$$(P) \quad \min_{x \in \mathcal{X}} f(x)$$

* A Taxonomy of Solutions to (P)

- Terminology: I will not distinguish b/w "solution" and "minimizer."

* Def. We say that $x^* \in \mathrm{dom}(f)$ is:

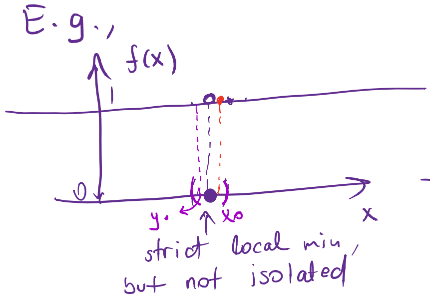
1) a local minimizer of (P) (or a local solution to (P)) if $\exists$ a neighborhood $\mathcal{N}_{x^*}$ of x^* s.t. $\forall x \in \mathcal{N}_{x^*} \cap \mathcal{X}$ we have $f(x) \geq f(x^*)$.

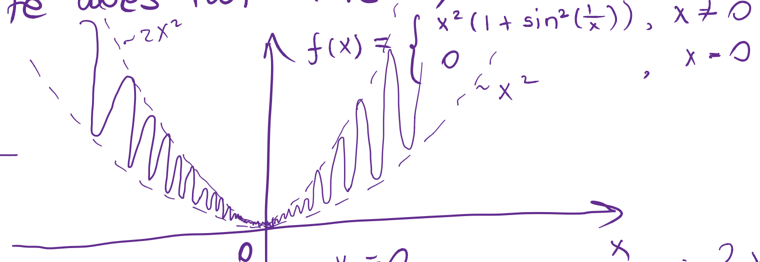
2) a strict local minimizer of (P) if it satisfies 1) but the inequality is strict: $f(x) > f(x^*)$.
3) a global minimizer of (P) if $\forall x \in \mathcal{X}$: $f(x) \geq f(x^*)$.
4) an isolated local minimizer of (P), if $\exists$ a neighborhood $\mathcal{N}_{x^*}$ s.t. $\forall x \in \mathcal{N}_{x^*} \cap \mathcal{X}$: $f(x) \geq f(x^*)$ and $\mathcal{N}_{x^*}$ does not contain any other local minimizers.

* Ex. Every isolated local minimizer is strict.

(But the opposite does not hold.)

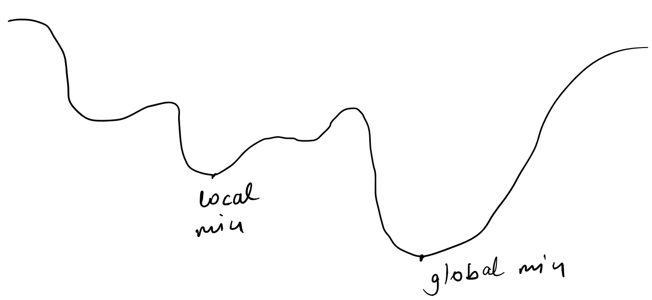
E.g.,





$x = 0$ is it a strict local min? Yes

Is it isolated? No.



* Taylor's Theorem:

→ For this part and until explicitly stated otherwise, we will be assuming that f is at least once cont.'ly diff.'able

→ Taylor's Theorem for 1D functions from calculus:

$f: \mathbb{R} \to \mathbb{R}$, f k-times cont.'ly diff.'able

$$\Rightarrow \forall x, y \in \mathbb{R}: f(y) = f(x) + \frac{1}{1!} f'(x)(y-x) + \frac{1}{2!} f''(x)(y-x)^2 + \dots + \frac{1}{k!} f^{(k)}(x)(y-x)^k + \underbrace{R_k(x)}_{\text{remainder}}$$

Typical forms of $R_k(x)$:
(assume that f is $(k+1)$-times cont.'ly diff.'able)

1) Lagrange (mean-value) remainder:

$$\exists \gamma \in (0,1): R_k(x) = \frac{1}{(k+1)!} f^{(k+1)}(x + \gamma(y-x))(y-x)^{k+1}$$

2) Integral remainder:

$$R_k(x) = \frac{1}{k!} \int_0^1 f^{(k+1)}(x + t(y-x))(y-x)^{k+1} \, dt$$

* Theorem (2.1 in WR) Let $f: \mathbb{R}^d \to \bar{\mathbb{R}}$ be a cont.'ly diff.'able function. Then, $\forall x, y \in \text{dom}(f)$ and s.t. $\forall \alpha \in (0,1)$ $(1-\alpha)x + \alpha y \in \text{dom}(f)$:

1) $f(y) = f(x) + \int_0^1 \langle \nabla f(x + t(y-x)), y - x \rangle \, dt$

2) $\exists \gamma \in (0,1): f(y) = f(x) + \langle \nabla f(x + \gamma(y-x)), y-x \rangle$.
(Mean Value Thm)

If f is twice cont.'ly diff.'able:

3) $\nabla f(y) = \nabla f(x) + \int_0^1 \underbrace{\nabla^2 f(x + t(y-x))}_{\text{Hessian matrix}} (y-x) \, dt$

" second-order derivative of f "

$$\nabla^2 f(x) = \begin{bmatrix} \frac{\partial^2 f}{\partial x_1^2}(x) & \frac{\partial^2 f}{\partial x_1 \partial x_2}(x) & \dots & \frac{\partial^2 f}{\partial x_1 \partial x_d}(x) \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial^2 f}{\partial x_d \partial x_1}(x) & \frac{\partial^2 f}{\partial x_d \partial x_2}(x) & \dots & \frac{\partial^2 f}{\partial x_d^2}(x) \end{bmatrix}$$

4) $\exists \gamma \in (0,1)$:

$$f(y) = f(x) + \langle \nabla f(x), y-x \rangle + \frac{1}{2}\langle \underbrace{\nabla^2 f(x+\gamma(y-x))(y-x), y-x}_{(y-x)^T \nabla^2 f(x+\gamma(y-x))(y-x)} \rangle$$

Q. Can you have:

$\cancel{\exists \gamma \in (0,1): \nabla f(y) = \nabla f(x) + \nabla^2 f(x+\gamma(y-x))(y-x)}$? NO

$\left(f(x) = \sum_i f_i(x_i) \right)$

* Properties of smooth functions:

* Terminology: f is L-smooth w.r.t. $\|\cdot\|$

$\iff \forall x, y \in \mathrm{dom}(f): \|\nabla f(x) - \nabla f(y)\|_* \leq L\|x-y\|$.

* Lemma 2.2. Let $f: \mathbb{R}^d \to \bar{\mathbb{R}}$ be an L-smooth function w.r.t. $\|\cdot\|$. Then, $\forall x, y \in \mathrm{dom}(f)$:

$$f(y) \boxed{\leq} f(x) + \langle \nabla f(x), y-x \rangle + \frac{L}{2}\|y-x\|^2.$$

Proof: From Part 1) of TT:

$$\begin{aligned}
f(y) - f(x) \underbrace{- \langle \nabla f(x), y-x \rangle} &= \int_0^1 \langle \nabla f(x+t(y-x)), y-x \rangle \, dt \\
&\quad - \int_0^1 \langle \nabla f(x), y-x \rangle \, dt \\
&= \int_0^1 \langle \nabla f(x+t(y-x)) - \nabla f(x), y-x \rangle \, dt \\
&\leq \int_0^1 \|\nabla f(x+t(y-x)) - \nabla f(x)\|_* \|y-x\| \, dt \\
&\leq \int_0^1 L t \|y-x\|^2 \, dt \\
&= \frac{L}{2}\|y-x\|^2.
\end{aligned}$$

* Ex. Prove that, under the same assumptions as in Lemma 2.2,

$$f(y) \geq f(x) + \langle \nabla f(x), y-x \rangle - \frac{L}{2}\|y-x\|^2.$$

* Lemma 2.3. Suppose that $f: \mathbb{R}^d \to \bar{\mathbb{R}}$ is twice cont.'ly diff.'able on dom(f). Then f is L-smooth w.r.t. $\|\cdot\|_2$ if and only if $-LI \boxed{\preccurlyeq} \nabla^2 f(x) \preccurlyeq LI$, $\forall x \in \text{dom}(f)$.

Löwner order

$A \succcurlyeq B$

$\Leftrightarrow A - B \succcurlyeq 0$

$\Leftrightarrow A - B$ is PSD.

$$\|A\|_2 = \sup_{x : \|x\|_2 \neq 0} \frac{\|Ax\|_2}{\|x\|_2}.$$

$$\Rightarrow \|Ax\|_2 \leq \|A\|_2 \|x\|_2.$$

Proof:

$\boxed{\Rightarrow}$ Suppose that f is L-smooth.

To show: $\nabla^2 f(x) \preccurlyeq LI$. ($-LI \preccurlyeq \nabla^2 f(x)$ left as an exercise)

Let $x, y \in \text{dom}(f)$, $y = \underbrace{x + \alpha p}_{\in \text{dom}(f)}$, $\alpha > 0$.

From Lemma 2.2:

$$(*) \quad f(x+\alpha p) \boxed{\leq} f(x) + \langle \nabla f(x), \alpha p \rangle + \frac{L}{2} \alpha^2 \|p\|_2^2.$$

From Part 4) of TT:

$$(**) \quad f(x+\alpha p) = f(x) + \langle \nabla f(x), \alpha p \rangle + \frac{\alpha^2}{2} p^T \nabla^2 f(x + \gamma \alpha p) p$$

for some $\gamma \in (0,1)$.

From $(*) + (**)$:

$$\cancel{\frac{\alpha^2}{2}} p^T \nabla^2 f(x + \underbrace{\gamma \alpha p}_{\to 0}) p \leq \cancel{\frac{L}{2}} \cancel{\alpha^2} \|p\|_2^2 \quad / \lim_{\alpha \downarrow 0}$$

$$\Rightarrow \quad p^T \nabla^2 f(x) p \leq L \|p\|_2^2$$

$\boxed{\Leftarrow}$ Suppose that $-LI \preceq \nabla^2 f(x) \preceq LI \Leftrightarrow \boxed{\|\nabla^2 f(x)\|_2 \leq L}$.

Want to show: $\forall x, y: \|\nabla f(x) - \nabla f(y)\|_2 \leq L \|x - y\|_2$.

From Part 3) of TT: $(\forall x, y \in \mathrm{dom}(f))$

$$\|\nabla f(y) - \nabla f(x)\|_2 = \left\| \int_0^1 \nabla^2 f(x + t(y-x))(y-x)\, dt \right\|_2$$

$$\overset{\text{Jensen}}{\leq} \int_0^1 \|\nabla^2 f(x + t(y-x))(y-x)\|_2 \, dt$$

$\lceil$ Ex. $\|\cdot\|_p$, $p \geq 1$, is convex $\rfloor$

$$\leq \int_0^1 \underbrace{\|\nabla^2 f(x + t(y-x))\|_2}_{\leq L} \|y - x\|_2 \, dt$$

$$\leq \int_0^1 L \|y - x\|_2 \, dt$$

$$= L \|y - x\|_2 .$$

* Characterizing minima of smooth functions:

Throughout this part: $\mathcal{X} = \mathbb{R}^d$

1) Necessary conditions:

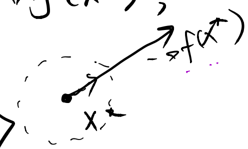
* Thm. 2.4

(a) Suppose that $f: \mathbb{R}^d \to \mathbb{R} \cup \{+\infty\}$ is cont.'ly diff.'able. If x^* is a local min of f, then $\nabla f(x^*) = 0$.

(b) Suppose that $f: \mathbb{R}^d \to \mathbb{R} \cup \{+\infty\}$ is twice cont.'ly diff.'able. Then, in addition to (a), $\nabla^2 f(x^*) \succeq 0$.

Proof:

(a) Suppose f.p.o.c. that $\nabla f(x^*) \neq 0$, but x^* is a local min. Apply Part 2) of TT w/ $y = x^* - \alpha \nabla f(x^*)$, $x = x^*$, $\alpha > 0$.



$$\rightarrow f(x^* - \alpha \nabla f(x^*)) = f(x^*) + \langle \nabla f(x^* - \gamma \alpha \nabla f(x^*)), -\alpha \nabla f(x^*) \rangle$$

for some $\gamma \in (0, 1)$.

Look at: $-\langle \nabla f(x^* - \gamma\alpha\nabla f(x^*)), \nabla f(x^*) \rangle$.

If it were the case that $\alpha = 0$, we would have:

$$- \langle \nabla f(x^* - \underbrace{\gamma\alpha\nabla f(x^*)}_{=0}), \nabla f(x^*) \rangle = - \|\nabla f(x^*)\|_2^2 .$$

Since ∇f is continuous, $\exists$ a sufficiently small $\alpha > 0$ s.t.

$$- \langle \nabla f(x^* - \gamma\alpha\nabla f(x^*)), \nabla f(x^*) \rangle \leq -\frac{1}{2}\|\nabla f(x^*)\|_2^2 .$$

$$\Rightarrow f(x^* - \alpha\nabla f(x^*)) \leq f(x^*) - \frac{\alpha}{2} \underbrace{\|\nabla f(x^*)\|_2^2}_{>0 \text{ by assumption}}$$

$$< f(x^*)$$

$\Rightarrow x^*$ cannot be a local min, a contradiction.

(b) Suppose f.p.o.c. that $\nabla^2 f(x^*)$ has negative evals.
Then, $\exists \lambda > 0$ s.t. $\nabla^2 f(x^*) \preceq -\lambda I$, and for all $v \in \mathbb{R}^d$, $\|v\|_2 = 1$, $v^T \nabla^2 f(x^*) v \leq -\lambda$.

Using Part 4) of TT w/ $x = x^*$, $y = x^* + \alpha v$, $\alpha > 0$.

$$f(x^* + \alpha v) = f(x^*) + \langle \underbrace{\nabla f(x^*)}_{0}, \alpha v \rangle + \frac{\alpha^2}{2} v^T \nabla^2 f(x^* + \gamma\alpha v) v ,$$

for some $\gamma \in (0,1)$.

If we had $\alpha = 0$, then:

$$v^T \nabla^2 f(x^* + \gamma\alpha v) v \leq -\lambda \|\overset{1}{v}\|_2^2$$

As $\nabla^2 f$ is continuous, $\exists$ suff. small $\alpha > 0$ s.t.

$$v^T \nabla^2 f(x^* + \gamma\alpha v) v \leq -\frac{\lambda}{2} .$$

$$f(x^* + \alpha v) \leq f(x^*) - \frac{\alpha^2 \lambda}{4} < f(x^*),$$ a contradiction. $\blacksquare$

* An alternative proof:

a) $\phi(\alpha) = f(x^* - \alpha \nabla f(x^*))$.

Fermat's thm: if x^* is a local min $\Rightarrow \phi'(\alpha) = 0$.

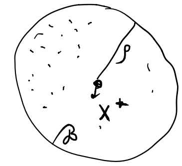
$$\phi'(\alpha) = \langle \nabla f(x^* - \alpha \nabla f(x^*)), -\nabla f(x^*) \rangle .$$

$$\phi'(0) = - \|\nabla f(x^*)\|_2^2 \Rightarrow \nabla f(x^*) = 0 .$$

b) $\phi_v(\alpha) = f(x^* + \alpha v)$; use 2nd derivative test $+\ \phi'(0) = 0$.

2) Sufficient conditions:

* Thm. 2.5. Let $f : \mathbb{R}^d \to \mathbb{R} \cup \{+\infty\}$ be twice cont.'ly diff.'able and assume that for some $x^* \in \mathrm{dom}(f)$ $\nabla f(x^*) = 0$ and $\nabla^2 f(x^*) \succ 0$. Then x^* is a strict local min of f.



* Proof: Let $\mathcal{B}$ be a ball centered at x^* and of radius ρ that is suff. small so that $\nabla^2 f(x^* + p) \succeq \varepsilon I$ for some $\varepsilon > 0$ and all p s.t. $\|p\|_2 \leq \rho$ (such a ball must exist b/c $\nabla^2 f(x^*) \succ 0$ and $\nabla^2 f$ is continuous).

Apply Part 4) of TT w/ $x = x^*$; $y = x^* + p$:

$$f(x^* + p) = f(x^*) + \langle \underbrace{\cancel{\nabla f(x^*)}}_{0}, p \rangle + \frac{1}{2} p^T \nabla^2 f(x^* + \gamma p) p,$$

for some $\gamma \in (0, 1)$.

$$f(x^* + p) \geq f(x^*) + \frac{\overset{>0}{\varepsilon}}{2} \|p\|_2^2 > f(x^*) \text{ if } \|p\|_2 \neq 0 .$$

$\Rightarrow x^*$ is a strict local min. ∎