Unconstrained Optimization

↪ Ch 2 of NW.

Overview

→ Identify a local minima → Taylor's thm

1st order 2nd order

↓

overview of algos

Line search Trust region

(define descent directions)

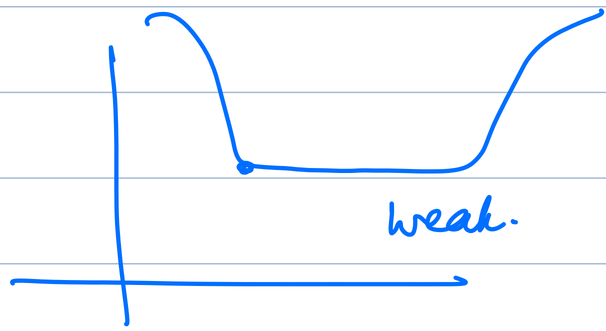
Identifying a local minima. → x^*

① Global minimizer, $f(x^*) \leq f(x) \quad \forall x \in \text{dom} f$.

② Local minimizer, $f(x^*) \leq f(x), \ \forall x \in \mathcal{N}$

weak ⟷ " $\leq$ "

strong ⟷ " $<$ "



↪ Convex fns: any local minimizer is also a global minimizer.

Taylor's Theorem

$f: \mathbb{R}^n \to \mathbb{R}$, $x, p \in \mathbb{R}^n$, $t \in (0,1)$

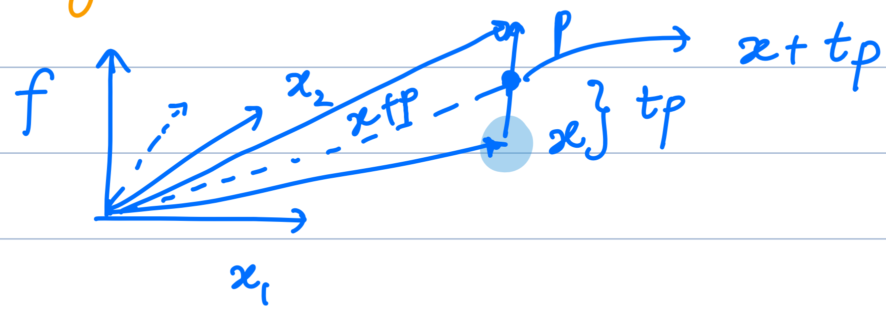
1) Continuously differentiable:

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

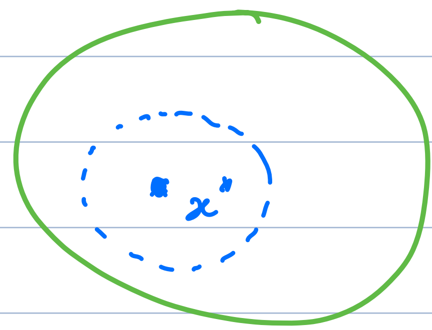
2) Twice continuously differentiable f:

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

Taylor series + remainder term.



" open neighbourhood of x^* "



1st order condition:

If x^* is a local minimizer & f is continuously differentiable in an open nbd of x^*, then $\nabla f(x^*) = 0$

[Necessary condn]

If $\nabla f(x^*) = 0$, then x^* is called a stationary point.

2nd order Condn

Require: $\nabla^2 f$ exists & is conts on an open nbd of x^*.

Necessary: If x^* is a minimizer then:

1) $\nabla f(x^*) = 0$

2) $\nabla^2 f(x^*)$ is pos. semdefinite

Sufficient: If both $\nabla f(x^*) = 0$ and $\nabla^2 f$ is pos. sem-def then x^* is a local minimizer.

Minimizers: weak & strong

$\updownarrow$ pos. semi def $\qquad$ $\updownarrow$ pos. def.

Proof of 1st order Condn

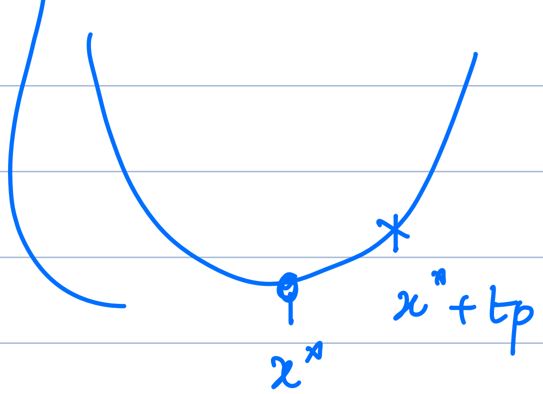
scalar fn.

one way: $g(t) = f(x^* + tp)$.

another way: Proof by contradiction.

$$f(x^*+p) = f(x^*) + \nabla f(x^* + tp)^T p$$

≥ 0



directional derivative

$D_p(f(x^* + tp))$

INTUITION

Assume x^* is a minimizer but $\boxed{\nabla f(x^*) \neq 0}$

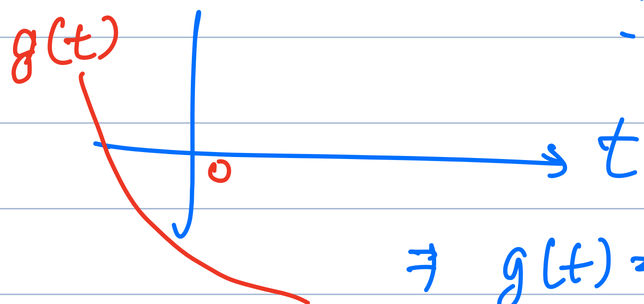
① Construct $g(t) = p^T \nabla f(x^* + tp)$. We can

② Choose any p. Choose $p = -\nabla f(x^*)$.

What $g(0) = -\|\nabla f(x^*)\|^2 < 0$.

③ $g(t)$ is a conts fn. $g(0) < 0$

$\therefore$ $g(t) < 0$ for some small $t < T$.



$\exists$ $g(t) = p^T \nabla f(x^* + tp) < 0$ for $0 < t < T$.

④ Go back to 1^{st} order Taylor thm:

$$f(x^* + Tp) = f(x^*) + \underbrace{\nabla f(x^* + tp)^T p}_{g(t) < 0} \quad \text{for } t < T.$$

$\Rightarrow f(x^* + Tp)$ is $< f(x^*)$

$\Rightarrow$ x^* is not a minimizer.

$\therefore$ Contradiction.

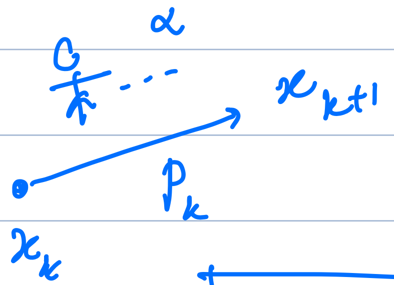
Overview of Algorithms.

start from $x_0 \longrightarrow \{x_k\} \longrightarrow$ Soln x^*

$\{x_k\} \downarrow x_{k+1}$

$x_{k+1} \swarrow$ Line Search

$x_{k+1} \searrow$ Trust region.



$$\boxed{\{\alpha^*, p_k^*\} = \underset{\alpha,\, p_k}{\arg\min}\ f(x_k + \alpha p_k)}$$

leading to

$$\underline{x_{k+1} = x_k + \alpha^* p_k^*}$$

4 families of methods for determining p_k:

① Steepest descent method:

p_k is chosen $\propto$ $\underline{-\nabla f(x_k)}$.

② Newton methods → Second order info is used. $H_k = \nabla^2 f_k$ if it pos. def. and

$$p_k = - \underbrace{(\nabla^2 f_k)^{-1}}\underbrace{\nabla f_k}.$$

③ Quasi-Newton methods → use approx for H_k.
Calc H_{k+1} from H_k.

④ Conjugate Gradient.