Lecture 5

HW2 posted

Office hours till pending

Quick coverage of univariate conditions

Major goal today: multivariate optimality conditions

We'll also see "test functions" or specific functions that are commonly used as examples.

Multidimensional taylor theorem.

And something brief on convexity (if time)

Software demos:

- $\leftrightarrow$ Cauchy-Schwarz function

1d Taylor

$f(x+h) = f(x) + hf'(x) + \frac{h^2}{2} f''(x) + o(h^2)$

$f(x+h) - f(x) + hf'(x) \approx \frac{h^2}{2} f''(x) + o(h^2)$

$= o(h^2)$

Necessary Cond. ($\Rightarrow$ Necessary) $\Leftrightarrow$ used to signal non-min

Sufficient Cond. ($\Leftarrow$ sufficient $\Rightarrow$ min) $\Leftrightarrow$ used to accept min.

Necessary Condition

Let x^* be a local min then for any dir. d cond.

$f'(x^*d) = 0$

Why?

$f(x^*+td) = f(x^*) + tf'(x^*d)$

if x^* is local min, so then

$tf'(x^*d) \geq 0$

$\Rightarrow f'(x^*d) \neq 0$ then

for t small cond. will be $f(x^*d) < f(x^*)$.

2nd order Necessary Cond

$f(x^*+td) = f(x^*) + tf'(x^*) + \frac{t^2}{2} f''(x^*) + o(t^2) \geq 0$

$f(x^*+td) = f(x^*) + \frac{t^2}{2} f''(x^*d) + o(t^2) \geq 0$

$\Rightarrow f(x^*+td) \geq 0$ as $t \to 0$

$\frac{t^2}{2} f''(x^*d) \geq 0 \Rightarrow f''(x^*d) \geq 0$ by cond $f''(x^*) \geq 0$.

Sufficient Condition

- $f \in \mathbb{R} \to \mathbb{R}$ twice cont. diff.
- $f'(x^*) = 0$
- $f''(x^*) > 0$

then x^* is a local min.

Why?

$f(x^*+td) = f(x^*) + \frac{t^2}{2} f''(x^*d) + o(t^2)$

But $\frac{t^2}{2} f''(x^*d) > 0$ $\Rightarrow$ $f(x^*+td) > f(x^*)$

$\Rightarrow$ x^* is local min.

$\Rightarrow$ the neighborhood $(x^*-\delta, x^*+\delta) \Rightarrow f(x) > f(x^*)$

$\Rightarrow$ x is local min.

Taylor theorem in vector form (d.)

Assume $f(x): \mathbb{R}^n \to \mathbb{R}$

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

Notation

Book	Me	Called
$\nabla f(x)$	$g(x)$	gradient
$\nabla^2 f(x)$	$H(x)$	Hessian

2nd order

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

Notation

Book	Me	Called
$\nabla f(x)$	$g(x)$	gradient
$\nabla^2 f(x)$	$H(x)$	Hessian

$g(x): \mathbb{R}^n \to \mathbb{R}^n$

$g(x) = \nabla f(x) = \begin{bmatrix} \partial f/\partial x_1(x) \\ \vdots \\ \partial f/\partial x_n(x) \end{bmatrix}$

$H(x): \mathbb{R}^n \to \mathbb{R}^{n \times n}$

$\nabla^2 f(x) = H(x) = \begin{bmatrix} \partial^2 f/\partial x_1^2 & \cdots & \partial^2 f/\partial x_1 \partial x_n \\ \vdots & & \\ \partial^2 f/\partial x_n \partial x_1 & \cdots & \partial^2 f/\partial x_n^2 \end{bmatrix}$

$\boxed{[H(x)]_{ij} = \frac{\partial^2 f}{\partial x_i x_j}}$

One fact about $H(x)$

$H(x)$ is symmetric.

$\partial^2 f/\partial x_i \partial x_j = \frac{\partial^2 f}{\partial x_j \partial x_i}$

Example

$f(x) = x_1^2 + x_2 x_1$

$g(x) = \begin{bmatrix} 2x_1 + x_2 \\ x_1 \end{bmatrix}$

$H(x) = \begin{bmatrix} 2 & 1 \\ 1 & 0 \end{bmatrix}$

So $H(x)$ doesn't depend on x!

Quadratic

$f(x) = (1-x_1)^2 + 100(x_2 - x_1^2)^2$

$g(x) = \begin{bmatrix} -2(1-x_1) - 400(x_2 - x_1^2) x_1 \\ 200(x_2 - x_1^2) \end{bmatrix}$

$H(x) = \begin{bmatrix} 2 - 400(x_2 - 3x_1^2) & -400 x_1 \\ -400 x_1 & 200 \end{bmatrix}$

$f(x) = (1-x_1)^2 + 100(x_2 - x_1^2)^2$

$g(x) = \begin{bmatrix} -2(1-x_1) - 400(x_2 - x_1^2) x_1 \\ 200(x_2 - x_1^2) \end{bmatrix}$

$H(x) = \begin{bmatrix} 2 - 400(x_2 - 3x_1^2) & -400 x_1 \\ -400 x_1 & 200 \end{bmatrix}$

If $f \in$ quadratic then $H(x)$ doesn't depend on x!

(why? $\Leftrightarrow$ H $\Rightarrow$ $\nabla^2 f(x+td) = \nabla^2 f(x) + t\nabla^2 f(x+td)$

$f(x+td) = f(x) + p^T g(x) + \frac{t^2}{2} p^T (\nabla^2 f(x+td)) p$

$= f(x) + tf_1 + \frac{t^2}{2} C$.

$f(x+tp) = \begin{bmatrix} f(x) \\ f_0 \end{bmatrix} + \begin{bmatrix} p^T g(x) \\ f_1 \end{bmatrix} t + \frac{t^2}{2} \begin{bmatrix} p^T H p \\ f_2 \end{bmatrix}$

Optimality Conditions

Let x^* be local min then

$g(x^*) = 0$

and $\nabla^2 f(x^*) \succeq 0$

positive semi-def.

SOC0

Stronger positive semi-definite matrix.

If $A: (1 \times n)$, $\boxed{A = A^T}$

$x^T A x \geq 0$ for all x, then we call A positive semi-definite

$\Leftrightarrow$ the det of any leading minor of $A \geq 0$.

$\Rightarrow$ all eigenvals are ≥ 0.

$\Rightarrow$ $A = BB^T$ for some B.

If $A = x^T x$, then A is square not pd.

If $A = BB^T$, then

B is not a square not unique B

B square.

$\Rightarrow$ A is pd.

Pf of $g(x^*) = 0$

Suppose not!

So let $g(x^*) = p^T \neq 0$.

Idea look at $f(x^*)$ in the direction p.

$f(x(t)) = f(x^* + tp) = f(x^*) + t p^T g(x^*) + \frac{t^2}{2} p^T H p$

By cond. $\geq \frac{1}{2} \geq 0$

$t > 0 \leq t \leq \delta$

So $m(t) \geq f(x^*) - \frac{t}{2}$

Contradiction!

$f(x(t)) = f(x^*) + t p^T g(x^*) + \frac{t^2}{2} p^T H p$

$f(x(t)) = f(x^*) + t(0)(x^* - x^*)^T$

$f(x(t)) = f(x^*) + \frac{t^2}{2} p^T H p$

$f(x(t)) - f(x^*) = \frac{t^2}{2} p^T H p$

$f(x(t)) - f(x^*) = o(t^2)$

$f(x(t)) - f(x^*) \geq \frac{t^2}{2} \lambda_{min} \|p\|^2$

$\Rightarrow$ $f(x(t)) - f(x^*) \geq \frac{t^2}{2} \lambda_{min} \|p\|^2 + o(t^2)$

$\Rightarrow$ $f(x(t)) - f(x^*) \geq \frac{t^2}{2} \lambda_{min} \|p\|^2$

$\Rightarrow$ $f(x(t)) - f(x^*) > 0$

$\Rightarrow$ $f(x^*) < f(x(t))$

$\Rightarrow$ x^* is local min.

See Theorem 2.3 in Nitsch.

2nd order sufficient cond are more relevant!

Quiz

Let $f(x): \mathbb{R}^n \to \mathbb{R}$ (no cont. assumptions)

Let $\bar{x}_0$ be a local min on all lines passing through $\bar{x}_0$.

Is $\bar{x}_0$ a local min of f?

"on a line" $\Rightarrow$

$f(\bar{x}_0 + \alpha \bar{p}) = \boxed{h(\alpha)}$

$\alpha = 0$ is local min for any choice of $\bar{p}$.