Lecture 2

Thursday, April 2, 2026 7:47 AM

Review:

→ x^* is a local min of f: $\exists \varepsilon > 0$: $f(x^*) \leq f(x)$ $\forall x: |x - x^*| < \varepsilon$

→ Technique: $g(\alpha) \triangleq f(x + \alpha d) \rightsquigarrow g(\alpha) = g(0) + g'(0)\alpha + \frac{1}{2} g''(0)\alpha^2 + O(\alpha^3)$

→ $x^* \in D$ is an interior point:

- Necessary conditions of optimality:

  1-st order : $\nabla f(x^*) = 0$ (if f is C^1)

  2nd order : $\nabla^2 f(x^*) \succcurlyeq 0$ P.S.D. (if f is C^2)

- sufficient conditions of optimality:

  $\nabla f(x^*) = 0$, $\nabla^2 f(x^*) \succ 0$ P.D. (if f is C^2)

→ $x^* \in \partial D$ on the boundary

- Nec. cond. of optim.:

  1st order : $\nabla f(x^*) \cdot d \geq 0$, $\forall$ feasible direc. d.

  2nd order : $d^T \nabla^2 f(x^*) d \geq 0$, " " " " .

- sufficient cond. (?!)

  Exercise: Supp. f is C^2, $x^* \in D$ and

  $\nabla f(x^*) \cdot d \geq 0$ and $d^T \nabla^2 f(x^*) d > 0$ for all admissible d ($d \neq 0$).

  Is x^* necessarily a local min of f?

  (prove or give a counter example).

Constrained optimization:

- supp. D is a lower dimentional subset of $\mathbb{R}^n$ defined by

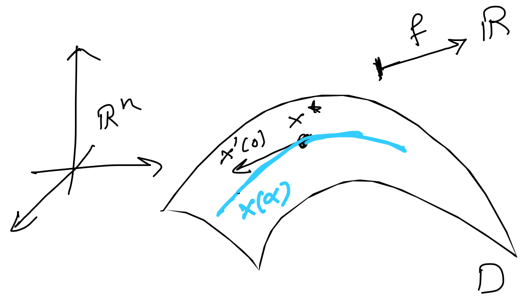
$$D = \{ x : h_1(x) = h_2(x) = \cdots = h_m(x) = 0 \} , \quad h_i : \mathbb{R}^n \to \mathbb{R} \text{ is } C^1$$

- f is also C^1 and $x^* \in D$ is a local min of f on D.

Def: x^* is a regular point of D iff $\{\nabla h_i\}_{i=1}^m$ are linearly independent

→ what is the necessary. cond. for opt.?

let $x(\alpha) \in D$, $\alpha \in [-\varepsilon, \varepsilon]$ be a curve passing through x^* at $\alpha = 0$



- with a similar argument:

$$g(\alpha) \triangleq f(x(\alpha)) \to g'(\alpha) = \nabla f(x(\alpha)) \cdot x'(\alpha)$$

$$\Rightarrow g'(0) = \nabla f(x^*) \cdot x'(0) = 0 \quad \text{as } x^* \text{ is a local min}$$

$x'(0)$: is the linear approx. of $x(\alpha)$ at $\alpha = 0$. why?

is "tangent" to D at x^*

Def: we call the set of all such $x'(0)$, the tangent space at x^* denoted by $T_{x^*}D$.

→ Can we characterize $T_{x^*}D$ more explicitly?

- if $x(\alpha) \in D$ is C^1 $\Rightarrow h_i(x(\alpha)) \equiv 0 \quad \forall \alpha \in [-\varepsilon, \varepsilon]$, $i = 1, \dots, m$.

$$\Rightarrow 0 = \frac{d}{d\alpha} h_i(x(\alpha)) = \nabla h_i(x(\alpha)) \cdot x'(\alpha) \quad \forall \alpha \in [-\varepsilon, \varepsilon].$$

$$\overset{(\alpha=0)}{\Rightarrow} \nabla h_i(x^*) \cdot x'(0) = 0, \quad i = 1, 2, \dots, m.$$

... too true.

• Exercise: Convince yourselve that the converse is also ...
for regular points of D.

So $T_{x^*} D = \{ d \in \mathbb{R}^n : \nabla h_i(x^*) \cdot d = 0 \text{ for } i=1,2,\dots,m \}$.

and the nec. cond. $\nabla f(x^*) \cdot x'(0) = 0$ for all $x'(0) \in T_{x^*} D$

reduces to $\boxed{\nabla f(x^*) \cdot d = 0 \quad \forall d : \nabla h_i(x^*) \cdot d = 0, \ i=1,2,\dots,m.}$

Exercise: • show that this is equivalent to

$$\nabla f(x^*) \in \text{Span} \{ \nabla h_i(x^*), \ i=1,2,\dots,m \}$$

meaning that: there exist real numbers $\lambda_1^*, \dots, \lambda_m^*$
such that

$$\boxed{\nabla f(x^*) + \lambda_1^* \nabla h_1(x^*) + \dots + \lambda_m^* \nabla h_m(x^*) = 0}$$

"1st-order nec. cond. of opt. of regular point $x^* \in D$"

Exercise: is this still a nec. cond. if x^* is not regular in D?
(argue or give a counter-example).

A useful interpretation of this nec. cond.:

Def: "Lagrangian" $\ell(x, \lambda) = f(x) + \sum_{i=1}^{m} \lambda_i h_i(x)$

"Lagrange multipliers"

$$\Rightarrow \nabla \ell(x^*,\lambda^*) = \begin{pmatrix} \nabla f(x^*) + \sum_{i=1}^{m} \lambda_i^* \nabla h_i(x^*) \\ h_1(x^*) \\ \vdots \\ h_m(x^*) \end{pmatrix} = 0$$

So if $x^* \in D$ is local min of f then (x^*,λ^*) is a <u>critical point</u> of Lagrangian $\ell(x,\lambda)$: $\nabla \ell(x^*,\lambda^*) = 0$

<u>2nd-order Conditions</u> :

- supp. f is C^2, $x^* \in D$ is a regular local min of f.

  then by 1-st. order nec. cond. $\exists\, \lambda^* \in \mathbb{R}^m$ s.t.

$$\nabla \ell(x^*,\lambda^*) = \nabla f(x^*) + \sum_{i=1}^{m} \lambda_i^* \nabla h_i(x^*) = 0$$

- what about the Hessian of ℓ w.r.t. x at (x^*,λ^*) :

$$\ell_{xx}(x^*,\lambda^*) = \nabla^2 f(x^*) + \sum_{i=1}^{m} \lambda_i^* \nabla^2 h_i(x^*).$$

with a similar argument as before:

$$\boxed{d^T \ell_{xx}(x^*,\lambda^*)\, d \geq 0 \;, \forall\, d \in T_{x^*} D = \{d : \nabla h_i(x^*) \cdot d = 0 \;,\; i=1,\dots,m\}}$$

"2nd-order Nec. Condition. of opt. of a regular $x^* \in D$"

$\ell \,(x^*,\lambda^*) \succeq 0$ P.S.D.

Note: this is <u>much weaker</u> than asking

2nd-order <u>sufficient cond.</u> for <u>any</u> (regular or irregular) $x^* \in D$

to be a <u>strict local min.</u>:

$$\boxed{\begin{array}{l} \exists\, \lambda^* \in \mathbb{R}^n \text{ s.t. } \nabla \ell(x^*, \lambda^*) = 0 \quad \text{and} \\ d^T \ell_{xx}(x^*, \lambda^*)\, d > 0 \,, \quad \forall\, d \in T_{x^*} D \end{array}}$$

Infinite-dimensional optimization:

$J : V \to \mathbb{R}$ is a functional on an infinite-dim vector space V.

<u>Example:</u> $V = C^0([0,1])$: Continuous functions of $t \in [0,1]$

$$J(y) = \int_0^1 y_t^2 \, dt \rightsquigarrow \text{the "energy of } y\text{"}.$$

- How to make sense of <u>y^* being a local min of J</u>?

$$\left\{ \begin{array}{l} \text{Recall: } y^* \text{ is a local min of } f \text{ if} \\ \exists\, \varepsilon > 0 : \; J(y^*) \le J(y) \quad \forall\, y : \|y - y^*\| < \varepsilon \end{array} \right\}$$

... define this norm!

– But what does $|y-y^*|<\varepsilon$ mean? we need to define ...

<u>Def:</u>
- $C^0([a,b], \mathbb{R}^n)$ $\quad y: \underbrace{[a,b]}_{\text{"time"}} \to \mathbb{R}^n$ continuous

  with $\|y\|_0 = \max_{t\in[a,b]} |y(t)|$.

- $C^1([a,b], \mathbb{R}^n)$ $\quad y:[a,b]\to\mathbb{R}^n$ continuously differentiable

  with $\|y\|_1 = \max_{t\in[a,b]} |y(t)| + \max_{t\in[a,b]} |y'(t)|$

For $k>1$, $C^k([a,b], \mathbb{R}^n)$ can be defined similarly, and

even we could replace $\|y\|_k$ with $\mathcal{L}_p$ norm:

$$\|y\|_{\mathcal{L}_p} = \left(\int_a^b |y(t)|^p \, dt\right)^{1/p}.$$

<u>Next complication:</u> if we want to carry the argument

$g(\alpha) \triangleq J(y^*+\alpha\eta)$, what is "$\frac{dJ}{dy}$" ?!

<u>First variation of J at y in the direction η</u>: (Gateaux derivative)

$$\delta J\big|_y(\eta) \triangleq \lim_{\alpha\to 0} \frac{J(y+\alpha\eta)-J(y)}{\alpha}.$$

... and is linear <u>in η</u>, we get that

Then, if this exists and

$$J(y+\alpha\eta) = J(y) + \delta J|_y(\eta)\,\alpha + o(\alpha).$$

So, we can upgrade our 1-st order nec. cond.:

If $y^* \in A \subset V$ is a local min of J on $A \subset V$, then we can look at all "admissible directions" η $(y+\alpha\eta \in A, |\alpha|<\varepsilon)$.

and argue that $g'(0) = 0$, for all such η.

<u>Exercise</u>: Argue $\boxed{g'(0) = \delta J|_{y^*}(\eta) = 0, \ \forall \eta \text{ admissible is a first order nec. cond. for opt. of } y^* \in A}$

<u>Exercise 1.5 from the book</u>.

<u>Second-order variations</u>:

- Consider a <u>bilinear functional</u> $B: V \times V \to \mathbb{R}$.
  → linear in each argument when the other one is fixed.

- Then, $Q(y) := B(y,y)$ is a quadratic functional on V.

familiar example: $B(d_1,d_2) = d_1^T \nabla^2 f(x^*)\, d_2 \rightsquigarrow Q(d) = d^T \nabla^2 f(x^*)\, d$.

$\delta^2 J|_y : V \to \mathbb{R}$ is called

- A quadratic functional ... the second variation of J at y if for all $\eta \in V$, $\forall \alpha \in \mathbb{R}$

$$J(y+\alpha\eta) = J(y) + \delta J|_y(\eta)\,\alpha + \delta^2 J|_y(\eta)\,\alpha^2 + o(\alpha^2).$$

- Then, a similar argument shows that:

$$\boxed{\begin{array}{c}\text{If } y^* \text{ is a local min of } J \text{ over } A \subset V, \text{ then} \\ \delta^2 J|_{y^*}(\eta) \geq 0 \quad \forall \eta \text{ admissible}\end{array}}$$

"2nd-order nec. cond."

→ Exercise 1.6 in the book.

What about 2nd-order sufficient condition?

Guess: $\delta J|_{y^*}(\eta) = 0$ and $\delta^2 J|_{y^*}(\eta) > 0$, $\forall \eta$ admissible.

→ This is not necessarily correct when V is infinite dimensional!

why?

Exercise: Argue what part of our argument fails!

hint: the ε-Ball around y^* in V is not compact!

For now we don't have a 2-nd order sufficient cond.!