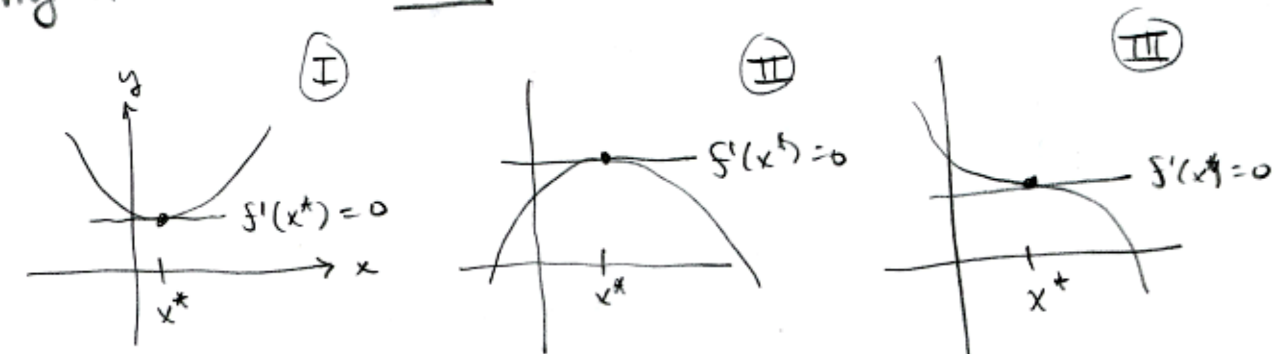
Optimality and convexity

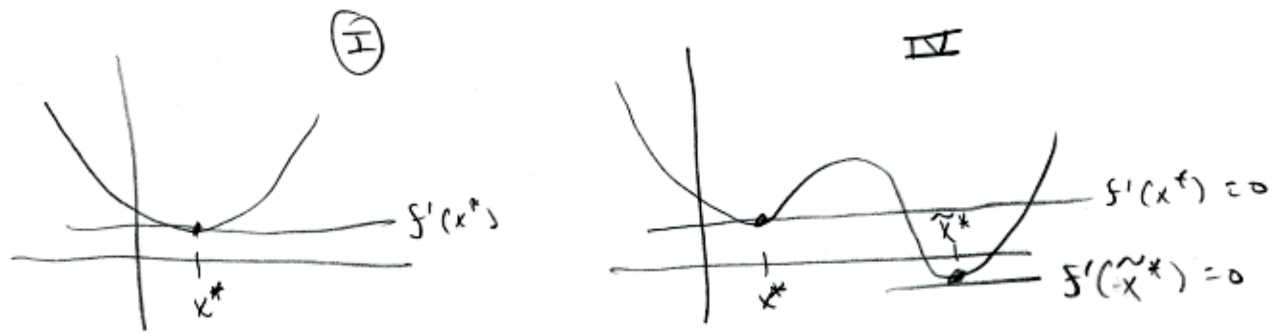
We've seen a few methods of solving minimization problems by this point. Let's be more precise about what we're trying to do. Previously, we just tried to find a point x^* s.t. $\nabla f(x^*) = 0$. This is called a <u>stationary point</u>. For x^* to be an optimum, this is a <u>necessary</u> but not <u>sufficient</u> condition. For this reason, we call this condition a <u>first-order necessary condition</u>. Remember from calculus why it is not <u>sufficient</u>:

Eg:



Among these cases, only case I is a minimizer. Further:

Eg:



So, it is clearly a local property — we also need to decide when a minimizer is the "true" minimizer. We say x^* is a <u>local minimum</u> if there exists some $\varepsilon = \varepsilon(x^*) > 0$ such that for all x s.t. $\|x - x^*\| < \varepsilon$ it holds that $f(x^*) \leq f(x)$. Similarly, we define a <u>strict local minimum</u> as being a point x^* s.t. there exists $\varepsilon = \varepsilon(x^*) > 0$ s.t. for all x s.t.

if $0 < \|x - x^*\| < \varepsilon$ then $f(x^*) < f(x)$. What is the difference between these two definitions? 2

E.g. "non-strict local minimum":

A global minimizer is x^* such that:

$$f(x^*) \leq f(x) \quad \text{for all } x \in \text{dom} f.$$

Likewise, strict global minimum:

$$f(x^*) < f(x) \quad \text{for all } x \in \text{dom} f.$$

Without knowing anything about f, even if we knew the location of a global minimum x^*, it would be nearly impossible to verify that it is a global minimum! For this reason, we will stick to algorithms for finding local optima; we will also look at properties of functions which ensure the existence of global optima.

How do we tell if a stationary point x^* is a local minimum? Is there an easier property to check? Since $\nabla f(x^*) = 0$, we can Taylor expand about x^* — letting $x = x^* + p$ — to get:

$$f(x) = f(x^* + p) = f(x^*) + \underbrace{\nabla f(x^*)^T p}_{=0 \text{ by F.O.N.C.}} + \frac{1}{2} p^T \nabla^2 f(x^*) p + O(\|p\|^3)$$

$$= f(x^*) + \frac{1}{2} p^T \nabla^2 f(x^*) p + O(\|p\|^3).$$

So, by our Taylor expansion, it is clear that for $\|p\|$ small enough, if $p^T \nabla^2 f(x^*) p \geq 0$, since then:

$$f(x) = f(x^*) + \frac{1}{2} p^T \nabla^2 f(x^*) p \geq f(x^*),$$

implying that x^* is a local minimum. The condition "$p^T \nabla^2 f(x^*) p \geq 0 \;\; \forall p \neq 0$" is the condition for $\nabla^2 f(x^*)$ to be positive semidefinite. In general, a matrix $A \in \mathbb{R}^{n \times n}$ is positive semidefinite if $x^T A x \geq 0$ for all $x \neq 0$. Similarly, A is positive definite if $x^T A x > 0$ for all $x \neq 0$. There are analogous definitions for negative definite and negative semidefinite matrices with "$<$" and "$\leq$" replacing "$>$" and "$\geq$", respectively. Using these definitions, we can define the second-order necessary conditions for optimality:

x^* is a ...	$\nabla f(x^*) = 0$ and $\nabla^2 f(x^*)$ is ...
local minimum	pos. semidef.
local maximum	neg. semidef.

Likewise, the second-order sufficient conditions are:

x^* is a ...	$\nabla f(x^*) = 0$ and $\nabla^2 f(x^*)$ is ...
strict local minimum	positive definite
strict local maximum	negative definite

How do we check the definiteness of a matrix?

If A is real and symmetric, as will be the case for all $\nabla^2 f$ in this class, we can look at the eigenvalues of A. Recall that λ is an eigenvalue of A if $Au = \lambda u$ for some $u \neq 0$. This u is called an eigenvector and may not be unique. Note that if A is symmetric, by the spectral theorem all λ will be real. We can characterize definiteness by:

A is...	if ...
pos def	$\lambda_i > 0 \ \forall i$
pos semidef	$\lambda_i \geq 0 \ \forall i$
neg def	$\lambda_i < 0 \ \forall i$
neg semidef	$\lambda_i \leq 0 \ \forall i$

To check whether a function has a single minimum, we can use convexity. Functions and sets can be convex. We will discuss convex functions here and return to sets later when we talk about constrained optimization. A function $f: \mathbb{R} \to \mathbb{R}$ is convex if:

$$f((1-\alpha)x + \alpha y) \leq (1-\alpha) f(x) + \alpha f(y)$$

$$\forall x, y, \text{ and } \forall \alpha \in [0,1].$$

and concave if:

$$f((1-\alpha)x + \alpha y) \geq (1-\alpha) f(x) + \alpha f(y)$$

$$\forall x, y \text{ and } \forall \alpha \in [0,1].$$

Strict convexity and strict concavity are defined by replacing "$\leq$" and "$\geq$" w/ "$<$" and "$>$", respectively.

Theorem: If f is convex (resp., strictly convex), and x^* is a local minimum (resp., strictly local minimum), then x^* is a global minimum (resp., unique global minimum).

Hard to tell in general whether a function is convex, but there is a "calculus of convex functions" which can be used (see "Convex Optimization" by Boyd).

We can relate convexity to our second-order conditions on the Hessian $\nabla^2 f$:

f is ...	... if and only if ...
convex	$\nabla^2 f(x)$ pos semidef $\forall x$
strictly convex	" pos def "
concave	" neg semidef "
strictly concave	" neg def ".