Lecture 1 Unconstrained Optimization

- Definitions
- Necessary first/second order optimality condition – variational approach
- Sufficient optimality condition
- Existence of optimal solutions
- Quadratic minimization – characterization/existence of optimal solutions
- Convexity

Definitions

$$\boxed{\begin{array}{ll} \text{minimize} & f(\boldsymbol{x}) \\ \text{subject to} & \boldsymbol{x} \in \mathbb{R}^n \end{array}}$$

- local minimum $\boldsymbol{x}^*$: $\exists \epsilon > 0$ s.t. $f(\boldsymbol{x}) \geq f(\boldsymbol{x}^*)$, for all $\|\boldsymbol{x} - \boldsymbol{x}^*\| \leq \epsilon$.
- Strict local minimum: $\exists \epsilon > 0$ s.t. $f(\boldsymbol{x}) > f(\boldsymbol{x}^*)$, for all $\|\boldsymbol{x} - \boldsymbol{x}^*\| \leq \epsilon$, $\boldsymbol{x} \neq \boldsymbol{x}^*$.
- Global minimum: $f(\boldsymbol{x}) \geq f(\boldsymbol{x}^*)$ for all $\boldsymbol{x} \in \mathbb{R}^n$.
- Strict global min: $f(\boldsymbol{x}) > f(\boldsymbol{x}^*)$, for all $\boldsymbol{x} \neq \boldsymbol{x}^*$.
- Geometrically ...

Checkable Conditions for Local Min

- Given a point $\boldsymbol{x}$, how do we know if it is a (strict) local/global min of a (twice) continuously differentiable function f?

- Need easily checkable **necessary optimality conditions** – variational approach

$$\boxed{\nabla f(\boldsymbol{x}^*) = \mathbf{0}, \quad \nabla^2 f(\boldsymbol{x}^*) \succeq \mathbf{0}.} \tag{1}$$

- One dimensional case: suppose x^* is a local min of a differentiable function $f : \mathbb{R} \mapsto \mathbb{R}$

$$0 \leq \lim_{x^r \downarrow x^*} \frac{f(x^r) - f(x^*)}{x^r - x^*} = f'(x^*) = \lim_{x^r \uparrow x^*} \frac{f(x^r) - f(x^*)}{x^r - x^*} \leq 0$$

$$0 \leq \lim_{x^r \to x^*} \frac{f(x^r) - f(x^*) - f'(x^*)(x^r - x^*)}{(x^r - x^*)^2} = \frac{1}{2} f''(x^*)$$

- For higher dimensions: fix any $\boldsymbol{d} \in \mathbb{R}^n$. Consider the one dimensional function $g(\alpha) = f(\boldsymbol{x}^* + \alpha \boldsymbol{d})$, which is minimized at $\alpha = 0$

$$\Longrightarrow \quad g'(0) = \nabla f(\boldsymbol{x}^*)'\boldsymbol{d} = 0, \quad g''(0) = \boldsymbol{d}'\nabla^2 f(\boldsymbol{x}^*)\boldsymbol{d} \geq 0, \quad \forall\ \boldsymbol{d} \in \mathbb{R}^n.$$

  implying (1).

- **Example:** $f(x) = |x|^3,\ x^3,\ -|x|^3$.
  Check the necessary conditions at $x = 0$. Plot f.

- **Sufficient condition** for local optimality:

$$\boxed{\nabla f(\boldsymbol{x}^*) = \boldsymbol{0}, \quad \nabla^2 f(\boldsymbol{x}^*) \succ \boldsymbol{0}.} \tag{2}$$

  since

$$f(\boldsymbol{x}) - f(\boldsymbol{x}^*) = \frac{1}{2}(\boldsymbol{x} - \boldsymbol{x}^*)'\nabla^2 f(\boldsymbol{x}^* + \alpha(\boldsymbol{x} - \boldsymbol{x}^*))(\boldsymbol{x} - \boldsymbol{x}^*) \geq 0, \quad 0 \leq \alpha \leq 1.$$

Why Optimality Conditions are Important?

- Optimality conditions are useful because:
  - ⋆ they provide a means of guaranteeing that a candidate solution is indeed optimal (sufficient conditions), and
  - ⋆ they indicate when a point is not optimal (necessary conditions)
  - ⋆ they help narrow down the list of potential solution candidates
- Furthermore they
  - ⋆ guide in the design of algorithms, since lack of optimality $\Longleftrightarrow$ indication of improvement

Use of Optimality Conditions

minimize	$f(\boldsymbol{y}) = e^{y_1} + e^{y_2} + \cdots + e^{y_n}$
subject to	$y_1 + y_2 + \cdots + y_n = s.$

- First we eliminate y_n by substituting $y_n = s - y_1 - y_2 - \cdots - y_{n-1}$ in the objective function. The new objective function is

$$g(\boldsymbol{y}) = e^{y_1} + e^{y_2} + \cdots + e^{y_{n-1}} + e^{s - y_1 - y_2 - \cdots - y_{n-1}}$$

- The first order optimality condition $\nabla g(\boldsymbol{y}^*) = \mathbf{0}$ implies, for $i = 1, 2, ..., n-1$,

$$\frac{\partial g}{\partial y_i} = e^{y_i^*} - e^{s - y_1^* - y_2^* - \cdots - y_{n-1}^*} = 0, \text{ or } y_i^* = s - y_1^* - y_2^* - \cdots - y_{n-1}^*.$$

  with the minimum $f^* = ne^{s/n}$. The 2nd order sufficient condition holds, so

$$e^{y_1} + e^{y_2} + \cdots + e^{y_n} \geq ne^{(y_1 + y_2 + \cdots + y_n)/n}.$$

- Using $x = e^y$, we obtain the well-known arithmetic-geometric inequality.

Use of Optimality Conditions

- **Example:** find the local/global mins of $f(x) = x^2 - x^4$.

$$
\begin{aligned}
& \nabla f(x) = f'(x) = 2x - 4x^3 = 0 \\
\Rightarrow \quad & x = 0,\ x = \pm\tfrac{\sqrt{2}}{2} \ \text{ candidates} \\
& f''(x) = 2 - 12x^2 \\
\Rightarrow \quad & f''(0) = 2 > 0, \quad f''\left(\pm\tfrac{\sqrt{2}}{2}\right) = 2 - 12 \times \tfrac{1}{2} < 0. \\
\Rightarrow \quad & x = 0 \text{ is a strict local min; } \pm\tfrac{\sqrt{2}}{2} \text{ are strict local max.}
\end{aligned}
$$

Given there is a unique local min $x = 0$, can we then conclude that $x = 0$ is also the unique global min?

- Global min does not exist. Plot.

Existence of Optimal Solution

- **Example:**

$$\inf_{x\in\mathbb{R}} e^{-|x|} = ?$$

  is the infimum attained?

- **Bolzano-Weierstrass Theorem**: every continuous function f attains its infimum over compact set X. That is, there exists an $\boldsymbol{x}^* \in X$ such that

$$f(\boldsymbol{x}^*) = \inf_{\boldsymbol{x}\in X} f(\boldsymbol{x}).$$

- Consequently, if the level set

$$f(\boldsymbol{x}) \le f(\boldsymbol{x}^0)$$

  of continuous function f is compact for some $\boldsymbol{x}^0$, then the global min of

$$\begin{array}{ll} \text{minimize} & f(\boldsymbol{x}) \\ \text{subject to} & \boldsymbol{x} \in \mathbb{R}^n \end{array}$$

  is attained. Check the level sets of $e^{-|x|}$.

- Another sufficient condition (**coercivity**): $f(\boldsymbol{x}) \to \infty$ as $\|\boldsymbol{x}\| \to \infty$.

Unconstrained Quadratic Optimization

$$\begin{array}{ll} \text{minimize} & \frac{1}{2}\boldsymbol{x}'\boldsymbol{Q}\boldsymbol{x} + \boldsymbol{b}'\boldsymbol{x} \\ \text{subject to} & \boldsymbol{x} \in \mathbb{R}^n \end{array}$$

- Necessary condition for optimality:

$$\nabla f(\boldsymbol{x}) = \boldsymbol{Q}\boldsymbol{x} + \boldsymbol{b} = \mathbf{0}, \quad \nabla^2 f(\boldsymbol{x}) = \boldsymbol{Q} \succeq \mathbf{0}. \tag{3}$$

- What if the linear system $\boldsymbol{Q}\boldsymbol{x} + \boldsymbol{b} = \mathbf{0}$ is infeasible? What if $\boldsymbol{Q} \not\succeq \mathbf{0}$?

- Sufficient condition requires $\boldsymbol{Q} \succ \mathbf{0}$.

- **Claim:** the necessary condition (3) is also sufficient; any local optimal solution is also globally optimal.

A 2-dimensional Example

$$\begin{array}{ll} \text{minimize} & f(x,y) = \frac{1}{2}(\alpha x^2 + \beta y^2) - x \\ \text{subject to} & (x,y) \in \mathbb{R}^2 \end{array}$$

- $\alpha > 0,\ \beta > 0$ (strongly convex): $(1/\alpha, 0)$ is the unique global minimum.
- $\alpha = 0$ (convex): There is no global minimum
- $\alpha > 0,\ \beta = 0$ (convex): $\{(1/\alpha, \xi) \mid \xi \in \mathbb{R}\}$ is the set of global minima
- $\alpha > 0,\ \beta < 0$ or $\alpha < 0$ (non-convex case): There is no global minimum
- Plot the level sets of all four cases

Linear Least Squares

$$\boxed{\begin{array}{ll} \text{minimize} & \frac{1}{2}\|\boldsymbol{A}\boldsymbol{x} - \boldsymbol{c}\|^2 \\ \text{subject to} & \boldsymbol{x} \in \mathbb{R}^n \end{array}}$$

- $\boldsymbol{A}$ may be fat (under-determined), tall (over-determined), or rank-deficient.
- Note that $\boldsymbol{Q} = \boldsymbol{A}'\boldsymbol{A}$, $\boldsymbol{b} = \boldsymbol{A}'\boldsymbol{c}$.
- Necessary & sufficient optimality condition:

$$\boldsymbol{A}'\boldsymbol{A}\boldsymbol{x}^* - \boldsymbol{A}'\boldsymbol{c} = \boldsymbol{0}$$

  which always has a solution.
- Linear least squares problem may have unbounded levels, but always admits a solution.

Role of Convexity

Suppose $f : \mathbb{R}^n \mapsto \mathbb{R}$ satisfies

$$f(\alpha \boldsymbol{x} + (1-\alpha)\boldsymbol{y}) \leq \alpha f(\boldsymbol{x}) + (1-\alpha) f(\boldsymbol{y}), \quad \forall\ \alpha \in [0,1],\ \boldsymbol{x},\ \boldsymbol{y}.$$

then f is called a convex function. [or $-f$ is called a concave function.]

- A set X is convex iff ι_X (the indicator function) is convex.
- If f is continuously differentiable, f is (strongly) convex iff
$$f(\boldsymbol{y}) \geq f(\boldsymbol{x}) + \nabla f(\boldsymbol{x})'(\boldsymbol{y}-\boldsymbol{x}) + \sigma \|\boldsymbol{y}-\boldsymbol{x}\|^2$$
(and $\sigma > 0$). If f is twice continuously differentiable, then
$$f \text{ is (strongly) convex} \quad \Longleftrightarrow \quad \nabla^2 f(\boldsymbol{x}) \succeq \mathbf{0} (\succ \mathbf{0}) \text{ for all } \boldsymbol{x}.$$
- Examples of convex functions: $\boldsymbol{a}'\boldsymbol{x} + b$, e^x, $-\ln x$, x^2, $\frac{1}{2}\boldsymbol{x}'\boldsymbol{Q}\boldsymbol{x} + \boldsymbol{b}'\boldsymbol{x}$, $\boldsymbol{Q} \succeq \mathbf{0}$.
- For convex differentiable f, each local min is also a global min (why?), so the necessary and sufficient optimality condition is
$$\nabla f(\boldsymbol{x}) = 0.$$
Claim: the set of minimizers of f is a convex set.

Applications of Convex Functions

The arithmetic-geometric inequality

$$(x_1 x_2 \cdots x_n)^{1/n} \leq \frac{1}{n}(x_1 + x_2 + \cdots + x_n), \quad \forall\ x_i \geq 0$$

can be derived from the convexity of $-\ln x$ function.

- First, the convexity of f is equivalent to

$$f(\alpha_1 \boldsymbol{x}^1 + \alpha_2 \boldsymbol{x}^2 + \cdots \alpha_r \boldsymbol{x}^r) \leq \sum_{i=1}^{r} \alpha_i f(\boldsymbol{x}^i), \quad \forall\ \boldsymbol{x}^i \text{ and } \alpha_i \geq 0,\ \sum_{i=1}^{r} \alpha_i = 1.$$

- Thus, the convexity of $f(x) = -\ln x$ implies

$$-\ln\left(\frac{1}{n}x_1 + \frac{1}{n}x_2 + \cdots + \frac{1}{n}x_n\right) \leq -\frac{1}{n}\sum_{i=1}^{n} \ln x_n.$$

  implying the arithmetic-geometric inequality.