Optimization - Lecture I

Qamrul Hasan Ansari
Aligarh Muslim University, India
KFUPM, Dhahran, Saudi Arabia
E-mail: qhansari@gmail.com

October 13, 2020

This talk is based on Chapter 2 of the following books:

- Q. H. Ansari, C. S. Lalitha and M. Mehta, *Generalized Convexity, Nonsmooth Variational Inequalities and Nonsmooth Optimization*, Taylor & Francis Group (2014)
- W. Forst and D. Hoffmann, *Optimization – Theory and Practice*, Springer Science & Business Media (2010)

Helped by:
Pradeep, Jauny, Ramsurat, Gourav, Debdas

Table of contents

1. Introduction
2. Convex Sets
3. Hyperplanes
4. Convex Functions
5. Optimality Conditions for Unconstrained Optimization
6. Optimality Conditions for Constrained Optimization I
7. Optimality Conditions for Constrained Optimization II
8. Constraint Qualifications

Mathematical Optimization

Unconstrained Optimization

$$\begin{array}{ll} \text{minimize} & f(x) \\ \text{subject to} & x \in \mathbb{R}^n \end{array}$$

where $f : \mathbb{R}^n \to \mathbb{R}$: *objective function*

- *Optimal solution* $\bar{x}$ has smallest value of f among all vectors in $\mathbb{R}^n$, that is, $f(\bar{x}) \leq f(y)$ for all $y \in \mathbb{R}^n$.

Mathematical Optimization

Constrained Optimization

$$\begin{array}{lll} \text{minimize} & f(x) & \\ \text{subject to} & g_i(x) \leq 0 & \text{for } i = 1, 2, \ldots, m \\ & h_j(x) = 0 & \text{for } j = 1, 2, \ldots, \ell \\ & x \in X & \end{array}$$

- $f : \mathbb{R}^n \to \mathbb{R}$: *objective function*
- $g_1, \ldots, g_m : \mathbb{R}^n \to \mathbb{R}$: *inequality constraint functions*
- $h_1, \ldots h_\ell : \mathbb{R}^n \to \mathbb{R}$: *equality constraint*
- $X \subseteq \mathbb{R}^n$

Feasible Region or Set

$$\mathscr{F} = \{x \in X : g_i(x) \leq 0 \text{ for } i = 1, 2, \ldots, m \text{ and } h_j(x) = 0 \text{ for } j = 1, 2, \ldots, \ell\}$$

A point of $\mathscr{F}$ is called a *feasible point*.

- *Optimal solution* $\bar{x}$ has smallest value of f among all vectors in $\mathscr{F}$, that is, $f(\bar{x}) \leq f(y)$ for all $y \in \mathscr{F}$.

If the objective function f is linear, all constraints functions g_i and h_j are linear and the set X can be represented by linear inequalities and / or linear equations, then the above problem is called a *linear program*, otherwise it called *nonlinear program*.

A *convex optimization problem* is one in which the objective and inequality constraint functions are convex and equality constraint functions are affine.

Feasible Region or Set

$$\mathscr{F} = \{x \in X : g_i(x) \leq 0 \text{ for } i = 1, 2, \ldots, m \text{ and } h_j(x) = 0 \text{ for } j = 1, 2, \ldots, \ell\}$$

A point of $\mathscr{F}$ is called a *feasible point*.

- *Optimal solution* $\bar{x}$ has smallest value of f among all vectors in $\mathscr{F}$, that is, $f(\bar{x}) \leq f(y)$ for all $y \in \mathscr{F}$.

If the objective function f is linear, all constraints functions g_i and h_j are linear and the set X can be represented by linear inequalities and / or linear equations, then the above problem is called a *linear program*, otherwise it called *nonlinear program*.

A *convex optimization problem* is one in which the objective and inequality constraint functions are convex and equality constraint functions are affine.

General Form of a Constrained Minimization Problem

$$\begin{array}{ll} \text{minimize} & f(x) \\ \text{subject to} & x \in X, \end{array} \tag{1}$$

where $f : \mathbb{R}^n \to \mathbb{R}$ and $X \subseteq \mathbb{R}^n$.

Optimal Solution

- $\bar{x} \in X$ is an *(strict) optimal solution* or *global optimal solution* of the minimization problem (1) if $f(\bar{x}) \leq f(x)$ $(f(\bar{x}) < f(x))$ for all $x \in X$.
- $\bar{x} \in X$ is a *(strict) local optimal solution* of the minimization problem (1) if there exists a neighbourhood $N_\varepsilon(\bar{x})$ of $\bar{x}$ such that $f(\bar{x}) \leq f(x)$ $(f(\bar{x}) < f(x))$ for all $x \in X \cap N_\varepsilon(\bar{x})$.

$N_\varepsilon(\bar{x}) = \{x \in \mathbb{R}^n : \|\bar{x} - x\| < \varepsilon\}$

General Form of a Constrained Minimization Problem

$$\begin{array}{ll} \text{minimize} & f(x) \\ \text{subject to} & x \in X, \end{array} \tag{1}$$

where $f : \mathbb{R}^n \to \mathbb{R}$ and $X \subseteq \mathbb{R}^n$.

Optimal Solution

- $\bar{x} \in X$ is an *(strict) optimal solution* or *global optimal solution* of the minimization problem (1) if $f(\bar{x}) \leq f(x)$ $(f(\bar{x}) < f(x))$ for all $x \in X$.
- $\bar{x} \in X$ is a *(strict) local optimal solution* of the minimization problem (1) if there exists a neighbourhood $N_\varepsilon(\bar{x})$ of $\bar{x}$ such that $f(\bar{x}) \leq f(x)$ $(f(\bar{x}) < f(x))$ for all $x \in X \cap N_\varepsilon(\bar{x})$.

$N_\varepsilon(\bar{x}) = \{x \in \mathbb{R}^n : \|\bar{x} - x\| < \varepsilon\}$

Hessian matrix

The *Hessian matrix* of f at $x = (x_1, x_2, \ldots, x_n)$ is given by

$$\nabla^2 f(x) \equiv H(x) = \begin{bmatrix} \dfrac{\partial^2 f(x)}{\partial x_1^2} & \cdots & \dfrac{\partial^2 f(x)}{\partial x_1 \partial x_n} \\ \vdots & & \vdots \\ \dfrac{\partial^2 f(x)}{\partial x_n \partial x_1} & \cdots & \dfrac{\partial^2 f(x)}{\partial x_n^2} \end{bmatrix}.$$

Positive (semi) definite matrix

An $n \times n$ symmetric matrix M of real numbers is said to be *positive semidefinite* if $\langle y, My \rangle \geq 0$ for all $y \in \mathbb{R}^n$.

It is called *positive definite* if $\langle y, My \rangle > 0$ for all $y \neq \mathbf{0}$.

Hessian matrix

The *Hessian matrix* of f at $x = (x_1, x_2, \ldots, x_n)$ is given by

$$\nabla^2 f(x) \equiv H(x) = \begin{bmatrix} \dfrac{\partial^2 f(x)}{\partial x_1^2} & \cdots & \dfrac{\partial^2 f(x)}{\partial x_1 \partial x_n} \\ \vdots & & \vdots \\ \dfrac{\partial^2 f(x)}{\partial x_n \partial x_1} & \cdots & \dfrac{\partial^2 f(x)}{\partial x_n^2} \end{bmatrix}.$$

Positive (semi) definite matrix

An $n \times n$ symmetric matrix M of real numbers is said to be *positive semidefinite* if $\langle y, My \rangle \geq 0$ for all $y \in \mathbb{R}^n$.
It is called *positive definite* if $\langle y, My \rangle > 0$ for all $y \neq \mathbf{0}$.

PART - I

Convex Sets

Definition

A subset K of $\mathbb{R}^n$ is said to be a *convex set* if for all $x, y \in K$ and $\lambda, \mu \geq 0$ such that $\lambda + \mu = 1$, we have $\lambda x + \mu y \in K$, that is, for all $x, y \in K$ and $\lambda \in [0, 1]$, we have $\lambda x + (1 - \lambda) y \in K$.

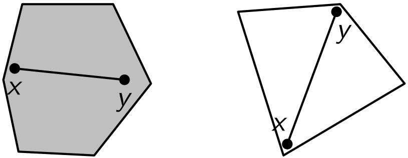


Figure: Convex sets

Nonconvex Sets

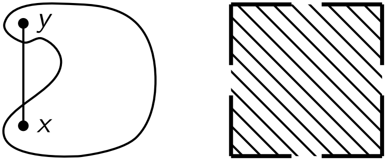


Figure: Nonconvex sets

Cone and Convex Cone

Definition

A subset C of $\mathbb{R}^n$ is said to be a *cone* if for all $x \in C$ and $\lambda \geq 0$, we have $\lambda x \in C$.

A subset C of $\mathbb{R}^n$ is said to be a *convex cone* if it is convex and a cone, that is, for all $x, y \in C$ and $\lambda, \mu \geq 0$, $\lambda x + \mu y \in C$.

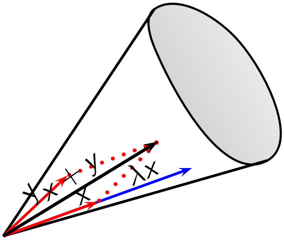


Figure: Convex cone

Cone but not convex

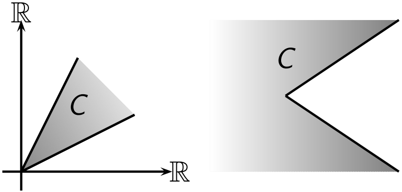
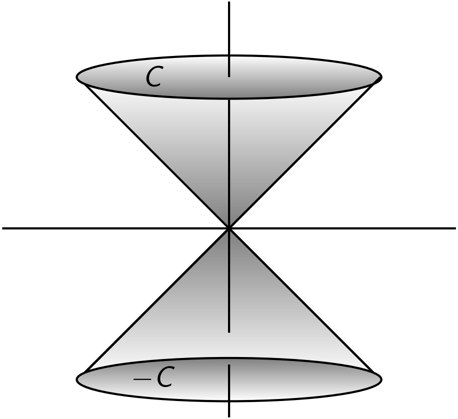


Figure: Cone with vertex at origin   Cone but not convex

Pointed Cone

Definition

A cone C is said to be *pointed* if $C \cap (-C) = \{\mathbf{0}\}$.



Dual cone

Definition

Let C be a closed convex pointed cone in $\mathbb{R}^n$. The *dual cone* C^* of C is defined by

$$C^* = \{y \in \mathbb{R}^n : \langle y, x \rangle \geq 0 \text{ for all } x \in C\}.$$

Geometrically, C^* consists of all those vectors which make a nonobtuse angle with every vector in C.

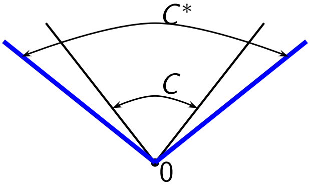


Figure: A cone and its dual

Hyperplanes

Definition

Given a nonzero vector $a = (a_1, a_2, \ldots, a_n) \in \mathbb{R}^n$ and $\lambda \in \mathbb{R}$, the set $H = \{x = (x_1, x_2, \ldots, x_n) \in \mathbb{R}^n : \langle a, x \rangle = a_1x_1 + a_2x_2 + \cdots + a_nx_n = \lambda\}$ is called a *hyperplane* in $\mathbb{R}^n$. The vector a is called a *normal* to the hyperplane H.

Every other normal to H is either a positive or a negative scalar multiple of a.

An $(n-1)$-dimensional affine set in $\mathbb{R}^n$ is a hyperplane.

In $\mathbb{R}^2$ the hyperplanes are the straight lines and in $\mathbb{R}^3$ such sets are planes.

A good interpretation of this is that every hyperplane has “two sides”, like one’s picture of a line in $\mathbb{R}^2$ or a plane in $\mathbb{R}^3$.

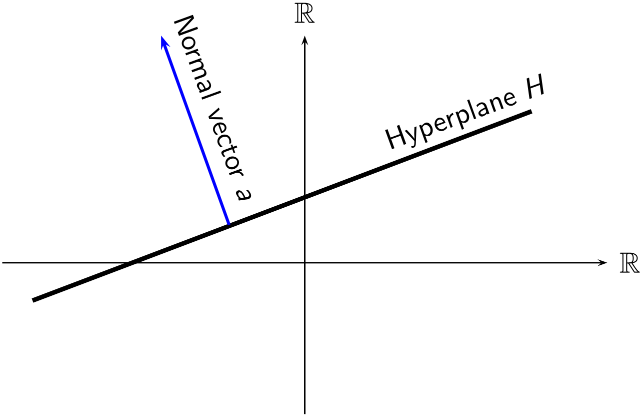


Figure: A hyperplane and a normal vector

Definition

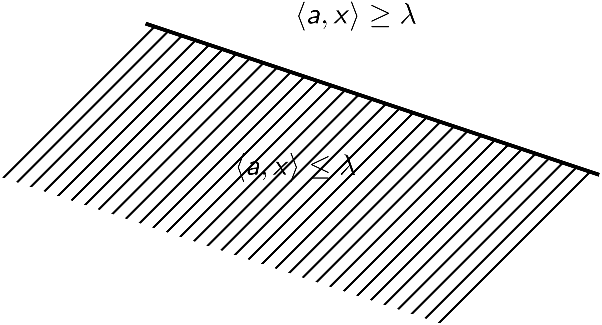
Let a be a nonzero vector in $\mathbb{R}^n$ and λ be a scalar.

(a) $H^{++} = \{x \in \mathbb{R}^n : \langle a, x \rangle > \lambda\}$ is called *upper open half-space*.

(b) $H^{--} = \{x \in \mathbb{R}^n : \langle a, x \rangle < \lambda\}$ is called *lower open half-space*.

(c) $H^{+} = \{x \in \mathbb{R}^n : \langle a, x \rangle \geq \lambda\}$ is called *upper closed half-space*.

(d) $H^{-} = \{x \in \mathbb{R}^n : \langle a, x \rangle \leq \lambda\}$ is called *lower closed half-space*.



(a) It is clear that the half-spaces are convex sets.

(b) The half-spaces and hyperplane partition $\mathbb{R}^n$ into disjoint sets as

$$\mathbb{R}^n = H^{--} \cup H \cup H^{++}.$$

Also,

$$\mathbb{R}^n = H^{-} \cup H^{+}.$$

(c) Note that any point in $\mathbb{R}^n$ lies in H^+, in H^- or in both.

Supporting hyperplane at a point

Let S be a nonempty set in $\mathbb{R}^n$ and $\bar{x} \in \mathrm{b}(S)$. A hyperplane $H = \{x \in \mathbb{R}^n : \langle a, x - \bar{x} \rangle = 0,\ a \neq \mathbf{0},\ a \in \mathbb{R}^n\}$ is called a *supporting hyperplane to S at $\bar{x}$* if either

$$S \subseteq H^+, \quad \text{that is,} \quad \langle a, x - \bar{x} \rangle \geq 0, \ \text{ for all } x \in S,$$

or

$$S \subseteq H^-, \quad \text{that is,} \quad \langle a, x - \bar{x} \rangle \leq 0, \ \text{ for all } x \in S.$$

H is called a *proper supporting hyperplane* to S at $\bar{x}$ if, in addition to the above said properties, it satisfies $S \not\subseteq H$.

Above definition can be stated equivalently as follows: The hyperplane $H = \{x \in \mathbb{R}^n : \langle a, x - \bar{x} \rangle = 0\}$ is a supporting hyperplane of S at $\bar{x} \in \mathrm{b}(S)$ if

$$\langle a, \bar{x} \rangle = \inf\{\langle a, x \rangle : x \in S\}, \quad \text{or else} \quad \langle a, \bar{x} \rangle = \sup\{\langle a, x \rangle : x \in S\}.$$

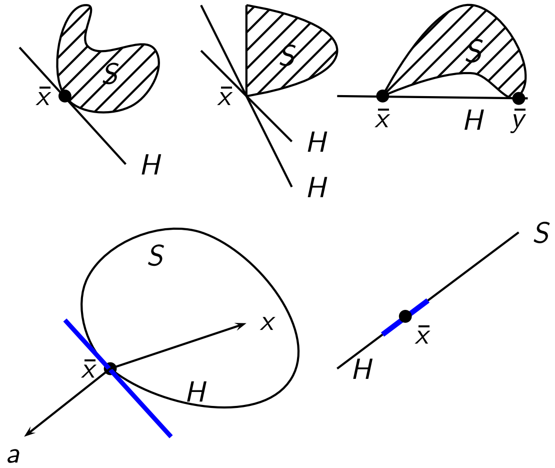


Figure: Supporting hyperplanes

Supporting hyperplane

Definition

Let S be a nonempty subset of $\mathbb{R}^n$. A hyperplane $H = \{x \in \mathbb{R}^n : \langle a, x\rangle = \lambda,\ a \neq \mathbf{0},\ a \in \mathbb{R}^n,\ \lambda \in \mathbb{R}\}$ is called a *supporting hyperplane of S* if either $S \subseteq H^+$ (or $S \subseteq H^-$) and $\mathrm{cl}(S) \cap H \neq \emptyset$.

H is called a *proper supporting hyperplane* of S if, in addition to the said properties, we have $\mathrm{cl}(S) \cap H \neq S$.

If $\bar{x} \in \mathrm{cl}(S) \cap H$, then H supports the set S at $\bar{x}$. It is obvious that a hyperplane H may support a set S at several distinct points.

The following result says that $\inf\{\langle a,x\rangle : x \in S\} = \lambda$ if the hyperplane $H = \{x \in \mathbb{R}^n : \langle a,x\rangle = \lambda,\ a \neq \mathbf{0},\ a \in \mathbb{R}^n,\ \lambda \in \mathbb{R}\}$ supports S and S is contained in H^+.

Theorem

Let $H = \{x \in \mathbb{R}^n : \langle a,x\rangle = \lambda,\ a \neq \mathbf{0},\ a \in \mathbb{R}^n,\ \lambda \in \mathbb{R}\}$ be a supporting hyperplane of a nonempty set S in $\mathbb{R}^n$ such that $S \subseteq H^+$. Then, $\inf\{\langle a,x\rangle : x \in S\} = \lambda$.

Corollary

If $H = \{x \in \mathbb{R}^n : \langle a,x\rangle = \lambda,\ a \neq \mathbf{0},\ a \in \mathbb{R}^n,\ \lambda \in \mathbb{R}\}$ is a supporting hyperplane for S such that $S \subseteq H^-$, then $\sup\{\langle a,x\rangle : x \in S\} = \lambda$.

Definition

Let K be a nonempty closed subset of $\mathbb{R}^n$ and $y \in \mathbb{R}^n$. A point $\bar{x} \in K$ is said to be the *projection of y on K* or *best approximation of y on K*, denoted by $\bar{x} = P_K(y)$, if

$$\|y - \bar{x}\| = \min_{x \in K} \|y - x\|.$$

Remark

If $y \in K$, then the projection is unique and $\bar{x} = y$. We note that the projection of y on K may not always exist (for example, if K is open) and when it exists it may not be unique;
For example, if $K = \{x \in \mathbb{R}^2 : \|x\| \geq 1\}$ and y is the origin.
However, under closedness and convexity assumptions the following assertion holds.

Theorem

Let K be a nonempty closed convex subset of $\mathbb{R}^n$ and y a point in $\mathbb{R}^n$ with $y \notin K$. Then, there exists a unique point $\bar{x} \in K$ such that

$$\|y - \bar{x}\| = \min_{x \in K} \|y - x\|. \tag{2}$$

Also, the unique point $\bar{x}$ satisfies the following inequality:

$$\langle y - \bar{x}, x - \bar{x} \rangle \leq 0, \quad \textit{for all } x \in K. \tag{3}$$

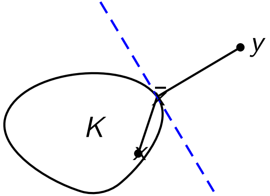


Figure: The projection of a point y onto K

Remark

The inequality (3) shows that $y - \bar{x}$ and $x - \bar{x}$ subtend a nonacute angle between them. The projection $P_K(y)$ of y on K can be interpreted as the result of applying to y the operator $P_K : \mathbb{R}^n \to K$, which is called *projection operator*. Note that $P_K(x) = x$ for all $x \in K$.

Corollary

Let K be a nonempty closed convex subset of $\mathbb{R}^n$. Then, for all $x, y \in \mathbb{R}^n$,

$$\|P_K(x) - P_K(y)\| \leq \|x - y\|, \tag{4}$$

that is, the projection operator P_K is nonexpansive. In particular, P_K is continuous on K.

The geometric interpretation of the nonexpansivity of P_K is given in the following figure. We observe that if strict inequality holds in (4), then the projection operator P_K reduces the distance. However, if the equality holds in (4) then the distance is conserved.

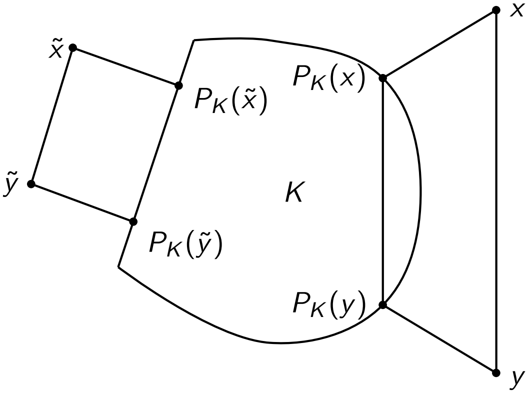


Figure: The nonexpansiveness of the projection operator

Remark

If K is a closed convex set and y is a point outside the set K, then we can find a hyperplane such that the convex set K lies in one of the half-spaces generated by the hyperplane and the point y lies in other open half-space.

Theorem (Separating hyperplane theorem)

Let K be a nonempty closed convex subset of $\mathbb{R}^n$ and $y \notin K$. Then there exists a nonzero vector $a \in \mathbb{R}^n$ such that $\inf_{x \in K} \langle a, x \rangle > \langle a, y \rangle$.

In other words, there exist a nonzero vector $a \in \mathbb{R}^n$ and $\alpha \in \mathbb{R}$ such that $\langle a, x \rangle \geq \alpha > \langle a, y \rangle$ for all $x \in K$,

that is, the hyperplane $H := \{x \in \mathbb{R}^n : \langle a, x \rangle = \alpha\}$ strictly separates K and y.

Remark

If K is a closed convex set and y is a point outside the set K, then we can find a hyperplane such that the convex set K lies in one of the half-spaces generated by the hyperplane and the point y lies in other open half-space.

Theorem (Separating hyperplane theorem)

Let K be a nonempty closed convex subset of $\mathbb{R}^n$ and $y \notin K$. Then there exists a nonzero vector $a \in \mathbb{R}^n$ such that $\inf_{x \in K} \langle a, x \rangle > \langle a, y \rangle$.

In other words, there exist a nonzero vector $a \in \mathbb{R}^n$ and $\alpha \in \mathbb{R}$ such that $\langle a, x \rangle \geq \alpha > \langle a, y \rangle$ for all $x \in K$,

that is, the hyperplane $H := \{x \in \mathbb{R}^n : \langle a, x \rangle = \alpha\}$ strictly separates K and y.

Remark

The result may not be valid if K is nonconvex or if K is not closed. For instances, consider the following examples.

(a) Take $K = \{(x_1, x_2) \in \mathbb{R}^2 : x_1^2 + x_2^2 < 1\}$ and $y = (1, 0) \in \mathbb{R}^2$. Note that there exists no line in $\mathbb{R}^2$ which can strictly separate y and K. The issue with this K is that it is not a closed set.

(b) Take $K = \{(x_1, x_2) \in \mathbb{R}^2 : x_2 \leq x_1^2\}$ and $y = (0, 1) \in \mathbb{R}^2$. Note that there exists no line in $\mathbb{R}^2$ which can separate y and K. The issue with this K is that it is a nonconvex set.

The following theorem is a consequence of the previous result and may be called "theorem of the supporting hyperplane".

Theorem

If y is a point on the boundary of a nonempty convex subset K of $\mathbb{R}^n$, then there exists a nonzero vector $a \in \mathbb{R}^n$ such that $\inf_{x \in K} \langle a, x \rangle = \langle a, y \rangle$, that is, there exists a hyperplane $H = \{x \in \mathbb{R}^n : \langle a, x - y \rangle = 0\}$ which supports K at y.

Remark

The above result says that at each boundary point of a convex set there is a supporting hyperplane passing through it.

Farkas theorem of alternatives

Theorem (Farkas' theorem of alternatives)

Let A be an $m \times n$ matrix and $b \in \mathbb{R}^m$. Then, exactly one of the following two systems has a solution:

- $\exists\ x \in \mathbb{R}_+$ *such that* $Ax = b$.
- $\exists\ y \in \mathbb{R}^m$ *such that* $A^\top y \geq 0$ *and* $\langle b, y \rangle < 0$.

Geometric illustration of Farkas' theorem

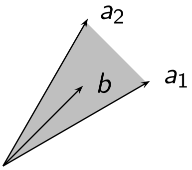
Assume that A is an $m \times n$ matrix with columns a_i.

First Alternative

$b = Ax = \sum_{i=1}^{n} a_i x_i$,

$x_i \geq 0, \ i = 1, 2, \ldots, n$

b is in the cone generated by the columns of A

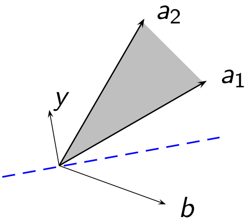


Second Alternative

$y^\top a_i \geq 0, \ i = 1, 2, \ldots, m$

$y^\top b < 0$

the hyperplane $\langle y, z \rangle$ separates b from $a_1, \ldots, a_m$



Convex Functions

Definition

Let K be a nonempty convex subset of $\mathbb{R}^n$. A function $f : K \to \mathbb{R}$ is said to be

(a) *convex* if for all $x, y \in K$ and all $\lambda \in [0, 1]$,

$$f(\lambda x + (1 - \lambda)y) \leq \lambda f(x) + (1 - \lambda)f(y);$$

(b) *strictly convex* if for all $x, y \in K$, $x \neq y$ and all $\lambda \in \,]0, 1[$,

$$f(\lambda x + (1 - \lambda)y) < \lambda f(x) + (1 - \lambda)f(y).$$

A function $f : K \to \mathbb{R}$ is said to be (*strictly*) *concave* if $-f$ is (strictly) convex.

Geometrically speaking, a function $f : K \to \mathbb{R}$ defined on a convex subset K of $\mathbb{R}^n$ is convex if the line segment joining any two points on the graph of the function lies on or above the portion of the graph between these points. Similarly, f is concave if the line segment joining any two points on the graph of the function lies on or below the portion of the graph between these points. Also, a function for which the line segment joining any two points on the graph of the function lies strictly above the portion of the graph between these points is referred to as strictly convex function.

Some of the examples of convex functions defined on $\mathbb{R}$ are $f(x) = e^x$, $f(x) = x$, $f(x) = |x|$, $f(x) = \max\{0, x\}$. The functions $f(x) = -\log x$ and $f(x) = x^\alpha$ for $\alpha < 0$, $\alpha > 1$ are strictly convex defined on the interval $]0, \infty[$. Clearly, every strictly convex function is convex but the converse may not be true. For example, the function $f(x) = x$ defined on $\mathbb{R}$ is not strictly convex. The function $f(x) = |x + x^3|$ is a nondifferentiable strictly convex function on $\mathbb{R}$.

Properties of convex functions

Theorem

Let $K \subset \mathbb{R}^n$ be a nonempty convex set and $f : K \to \mathbb{R}$ be a convex function. Then, the following assertions hold.

(a) *f is continuous in $int(K)$.*

(b) *f is locally Lipschitz on $int(K)$.*

(c) *For any point $\bar{x} \in int(K)$ and any direction $d \in \mathbb{R}$, the directional derivative $f'(\bar{x}; d)$ of f exists at $\bar{x}$ along d.*

The directional derivative $f'(\bar{x}; d)$ of f exists at $\bar{x}$ along d is

$$f'(x; d) = \lim_{t \to 0} \frac{f(x + td) - f(x)}{t}.$$

Properties of convex functions

Theorem

Every local minimum of a (strictly) convex function $f : K \to \mathbb{R}$ defined on a convex set $K \subseteq \mathbb{R}^n$ is a (unique) global minimum of f over K. Further, the set of points at which a convex function attains its global minimum on K is a convex set.

Theorem

Let K be a nonempty open convex subset of $\mathbb{R}^n$ and $f : K \to \mathbb{R}$ be a differentiable (strictly) convex function. If $\nabla f(x) = \mathbf{0}$, then x is a (unique) global minimum of f over K; that is, every stationary point of f is a (unique) global minimum of f over K.

Properties of convex functions

Theorem

Every local minimum of a (strictly) convex function $f : K \to \mathbb{R}$ defined on a convex set $K \subseteq \mathbb{R}^n$ is a (unique) global minimum of f over K. Further, the set of points at which a convex function attains its global minimum on K is a convex set.

Theorem

Let K be a nonempty open convex subset of $\mathbb{R}^n$ and $f : K \to \mathbb{R}$ be a differentiable (strictly) convex function. If $\nabla f(x) = \mathbf{0}$, then x is a (unique) global minimum of f over K; that is, every stationary point of f is a (unique) global minimum of f over K.

We observe here that a stationary point of a function can be a global minimum without the function being convex. For example, $x = 0$ is a stationary point of the nonconvex function $f(x) = 1 - e^{-x^2}$ defined on $\mathbb{R}$.

Properties of convex functions

Theorem

Every local minimum of a (strictly) convex function $f : K \to \mathbb{R}$ defined on a convex set $K \subseteq \mathbb{R}^n$ is a (unique) global minimum of f over K. Further, the set of points at which a convex function attains its global minimum on K is a convex set.

Theorem

Let K be a nonempty open convex subset of $\mathbb{R}^n$ and $f : K \to \mathbb{R}$ be a differentiable (strictly) convex function. If $\nabla f(x) = \mathbf{0}$, then x is a (unique) global minimum of f over K; that is, every stationary point of f is a (unique) global minimum of f over K.

We observe here that a stationary point of a function can be a global minimum without the function being convex. For example, $x = 0$ is a stationary point of the nonconvex function $f(x) = 1 - e^{-x^2}$ defined on $\mathbb{R}$.

Characterization for convex functions

Proposition

A function $f : K \to \mathbb{R}$ defined on a nonempty convex subset K of $\mathbb{R}^n$ is convex if and only its epigraph $\mathrm{epi}(f) := \{(x, \alpha) \in K \times \mathbb{R} : f(x) \leq \alpha\}$ is a convex set.

If f is a convex function defined on K, then for every $\alpha \in \mathbb{R}$ the lower level set $L(f, \alpha) = \{x \in K : f(x) \leq \alpha\}$ is also convex. However, the converse is not true. For example, the function $f(x) = x^3$ defined on $\mathbb{R}$ is not convex, but the lower level set $L(f, \alpha) = \{x \in \mathbb{R} : x \leq \alpha^{1/3}\}$ is convex for all $\alpha \in \mathbb{R}$.

Characterization for convex functions

Proposition

A function $f : K \to \mathbb{R}$ defined on a nonempty convex subset K of $\mathbb{R}^n$ is convex if and only its epigraph $\text{epi}(f) := \{(x, \alpha) \in K \times \mathbb{R} : f(x) \leq \alpha\}$ is a convex set.

If f is a convex function defined on K, then for every $\alpha \in \mathbb{R}$ the lower level set $L(f, \alpha) = \{x \in K : f(x) \leq \alpha\}$ is also convex. However, the converse is not true. For example, the function $f(x) = x^3$ defined on $\mathbb{R}$ is not convex, but the lower level set $L(f, \alpha) = \{x \in \mathbb{R} : x \leq \alpha^{1/3}\}$ is convex for all $\alpha \in \mathbb{R}$.

Characterization for convex functions

Theorem

Let K be a nonempty open convex subset of $\mathbb{R}^n$ and $f : K \to \mathbb{R}$ be a function. Then, f is convex if and only if for every $x \in K$, there exists a vector $\xi \in \mathbb{R}^n$ such that

$$\langle \xi, y - x \rangle \leq f(y) - f(x), \quad \text{for all } y \in K. \tag{5}$$

Corollary

Let K be a nonempty open convex subset of $\mathbb{R}^n$ and $f : K \to \mathbb{R}$ be a function. Then, f is strictly convex if and only if for every $x \in K$, there exists a vector $\xi \in \mathbb{R}^n$ such that

$$\langle \xi, y - x \rangle < f(y) - f(x), \quad \text{for all } y \in K, \ y \neq x.$$

Remark

If in Theorem 20, the set K is not open then the sufficient part of this theorem may not be true. That is, if for every $x \in K$, there exists a vector $\xi \in \mathbb{R}^n$ such that

$$\langle \xi, y - x \rangle \leq f(y) - f(x), \quad \text{for all } y \in K,$$

then f is not necessarily convex.
Indeed, consider the set $K = \{(x_1, x_2) \in \mathbb{R}^2 : 0 \leq x_1 \leq 1,\ 0 \leq x_2 \leq 1\}$ and the function $f : K \to \mathbb{R}$ defined by

$$f(x_1, x_2) = \begin{cases} 0, & \text{if } 0 \leq x_1 \leq 1,\ 0 < x_2 \leq 1, \\ \frac{1}{4} - \left(x_1 - \frac{1}{2}\right)^2, & \text{if } 0 \leq x_1 \leq 1,\ x_2 = 0. \end{cases}$$

Then, for each point in the interior of K, $\xi = \mathbf{0}$ satisfies the inequality (5). However, f is not convex on K since $\operatorname{epi}(f)$ is clearly not a convex set.

Characterization for differentiable convex functions

Theorem

Let K be a nonempty open convex subset of $\mathbb{R}^n$ and $f : K \to \mathbb{R}$ be a differentiable function. Then,

(a) *f is convex if and only if for all $x, y \in K$,*

$$\langle \nabla f(x), y - x \rangle \leq f(y) - f(x). \tag{6}$$

(b) *f is strictly convex if and only if the inequality is strict in (6) for $x \neq y$.*

Characterizations for differentiable convex functions

Theorem

Let K be a nonempty open convex subset of $\mathbb{R}^n$ and $f : K \to \mathbb{R}$ be a differentiable function. Then, f is convex (strictly convex) if and only if for all $x, y \in K$ $(x \neq y)$,

$$\langle \nabla f(y) - \nabla f(x), y - x \rangle \geq 0 \, (> 0).$$

The following example illustrates the above theorem.

Example

The function $f(x) = x_1^2 + x_2^2$, where $x = (x_1, x_2) \in \mathbb{R}^2$ is a convex function on $\mathbb{R}^2$ and $\nabla f(x) = 2(x_1, x_2)$. For $x, y \in \mathbb{R}^2$,

$$\begin{aligned}
\langle \nabla f(y) - \nabla f(x), y - x \rangle &= \langle 2(y_1 - x_1, y_2 - x_2), (y_1 - x_1, y_2 - x_2) \rangle \\
&= 2(y_1 - x_1)^2 + 2(y_2 - x_2)^2 \geq 0.
\end{aligned}$$

Characterization for twice differentiable convex functions

The next theorem provides a characterization for a twice differentiable convex function in terms of its Hessian matrix. However, for a strictly convex function we have only one way implication.

Theorem

Let K be a nonempty open convex subset of $\mathbb{R}^n$.

(a) *A twice differentiable function $f : K \to \mathbb{R}$ is convex if and only if its Hessian matrix $H(x) = \nabla^2 f(x)$ is positive semidefinite for every $x \in K$.*

(b) *If the Hessian matrix $H(x)$ of a twice differentiable function $f : K \to \mathbb{R}$ is positive definite for every $x \in K$, then f is strictly convex.*

Characterization for twice differentiable convex functions

The next theorem provides a characterization for a twice differentiable convex function in terms of its Hessian matrix. However, for a strictly convex function we have only one way implication.

Theorem

Let K be a nonempty open convex subset of $\mathbb{R}^n$.

(a) *A twice differentiable function $f : K \to \mathbb{R}$ is convex if and only if its Hessian matrix $H(x) = \nabla^2 f(x)$ is positive semidefinite for every $x \in K$.*

(b) *If the Hessian matrix $H(x)$ of a twice differentiable function $f : K \to \mathbb{R}$ is positive definite for every $x \in K$, then f is strictly convex.*

Necessary optimality condition for unconstrained problems

Consider the optimization problem $\min_{x \in \mathbb{R}^n} f(x)$. Let $\bar{x}$ be a local minimum of f. Then, the following assertions hold.

(a) If f is differentiable at $\bar{x}$, then $\nabla f(\bar{x}) = 0$.

(b) If f is twice continuously differentiable in the neighborhood of $\bar{x}$, then the Hessian $\nabla^2 f(\bar{x})$ is positive semidefinite.

Sufficient optimality condition for unconstrained problems

Suppose that the function f is twice continuously differentiable in the neighborhood of $\bar{x}$. If $\nabla f(\bar{x}) = 0$ and $\nabla^2 f(\bar{x})$ is positive definite, then f has a strict local minimum at $\bar{x}$.

Necessary optimality condition for unconstrained problems

Consider the optimization problem $\min_{x \in \mathbb{R}^n} \mathrm{f(x)}$. Let $\bar{x}$ be a local minimum of f. Then, the following assertions hold.

a. If f is differentiable at $\bar{x}$, then $\nabla f(\bar{x}) = 0$.
b. If f is twice continuously differentiable in the neighborhood of $\bar{x}$, then the Hessian $\nabla^2 f(\bar{x})$ is positive semidefinite.

Sufficient optimality condition for unconstrained problems

Suppose that the function f is twice continuously differentiable in the neighborhood of $\bar{x}$. If $\nabla f(\bar{x}) = 0$ and $\nabla^2 f(\bar{x})$ is positive definite, then f has a strict local minimum at $\bar{x}$.

We say that $\alpha = \inf_{x \in X} f(x)$ if α is the greatest lower bound of f on X, that is, $\alpha \leq f(x)$ for all $x \in X$ and there is no $\beta > \alpha$ such that $\beta \leq f(x)$ for all $x \in X$.

Lower Semicontinuous Function

A function $f : X \subseteq \mathbb{R}^n \to \mathbb{R}$ is said to be *lower semicontinuous* at a point $x \in X$ if $f(x) \leq \liminf_{n \to \infty} f(x_n)$ whenever $x_n \to x$ as $n \to \infty$. f is said to be *lower semicontinuous on* X if it is lower semicontinuous at each point of X.

We say that $\alpha = \inf_{x \in X} f(x)$ if α is the greatest lower bound of f on X, that is, $\alpha \leq f(x)$ for all $x \in X$ and there is no $\beta > \alpha$ such that $\beta \leq f(x)$ for all $x \in X$.

Lower Semicontinuous Function

A function $f : X \subseteq \mathbb{R}^n \to \mathbb{R}$ is said to be *lower semicontinuous* at a point $x \in X$ if $f(x) \leq \liminf_{n \to \infty} f(x_n)$ whenever $x_n \to x$ as $n \to \infty$. f is said to be *lower semicontinuous on* X if it is lower semicontinuous at each point of X.

Weierstrass's Theorem

Let X be a nonempty compact subset of $\mathbb{R}^n$ and $f : X \to \mathbb{R}$ be a lower semicontinuous function. Then the minimization problem

$$\min\{f(x) : x \in X\}$$

attaints its minimum, that is, there exists a $\bar{x} \in X$ such that $f(\bar{x}) \leq f(x)$ for all $x \in X$.

We say that $\alpha = \inf_{x \in X} f(x)$ if α is the greatest lower bound of f on X, that is, $\alpha \leq f(x)$ for all $x \in X$ and there is no $\beta > \alpha$ such that $\beta \leq f(x)$ for all $x \in X$.

Lower Semicontinuous Function

A function $f : X \subseteq \mathbb{R}^n \to \mathbb{R}$ is said to be *lower semicontinuous* at a point $x \in X$ if $f(x) \leq \liminf_{n \to \infty} f(x_n)$ whenever $x_n \to x$ as $n \to \infty$. f is said to be *lower semicontinuous on* X if it is lower semicontinuous at each point of X.

Weierstrass's Theorem

Let X be a nonempty compact subset of $\mathbb{R}^n$ and $f : X \to \mathbb{R}$ be a lower semicontinuous function. Then the minimization problem

$$\min\{f(x) : x \in X\}$$

attaints its minimum, that is, there exists a $\bar{x} \in X$ such that $f(\bar{x}) \leq f(x)$ for all $x \in X$.

Sufficient & Sufficient optimality condition for constrained problems

Let K be a nonempty convex subset of $\mathbb{R}^n$ and $f : K \to \mathbb{R}$ be a convex differentiable function. A point $\bar{x} \in K$ is a solution of the constrained minimization problem (1) if and only if

$$\langle \nabla f(\bar{x}), y - \bar{x} \rangle \geq 0, \quad \forall y \in K. \tag{7}$$

Roughly speaking, the *variational inequality* (7) states that the vector $\nabla f(\bar{x})$ must be at a non-obtuse angle with all the feasible vectors emanating from $\bar{x}$. In other words, the vector $\bar{x}$ is a solution of the variational inequality (7) if and only if $\nabla f(\bar{x})$ forms a non-obtuse angle with every vector of the form $y - \bar{x}$, for all $y \in K$.

Sufficient & Sufficient optimality condition for constrained problems

Let K be a nonempty convex subset of $\mathbb{R}^n$ and $f : K \to \mathbb{R}$ be a convex differentiable function. A point $\bar{x} \in K$ is a solution of the constrained minimization problem (1) if and only if

$$\langle \nabla f(\bar{x}), y - \bar{x} \rangle \geq 0, \quad \forall y \in K. \tag{7}$$

Roughly speaking, the *variational inequality* (7) states that the vector $\nabla f(\bar{x})$ must be at a non-obtuse angle with all the feasible vectors emanating from $\bar{x}$. In other words, the vector $\bar{x}$ is a solution of the variational inequality (7) if and only if $\nabla f(\bar{x})$ forms a non-obtuse angle with every vector of the form $y - \bar{x}$, for all $y \in K$.

Normal Cone

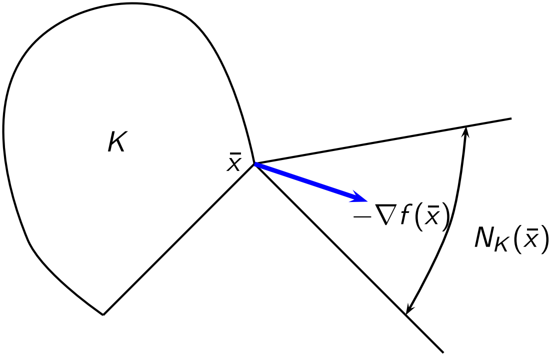
The *normal cone* to K at the point $\bar{x}$, denoted by $N_K(\bar{x})$, is defined by

$$N_K(x) = \begin{cases} \{d \in \mathbb{R}^n : \langle d, y - x \rangle \leq 0 \text{ for all } y \in K\}, & \text{if } x \in K \\ \emptyset, & \text{otherwise.} \end{cases}$$

Geometrically, a vector $\bar{x}$ is a solution of variational inequality if and only if $-\nabla f(\bar{x}) \in N_K(\bar{x})$, where Clearly, $\bar{x} \in K$ is a solution of VIP if and only if

$$\mathbf{0} \in \nabla f(\bar{x}) + N_K(\bar{x}). \tag{8}$$

The inclusion (8) is called a *generalized equation*.



Proposition

Let K be a nonempty closed convex subset of $\mathbb{R}^n$ and $f : K \to \mathbb{R}$ be a convex differentiable function. An element $\bar{x} \in K$ is a solution of the constrained minimization problem (1) if and only if for any $\gamma > 0$, $\bar{x}$ is a fixed point of the mapping $P_K(I - \gamma \nabla f) : K \to K$, that is, $\bar{x} = P_K(\bar{x} - \gamma \nabla f(\bar{x}))$, where $P_K(\bar{x} - \gamma \nabla f(\bar{x}))$ denotes the projection of $\bar{x} - \gamma \nabla f(\bar{x})$ onto K.

This proposition suggests the following projection method for solving constrained minimization problem (1).

Projection Method

For a given $x_0 \in \mathbb{R}^n$, compute x_{m+1} by the rule:

$$x_{m+1} = P_K(x_m - \gamma \nabla f(x_m)), \quad m = 0, 1, 2, \ldots, \tag{9}$$

where P_K denotes the projection operator and $\gamma > 0$ is a constant.

Strongly convex function

A function $f : K \to \mathbb{R}$ defined on a nonempty convex subset K of $\mathbb{R}^n$ is said to be *strongly convex with modulus σ* if there exists a real number $\rho > 0$ such that for all $x, y \in K$ and $\lambda \in [0, 1]$, we have

$$f((1-\lambda)x + \lambda y) + \sigma\lambda(1-\lambda)\|y - x\|^2 \leq (1-\lambda)f(x) + \lambda f(y). \quad (10)$$

Theorem

Let K be a nonempty closed convex subset of $\mathbb{R}^n$ and $f : K \to \mathbb{R}^n$ be strongly convex with constant $\sigma > 0$ and Lipschitz continuous with constant $k > 0$. If $\{x_m\}$ is a sequence generated by the iteration (9) with $\gamma \in \,]0, \sigma/k^2[$, then it converges to a unique solution $\bar{x}$ of the constrained minimization problem (1).

Strongly convex function

A function $f : K \to \mathbb{R}$ defined on a nonempty convex subset K of $\mathbb{R}^n$ is said to be *strongly convex with modulus σ* if there exists a real number $\rho > 0$ such that for all $x, y \in K$ and $\lambda \in [0, 1]$, we have

$$f((1-\lambda)x + \lambda y) + \sigma\lambda(1-\lambda)\|y - x\|^2 \leq (1-\lambda)f(x) + \lambda f(y). \tag{10}$$

Theorem

Let K be a nonempty closed convex subset of $\mathbb{R}^n$ and $f : K \to \mathbb{R}^n$ be strongly convex with constant $\sigma > 0$ and Lipschitz continuous with constant $k > 0$. If $\{x_m\}$ is a sequence generated by the iteration (9) with $\gamma \in \ ]0, \sigma/k^2[$, then it converges to a unique solution $\bar{x}$ of the constrained minimization problem (1).

PART - II

Constrained Optimization

In this section, we study first order (that uses gradient) and second order (that uses Hessian) optimality conditions for the following constrained optimization problem:

$$\begin{cases} \text{minimize} & f(x) \\ \text{subject to} & g_i(x) \leq 0,\ i = 1, 2, \dots, m \\ & h_j(x) = 0,\ j = 1, 2, \dots, \ell, \\ & x \in X\,(\neq \emptyset) \subseteq \mathbb{R}^n, \end{cases} \tag{11}$$

where f and all of g_i's and h_j's are continuously differentiable on X.

$$\mathscr{F} = \{x \in X : g_i(x) \leq 0 \text{ for } i = 1, 2, \dots, m \text{ and } h_i(x) = 0 \text{ for } j = 1, 2, \dots, \ell\}$$

Given a feasible point x_0, a constraint $g_i(x) \leq 0$ is called

- *active* if $g_i(x_0) = 0$
- *inactive* if $g_i(x_0) < 0$.

(An equality constraint $h_j(x) = 0$ is active at any feasible point.)

For a feasible point x_0, we set

$$\mathscr{A}(x_0) := \{i \in \{1, 2, \ldots, m\} : g_i(x_0) = 0\},$$

which describes the inequality restrictions that are active at x_0.

The active constraints have a special significance: if a constraint is inactive ($g_i(x_0) < 0$) at the feasible point x_0, it is possible to move from x_0 a bit in any direction without violating this constraint.

Cone of feasible direction

Definition

Let $d \in \mathbb{R}^n$ and $x_0 \in \mathscr{F}$. A vector d is called the *feasible direction* of $\mathscr{F}$ at x_0 if

$$\exists\ \delta > 0 \text{ such that } x_0 + \tau d \in \mathscr{F}, \quad \forall \tau \in [0, \delta].$$

(That is, a 'small' movement from x_0 along such a direction gives feasible points.)

Set of all feasible directions of $\mathscr{F}$ at x_0 constitutes a cone, which is denoted by $\mathcal{C}_{fd}(x_0)$ and it is called *cone of feasible directions*. That is,

$$\mathcal{C}_{fd}(x_0) = \{d \in \mathbb{R}^n : \exists\ \delta > 0 \text{ such that } x_0 + \tau d \in \mathscr{F}, \forall \tau \in [0, \delta]\}. \quad (12)$$

The cone $\mathcal{C}_{fd}(x_0)$ is also referred as *radial cone* at x_0.

Let d be a feasible direction of $\mathscr{F}$ at x_0. Then, for an $i \in \mathscr{A}(x_0)$, there exists $\delta > 0$ such that for $0 < \tau \leq \delta$, we have

$$\underbrace{g_i(x_0 + \tau d)}_{\leq 0} = \underbrace{g_i(x_0)}_{=0} + \tau \langle \nabla g_i(x_0), d \rangle + o(\tau).$$

Dividing by τ and passing the limit as $\tau \to 0^+$ gives $\langle \nabla g_i(x_0), d \rangle \leq 0$. In the same way, we get $\langle \nabla h_j(x_0), d \rangle = 0$ for all $j = 1, 2, \ldots, \ell$.

Linearizing cone

Definition

For any $x_0 \in \mathscr{F}$, the set (cone)

$$\mathcal{C}_l(x_0) := \Big\{ d \in \mathbb{R}^n : \langle \nabla g_i(x_0), d \rangle \leq 0 \, \forall i \in \mathscr{A}(x_0), \\ \langle \nabla h_j(x_0), d \rangle = 0 \, \forall j = 1, 2, \ldots, \ell \Big\}$$

is called *linearizing cone*, of the constrained optimization problem (11), at x_0.

Evidently, $\mathcal{C}_l(x_0)$ contains at least all feasible directions of $\mathscr{F}$ at x_0, i.e., $\mathcal{C}_{fd}(x_0) \subset \mathcal{C}_l(x_0)$.

Cone of descent directions

Definition

For any $x_0 \in X$, the set (cone)

$$\mathcal{C}_{dd}(x_0) := \{d \in \mathbb{R}^n : \langle \nabla f(x_0), d \rangle < 0\}$$

is called *cone of descent directions* of f at x_0.

Note that for any $d \in \mathcal{C}_{dd}(x_0)$,

$$f(x_0 + \tau d) = f(x_0) + \tau \underbrace{\langle \nabla f(x_0), d \rangle}_{<\,0} + o(\tau)$$

$$\Longrightarrow \lim_{\tau \to 0+} \frac{1}{\tau}(f(x_0 + \tau d) - f(x_0)) = \langle \nabla f(x_0), d \rangle < 0$$

$$\Longrightarrow \exists \delta > 0 \text{ such that } \frac{1}{\tau}(f(x_0 + \tau d) - f(x_0)) < 0 \, \forall \tau \in (0, \delta]$$

$$(\text{since } \frac{1}{\tau}(f(x_0 + \tau d) - f(x_0) \text{ is continuous for } \tau > 0)$$

$$\Longrightarrow \exists \delta > 0 \text{ such that } f(x_0 + \tau d) < f(x_0), \, \forall \tau \in (0, \delta]. \tag{13}$$

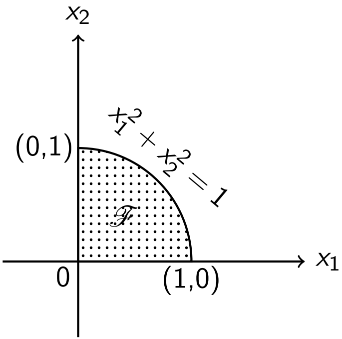
Thus, $d \in \mathcal{C}_{dd}(x_0)$ guarantees that the objective function f can be reduced along this direction.

Examples of $\mathcal{C}_{fd}(x_0)$, $\mathcal{C}_l(x_0)$ and $\mathcal{C}_{dd}(x_0)$

Let $f(x) := x_1 + x_2$ and

$$\mathscr{F} = \left\{x = (x_1, x_2) \in \mathbb{R}^2 : x_1^2 + x_2^2 - 1 \leq 0,\ -x_1 \leq 0,\ -x_2 \leq 0\right\},$$

where $g_1(x) := x_1^2 + x_2^2 - 1$, $g_2(x) := -x_1$ and $g_3(x) := -x_2$.
Evidently, f attains a (strict, global) minimum at $(0, 0)$.

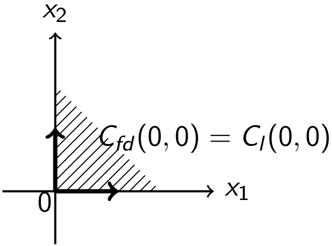


Examples of $\mathcal{C}_{fd}(x_0)$ and $\mathcal{C}_l(x_0)$

Take $x_0 = (0,0)$. Then, $\mathcal{A}(x_0) = \{2,3\}$,

$$\mathcal{C}_{fd}(x_0) = \{(d_1, d_2) \in \mathbb{R}^2 : d_1 \geq 0,\ d_2 \geq 0\}$$

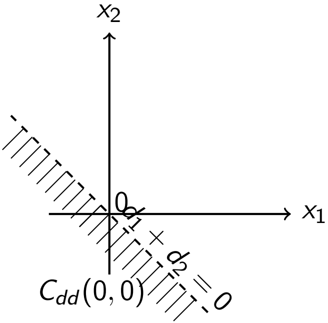
$$\begin{aligned} \mathcal{C}_l(x_0) &= \{d \in \mathbb{R}^2 : \langle \nabla g_2(x_0), d \rangle \leq 0,\ \langle \nabla g_3(x_0), d \rangle \leq 0\} \\ &= \left\{(d_1, d_2) \in \mathbb{R}^2 : d_1 \geq 0, d_2 \geq 0\right\} \end{aligned}$$



Examples of $\mathcal{C}_{dd}(x_0)$

At $x_0 = (0,0)$,

$$\begin{aligned}\mathcal{C}_{dd}(x_0) &= \{d \in \mathbb{R}^2 : \langle \nabla f(x_0), d\rangle < 0\} \\ &= \{(d_1, d_2) \in \mathbb{R}^2 : d_1 + d_2 < 0\}.\end{aligned}$$

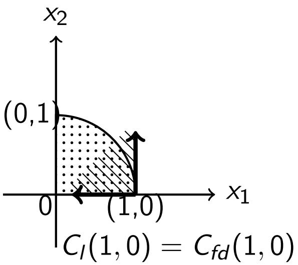


Examples of $\mathcal{C}_{fd}(x_0)$ and $\mathcal{C}_l(x_0)$

Take $x_0 = (1,0)$. Then, $\mathcal{A}(x_0) = \{1,3\}$,

$$\mathcal{C}_{fd}(x_0) = \{(d_1,d_2) \in \mathbb{R}^2 : d_1 < 0, d_2 \geq 0\},$$

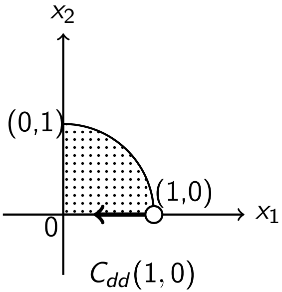
$$\begin{aligned} \mathcal{C}_l(x_0) &= \{d \in \mathbb{R}^2 : \langle \nabla g_1(x_0), d \rangle \leq 0,\ \langle \nabla g_3(x_0) d \rangle \leq 0\} \\ &= \left\{(d_1,d_2) \in \mathbb{R}^2 : d_1 \leq 0, d_2 \geq 0\right\}. \end{aligned}$$



Examples of $\mathcal{C}_{dd}(x_0)$

At $x_0 = (1, 0)$,

$$\mathcal{C}_{dd}(x_0) = \{d \in \mathbb{R}^2 : \langle \nabla f(x_0), d \rangle < 0\} = \{(d_1, d_2) \in \mathbb{R}^2 : d_1 < 0\}.$$



Geometric necessary optimality condition 1

Theorem

If x_0 is a local minimizer of the constrained optimization problem (11), *then*

$$\mathcal{C}_{dd}(x_0) \cap \mathcal{C}_{fd}(x_0) = \emptyset.$$

Proof

Assume contrary that there exists $d \in \mathcal{C}_{dd}(x_0) \cap \mathcal{C}_{fd}(x_0)$. Then, from (12) and (13), there exists $\delta > 0$ such that

$$x_0 + \tau d \in \mathscr{F} \text{ and } f(x_0 + \tau d) < f(x_0) \text{ for all } \tau \in (0, \delta],$$

which is contradictory to x_0 being a local minimizer of (11). Hence, $\mathcal{C}_{dd}(x_0) \cap \mathcal{C}_{fd}(x_0) = \emptyset$.

Remark

The condition $\mathcal{C}_{dd}(x_0) \cap \mathcal{C}_{fd}(x_0) = \emptyset$ is only a necessary optimality condition not sufficient. For example, consider

$$\begin{cases} \text{minimize} & f(x_1, x_2) := x_2 \\ \text{subject to} & h_1(x) := x_1^2 + x_2^2 - 1 = 0. \end{cases}$$

Take $x_0 = (0, 1)$. In this problem, $\mathcal{C}_{fd}(x_0) = \{\mathbf{0}\}$ and $\mathcal{C}_{dd}(x_0) = \{(d_1, d_2) \in \mathbb{R}^2 : d_2 < 0\}$.
Hence, $\mathcal{C}_{dd}(x_0) \cap \mathcal{C}_{fd}(x_0) = \emptyset$ but x_0 is not a minimizer.

Question arises whether not just $\mathcal{C}_{dd}(x_0) \cap \mathcal{C}_{fd}(x_0) = \emptyset$ but even $\mathcal{C}_{dd}(x_0) \cap \mathcal{C}_{l}(x_0) = \emptyset$ is true for any local minimizer $x_0 \in \mathscr{F}$.

The next example gives a negative answer to this question.

$\mathcal{C}_l(x_0) \cap \mathcal{C}_{dd}(x_0) = \emptyset$ is not necessary for optimality

Consider $X = \mathbb{R}^2$ and the following optimization problem:

$$\begin{cases} \text{minimize} & f(x) := -x_1 \\ \text{subject to} & g_1(x) := x_2 + (x_1 - 1)^3 \leq 0 \\ & g_2(x) := -x_1 \leq 0 \\ & g_3(x) := -x_2 \leq 0 \\ & x \in X. \end{cases}$$

Here $x_0 = (1, 0)$ is a minimizer. At x_0, we have $\mathcal{A}(x_0) = \{1, 3\}$,

$$\nabla f(x_0) = (-1, 0),\ \nabla g_1(x_0) = (0, 1),\ \nabla g_2(x_0) = (-1, 0) \text{ and } \nabla g_3(x_0) = (0, -1).$$

Thus, $\mathcal{C}_l(x_0) = \{(d_1, d_2) \in \mathbb{R}^2 : d_2 = 0\}$ and $\mathcal{C}_{dd}(x_0) = \{(d_1, d_2) \in \mathbb{R}^2 : d_1 > 0\}$.
Evidently, $\mathcal{C}_l(x_0) \cap \mathcal{C}_{dd}(x_0) = \{(d_1, d_2) \in \mathbb{R}^2 : d_1 > 0,\ d_2 = 0\} \neq \emptyset$.

As feasible directions allows movements along straight line segments those are embedded in $\mathscr{F}$, the cone of feasible direction turns out to be the singleton set containing the null vector if the feasible set $\mathscr{F}$ is a curved surface.

For example, consider $\mathscr{F} = \{(x_1, x_2) \in \mathbb{R}^2 : x_1^2 + x_2^2 = 1\}$ and any $x_0 \in \mathscr{F}$.
Note that $\mathcal{C}_{fd}(x_0) = \{\mathbf{0}\} = \{(\mathbf{0}, \mathbf{0})\}$.
Thus, often there would be no point of $\mathscr{F}$ from which progress could be made by a descent method for minimization.

As feasible directions allows movements along straight line segments those are embedded in $\mathscr{F}$, the cone of feasible direction turns out to be the singleton set containing the null vector if the feasible set $\mathscr{F}$ is a curved surface.

For example, consider $\mathscr{F} = \{(x_1, x_2) \in \mathbb{R}^2 : x_1^2 + x_2^2 = 1\}$ and any $x_0 \in \mathscr{F}$.
Note that $\mathcal{C}_{fd}(x_0) = \{\mathbf{0}\} = \{(\mathbf{0}, \mathbf{0})\}$.
Thus, often there would be no point of $\mathscr{F}$ from which progress could be made by a descent method for minimization.

As $\mathcal{C}_{fd}(x_0)$ is too limited to include quite a few descent direction, to development of a descent algorithms and to establish optimality conditions that are general enough, the notion of tangent cone is introduced.
Although the notion of tangent cone is less intuitive but contains $\mathcal{C}_{fd}(x_0)$.

As feasible directions allows movements along straight line segments those are embedded in $\mathscr{F}$, the cone of feasible direction turns out to be the singleton set containing the null vector if the feasible set $\mathscr{F}$ is a curved surface.

For example, consider $\mathscr{F} = \{(x_1, x_2) \in \mathbb{R}^2 : x_1^2 + x_2^2 = 1\}$ and any $x_0 \in \mathscr{F}$. Note that $\mathcal{C}_{fd}(x_0) = \{\mathbf{0}\} = \{(\mathbf{0}, \mathbf{0})\}$.
Thus, often there would be no point of $\mathscr{F}$ from which progress could be made by a descent method for minimization.

As $\mathcal{C}_{fd}(x_0)$ is too limited to include quite a few descent direction, to development of a descent algorithms and to establish optimality conditions that are general enough, the notion of tangent cone is introduced.
Although the notion of tangent cone is less intuitive but contains $\mathcal{C}_{fd}(x_0)$.

We say that a sequence $\{x_k\}$ in $\mathbb{R}^n$ converges in direction $d \in \mathbb{R}^n$ to x_0 if there exist sequences $\{r_k\}$ in $\mathbb{R}^n$ and $\{\alpha_k\}$ in $\mathbb{R}_+$ with $r_k \to 0$ and $\alpha_k \downarrow 0$ such that

$$x_k = x_0 + \alpha_k(d + r_k).$$

Then, we use the *notation*: $x_k \xrightarrow{d} x_0$.
Note that $x_k \xrightarrow{d} x_0$ simply means there exists a sequence of positive numbers $\{\alpha_k\}$ such that $\alpha_k \downarrow 0$ and

$$\frac{1}{\alpha_k}(x_k - x_0) \to d.$$

Tangent Cone

Definition

Let M be a nonempty subset of $\mathbb{R}^n$ and $x_0 \in M$. The set

$$\mathcal{C}_t(x_0) := \left\{ d \in \mathbb{R}^n : \exists \ \{x_k\} \text{ in } M \text{ such that } x_k \xrightarrow{d} x_0 \right\}$$

is called *tangent cone* of M at x_0.

The vectors of $\mathcal{C}_t(x_0)$ are called *tangents* or *tangent directions* of M at x_0.

The cone $\mathcal{C}_t(x_0)$ is also referred as *Bouligand's contingent cone*.

Examples of tangent cone $\mathcal{C}_t(x_0)$

Example

Consider the set

$$M_1 := \left\{x = (x_1, x_2) \in \mathbb{R}^2 : x_1 \geq 0,\ x_1^2 \geq x_2 \geq x_1^2(x_1 - 1)\right\}.$$

At $x_0 = (2, 4)$, $\mathcal{C}_t(M_1, x_0) = \{(d_1, d_2) \in \mathbb{R}^2 : 8d_1 - 12 \leq d_2 \leq 4d_1 - 4\}$.

At $x_0 = (0, 0)$, $\mathcal{C}_t(M_1, x_0) = \{(d_1, d_2) \in \mathbb{R}^2 : d_1 \geq 0, d_2 = 0\}$.

At $x_0 = (1, 0)$, $\mathcal{C}_t(M_1, x_0) = \{(d_1, d_2) \in \mathbb{R}^2 : d_2 \geq d_1 - 1\}$.

For the convenience of interrelating the location of x_0 in M_1 and $\mathcal{C}_t(M_1, x_0)$, figures of tangent cones after translating the origin to x_0 are given in the book by Forst and Hoffmann, pp. 49.

Examples of tangent cone $\mathcal{C}_t(x_0)$

Consider $M_2 := \left\{x \in \mathbb{R}^2 : -x_1^3 + x_2 \leq 0, -x_2 \leq 0\right\}$.

At $x_0 = (0, 0)$, $\mathcal{C}_t(x_0) = \left\{d \in \mathbb{R}^2 : d_1 \geq 0,\ d_2 = 0\right\}$.

Consider $M_3 := \left\{x \in \mathbb{R}^2 : -x_1^3 + x_2 \leq 0, -x_1 \leq 0, -x_2 \leq 0\right\}$.

At $x_0 = (0, 0)$, $\mathcal{C}_t(x_0) = \left\{d \in \mathbb{R}^2 : d_1 \geq 0,\ d_2 = 0\right\}$.

Theorem (Relationship among $\mathcal{C}_{fd}(x_0)$, $\mathcal{C}_l(x_0)$ and $\mathcal{C}_t(x_0)$)

$$\overline{\mathcal{C}_{fd}(x_0)} \subseteq \mathcal{C}_t(x_0) \subseteq \mathcal{C}_l(x_0).$$

Examples of tangent cone $\mathcal{C}_t(x_0)$

Consider $M_2 := \left\{x \in \mathbb{R}^2 : -x_1^3 + x_2 \leq 0, -x_2 \leq 0\right\}$.

At $x_0 = (0,0)$, $\mathcal{C}_t(x_0) = \left\{d \in \mathbb{R}^2 : d_1 \geq 0,\ d_2 = 0\right\}$.

Consider $M_3 := \left\{x \in \mathbb{R}^2 : -x_1^3 + x_2 \leq 0, -x_1 \leq 0, -x_2 \leq 0\right\}$.

At $x_0 = (0,0)$, $\mathcal{C}_t(x_0) = \left\{d \in \mathbb{R}^2 : d_1 \geq 0,\ d_2 = 0\right\}$.

Theorem (Relationship among $\mathcal{C}_{fd}(x_0)$, $\mathcal{C}_l(x_0)$ and $\mathcal{C}_t(x_0)$)

$$\overline{\mathcal{C}_{fd}(x_0)} \subseteq \mathcal{C}_t(x_0) \subseteq \mathcal{C}_l(x_0).$$

$\mathcal{C}_t(x_0)$ is strictly a subset of $\mathcal{C}_l(x_0)$

Example

Consider $M_2 := \left\{x \in \mathbb{R}^2 : -x_1^3 + x_2 \leq 0, -x_2 \leq 0\right\}$.

At $x_0 = (0,0)$, $\mathcal{C}_t(x_0) = \left\{d \in \mathbb{R}^2 : d_1 \geq 0,\ d_2 = 0\right\}$.

At $x_0 = (0,0)$, $\mathscr{A}(x_0) = \{1, 2\}$,

which gives $\mathcal{C}_l(x_0) = \left\{d \in \mathbb{R}^2 : d_2 = 0\right\}$. Hence, in this example, $\mathcal{C}_t(x_0) \subset \mathcal{C}_l(x_0)$.

$\mathcal{C}_t(x_0)$ is identical to $\mathcal{C}_l(x_0)$

Example

Consider $M_3 := \left\{x \in \mathbb{R}^2 : -x_1^3 + x_2 \leq 0,\ -x_1 \leq 0,\ -x_2 \leq 0\right\}$ and $x_0 = (0,0)$. Then, $\mathscr{A}(x_0) = \{1,2,3\}$ and therefore

$$\mathcal{C}_l(x_0) = \left\{d \in \mathbb{R}^2 : d_1 \geq 0,\ d_2 = 0\right\} = \mathcal{C}_t(x_0).$$

Remark

From the above two examples, we notice that although $M_2 = M_3$, the linearizing cones are different. Thus, linearizing cone is dependent on the algebraic representation of the set of feasible points! However, it is true in general that tangent cone depends only on the geometric representation and not on algebraic representation of the set.

Theorem (Geometric necessary optimality condition 2)

If x_0 is a local minimizer of the optimization problem (11)*, then* $\nabla f(x_0) \in \mathcal{C}_t(x_0)^*$*, and hence*

$$\mathcal{C}_{dd}(x_0) \cap \mathcal{C}_t(x_0) = \emptyset.$$

Theorem (First order Karush-Kuhn-Tucker necessary optimality condition)

Suppose that x_0 is a local minimizer of (11)*, and (the constraint qualification) $\mathcal{C}_\ell(x_0)^* = \mathcal{C}_t(x_0)^*$ is fulfilled. Then, there exists vectors $\lambda \in \mathbb{R}^m_+$ and $\mu \in \mathbb{R}^p$ such that*

$$\nabla f(x_0) + \sum_{i=1}^{m} \lambda_i \nabla g_i(x_0) + \sum_{j=1}^{p} \mu_j \nabla h_j(x_0) = 0 \tag{14}$$

$$\text{and } \lambda_i g_i(x_0) = 0 \text{ for } i = 1, 2, \dots, m. \tag{15}$$

Theorem (Geometric necessary optimality condition 2)

If x_0 *is a local minimizer of the optimization problem* (11), *then* $\nabla f(x_0) \in \mathcal{C}_t(x_0)^*$, *and hence*

$$\mathcal{C}_{dd}(x_0) \cap \mathcal{C}_t(x_0) = \emptyset.$$

Theorem (First order Karush-Kuhn-Tucker necessary optimality condition)

Suppose that x_0 *is a local minimizer of* (11), *and (the constraint qualification)* $\mathcal{C}_\ell(x_0)^* = \mathcal{C}_t(x_0)^*$ *is fulfilled. Then, there exists vectors* $\lambda \in \mathbb{R}^m_+$ *and* $\mu \in \mathbb{R}^p$ *such that*

$$\nabla f(x_0) + \sum_{i=1}^{m} \lambda_i \nabla g_i(x_0) + \sum_{j=1}^{p} \mu_j \nabla h_j(x_0) = 0 \tag{14}$$

$$\text{and } \lambda_i g_i(x_0) = 0 \text{ for } i = 1, 2, \ldots, m. \tag{15}$$

Proof

If x_0 is a local minimizer of (11), it follow from Lemma 35 and $\mathcal{C}_l(x_0)^* = \mathcal{C}_t(x_0)^*$ that $\nabla f(x_0) \in \mathcal{C}_\ell(x_0)^*$.
Note that for any $x_0 \in \mathscr{F}$,

$$\begin{aligned} \nabla f(x_0) \in \mathcal{C}_\ell(x_0)^* \quad &\Longleftrightarrow \quad \forall d \in \mathcal{C}_l(x_0), \nabla f(x_0)^\top d \geq 0 \\ &\Longleftrightarrow \quad \mathcal{C}_l(x_0) \cap \mathcal{C}_{dd}(x_0) = \emptyset. \end{aligned}$$

Proof continue ....

Hence, by definition of $\mathcal{C}_l(x_0)$ and $\mathcal{C}_{dd}(x_0)$, the system

$$\begin{cases} \langle \nabla f(x_0), d \rangle < 0 \\ \langle \nabla g_i(x_0), d \rangle \leq 0 \ \forall \ i \in \mathscr{A}(x_0) \\ \langle \nabla h_j(x_0), d \rangle = 0 \ \forall \ j = 1, 2, \ldots, \ell \end{cases}$$

that is,

$$\begin{cases} \langle \nabla f(x_0), d \rangle < 0 \\ -\langle \nabla g_i(x_0), d \rangle \geq 0 \ \forall \ i \in \mathscr{A}(x_0) \\ -\langle \nabla h_j(x_0), d \rangle \geq 0 \ \forall \ j = 1, 2, \ldots, \ell \\ \langle \nabla h_j(x_0), d \rangle \geq 0 \ \forall \ j = 1, 2, \ldots, \ell \end{cases}$$

does not have any solution.

Proof continue ....

Therefore, with the help of Theorem 15, there exists $\lambda_i \geq 0$ for each $i \in \mathscr{A}(x_0)$ and $\mu'_j \geq 0$, $\mu''_j \geq 0$ for each $j = 1, 2, \ldots, \ell$ such that

$$\nabla f(x_0) = \sum_{i \in \mathscr{A}(x_0)} \lambda_i \left(-\nabla g_i(x_0)\right) + \sum_{j=1}^{p} \mu'_j \left(-\nabla h_j(x_0)\right) + \sum_{j=1}^{p} \mu''_j \nabla h_j(x_0).$$

Setting $\lambda_i := 0$ for $i \in \{1, 2, \ldots, m\} \setminus \mathscr{A}(x_0)$ and $\mu_j := \mu'_j - \mu''_j$ for $j 1, 2, \ldots, m$, we see that there exists $\lambda_i \geq 0$ for $i \in \mathcal{I}$ and $\mu_j \in \mathbb{R}$ for $j \in \mathcal{E}$ with

$$\nabla f(x_0) + \sum_{i=1}^{m} \lambda_i \nabla g_i(x_0) + \sum_{j=1}^{p} \mu_j \nabla h_j(x_0) = 0$$

and

$$\lambda_i g_i(x_0) = 0 \quad \text{for all } i = 1, 2, \ldots, m.$$

KKT point

Definition

A point $x_0 \in X$ is called a *KKT point* for the problem (11) if it satisfies the following four (KKT) conditions:

1. (Primal Feasibility). x_0 is a feasible point of (11).
2. (Stationarity). $\nabla f(x_0) + \sum_{i=1}^{m} \lambda_i \nabla g_i(x_0) + \sum_{j=1}^{p} \mu_j \nabla h_j(x_0) = 0$ for some $\lambda \in \mathbb{R}^m$ and $\mu \in \mathbb{R}^l$.
3. (Dual Feasibility). The stationarity condition is satisfied for a $\lambda \in \mathbb{R}^m_+$ whose all the components are nonnegative.
4. (Complementary Slackness). With λ in dual feasibility, $\lambda_i g_i(x_0) = 0$.

Examples of a KKT point

Consider the problem

$$\begin{cases} \text{minimize} & f(x_1, x_2) := (x_1 - 3)^2 + (x_2 - 2)^2 \\ \text{subject to} & g_1(x_1, x_2) := x_1^2 + x_2^2 - 5 \leq \\ & g_2(x_1, x_2) := -x_1 \leq 0 \\ & g_3(x_1, x_2) := -x_2 \leq 0 \\ & h_1(x_1, x_2) := x_1 + 2x_2 = 4. \end{cases}$$

Let us check if the points $x_0 = (2, 1)$ is a KKT points for this problem.
The point x_0 is a (primal) feasible point.

Examples of a KKT point continue ....

We see that $\mathcal{A}(x_0) = \{1\}$ and

$$\nabla f(x_0) = \begin{bmatrix} -2 \\ -2 \end{bmatrix}, \nabla g_1(x_0) = \begin{bmatrix} 4 \\ 2 \end{bmatrix} \text{ and } \nabla h_1(x_0) = \begin{bmatrix} 1 \\ 2 \end{bmatrix}.$$

Let us check if the stationarity holds at x_0. For this, consider the system

$$\begin{aligned} \nabla f(x_0) + \lambda_1 \nabla g_1(x_0) + \mu_1 \nabla h_1(x_0) &= 0 \\ \text{i.e., } \begin{bmatrix} -2 \\ -2 \end{bmatrix} + \lambda_1 \begin{bmatrix} 4 \\ 2 \end{bmatrix} + \mu_1 \begin{bmatrix} 1 \\ 2 \end{bmatrix} &= \begin{bmatrix} 0 \\ 0 \end{bmatrix}. \end{aligned} \tag{16}$$

Notice that $\lambda_1 = \frac{1}{3}$ and $\mu_1 = \frac{2}{3}$ is a solution to (16). Thus, x_0 satisfies stationarity.

Examples of a KKT point continue ....

Taking $\lambda_2 = \lambda_3 = 0$, we see that

$$\lambda_1 g_1(x_0) = 0, \lambda_2 g_2(x_0) = 0 \text{ and } \lambda_3 g_3(x_0) = 0,$$

which shows that complementary slackness holds at x_0.
Further notice that $\lambda_1 \geq 0$, $\lambda_2 \geq 0$, and $\lambda_3 \geq 0$. Thus, dual feasibility holds.
Therefore, x_0 is a KKT point for the considered problem.

First order KKT sufficient optimality condition

Theorem

Consider the problem (11) where f and g_i, $i = 1, 2, \ldots, m$ are convex and $h_j(x) = \langle a_j, x \rangle - b_j$ for some $a_j \in \mathbb{R}^n$ and $b_j \in \mathbb{R}$, $j = 1, 2, \ldots, \ell$. Suppose at an $x_0 \in \mathscr{F}$ the KKT conditions are satisfied. Then, x_0 is a global minimum point.

Remark

There is one important point to note in regard to KKT conditions that is often a source of error. Namely, despite the usually the well behaved nature of convex programming problems and the sufficiency of KKT conditions under convexity assumptions, the KKT conditions are not necessary for optimality for convex programming problems. For example, consider

$$\begin{cases} \text{minimize} & x_1 \\ \text{subject to} & (x_1-1)^2+(x_2-1)^2 \leq 1 \\ & (x_1-1)^2+(x_2+1)^2 \leq 1. \end{cases}$$

The only feasible point $(1,0)$ is naturally optimal. However, this is not a KKT point.

For the problem in Remark 82, a careful reader must have observed that the point $(1,0)$ does not satisfy the hypothesis of Theorem 36 since $\mathcal{C}_l(x_0)^* = \{(0,0)\} \neq \mathbb{R}^2 = \mathcal{C}_t(x_0)^*$. Hence, although $(1,0)$ is optimizer, it is not a KKT point.

In order to ensure that the KKT conditions are necessary for optimality, a constraint qualification (CQ) is needed. A CQ is an assumption made about the constraint functions that, when satisfied by a local minimizer, ensures stationarity.

However, the CQ $\mathcal{C}_l(x_0)^* = \mathcal{C}_t(x_0)^*$ is very abstract, extremely general, but not easily verifiable. Therefore, for practical problems, we will try to find regularity assumptions called *constraint qualification* (CQ) which are more specific, easily verifiable, but also somewhat restrictive.
For the moment, we consider the case that we only have inequality constraints.
As affine constraints poses fewer problems than nonlinear constraints, we the partition the set inequality indices $\mathcal{I} = \{1, 2, \dots, m\}$ by

$$\mathcal{I} = \mathcal{I}_1 \cup \mathcal{I}_2,$$

where $\mathcal{I}_2$ consists the indices corresponding to the affine inequalities and $\mathcal{I}_1$ are for nonaffine constraints.
Corresponding to this partition, we will also split up the set of active constraints $\mathcal{A}(x_0)$ for $x_0 \in \mathcal{F}$ into

$$\mathcal{A}_j(x_0) := \mathcal{I}_j \cap \mathcal{A}(x_0) \text{ for } j = 1, 2.$$

We now will focus on following Constraint Qualifications:
(GCQ) Guignard Constraint Qualification: $\mathcal{C}_l(x_0)^* = \mathcal{C}_t(x_0)^*$
(ACQ) Abadie Constraint Qualification: $\mathcal{C}_l(x_0) = \mathcal{C}_t(x_0)$
(MFCQ) Mangasarian-Fromovitz Constraint Qualification:

$$\exists\ d \in \mathbb{R}^n \text{ such that } \begin{cases} g_i'(x_0)d < 0 & \text{for } i \in \mathcal{A}_1(x_0) \\ g_i'(x_0)d \leq 0 & \text{for } i \in \mathcal{A}_2(x_0) \end{cases}$$

(SCQ) Slater Constraint Qualification: The functions g_i are convex for all $i \in \mathcal{I}$ and

$$\exists\ \tilde{x} \in \mathcal{F} \text{ such that } g_i(\tilde{x}) < 0 \text{ for } i \in \mathcal{I}_1.$$

The conditions $g_i'(x_0)d < 0$ and $g_i'(x_0)d \leq 0$ each define half space. (MFCQ) means nothing else but that the intersection of all of these half spaces is nonempty.

Theorem

(SCQ) $\implies$ *(MFCQ)* $\implies$ *(ACQ)* $\implies$ *(GCQ).*

Now we will consider the general case, where there may also occur equality constraints. In this context, one often finds the following Linear Independence Constraint Qualification in the literature:

(LICQ) The vectors $(\nabla g_i(x_0) : i \in \mathcal{A}(x_0))$ and $(\nabla h_j(x_0) : j \in 1, 2, \ldots, l)$ are linearly independent.

(LICQ) greatly reduces the number of active inequality constraints. Instead of (LICQ) we will now consider the following weaker constraint qualification which is a variant of (MFCQ), and is often cited as the Arrow-Hurwitz-Uzawa Constraint Qualification:

(AHUCQ) There exists a $d \in \mathbb{R}^n$ such that

$$\begin{cases} g_i'(x_0)d < 0 & \text{for } i \in \mathcal{A}_1(x_0) \\ h_j'(x_0)d = 0 & \text{for } j \in \mathcal{E}, \end{cases}$$

and the vectors $(\nabla h_j(x_0) : j \in \mathcal{E})$ are linearly independent.

Theorem

(LICQ) $\implies$ *(AHUCQ)* $\implies$ *(ACQ)* $\implies$ *(GCQ).*

Questions?

&

Questions?

&

Thanks!

Questions?

&

Thanks!