Last time:

Theorem: Necessary conditions for optimality

(1) If f is continuously differentiable & x^* is a local min then

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

(2) If $\nabla^2 f$ is continuous and x^* is a local minimum

$$\nabla^2 f(x^*) \succcurlyeq 0$$

notation:

$A \succcurlyeq 0$

$\iff$ A is a PSD matrix

This theorem means that if f is cont. differentiable & has a continuous Hessian then we must have

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

<u>Proof</u>: (1) Suppose by way of contradiction that $\nabla f(x^*) \neq 0$, then $\vec{v} = -\nabla f(x^*)$ is a descent direction, so x^* is not a local minimum, and we have a contradiction.

(2) Since x^* is a local minimum then $f(x^* + t\vec{v}) \geq f(x^*)$ $\forall \vec{v}$ and for all t that are small enough.

Now, Taylor's theorem tells us that

$$f(x^* + tv) = f(x^*) + t v^T \overset{0 \text{ by } (1)}{\cancel{\nabla f(x^*)}} + \frac{1}{2} t^2 v^T \nabla^2 f(x^* + \tilde{t} v) v$$

for $\tilde{t} \in (0, t)$.

$$\Rightarrow \overbrace{f(x^* + tv) - f(x^*)}^{\geq 0 \text{ bec } x^* \text{ is local min}} = \frac{1}{2} t^2 v^T \nabla^2 f(x^* + \tilde{t} v) v$$

$\forall v$

$$\Rightarrow \frac{1}{2} t^2 v^T \nabla^2 f(x^* + \tilde{t} v) v \geq 0 \quad \forall v$$

Taking limits as $t \to 0$ we also have $\tilde{\varepsilon} \to 0$

and this gives $v^T \nabla^2 f(x^*) v \geq 0$ $\forall v$

which is equivalent to saying $\nabla^2 f(x^*) \succcurlyeq 0$.

$\blacksquare$

Theorem:

(Sufficient cond. for optimality).

If f is twice continuously differentiable, and

x^* satisfies $\begin{cases} \nabla f(x^*) = 0 \\ \nabla^2 f(x^*) \succ 0 \end{cases}$

positive def.

then x^* is a local min.

$$\boxed{\text{So having twice cont. diff } f. \text{ with } \begin{cases} \nabla f(x^*) = 0 \\ \nabla^2 f(x^*) \succ 0 \end{cases} \text{ guarantees that } x^* \text{ is a local min.}}$$

Proof: $\forall \vec{h}$ with $\|h\|$ small:

$$f(x^*+h) = f(x^*) + h^T \cancelto{0}{\nabla f(x^*)} + \underbrace{\frac{1}{2} h^T \nabla^2 f(x^* + \alpha h) h}_{>0 \text{ by pos. def. of } \nabla^2 f(x^*) \text{ and continuity of } \nabla^2 f}$$

$\hookrightarrow \alpha \in (0,1)$

$$\geq f(x^*)$$

$\Rightarrow$ local min at x^* ∎

Example: Let $f(x) = x_1^3 + 2x_2^2$

Here $\nabla f(x) = (3x_1^2 ; 4x_2)$

So $\nabla f(x^*) = 0 \Rightarrow x^* = (0,0)$

& $\nabla^2 f(x) = \begin{pmatrix} 6x_1 & 0 \\ 0 & 4 \end{pmatrix}$

$$\nabla^2 f(x^*) = \begin{pmatrix} 0 & 0 \\ 0 & 4 \end{pmatrix} \succeq 0$$

But $f(x^*) = f(0,0) = 0$

& $f(-\varepsilon, 0) = -\varepsilon^3 < 0 \quad \forall \varepsilon > 0$

So x^* is not a local min.

Example:

$$f(x) = \frac{1}{2}x_1^2 + x_1x_2 + 2x_2^2 - 4x_1 - 4x_2 - x_2^3$$

$$\nabla f(x) = \begin{pmatrix} x_1 + x_2 - 4 \\ x_1 + 4x_2 - 4 - 3x_2^2 \end{pmatrix}$$

$$\nabla^2 f(x) = \begin{pmatrix} 1 & 1 \\ 1 & 4 - 6x_2 \end{pmatrix}$$

$$\nabla f(x) = 0 \implies x^* = (4, 0)$$
$$x^{**} = (3, 1)$$

(by ~~the~~ solving $\nabla f(x) = 0$)

are candidates for local minimizers.

But are they really?

Let's start with x^{**}:

$$\nabla^2 f(x^{**}) = \begin{pmatrix} 1 & 1 \\ 1 & -2 \end{pmatrix}$$

which has eigenvalues $\lambda_1 < 0$
$\lambda_2 > 0$

(check this)

so $\nabla^2 f(x^{**})$ is not pos. semidef.

$\Rightarrow x^{**}$ is not a local min.

How about x^*:

$$\nabla^2 f(x^*) = \begin{pmatrix} 1 & 1 \\ 1 & 4 \end{pmatrix}$$

which is now positive def.
(check)

So x^* satisfies the conditions of the suff. cond. theorem.

$\Rightarrow x^*$ is a local min.

The previous theorems did not assume f was convex, which we now will assume:

$\boxed{\begin{array}{l}\text{Theorem: Let } f: \mathbb{R}^n \to \mathbb{R} \text{ be convex \& cont. differentiable,} \\ \text{then } x^* \text{ is a global minimizer if and only if } \nabla f(x^*) = 0.\end{array}}$

Proof: (1) Suppose x^* is a global minimizer, then there is no descent direction from x^*
$$\Rightarrow \quad \nabla f(x^*) = 0$$

(2) Suppose $\underline{\nabla f(x^*)}=0$, then convexity gives $\forall x$

$$f(x) \geq f(x^*) + \underbrace{\nabla f(x^*)^T}_{0}(x-x^*)$$

$\Rightarrow f(x) \geq f(x^*)$,

$\Rightarrow x^*$ is a (global) min. $\blacksquare$

Remark: If f satisfies conditions,

$$\nabla f(x) = 0 \iff x \text{ global min.}$$

When stop GD?

Recall $x_t - x_{t-1} = -\mu \nabla f(x_{t-1})$

$\Rightarrow \|x_t - x_{t-1}\| = \mu \|\nabla f(x_{t-1})\|$

$\Rightarrow$ when at global min,

$\|x_t - x_{t-1}\| = 0. \Rightarrow$ stop moving at solution!