MS-E2122 - Nonlinear Optimization
Lecture 4

Fabricio Oliveira

Systems Analysis Laboratory
Department of Mathematics and Systems Analysis

Aalto University
School of Science

October 7, 2021

Outline of this lecture

Outline of this lecture

Recognising optimality

- Minima and maxima in optimisation
- Optimality conditions
- First- and second-order conditions

Preliminary definitions

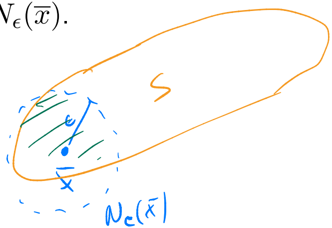
Let $f : \mathbb{R}^n \to \mathbb{R}$. Consider the problem (P) : min. $\{f(x) : x \in S\}$.
Some important terminology:

- feasible solution: $\overline{x} \in S$;

Preliminary definitions

Let $f : \mathbb{R}^n \to \mathbb{R}$. Consider the problem (P) : min. $\{f(x) : x \in S\}$.
Some important terminology:

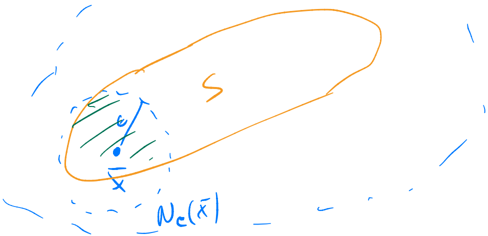
- feasible solution: $\overline{x} \in S$;
- local optimal solution: $\overline{x} \in S$ has a neighbourhood $N_\epsilon(\overline{x}) = \{x : ||x - \overline{x}|| \leq \epsilon\}$ for some $\epsilon > 0$ such that $f(\overline{x}) \leq f(x)$ for each $x \in S \cap N_\epsilon(\overline{x})$.



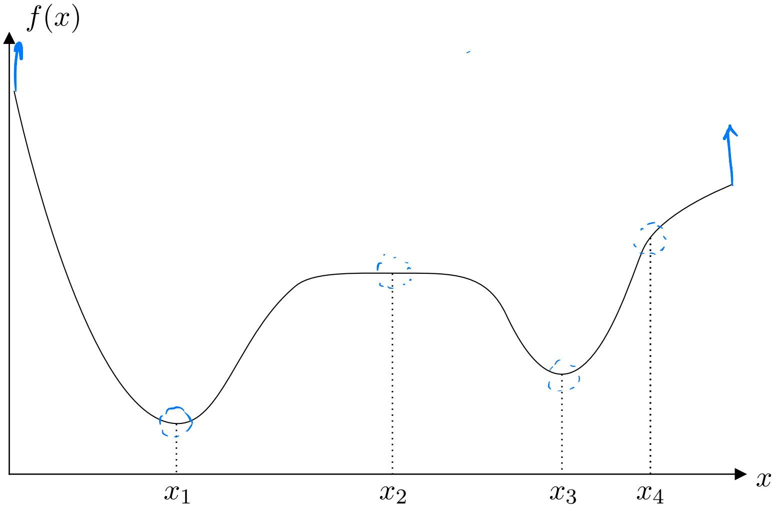
Preliminary definitions

Let $f : \mathbb{R}^n \to \mathbb{R}$. Consider the problem (P) : min. $\{f(x) : x \in S\}$.
Some important terminology:

- feasible solution: $\overline{x} \in S$;
- local optimal solution: $\overline{x} \in S$ has a neighbourhood $N_\epsilon(\overline{x}) = \{x : ||x - \overline{x}|| \leq \epsilon\}$ for some $\epsilon > 0$ such that $f(\overline{x}) \leq f(x)$ for each $x \in S \cap N_\epsilon(\overline{x})$.
- global optimal solution: $\overline{x} \in S$ with $f(\overline{x}) \leq f(x)$ for all $x \in S$.

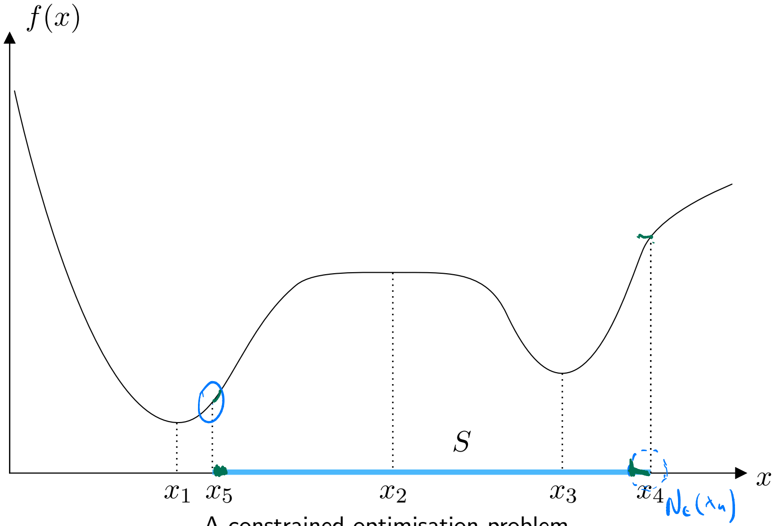


Preliminary definitions



An unconstrained optimisation problem

Preliminary definitions



A constrained optimisation problem

The importance of convexity

The following is the most fundamental result in optimisation:

Theorem 1 (global optimality of convex problems)

Let $S \subseteq \mathbb{R}^n$ be a nonempty convex set and $f : S \to \mathbb{R}$ convex on S. Consider the problem (P) : min. $\{f(x) : x \in S\}$. Suppose $\overline{x}$ is a local optimal solution to P. Then $\overline{x}$ is a global optimal solution.

The importance of convexity

The following is the most fundamental result in optimisation:

Theorem 1 (global optimality of convex problems)

Let $S \subseteq \mathbb{R}^n$ be a nonempty convex set and $f : S \to \mathbb{R}$ convex on S. Consider the problem (P) : min. $\{f(x) : x \in S\}$. Suppose $\overline{x}$ is a local optimal solution to P. Then $\overline{x}$ is a global optimal solution.

Proof.

Since $\overline{x}$ is a local optimal solution, there exists $N_\epsilon(\overline{x})$ such that, for each $x \in S \cap N_\epsilon(\overline{x})$, $f(\overline{x}) \leq f(x)$. By contradiction, suppose $\overline{x}$ is not a global optimal solution. Then, there exists a solution $\hat{x} \in S$ so that $f(\hat{x}) < f(\overline{x})$. Now, for any $\lambda \in [0,1]$, the convexity of f implies:

$$\underbrace{f(\lambda\hat{x} + (1-\lambda)\overline{x}) \leq \lambda f(\hat{x}) + (1-\lambda)f(\overline{x})}_{\text{convexity}} < \lambda f(\overline{x}) + (1-\lambda)f(\overline{x}) = f(\overline{x})$$

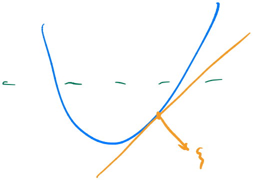
However, for $\lambda > 0$ sufficiently small, $\lambda\hat{x} + (1-\lambda)\overline{x} \in S \cap N_\epsilon(\overline{x})$, which contradicts the local optimality of $\overline{x}$. Thus, $\overline{x}$ is a global optimum. $\square$

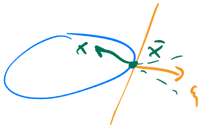
Optimality conditions

Theorem 2 gives a certificate for global optimal solutions for convex optimisation problems.

Theorem 2 (optimality condition)

Let $S \subseteq \mathbb{R}^n$ be a nonempty convex set and $f : S \to \mathbb{R}$ convex on S. Consider the problem (P) : min. $\{f(x) : x \in S\}$. Then, $\overline{x} \in S$ is an optimal solution to P if and only if f has a subgradient ξ at $\overline{x}$ such that $\xi^\top(x - \overline{x}) \geq 0$ for all $x \in S$.





Optimality conditions

Theorem 2 gives a certificate for global optimal solutions for convex optimisation problems.

Theorem 2 (optimality condition)

Let $S \subseteq \mathbb{R}^n$ be a nonempty convex set and $f : S \to \mathbb{R}$ convex on S. Consider the problem (P) : min. $\{f(x) : x \in S\}$. Then, $\overline{x} \in S$ is an optimal solution I *to P if and only if f has a subgradient ξ at $\overline{x}$ such that $\xi^\top(x - \overline{x}) \geq 0$ for all $x \in S$.* II

Proof.

Suppose that $\xi^\top(x - \overline{x}) \geq 0$ for all $x \in S$, where ξ is a subgradient of f at $\overline{x}$. By convexity of f, we have, for all $x \in S$

II → I
$$f(x) \geq f(\overline{x}) + \xi^\top(x - \overline{x}) \geq f(\overline{x})$$
⇝ subgradient ineq.

and hence $\overline{x}$ is optimal.

Optimality conditions

Proof (cont.)

Conversely, suppose that $\overline{x}$ is optimal for P. Construct the sets:

$\mathrm{I} \dashrightarrow \mathrm{II}$

$$\Lambda_1 = \{(x - \overline{x}, y) : x \in \mathbb{R}^n, y > \overbrace{f(x) - f(\overline{x})}^{\geq 0}\}$$

$\zeta^T(x - \bar{x}) \geq 0$

$$\Lambda_2 = \{(x - \overline{x}, y) : x \in S, y \leq 0\}$$

Optimality conditions

Proof (cont.)

Conversely, suppose that $\overline{x}$ is optimal for P. Construct the sets:

$$\Lambda_1 = \{(x - \overline{x}, y) : x \in \mathbb{R}^n, y > f(x) - f(\overline{x})\}$$
$$\Lambda_2 = \{\underbrace{(x - \overline{x}, y)}_{z} : x \in S, y \leq 0\}$$

Note that Λ_1 and Λ_2 are convex. By optimality of $\overline{x}$, $\Lambda_1 \cap \Lambda_2 = \emptyset$. Using the separation theorem, there exists a hyperplane defined by $\underbrace{(\xi_0, \mu)}_{p} \neq 0$ and α that separates Λ_1 and Λ_2:

$p^\top z = \alpha$

$$\xi_0^\top(x - \overline{x}) + \mu y \leq \alpha, \ \forall x \in \mathbb{R}^n, \ y > f(x) - f(\overline{x}) \quad (1)$$
$$\xi_0^\top(x - \overline{x}) + \mu y \geq \alpha, \ \forall x \in S, \quad y \leq 0. \quad (2)$$

Optimality conditions

Proof (cont.)

Conversely, suppose that $\overline{x}$ is optimal for P. Construct the sets:

$$\Lambda_1 = \{(x - \overline{x}, y) : x \in \mathbb{R}^n, y > f(x) - f(\overline{x})\}$$
$$\Lambda_2 = \{(x - \overline{x}, y) : x \in S, y \leq 0\}$$

Note that Λ_1 and Λ_2 are convex. By optimality of $\overline{x}$, $\Lambda_1 \cap \Lambda_2 = \emptyset$. Using the separation theorem, there exists a hyperplane defined by $(\xi_0, \mu) \neq 0$ and α that separates Λ_1 and Λ_2:

$$\xi_0^\top (x - \overline{x}) + \mu y \leq \alpha, \ \forall x \in \mathbb{R}^n, \ y > f(x) - f(\overline{x}) \tag{1}$$
$$\xi_0^\top (x - \overline{x}) + \mu y \geq \alpha, \ \forall x \in S, \quad y \leq 0. \tag{2}$$

Letting $x = \overline{x}$ and $y = 0$ in (2), we get $\alpha \leq 0$. Next, letting $x = \overline{x}$ and $y = \epsilon > 0$ in (1), we obtain $\alpha \geq \mu\epsilon$. As this holds for any $\epsilon > 0$, we must have $\mu \leq 0$ and $\alpha \geq 0$, the latter implying $\alpha = 0$.

Optimality conditions

$\xi_0^\top \xi_0 \le 0 \Rightarrow$

$\xi_0 = 0 \Rightarrow \rho = 0$

$\mu = 0$

Proof (cont.)

If $\mu = 0$, we get from (1) that $\xi_0^\top(x - \overline{x}) \le 0$ for all $x \in \mathbb{R}^n$. Now, by letting $x = \overline{x} + \xi_0$, it follows that $\xi_0^\top(x - \overline{x}) = ||\xi_0||^2 \le 0$, and thus $\xi_0 = 0$. Since $(\xi_0, \mu) \neq 0$, we must have $\mu < 0$.

Dividing (1) and (2) by $-\mu$ and denoting $\xi = \frac{-\xi_0}{\mu}$, we obtain:

$$\xi^\top(x - \overline{x}) \le y, \ \forall x \in \mathbb{R}^n, \ y > f(x) - f(\overline{x}) \tag{3}$$

$$\xi^\top(x - \overline{x}) \ge y, \ \forall x \in S, \quad y \le 0 \tag{4}$$

Optimality conditions

Proof (cont.)

If $\mu = 0$, we get from (1) that $\xi_0^\top(x - \overline{x}) \leq 0$ for all $x \in \mathbb{R}^n$. Now, by letting $x = \overline{x} + \xi_0$, it follows that $\xi_0^\top(x - \overline{x}) = ||\xi_0||^2 \leq 0$, and thus $\xi_0 = 0$. Since $(\xi_0, \mu) \neq 0$, we must have $\mu < 0$.

Dividing (1) and (2) by $-\mu$ and denoting $\xi = \frac{-\xi_0}{\mu}$, we obtain:

$$\xi^\top(x - \overline{x}) \leq y, \ \forall x \in \mathbb{R}^n, \ y > f(x) - f(\overline{x}) \tag{3}$$

$$\xi^\top(x - \overline{x}) \geq y, \ \forall x \in S, \quad y \leq 0 \tag{4}$$

$y > f(x) - f(\bar{x})$ '

$y \geq \xi^T(x - \bar{x})$ ''

Letting $y = 0$ in (4), we get $\xi^\top(x - \overline{x}) \geq 0$ for all $x \in S$. From (3), we can see that $y > f(x) - f(\overline{x})$ and $y \geq \xi^\top(x - \overline{x})$. Thus, $f(x) - f(\overline{x}) \geq \xi^\top(x - \overline{x})$, which is the subgradient inequality.

Thus ξ is a subgradient at $\overline{x}$ with $\xi^\top(x - \overline{x}) \geq 0$ for all $x \in S$. $\square$

$f(x) - f(\bar{x}) \geq \xi^T(x - \bar{x})$

$f(x) \geq f(\bar{x}) + \xi^T(x - \bar{x})$

Optimality conditions

Theorem 2 leads to two important corollaries:

Corollary 3 (optimality in open sets)

Under the conditions of Theorem 2, if S is open, $\overline{x}$ is an optimal solution to P if and only if $0 \in \partial f(\overline{x})$.

Optimality conditions

Theorem 2 leads to two important corollaries:

Corollary 3 (optimality in open sets)

Under the conditions of Theorem 2, if S is open, $\overline{x}$ is an optimal solution to P if and only if $0 \in \partial f(\overline{x})$.

Proof.

From Theorem 2, $\overline{x}$ is optimal if and only if ξ is a subgradient at $\overline{x}$ with $\xi^\top(x - \overline{x}) \geq 0$ for all $x \in S$. Since S is open, $x = \overline{x} - \lambda\xi \in S$ for some $\lambda > 0$, and thus $-\lambda||\xi||^2 \geq 0$, which implies $\xi = 0$. $\square$

$(\bar{x} - \lambda\xi)$

Optimality conditions

Theorem 2 leads to two important corollaries:

Corollary 3 (optimality in open sets)

Under the conditions of Theorem 2, if S is open, $\overline{x}$ is an optimal solution to P if and only if $0 \in \partial f(\overline{x})$.

Proof.

From Theorem 2, $\overline{x}$ is optimal if and only if ξ is a subgradient at $\overline{x}$ with $\xi^\top(x - \overline{x}) \geq 0$ for all $x \in S$. Since S is open, $x = \overline{x} - \lambda\xi \in S$ for some $\lambda > 0$, and thus $-\lambda||\xi||^2 \geq 0$, which implies $\xi = 0$. □

Corollary 4 (optimality for differentiable functions)

Suppose that $S \subseteq \mathbb{R}^n$ is a nonempty convex set and $f : S \to \mathbb{R}$ a differentiable convex function on S. Then $\overline{x} \in S$ is optimal if and only if $\nabla f(\overline{x})^\top(x - \overline{x}) \geq 0$ for all $x \in S$. Moreover, if S is open, then $\overline{x}$ is optimal if and only if $\nabla f(\overline{x}) = 0$.

Optimality conditions

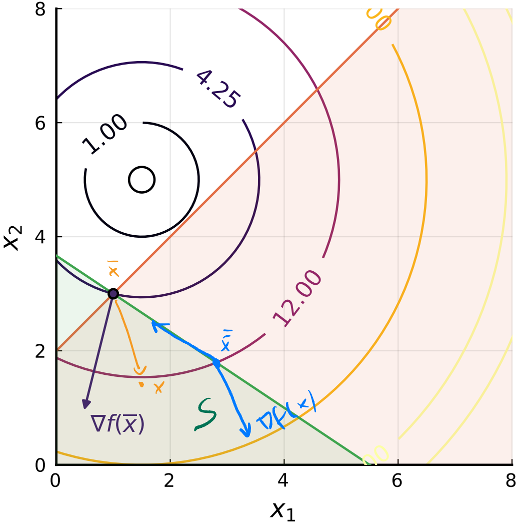
Example 1:

$$\text{min. } \left(x_1 - \frac{3}{2}\right)^2 + (x_2 - 5)^2$$

$$\begin{aligned} \text{subject to: } & -x_1 + x_2 \leq 2 \\ & 2x_1 + 3x_2 \leq 11 \\ & x_1 \geq 0 \\ & x_2 \geq 0 \end{aligned}$$

$\nabla f(x)^\top (x - \bar{x}) \geq 0$

$x \in S$.

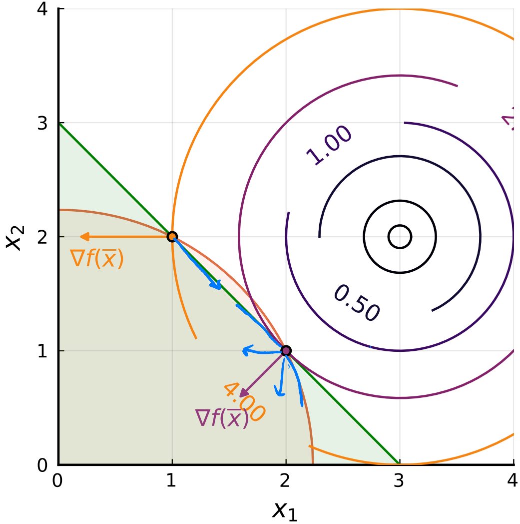


Optimality conditions

Example 2:

$$
\begin{aligned}
\text{min. } & (x_1 - 3)^2 + (x_2 - 2)^2 \\
\text{subject to: } & x_1 + x_2 \leq 3 \\
& x_1^2 + x_2^2 \leq 5 \\
& x_1 \geq 0 \\
& x_2 \geq 0
\end{aligned}
$$

$\nabla f(x)^T (x - \bar{x}) \geq 0$



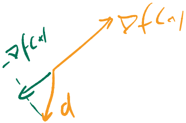
Optimality for unconstrained problems

We can derive necessary first- and second-order optimality conditions for unconstrained problems assuming differentiability.

Theorem 5 (descent direction)

Suppose $f : \mathbb{R}^n \to \mathbb{R}$ is differentiable at $\overline{x}$. If there is d such that $\nabla f(\overline{x})^\top d < 0$, there exists $\delta > 0$ such that $f(\overline{x} + \lambda d) < f(\overline{x})$ for each $\lambda \in (0, \delta)$, so that d is a descent direction of f at $\overline{x}$.

$$\nabla f(x)^\top \overbrace{(x - \bar{x})}^{d} \geq 0$$



Optimality for unconstrained problems

We can derive necessary first- and second-order optimality conditions for unconstrained problems assuming differentiability.

Theorem 5 (descent direction)

Suppose $f : \mathbb{R}^n \to \mathbb{R}$ is differentiable at $\overline{x}$. If there is d such that $\nabla f(\overline{x})^\top d < 0$, there exists $\delta > 0$ such that $f(\overline{x} + \lambda d) < f(\overline{x})$ for each $\lambda \in (0, \delta)$, so that d is a descent direction of f at $\overline{x}$.

Proof.

By differentiability of f at $\overline{x}$, we have that

directional diff.

$$\frac{f(\overline{x} + \lambda d) - f(\overline{x})}{\lambda} = \underbrace{\nabla f(\overline{x})^\top d}_{<0} + ||d||\alpha(\overline{x}; \lambda d).$$

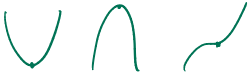
Since $\nabla f(\overline{x})^\top d < 0$ and $\alpha(\overline{x}; \lambda d) \to 0$ when $\lambda \to 0$ for some $\lambda \in (0, \delta)$, we must have $f(\overline{x} + \lambda d) - f(\overline{x}) < 0$. $\square$

Optimality for unconstrained problems

The first-order necessary condition follows from Theorem 5.

Corollary 6 (first-order necessary condition)

Suppose $f : \mathbb{R}^n \to \mathbb{R}$ is differentiable at $\overline{x}$. If $\overline{x}$ is a local minimum, then $\nabla f(\overline{x}) = 0$.



Optimality for unconstrained problems

The first-order necessary condition follows from Theorem 5.

Corollary 6 (first-order necessary condition)

Suppose $f : \mathbb{R}^n \to \mathbb{R}$ *is differentiable at* $\overline{x}$. *If* $\overline{x}$ *is a local minimum, then* $\nabla f(\overline{x}) = 0$.

Proof.

By contradiction, suppose that $\overline{x}$ is a local minimum with $\nabla f(\overline{x}) \neq 0$. Then, $\nabla f(\overline{x})^\top d = -||\nabla f(\overline{x})||^2 < 0$ for $d = -\nabla f(\overline{x})$. By Theorem 5, there exists a $\delta > 0$ such that $f(\overline{x} + \lambda d) < f(\overline{x})$ for all $\lambda \in (0, \delta)$, thus contradicting the optimality of $\overline{x}$. □

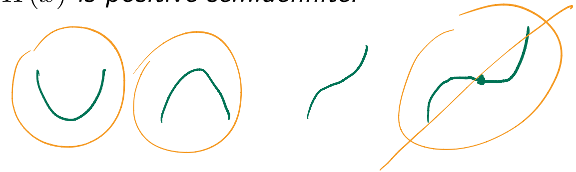
Remark: Theorem 5 and Corollary 6 can be combined to design a rudimentary optimisation algorithm.

Optimality for unconstrained problems

The second-order necessary condition is based on semi-definiteness of the Hessian of f, $H(\overline{x})$, at $\overline{x}$.

Theorem 7 (second-order necessary condition)

Suppose $f : \mathbb{R}^n \to \mathbb{R}$ is twice differentiable at $\overline{x}$. If $\overline{x}$ is a local minimum, then $H(\overline{x})$ is positive semidefinite.



Optimality for unconstrained problems

The second-order necessary condition is based on semi-definiteness of the Hessian of f, $H(\overline{x})$, at $\overline{x}$.

Theorem 7 (second-order necessary condition)

Suppose $f : \mathbb{R}^n \to \mathbb{R}$ is twice differentiable at $\overline{x}$. If $\overline{x}$ is a local minimum, then $H(\overline{x})$ is positive semidefinite.

Proof.

Take an arbitrary direction d. As f is twice differentiable, we have:

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

since $\overline{x}$ is a local minimum, Corollary 6 implies that $\nabla f(\overline{x}) = 0$ and $f(\overline{x} + \lambda d) \geq f(\overline{x})$.

Optimality for unconstrained problems

Proof (Cont.)

Rearranging terms and dividing by $\lambda^2 > 0$ we obtain

$$\frac{f(\overline{x} + \lambda d) - f(\overline{x})}{\lambda^2} = \frac{1}{2} d^\top H(\overline{x}) d + ||d||^2 \alpha(\overline{x}; \lambda d).$$

Since $\alpha(\overline{x}; \lambda d) \to 0$ as $\lambda \to 0$, we have that $d^\top H(\overline{x}) d \geq 0$. $\square$

Optimality for unconstrained problems

Proof (Cont.)

Rearranging terms and dividing by $\lambda^2 > 0$ we obtain

$$\frac{f(\overline{x} + \lambda d) - f(\overline{x})}{\lambda^2} = \frac{1}{2} d^\top H(\overline{x}) d + ||d||^2 \alpha(\overline{x}; \lambda d).$$

Since $\alpha(\overline{x}; \lambda d) \to 0$ as $\lambda \to 0$, we have that $d^\top H(\overline{x}) d \geq 0$. $\square$

These conditions are also sufficient in the following cases:

1. If $H(\overline{x})$ is positive definite, the second-order condition becomes sufficient for local optimality.
2. If f is convex, the first-order condition becomes necessary and sufficient for global optimality.

Optimality for unconstrained problems

The convexity of f implies that the first-order conditions are necessary and sufficient for global optimality.

Theorem 8

Let $f : \mathbb{R}^n \mapsto \mathbb{R}$ be convex. Then $\overline{x}$ is a global minimum if and only if $\nabla f(\overline{x}) = 0$.

Optimality for unconstrained problems

The convexity of f implies that the first-order conditions are necessary and sufficient for global optimality.

Theorem 8

Let $f : \mathbb{R}^n \mapsto \mathbb{R}$ be convex. Then $\overline{x}$ is a global minimum if and only if $\nabla f(\overline{x}) = 0$.

Proof.

From Corollary 6, if $\overline{x}$ is a global minimum, then $\nabla f(\overline{x}) = 0$. Now, since f is convex, we have that

$$f(x) \geq f(\overline{x}) + \nabla f(\overline{x})^\top (x - \overline{x})$$

Suppose that $\nabla f(\overline{x}) = 0$. This implies that $\nabla f(\overline{x})^\top (x - \overline{x}) = 0$ for each $x \in \mathbb{R}^n$, thus implying that $f(\overline{x}) \leq f(x)$ for all $x \in \mathbb{R}^n$. $\square$