CHAPTER IV: Convex Optimization Problems

1 Basic definitions

Consider the optimization problem

$$\begin{aligned} \underset{\boldsymbol{x}\in\mathbb{R}^n}{\textbf{minimize}} \quad & f_0(\boldsymbol{x}) \\ \textit{s.t.} \quad & f_i(\boldsymbol{x}) \leq 0 \qquad (\forall i \in [m]). \end{aligned} \tag{P}$$

- The function f_0 is called the *objective function*.
- The inequalities $f_i(\boldsymbol{x}) \leq 0$ (for $i \geq 1$) are called *constraints*.
- A vector $\boldsymbol{x} \in \mathbb{R}^n$ is called *feasible*, or a *feasible solution*, if it satisfies all the constraints, i.e.,

$$f_i(\boldsymbol{x}) \leq 0 \quad (\forall i \in [m]).$$

- The set $\mathcal{F} = \{\boldsymbol{x} \in \mathbb{R}^n |\ f_1(\boldsymbol{x}) \leq 0, \ldots, f_m(\boldsymbol{x}) \leq 0\}$ of all feasible solutions is called the *feasible set*.
- The value $p^* = \inf\{f_0(\boldsymbol{x}) |\ \boldsymbol{x} \in \mathcal{F}\} \in \mathbb{R} \cup \{-\infty, +\infty\}$ is the *optimal value* of the problem.
- $p^* = +\infty$ iff the feasible set $\mathcal{F}$ is empty; then we say that Problem (P) is *infeasible*.
- $p^* = -\infty$ iff there is a sequence $\boldsymbol{x}_i \in \mathcal{F}$ with $f_0(\boldsymbol{x}_i) \to -\infty$; in this case, (P) is *unbounded from below*.
- When f_0 is a constant function, the problem asks to find any feasible vector $\boldsymbol{x} \in \mathcal{F}$. In this case (P) is a *feasibility problem*.
- If $\boldsymbol{x}^* \in \mathcal{F}$ satisfies $f_0(\boldsymbol{x}^*) = p^*$, we say that $\boldsymbol{x}^*$ is an *optimal solution* of the problem, or that $\boldsymbol{x}^*$ *solves* (P). Sometimes, we'll say that $\boldsymbol{x}^*$ is a *global optimum*, to stress the difference with *local optima* (cf. definition below). The set of of all optimal solutions of (P) is called the *optimal set*.
- The vector $\boldsymbol{x}$ is called *ϵ-suboptimal* if it is feasible and $f_0(\boldsymbol{x}) \leq p^* + \epsilon$. The set of all ϵ-suboptimal solutions is called the *ϵ-suboptimal set*.

Remark: The constraints are understood in the sense of extended values of the functions, i.e., $\boldsymbol{x} \notin \mathbf{dom}\, f_i \implies f_i(\boldsymbol{x}) = \infty$. Hence,

$$\mathcal{F} \subseteq \bigcap_{i=0}^{m} \mathbf{dom}\, f_i.$$

Definition 1 (local optimum)**.** The vector $\boldsymbol{x}$ is called a *local optimum* for Problem (P) if it solves the problem

$$\begin{aligned} \underset{\boldsymbol{z}\in\mathbb{R}^n}{\textbf{minimize}} \quad & f_0(\boldsymbol{z}) \\ \textit{s.t.} \quad & f_i(\boldsymbol{z}) \leq 0 \qquad (\forall i \in [m]); \\ & \|\boldsymbol{z} - \boldsymbol{x}\| \leq R \end{aligned} \tag{P$_\text{R}$}$$

for some $R > 0$. In other words, there is a neighbourhood of $\boldsymbol{x}$ in $\mathcal{F}$ in which f_0 is minimized at $\boldsymbol{x}$.

The next proposition characterizes the local optima that occur **in the interior** of the feasible region:

Proposition 1 (Differential characterization of local optima)**.** *Assume the objective function f_0 is twice differentiable, and let $\boldsymbol{x} \in \mathbf{int}\,\mathcal{F}$. Then, the following holds:*

- *If $\nabla f_0(\boldsymbol{x}) = \mathbf{0}$ and $\nabla^2 f_0(\boldsymbol{x}) \succ 0$, then $\boldsymbol{x}$ is a local optimum.*
- *Conversely, if $\boldsymbol{x}$ is a local optimum, then $\nabla f_0(\boldsymbol{x}) = \mathbf{0}$ and $\nabla^2 f_0(\boldsymbol{x}) \succeq 0$.*

Note that the converse statement is a bit weaker ($\succeq$ instead of $\succ$). The following examples show that we cannot strengthen the above result:

- Consider $f(x) = x^3$ over $\mathbb{R}$. Then, $\nabla^2 f(0) \succeq 0$ (we have $f''(0) = 0$), but x is not a local optimum.
- Consider $f(x) = x^4$ over $\mathbb{R}$. Then, $x = 0$ is a local (and even, global) optimum, but $\nabla^2 f(0) = f''(0) = 0$;

Let us now prove this proposition.

Proof. For the first statement, we use the Taylor expansion of f_0. For all $\boldsymbol{h} \in \mathbb{R}^n$ with $\|\boldsymbol{h}\| = 1$,

$$f_0(\boldsymbol{x} + \epsilon\boldsymbol{h}) = f_0(\boldsymbol{x}) + \epsilon \nabla f_0(\boldsymbol{x})^T \boldsymbol{h} + \frac{\epsilon^2}{2} \boldsymbol{h}^T \nabla^2 f_0(\boldsymbol{x}) \boldsymbol{h} + o(\epsilon^2).$$

(For those not familiar with the "little-o" notation, you can replace the $o(\epsilon^2)$ by $\epsilon^2\eta(\epsilon)$, where η is a function converging to 0 as $\epsilon \to 0$). Now, we use $\nabla f_0(\boldsymbol{x}) = \mathbf{0}$ and $\nabla^2 f_0(\boldsymbol{x}) \succ 0$, which gives

$$f_0(\boldsymbol{x} + \epsilon\boldsymbol{h}) - f_0(\boldsymbol{x}) = \epsilon^2 \left(\frac{1}{2} \underbrace{\boldsymbol{h}^T \nabla^2 f_0(\boldsymbol{x}) \boldsymbol{h}}_{>0} + \underbrace{o(1)}_{=\eta(\epsilon)\to 0} \right).$$

So, for all $\boldsymbol{h}$, there exists $\epsilon(\boldsymbol{h}) > 0$ such that $f_0(\boldsymbol{x}) \leq f_0(\boldsymbol{x} + t\boldsymbol{h})$, for all $0 \leq t \leq \epsilon(\boldsymbol{h})$. Finally, setting $R > 0$ to the minimum value of $\epsilon(\boldsymbol{h})$ over the unit sphere shows that $\boldsymbol{x}$ is a local minimum.

For the converse statement, let $\boldsymbol{h} \in \mathbb{R}^n$, and consider the restriction to a line:

$$g : t \mapsto f_0(\boldsymbol{x} + t\boldsymbol{h}).$$

so that $g'(0) = \nabla f_0(\boldsymbol{x})^T \boldsymbol{h}$ and $g''(0) = \boldsymbol{h}^T \nabla^2 f_0(\boldsymbol{x}) \boldsymbol{h}$. Since $\boldsymbol{x}$ is a local minimum, $g(t) \geq g(0)$ for $t > 0$ small enough, hence $g'(0) = \lim_{t\to 0^+} \frac{g(t)-g(0)}{t} \geq 0$. Similarly, $g(t) \geq g(0)$ for $t < 0$ with $|t|$ small enough, hence $g'(0) = \lim_{t\to 0^-} \frac{g(t)-g(0)}{t} \leq 0$. So, the equality $g'(0) = \nabla f_0(\boldsymbol{x})^T \boldsymbol{h} = 0$ must hold for all $\boldsymbol{h} \in \mathbb{R}^n$, which implies $\nabla f_0(\boldsymbol{x}) = \mathbf{0}$.

Then, we note that both $g(t) \geq g(0)$ and $g(-t) \geq g(0)$ for small enough $|t|$, which implies that $g(t)+g(-t)-2g(0) \geq 0$ for small enough $|t|$. Hence, $\boldsymbol{h}^T \nabla^2 f_0(\boldsymbol{x}) \boldsymbol{h} = g''(0) = \lim_{t\to 0} \frac{g(t)+g(-t)-2g(0)}{t^2} \geq 0$ for all $\boldsymbol{h} \in \mathbb{R}^n$, which corresponds to the definition of $\nabla^2 f_0(\boldsymbol{x}) \succeq 0$. □

It is possible to include an *equality constraint* in a problem of the form (P), by considering two reversed inequalities $f(\boldsymbol{x}) \leq 0$ and $-f(\boldsymbol{x}) \leq 0$. If all the functions $f_0, \ldots, f_m$ are convex, then we say that (P) is a *convex optimization problem.* If there is an equality constraint $f_i(\boldsymbol{x}) = 0$ in a convex optimization problem, then f_i must be both concave and convex, hence f_i is affine. Therefore, it is often convenient to write the equality constraints separately in a convex optimization problem:

$$\begin{aligned} \underset{\boldsymbol{x}\in\mathbb{R}^n}{\mathbf{minimize}} \quad & f_0(\boldsymbol{x}) \\ s.t. \quad & f_i(\boldsymbol{x}) \leq 0 \qquad (\forall i \in [m]); \\ & A\boldsymbol{x} = \boldsymbol{b}. \end{aligned} \tag{P$_{\mathrm{Eq}}$}$$

Remark: The above definitions can be extended in the obvious manner for maximization problems (since maximizing $f_0(\boldsymbol{x})$ is the same as minimizing $-f_0(\boldsymbol{x})$). Hence, the problem "$\mathbf{maximize}_{\boldsymbol{x}\in\mathcal{F}} f_0(\boldsymbol{x})$" is called *convex* if f_0 is concave and the $f_i's$ are convex ($\forall i \geq 1$).

The next result states an essential property of convex optimization problems, which contributes to make them efficiently solvable:

> **Theorem 2.** *Let* (P) *be a convex optimization problem. Then, any local optimum* $\boldsymbol{x}^*$ *is also a global optimum.*

Proof. Let $\boldsymbol{x}^*$ be a local optimum. Then, there exist a $R > 0$ such that

$$\boldsymbol{z} \in \mathcal{F},\ \|\boldsymbol{z} - \boldsymbol{x}^*\| \leq R \implies f_0(\boldsymbol{z}) \geq f_0(\boldsymbol{x}^*).$$

Now, we assume (by contradiction) that $\boldsymbol{x}^*$ is not globally optimal, i.e., there exists an $\boldsymbol{y} \in \mathcal{F}$ such that $f_0(\boldsymbol{y}) < f_0(\boldsymbol{x}^*)$. Let $\boldsymbol{z} = \theta\boldsymbol{y} + (1-\theta)\boldsymbol{x}^*$, where $\theta = \frac{R}{2\|\boldsymbol{y}-\boldsymbol{x}^*\|}$ is chosen so that $\|\boldsymbol{z} - \boldsymbol{x}^*\| = \frac{R}{2}$. We have $\boldsymbol{z} \in \mathcal{F}$ by convexity of $\mathcal{F}$; it then follows that $f_0(\boldsymbol{z}) \geq f_0(\boldsymbol{x}^*)$, but this contradicts the convexity inequality of f_0:

$$f_0(\boldsymbol{z}) \leq \theta f_0(\boldsymbol{y}) + (1-\theta) f_0(\boldsymbol{x}^*) < f_0(\boldsymbol{x}^*).$$

□

2 Problem reformulations

In this section, we review a few techniques useful to reformulate a problem to another *equivalent problem.* Sometimes, this is necessary, because the new problem will have better properties (e.g. solution can be characterized more easily, or it can be solved more efficiently). This, of course, would require a proper definition of what it means for two problems to be equivalent. While this can be stated in a more rigorous manner, we'll just say that Problems P_1 and P_2 are *equivalent* if an optimal solution of P_1 can be transformed to an optimal solution of P_2, and vice versa. This will become clear with the following examples. We use the (nonstandard) notation $P_1 \dot{\sim} P_2$ to say that P_1 and P_2 are equivalent.

(a) **Eliminating (or adding) equality constraints:** When a problem contains equality constraints "$A\boldsymbol{x} = \boldsymbol{b}$", we can remove these constraints by using a simple change of variables. Indeed, $A\boldsymbol{x} = \boldsymbol{b}$ means that $\boldsymbol{x}$ belongs to some affine set L, and we know that L admits an alternative representation of the form $L = \{C\boldsymbol{z} + \boldsymbol{d} : \boldsymbol{z} \in \mathbb{R}^r\}$. (To obtain such a representation, take for $\boldsymbol{d}$ a particular solution[1] of the equation $A\boldsymbol{x} = \boldsymbol{b}$, and let $C \in \mathbb{R}^{n\times r}$ be a matrix whose columns form a basis of $\mathbf{Ker}\, A$.) Then, with the change of variables $\boldsymbol{x} = C\boldsymbol{z} + \boldsymbol{d}$, Problem ($\mathrm{P_{Eq}}$) can be seen to be equivalent to

$$\begin{aligned} \underset{\boldsymbol{z}\in\mathbb{R}^r}{\mathbf{minimize}} \quad & f_0(C\boldsymbol{z} + \boldsymbol{d}) \\ s.t. \quad & f_i(C\boldsymbol{z} + \boldsymbol{d}) \leq 0 \qquad (\forall i \in [m]); \end{aligned}$$

If ($\mathrm{P_{Eq}}$) is convex, then the new problem is convex, too (composition with an affine mapping). An optimal solution $\boldsymbol{z}^*$ of the above problem readily gives an optimal solution $\boldsymbol{x}^* = C\boldsymbol{z}^* + \boldsymbol{d}$ of ($\mathrm{P_{Eq}}$), and conversely, an optimal solution $\boldsymbol{x}^*$ can be transformed to an optimal $\boldsymbol{z}^*$ by solving the equation $C\boldsymbol{z}^* = \boldsymbol{x}^* - \boldsymbol{d}$ for $\boldsymbol{z}^*$.

(b) **Slack variables** One can replace linear inequalities by linear equalities by introducing a *slack variable.* Indeed,

$$\boldsymbol{a}_i^T\boldsymbol{x} \leq b \iff \exists s \geq 0 : \boldsymbol{a}_i^T\boldsymbol{x} + s = b.$$

[1] For example, one solution is given by $\boldsymbol{d} = A^\dagger\boldsymbol{b}$, where $A^\dagger$ is the Moore-Penrose pseudo inverse of A; when A has full column rank, $A^\dagger = (A^TA)^{-1}A^T$.

This means that we obtain an equivalent problem when we replace the constraint "$\boldsymbol{a}_i^T\boldsymbol{x} \leq b$" by the system of two constraints "$\boldsymbol{a}_i^T\boldsymbol{x} + s = b,\ s \geq 0$", which involves the new set of variables $(\boldsymbol{x}, s) \in \mathbb{R}^{n+1}$.

(c) **Change of variables:** If $\phi : \mathbb{R}^n \to \mathbb{R}^n$ is one-to-one, and every feasible $\boldsymbol{x}$ can be written as $\boldsymbol{x} = \phi(\boldsymbol{z})$ for some $\boldsymbol{z}$, i.e. $\phi(\mathbf{dom}\,\phi) \supseteq \mathcal{F}$, then we can make the substitution $\boldsymbol{x} = \phi(\boldsymbol{z})$ to obtain an equivalent problem. That is, Problem (P) is equivalent to

$$\begin{aligned} \underset{\boldsymbol{z}\in\mathbb{R}^n}{\textbf{minimize}} \quad & f_0(\phi(\boldsymbol{z})) \\ s.t. \quad & f_i(\phi(\boldsymbol{z})) \leq 0 \qquad (\forall i \in [m]). \end{aligned}$$

(d) **Transformation of objective or constraints:** If $\psi_0 : \mathbb{R} \to \mathbb{R}$ is strictly increasing and for all $i \in [m]$, $\psi_i : \mathbb{R} \to \mathbb{R}$ satisfies $\psi_i(u) \leq 0 \iff u \leq 0$, then Problem (P) is equivalent to

$$\begin{aligned} \underset{\boldsymbol{x}\in\mathbb{R}^n}{\textbf{minimize}} \quad & \psi_0(f_0(\boldsymbol{x})) \\ s.t. \quad & \psi_i(f_i(\boldsymbol{x})) \leq 0 \qquad (\forall i \in [m]). \end{aligned}$$

(e) **Epigraph reformulation:** It is always possible to assume that the objective function of a problem is linear. Indeed, Problem (P) is equivalent to

$$\begin{aligned} \underset{\boldsymbol{x}\in\mathbb{R}^n, t\in\mathbb{R}}{\textbf{minimize}} \quad & t \\ s.t. \quad & f_i(\boldsymbol{x}) \leq 0 \qquad (\forall i \in [m]); \\ & f_0(\boldsymbol{x}) \leq t. \end{aligned}$$

(f) **Partial minimization:** When the constraints of the problem involve different blocks of variables, it is possible to reduce the problem by solving it (partially) for some blocks of variables. For example, the following two problems are equivalent:

$$\begin{aligned} \underset{\boldsymbol{x}_1\in\mathbb{R}^{n_1}, \boldsymbol{x}_2\in\mathbb{R}^{n_2}}{\textbf{minimize}} \quad & f_0(\boldsymbol{x}_1, \boldsymbol{x}_2) \\ s.t. \quad & f_i(\boldsymbol{x}_1) \leq 0 \quad (\forall i \in [m_1]) \\ & g_j(\boldsymbol{x}_1, \boldsymbol{x}_2) \leq 0 \quad (\forall j \in [m_2]) \end{aligned} \qquad \dot\sim \qquad \begin{aligned} \underset{\boldsymbol{x}_1\in\mathbb{R}^{n_1}}{\textbf{minimize}} \quad & \tilde{f}_0(\boldsymbol{x}_1) \\ s.t. \quad & f_i(\boldsymbol{x}_1) \leq 0 \quad (\forall i \in [m_1]), \end{aligned}$$

where we have defined $\tilde{f}_0(\boldsymbol{x}_1) := \inf\big\{f_0(\boldsymbol{x}_1, \boldsymbol{x}_2) \mid\ \boldsymbol{x}_2 \in \mathbb{R}^{n_2},\ g_j(\boldsymbol{x}_1, \boldsymbol{x}_2) \leq 0\ \ \forall j \in [m_2]\big\}$.

3 First order optimality conditions

When the objective function f_0 is differentiable and Problem (P) is convex, we can derive a simple optimality condition, which depends only on ∇f_0 and the feasibility set $\mathcal{F} = \{\boldsymbol{x} \in \mathbb{R}^n \mid\ f_1(\boldsymbol{x}) \leq 0, \ldots, f_m(\boldsymbol{x}) \leq 0\}$.

Theorem 3. *Let f_0 be differentiable. Then, a vector $\boldsymbol{x}^*$ is optimal for the convex problem* (P) *if and only if $\boldsymbol{x}^*$ is feasible (i.e., $\boldsymbol{x}^* \in \mathcal{F}$), and*

$$\forall \boldsymbol{y} \in \mathcal{F}, \quad \nabla f_0(\boldsymbol{x}^*)^T(\boldsymbol{y} - \boldsymbol{x}^*) \geq 0.$$

Proof. $\Longleftarrow$: We use the first order condition of convexity for f_0:

$$\forall \boldsymbol{y} \in \mathcal{F},\ f_0(\boldsymbol{y}) \geq f_0(\boldsymbol{x}^*) + \underbrace{\nabla f_0(\boldsymbol{x}^*)^T(\boldsymbol{y} - \boldsymbol{x}^*)}_{\geq 0} \geq f_0(\boldsymbol{x}^*).$$

Hence, $\boldsymbol{x}^*$ is optimal.

$\Longrightarrow$: We will prove the contrapositive statement. Assume there exists $\boldsymbol{y} \in \mathcal{F}$ such that $\nabla f_0(\boldsymbol{x}^*)^T(\boldsymbol{y}-\boldsymbol{x}^*) < 0$. We recognize that $\nabla f_0(\boldsymbol{x}^*)^T(\boldsymbol{y}-\boldsymbol{x}^*)$ is the directional derivative of f_0 is the direction of $(\boldsymbol{y}-\boldsymbol{x}^*)$, that is,

$$\nabla f_0(\boldsymbol{x}^*)^T(\boldsymbol{y}-\boldsymbol{x}^*) \;=\; \frac{d}{dt}\left(t \mapsto f_0(\boldsymbol{x}^* + t(\boldsymbol{y}-\boldsymbol{x}^*))\right)\Big|_{t=0} \;=\; \lim_{t\to 0} \frac{f_0(\boldsymbol{x}^* + t(\boldsymbol{y}-\boldsymbol{x}^*)) - f_0(\boldsymbol{x}^*)}{t} < 0.$$

Hence, for $t>0$ small enough, we have $f_0(\boldsymbol{x}^* + t(\boldsymbol{y}-\boldsymbol{x}^*)) < f(\boldsymbol{x}^*)$, and $\boldsymbol{x}^* + t(\boldsymbol{y}-\boldsymbol{x}^*) \in \mathcal{F}$ by convexity of $\mathcal{F}$. This shows that $\boldsymbol{x}^*$ is not optimal. ☐

Remark: This theorem has a simple geometric interpretation. It says that $\boldsymbol{x}^*$ is optimal if and only if $\nabla f_0(\boldsymbol{x}^*) = 0$, or if $\nabla f_0(\boldsymbol{x}^*)$ defines a supporting hyperplane of $\mathcal{F}$ at $\boldsymbol{x}^*$.

The theorem above already allows us to characterize optimal solutions in a number of simple situations. For example, when the problem is unconstrained (i.e., $\mathcal{F} = \mathbb{R}^n$), we get the well-known fact that $\boldsymbol{x}^*$ is optimal iff it solves the equation $\nabla f_0(\boldsymbol{x}^*) = 0$. The next result characterizes optimal solutions of a problem with equality constraints only

Proposition 4. *Consider the optimization problem*

$$\begin{aligned} \underset{\boldsymbol{x}\in\mathbb{R}^n}{\textbf{minimize}} \quad & f_0(\boldsymbol{x}) \\ s.t. \quad & A\boldsymbol{x} = \boldsymbol{b}, \end{aligned}$$

where f_0 is convex and differentiable. Then, $\boldsymbol{x}^*$ is optimal iff $\nabla f_0(\boldsymbol{x}^*) \in \mathbf{Im}\, A^T$.

Proof. For any $\boldsymbol{x}^*$ in the affine set $\mathcal{F} = \{\boldsymbol{x} : A\boldsymbol{x} = \boldsymbol{b}\}$, it holds ($\boldsymbol{y} \in \mathcal{F} \iff \boldsymbol{y}-\boldsymbol{x}^* \in \mathbf{Ker}\, A$). So, the first order optimality condition for $\boldsymbol{x}^*$ is

$$\forall \boldsymbol{y} \in \mathcal{F},\ \nabla f_0(\boldsymbol{x}^*)^T(\boldsymbol{y}-\boldsymbol{x}^*) \geq 0 \iff \forall \boldsymbol{v} \in \mathbf{Ker}\, A,\ \nabla f_0(\boldsymbol{x}^*)^T \boldsymbol{v} \geq 0.$$

A linear function which is nonnegative everywhere on a vector space is necessarily constant. So, since $\mathbf{0} \in \mathbf{Ker}\, A$, the optimality condition can be rewritten as $\forall \boldsymbol{v} \in \mathbf{Ker}\, A,\ \nabla f_0(\boldsymbol{x}^*)^T \boldsymbol{v} = 0$. This is equivalent to saying that $\nabla f_0(\boldsymbol{x}^*)$ is orthogonal to $\mathbf{Ker}\, A$, that is, $\nabla f_0(\boldsymbol{x}^*) \in \mathbf{Im}\, A^T$. ☐

Another simple case, which occurs often in practice, is the case of optimization over the nonnegative orthant $\mathbb{R}^n_+$.

Proposition 5. *Consider the optimization problem*

$$\begin{aligned} \underset{\boldsymbol{x}\in\mathbb{R}^n}{\textbf{minimize}} \quad & f_0(\boldsymbol{x}) \\ s.t. \quad & \boldsymbol{x} \geq \mathbf{0}, \end{aligned}$$

where f_0 is convex and differentiable. Then, $\boldsymbol{x}^*$ is optimal iff the following condition holds:

$$\boldsymbol{x}^* \geq \mathbf{0}, \quad \nabla f_0(\boldsymbol{x}^*) \geq \mathbf{0}, \quad \textit{and} \quad \forall i \in [n], \left(x_i = 0 \ \textit{ or } \ \frac{\partial f_0}{\partial x_i}(\boldsymbol{x}^*) = 0\right).$$

Proof. The first order condition for $\boldsymbol{x}^* \in \mathcal{F}$ reads

$$\forall \boldsymbol{y} \geq 0,\ \nabla f_0(\boldsymbol{x}^*)^T(\boldsymbol{y}-\boldsymbol{x}^*) \geq 0.$$

This already implies $\nabla f_0(\boldsymbol{x}^*)^T \geq 0$ (Otherwise, if $\frac{\partial f_0}{\partial x_i}(\boldsymbol{x}^*) < 0$ for some i, the above condition would be violated when $y_i \to \infty$.) Now, for $\boldsymbol{y} = \mathbf{0}$, we obtain

$$\nabla f_0(\boldsymbol{x}^*)^T \boldsymbol{x}^* = \sum_i x_i^* \frac{\partial f_0}{\partial x_i}(\boldsymbol{x}^*) \leq 0.$$

Each term of this sum is the product of two nonnegative numbers, so it is nonnegative. It follows that the above inequality can only hold if for all $i \in [n]$, $x_i^* \cdot \frac{\partial f_0}{\partial x_i}(\boldsymbol{x}^*) = 0$. This yields the condition of the proposition.

Conversely, assume that the condition of the proposition holds, and let $\boldsymbol{y} \geq \mathbf{0}$. Then, for all i we have

$$\frac{\partial f_0}{\partial x_i}(\boldsymbol{x}^*) \cdot (y_i - x_i^*) \geq 0,$$

because this expression is either the product of two nonnegative terms, or it is equal to 0 whenever $\frac{\partial f_0}{\partial x_i}(\boldsymbol{x}^*) = 0$. Finally, summing over i yields the first order optimality condition for $\boldsymbol{x}^*$. □