UNCONSTRAINED OPTIMIZATION

An unconstrained optimization problem has the form

$$\min_{x \in \mathbb{R}^d} \quad f(x)$$

where $f: \mathbb{R}^d \to \mathbb{R}$ is called the objective function.

A point $x^* \in \mathbb{R}^d$ is called a local minimizer if $\exists r > 0$ such that $f(x^*) \le f(x)$ $\forall x$ satisfying $\|x - x^*\| < r$. x^* is called a global minimizer if $f(x^*) \le f(x)$ $\forall x \in \mathbb{R}^d$.

First and Second Order Necessary Conditions

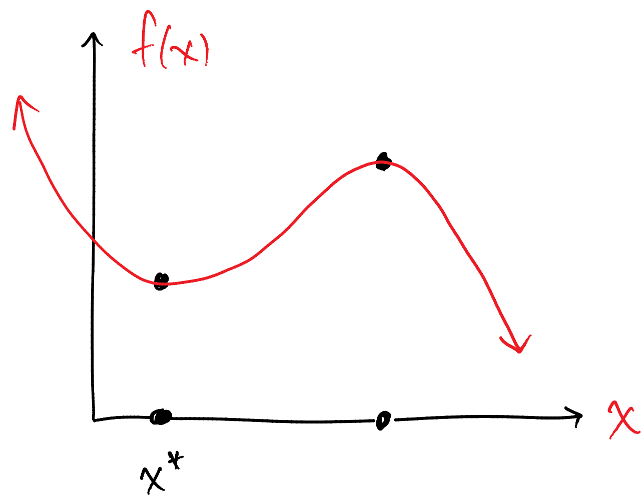
Given a function $f: \mathbb{R}^d \to \mathbb{R}$, the gradient and Hessian of f at $x = [x_1 \cdots x_d]^T \in \mathbb{R}^d$ are defined by

$$\nabla f(x) = \begin{bmatrix} \frac{\partial f(x)}{\partial x_1} \\ \vdots \\ \frac{\partial f(x)}{\partial x_d} \end{bmatrix} \quad \text{and} \quad \nabla^2 f(x) = \begin{bmatrix} \frac{\partial^2 f(x)}{\partial x_1^2} & \cdots & \frac{\partial^2 f(x)}{\partial x_1 \partial x_d} \\ \vdots & \ddots & \\ \frac{\partial^2 f(x)}{\partial x_d \partial x_1} & \cdots & \frac{\partial^2 f(x)}{\partial x_d^2} \end{bmatrix}.$$

We say f is differentiable if $\nabla f(x)$ exists $\forall x \in \mathbb{R}^d$, and twice differentiable if $\nabla^2 f(x)$ exists $\forall x \in \mathbb{R}^d$. We say f is twice continuously differentiable if it is twice differentiable and all of the second derivatives are continuous. If f is twice continuously differentiable, then $\nabla^2 f(x)$ is symmetric $\forall x$, i.e.,

$$\frac{\partial f(x)}{\partial x_i \partial x_j} = \frac{\partial f(x)}{\partial x_j \partial x_i} \qquad \forall x \in \mathbb{R}^d, \ \forall i,j = 1, \dots, d$$

Property 1 If f is differentiable and x^* is a local minimizer of f, then $\nabla f(x^*) = 0$.



$\nabla f(x^*) = 0$ is necessary but not sufficient

Proof Define the scalar valued function $\varphi(t) = f(x^* + yt)$, where $y \in \mathbb{R}^d$ is arbitrary. Then

$$\varphi'(0) = \lim_{t \searrow 0} \frac{f(x^* + yt) - f(x^*)}{t}$$

[def of directional derivative]

$$= \langle \nabla f(x^*), y \rangle \qquad \text{[chain rule]}$$

Since x^* is a local min, we know

$$f(x^* + yt) \geq f(x^*)$$

for t sufficiently small. Therefore $\langle \nabla f(x^*), y \rangle \geq 0$.

Now choose $y = -\nabla f(x^*)$. Then

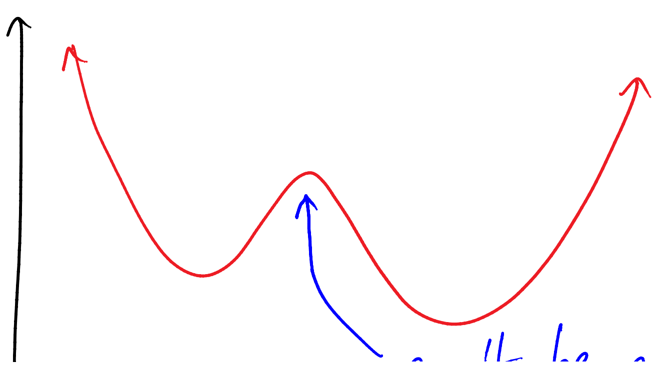
$$0 \leq \langle \nabla f(x^*), -\nabla f(x^*) \rangle = -\|\nabla f(x)\|^2 \leq 0,$$

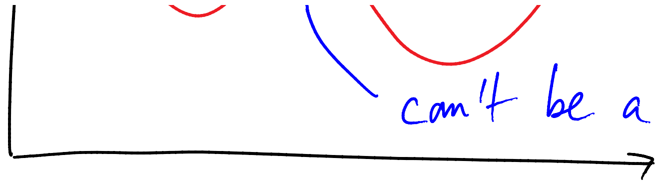
so we must have $\nabla f(x^*) = 0$. ▨

Property 2 | If f is twice continuously differentiable and x^* is a local min, then $\nabla^2 f(x)$ is positive semi-definite, i.e., $z^T \nabla^2 f(x^*) z \geq 0 \quad \forall z \in \mathbb{R}^d$.

Proof | Homework problem

This result generalizes the result from single-variable calculus that the second derivative is nonnegative at a local min.





can't be a local min!

Exercise Give an example of a function f such that $\nabla^2 f(x^*)$ is positive semi-definite at some x^*, but x^* is not a local minimizer.

Convexity

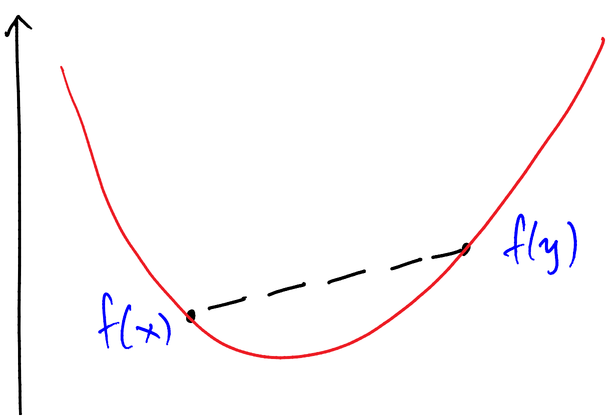
We say f is convex if

$$f(tx + (1-t)y) \le t f(x) + (1-t) f(y)$$

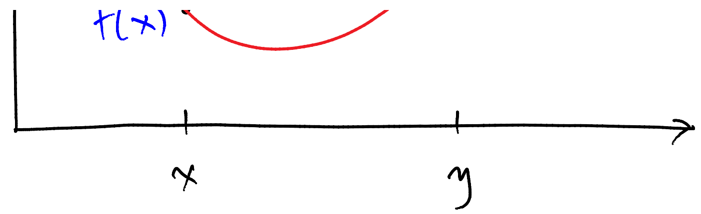
$\forall x, y \in \mathbb{R}^d$ and $t \in [0,1]$. We say f is strictly convex if

$$f(tx + (1-t)y) < t f(x) + (1-t) f(y)$$

$\forall x \neq y$ and $t \in (0,1)$.



line segment is above graph



Exercise | Give an example of a function that is convex but not strictly convex.

If f is convex, the problem of minimizing f becomes easier to understand. Let's look at some basic properties.

Property 3 | If f is convex, then every local min is a global min.

Proof | Suppose x^* is a local min but not a global min. Then $\exists\, y^* \in \mathbb{R}^d$ s.t. $f(y^*) < f(x^*)$. By convexity, $\forall\, t \in [0,1)$ we have

$$\begin{aligned} f(tx^* + (1-t)y^*) &\leq t f(x^*) + (1-t) f(y^*) \\ &< t f(x^*) + (1-t) f(x^*) \\ &= f(x^*). \end{aligned}$$

Taking $t \nearrow 1$, the above strict inequality contradicts local minimality of x^*. Thus x^* is a global min.

local minimality of x. Thus x is a global min.

Property 4 | If f is strictly convex, then f has at most one global min.

Proof | Homework

Exercise | Give an example of f that is

- convex and has more than one global min
- strictly convex and has no global min

The following is a first-order characterization of convexity.

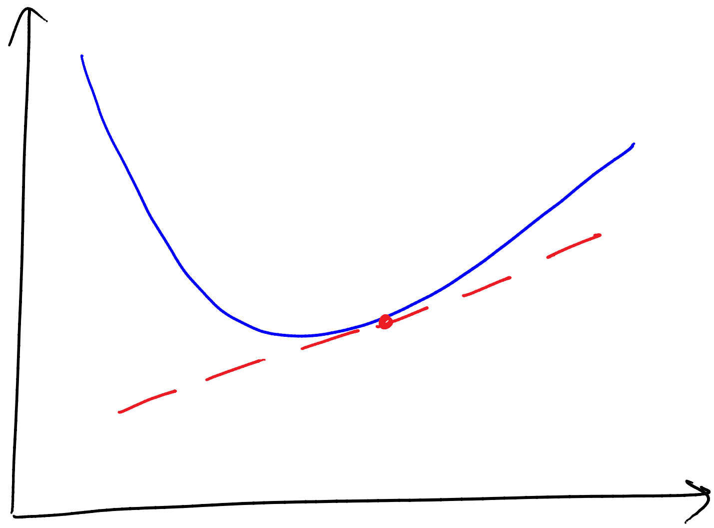
Property 5 | Suppose $f : \mathbb{R}^d \longrightarrow \mathbb{R}$ is differentiable.

Then f is convex iff $\forall x, y \in \mathbb{R}^d$,

$$f(y) \geq f(x) + \langle \nabla f(x), y - x \rangle.$$

Similarly, f is strictly convex iff $\forall x \neq y$,

$$f(y) > f(x) + \langle \nabla f(x), y - x \rangle.$$



the graph is above the tangent line

Proof First, assume x is convex. For any $x, y \in \mathbb{R}^d$, $t \in [0,1]$,

$$f(ty + (1-t)x) \le t f(y) + (1-t) f(x)$$
$$= f(x) + t\big(f(y) - f(x)\big).$$

Rearranging,

$$\frac{f(x + t(y-x)) - f(x)}{t} \le f(y) - f(x).$$

The limit of the LHS is a directional derivative and equal to $\langle \nabla f(x), y-x \rangle$ by the chain rule.

Therefore $f(y) \ge f(x) + \langle \nabla f(x), y - x \rangle$.

Now suppose conversely that $\forall x, y$

Now suppose conversely that $\forall x, y$

$$f(y) \geq f(x) + \langle \nabla f(x), y - x \rangle. \quad (\star)$$

Let $x, y \in \mathbb{R}^d$ and $t \in [0,1]$. Denote $z = tx + (1-t)y$.

Applying $(\star)$ twice, we have

$$f(x) \geq f(z) + \langle \nabla f(z), x - z \rangle \quad (\star a)$$

$$f(y) \geq f(z) + \langle \nabla f(z), y - z \rangle. \quad (\star b)$$

Now multiply $(\star a)$ by t, $(\star b)$ by $1-t$, and add:

$$t f(x) + (1-t) f(y) \geq f(z) + \langle \nabla f(z), \underbrace{tx + (1-t)y - z}_{=0} \rangle$$

$$= f(tx + (1-t)y).$$

This establishes convexity.

The proof of the second statement (strict convexity) is similar. $\blacksquare$

For convex and differentiable f, the first order necessary condition is also sufficient.

Property 6 Let f be convex and continuously differentiable.

Then x^* is a global min iff $\nabla f(x^*) = 0$.

<u>Proof</u> The forward implication follows from Property 1.

The reverse implication follows immediately from Property 5. ∎

There is also a second-order characterization of convexity.

<u>Property 7</u> Let f be twice continuously differentiable. Then

(a) f is convex $\iff$ $\nabla^2 f(x)$ is positive semi-definite $\forall x \in \mathbb{R}^d$

(b) f is strictly convex $\Longleftarrow$ $\nabla^2 f(x)$ is positive definite $\forall x \in \mathbb{R}^d$

<u>Proof</u> Part (a) is a homework problem. Part (b) follows similarly. ∎

<u>Exercise</u> Give an example of an f that is strictly convex but for which $\nabla^2 f(x)$ is not positive-definite $\forall x$.