1. Unconstrained Optimization

1.1. Taylor's Theorem

Theorem 2.1 (N&W)

1.1.1

A. Suppose $f: \mathbb{R}^n \to \mathbb{R}$ is continuously differentiable and $p \in \mathbb{R}^n$. Then we have

$$f(x+p) = f(x) + \int_0^1 \nabla f(x+\tau p)^T p \, d\tau \overset{MVT}{=}$$

$$f(x) + \nabla f(x+tp) p \quad \text{for some } t \in (0,1)$$

B. Moreover, if f is twice continuously differentiable, we have that

$$\nabla f(x+p) = \nabla f(x) + \int_0^1 \nabla^2 f(x+\tau p) p \, d\tau$$

It works by Leibniz-Newton but MVT does not apply!

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

for some $t \in (0,1)$

Mihai Anitescu Notes

1.1.2 Again by Leibniz-Newton and $\int$ by parts on the scalar function $f(x+tp)$.

Mihai Anitescu Notes

1. UNCONSTRAINED OPTIMIZATION

1.2 Necessary Conditions for optimality:

Theorem 2.3

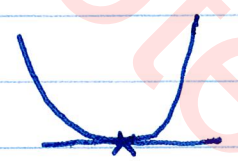
If x^* is a local minimizer of $f \in C^2$ in a neighborhood of x^*, then $\nabla f(x^*) = 0$ and $\nabla^2 f(x^*) \succeq 0$ is PSD

~~Sufficient Conditions~~

The conditions: $\begin{cases} \nabla f(x^*) = 0 \\ \nabla^2_{xx} f(x^*) \succeq 0 \end{cases}$

are called the necessary first and second-order optimality conditions.

"Proof"



Euler's principle;

Mihai Antitescu Notes

1. UNCONSTRAINED OPTIMISATION

1.3. ~~Necessary conditions~~

1.3.1 Sufficient conditions → Sometimes

~~Theorem 2.4~~

called second-order sufficient conditions even if there are no first-order sufficient conditions.: $\begin{cases} \nabla f(x^*) = 0 \\ \nabla^2 f(x^*) \succ 0 \end{cases}$ (SOSC)

<u>Theorem 2.4 (NfW)</u>

Suppose ~~$f \in C^2$~~ $\nabla^2_{xx} f$ is continuous in $\mathcal{N}(x^*)$; that $\nabla f(x^*) = 0$ and $\nabla^2_{xx} f(x^*)$ is positive definite. Then x^* is a local minimizer.

<u>Proof</u> Use Taylor Theorem C

Since $\nabla^2 f(x)$ is cts, $\exists$ r ~~such that~~ and $\varepsilon > 0$ such that

Mihai Anitescu Notes

1.3.2

$\nabla^2 f(x) \geq \varepsilon I \quad \forall\, x \in B(x^*, r).$

Define $p = x - x^*$.

By Taylor c, we have

$$f(x) = f(x^* + p) = f(x^*) + \underbrace{\nabla f(x^*)}_{=0} p + \frac{1}{2} p^T \nabla^2 f(x^* + tp) p$$

$$x^* + tp = x^* + t(x - x^*) = (1-t)x^* + tx \in B(x^*, r)$$

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

$$f(x^*) + \frac{1}{2} \varepsilon \cdot \|p\|^2 =$$

$$f(x^*) + \frac{1}{2} \varepsilon \cdot \|x - x^*\|^2$$

↳ "Quadratic Growth Condition"

and $\Rightarrow f(x) > f(x^*) \quad \forall\, x \in B(x^*, r)$

$x \neq x^*$. QED ∎

Mihai Anitescu Notes

1.3.3 It is strict..., but is it isolated?

Theorem 2.4 A.

SOSC $\Rightarrow$ x^* is ILM.

Proof

Use Taylor Theorem B. Continue Proof

$$\nabla f(x) = \nabla f(x^*) + \int_0^1 \nabla^2 f(x+\tau p)\, p \, d\tau$$

$$\Rightarrow p^T(\nabla f(x) - \nabla f(x^*)) = \int_0^1 p^T \nabla^2 f(x+\tau p)\, p \, d\tau$$

$$\geq \varepsilon \cdot \|p\|^2$$

$$\Rightarrow (x-x^*)^T(\nabla f(x) - \nabla f(x^*)) \geq \varepsilon \|x-x^*\|^2$$

$\Rightarrow \nabla f(x)$ is "Monotone"

But $\nabla f(x^*) = 0 \overset{x \neq x^*}{\Rightarrow} \nabla f(x) \neq 0$ so

x cannot be a LM $\Rightarrow$ x^* is ILM. $\blacksquare$

Mihai Anitescu Notes

1. UNCONSTRAINED OPTIMIZATION

1.4. ~~Ex~~ How large is gap btwn. Nec Cond

1.4.1. and Suff Conditions?

Note: SOSC $\Rightarrow$ LM is both ILM and SLM so no gap B, C, D...

A. Say x^* is LM and satisfies NC

$\Rightarrow \tilde{f}(x) = f(x) + \varepsilon (x-x^*)^T (x-x^*)$ satisfies SC at x^* $\Rightarrow$ so it seems that any problem can be perturbed to have only LM that satisfy SOSC.

B. On the other hand; if (P) satisfies SOSC at any LM, so will any suff small perturbation.

Mihai Anitescu Notes

1.4.2

c) So if $\{f\} \subset (X, \mathcal{T})$

$\Rightarrow \{\{f\} \mid$ any LM has SOSC$\}$

~~seem~~ is open and dense.

Problem. w

$\Rightarrow$ So LM that are "degenerate" have measure 0.

This statement can be made precise using transversality arguments, but it indicates gap is negligible in this model.

So we will assume SOSC from here on at x^*

$\boxed{\text{J-J-T, Theorem 2.1}}$

<u>Exact Statement</u>: The set $\mathcal{J} = \{f \in C^3(\mathbb{R}^n, \mathbb{R}) \mid$ any NC point is not degenerate$\}$ is open and dense in the Whitney C^3_s topology

Mihai Anitescu Notes

1.5. Newton's method for optimization

1.5.1 ~~Newto~~ and nonlinear equation

Idea: Represent the optimality conditions as a nonlinear equation:

$$\nabla f(x) = 0 = F(x)$$

How do I solve it? Iteratively.

$$x^k \longrightarrow m_{F,x^k}(x) = F(x^k) + \nabla F(x^k)\cdot(x - x^k)$$

$$: \mathbb{R}^n \to \mathbb{R}^n$$

Therefore $\boxed{m_{F,x^k}(x^{k+1}) = 0}$

Defines an iterative process

$$F(x^k) + \nabla F(x^k)(x^{k+1} - x^k) = 0$$

$$\Rightarrow x^{k+1} = x^k - (\nabla F(x^k))^{-1} \cdot F(x^k)$$

Digression: **Notation** convention:

Sequence index: superscript. Component: subscript

Mihai Anitescu Notes

1.6. Rate of convergence

Digression: $G(x): \mathbb{R}^n \to \mathbb{R}^m$ is LC on $W(x^*)$ if
$$\|G(x^1) - G(x^2)\| \le L_{G,W} \|x^1 - x^2\| \quad \forall\, x^1, x^2 \in W$$

1.6.1 Suppose that:

1. $\nabla F(x)$ is Lipschitz continuous in
2. a neighborhood of x^*; s.t. $F(x^*) = 0$.
2. $\nabla F(x^*)$ is invertible

Then

α) If x^0 is sufficiently close to x^* then $x^k \to x^*$

β) Rate of convergence is quadratic

γ) $\|\nabla f(x^k)\| \to$ quadratically at 0.

Digression. Rate of convergence

$\{z^k\}_{k \ge 0}$ 1. $0 \le z^{k+1} \le c z^k \qquad c < 1$

(Q) Linearly convergent at 0 w.c

3. $0 \le z^{k+1} \le w^k z^k$, $w^k \to 0$ superlinear

2. $0 \le z^{k+1} \le c\{z^k\}^2$; $c z^0 < 1$

$\Rightarrow z^k$ is quadr convergent

Mihai Anitescu Notes

1.5.2 a)

$$x^{k+1} - x^* = x^k - x^* - \left[\nabla F(x^k)\right]^{-1} \cdot F(x^k)$$

$$= \underbrace{\left[\nabla F(x^k)\right]^{-1}}_{3} \cdot \Big(\underbrace{\nabla F(x^k) \cdot (x^k - x^*)}_{2} - \underbrace{\left(F(x^k) - F(x^*)\right)}_{1} \Big)$$

But

1. $$F(x^k) - F(x^*) = \int_0^1 \nabla F(x^k + t(x^* - x^k)) \cdot (x^k - x^*)\, dt$$

2. $$\nabla F(x^k) \cdot (x^k - x^*) = \int_0^1 \nabla F(x^k) \cdot (x^k - x^*)$$

$$\Rightarrow x^{k+1} - x^* = \int_0^1 \left[\nabla F(x^k) - \nabla F(x^k + t(x^* - x^k))\right] (x^k - x^*)\, dt$$

$$\|x^{k+1} - x^*\| \le \int_0^1 \|\nabla F(x^k) - \nabla F(x^k + t(x^* - x^k))\| \, \|x^k - x^*\|\, dt$$

$$\le \left(\int_0^1 Lt\right) \cdot \|x^k - x^*\|^2 = \frac{1}{2} L \|x^k - x^*\|^2$$

3. $B(x^*, r)$; s.t $\|\nabla F(x^k)^{-1}\| \le 2 \cdot \|\nabla F(x^*)^{-1}\|$

Mihai Anitescu Notes

1.6.3. $\Rightarrow \|x^{k+1} - x^*\| \leq L \cdot \|\nabla F(x^*)\| \cdot \|x^k - x^*\|^2$

So

$x^k \in B(x^*, r)$ s.t. $\begin{cases} L \text{ for all of } \nabla F(x) \\ \nabla f(x)^{-1} \leq 2 \nabla f(x) \end{cases}$

the above holds

Then

1) x^0 is such that $\boxed{\text{induction}}$

$\|x^0\| < r$; $\|x^0 - x^*\| \leq \dfrac{1}{2L\|\nabla F(x^*)\|}$ $\Rightarrow$ $\|x^k\| < r$, $\|x^k - x^*\| \leq \dfrac{1}{2L\|\nabla F(x^*)\|}$

and

$$\|x^{k+1} - x^*\| < \frac{1}{2}\|x^k - x^*\|$$

So α) $\|x^k - x^*\| \to 0$; and

β) quad convergent

QED

Mihai Anitescu Notes

1.6.4

How does this apply to optimization?

$\nabla f(x^*) = 0$; $\nabla f(x^*)$ is PD

$\Rightarrow F(x) = \nabla f(x)$ satisfies conditions.

$\boxed{\text{Goal 2: Have my iterative algorithms achieve quadratic convergence}}$

(or superlinear convergence) since linear convergence can be very slow (see steepest descent).

Mihai Anitescu Notes

1.7.

What is the problem with Newton's Method?

We have proved we achieve goal 2, but not a word about goal 1 !! So if x^0 is not in the right neighborhood of x^* this iteration may truly go anywhere!! Divergence.

Example of nonconvergence.

How do I achive both goals? Answer globalization (global convergence to a stationary point). But make sure I do not destroy second goal!!

Mihai Anitescu Notes