MSE2122 - Nonlinear Optimization Lecture Notes IV

Fernando Dias (based on previous version by Fabricio Oliveira)

September 26, 2023

Abstract

In this lecture, we use the concepts of convexity and differentiability to derive optimality conditions. We start by looking into how the presence of convexity can be used to derive general optimality conditions for unconstrained (and also constrained, which we will consider later in the course) problems. Next, we use the more general and easier to verify concept of differentiability to derive necessary and sufficient local optimality conditions and conclude showing how convexity can be used to show that first-order optimality conditions are in fact necessary and sufficient for global optimality.

Contents

1 Recognising optimality

We now turn our focus to recognising whether a given point satisfy necessary and/or sufficient conditions for optimality. Even though these conditions can be used to test if a candidate point is optimal for a problem, its most important use is serving as a framework for directing solution methods in their search for optimal solutions.

Before we proceed, let us define the terminology we will use to refer to solutions. Let $f : \mathbb{R}^n \mapsto \mathbb{R}$. Consider the problem $(P) : \min. \{f(x) : x \in S\}$.

1. a *feasible solution* is any solution $\overline{x} \in S$;

2. a *local optimal solution* is a feasible solution $\overline{x} \in S$ that 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 $x \in S \cap N_\epsilon(\overline{x})$.

3. a *global optimal solution* is a feasible solution $\overline{x} \in S$ with $f(\overline{x}) \leq f(x)$ for all $x \in S$. Or alternatively, is a local optimal solution for which $S \subseteq N_\epsilon(\overline{x})$.

Figure 2 illustrates the concepts above. Solution x_1 is an unconstrained global minimum, but is not a feasible solution considering the feasibility set S. Solution x_2 is a local optima, for any neighbourhood $N_\epsilon(x_2)$ only encompassing the points within the same plateau. Solution x_3 is a local optimum, while x_4 is neither a local or a global optimum in the unconstrained case, but is a global maximum in the constrained case. Finally, x_5 is the global minimum in the constrained case.
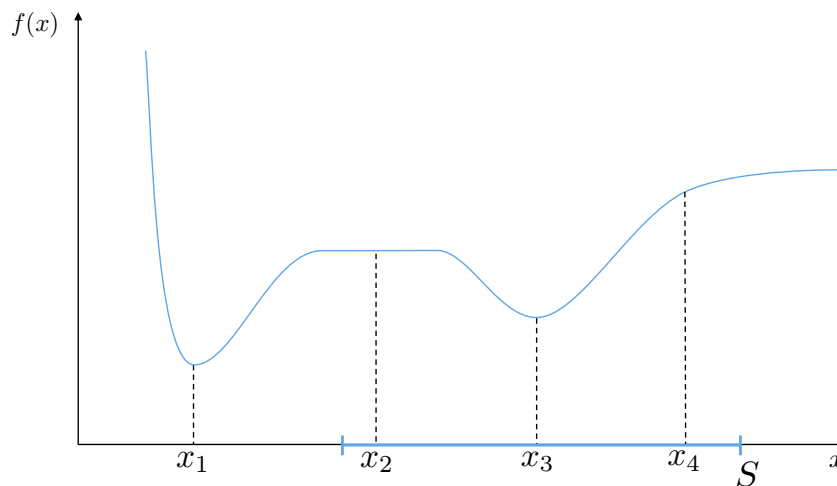

Figure 1: Examples of local and global optima

2 The role of convexity in optimality conditions

We can now state what is possibly the most important result in optimisation. In a nutshell, this results allows one promote local optimality to global optimality in the presence of convexity.

> **Theorem 2.1. Global optimality of convex problems**
>
> Let $S \subseteq \mathbb{R}^n$ be a nonempty convex set and $f : S \mapsto \mathbb{R}$ 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
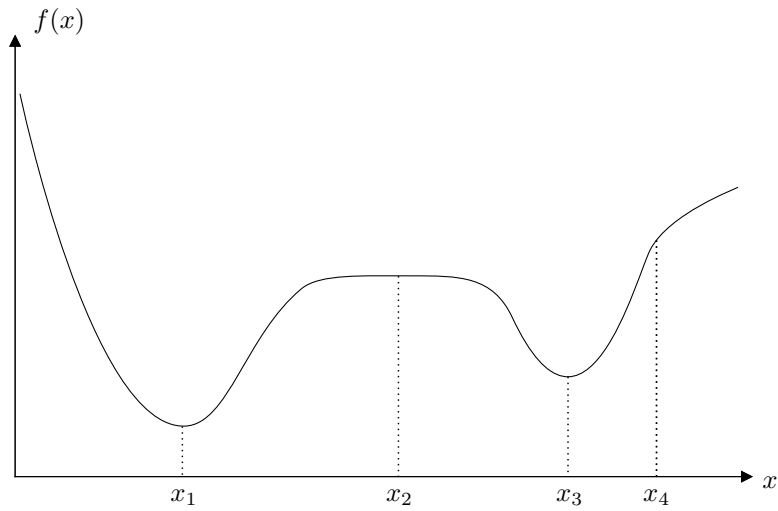
(a) Unconstrained optimisation problem
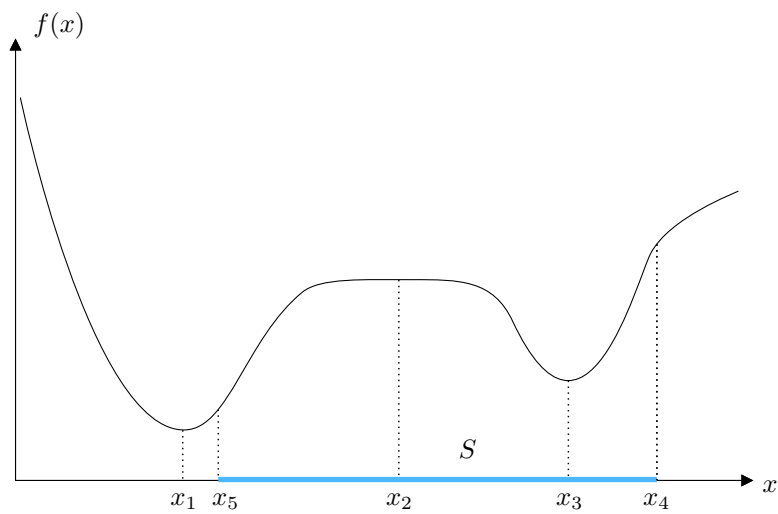
(b) Constrained optimisation problem

Figure 2: Points of interest in optimisation. Points x_1, x_2 and x_3 are local optima in the unconstrained problem. Once a constraint set S is imposed, x_4 and x_5 become points of interest and x_1 becomes infeasible.

> 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:
>
> $$f(\lambda \hat{x} + (1 - \lambda)\overline{x}) \leq \lambda f(\hat{x}) + (1 - \lambda)f(\overline{x}) < \lambda f(\overline{x}) + (1 - \lambda)f(\overline{x}) = f(\overline{x})$$
>
> However, for $\lambda > 0$ sufficiently small, $\lambda \hat{x} + (1 - \lambda)\overline{x} \in S \cap N_\epsilon(\overline{x})$ due to the convexity of S, which contradicts the local optimality of $\overline{x}$. Thus, $\overline{x}$ is a global optimum.
>
> 😎

The proof is built using contradiction. That is, we show that for a solution to be a local optimum in a convex problem, not being a global solution contradicts its local optimality, originally true by assumption. This is achieved using the convexity of f and showing that the convex combination between hypothetical better solution $\hat{x}$ and $\overline{x}$ would have to be both in $N_\epsilon(\overline{x})$ and better than $\overline{x}$, contradicting the local optimality of $\overline{x}$.

3

3 Optimality condition of convex problems

We first look at optimality conditions in a general sense to then translate the concept to unconstrained and constrained problems specifically. Taking this more general standpoint is also helpful to understand how these can be specialised in the absence of a closed domain or in the presence of differentiability. We assume convexity for now, and later we will discuss further the consequences of the absence of convexity. Note that unconstrained problems have convex feasibility set (i.e., the whole $\mathbb{R}^n$), and thus what follows can be generalised to unconstrained optimisation problems.

> **Theorem 3.1. Optimality condition for convex problems**
>
> Let $S \subseteq \mathbb{R}^n$ be a nonempty convex set and $f : \mathbb{R}^n \mapsto \mathbb{R}$ 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 $x \in S$.

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 $x \in S$

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

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

Conversely, suppose that $\overline{x}$ is a global 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:

$$\xi_0^\top(x - \overline{x}) + \mu y \leq \alpha, \ \forall x \in \mathbb{R}^n, \ y > f(x) - f(\overline{x}) \tag{1}$$
$$\xi_0^\top(x - \overline{x}) + \mu y \geq \alpha, \ \forall x \in S, \ y \leq 0. \tag{2}$$
$$\tag{3}$$

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 $\alpha = 0$.

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:

$$\xi^\top(x - \overline{x}) \leq y, \ \forall x \in \mathbb{R}^n, \ y > f(x) - f(\overline{x}) \tag{4}$$
$$\xi^\top(x - \overline{x}) \geq y, \ \forall x \in S, \ y \leq 0 \tag{5}$$

Letting $y = 0$ in (5), we get $\xi^\top(x - \overline{x}) \geq 0$ for all $x \in S$. From (4), 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 $x \in S$. 😎

In the first part of the proof, we use the definition of convexity based on the subgradient inequality to show that $\xi^\top(x - \overline{x}) \geq 0$ implies that $f(\overline{x}) \leq f(x)$ for all $x \in S$. The second part of the proof uses the separation theorem in a creative way to show that the subgradient inequality must hold if $\overline{x}$ is optimal. This is achieved by using the two sets Λ_1 and Λ_2. Notice that, $\overline{x}$ being optimal implies that $y > f(x) - f(\overline{x}) \geq 0$, which leads to the conclusion that $\Lambda_1 \cap \Lambda_2 = \emptyset$, demonstrating the existence of a separating hyperplane between them, as shown in (1) and (2). We can show that α in those has to be 0 by noticing that $\mu\epsilon \leq 0$ must hold for $\epsilon > 0$ to be a bounded constant.

The second part is dedicated to show that $\mu < 0$, so we can divide (1) and (2) by μ to obtain the subgradient inequality as we have seen. We show that by contradiction, since $\mu = 0$ would imply $\xi_0 = 0$, which disagrees with existence of a $(\xi, \mu) \neq 0$ in the separation theorem. Finally, as $y > f(x) - f(\overline{x})$ and $y \geq \xi^\top(x - \overline{x})$, for any given y, we have that $f(x) - f(\overline{x}) \geq \xi^\top(x - \overline{x})$ [1], which leads to the subgradient inequality.

Notice that this result provides necessary and sufficient conditions for optimality for convex problems. These conditions can be extended to the unconstrained case as well, which is presented in Corollary 3.2.

Corollary 3.2 (Optimality in open sets). Under the conditions of Theorem 3.1, 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 3.1, $\overline{x}$ 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 $-\lambda||\xi||^2 \geq 0$, implying $\xi = 0$. 😎

Notice that, if S is open, then the only way to attain the condition $\xi^\top(x - \overline{x}) \geq 0$ is if $\xi = 0$ itself. This is particularly relevant in the context of nondifferentiable functions, as we will see later. Another important corollary is the classic optimality condition $\nabla f(\overline{x}) = 0$, which we state below for completeness.

Corollary 3.3 (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 $\nabla f(\overline{x}) = 0$.

The proof for Corollary 3.3 is the same as Theorem 3.1 under a setting where $\partial(x) = \{\nabla f(x)\}$ due to the differentiability of f.

Let us consider two examples. First, consider the problem:

$$\text{min.} \ \left(x_1 - \frac{3}{2}\right)^2 + (x_2 - 5)^2$$
$$\text{subject to:} \ -x_1 + x_2 \leq 2$$
$$2x_1 + 3x_2 \leq 11$$
$$x_1 \geq 0$$
$$x_2 \geq 0$$

Figure 3 presents a plot of the feasible region S, which is form by the intersection of the two half-spaces, and the level curves of the objective function, with some of the values indicated in the curves. Notice that that this is a convex problem.

The arrow shows the gradient $\nabla f(\overline{x})$ at $\overline{x} = (1, 3)$. Notice that this point is special since at that point, no vector $x - \overline{x}$ can be found forming an angle greater than 90° with $\nabla f(\overline{x})$, that is $\nabla f(\overline{x})^\top(x - \overline{x}) \geq 0$ for any $x \in S$, which means that $\overline{x}$ is optimal. Since the problem is convex, that is in fact the global optimum for this problem.

Figure 4 shows a similar situation, but now with one of the constraints being nonlinear. Notice that of the two points highlighted ((1,2) in orange and (2,1) in purple), the optimality condition only holds for (2,1). For example, for $x = (2, 1)$ and $\overline{x} = (1, 2)$ the vector $x - \overline{x}$ forms a angle greater than 90° with the gradient of f at $\overline{x}$, $\nabla f(\overline{x})$, and thus the condition $\nabla f(\overline{x})^{(}x - \overline{x}) \geq 0$ does not hold for all S. The condition does hold for $\overline{x} = (2, 1)$, as can be seen in Figure 4.

A geometrical interpretation of the optimality condition $\xi^\top(x - \overline{x}) \geq 0$ is as follows. If there exists a subgradient ξ (or a gradient $\nabla f(\overline{x})$ if f is differentiable) that serves as a separating hyperplane between the level curve of f at $\overline{x}$ and the feasible region S, then there can be no feasible point further into the

---

[1] Notice that, on the line of nonnegative reals, for a same y, $f(x) - f(\overline{x})$ is always on the 'right side' of $\xi^\top(x - \overline{x})$ because it is an open interval.
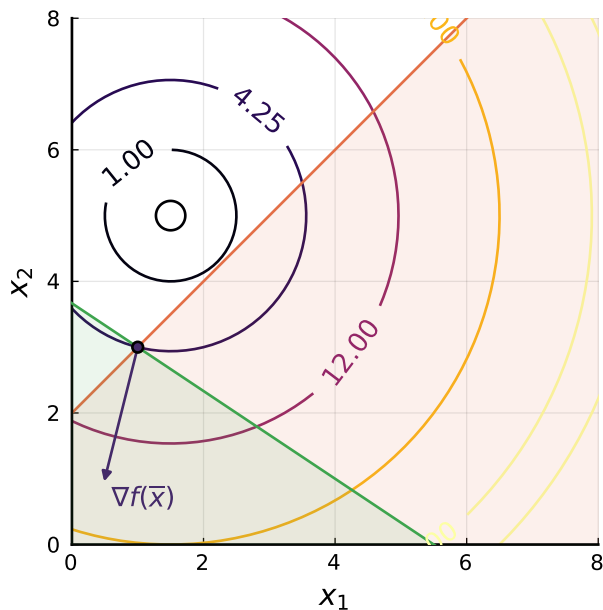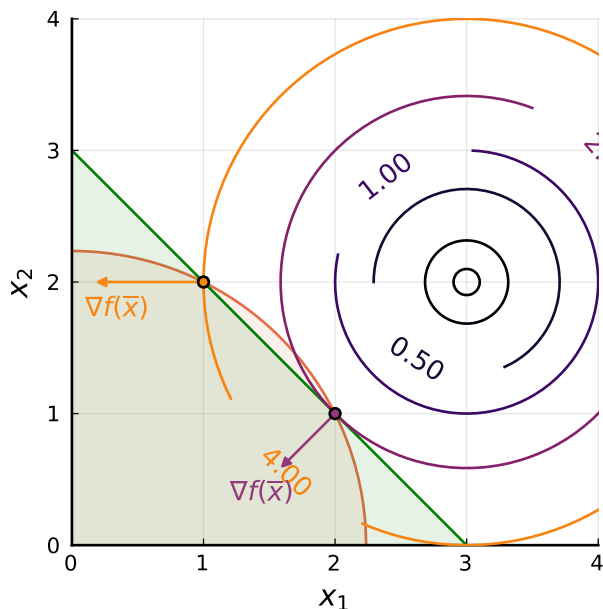
Figure 3: Example 1

Figure 4: Example 2

lower level set defined by that level curve. Ultimately, this means that there is no feasible point with smaller objective function value to be found. This is why the separation theorem from Lecture 2 plays an important role here, since it can be used to state that the feasible options have been exhausted in terms of potential directions of decrease of objective function value.

3.1 Optimality conditions for unconstrained problems

We have developed most of the concepts required to state optimality conditions for unconstrained optimisation problems, as presented in Corollaries 3.2 and 3.3. We now take an alternative route in which we do not take into account the feasibility set, but only the differentiability of f. This will be useful as it will allow us to momentarily depart from the assumption of convexity, which was used to state Theorem 3.1.

3.1.1 First-order optimality conditions

Let us start defining what it means to be a *descent direction*.

> **Theorem 3.4. Descent direction**
>
> Suppose $f : \mathbb{R}^n \mapsto \mathbb{R}$ 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:

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

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 $f(\overline{x} + \lambda d) - f(\overline{x}) < 0$.

😎

The proof uses the first-order expansion around $\overline{x}$ to show that, f being differentiable, the condition $\nabla f(\overline{x})^\top d < 0$ implies that $f(\overline{x} + \lambda d) < f(\overline{x})$, or put in words, that a step in the direction d decreases the objective function value.

We can derive the first-order optimality condition in Corollary 3.3 as a consequence from Theorem 3.4. Notice, however, that since convexity is not assumed, all we can say is that this condition is necessary (but not sufficient) for local optimality.

> **Corollary 3.5** (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 $\nabla f(\overline{x}) \neq 0$. Letting $d = -\nabla f(\overline{x})$, we have that $\nabla f(\overline{x})^\top d = -||\nabla f(\overline{x})||^2 < 0$. By Theorem 3.4, there exists a $\delta > 0$ such that $f(\overline{x} + \lambda d) < f(\overline{x})$ for all $\lambda \in (0, \delta)$, thus contradicting the local optimality of $\overline{x}$.

😎

Notice that Corollary 3.5 only holds in one direction. The proof uses contradiction once again, where we assume local optimality of $\overline{x}$ and show that having $\nabla f(\overline{x}) \neq 0$ contradicts the local optimality of $\overline{x}$, our initial assumption. To do that, we simply show that having any descent direction d (we use $-\nabla f(\overline{x})$ since in this setting it is guaranteed to exist as $\nabla f(\overline{x}) \neq 0$) would mean that small step λ can reduce the objective function value, contradicting the local optimality of $\overline{x}$.

3.1.2 Second-order optimality conditions

We now derive necessary conditions for local optimality of $\overline{x}$ based on second-order differentiability. As we will see, it requires that the Hessian $H(\overline{x})$ of $f(x)$ at $\overline{x}$ is positive semidefinite.

> **Theorem 3.6. 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 3.5 implies that $\nabla f(\overline{x}) = 0$ and $f(\overline{x} + \lambda d) \geq f(\overline{x})$.

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 $d^\top H(\overline{x}) d \geq 0$.

😎

The second-order conditions can be used to attest local optimality of $\overline{x}$. In the case where $H(\overline{x})$ is positive definite, then this second order condition becomes *sufficient* for local optimality, since it implies that the function is 'locally convex' for a small enough neighbourhood $N_\epsilon(\overline{x})$.

In case f is convex, then the first-order condition $\nabla f(x) = 0$ becomes also sufficient for attesting the global optimality of $\overline{x}$. Recall that f is convex if and only if $H(x)$ is positive semidefinite for all $x \in \mathbb{R}^n$, meaning that in this case the second-order necessary conditions are also satisfied at $\overline{x}$.

Theorem 3.7

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

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

Notice that $\nabla f(\overline{x}) = 0$ 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 $x \in \mathbb{R}^n$.
😎