - b = (\mathbf{y} - \bar{\mathbf{x}})^T\mathbf{y} - \bar{\mathbf{x}}^T(\mathbf{y} - \bar{\mathbf{x}}) = \|\mathbf{y} - \bar{\mathbf{x}}\|^2 > 0$, which completes the proof. $\square$

Supporting hyperplane

Definition (Supporting hyperplane). Let S be nonempty set in $\mathbb{R}^n$ and let $\bar{\mathbf{x}} \in \mathbf{bd}(S)$. A hyperplane $H = \{\mathbf{x} \mid \mathbf{a}^T(\mathbf{x} - \bar{\mathbf{x}}) = 0\}$ is called a supporting hyperplane of S at $\bar{\mathbf{x}}$. Equivalently, $H = \{\mathbf{x} \mid \mathbf{a}^T(\mathbf{x} - \bar{\mathbf{x}}) = 0\}$ is a supporting hyperplane of S at $\bar{\mathbf{x}} \in \mathbf{bd}(S)$ if $\mathbf{a}^T\bar{\mathbf{x}} = \inf\{\mathbf{a}^T\mathbf{x} \mid \mathbf{x} \in S\}$ or $\mathbf{a}^T\bar{\mathbf{x}} = \sup\{\mathbf{a}^T\mathbf{x} \mid \mathbf{x} \in S\}$

Theorem

Let S be a nonempty convex set in $\mathbb{R}^n$ and let $\bar{\mathbf{x}} \in \mathbf{bd}(S)$. Then there exists a hyperplane that supports S at $\bar{\mathbf{x}}$; i.e., there exists a nonzero vector $\mathbf{a}$ such that $\mathbf{a}^T(\mathbf{x} - \bar{\mathbf{x}}) \leq 0, \forall \mathbf{x} \in \mathbf{cl}(S)$.

Polyhedra

Definition (Hyperplane). $\{\mathbf{x} \in \mathbb{R}^n : \mathbf{a}^T\mathbf{x} = b, \mathbf{a} \neq 0\}$

Definition (Halfspace). $\{\mathbf{x} \in \mathbb{R}^n : \mathbf{a}^T\mathbf{x} \geq b, \mathbf{a} \neq 0\}$

Definition (Polyhedron). A set $P \subseteq \mathbb{R}^n$ is called a polyhedron if P is the intersection of a finite number of halfspaces. $P = \{\mathbf{x} \in \mathbb{R}^n : A\mathbf{x} \leq \mathbf{b}\}$

Definition (Polytope). A bounded polyhedron is called a polytope.
Question Is $\{\mathbf{x} \in \mathbb{R}^n : A\mathbf{x} = \mathbf{b}, \mathbf{x} \geq 0\}$ a polyhedron?

Definition (Extreme point). Let P be a polyhedron. Then, $\mathbf{x} \in P$ is an extreme point of P if we cannot express $\mathbf{x}$ as a convex combination of other points in P.
Question Is $P = \{\mathbf{x} \in \mathbb{R}^n \mid A\mathbf{x} = \mathbf{b}\}$, where $A \in \mathbb{R}^{m \times n}$ and $\mathbf{b} \in \mathbb{R}^m$ a convex set?

Definition (Ray). Let P be a polyhedron. Then, $\mathbf{r}$ is a recession direction or extreme ray of P, if, for every $\bar{\mathbf{x}} \in P$, $\bar{\mathbf{x}} + \lambda\mathbf{r} \in P, \forall \lambda \geq 0$.

Definition (Extreme ray). Let P be a polyhedron. Then, $\mathbf{r} \in P$ is an extreme ray of P if we cannot express $\mathbf{r}$ as a conic combination of other rays in P.

Minkowski-Weyl (representation) theorem for polyhedra

Theorem

Let $P = \{\mathbf{x} \in \mathbb{R}^n \mid A\mathbf{x} = \mathbf{b}\}$, *where* $A \in \mathbb{R}^{m\times n}, \mathbf{b} \in \mathbb{R}^m$. *Further, let* $\mathbf{v}^1, \cdots, \mathbf{v}^k$ *be the extreme points of* P *and* $\mathbf{r}^1, \mathbf{r}^2, \cdots, \mathbf{r}^h$ *be the extreme rays of* S. *Then,* $\mathbf{x} \in S$ *if and only if* $\mathbf{x}$ *can be expressed as*

$$\mathbf{x} = \sum_{j=1}^{k} \lambda_j \mathbf{v}^j + \sum_{l=1}^{h} \mu_l \mathbf{r}^l$$

$$\sum_{j=1}^{k} \lambda_j = 1$$

$$\lambda_j \geq 0, \forall j = 1, \cdots, k$$

$$\mu_l \geq 0, \forall l = 1, \cdots, h$$

Remark. In case of a polyhedra corresponding to a network flow problem, any feasible flow in a network can be decomposed into a sum of path flows and cycle (circulation) flows. This result is also known as flow decomposition theorem.

Functions

Consider a multivariable function $f : \mathbb{R}^n \mapsto \mathbb{R}$

- Gradient of f at $\mathbf{x}$

$$\nabla f(\mathbf{x}) = \begin{bmatrix} \frac{\partial f(\mathbf{x})}{\partial x_1} \\ \vdots \\ \frac{\partial f(\mathbf{x})}{\partial x_n} \end{bmatrix}$$

  with $\frac{\partial f(\mathbf{x})}{\partial x_i} = \lim_{h \mapsto 0} \frac{f(\mathbf{x}+h\mathbf{e}_i)-f(\mathbf{x})}{h}$, where $\mathbf{e}_i$ is the i^{th} unit vector

- Hessian matrix of f at $\mathbf{x}$

$$\nabla^2 f(\mathbf{x}) = \left[\frac{\partial^2 f(\mathbf{x})}{\partial x_i \partial x_j}\right]_{n \times n}$$

  Remark. If f is twice continuously differentiable then $\nabla^2 f$ is a symmetric matrix.

- Jacobian of a vector-valued function $f(\mathbf{x}) = \begin{bmatrix} f_1(\mathbf{x}) \\ \vdots \\ f_p(\mathbf{x}) \end{bmatrix}$ is

$$\begin{bmatrix} \frac{\partial f_1(\mathbf{x})}{\partial x_1} & \frac{\partial f_1(\mathbf{x})}{\partial x_2} & \dots & \frac{\partial f_1(\mathbf{x})}{\partial x_n} \\ \frac{\partial f_2(\mathbf{x})}{\partial x_1} & \frac{\partial f_2(\mathbf{x})}{\partial x_2} & \dots & \frac{\partial f_2(\mathbf{x})}{\partial x_n} \\ & \vdots & & \\ \frac{\partial f_p(\mathbf{x})}{\partial x_1} & \frac{\partial f_p(\mathbf{x})}{\partial x_2} & \dots & \frac{\partial f_p(\mathbf{x})}{\partial x_n} \end{bmatrix}$$

Convex function

- A function $f : \mathbb{R}^n \mapsto \mathbb{R}$ is a convex function if $\textbf{dom}(f)$ is convex set and if for all $\mathbf{x}_1, \mathbf{x}_2 \in \textbf{dom}(f)$ and $0 \leq \lambda \leq 1$, we have

$$\boxed{f(\lambda \mathbf{x} + (1-\lambda)\mathbf{y}) \leq \lambda f(\mathbf{x}) + (1-\lambda) f(\mathbf{y})}$$

- (First order conditions) A differentiable function $f : \mathbb{R}^n \mapsto \mathbb{R}$ is a convex function if and only if $\textbf{dom}(f)$ is convex set and

$$\boxed{f(\mathbf{x}_2) \geq f(\mathbf{x}_1) + \nabla f(\mathbf{x}_1)^T (\mathbf{x}_2 - \mathbf{x}_1), \forall \mathbf{x}_1, \mathbf{x}_2 \in \textbf{dom}(f)}$$

  The first order Taylor series approximation of f is a global underestimator this function.
- (Second order conditions) A twice differentiable function $f : \mathbb{R}^n \mapsto \mathbb{R}$ is a convex function if and only if $\textbf{dom}(f)$ is convex set and its Hessian is positive semidefinite, i.e.,

$$\boxed{\nabla^2 f(\mathbf{x}) \succcurlyeq 0, \forall \mathbf{x} \in \textbf{dom}(f)}$$

Remark. A function is concave is $-f$ is a convex function.

Optimization Problem

Components of an optimization problem

- Decisions
- Constraints
- Objective

Optimization seeks to choose some decisions to optimize (maximize or minimize) an objective subject to certain constraints.

Common Framework

Given $f, g_i, h_i : \mathbb{R}^n \mapsto \mathbb{R}$

$$Z = \underset{\mathbf{x}}{\text{minimize}} \quad f(\mathbf{x}) \tag{7a}$$

$$\text{subject to} \quad g_i(\mathbf{x}) \leq 0, \forall i = 1, 2, ..., p \tag{7b}$$

$$g_j(\mathbf{x}) \geq 0, \forall j = 1, 2, ..., q \tag{7c}$$

$$h_k(\mathbf{x}) = 0, \forall k = 1, 2, ..., r \tag{7d}$$

- Decisions: $\mathbf{x}$, Objective: $f(\mathbf{x})$, and Constraints: (7b)-(7d)
- (7b), (7c), and (7d): set of "$\leq$", "$\geq$", and equality constraints
- $\mathcal{X} = \{\mathbf{x} \in \mathbb{R}^n :$ (7b) $-$ (7d)$\} \cap \textbf{dom}(f) \cap_{i=1}^{p} \textbf{dom}(g_i) \cap_{j=1}^{q} \textbf{dom}(g_j) \cap_{k=1}^{r} \textbf{dom}(h_k)$ define the feasible region.
- Any $\hat{\mathbf{x}}$ satisfying all the constraints is a feasible solution.
- Any $\mathbf{x}^* \in \mathcal{X}$ satisfying $f(\mathbf{x}^*) \leq f(\mathbf{x}), \forall \mathbf{x} \in \mathcal{X}$ is an optimal solution.
- $f(\mathbf{x}^*)$ is known as optimal objective value.

Remark. Above problem is a convex optimization problem if all functions are convex and feasible region is a convex set.

For convex problems, local optimal $\implies$ global optimal

Definition (Local optimal solution). For an optimization problem $\min_{\mathbf{x}}\{f(\mathbf{x}) \mid x \in S\}$, $\mathbf{x}^*$ is a local optimal solution if $\exists \epsilon > 0$, $f(\mathbf{x}^*) \leq f(\mathbf{x})$, $\forall \mathbf{x} \in S \cap \{\mathbf{x} \mid \|\mathbf{x} - \mathbf{x}^*\| \leq \epsilon\}$

Theorem

For a convex optimization problem $\min_{\mathbf{x}}\{f(\mathbf{x}) \mid x \in S\}$, a local optimal solution $\mathbf{x}^*$ is also a global optimal solution (i.e., $f(\mathbf{x}^*) \leq f(\mathbf{x})$, $\forall \mathbf{x} \in S$).

Proof.

Let's assume that for a convex optimization problem, $\mathbf{x}^*$ is local optimal solution but it is not global optimal, i.e., $\exists \hat{x} \in S$ such that $f(\hat{\mathbf{x}}) < f(\mathbf{x}^*)$. Let $0 < \lambda < 1$, consider a point $(\lambda \hat{\mathbf{x}} + (1-\lambda)\mathbf{x}^*)$ such that $\|(\lambda \hat{\mathbf{x}} + (1-\lambda)\mathbf{x}^*) - \mathbf{x}^*\| < \epsilon$. Note that $(\lambda \hat{\mathbf{x}} + (1-\lambda)\mathbf{x}^*) \in S$ since S is a convex set. Since $\mathbf{x}^*$ is local optimal solution, we have

$$f(\lambda \hat{\mathbf{x}} + (1-\lambda)\mathbf{x}^*) \geq f(\mathbf{x}^*) \tag{8}$$

Also, since f is a convex function,

$$f(\lambda \hat{\mathbf{x}} + (1-\lambda)\mathbf{x}^*) \leq \lambda f(\hat{\mathbf{x}}) + (1-\lambda) f(\mathbf{x}^*) < \lambda f(\mathbf{x}^*) + (1-\lambda) f(\mathbf{x}^*) = f(\mathbf{x}^*)$$

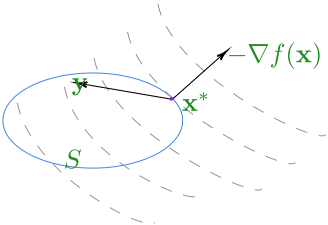
which is a contradiction from (8). $\square$

Optimality criterion for convex optimization problem with differentiable objective function

Theorem

For a convex optimization problem $\min_{\mathbf{x}}\{f(\mathbf{x}) \mid \mathbf{x} \in S\}$ *with differentiable* f, $\mathbf{x}^* \in S$ *is optimal if and only if*

$$\boxed{f(y) \geq f(x) + \nabla f(x)^T(\mathbf{y}-\mathbf{x}), \forall \mathbf{y} \in S}$$



Remark. For unconstrained problems, we can choose sufficiently close $\mathbf{y} = \mathbf{x} - t\nabla f(\mathbf{x})$ to $\mathbf{x}$, the above condition reduces to $\nabla f(\mathbf{x}) = 0$ (the well known necessary and sufficient condition).

The Lagrangian

Consider the following convex optimization problem

$$
\begin{aligned}
Z_P^* = \underset{\mathbf{x}}{\text{minimize}} \quad & f_0(\mathbf{x}) && \text{(9a)} \\
\text{subject to} \quad & f_i(\mathbf{x}) \le 0, \forall i = 1, 2, ..., m && \text{(9b)} \\
& h_k(\mathbf{x}) = 0, \forall k = 1, 2, ..., p && \text{(9c)}
\end{aligned}
$$

We define the Lagrangian $L : \mathbb{R}^n \times \mathbb{R}^m \times \mathbb{R}^p \mapsto \mathbb{R}$ associated with (9) as

$$L(\mathbf{x}, \lambda, \nu) = f_0(\mathbf{x}) + \sum_{i=1}^{m} \lambda_i f_i(\mathbf{x}) + \sum_{k=1}^{p} \nu_k h_k(\mathbf{x})$$

where, $\{\lambda_i\}_{i=1}^m$ and $\{\nu_k\}_{k=1}^p$ are the Lagrangian multipliers or dual variables associated to constraints (9b) and (9c) respectively. We will refer to (9) as the Primal problem.

Lagrange dual function

Definition (Lagrange dual function). The Lagrange dual function $g : \mathbb{R}^m \times \mathbb{R}^p \mapsto \mathbb{R}$ is defined as minimum value of $L(\mathbf{x}, \lambda, \nu)$ over $\mathbf{x}$

$$g(\lambda, \nu) = \inf_{\mathbf{x} \in \mathcal{F}} \left\{ f_0(\mathbf{x}) + \sum_{i=1}^{m} \lambda_i f_i(\mathbf{x}) + \sum_{k=1}^{p} \nu_k h_k(\mathbf{x}) \right\} \tag{10}$$

Remark. The Lagrange dual function provides a lower bound on the optimal value of (9), i.e.,

$$Z_P^* \geq g(\lambda, \nu)$$

Remark. The dual function is always (since it is affine function of (λ^*, ν^*)) concave even when the primal problem is not convex. Definition (Lagrange Dual problem).

$$Z_D^* = \underset{\lambda, \nu}{\text{maximize}} \quad g(\lambda, \nu) \tag{11a}$$

$$\text{subject to} \quad \lambda \succcurlyeq 0 \tag{11b}$$

Remark. (Weak Duality) $Z_P^* \geq Z_D^*$. The difference $Z_P^* - Z_D^*$ is called duality gap (Useful from algorithmic perspective.)

Remark. (Strong Duality) $Z_P^* = Z_D^*$ For convex problems it usually (not always) holds. There are some constraint qualifications under which strong duality holds. One such constraint qualification is Slater's condition.

Complementary slackness

Suppose $\mathbf{x}^*$ and (λ^*, ν^*) are optimal primal and dual values respectively. Further suppose that strong duality holds, i.e., $Z_P^* = Z_D^*$.

$$\begin{aligned} f_0(\mathbf{x}) &= g(\lambda^*, \nu^*) \\ &= \inf_{\mathbf{x}} \left\{ f_0(\mathbf{x}) + \sum_{i=1}^{m} \lambda_i^* f_i(\mathbf{x}) + \sum_{k=1}^{p} \nu_k^* h_k(\mathbf{x}) \right\} \\ &\leq f_0(\mathbf{x}^*) + \sum_{i=1}^{m} \lambda_i^* f_i(\mathbf{x}^*) + \sum_{k=1}^{p} \nu_k^* h_k(\mathbf{x}^*) \\ &\leq f_0(\mathbf{x}^*) \end{aligned}$$

Above equation implies $\sum_{i=1}^{m} \lambda_i^* f_i(\mathbf{x}^*) = 0$. Since each term in this summation is non positive, we conclude that

$$\boxed{\lambda_i^* f_i(\mathbf{x}^*) = 0}, \ \forall i = 1, \cdots, m$$

This condition is called complementary slackness. It holds for any primal and dual optimal values (when strong duality holds). It implies that when $\lambda_i^* > 0 \implies f_i(\mathbf{x}^*) = 0$ or equivalently, $f_i(\mathbf{x}^*) < 0 \implies \lambda_i^* = 0$.

Karush Kuhn Tucker (KKT) conditions

Suppose $f_0, \{f_i\}_{i=1}^m \{h_k\}_{k=1}^p$ are differentiable functions and $\mathbf{x}^*$ and (λ^*, ν^*) are pair of primal and dual values with zero duality gap. Then, the problem must satisfy the following conditions which are famously called KKT conditions.

1. Primal feasibility

$$f_i(\mathbf{x}) \leq 0, \forall i = 1, \cdots, m$$
$$h_k(\mathbf{x}^*) = 0, \forall k = 1, \cdots, p$$

2. Dual feasibility

$$\lambda_i^* \geq 0, \forall i = 1, \cdots, m$$

3. Dual feasibility

$$\lambda_i^* f_i(\mathbf{x}^*) = 0, \ \forall i = 1, \cdots, m$$

4. Gradient of the Lagrangian must vanish at $\mathbf{x}^*$

$$\nabla f_0(\mathbf{x}^*) + \sum_{i=1}^m \lambda_i^* \nabla f_i(\mathbf{x}^*) + \sum_{k=1}^p \nu_k^* \nabla h_k(\mathbf{x}^*) = 0$$

Remark. For convex problems with differentiable objective and constraint functions satisfying Slater's condition, KKT conditions are both necessary and sufficient conditions.

Suggested reading

Boyd, Stephen P., and Lieven Vandenberghe. Convex optimization. Cambridge university press, 2004.

Thank you!