Optimality Conditions

María M. Seron

September 2004

The University of Newcastle Australia

Centre for Complex Dynamic Systems and Control

Outline

1. Unconstrained Optimisation
   - Local and Global Minima
   - Descent Direction
   - Necessary Conditions for a Minimum
   - Necessary and Sufficient Conditions for a Minimum

2. Constrained Optimisation
   - Geometric Necessary Optimality Conditions
   - Problems with Inequality and Equality Constraints
   - The Fritz John Necessary Conditions
   - Karush–Kuhn–Tucker Necessary Conditions
   - Karush–Kuhn–Tucker Sufficient Conditions
   - Quadratic Programs

Unconstrained Optimisation

An unconstrained optimisation problem is a problem of the form

$$\text{minimise } f(x), \tag{1}$$

without any constraint on the vector x.

Definition (Local and Global Minima)

Consider the problem of minimising $f(x)$ over $\mathbb{R}^n$ and let $\bar{x} \in \mathbb{R}^n$.

- *If $f(\bar{x}) \leq f(x)$ for all $x \in \mathbb{R}^n$, then $\bar{x}$ is called a **global minimum**.*
- *If there exists an ε-neighbourhood $N_\varepsilon(\bar{x})$ around $\bar{x}$ such that $f(\bar{x}) \leq f(x)$ for all $x \in N_\varepsilon(\bar{x})$, then $\bar{x}$ is called a **local minimum**.*
- *If $f(\bar{x}) < f(x)$ for all $x \in N_\varepsilon(\bar{x})$, $x \neq \bar{x}$, for some $\varepsilon > 0$, then $\bar{x}$ is called a **strict local minimum**.*

Local and Global Minima

The figure illustrates local and global minima of a function f over the reals.

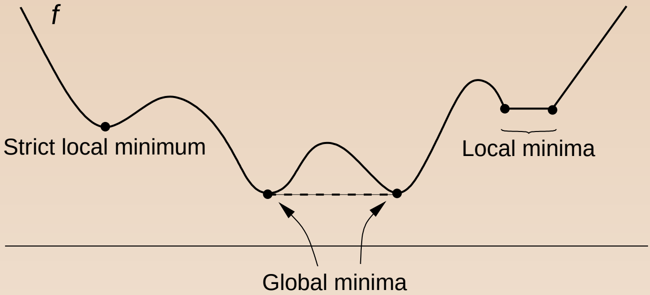


Figure: Local and global minima

Clearly, a global minimum is also a local minimum.

Descent Direction

Given a point $x \in \mathbb{R}^n$, we wish to determine, if possible, whether or not the point is a local or global minimum of a function f.

For differentiable functions, there exist conditions that provide this characterisation, as we will see below.

We start by characterising *descent directions*.

Theorem (Descent Direction)

Let $f : \mathbb{R}^n \to \mathbb{R}$ be differentiable at $\bar{x}$. If there exists a vector d such that

$$\nabla f(\bar{x}) d < 0,$$

then there exists a $\delta > 0$ such that

$$f(\bar{x} + \lambda d) < f(\bar{x}) \quad \text{for each} \quad \lambda \in (0, \delta),$$

*so that d is a **descent direction** of f at $\bar{x}$.*

Descent Direction

Proof.

By the differentiability of f at $\bar{x}$, we have

$$f(\bar{x} + \lambda d) = f(\bar{x}) + \lambda \nabla f(\bar{x}) d + \lambda \|d\| \alpha(\bar{x}, \lambda d),$$

where $\alpha(\bar{x}, \lambda d) \to 0$ as $\lambda \to 0$.

Rearranging and dividing by $\lambda \neq 0$:

$$\frac{f(\bar{x} + \lambda d) - f(\bar{x})}{\lambda} = \nabla f(\bar{x}) d + \|d\| \alpha(\bar{x}, \lambda d).$$

Since $\nabla f(\bar{x}) d < 0$ and $\alpha(\bar{x}, \lambda d) \to 0$ as $\lambda \to 0$, there exists a $\delta > 0$ such that the right hand side above is negative for all $\lambda \in (0, \delta)$. $\square$

Necessary Conditions for a Minimum

We then have a *first-order* necessary condition for a minimum.

Corollary (First Order Necessary Condition for a Minimum)

Suppose that $f : \mathbb{R}^n \to \mathbb{R}$ is differentiable at $\bar{x}$. If $\bar{x}$ is a local minimum, then

$$\nabla f(\bar{x}) = 0.$$

Proof.

Suppose that $\nabla f(\bar{x}) \neq 0$. Then, letting $d = -\nabla f(\bar{x})^T$, we get

$$\nabla f(\bar{x}) d = -\|\nabla f(\bar{x})\|^2 < 0,$$

and by Theorem 2.1 (Descent Direction) there is a $\delta > 0$ such that $f(\bar{x} + \lambda d) < f(\bar{x})$ for each $\lambda \in (0, \delta)$, contradicting the assumption that $\bar{x}$ is a local minimum. Hence, $\nabla f(\bar{x}) = 0$. □

Necessary Conditions for a Minimum

A *second-order* necessary condition for a minimum can be given in terms of the Hessian matrix.

Theorem (Second Order Necessary Condition for a Minimum)

Suppose that $f : \mathbb{R}^n \to \mathbb{R}$ *is twice-differentiable at* $\bar{x}$*. If* $\bar{x}$ *is a local minimum, then*

$$\nabla f(\bar{x}) = 0$$

and

$$H(\bar{x}) \quad \text{is positive semidefinite.}$$

Proof.

Consider an arbitrary direction d. Then, since by assumption f is twice-differentiable at $\bar{x}$, we have

$$f(\bar{x} + \lambda d) = f(\bar{x}) + \lambda \nabla f(\bar{x}) d + \frac{1}{2}\lambda^2 d^{\top} H(\bar{x}) d + \lambda^2 \|d\|^2 \alpha(\bar{x}, \lambda d), \quad (2)$$

where $\alpha(\bar{x}, \lambda d) \to 0$ as $\lambda \to 0$. Since $\bar{x}$ is a local minimum, from Corollary 2.2 we have $\nabla f(\bar{x}) = 0$. Rearranging the terms in (2) and dividing by $\lambda^2 > 0$, we obtain

$$\frac{f(\bar{x} + \lambda d) - f(\bar{x})}{\lambda^2} = \frac{1}{2} d^{\top} H(\bar{x}) d + \|d\|^2 \alpha(\bar{x}, \lambda d)\,. \quad (3)$$

Since $\bar{x}$ is a local minimum, $f(\bar{x} + \lambda d) \geq f(\bar{x})$ for sufficiently small λ. From (3), $\frac{1}{2} d^{\top} H(\bar{x}) d + \|d\|^2 \alpha(\bar{x}, \lambda d) \geq 0$ for sufficiently small λ. By taking the limit as $\lambda \to 0$, it follows that $d^{\top} H(\bar{x}) d \geq 0$; and, hence, $H(\bar{x})$ is positive semidefinite. □

Necessary and Suffi cient Conditions for a Minimum

We now give, without proof, a sufficient condition for a local minimum.

Theorem (Sufficient Condition for a Local Minimum)

Suppose that $f : \mathbb{R}^n \to \mathbb{R}$ is twice-differentiable at $\bar{x}$. If $\nabla f(\bar{x}) = 0$ and $H(\bar{x})$ is positive defi nite, then $\bar{x}$ is a strict local minimum.

As is generally the case with optimisation problems, more powerful results exist under (generalised) convexity conditions.

The next result shows that the necessary condition $\nabla f(\bar{x}) = 0$ is also sufficient for $\bar{x}$ to be a global minimum if f is pseudoconvex at $\bar{x}$.

Theorem (Nec. and Suff. Condition for Pseudoconvex Functions)

Let $f : \mathbb{R}^n \to \mathbb{R}$ be pseudoconvex at $\bar{x}$. Then $\bar{x}$ is a global minimum if and only if $\nabla f(\bar{x}) = 0$.

Constrained Optimisation

We first derive optimality conditions for a problem of the following form:

$$\begin{aligned} &\text{minimise } f(x), \qquad (4)\\ &\text{subject to:}\\ &x \in S. \end{aligned}$$

We will first consider a general constraint set S.

Later, the set S will be more explicitly defined by a set of equality and inequality constraints.

For constrained optimisation problems we have the following definitions.

Feasible and Optimal Solutions

Definition (Feasible and Optimal Solutions)

Let $f : \mathbb{R}^n \to \mathbb{R}$ and consider the constrained optimisation problem (4), where S is a nonempty set in $\mathbb{R}^n$.

- *A point $x \in S$ is called a **feasible solution** to problem (4).*
- *If $\bar{x} \in S$ and $f(x) \geq f(\bar{x})$ for each $x \in S$, then $\bar{x}$ is called an **optimal solution**, a **global optimal solution**, or simply a **solution** to the problem.*
- *The collection of optimal solutions is called the set of **alternative optimal solutions**.*
- *If $\bar{x} \in S$ and if there exists an ε-neighbourhood $N_\varepsilon(\bar{x})$ around $\bar{x}$ such that $f(x) \geq f(\bar{x})$ for each $x \in S \cap N_\varepsilon(\bar{x})$, then $\bar{x}$ is called a **local optimal solution**.*
- *If $\bar{x} \in S$ and if $f(x) > f(\bar{x})$ for each $x \in S \cap N_\varepsilon(\bar{x})$, $x \neq \bar{x}$, for some $\varepsilon > 0$, then $\bar{x}$ is called a **strict local optimal solution**.*

Local and global minima

The figure illustrates examples of local and global minima.

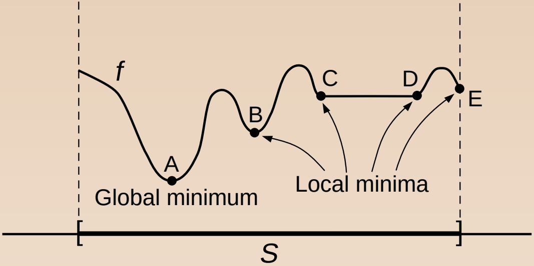


Figure: Local and global minima

The points in S corresponding to A, B and E are also strict local minima, whereas those corresponding to the flat segment of the graph between C and D are local minima that are not strict.

Convex Programs

A *convex program* is a problem of the form

$$\text{minimise } f(x), \tag{5}$$
$$\text{subject to:}$$
$$x \in S.$$

in which the function f and set S are, respectively, a convex function and a convex set.

The following is important property of convex programs.

Theorem (Local Minima of Convex Programs are Global Minima)

Consider problem (5), *where S is a nonempty convex set in $\mathbb{R}^n$, and $f : S \to \mathbb{R}$ is convex on S. If $\bar{x} \in S$ is a local optimal solution to the problem, then $\bar{x}$ is a global optimal solution. Furthermore, if either $\bar{x}$ is a strict local minimum, or if f is strictly convex, then $\bar{x}$ is the unique global optimal solution.*

In this section we give a necessary optimality condition for problem

$$\begin{aligned} &\text{minimise } f(x), \\ &\text{subject to:} \\ &x \in S \end{aligned} \tag{6}$$

using the *cone of feasible directions* defined below.

We do not assume problem (6) to be a convex program.

As a consequence of this generality, only *necessary* conditions for optimality will be derived.

In a later section we will impose suitable convexity conditions to the problem in order to obtain sufficiency conditions for optimality.

Cones of Feasible Directions and of Improving Directions

Definition (Cones of Feasible and Improving Directions)

- *Let S be a nonempty set in $\mathbb{R}^n$ and let $\bar{x} \in \operatorname{cl} S$. The* ***cone of feasible directions*** *of S at $\bar{x}$, denoted by D, is given by*

$$D = \{d : d \neq 0, \text{ and } \bar{x} + \lambda d \in S \text{ for all } \lambda \in (0, \delta) \text{ for some } \delta > 0\}.$$

Each nonzero vector $d \in D$ is called a ***feasible direction****.*

- *Given a function $f : \mathbb{R}^n \to \mathbb{R}$, the* ***cone of improving directions*** *at $\bar{x}$, denoted by F, is given by*

$$F = \{d : f(\bar{x} + \lambda d) < f(\bar{x}) \text{ for all } \lambda \in (0, \delta) \text{ for some } \delta > 0\}.$$

Each direction $d \in F$ is called an ***improving direction****, or a* ***descent direction*** *of f at $\bar{x}$.*

Illustration: Cone of Feasible Directions

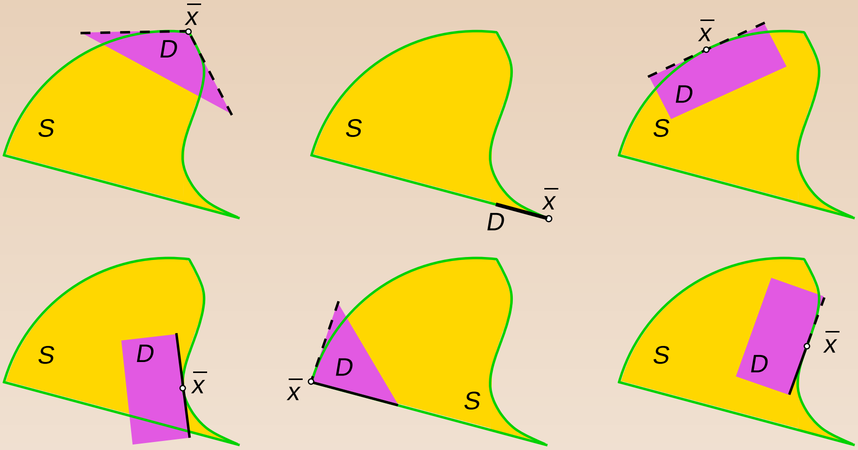
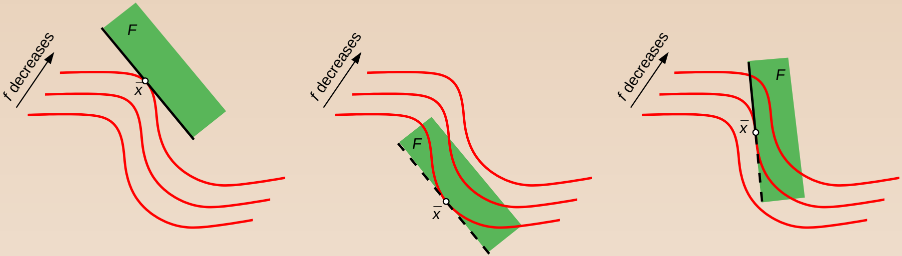


Illustration: Cone of Improving Directions



Algebraic Description of the Cone of Improving Directions

We will now consider the function f to be differentiable at the point $\bar{x}$. We can then define the sets

$$F_0 \triangleq \{d : \nabla f(\bar{x})d < 0\}, \quad (7)$$

$$F_0' \triangleq \{d \neq 0 : \nabla f(\bar{x})d \leq 0\}. \quad (8)$$

From Theorem 2.1 (Descent Direction), if $\nabla f(\bar{x})d < 0$, then d is an improving direction. It then follows that $F_0 \subseteq F$.

Also, if $d \in F$, we must have $\nabla f(\bar{x})d \leq 0$, or else, analogous to Theorem 2.1, $\nabla f(\bar{x})d > 0$ would imply that d is an *ascent direction*.

Hence, we have

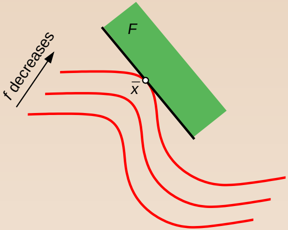
$$F_0 \subseteq F \subseteq F_0'. \quad (9)$$

Algebraic Description of the Cone of Improving Directions

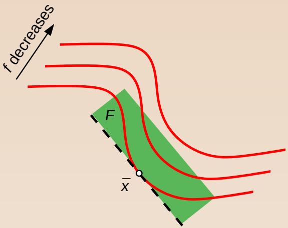
$$F_0 \subseteq F \subseteq F_0'$$

where

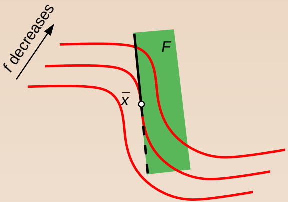
$$F_0 \triangleq \{d : \nabla f(\bar{x})d < 0\} \qquad F_0' \triangleq \{d \neq 0 : \nabla f(\bar{x})d \leq 0\}.$$



$F_0 \subset F = F_0'$



$F_0 = F \subset F_0'$



$F_0 \subset F \subset F_0'$

Geometric Necessary Optimality Conditions

The following theorem states that a necessary condition for local optimality is that every improving direction in F_0 is not a feasible direction.

Theorem (Geometric Necessary Condition for Local Optimality)

Consider the problem to minimise $f(x)$ *subject to* $x \in S$*, where* $f : \mathbb{R}^n \to \mathbb{R}$ *and* S *is a nonempty set in* $\mathbb{R}^n$*. Suppose that* f *is differentiable at a point* $\bar{x} \in S$*. If* $\bar{x}$ *is a local optimal solution then*

$$F_0 \cap D = \emptyset, \tag{10}$$

where $F_0 = \{d : \nabla f(\bar{x})d < 0\}$ *and* D *is the cone of feasible directions of* S *at* $\bar{x}$*, that is*

$$D = \{d : d \neq 0, \text{ and } \bar{x} + \lambda d \in S \text{ for all } \lambda \in (0, \delta) \text{ for some } \delta > 0\}.$$

Proof.

Suppose, by contradiction, that there exists a vector $d \in F_0 \cap D$. Since $d \in F_0$, then, by Theorem 2.1 (Descent Direction), there exists a $\delta_1 > 0$ such that

$$f(\bar{x} + \lambda d) < f(\bar{x}) \quad \text{for each } \lambda \in (0, \delta_1). \tag{11}$$

Also, since $d \in D$, by Definition 3.2, there exists a $\delta_2 > 0$ such that

$$\bar{x} + \lambda d \in S \quad \text{for each } \lambda \in (0, \delta_2). \tag{12}$$

The assumption that $\bar{x}$ is a local optimal solution is not compatible with (11) and (12). Thus, $F_0 \cap D = \emptyset$. □

Geometric Necessary Optimality Conditions

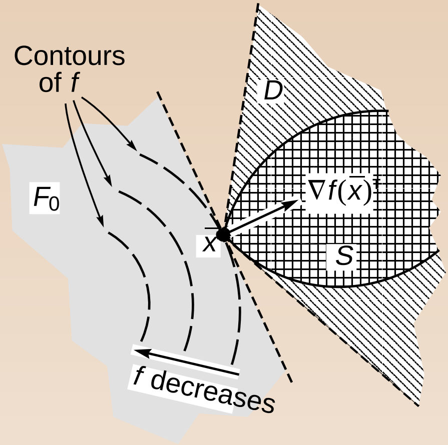


Figure: Illustration of the necessary condition $F_0 \cap D = \emptyset$.

Problems with Inequality and Equality Constraints

We next consider a specific description for the feasible region S as follows:

$$S = \{x \in X : g_i(x) \leq 0,\ i = 1, \ldots, m,\ h_i(x) = 0,\ i = 1, \ldots, l\},$$

where $g_i : \mathbb{R}^n \to \mathbb{R}$ for $i = 1, \ldots, m$, $h_i : \mathbb{R}^n \to \mathbb{R}$ for $i = 1, \ldots, \ell$, and X is a nonempty open set in $\mathbb{R}^n$.

This gives the following *nonlinear programming* problem with inequality and equality constraints:

$$\begin{aligned} &\text{minimise } f(x), \\ &\text{subject to:} \\ &g_i(x) \leq 0 \quad \text{for } i = 1, \ldots, m, \\ &h_i(x) = 0 \quad \text{for } i = 1, \ldots, \ell, \\ &x \in X. \end{aligned} \tag{13}$$

Suppose that $\bar{x}$ is a feasible solution of problem (13), and let $I = \{i : g_i(\bar{x}) = 0\}$ be the index set for the *binding* or *active* constraints. Suppose that there are no equality constraints.

Furthermore, suppose that each g_i for $i \notin I$ is continuous at $\bar{x}$, that f and g_i for $i \in I$ are differentiable at $\bar{x}$.

Let

$$G_0 \triangleq \{d : \nabla g_i(\bar{x})d < 0 \quad \text{for } i \in I\},$$
$$G_0' \triangleq \{d \neq 0 : \nabla g_i(\bar{x})d \leq 0 \quad \text{for } i \in I\}.$$

Recall the cone of feasible directions of S at $\bar{x}$:

$$D = \{d : d \neq 0, \text{ and } \bar{x} + \lambda d \in S \text{ for all } \lambda \in (0, \delta) \text{ for some } \delta > 0\}.$$

Then

$$G_0 \subseteq D \subseteq G_0'. \tag{14}$$

To see the first inclusion, let $d \in G_0$. Since $\overline{x} \in X$, and X is open, there exists $\delta_1 > 0$ such that

$$\overline{x} + \lambda d \in X \quad \text{for } \lambda \in (0, \delta_1).$$

Also, since g_i, $i \notin I$ is continuous at $\overline{x}$, there exists $\delta_2 > 0$ such that

$$g_i(\overline{x} + \lambda d) < 0 \quad \text{for } \lambda \in (0, \delta_2) \text{ and for } i \notin I.$$

Furthermore, since $d \in G_0$, then $\nabla g_i(\overline{x})d < 0$ for each $i \in I$. By Theorem 2.1 (Descent Direction) there exists $\delta_3 > 0$ such that

$$g_i(\overline{x} + \lambda d) < g_i(\overline{x}) = 0 \quad \text{for } \lambda \in (0, \delta_3) \text{ and for } i \in I.$$

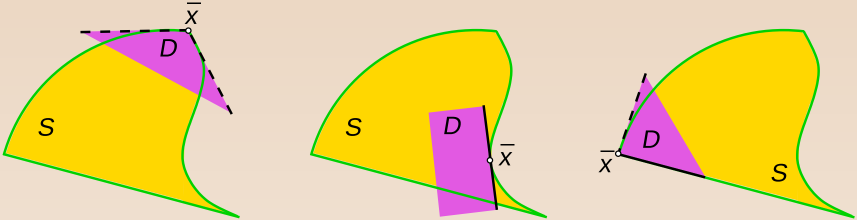
It is then clear that points of the form $\overline{x} + \lambda d$ are feasible to S for each $\lambda \in (0, \delta)$, where $\delta = \min\{\delta_1, \delta_2, \delta_3\}$. Thus $d \in D$ and hence $G_0 \subseteq D$.

Algebraic Description of the Cone of Feasible Directions

$$G_0 \subseteq D \subseteq G_0'$$

where

$$G_0 \triangleq \{d : \nabla g_i(\bar{x})d < 0 \quad \text{for } i \in I\},$$
$$G_0' \triangleq \{d \neq 0 : \nabla g_i(\bar{x})d \leq 0 \quad \text{for } i \in I\}.$$



$G_0 = D \subset G_0'$ $\qquad$ $G_0 \subset D = G_0'$ $\qquad$ $G_0 \subset D \subset G_0'$

Theorem (Geometric Necessary Condition for Problems with Inequality and Equality Constraints)

Let X be a nonempty open set in $\mathbb{R}^n$, and let $f : \mathbb{R}^n \to \mathbb{R}$, $g_i : \mathbb{R}^n \to \mathbb{R}$ for $i = 1, \ldots, m$, $h_i : \mathbb{R}^n \to \mathbb{R}$ for $i = 1, \ldots, \ell$. Consider the problem defi ned in (13). Suppose that $\bar{x}$ is a local optimal solution, and let $I = \{i : g_i(\bar{x}) = 0\}$ be the index set for the binding or active constraints. Furthermore, suppose that each g_i for $i \notin I$ is continuous at $\bar{x}$, that f and g_i for $i \in I$ are differentiable at $\bar{x}$, and that each h_i for $i = 1, \ldots, \ell$ is continuously differentiable at $\bar{x}$. If $\nabla h_i(\bar{x})^T$ for $i = 1, \ldots, \ell$ are linearly independent, then $F_0 \cap G_0 \cap H_0 = \emptyset$, where

$$
\begin{aligned}
F_0 &= \{d : \nabla f(\bar{x}) d < 0\}, \\
G_0 &= \{d : \nabla g_i(\bar{x}) d < 0 \quad \text{for } i \in I\}, \\
H_0 &= \{d : \nabla h_i(\bar{x}) d = 0 \quad \text{for } i = 1, \ldots, \ell\}.
\end{aligned}
\tag{15}
$$

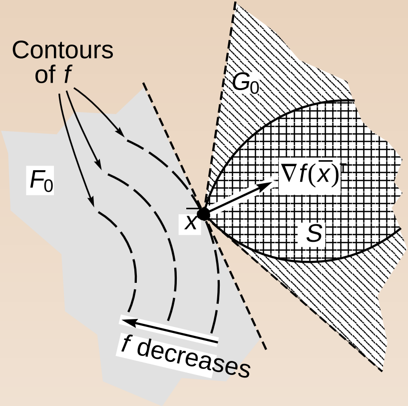
Problems with Inequality and Equality Constraints

Proof.

(Only for inequality constraints.) Let $\bar{x}$ be a local minimum. We then have the following implications from (10) and (14):

$$\bar{x} \text{ is a local minimum}$$
$$\implies F_0 \cap D = \emptyset$$
$$\implies F_0 \cap G_0 = \emptyset.$$

□



The Fritz John Necessary Conditions

We will now express the geometric condition $F_0 \cap G_0 \cap H_0 = \emptyset$ in an algebraic form known as the Fritz John conditions.

Theorem (The Fritz John Necessary Conditions)

Let X be a nonempty open set in $\mathbb{R}^n$, and let $f : \mathbb{R}^n \to \mathbb{R}$, $g_i : \mathbb{R}^n \to \mathbb{R}$ for $i = 1, \ldots, m$, $h_i : \mathbb{R}^n \to \mathbb{R}$ for $i = 1, \ldots, \ell$. Let $\overline{x}$ be a feasible solution of (13), and let $I = \{i : g_i(\overline{x}) = 0\}$. Suppose that g_i for $i \notin I$ is continuous at $\overline{x}$, that f and g_i for $i \in I$ are differentiable at $\overline{x}$, and that h_i for $i = 1, \ldots, \ell$ is continuously differentiable at $\overline{x}$. If $\overline{x}$ locally solves problem (13), then there exist scalars u_0 and u_i for $i \in I$, and v_i for $i = 1, \ldots, \ell$, such that

$$
\begin{aligned}
u_0 \nabla f(\overline{x})^\mathsf{T} + \sum_{i \in I} u_i \nabla g_i(\overline{x})^\mathsf{T} + \sum_{i=1}^{\ell} v_i \nabla h_i(\overline{x})^\mathsf{T} &= 0, \\
u_0, u_i \geq 0 \quad \text{for } i \in I&, \\
\{u_0, u_i, i \in I, v_1, \ldots, v_\ell\} \text{ not all zero}&\,.
\end{aligned}
\tag{16}
$$

Theorem (The FJ Necessary Conditions, continued)

Furthermore, if g_i, $i \notin I$ are also differentiable at $\bar{x}$, then the above conditions can be written as

$$\begin{aligned} u_0 \nabla f(\bar{x})^\top + \sum_{i=1}^{m} u_i \nabla g_i(\bar{x})^\top + \sum_{i=1}^{\ell} v_i \nabla h_i(\bar{x})^\top &= 0, \\ u_i g_i(\bar{x}) &= 0 \quad \text{for } i = 1, \ldots, m, \\ u_0, u_i &\geq 0 \quad \text{for } i = 1, \ldots, m, \\ (u_0, u, v) &\neq (0, 0, 0), \end{aligned} \tag{17}$$

where u and v are vectors whose components are u_i, $i = 1, \ldots, m$, and v_i, $i = 1, \ldots, \ell$, respectively.

Proof:

If the vectors $\nabla h_i(\bar{x})^\intercal$ for $i = 1, \ldots, \ell$ are linearly dependent, then one can find scalars $v_1, \ldots, v_\ell$, not all zero, such that $\sum_{i=1}^{\ell} v_i \nabla h_i(\bar{x})^\intercal = 0$. Letting u_0 and u_i for $i \in I$ equal to zero, conditions (16) hold trivially.

Now suppose that $\nabla h_i(\bar{x})^\intercal$ for $i = 1, \ldots, \ell$ are linearly independent. Then, from Theorem 3.3 (Geometric Necessary Condition), local optimality of $\bar{x}$ implies that the sets defined in (15) satisfy:

$$F_0 \cap G_0 \cap H_0 = \emptyset. \tag{18}$$

Let A_1 be the matrix whose rows are $\nabla f(\bar{x})$ and $\nabla g_i(\bar{x})$ for $i \in I$, and let A_2 be the matrix whose rows are $\nabla h_i(\bar{x})$ for $i = 1, \ldots, \ell$. Then, (18) is satisfied if and only if the following system is inconsistent:

$$A_1 d < 0,$$
$$A_2 d = 0.$$

Proof (continued):

Now consider the following two sets:

$$S_1 = \{(z_1, z_2) : z_1 = A_1 d,\ z_2 = A_2 d,\ d \in \mathbb{R}^n\},$$
$$S_2 = \{(z_1, z_2) : z_1 < 0,\ z_2 = 0\}.$$

Note that S_1 and S_2 are nonempty convex sets and, since the system $A_1 d < 0$, $A_2 d = 0$ has no solution, then $S_1 \cap S_2 = \emptyset$.

Then, by the theorem of separation of two disjoint convex sets, there exists a nonzero vector $p^\top = (p_1^\top, p_2^\top)$ such that

$$p_1^\top A_1 d + p_2^\top A_2 d \geq p_1^\top z_1 + p_2^\top z_2,$$

for each $d \in \mathbb{R}^n$ and $(z_1, z_2) \in \operatorname{cl} S_2$.

Proof (continued):

Hence

$$p_1^\intercal A_1 d + p_2^\intercal A_2 d \geq p_1^\intercal z_1 + p_2^\intercal z_2,$$

for each $d \in \mathbb{R}^n$ and $(z_1, z_2) \in \operatorname{cl} S_2 = \{(z_1, z_2) : z_1 < 0,\ z_2 = 0\}$.

Noting that $z_2 = 0$ and since each component of z_1 can be made an arbitrarily large negative number, it follows that $p_1 \geq 0$.

Also, letting $(z_1, z_2) = (0, 0) \in \operatorname{cl} S_2$, we must have $(p_1^\intercal A_1 + p_2^\intercal A_2)d \geq 0$ for each $d \in \mathbb{R}^n$.

Letting $d = -(A_1^\intercal p_1 + A_2^\intercal p_2)$, it follows that $-\|A_1^\intercal p_1 + A_2^\intercal p_2\|^2 \geq 0$, and thus $A_1^\intercal p_1 + A_2^\intercal p_2 = 0$.

Proof (continued):

Summarising, we have found a nonzero vector $p^{\mathsf{T}} = (p_1^{\mathsf{T}}, p_2^{\mathsf{T}})$ with $p_1 \geq 0$ such that $A_1^{\mathsf{T}} p_1 + A_2^{\mathsf{T}} p_2 = 0$, where A_1 is the matrix whose rows are $\nabla f(\bar{x})$ and $\nabla g_i(\bar{x})$ for $i \in I$, and A_2 is the matrix whose rows are $\nabla h_i(\bar{x})$ for $i = 1, \ldots, \ell$.

Denoting the components of p_1 by u_0 and u_i, $i \in I$, and letting $p_2 = v$, conditions (16) follow.

The equivalent form (17) is readily obtained by letting $u_i = 0$ for $i \notin I$, and the proof is complete. $\square$

The Fritz John Necessary Conditions

The scalars u_0, u_i for $i = 1, \ldots, m$, and v_i for $i = 1, \ldots, \ell$, are called the *Lagrange multipliers* associated, respectively, with the objective function, the inequality constraints $g_i(x) \le 0$, $i = 1, \ldots, m$, and the equality constraints $h_i(x) = 0$, $i = 1, \ldots, \ell$.

The condition that $\bar{x}$ be feasible for the optimisation problem (13) is called the *primal feasibility* [PF] condition.

The requirements $u_0 \nabla f(\bar{x})^\intercal + \sum_{i=1}^{m} u_i \nabla g_i(\bar{x})^\intercal + \sum_{i=1}^{\ell} v_i \nabla h_i(\bar{x})^\intercal = 0$, with $u_0, u_i \ge 0$ for $i = 1, \ldots, m$, and $(u_0, u, v) \ne (0, 0, 0)$ are called the *dual feasibility* [DF] conditions.

The condition $u_i g_i(\bar{x}) = 0$ for $i = 1, \ldots, m$ is called the *complementary slackness* [CS] condition; it requires that $u_i = 0$ if the corresponding inequality is nonbinding (that is, $g_i(\bar{x}) < 0$), and allows for $u_i > 0$ only for those constraints that are binding.

The Fritz John Necessary Conditions

The FJ conditions can also be written in vector form as follows:

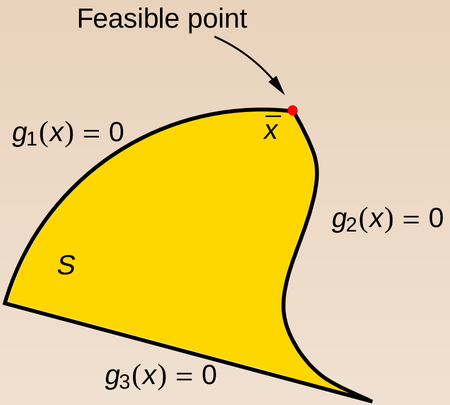
$$\begin{aligned} \nabla f(\bar{x})^\intercal u_0 + \nabla g(\bar{x})^\intercal u + \nabla h(\bar{x})^\intercal v &= 0, \\ u^\intercal g(\bar{x}) &= 0, \\ (u_0, u) &\geq (0, 0), \\ (u_0, u, v) &\neq (0, 0, 0), \end{aligned} \tag{19}$$

where

- $\nabla g(\bar{x})$ is the $m \times n$ Jacobian matrix whose ith row is $\nabla g_i(\bar{x})$,
- $\nabla h(\bar{x})$ is the $\ell \times n$ Jacobian matrix whose ith row is $\nabla h_i(\bar{x})$,
- $g(\bar{x})$ is the m vector function whose ith component is $g_i(\bar{x})$.

Any point $\bar{x}$ for which there exist Lagrange multipliers such that the FJ conditions are satisfied is called an *FJ point*.

Illustration: FJ conditions

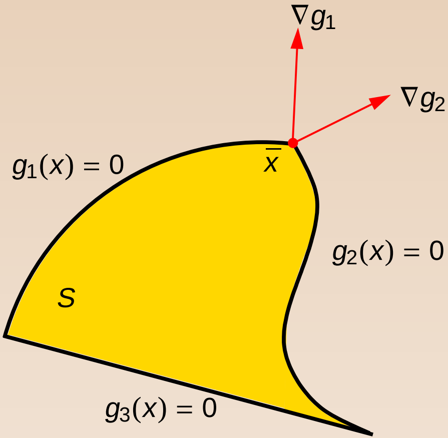


The constraint set S is:

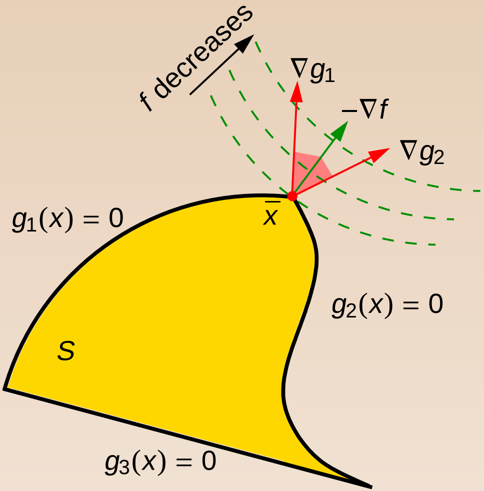
$$S = \{x \in \mathbb{R}^2 : \\ g_1(x) \leq 0, \\ g_2(x) \leq 0, \\ g_3(x) \leq 0\}$$

Consider the feasible point $\bar{x}$.

Illustration: FJ conditions

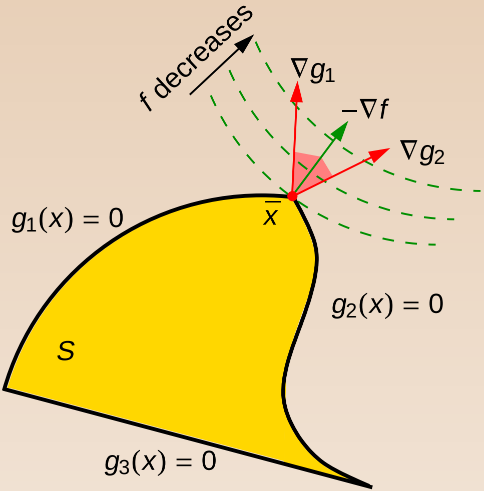


Consider the gradients of the active constraints at $\bar{x}$, $\nabla g_1(\bar{x})$ and $\nabla g_2(\bar{x})$.



For the given contours of the objective function f, we have that $u_0(-\nabla f(\bar{x}))$ is in the cone spanned by $\nabla g_1(\bar{x})$ and $\nabla g_2(\bar{x})$ with $u_0 > 0$.

Illustration: FJ conditions



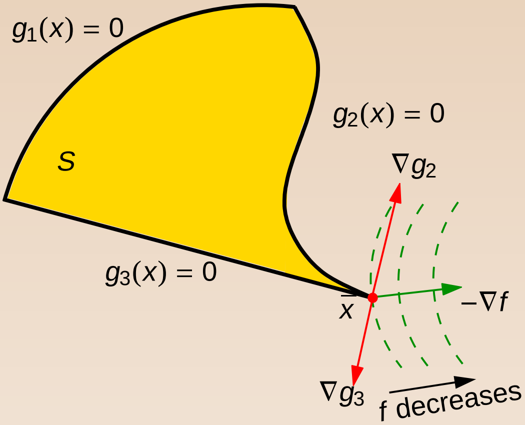
The FJ conditions are

$$\nabla f(\bar{x})^\intercal u_0 + \nabla g(\bar{x})^\intercal u = 0,$$
$$u^\intercal g(\bar{x}) = 0,$$
$$(u_0, u) \geq (0, 0),$$
$$(u_0, u, v) \neq (0, 0, 0),$$

$\bar{x}$ is an FJ point with $u_0 > 0$.

It is also a local minimum.

Illustration: FJ conditions

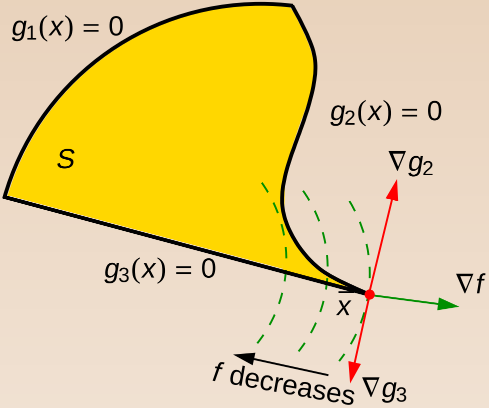


For the given contours of f, we have that $u_0(-\nabla f(\bar{x}))$ is in the cone spanned by $\nabla g_1(\bar{x})$ and $\nabla g_2(\bar{x})$ only if $u_0 = 0$.

$\bar{x}$ is an FJ point with $u_0 = 0$.

It is also a local minimum.

Illustration: FJ conditions



$\bar{x}$ is an FJ point with $u_0 = 0$.

It is also a local maximum.

The Fritz John Necessary Conditions

Given an optimisation problem, there might be points that satisfy the FJ conditions trivially. For example:

- if a feasible point $\bar{x}$ (not necessarily an optimum) satisfies $\nabla f(\bar{x}) = 0$, or $\nabla g_i(\bar{x}) = 0$ for some $i \in I$, or $\nabla h_i(\bar{x}) = 0$ for some $i = 1, \dots, \ell$, then we can let the corresponding Lagrange multiplier be any positive number, set all the other multipliers equal to zero, and satisfy conditions (16).
- In fact, given *any* feasible solution $\bar{x}$ we can always add a redundant constraint to the problem to make $\bar{x}$ an FJ point. For example, we can add the constraint $\|x - \bar{x}\|^2 \geq 0$, which holds true for all $x \in \mathbb{R}^n$, is a binding constraint at $\bar{x}$ and whose gradient is zero at $\bar{x}$.

- Moreover, it is also possible that, at some feasible point $\bar{x}$, the FJ conditions (16) are satisfied with Lagrange multiplier associated with the objective function $u_0 = 0$.

  In those cases, the objective function gradient does not play a role in the optimality conditions (16) and the conditions merely state that the gradients of the binding inequality constraints and of the equality constraints are linearly dependent.

  Thus, if $u_0 = 0$, the FJ conditions are of no practical value in locating an optimal point.

Constraint Qualifi cation

Under suitable assumptions, referred to as *constraint qualifi cations*, u_0 is guaranteed to be positive and the FJ conditions become the Karush–Kuhn–Tucker [KKT] conditions, which will be presented next.

There exist various constraint qualifications for problems with inequality and equality constraints.

Here, we use a typical constraint qualification that requires that the gradients of the inequality constraints for $i \in I$ and the gradients of the equality constraints at $\bar{x}$ be linearly independent.

Karush–Kuhn–Tucker Necessary Conditions

Theorem (Karush–Kuhn–Tucker Necessary Conditions)

Let X be a nonempty open set in $\mathbb{R}^n$, and let $f : \mathbb{R}^n \to \mathbb{R}$, $g_i : \mathbb{R}^n \to \mathbb{R}$ for $i = 1, \dots, m$, $h_i : \mathbb{R}^n \to \mathbb{R}$ for $i = 1, \dots, \ell$. Consider the problem defined in (13). Let $\bar{x}$ be a feasible solution, and let $I = \{i : g_i(\bar{x}) = 0\}$. Suppose that f and g_i for $i \in I$ are differentiable at $\bar{x}$, that each g_i for $i \notin I$ is continuous at $\bar{x}$, and that each h_i for $i = 1, \dots, \ell$ is continuously differentiable at $\bar{x}$. Furthermore, suppose that $\nabla g_i(\bar{x})^\intercal$ for $i \in I$ and $\nabla h_i(\bar{x})^\intercal$ for $i = 1, \dots, \ell$ are linearly independent. If $\bar{x}$ is a local optimal solution, then there exist unique scalars u_i for $i \in I$, and v_i for $i = 1, \dots, \ell$, such that

$$\begin{aligned} \nabla f(\bar{x})^\intercal + \sum_{i \in I} u_i \nabla g_i(\bar{x})^\intercal + \sum_{i=1}^{\ell} v_i \nabla h_i(\bar{x})^\intercal &= 0, \\ u_i \geq 0 \quad \text{for } i \in I. \end{aligned} \tag{20}$$

Karush–Kuhn–Tucker Necessary Conditions

Theorem (KKT Necessary Conditions, continued)

Furthermore, if g_i, $i \notin I$ are also differentiable at $\bar{x}$, then the above conditions can be written as

$$\begin{aligned} \nabla f(\bar{x})^\top + \sum_{i=1}^{m} u_i \nabla g_i(\bar{x})^\top + \sum_{i=1}^{\ell} v_i \nabla h_i(\bar{x})^\top = 0, \\ u_i g_i(\bar{x}) = 0 \quad \text{for } i = 1, \dots, m, \\ u_i \geq 0 \quad \text{for } i = 1, \dots, m. \end{aligned} \tag{21}$$

Proof.

We have, from the FJ conditions, that there exist scalars $\hat{u}_0$ and $\hat{u}_i$, $i \in I$, and $\hat{v}_i$, $i = 1, \ldots, \ell$, not all zero, such that

$$\hat{u}_0 \nabla f(\bar{x})^{\mathsf{T}} + \sum_{i \in I} \hat{u}_i \nabla g_i(\bar{x})^{\mathsf{T}} + \sum_{i=1}^{\ell} \hat{v}_i \nabla h_i(\bar{x})^{\mathsf{T}} = 0, \qquad \hat{u}_0, \hat{u}_i \geq 0 \quad \text{for } i \in I. \tag{22}$$

Note that the assumption of linear independence of $\nabla g_i(\bar{x})^{\mathsf{T}}$ for $i \in I$ and $\nabla h_i(\bar{x})^{\mathsf{T}}$ for $i = 1, \ldots, \ell$, together with (22) and the fact that at least one of the multipliers is nonzero, implies that $\hat{u}_0 > 0$.

Then, letting $u_i = \hat{u}_i / \hat{u}_0$ for $i \in I$, and $v_i = \hat{v}_i / \hat{u}_0$ for $i = 1, \ldots, \ell$ we obtain conditions (20).

Furthermore, the linear independence assumption implies the uniqueness of these Lagrange multipliers. □

As in the FJ conditions, the scalars u_i and v_i are called the *Lagrange multipliers*.

The condition that $\overline{x}$ be feasible for the optimisation problem (13) is called the *primal feasibility* [PF] condition.

The requirement that $\nabla f(\overline{x})^\intercal + \sum_{i=1}^{m} u_i \nabla g_i(\overline{x})^\intercal + \sum_{i=1}^{\ell} v_i \nabla h_i(\overline{x})^\intercal = 0$, with $u_i \geq 0$ for $i = 1, \ldots, m$ is called the *dual feasibility* [DF] condition.

The condition $u_i g_i(\overline{x}) = 0$ for $i = 1, \ldots, m$ is called the *complementary slackness* [CS] condition

Karush–Kuhn–Tucker Necessary Conditions

The KKT conditions can also be written in vector form as follows:

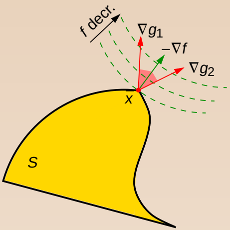
$$\begin{aligned} \nabla f(\bar{x})^\intercal + \nabla g(\bar{x})^\intercal u + \nabla h(\bar{x})^\intercal v &= 0, \\ u^\intercal g(\bar{x}) &= 0, \\ u &\geq 0, \end{aligned} \tag{23}$$

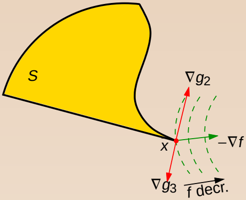
where

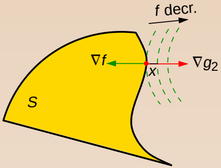
- $\nabla g(\bar{x})$ is the $m \times n$ Jacobian matrix whose ith row is $\nabla g_i(\bar{x})$,
- $\nabla h(\bar{x})$ is the $\ell \times n$ Jacobian matrix whose ith row is $\nabla h_i(\bar{x})$,
- $g(\bar{x})$ is the m vector function whose ith component is $g_i(\bar{x})$.

Any point $\bar{x}$ for which there exist Lagrange multipliers that satisfy the KKT conditions (23) is called a *KKT point*.

Illustration: KKT conditions







$\bar{x}$ is a KKT point

$\bar{x}$ is not a KKT point

$\bar{x}$ is a KKT point

Constraint Qualifi cations

The linear independence constraint qualification is a *suffi cient condition* placed on the behaviour of the constraints to ensure that an FJ point (and hence any local optimum) be a KKT point.

> Thus, the importance of the constraint qualifications is to guarantee that, by examining only KKT points, we do not lose out on optimal solutions.

There is an important special case:

> When the constraints are linear the KKT conditions are always necessary optimality conditions irrespective of the objective function.

This is because Abadie's constraint qualification is automatically satisfied for linear constraints.

However, we are still left with the problem of determining, among all the points that satisfy the KKT conditions, which ones constitute local optimal solutions.

The following result shows that, under moderate convexity assumptions, the KKT conditions are also sufficient for local optimality.

Theorem (Karush–Kuhn–Tucker Sufficient Conditions)

Let X be a nonempty open set in $\mathbb{R}^n$, and let $f : \mathbb{R}^n \to \mathbb{R}$, $g_i : \mathbb{R}^n \to \mathbb{R}$ for $i = 1, \ldots, m$, $h_i : \mathbb{R}^n \to \mathbb{R}$ for $i = 1, \ldots, \ell$. Consider the problem defi ned in (13). Let $\bar{x}$ be a feasible solution, and let $I = \{i : g_i(\bar{x}) = 0\}$. Suppose that the KKT conditions hold at $\bar{x}$; that is, there exist scalars $\bar{u}_i \geq 0$ for $i \in I$, and $\bar{v}_i$ for $i = 1, \ldots, \ell$, such that

$$\nabla f(\bar{x})^\intercal + \sum_{i \in I} \bar{u}_i \nabla g_i(\bar{x})^\intercal + \sum_{i=1}^{\ell} \bar{v}_i \nabla h_i(\bar{x})^\intercal = 0. \tag{24}$$

Let $J = \{i : \bar{v}_i > 0\}$ and $K = \{i : \bar{v}_i < 0\}$. Further, suppose that f is pseudoconvex at $\bar{x}$, g_i is quasiconvex at $\bar{x}$ for $i \in I$, h_i is quasiconvex at $\bar{x}$ for $i \in J$, and h_i is quasiconcave at $\bar{x}$ (that is, $-h_i$ is quasiconvex at $\bar{x}$) for $i \in K$. Then $\bar{x}$ is a global optimal solution to problem (13).

Quadratic Programs

Quadratic programs are a special class of nonlinear programs in which the objective function is quadratic and the constraints are linear.

Thus, a quadratic programming [QP] problem can be written as

$$\text{minimise } \frac{1}{2}x^\intercal Hx + x^\intercal c, \tag{25}$$

subject to:

$$A_I^\intercal x \leq b_I,$$

$$A_E^\intercal x = b_E,$$

where H is an $n \times n$ matrix, c is an n vector, A_I is an $n \times m_I$ matrix, b_I is an m_I vector, A_E is an $n \times m_E$ matrix and b_E is an m_E vector.

The constraints are linear, hence

- $\bar{x}$ is a local minimum $\implies$ $\bar{x}$ is a KKT point.
- the constraint set $S = \{x : A_I^T x \leq b_I, A_E^T x = b_E\}$ is convex.

Thus,

the QP is convex $\iff$ the objective function is convex
$\iff$ H is symmetric and positive semidefinite

In this case:

$$\bar{x} \text{ is a local min} \iff \bar{x} \text{ is a global min} \iff \bar{x} \text{ is a KKT point}$$

Furthermore, if $H > 0$, then $\bar{x}$ is the unique global minimum.

KKT Conditions for QP

The KKT conditions (23) for the QP problem defined in (25) are:

$$
\begin{array}{llcl}
\text{PF:} & A_I^\mathsf{T}\overline{x} & \leq & b_I, \\
 & A_E^\mathsf{T}\overline{x} & = & b_E, \\
\text{DF:} & H\overline{x} + c + A_I u + A_E v & = & 0, \\
 & u & \geq & 0, \\
\text{CS:} & u^\mathsf{T}(A_I^\mathsf{T}\overline{x} - b_I) & = & 0,
\end{array}
\tag{26}
$$

where u is an m_I vector of Lagrange multipliers corresponding to the inequality constraints and v is an m_E vector of Lagrange multipliers corresponding to the equality constraints.