TANGENT CONE AND CONSTRAINT QUALIFICATIONS

Saurav Samantaray

Department of Mathematics

Indian Institute of Technology Madras

April 2, 2024

TANGENT CONE

- We determined whether or not it was possible to take a feasible descent step away from a given feasible point x;

TANGENT CONE

- We determined whether or not it was possible to take a feasible descent step away from a given feasible point x;
- by examining the first derivatives of f and;

TANGENT CONE

- We determined whether or not it was possible to take a feasible descent step away from a given feasible point x;
- by examining the first derivatives of f and;
- the constraint functions c_i.

TANGENT CONE

- We determined whether or not it was possible to take a feasible descent step away from a given feasible point x;
- by examining the first derivatives of f and;
- the constraint functions c_i.
- The first-order Taylor series expansion of these functions about x was used to form an approximate problem in which both objective and constraints are linear.

TANGENT CONE

- We determined whether or not it was possible to take a feasible descent step away from a given feasible point x;
- by examining the first derivatives of f and;
- the constraint functions c_i.
- The first-order Taylor series expansion of these functions about x was used to form an approximate problem in which both objective and constraints are linear.
- Makes sense if the linearised approximation captures the essential geometric features of the feasible set near the point x in question.

TANGENT CONE

- We determined whether or not it was possible to take a feasible descent step away from a given feasible point x;
- by examining the first derivatives of f and;
- the constraint functions c_i.
- The first-order Taylor series expansion of these functions about x was used to form an approximate problem in which both objective and constraints are linear.
- Makes sense if the linearised approximation captures the essential geometric features of the feasible set near the point x in question.
- Assumptions about the nature of the constraints c_i that are active at x are needed to be made to ensure that the linearised approximation is similar to the feasible set, near x.

Cone

Definition

A cone is a set $\mathscr{F}$ with the property that for all $x \in \mathscr{F}$ we have

$$x \in \mathscr{F} \implies \alpha x \in \mathscr{F}, \text{ for all } \alpha > 0.$$

Cone

Definition

A cone is a set $\mathscr{F}$ with the property that for all $x \in \mathscr{F}$ we have

$$x \in \mathscr{F} \implies \alpha x \in \mathscr{F}, \text{ for all } \alpha > 0.$$

For example, the set $\mathscr{F} \subset \mathbb{R}^2$ defined by

$$\{(x_1, x_2)^T | x_1 > 0, x_2 \geq 0\}$$

is a cone in $\mathbb{R}^2$.

TANGENT CONE

- Given a feasible point x, $\{z_k\}$ is called a feasible sequence approaching x, if $z_k \in \Omega$ for all k, sufficiently large and $z_k \to x$.

Definition

The vector d is said to be a tangent (or tangent vector) to Ω at a point x if there are a feasible sequence $\{z_k\}$ approaching x and a sequence of positive scalars $\{t_k\}$ with $t_k \to 0$ such that

$$\lim_{k\to\infty} \frac{z_k - x}{t_k} = d. \tag{1}$$

The set of all tangents to Ω at x^* is called the tangent cone and is denoted by $T_\Omega(x^*)$.

TANGENT CONE

- Let d be a tangent vector with corresponding sequences $\{z_k\}$ and $\{t_k\}$.

TANGENT CONE

- Let d be a tangent vector with corresponding sequences $\{z_k\}$ and $\{t_k\}$.
- Consider

$$\lim_{k\to\infty} \frac{z_k - x}{\alpha^{-1} t_k} = \alpha \lim_{k\to\infty} \frac{z_k - x}{t_k} = \alpha d$$

TANGENT CONE

- Let d be a tangent vector with corresponding sequences $\{z_k\}$ and $\{t_k\}$.
- Consider

$$\lim_{k\to\infty} \frac{z_k - x}{\alpha^{-1} t_k} = \alpha \lim_{k\to\infty} \frac{z_k - x}{t_k} = \alpha d$$

- Therefore, for any $\alpha > 0$, if d is a tangent vector then αd also is i.e.

$$\text{if } d \in T_\Omega(x^*) \qquad \Longrightarrow \alpha d \in T_\Omega(x^*)$$

TANGENT CONE

- Let d be a tangent vector with corresponding sequences $\{z_k\}$ and $\{t_k\}$.
- Consider
$$\lim_{k\to\infty} \frac{z_k - x}{\alpha^{-1} t_k} = \alpha \lim_{k\to\infty} \frac{z_k - x}{t_k} = \alpha d$$
- Therefore, for any $\alpha > 0$, if d is a tangent vector then αd also is i.e.
$$\text{if } d \in T_\Omega(x^*) \qquad \implies \alpha d \in T_\Omega(x^*)$$
- By setting $z_k \equiv x$ the constant sequence, implies $0 \in T_\Omega(x^*)$.

Linearised Feasible Direction

Definition

Given a feasible point x and the active constraint set $\mathscr{A}(x)$, the set of linearised feasible directions $\mathscr{F}(x)$ is

$$\mathscr{F}(x) = \left\{ d \middle| \begin{array}{ll} d^T \nabla c_i(x) = 0, & \text{for all } i \in \mathscr{E} \\ d^T \nabla c_i(x) \geq 0, & \text{for all } i \in \mathscr{A}(x) \cap \mathscr{I} \end{array} \right\} \tag{2}$$

Linearised Feasible Direction

Definition

Given a feasible point x and the active constraint set $\mathscr{A}(x)$, the set of linearised feasible directions $\mathscr{F}(x)$ is

$$\mathscr{F}(x) = \left\{ d \middle| \begin{array}{ll} d^T \nabla c_i(x) = 0, & \text{for all } i \in \mathscr{E} \\ d^T \nabla c_i(x) \geq 0, & \text{for all } i \in \mathscr{A}(x) \cap \mathscr{I} \end{array} \right\} \tag{2}$$

- $\mathscr{F}(x)$ is also a cone.

Linearised Feasible Direction

Definition

Given a feasible point x and the active constraint set $\mathscr{A}(x)$, the set of linearised feasible directions $\mathscr{F}(x)$ is

$$\mathscr{F}(x) = \left\{ d \middle| \begin{array}{ll} d^T \nabla c_i(x) = 0, & \text{for all } i \in \mathscr{E} \\ d^T \nabla c_i(x) \geq 0, & \text{for all } i \in \mathscr{A}(x) \cap \mathscr{I} \end{array} \right\} \tag{2}$$

- $\mathscr{F}(x)$ is also a cone.
- The definition of tangent cone does not explicitly depend on the constraints c_i it depends on the geometry of Ω.

Linearised Feasible Direction

Definition

Given a feasible point x and the active constraint set $\mathscr{A}(x)$, the set of linearised feasible directions $\mathscr{F}(x)$ is

$$\mathscr{F}(x) = \left\{ d \middle| \begin{array}{ll} d^T \nabla c_i(x) = 0, & \text{for all } i \in \mathscr{E} \\ d^T \nabla c_i(x) \geq 0, & \text{for all } i \in \mathscr{A}(x) \cap \mathscr{I} \end{array} \right\} \tag{2}$$

- $\mathscr{F}(x)$ is also a cone.
- The definition of tangent cone does not explicitly depend on the constraints c_i it depends on the geometry of Ω.
- The linearised feasible direction set does, however, depend on the definition of the constraint functions c_i , $i \in \mathscr{E} \cup \mathscr{I}$.

Tangent Cone and Feasible Direction for One Equality Constraint

- Consider the problem with one equality constraint.

Tangent Cone and Feasible Direction for One Equality Constraint

- Consider the problem with one equality constraint.
- The objective function $f(x) = x_1 + x_2$, $\mathscr{E} = \{1\}$, $\mathscr{I} = \phi$

Tangent Cone and Feasible Direction for One Equality Constraint

- Consider the problem with one equality constraint.
- The objective function $f(x) = x_1 + x_2$, $\mathscr{E} = \{1\}$, $\mathscr{I} = \phi$
- $c_1(x) = x_1^2 + x_2^2 - 2$

Tangent Cone and Feasible Direction for One Equality Constraint

- Consider the problem with one equality constraint.
- The objective function $f(x) = x_1 + x_2$, $\mathscr{E} = \{1\}$, $\mathscr{I} = \phi$
- $c_1(x) = x_1^2 + x_2^2 - 2$
- The feasible set for this problem is the circle of radius $\sqrt{2}$ centered at the origin.

Tangent Cone and Feasible Direction for One Equality Constraint

- Consider the problem with one equality constraint.
- The objective function $f(x) = x_1 + x_2$, $\mathscr{E} = \{1\}$, $\mathscr{I} = \phi$
- $c_1(x) = x_1^2 + x_2^2 - 2$
- The feasible set for this problem is the circle of radius $\sqrt{2}$ centered at the origin.

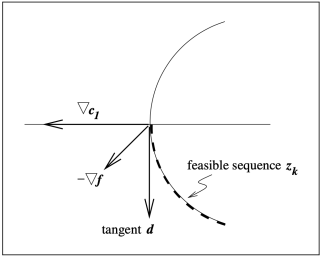

Figure: Constraint normal, objective gradient, and feasible sequence

Tangent Cone and Feasible Direction

- Consider the non-optimal point $x = (\sqrt{2}, 0)^T$.

Tangent Cone and Feasible Direction

- Consider the non-optimal point $x = (\sqrt{2}, 0)^T$.
- The figure also shows a feasible sequence approaching x.

$$z_k = \begin{bmatrix} -\sqrt{2 - 1/k^2} \\ -1/k \end{bmatrix}$$

Tangent Cone and Feasible Direction

- Consider the non-optimal point $x = (\sqrt{2}, 0)^T$.
- The figure also shows a feasible sequence approaching x.

$$z_k = \begin{bmatrix} -\sqrt{2 - 1/k^2} \\ -1/k \end{bmatrix}$$

- Choose $t_k = ||z_k - x||$, to get $d = (0, -1)^T$ is a tangent.

Tangent Cone and Feasible Direction

- Consider the non-optimal point $x = (\sqrt{2}, 0)^T$.
- The figure also shows a feasible sequence approaching x.

$$z_k = \begin{bmatrix} -\sqrt{2 - 1/k^2} \\ -1/k \end{bmatrix}$$

- Choose $t_k = ||z_k - x||$, to get $d = (0, -1)^T$ is a tangent.
- f increases as we move along z_k, i.e. $f(z_{k+1}) > f(z_k)$ for all $k = 2, 3, \ldots$.

Tangent Cone and Feasible Direction

- Consider the non-optimal point $x = (\sqrt{2}, 0)^T$.
- The figure also shows a feasible sequence approaching x.

$$z_k = \begin{bmatrix} -\sqrt{2 - 1/k^2} \\ -1/k \end{bmatrix}$$

- Choose $t_k = ||z_k - x||$, to get $d = (0, -1)^T$ is a tangent.
- f increases as we move along z_k, i.e. $f(z_{k+1}) > f(z_k)$ for all $k = 2, 3, \ldots$.
- $f(z_k) < f(x)$ for $k = 2, 3, \ldots$, so x cannot be a minimiser.

Tangent Cone and Feasible Direction

- Another feasible sequence is one that approaches $x = (-\sqrt{2}, 0)^T$ from the opposite direction.

Tangent Cone and Feasible Direction

- Another feasible sequence is one that approaches $x = (-\sqrt{2}, 0)^T$ from the opposite direction.
- $$z_k = \begin{bmatrix} -\sqrt{2 - 1/k^2} \\ 1/k \end{bmatrix}$$

Tangent Cone and Feasible Direction

- Another feasible sequence is one that approaches $x = (-\sqrt{2}, 0)^T$ from the opposite direction.
- $$z_k = \begin{bmatrix} -\sqrt{2 - 1/k^2} \\ 1/k \end{bmatrix}$$
- f decreases along this sequence.

Tangent Cone and Feasible Direction

- Another feasible sequence is one that approaches $x = (-\sqrt{2}, 0)^T$ from the opposite direction.
- $$z_k = \begin{bmatrix} -\sqrt{2 - 1/k^2} \\ 1/k \end{bmatrix}$$
- f decreases along this sequence.
- The tangents corresponding to this sequence are $d = (0, \alpha)^T$.

Tangent Cone and Feasible Direction

- Another feasible sequence is one that approaches $x = (-\sqrt{2}, 0)^T$ from the opposite direction.
- $$z_k = \begin{bmatrix} -\sqrt{2 - 1/k^2} \\ 1/k \end{bmatrix}$$
- f decreases along this sequence.
- The tangents corresponding to this sequence are $d = (0, \alpha)^T$.
- In summary, the tangent cone at $x = (-\sqrt{2}, 0)^T$ is $\{(0, d_2)^T | d_2 \in \mathbb{R}\}$.

Tangent Cone and Feasible Direction

- For the set of linearised feasible directions $\mathscr{F}(x)$, $d = (d_1, d_2)^T \in \mathscr{F}(x)$ if

$$0 = \nabla c_1(x)^T d = \begin{bmatrix} 2x_1 \\ 2x_2 \end{bmatrix}^T \begin{bmatrix} d_1 \\ d_2 \end{bmatrix} = -2\sqrt{2} d_1$$

Tangent Cone and Feasible Direction

- For the set of linearised feasible directions $\mathscr{F}(x)$, $d = (d_1, d_2)^T \in \mathscr{F}(x)$ if

$$0 = \nabla c_1(x)^T d = \begin{bmatrix} 2x_1 \\ 2x_2 \end{bmatrix}^T \begin{bmatrix} d_1 \\ d_2 \end{bmatrix} = -2\sqrt{2} d_1$$

- $\mathscr{F}(x) = \{(0, d_2)^T | d_2 \in \mathbb{R}\}$.

Tangent Cone and Feasible Direction

- For the set of linearised feasible directions $\mathscr{F}(x)$, $d = (d_1, d_2)^T \in \mathscr{F}(x)$ if

$$0 = \nabla c_1(x)^T d = \begin{bmatrix} 2x_1 \\ 2x_2 \end{bmatrix}^T \begin{bmatrix} d_1 \\ d_2 \end{bmatrix} = -2\sqrt{2}d_1$$

- $\mathscr{F}(x) = \{(0, d_2)^T | d_2 \in \mathbb{R}\}$.
- In this case $T_\Omega(x) = \mathscr{F}(x)$.

Tangent Cone and Feasible Direction

- For the set of linearised feasible directions $\mathscr{F}(x)$, $d = (d_1, d_2)^T \in \mathscr{F}(x)$ if

$$0 = \nabla c_1(x)^T d = \begin{bmatrix} 2x_1 \\ 2x_2 \end{bmatrix}^T \begin{bmatrix} d_1 \\ d_2 \end{bmatrix} = -2\sqrt{2} d_1$$

- $\mathscr{F}(x) = \{(0, d_2)^T | d_2 \in \mathbb{R}\}$.
- In this case $T_\Omega(x) = \mathscr{F}(x)$.
- Suppose that the feasible set is defined instead by the formula

$$\Omega = \{x | c_1(x) = 0\}, \qquad \text{where } c_1(x) = (x_1^2 + x_2^2 - 2)^2 = 0$$

- Ω is geometrically the same, but with a different algebraic specification.

Tangent Cone and Feasible Direction

- Then d belongs to the linearised feasible set if:

$$0 = \nabla c_1(x)^T d = \begin{bmatrix} 4(x_1^2 + x_2^2 - 2)x_1 \\ 4(x_1^2 + x_2^2 - 2)x_2 \end{bmatrix}^T \begin{bmatrix} d_1 \\ d_2 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}^T \begin{bmatrix} d_1 \\ d_2 \end{bmatrix}$$

Tangent Cone and Feasible Direction

- Then d belongs to the linearised feasible set if:

$$0 = \nabla c_1(x)^T d = \begin{bmatrix} 4(x_1^2 + x_2^2 - 2)x_1 \\ 4(x_1^2 + x_2^2 - 2)x_2 \end{bmatrix}^T \begin{bmatrix} d_1 \\ d_2 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}^T \begin{bmatrix} d_1 \\ d_2 \end{bmatrix}$$

- which is true for all $(d_1, d_2)^T$.

Tangent Cone and Feasible Direction

- Then d belongs to the linearised feasible set if:

$$0 = \nabla c_1(x)^T d = \begin{bmatrix} 4(x_1^2 + x_2^2 - 2)x_1 \\ 4(x_1^2 + x_2^2 - 2)x_2 \end{bmatrix}^T \begin{bmatrix} d_1 \\ d_2 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}^T \begin{bmatrix} d_1 \\ d_2 \end{bmatrix}$$

- which is true for all $(d_1, d_2)^T$.
- $\mathscr{F}(x) = \mathbb{R}^2$.

Tangent Cone and Feasible Direction

- Then d belongs to the linearised feasible set if:

$$0 = \nabla c_1(x)^T d = \begin{bmatrix} 4(x_1^2 + x_2^2 - 2)x_1 \\ 4(x_1^2 + x_2^2 - 2)x_2 \end{bmatrix}^T \begin{bmatrix} d_1 \\ d_2 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}^T \begin{bmatrix} d_1 \\ d_2 \end{bmatrix}$$

- which is true for all $(d_1, d_2)^T$.
- $\mathscr{F}(x) = \mathbb{R}^2$.
- So for this algebraic specification of Ω, the tangent cone and linearised feasible sets differ.

Tangent Cone and Feasible Direction for One In-Equality Constraint

- The solution $x = (-1, -1)^T$ is the same as in the equality-constrained case.

Tangent Cone and Feasible Direction for One In-Equality Constraint

- The solution $x = (-1, -1)^T$ is the same as in the equality-constrained case.
- But, there is a much more extensive collection of feasible sequences that converge to any given feasible point.

Tangent Cone and Feasible Direction for One In-Equality Constraint

- The solution $x = (-1, -1)^T$ is the same as in the equality-constrained case.
- But, there is a much more extensive collection of feasible sequences that converge to any given feasible point.

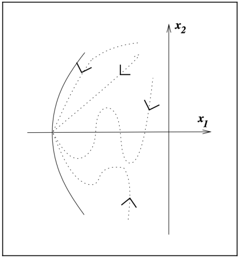

Figure: Feasible sequences converging to a particular feasible point for the region defined by $x_1^2 + x_2^2 \leq 2$

Tangent Cone and Feasible Direction

- From the point $x = (-\sqrt{2}, 0)^T$, all the feasible sequences defined above for the equality-constrained problem are still feasible.

Tangent Cone and Feasible Direction

- From the point $x = (-\sqrt{2}, 0)^T$, all the feasible sequences defined above for the equality-constrained problem are still feasible.
- There are also infinitely many feasible sequences that converge to x, along a straight line from the interior of the circle.

Tangent Cone and Feasible Direction

- From the point $x = (-\sqrt{2}, 0)^T$, all the feasible sequences defined above for the equality-constrained problem are still feasible.
- There are also infinitely many feasible sequences that converge to x, along a straight line from the interior of the circle.
- $$z_k = (-\sqrt{2}, 0)^T + (1/k)w,$$
  where w is any vector whose first component is positive $(w_1 > 0)$.

Tangent Cone and Feasible Direction

- From the point $x = (-\sqrt{2}, 0)^T$, all the feasible sequences defined above for the equality-constrained problem are still feasible.
- There are also infinitely many feasible sequences that converge to x, along a straight line from the interior of the circle.
- $$z_k = (-\sqrt{2}, 0)^T + (1/k)w,$$
  where w is any vector whose first component is positive $(w_1 > 0)$.
- z_k is feasible provided that $||z_k|| \leq \sqrt{2}$ i.e.
  $$(-\sqrt{2} + w_1/k)^2 + (w_2/k)^2 \leq 2,$$

Tangent Cone and Feasible Direction

- From the point $x = (-\sqrt{2}, 0)^T$, all the feasible sequences defined above for the equality-constrained problem are still feasible.
- There are also infinitely many feasible sequences that converge to x, along a straight line from the interior of the circle.
- $$z_k = (-\sqrt{2}, 0)^T + (1/k)w,$$

  where w is any vector whose first component is positive ($w_1 > 0$).
- z_k is feasible provided that $||z_k|| \leq \sqrt{2}$ i.e.

  $$(-\sqrt{2} + w_1/k)^2 + (w_2/k)^2 \leq 2,$$

- Which is true when $k \geq (w_1^2 + w_2^2)/(2\sqrt{2}w_1)$

Tangent Cone and Feasible Direction

- we can also define an infinite variety of sequences that approach x along a curve from the interior of the circle.

Tangent Cone and Feasible Direction

- we can also define an infinite variety of sequences that approach x along a curve from the interior of the circle.
- To summarize, the tangent cone to this set at $(-\sqrt{2}, 0)^T$ is $\{(w_1, w_2)^T | w_1 \geq 0\}$.

Tangent Cone and Feasible Direction

- we can also define an infinite variety of sequences that approach x along a curve from the interior of the circle.
- To summarize, the tangent cone to this set at $(-\sqrt{2}, 0)^T$ is $\{(w_1, w_2)^T | w_1 \geq 0\}$.
- For the feasibility set $\mathscr{F}(x)$ let us consider:

$$0 \leq \nabla c_1(x)^T d = \begin{bmatrix} -2x_1 \\ -2x_2 \end{bmatrix}^T \begin{bmatrix} d_1 \\ d_2 \end{bmatrix} = 2\sqrt{2} d_1$$

Tangent Cone and Feasible Direction

- we can also define an infinite variety of sequences that approach x along a curve from the interior of the circle.
- To summarize, the tangent cone to this set at $(-\sqrt{2}, 0)^T$ is $\{(w_1, w_2)^T | w_1 \geq 0\}$.
- For the feasibility set $\mathscr{F}(x)$ let us consider:

$$0 \leq \nabla c_1(x)^T d = \begin{bmatrix} -2x_1 \\ -2x_2 \end{bmatrix}^T \begin{bmatrix} d_1 \\ d_2 \end{bmatrix} = 2\sqrt{2} d_1$$

- Hence, we obtain $\mathscr{F}(x) = T_\Omega(x)$ for this particular algebraic specification of the feasible set.

Constraint qualifications

- Constraint qualifications are conditions under which the linearised feasible set $\mathscr{F}(x)$ is similar to the tangent cone $T_{\Omega}(x)$.

Constraint qualifications

- Constraint qualifications are conditions under which the linearised feasible set $\mathscr{F}(x)$ is similar to the tangent cone $T_\Omega(x)$.
- Most constraint qualifications ensure that these two sets are identical.

Constraint qualifications

- Constraint qualifications are conditions under which the linearised feasible set $\mathscr{F}(x)$ is similar to the tangent cone $T_\Omega(x)$.
- Most constraint qualifications ensure that these two sets are identical.
- These conditions ensure that the $\mathscr{F}(x)$, which is constructed by linearising the algebraic description of the set Ω at x, captures the essential geometric features of the set Ω in the vicinity of x, as represented by T_Ω.

Constraint qualifications

- Constraint qualifications are conditions under which the linearised feasible set $\mathscr{F}(x)$ is similar to the tangent cone $T_\Omega(x)$.
- Most constraint qualifications ensure that these two sets are identical.
- These conditions ensure that the $\mathscr{F}(x)$, which is constructed by linearising the algebraic description of the set Ω at x, captures the essential geometric features of the set Ω in the vicinity of x, as represented by T_Ω.

Definition

Given the point x and the active set $\mathscr{A}(x)$, we say that the linear independence constraint qualification (LICQ) holds if the set of active constraint gradients $\{\nabla c_i(x) | i \in \mathscr{A}(x)\}$ is linearly independent.

Constraint qualifications

- Constraint qualifications are conditions under which the linearised feasible set $\mathscr{F}(x)$ is similar to the tangent cone $T_\Omega(x)$.
- Most constraint qualifications ensure that these two sets are identical.
- These conditions ensure that the $\mathscr{F}(x)$, which is constructed by linearising the algebraic description of the set Ω at x, captures the essential geometric features of the set Ω in the vicinity of x, as represented by T_Ω.

Definition

Given the point x and the active set $\mathscr{A}(x)$, we say that the linear independence constraint qualification (LICQ) holds if the set of active constraint gradients $\{\nabla c_i(x) | i \in \mathscr{A}(x)\}$ is linearly independent.

In general, if LICQ holds, none of the active constraint gradients can be zero.

FIRST-ORDER OPTIMALITY CONDITIONS

Consider the constrained optimisation problem

$$\min_{x\in\mathbb{R}^n} f(x) \qquad \text{subject to} \qquad \begin{cases} c_i(x) & = 0, \quad i \in \mathscr{E} \\ c_j(x) & \geq 0, \quad j \in \mathscr{I} \end{cases} \tag{3}$$

f and c_i are scalar valued functions of the vector of unknowns x and $\mathscr{E}$ and $\mathscr{I}$ are set of indices.

FIRST-ORDER OPTIMALITY CONDITIONS

Consider the constrained optimisation problem

$$\min_{x\in\mathbb{R}^n} f(x) \qquad \text{subject to} \qquad \begin{cases} c_i(x) & = 0, \quad i \in \mathscr{E} \\ c_j(x) & \geq 0, \quad j \in \mathscr{I} \end{cases} \tag{3}$$

f and c_i are scalar valued functions of the vector of unknowns x and $\mathscr{E}$ and $\mathscr{I}$ are set of indices. Define the Lagrangian function for the general problem as

$$\mathscr{L}(x,\lambda) = f(x) - \sum_{i\in\mathscr{E}\cup\mathscr{I}} \lambda_i c_i(x). \tag{4}$$

FIRST-ORDER OPTIMALITY CONDITIONS

Consider the constrained optimisation problem

$$\min_{x \in \mathbb{R}^n} f(x) \qquad \text{subject to} \qquad \begin{cases} c_i(x) & = 0, \quad i \in \mathscr{E} \\ c_j(x) & \geq 0, \quad j \in \mathscr{I} \end{cases} \tag{3}$$

f and c_i are scalar valued functions of the vector of unknowns x and $\mathscr{E}$ and $\mathscr{I}$ are set of indices. Define the Lagrangian function for the general problem as

$$\mathscr{L}(x, \lambda) = f(x) - \sum_{i \in \mathscr{E} \cup \mathscr{I}} \lambda_i c_i(x). \tag{4}$$

- The necessary conditions defined in the following theorem are called first-order conditions.

FIRST-ORDER OPTIMALITY CONDITIONS

Consider the constrained optimisation problem

$$\min_{x\in\mathbb{R}^n} f(x) \qquad \text{subject to} \qquad \begin{cases} c_i(x) & = 0, \quad i \in \mathscr{E} \\ c_j(x) & \geq 0, \quad j \in \mathscr{I} \end{cases} \tag{3}$$

f and c_i are scalar valued functions of the vector of unknowns x and $\mathscr{E}$ and $\mathscr{I}$ are set of indices. Define the Lagrangian function for the general problem as

$$\mathscr{L}(x,\lambda) = f(x) - \sum_{i\in\mathscr{E}\cup\mathscr{I}} \lambda_i c_i(x). \tag{4}$$

- The necessary conditions defined in the following theorem are called first-order conditions.
- They are named so owing to their association with gradients (first-derivative vectors) of the objective and constraint functions.

FIRST-ORDER OPTIMALITY CONDITIONS

Consider the constrained optimisation problem

$$\min_{x\in\mathbb{R}^n} f(x) \qquad \text{subject to} \qquad \begin{cases} c_i(x) & = 0, \quad i \in \mathscr{E} \\ c_j(x) & \geq 0, \quad j \in \mathscr{I} \end{cases} \tag{3}$$

f and c_i are scalar valued functions of the vector of unknowns x and $\mathscr{E}$ and $\mathscr{I}$ are set of indices. Define the Lagrangian function for the general problem as

$$\mathscr{L}(x,\lambda) = f(x) - \sum_{i\in\mathscr{E}\cup\mathscr{I}} \lambda_i c_i(x). \tag{4}$$

- The necessary conditions defined in the following theorem are called first-order conditions.
- They are named so owing to their association with gradients (first-derivative vectors) of the objective and constraint functions.
- They act as a foundation for many of the algorithms.

First-Order Necessary Conditions

Theorem

Suppose that x^* is a local solution of the optimisation problem (3), that the functions f and c_i's in (3) are continuously differentiable, and that the LICQ holds at x^*. Then there is a Lagrange multiplier vector λ^*, with components λ_i^*, $i \in \mathscr{E} \cup \mathscr{I}$, such that the following conditions are satisfied at (x^*, λ^*)

$$\begin{aligned} \nabla_x \mathscr{L}(x^*, \lambda^*) &= 0, && && (5)\\ c_i(x^*) &= 0, && \text{for all } i \in \mathscr{E}, && (6)\\ c_i(x^*) &\geq 0, && \text{for all } i \in \mathscr{I}, && (7)\\ \lambda_i^* &\geq 0, && \text{for all } i \in \mathscr{I}, && (8)\\ \lambda_i^* c_i(x^*) &= 0, && \text{for all } i \in \mathscr{E} \cup \mathscr{I}. && (9) \end{aligned}$$

First-Order Necessary Conditions

- The above stated conditions are often known as the *Karush–Kuhn–Tucker* conditions, or *KKT* conditions for short.

First-Order Necessary Conditions

- The above stated conditions are often known as the *Karush–Kuhn–Tucker* conditions, or *KKT* conditions for short.
- The last set of conditions comprises of conditions that are the complementarity conditions; they imply that either constraint i is active or $\lambda_i^* = 0$, or possibly both.

First-Order Necessary Conditions

- The above stated conditions are often known as the *Karush–Kuhn–Tucker* conditions, or *KKT* conditions for short.
- The last set of conditions comprises of conditions that are the complementarity conditions; they imply that either constraint i is active or $\lambda_i^* = 0$, or possibly both.
- The Lagrange multipliers corresponding to inactive inequality constraints are zero.

First-Order Necessary Conditions

- The above stated conditions are often known as the *Karush–Kuhn–Tucker* conditions, or *KKT* conditions for short.
- The last set of conditions comprises of conditions that are the complementarity conditions; they imply that either constraint i is active or $\lambda_i^* = 0$, or possibly both.
- The Lagrange multipliers corresponding to inactive inequality constraints are zero.
- We can omit the terms for indices $i \notin \mathscr{A}(x^*)$ and rewrite the first condition as

First-Order Necessary Conditions

- The above stated conditions are often known as the *Karush–Kuhn–Tucker* conditions, or *KKT* conditions for short.
- The last set of conditions comprises of conditions that are the complementarity conditions; they imply that either constraint i is active or $\lambda_i^* = 0$, or possibly both.
- The Lagrange multipliers corresponding to inactive inequality constraints are zero.
- We can omit the terms for indices $i \notin \mathscr{A}(x^*)$ and rewrite the first condition as

$$0 = \nabla_x \mathscr{L}(x^*, \lambda^*) = \nabla f(x^*) - \sum_{i \in \mathscr{A}(x^*)} \lambda_i^* \nabla c_i(x^*). \tag{10}$$

Strict Complementarity

Definition

Given a local solution x^* of the optimisation problem and a vector λ^* satisfying the KKT conditions, we say that the *strict complementarity condition* holds if exactly one of λ_i^* and $c_i(x^*)$ is zero for each index $i \in \mathscr{I}$. In other words, we have that $\lambda_i^* > 0$ for each $i \in \mathscr{I} \cup \mathscr{A}(x^*)$.

Strict Complementarity

Definition

Given a local solution x^* of the optimisation problem and a vector λ^* satisfying the KKT conditions, we say that the *strict complementarity condition* holds if exactly one of λ_i^* and $c_i(x^*)$ is zero for each index $i \in \mathscr{I}$. In other words, we have that $\lambda_i^* > 0$ for each $i \in \mathscr{I} \cup \mathscr{A}(x^*)$.

- Satisfaction of the strict complementarity property usually makes it easier for algorithms to determine the active set $\mathscr{A}(x^*)$ and converge rapidly to the solution x^*.

Strict Complementarity

Definition

Given a local solution x^* of the optimisation problem and a vector λ^* satisfying the KKT conditions, we say that the *strict complementarity condition* holds if exactly one of λ_i^* and $c_i(x^*)$ is zero for each index $i \in \mathscr{I}$. In other words, we have that $\lambda_i^* > 0$ for each $i \in \mathscr{I} \cup \mathscr{A}(x^*)$.

- Satisfaction of the strict complementarity property usually makes it easier for algorithms to determine the active set $\mathscr{A}(x^*)$ and converge rapidly to the solution x^*.
- For a given problem and solution point x^*, there may be many vectors λ^* for which the KKT conditions are satisfied.

Strict Complementarity

Definition

Given a local solution x^* of the optimisation problem and a vector λ^* satisfying the KKT conditions, we say that the *strict complementarity condition* holds if exactly one of λ_i^* and $c_i(x^*)$ is zero for each index $i \in \mathscr{I}$. In other words, we have that $\lambda_i^* > 0$ for each $i \in \mathscr{I} \cup \mathscr{A}(x^*)$.

- Satisfaction of the strict complementarity property usually makes it easier for algorithms to determine the active set $\mathscr{A}(x^*)$ and converge rapidly to the solution x^*.
- For a given problem and solution point x^*, there may be many vectors λ^* for which the KKT conditions are satisfied.
- When the LICQ holds, however, the optimal λ^* is unique.

KKT Conditions With an Example

- Consider the feasible region illustrated in the figure below described by the four constraints of the ensuing optimization problem.

KKT Conditions With an Example

- Consider the feasible region illustrated in the figure below described by the four constraints of the ensuing optimization problem.

$$\min_x \left(x_1 - \frac{3}{2}\right)^2 + \left(x_2 - \frac{1}{2}\right)^2 \quad \text{s.t.} \quad \begin{bmatrix} 1 - x_1 - x_2 \\ 1 - x_1 + x_2 \\ 1 + x_1 - x_2 \\ 1 + x_1 + x_2 \end{bmatrix} \geq 0. \quad (11)$$

KKT Conditions With an Example

- Consider the feasible region illustrated in the figure below described by the four constraints of the ensuing optimization problem.

$$\min_x \left(x_1 - \frac{3}{2}\right)^2 + \left(x_2 - \frac{1}{2}\right)^2 \quad \text{s.t.} \quad \begin{bmatrix} 1 - x_1 - x_2 \\ 1 - x_1 + x_2 \\ 1 + x_1 - x_2 \\ 1 + x_1 + x_2 \end{bmatrix} \geq 0. \quad (11)$$

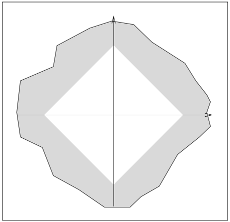

Figure: Four constraints

KKT Conditions With an Example

- It can be shown that the solution is $x^* = (1, 0)^T$.

KKT Conditions With an Example

- It can be shown that the solution is $x^* = (1, 0)^T$.
- The first and second constraints are active at this point.

KKT Conditions With an Example

- It can be shown that the solution is $x^* = (1,0)^T$.
- The first and second constraints are active at this point.
- Denoting them by c_1 and c_2 (and the inactive constraints by c_3 and c_4).

KKT Conditions With an Example

- It can be shown that the solution is $x^* = (1, 0)^T$.
- The first and second constraints are active at this point.
- Denoting them by c_1 and c_2 (and the inactive constraints by c_3 and c_4).
- To verify the KKT conditions compute

$$\nabla f(x^*) = \begin{bmatrix} -1 \\ -\frac{1}{2} \end{bmatrix},$$

KKT Conditions With an Example

- It can be shown that the solution is $x^* = (1, 0)^T$.
- The first and second constraints are active at this point.
- Denoting them by c_1 and c_2 (and the inactive constraints by c_3 and c_4).
- To verify the KKT conditions compute

$$\nabla f(x^*) = \begin{bmatrix} -1 \\ -\frac{1}{2} \end{bmatrix}, \quad \nabla c_1(x^*) = \begin{bmatrix} -1 \\ -1 \end{bmatrix},$$

KKT Conditions With an Example

- It can be shown that the solution is $x^* = (1, 0)^T$.
- The first and second constraints are active at this point.
- Denoting them by c_1 and c_2 (and the inactive constraints by c_3 and c_4).
- To verify the KKT conditions compute

$$\nabla f(x^*) = \begin{bmatrix} -1 \\ -\frac{1}{2} \end{bmatrix}, \quad \nabla c_1(x^*) = \begin{bmatrix} -1 \\ -1 \end{bmatrix}, \quad \nabla c_2(x^*) = \begin{bmatrix} -1 \\ 1 \end{bmatrix}.$$

KKT Conditions With an Example

- It can be shown that the solution is $x^* = (1, 0)^T$.
- The first and second constraints are active at this point.
- Denoting them by c_1 and c_2 (and the inactive constraints by c_3 and c_4).
- To verify the KKT conditions compute

$$\nabla f(x^*) = \begin{bmatrix} -1 \\ -\frac{1}{2} \end{bmatrix}, \quad \nabla c_1(x^*) = \begin{bmatrix} -1 \\ -1 \end{bmatrix}, \quad \nabla c_2(x^*) = \begin{bmatrix} -1 \\ 1 \end{bmatrix}.$$

- Therefore, the KKT conditions are satisfied when we set

$$\lambda^* = (\frac{3}{4}, \frac{1}{4}, 0, 0)^T.$$

FIRST-ORDER OPTIMALITY CONDITIONS

Lemma

Let x^* be a feasible point. The following two statements are true.

1. $T_\Omega(x^*) \subset \mathscr{F}(x^*)$.

FIRST-ORDER OPTIMALITY CONDITIONS

Lemma

Let x^* be a feasible point. The following two statements are true.

1. $T_\Omega(x^*) \subset \mathscr{F}(x^*)$.
2. If the LICQ condition is satisfied at x^*, $T_\Omega(x^*) = \mathscr{F}(x^*)$.

The above Lemma uses a constraint qualification (LICQ) to relate the tangent cone T_Ω to the set $\mathscr{F}$ of first-order feasible directions.

A FUNDAMENTAL NECESSARY CONDITION

Definition (Local Solution)

A local solution of the optimisation problem is a point x at which all feasible sequences have the property that $f(z_k) \geq f(x)$ for all k sufficiently large.

A FUNDAMENTAL NECESSARY CONDITION

Definition (Local Solution)

A local solution of the optimisation problem is a point x at which all feasible sequences have the property that $f(z_k) \geq f(x)$ for all k sufficiently large.

Theorem

If x^* is a local solution of the optimization problem (3), then we have

A FUNDAMENTAL NECESSARY CONDITION

Definition (Local Solution)

A local solution of the optimisation problem is a point x at which all feasible sequences have the property that $f(z_k) \geq f(x)$ for all k sufficiently large.

Theorem

If x^* is a local solution of the optimization problem (3), then we have

$$\nabla f(x^*)^T d \geq 0, \quad \text{for all } d \in T_\Omega(x^*) \tag{12}$$

A FUNDAMENTAL NECESSARY CONDITION

Definition (Local Solution)

A local solution of the optimisation problem is a point x at which all feasible sequences have the property that $f(z_k) \geq f(x)$ for all k sufficiently large.

Theorem

If x^* is a local solution of the optimization problem (3), then we have

$$\nabla f(x^*)^T d \geq 0, \quad \text{for all } d \in T_\Omega(x^*) \tag{12}$$

- Therefore the theorem says if a sequence z_k as considered above exists, then its limiting directions must make a non-negative inner product with the gradient of the objective function.

A FUNDAMENTAL NECESSARY CONDITION

Proof of Theorem

- To contradict lets assume that there is a tangent d for which $\nabla f(x^*)^T d < 0$.

A FUNDAMENTAL NECESSARY CONDITION

Proof of Theorem

- To contradict lets assume that there is a tangent d for which $\nabla f(x^*)^T d < 0$.
- Let $\{z_k\}$ and $\{t_k\}$ be the sequences satisfying definition of tangent vector for this d.

A FUNDAMENTAL NECESSARY CONDITION

Proof of Theorem

- To contradict lets assume that there is a tangent d for which $\nabla f(x^*)^T d < 0$.
- Let $\{z_k\}$ and $\{t_k\}$ be the sequences satisfying definition of tangent vector for this d.
- Then we have:

$$\lim_{k \to \infty} \frac{z_k - x^*}{t_k} = d$$

$$z_k = x^* + t_k d + o(t_k).$$

. for k sufficiently large.

A FUNDAMENTAL NECESSARY CONDITION

Proof of Theorem

- To contradict lets assume that there is a tangent d for which $\nabla f(x^*)^T d < 0$.
- Let $\{z_k\}$ and $\{t_k\}$ be the sequences satisfying definition of tangent vector for this d.
- Then we have:
$$\lim_{k\to\infty} \frac{z_k - x^*}{t_k} = d$$
$$z_k = x^* + t_k d + o(t_k).$$
. for k sufficiently large.
- We have:
$$\begin{aligned} f(z_k) &= f(x^*) + (z_k - x^*)^T \nabla f(x^*) + o(||z_k - x^*||) \\ &= f(x^*) + t_k d^T \nabla f(x^*) + o(t_k) \end{aligned}$$

A FUNDAMENTAL NECESSARY CONDITION

Proof of Theorem

- Since $d^T \nabla f(x^*) < 0$, and the remainder term eventually gets dominated by the first-order term we have

$$f(z_k) < f(x^*) + \frac{1}{2} t_k d^T \nabla f(x^*), \quad \text{for all } k \text{ sufficiently large.}$$

A FUNDAMENTAL NECESSARY CONDITION

Proof of Theorem

- Since $d^T \nabla f(x^*) < 0$, and the remainder term eventually gets dominated by the first-order term we have

$$f(z_k) < f(x^*) + \frac{1}{2} t_k d^T \nabla f(x^*), \quad \text{for all } k \text{ sufficiently large.}$$

- This implies given an open nbhd of x^*, a k sufficiently large can be chosen such that z_k lies in this nbhd and has a lower value lower value of the objective f .

A FUNDAMENTAL NECESSARY CONDITION

Proof of Theorem

- Since $d^T \nabla f(x^*) < 0$, and the remainder term eventually gets dominated by the first-order term we have

$$f(z_k) < f(x^*) + \frac{1}{2} t_k d^T \nabla f(x^*), \quad \text{for all } k \text{ sufficiently large.}$$

- This implies given an open nbhd of x^*, a k sufficiently large can be chosen such that z_k lies in this nbhd and has a lower value lower value of the objective f .
- Therefore, x^* is not a local solution.

Is the Converse True?

- The converse of this result is not necessarily true.

Is the Converse True?

- The converse of this result is not necessarily true.
- We may have $\nabla f(x^*)^T d \geq 0$ for all $d \in T_\Omega(x^*)$, yet x^* not being a local minimiser.

Is the Converse True?

- The converse of this result is not necessarily true.
- We may have $\nabla f(x^*)^T d \geq 0$ for all $d \in T_\Omega(x^*)$, yet x^* not being a local minimiser.
- Consider the problem

$$\min\ x_2 \qquad \text{subject to } x_2 \geq -x_1^2.$$

- The problem is unbounded.

Is the Converse True?

- The converse of this result is not necessarily true.
- We may have $\nabla f(x^*)^T d \geq 0$ for all $d \in T_\Omega(x^*)$, yet x^* not being a local minimiser.
- Consider the problem

$$\min\ x_2 \qquad \text{subject to } x_2 \geq -x_1^2.$$

- The problem is unbounded.
- Let us examine its behavior at $x^* = (0,0)^T$.

Is the Converse True?

- The converse of this result is not necessarily true.
- We may have $\nabla f(x^*)^T d \geq 0$ for all $d \in T_\Omega(x^*)$, yet x^* not being a local minimiser.
- Consider the problem

$$\min\ x_2 \qquad \text{subject to } x_2 \geq -x_1^2.$$

- The problem is unbounded.
- Let us examine its behavior at $x^* = (0,0)^T$.
- All limiting directions d of feasible sequences must have $d_2 \geq 0$, so that $\nabla f(x^*)^T d = d_2 \geq 0$.

Is the Converse True?

- The converse of this result is not necessarily true.
- We may have $\nabla f(x^*)^T d \geq 0$ for all $d \in T_\Omega(x^*)$, yet x^* not being a local minimiser.
- Consider the problem

$$\min\ x_2 \qquad \text{subject to } x_2 \geq -x_1^2.$$

- The problem is unbounded.
- Let us examine its behavior at $x^* = (0,0)^T$.
- All limiting directions d of feasible sequences must have $d_2 \geq 0$, so that $\nabla f(x^*)^T d = d_2 \geq 0$.
- x^* is clearly not a local minimiser.

Is the Converse True?

- The converse of this result is not necessarily true.
- We may have $\nabla f(x^*)^T d \geq 0$ for all $d \in T_\Omega(x^*)$, yet x^* not being a local minimiser.
- Consider the problem

$$\min \ x_2 \qquad \text{subject to } x_2 \geq -x_1^2.$$

- The problem is unbounded.
- Let us examine its behavior at $x^* = (0, 0)^T$.
- All limiting directions d of feasible sequences must have $d_2 \geq 0$, so that $\nabla f(x^*)^T d = d_2 \geq 0$.
- x^* is clearly not a local minimiser.
- The point $(\alpha, -\alpha^2)^T$ for $\alpha > 0$ has a smaller function value than x^*, and can be brought arbitrarily close to x^* by setting α sufficiently small.

Is the Converse True?

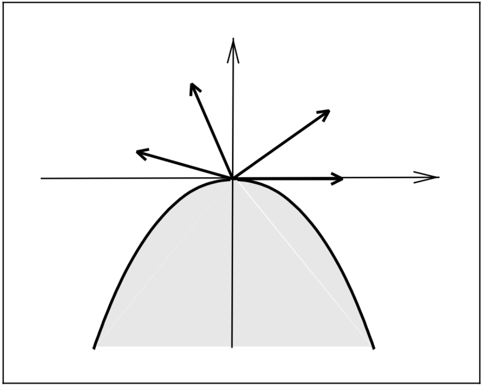

Figure: showing various limiting directions of feasible sequences at the point $(0,0)^T$.

FARKAS' LEMMA

- The most important step in proving the KKT theorem.

FARKAS' LEMMA

- The most important step in proving the KKT theorem.
- This lemma considers a cone K defined as follows:

$$K = \{By + Cw \mid y \geq 0\}, \tag{13}$$

  where B and C are matrices of dimension $n \times m$ and $n \times p$, respectively, and y and w are vectors of appropriate dimensions.
- Given $g \in \mathbb{R}^n$, Farkas' Lemmma states that one (and only one) of the two alternatives is true.
  1. Either $g \in K$, or else

FARKAS' LEMMA

- The most important step in proving the KKT theorem.
- This lemma considers a cone K defined as follows:

$$K = \{By + Cw \mid y \geq 0\}, \tag{13}$$

  where B and C are matrices of dimension $n \times m$ and $n \times p$, respectively, and y and w are vectors of appropriate dimensions.
- Given $g \in \mathbb{R}^n$, Farkas' Lemmma states that one (and only one) of the two alternatives is true.
  1. Either $g \in K$, or else
  2. there is a vector $d \in \mathbb{R}^n$ such that

$$g^T d < 0, \quad B^T d \geq 0, \quad C^T d = 0. \tag{14}$$

FARKAS' LEMMA

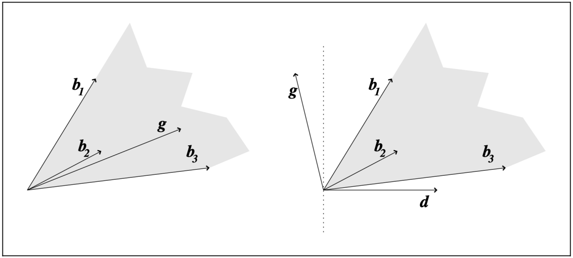

Figure: Farkas' Lemma: Either $g \in L$ (left) or there is a separating hyperplane (right).

- In the above figure B has three columns, C is null and $n = 2$.

FARKAS' LEMMA

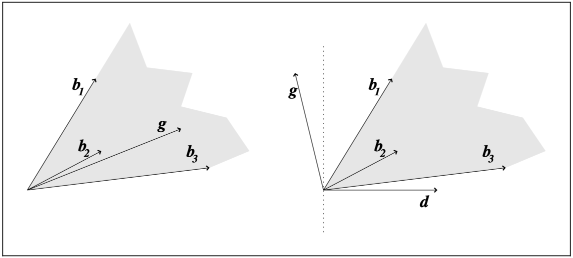

Figure: Farkas' Lemma: Either $g \in L$ (left) or there is a separating hyperplane (right).

- In the above figure B has three columns, C is null and $n = 2$.
- Note that in the second case, the vector d defines a *separating hyperplane*, which is a plane in $\mathbb{R}^n$ that separates the vector g from the cone K.

FARKAS' LEMMA

Farkas' Lemma

Let the cone K be defined as above. Given any vector $g \in \mathbb{R}^n$, we have either that $g \in K$ or that there exist $d \in \mathbb{R}^n$ satisfying (14), but not both.

Proof of First-Order Necessary Conditions (KKT)

- Suppose that $x^* \in \mathbb{R}^n$ is a feasible point at which the LICQ holds.

Proof of First-Order Necessary Conditions (KKT)

- Suppose that $x^* \in \mathbb{R}^n$ is a feasible point at which the LICQ holds.
- The theorem claims that if x^* is a solution for the optimisation problem, then there is a vector $\lambda^* \in \mathbb{R}^m$ that satisfies the KKT conditions.

Proof of First-Order Necessary Conditions (KKT)

- Suppose that $x^* \in \mathbb{R}^n$ is a feasible point at which the LICQ holds.
- The theorem claims that if x^* is a solution for the optimisation problem, then there is a vector $\lambda^* \in \mathbb{R}^m$ that satisfies the KKT conditions.
- We first show that there are multipliers λ_i, $i \in \mathscr{A}(x^*)$, such that the following is satisfied:

$$\nabla f(x^*) = \sum_{i \in \mathscr{A}(x^*)} \lambda_i \nabla c_i(x^*)$$

Proof of First-Order Necessary Conditions (KKT)

- Suppose that $x^* \in \mathbb{R}^n$ is a feasible point at which the LICQ holds.
- The theorem claims that if x^* is a solution for the optimisation problem, then there is a vector $\lambda^* \in \mathbb{R}^m$ that satisfies the KKT conditions.
- We first show that there are multipliers λ_i, $i \in \mathscr{A}(x^*)$, such that the following is satisfied:

$$\nabla f(x^*) = \sum_{i \in \mathscr{A}(x^*)} \lambda_i \nabla c_i(x^*)$$

- We have from the previous theorem

$$d^T \nabla f(x^*) \geq 0, \quad \text{for all tangent vectors } d \in T_\Omega(x^*).$$

Proof of First-Order Necessary Conditions (KKT)

- Suppose that $x^* \in \mathbb{R}^n$ is a feasible point at which the LICQ holds.
- The theorem claims that if x^* is a solution for the optimisation problem, then there is a vector $\lambda^* \in \mathbb{R}^m$ that satisfies the KKT conditions.
- We first show that there are multipliers λ_i, $i \in \mathscr{A}(x^*)$, such that the following is satisfied:

$$\nabla f(x^*) = \sum_{i \in \mathscr{A}(x^*)} \lambda_i \nabla c_i(x^*)$$

- We have from the previous theorem

$$d^T \nabla f(x^*) \geq 0, \quad \text{for all tangent vectors } d \in T_\Omega(x^*).$$

- We also have the equivalence of $\mathscr{F}(x^*)$ and $T_\Omega(x^*)$, whenever LICQ holds

- But putting together the above two results we have

$$d^T \nabla f(x^*) \geq 0 \quad \text{for all } d \in \mathscr{F}(x^*).$$

- But putting together the above two results we have

$$d^T \nabla f(x^*) \geq 0 \quad \text{for all } d \in \mathscr{F}(x^*).$$

- Consider the cone N defined by:

$$N = \{ \sum_{i \in \mathscr{A}(x^*)} \lambda_i \nabla c_i(x^*), \quad \lambda_i \geq 0 \text{ for } i \in \mathscr{A}(x^*) \cap \mathscr{I} \} \quad (15)$$

- But putting together the above two results we have

$$d^T \nabla f(x^*) \geq 0 \quad \text{for all } d \in \mathscr{F}(x^*).$$

- Consider the cone N defined by:

$$N = \{ \sum_{i \in \mathscr{A}(x^*)} \lambda_i \nabla c_i(x^*), \quad \lambda_i \geq 0 \text{ for } i \in \mathscr{A}(x^*) \cap \mathscr{I} \} \quad (15)$$

- Set $g = \nabla f(x^*)$.

- But putting together the above two results we have

$$d^T \nabla f(x^*) \geq 0 \quad \text{for all } d \in \mathscr{F}(x^*).$$

- Consider the cone N defined by:

$$N = \{ \sum_{i \in \mathscr{A}(x^*)} \lambda_i \nabla c_i(x^*), \quad \lambda_i \geq 0 \text{ for } i \in \mathscr{A}(x^*) \cap \mathscr{I} \} \tag{15}$$

- Set $g = \nabla f(x^*)$.
- Now, Farkas' Lemma implies either

$$\nabla f(x^*) = \sum_{i \in \mathscr{A}(x^*)} \lambda_i \nabla c_i(x^*) = A(x^*)^T \lambda^*, \quad \lambda_i \geq 0 \text{ for } i \in \mathscr{A}(x^*) \cap \mathscr{I} \tag{16}$$

- But putting together the above two results we have

$$d^T \nabla f(x^*) \geq 0 \quad \text{for all } d \in \mathscr{F}(x^*).$$

- Consider the cone N defined by:

$$N = \{ \sum_{i \in \mathscr{A}(x^*)} \lambda_i \nabla c_i(x^*), \quad \lambda_i \geq 0 \text{ for } i \in \mathscr{A}(x^*) \cap \mathscr{I} \} \quad (15)$$

- Set $g = \nabla f(x^*)$.
- Now, Farkas' Lemma implies either

$$\nabla f(x^*) = \sum_{i \in \mathscr{A}(x^*)} \lambda_i \nabla c_i(x^*) = A(x^*)^T \lambda^*, \quad \lambda_i \geq 0 \text{ for } i \in \mathscr{A}(x^*) \cap \mathscr{I} \quad (16)$$

- or else there is a direction d such that $d^T \nabla f(x^*) < 0$ and $d \in \mathscr{F}(x^*)$.

- We have as a consequence of the previously stated results, that (16) holds true.

- We have as a consequence of the previously stated results, that (16) holds true.
- We now define the vector λ^* as

$$\lambda_i^* = \begin{cases} \lambda_i, & i \in \mathscr{A}(x^*), \\ 0, & i \in \mathscr{I} \setminus \mathscr{A}(x^*), \end{cases} \tag{17}$$

and show that the this choice of λ^*, together with out local solution x^*, satisfies the KKT conditions.

- We have as a consequence of the previously stated results, that (16) holds true.
- We now define the vector λ^* as

$$\lambda_i^* = \begin{cases} \lambda_i, & i \in \mathscr{A}(x^*), \\ 0, & i \in \mathscr{I} \setminus \mathscr{A}(x^*), \end{cases} \tag{17}$$

  and show that the this choice of λ^*, together with out local solution x^*, satisfies the KKT conditions.
- The stationary point condition for the Lagrangian function follows immediately from (16) and the definitions of Lagrangian function and the definition of λ^* above.

- We have as a consequence of the previously stated results, that (16) holds true.
- We now define the vector λ^* as

$$\lambda_i^* = \begin{cases} \lambda_i, & i \in \mathscr{A}(x^*), \\ 0, & i \in \mathscr{I} \setminus \mathscr{A}(x^*), \end{cases} \tag{17}$$

  and show that the this choice of λ^*, together with out local solution x^*, satisfies the KKT conditions.
- The stationary point condition for the Lagrangian function follows immediately from (16) and the definitions of Lagrangian function and the definition of λ^* above.
- Since x^* is feasible, the two feasibility conditions are satisfied.

- We have as a consequence of the previously stated results, that (16) holds true.
- We now define the vector λ^* as

$$\lambda_i^* = \begin{cases} \lambda_i, & i \in \mathscr{A}(x^*), \\ 0, & i \in \mathscr{I} \setminus \mathscr{A}(x^*), \end{cases} \tag{17}$$

  and show that the this choice of λ^*, together with out local solution x^*, satisfies the KKT conditions.
- The stationary point condition for the Lagrangian function follows immediately from (16) and the definitions of Lagrangian function and the definition of λ^* above.
- Since x^* is feasible, the two feasibility conditions are satisfied.
- $\lambda_i^* \geq 0$ for $i \in \mathscr{A}(x^*) \cap \mathscr{I}$, while from the definition of λ^*, $\lambda_i^* = 0$ for $i \in \mathscr{I} \setminus \mathscr{A}(x^*)$.

- Hence, $\lambda_i^* \geq 0$ for $i \in \mathscr{I}$.

- Hence, $\lambda_i^* \geq 0$ for $i \in \mathscr{I}$.
- We have for $i \in \mathscr{A}(x^*) \cap \mathscr{I}$ that $c_i(x^*) = 0$, while for $i \in \mathscr{I} \setminus \mathscr{A}(x^*)$, we have $\lambda_i^* = 0$.

- Hence, $\lambda_i^* \geq 0$ for $i \in \mathscr{I}$.
- We have for $i \in \mathscr{A}(x^*) \cap \mathscr{I}$ that $c_i(x^*) = 0$, while for $i \in \mathscr{I} \setminus \mathscr{A}(x^*)$, we have $\lambda_i^* = 0$.
- Hence $\lambda_i^* c_i(x^*) = 0$, for $i \in \mathscr{I}$.

SECOND-ORDER CONDITIONS

- The KKT conditions tell us how the first derivatives of f and the active constraints c_i are related to each other at a solution x^*.

SECOND-ORDER CONDITIONS

- The KKT conditions tell us how the first derivatives of f and the active constraints c_i are related to each other at a solution x^*.
- When these conditions are satisfied, any movement along any vector $w \in \mathscr{F}(x^*)$ either increases the first-order approximation to the objective function ($\nabla f(x^*)^T w > 0$) or else keeps this value the same ($\nabla f(x^*)^T w = 0$).

SECOND-ORDER CONDITIONS

- The KKT conditions tell us how the first derivatives of f and the active constraints c_i are related to each other at a solution x^*.
- When these conditions are satisfied, any movement along any vector $w \in \mathscr{F}(x^*)$ either increases the first-order approximation to the objective function ($\nabla f(x^*)^T w > 0$) or else keeps this value the same ($\nabla f(x^*)^T w = 0$).
- For the directions $w \in \mathscr{F}(x^*)$ for which $\nabla f(x^*)^T w = 0$ one cannot determine from first derivative information alone whether a move along this direction will increase or decrease the objective function f.

SECOND-ORDER CONDITIONS

- The KKT conditions tell us how the first derivatives of f and the active constraints c_i are related to each other at a solution x^*.
- When these conditions are satisfied, any movement along any vector $w \in \mathscr{F}(x^*)$ either increases the first-order approximation to the objective function $(\nabla f(x^*)^T w > 0)$ or else keeps this value the same $(\nabla f(x^*)^T w = 0)$.
- For the directions $w \in \mathscr{F}(x^*)$ for which $\nabla f(x^*)^T w = 0$ one cannot determine from first derivative information alone whether a move along this direction will increase or decrease the objective function f.
- Second derivatives play a "tiebreaking" role.

SECOND-ORDER CONDITIONS

- The second derivative terms in the Taylor series expansions of f and c_i are examined by the second-order conditions.

SECOND-ORDER CONDITIONS

- The second derivative terms in the Taylor series expansions of f and c_i are examined by the second-order conditions.
- The approach is to see whether this extra information resolves the issue of increase or decrease in f.

SECOND-ORDER CONDITIONS

- The second derivative terms in the Taylor series expansions of f and c_i are examined by the second-order conditions.
- The approach is to see whether this extra information resolves the issue of increase or decrease in f.
- These conditions are concerned with the curvature of the Lagrangian function in the "undecided" directions ($w \in \mathscr{F}(x^*)$ for which $\nabla f(x^*)^T w = 0$).

SECOND-ORDER CONDITIONS

- The second derivative terms in the Taylor series expansions of f and c_i are examined by the second-order conditions.
- The approach is to see whether this extra information resolves the issue of increase or decrease in f.
- These conditions are concerned with the curvature of the Lagrangian function in the "undecided" directions ($w \in \mathscr{F}(x^*)$ for which $\nabla f(x^*)^T w = 0$).
- For second derivatives stronger smoothness assumptions are needed, f and c_i, $i \in \mathscr{I} \cup \mathscr{E}$, are all assumed to be twice continuously differentiable.

SECOND-ORDER CONDITIONS

Definition

Given $\mathscr{F}(x^*)$ and some Lagrange multiplier vector λ^* satisfying the KKT conditions, we define the critical cone $\mathscr{C}(x^*, \lambda^*)$ as follows:

$$\begin{aligned}\mathscr{C}(x^*, \lambda^*) = \{w \in \mathscr{F}(x^*) | \nabla c_i(x^*)^T w = 0, \ &\text{for all} \\ i \in \mathscr{A}(x^*) \cap \mathscr{I} \ &\text{with } \lambda_i^* > 0\}\end{aligned}$$

Equivalently,

$$w \in \mathscr{C}(x^*, \lambda^*) \Leftrightarrow$$
$$\begin{cases} \nabla c_i(x^*)^T w = 0, & \text{for all } i \in \mathscr{E}, \\ \nabla c_i(x^*)^T w = 0, & \text{for all } i \in \mathscr{A}(x^*) \cap \mathscr{I} \text{ with } \lambda_i^* > 0, \\ \nabla c_i(x^*)^T w \geq 0, & \text{for all } i \in \mathscr{A}(x^*) \cap \mathscr{I} \text{ with } \lambda_i^* = 0. \end{cases}$$

SECOND-ORDER CONDITIONS

- From the above definition, and the fact that $\lambda_i^* = 0$ for all inactive components $i \in \mathscr{I} \setminus \mathscr{A}(x^*)$, it follows that

$$w \in \mathscr{C}(x^*, \lambda^*) \implies \lambda_i^* \nabla c_i(x^*)^T w = 0, \text{ for all } i \in \mathscr{E} \cup \mathscr{I}.$$

SECOND-ORDER CONDITIONS

- From the above definition, and the fact that $\lambda_i^* = 0$ for all inactive components $i \in \mathscr{I} \setminus \mathscr{A}(x^*)$, it follows that

$$w \in \mathscr{C}(x^*, \lambda^*) \implies \lambda_i^* \nabla c_i(x^*)^T w = 0, \text{ for all } i \in \mathscr{E} \cup \mathscr{I}.$$

- Now from the first KKT condition and from the definition of the Lagrangian function, we have

$$w \in \mathscr{C}(x^*, \lambda^*) \implies w^T \nabla f(x^*) = \sum_{i \in \mathscr{E} \cup \mathscr{I}} \lambda_i^* w^T \nabla c_i(x^*) = 0.$$

SECOND-ORDER CONDITIONS

- From the above definition, and the fact that $\lambda_i^* = 0$ for all inactive components $i \in \mathscr{I} \setminus \mathscr{A}(x^*)$, it follows that

$$w \in \mathscr{C}(x^*, \lambda^*) \implies \lambda_i^* \nabla c_i(x^*)^T w = 0, \text{ for all } i \in \mathscr{E} \cup \mathscr{I}.$$

- Now from the first KKT condition and from the definition of the Lagrangian function, we have

$$w \in \mathscr{C}(x^*, \lambda^*) \implies w^T \nabla f(x^*) = \sum_{i \in \mathscr{E} \cup \mathscr{I}} \lambda_i^* w^T \nabla c_i(x^*) = 0.$$

- Hence the critical cone contains directions from $\mathscr{F}(x^*)$ for which it is not clear from first derivative information alone whether f will increase or decrease.

Second-Order Necessary Conditions

Theorem

Suppose that x^* is a local solution of the optimisation problem and that the LICQ condition is satisfied. Let λ^* be the Lagrange multiplier vector for which the KKT conditions are satisfied. Then

$$w^T \nabla^2_{xx} \mathscr{L}(x^*, \lambda^*) w \geq 0, \quad \text{for all } w \in \mathscr{C}(x^*, \lambda^*). \tag{18}$$

Second-Order Sufficient Conditions

Theorem

Suppose that for some feasible point $x^* \in \mathbb{R}^n$ there is a Lagrange multiplier vector λ^* such that the KKT conditions are satisfied. Suppose also that

$$w^T \nabla^2_{xx} \mathscr{L}(x^*, \lambda^*) w > 0, \qquad \text{for all } w \in \mathscr{C}(x^*, \lambda^*), w \neq 0. \tag{19}$$

Then x^* is a strict local solution for the optimisation problem.

Example

- $f(x) = x_1 + x_2,\ c_1(x) = 2 - x_1^2 - x_2^2$

Example

- $f(x) = x_1 + x_2$, $c_1(x) = 2 - x_1^2 - x_2^2$
- $\mathscr{E} = \phi$, $\mathscr{I} = \{1\}$

Example

- $f(x) = x_1 + x_2$, $c_1(x) = 2 - x_1^2 - x_2^2$
- $\mathscr{E} = \phi$, $\mathscr{I} = \{1\}$
- The Lagrangian is

$$\mathscr{L}(x, \lambda) = (x_1 + x_2) - \lambda_1(2 - x_1^2 - x_2^2),$$

Example

- $f(x) = x_1 + x_2$, $c_1(x) = 2 - x_1^2 - x_2^2$
- $\mathscr{E} = \phi$, $\mathscr{I} = \{1\}$
- The Lagrangian is

$$\mathscr{L}(x, \lambda) = (x_1 + x_2) - \lambda_1(2 - x_1^2 - x_2^2),$$

- It can be verified that the KKT conditions are satisfied at $x^* = (-1, -1)^T$, with $\lambda_1^* = \frac{1}{2}$.

Example

- $f(x) = x_1 + x_2$, $c_1(x) = 2 - x_1^2 - x_2^2$
- $\mathscr{E} = \phi$, $\mathscr{I} = \{1\}$
- The Lagrangian is

$$\mathscr{L}(x, \lambda) = (x_1 + x_2) - \lambda_1(2 - x_1^2 - x_2^2),$$

- It can be verified that the KKT conditions are satisfied at $x^* = (-1, -1)^T$, with $\lambda_1^* = \frac{1}{2}$.
- The Lagrangian Hessian at this point is

$$\nabla_{xx}^2 \mathscr{L}(x^*, \lambda^*) = \begin{bmatrix} 2\lambda_1^* & 0 \\ 0 & 2\lambda_1^* \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$$

Example

- $f(x) = x_1 + x_2$, $c_1(x) = 2 - x_1^2 - x_2^2$
- $\mathscr{E} = \phi$, $\mathscr{I} = \{1\}$
- The Lagrangian is

$$\mathscr{L}(x, \lambda) = (x_1 + x_2) - \lambda_1(2 - x_1^2 - x_2^2),$$

- It can be verified that the KKT conditions are satisfied at $x^* = (-1, -1)^T$, with $\lambda_1^* = \frac{1}{2}$.
- The Lagrangian Hessian at this point is

$$\nabla_{xx}^2 \mathscr{L}(x^*, \lambda^*) = \begin{bmatrix} 2\lambda_1^* & 0 \\ 0 & 2\lambda_1^* \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$$

- This matrix is positive definite, so it certainly satisfies the conditions of the above theorem, $x^* = (-1, -1)^T$ is a strict local solution.