Problem: $\min_{x \in \mathcal{X}} f(x)$. $\mathcal{X} \subseteq \mathrm{dom}(f) \subseteq \mathbb{R}^n$. a closed set.

1 A Taxonomy of Solutions to (P)

Will use "solution" and "minimizer" interchangeably.

Definition 1. We say that $x^* \in \mathcal{X} \subseteq \mathrm{dom}(f)$ is

1. a *local minimizer/solution* of (P) if there exists a neighborhood $\mathcal{N}_{x^*}$ of x^* such that for all $x \in \mathcal{N}_{x^*} \cap \mathcal{X}$ we have $f(x) \geq f(x^*)$;
2. a *global minimizer* of (P) if $\forall x \in \mathcal{X}$: $f(x) \geq f(x^*)$
3. a *strict local minimizer* of (P) if there exists a neighborhood $\mathcal{N}_{x^*}$ of x^* such that for all $x \in \mathcal{N}_{x^*} \cap \mathcal{X}$ and $x \neq x^*$ we have $f(x) > f(x^*)$; (i.e., satisfies part 1 with a strict inequality)
4. an *isolated local minimizer* of (P) if there exists a neighborhood $\mathcal{N}_{x^*}$ such that $\forall x \in \mathcal{N}_{x^*} \cap \mathcal{X}$: $f(x) \geq f(x^*)$ and $\mathcal{N}_{x^*}$ does not contain any other local minimizer.
5. a *unique minimizer* if it is the only global minimizer.

Ex. prove isolated local min is strict.

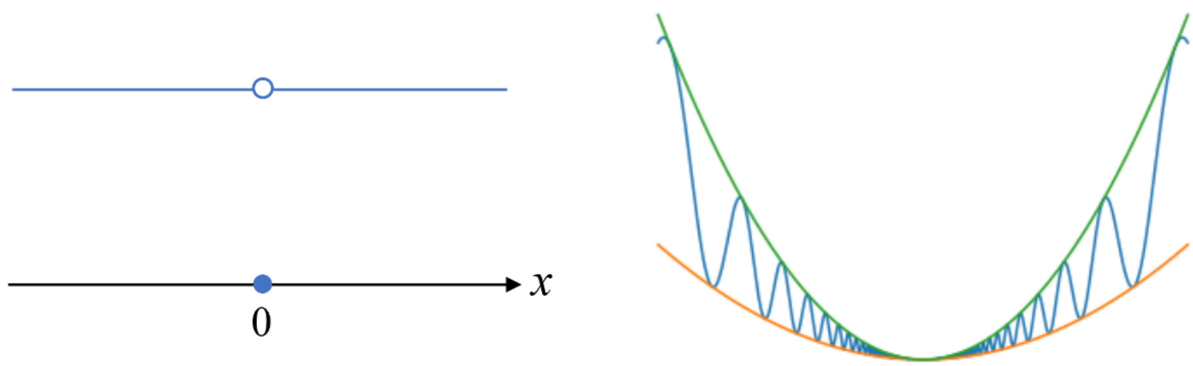
Pf: Suppose for contra, strict inequality does not hold. Then $\exists\, x^{*\prime} \neq x^*$ s.t. $f(x^{*\prime}) = f(x^*)$. And $x^{*\prime} \in N_{x^*} \cap \mathcal{X}$. This contradicts with x^* is an isolated min.

Strict min $\not\Rightarrow$ isolated.

Example 3. A strict minimizer that is not isolated:

- (not continuous) $f_1(x) = \begin{cases} 1 & x \neq 0 \\ 0 & x = 0 \end{cases}$ and $x^* = 0$.
- (continuous) $f_2(x) = \begin{cases} x^2\left(1 + \sin^2(\frac{1}{x})\right) & x \neq 0 \\ 0 & x = 0 \end{cases}$ and $x^* = 0$.

Illustration: Left f_1. Right: f_2.



Determine a pt is local / global min $\Rightarrow$ Taylor's Thm.

Theorem 1 (Taylor's Theorem; Thm 2.1 in Wright-Recht). *Let $f : \mathbb{R}^d \to \bar{\mathbb{R}}$ be a continuously differentiable function. Then, for all $x, y \in \mathrm{dom}(f)$ such that $\{(1-\alpha)x + \alpha y : \alpha \in (0,1)\} \subseteq \mathrm{dom}(f)$, we have*

1. $f(y) = f(x) + \int_0^1 \langle \nabla f(x + t(y-x)), y - x \rangle \, \mathrm{d}t$
2. $f(y) = f(x) + \langle \nabla f(x + \gamma(y-x)), y - x \rangle$ *for some $\gamma \in (0,1)$ (a.k.a. Mean Value Thm).*

If f is twice continuously differentiable:

3. $\nabla f(y) = \nabla f(x) + \int_0^1 \nabla^2 f(x + t(y-x))(y-x)\mathrm{d}t$. *Here*

$$\nabla^2 f(x) = \begin{bmatrix} & \cdots & \\ \vdots & \frac{\partial^2 f}{\partial x_i \partial x_j}(x) & \vdots \\ & \cdots & \end{bmatrix} \in \mathbb{R}^{d\times d}$$

denotes the Hessian matrix ("second-order derivative") of f at x.

MVT.

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

$$\begin{aligned} f(y) &= f(x) + \langle \nabla f(x), y - x \rangle + \frac{1}{2}\langle \nabla^2 f(x + \gamma(y-x))(y-x), y-x \rangle \\ &= f(x) + \langle \nabla f(x), y - x \rangle + \frac{1}{2}(y-x)^\top \nabla^2 f(x + \gamma(y-x))(y-x). \end{aligned}$$

Remark 1. A common mistake is to write down the following "Mean-Value Thm" for the gradient:

$$\exists \gamma \in (0,1) : \nabla f(y) = \nabla f(x) + \nabla^2 f(x + \gamma(y-x))(y-x)? \longleftarrow \textbf{This is wrong!}$$

4. Illustrate:

Start from 2nd Taylor expansion:

$$f(y) = f(x) + \langle \nabla f(x), y-x \rangle + \int_0^1 (1-t) \langle \nabla^2 f(x+t(y-x))(y-x), y-x \rangle \, dt$$

Integral MVT: $\int_a^b f(x) g(x) \, dx = f(c) \int_a^b g(x) \, dx$ $\quad (g(x) = 1 \Rightarrow$ std MVT$)$

$$\exists \nu \in (0,1) \quad = f(x) + \langle \nabla f(x), y-x \rangle + \langle \nabla^2 f(x+\nu(y-x))(y-x), y-x \rangle \underbrace{\int_0^1 (1-t) \, dt}_{= \frac{1}{2}}$$

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

For f continuously differentiable at x, we have

$$f(x+p) = f(x) + \nabla f(x)^T p + o(\|p\|)$$

or equivalently,

$$\lim_{\|p\|\to 0} \frac{|f(x+p)-f(x)-\nabla f(x)^T p|}{\|p\|} = 0.$$

Pf: By part 1 of Taylor Thm above,

$$f(x+p) \overset{\exists \gamma\in(0,1)}{=\!=} f(x) + \langle \nabla f(x+\gamma p), p\rangle$$

$$= f(x) + \nabla f(x)^T p + (\nabla f(x+\gamma p) - \nabla f(x))^T p$$

$$\le f(x) + \nabla f(x)^T p + \|\nabla f(x+\gamma p) - \nabla f(x)\| \, \|p\|$$

As $\|p\|\to 0$ $\nabla f(x+\gamma p) \to \nabla f(x)$. By sandwich we are done.

$\square$.