Lecture 4: Optimality Conditions for Unconstrained Problems

goals:

- discuss general optimality conditions for unconstrained problems
- discuss their limitations, which motivates our study of convex optimization problems

---

Let's consider the unconstrained optimization problem

$$\text{minimize} \quad f(x)$$

- no constraints, which means $\mathcal{X} = \mathbb{R}^n$

First Order Necessary Condition for Optimality

Theorem: If $\bar{x}$ is an unconstrained local minimum of a differentiable function $f: \mathbb{R}^n \to \mathbb{R}$, then $\nabla f(\bar{x}) = 0$.

Proof: If $\nabla f(\bar{x}) \neq 0$, then $\exists\, i$ such that $\frac{\partial f}{\partial x_i}(\bar{x}) \neq 0$.

This means the function value can be reduced in this direction, so $\bar{x}$ cannot be a local minimum ∎

For $\alpha > 0$ look at first order Taylor expansion around $\bar{x}$

$$f(\bar{x} + \alpha e_i) = f(\bar{x}) + \alpha e_i^T \nabla f(\bar{x}) + o(\alpha)$$

$$\Longrightarrow \quad \frac{f(\bar{x} + \alpha e_i) - f(\bar{x})}{\alpha} = e_i^T \nabla f(x) + \frac{o(\alpha)}{\alpha}$$

Note that $\lim_{\alpha \downarrow 0}$ RHS < 0, which means that there are points close to $\bar{x}$ with lower value, i.e., $\bar{x}$ cannot be a local minimum. $\blacksquare$

Here e_i is any direction where $\nabla f(\bar{x})$ decreases.

a bit more formally

- The converse is false in general, since $\nabla f = 0$ at a local maximum (or saddle point)
- All unconstrained optimization algorithms seek a point where $\nabla f = 0$

Second Order Optimality Conditions

① A necessary condition for local optimality

Theorem: If $\bar{x}$ is an unconstrained local minimum of a twice differentiable function $f : \mathbb{R}^n \to \mathbb{R}$, then in addition to $\nabla f(\bar{x}) = 0$, we have

$$\nabla^2 f(\bar{x}) \succeq 0$$

i.e. the Hessian of f at $\bar{x}$ is positive semidefinite.

Proof: Consider any $y \in \mathbb{R}^n$, and for $\alpha > 0$ look at the second order Taylor expansion of f around $\bar{x}$

$$f(\bar{x} + \alpha y) = f(\bar{x}) + \alpha y^T \nabla f(\bar{x}) + \frac{\alpha^2}{2} y^T \nabla^2 f(\bar{x}) y + o(\alpha^2).$$

Since $\nabla f(\bar{x})$ must be zero (from the previous Theorem),

$$\frac{f(\bar{x} + \alpha y) - f(\bar{x})}{\alpha^2} = \frac{1}{2} y^T \nabla^2 f(\bar{x}) y + \frac{o(\alpha^2)}{\alpha^2}$$

Since $\bar{x}$ is a local minimum, the LHS is nonnegative for sufficiently small α by definition. Thus

$$\lim_{\alpha \downarrow 0} \frac{1}{2} y^T \nabla^2 f(\bar{x}) y + \frac{o(\alpha^2)}{\alpha^2} \geq 0$$

But since $\lim_{\alpha \downarrow 0} \frac{o(\alpha^2)}{\alpha^2} = 0$ by definition of $o(\alpha^2)$

then

$$y^T \nabla^2 f(\bar{x}) y \geq 0 .$$

Since y was arbitrary, then $\nabla^2 f(\bar{x}) \succeq 0$.

∎

- The converse of this Theorem is also false

Ex Consider $f(x) = -x^4$

$$\nabla f(x) = -4x^3$$

$\Rightarrow$ $x = 0$ is a stationary point

$$\nabla^2 f(x) = -12x^2 \Rightarrow \nabla^2 f(0) = 0$$

However, clearly $x = 0$ is a maximum

since $f(x) < 0 \quad \forall x \neq 0$

---

② A sufficient condition for local optimality

Theorem: Suppose $f: \mathbb{R}^n \to \mathbb{R}$ is twice differentiable and there is a point $\bar{x}$ such that $\nabla f(\bar{x}) = 0$ and

$$\nabla^2 f(\bar{x}) \succ 0$$

i.e. the Hessian of f at $\bar{x}$ is positive definite.

Then $\bar{x}$ is a strict local minimum of f.

Proof: Let $\lambda > 0$ be the minimum eigenvalue of $\nabla^2 f(\bar{x})$.
Then $\nabla^2 f(\bar{x}) - \lambda I \succeq 0$, which implies

$$y^T \nabla^2 f(\bar{x}) y \geq \lambda \|y\|^2 \quad \forall y \in \mathbb{R}^n$$

Once again, a Taylor expansion around $\bar{x}$ yields

$$\begin{aligned} f(\bar{x}+y) - f(\bar{x}) &= y^T \nabla f(\bar{x}) + \frac{1}{2} y^T \nabla^2 f(\bar{x}) y + o(\|y\|^2) \\ &\geq \frac{1}{2} \lambda \|y\|^2 + o(\|y\|^2) \\ &= \|y\|^2 \left(\frac{\lambda}{2} + \frac{o(\|y\|^2)}{\|y\|^2} \right) \end{aligned}$$

Since $\displaystyle\lim_{\|y\| \to 0} \frac{o(\|y\|^2)}{\|y\|^2} = 0$, then we can always

make $\|y\|$ small enough so that the RHS is strictly positive,
i.e. $\exists \delta > 0$ such that $\displaystyle\frac{\lambda}{2} > \frac{o(\|y\|^2)}{\|y\|^2}$ $\forall y$ with $\|y\| \leq \delta$.

Thus,

$$f(\bar{x}+y) > f(\bar{x}) \quad \forall y \in B(0, \delta)$$

which by definition means that $\bar{x}$ is a strict local minimum.

$\boxtimes$

- The converse of this Theorem is also false

Ex Consider $f(x) = x^4$

$$\nabla f(x) = 4x^3$$

$\Longrightarrow$ $x = 0$ is a stationary point

$$\nabla^2 f(x) = 12x^2 \Longrightarrow \nabla^2 f(0) = 0$$

But clearly $x = 0$ is a strict minimum since $f(x) > 0$ $\forall x \neq 0$.

Ex Quadratic objective

$$\text{minimize} \quad \frac{1}{2} x^T P x + q^T x$$

$$\nabla f(x) = Px + q = 0 \Longrightarrow x = -P^{-1} q$$

is a candidate solution, if P is invertible.

$$\nabla^2 f(x) = P$$

- If $P \succ 0$, the last Theorem says that $x = -P^{-1}q$ is a strict local min
- If $P \prec 0$, it says $x = -P^{-1}q$ is a strict local max
- If P is indefinite, $x = P^{-1}q$ is a saddle point

- If $P \succeq 0$ (or $P \preceq 0$), the Theorems as stated say <u>nothing</u> about candidate solutions

---

<u>Summary</u>

- These results give basis for wide array of general optimization algorithms, essentially all of which search for points where $\nabla f(x) = 0$ and <u>hope</u> these are <u>local</u> minima
- However, it's possible for all three conditions to be <u>inconclusive</u> about testing local optimality
- Also, they say <u>absolutely nothing</u> about global optimality of solutions
- We'll see that <u>convexity</u> allows us do make statements about global optimality. Convex problems:
  - are roughly the broadest class we can (globally) solve efficiently
  - have nice geometric structure
  - have excellent + widely available software tools
  - are <u>surprisingly</u> common in many, many applications!