MATH 4211/6211 – Optimization

Basics of Optimization Problems

Xiaojing Ye
Department of Mathematics & Statistics
Georgia State University

A standard minimization problem is written as

$$\underset{\boldsymbol{x}\in\Omega}{\text{minimize}}\ f(\boldsymbol{x})$$

where Ω is a subset of $\mathbb{R}^n$.

- $f(x)$ is called the **objective function** or **cost function**;

- $\Omega = \mathbb{R}^n$: unconstrained minimization;

- $\Omega \subset \mathbb{R}^n$ explicitly given: set constrained minimization;

- $\Omega = \{\boldsymbol{x} \in \mathbb{R}^n : g(\boldsymbol{x}) = 0, h(\boldsymbol{x}) \leq 0\}$: functional constrained minimization.

- **Local minimizer**: $\boldsymbol{x}^*$ is called a local minimizer of f if $\exists\epsilon > 0$ such that $f(\boldsymbol{x}) \geq f(\boldsymbol{x}^*)$ for all $\boldsymbol{x} \in B(\boldsymbol{x}^*, \epsilon)$.

- **Global minimizer**: $\boldsymbol{x}^*$ is called a global minimizer of f if $f(\boldsymbol{x}) \geq f(\boldsymbol{x}^*)$ for all $\boldsymbol{x} \in \Omega$.

- **Strict local minimizer**: $\boldsymbol{x}^*$ is called a strict local minimizer of f if $\exists\epsilon > 0$ such that $f(\boldsymbol{x}) > f(\boldsymbol{x}^*)$ for all $\boldsymbol{x} \in B(\boldsymbol{x}^*, \epsilon) \setminus \{\boldsymbol{x}^*\}$.

- **Strict global minimizer**: $\boldsymbol{x}^*$ is called a strict global minimizer of f if $f(\boldsymbol{x}) > f(\boldsymbol{x}^*)$ for all $\boldsymbol{x} \in \Omega \setminus \{\boldsymbol{x}^*\}$.

- **(Strict) local/global maximizers** are defined similarly.

- $\boldsymbol{d} \in \mathbb{R}^n$ is called a **feasible direction** at $\boldsymbol{x} \in \Omega$ if $\exists \epsilon > 0$ such that $\boldsymbol{x} + \alpha \boldsymbol{d} \in \Omega$ for all $\alpha \in [0, \epsilon]$.

- **Directional derivative** of f in the direction $\boldsymbol{d}$ is defined by

$$\lim_{\alpha \to 0} \frac{f(\boldsymbol{x} + \alpha \boldsymbol{d}) - f(\boldsymbol{x})}{\alpha} = (\nabla f(\boldsymbol{x}))^\top \boldsymbol{d}$$

Example. Let $f : \mathbb{R}^3 \to \mathbb{R}$ defined by $f(\boldsymbol{x}) = x_1 x_2 x_3$, and $\boldsymbol{d} = \left(\frac{1}{2}, \frac{1}{2}, \frac{1}{\sqrt{2}} \right)^T$.
Compute the directional derivative of f in the direction of $\boldsymbol{d}$:

$$\nabla f(\boldsymbol{x})^T \boldsymbol{d} = (x_2 x_3, x_1 x_3, x_1 x_2) \left(\frac{1}{2}, \frac{1}{2}, \frac{1}{\sqrt{2}} \right)^T = \frac{x_2 x_3 + x_1 x_3 + \sqrt{2} x_1 x_2}{2}$$

Suppose f is differentiable.

- **First order necessary condition (FONC)**: $\boldsymbol{x}^*$ is a local minimizer, then for any feasible direction $\boldsymbol{d}$ at $\boldsymbol{x}^*$, there is $\boldsymbol{d}^\top \nabla f(\boldsymbol{x}^*) \geq 0$.

- If $\boldsymbol{x}^*$ is an interior point of Ω, then FONC reduces to $\nabla f(\boldsymbol{x}^*) = 0$.

Example. Consider the problem

$$\text{minimize} \quad f(\boldsymbol{x}) := x_1^2 + 0.5x_2^2 + 3x_2 + 4.5, \quad \text{subject to } x_1, x_2 \geq 0.$$

1. Is the FONC for a local minimizer satisfied at $\boldsymbol{x} = (1, 3)^\top$?

2. Is the FONC for a local minimizer satisfied at $\boldsymbol{x} = (0, 3)^\top$?

3. Is the FONC for a local minimizer satisfied at $\boldsymbol{x} = (1, 0)^\top$?

4. Is the FONC for a local minimizer satisfied at $\boldsymbol{x} = (0, 0)^\top$?

Idea: First compute $\nabla f(\boldsymbol{x}) = (2x_1, x_2 + 3)^\top$. If $\boldsymbol{x}$ is interior point, check if $\nabla f(\boldsymbol{x}) = 0$; otherwise, check if $\boldsymbol{d}^\top \nabla f(\boldsymbol{x}) \geq 0$ for all feasible direction $\boldsymbol{d}$.

Suppose f is twice differentiable.

- **Second order necessary condition (SONC)**: $\boldsymbol{x}^*$ is a local minimizer, then for any feasible direction $\boldsymbol{d}$ at $\boldsymbol{x}^*$ such that $\boldsymbol{d}^\top \nabla f(\boldsymbol{x}^*) = 0$, there is $\boldsymbol{d}^\top \nabla^2 f(\boldsymbol{x}^*) \boldsymbol{d} \geq 0$.

- If $\boldsymbol{x}^*$ is an interior point of Ω, then SONC reduces to $\nabla f(\boldsymbol{x}^*) = 0$ and $\nabla^2 f(\boldsymbol{x}^*) \succeq 0$.

Note that these conditions are necessary but not sufficient. This means $\boldsymbol{x}$ may not be a local minimizer even if FONC or SONC is satisfied.

Example. Consider $f(x) = x^3$ for $x \in \mathbb{R}$, then $f'(0) = 0, f''(0) = 0$. So $x = 0$ satisfies FONC (and even SONC), but it is not a local minimizer.

Example. Let $f(\boldsymbol{x}) = x_1^2 - x_2^2$, then the gradient is $\nabla f(\boldsymbol{x}) = (2x_1, -2x_2)^\top$ and the Hessian is

$$\nabla^2 f(\boldsymbol{x}) = \begin{bmatrix} 2 & 0 \\ 0 & -2 \end{bmatrix}$$

which is not positive semi-definite. So $\boldsymbol{x} = (0, 0)^\top$ does not satisfy SONC and hence is not a local minimizer of f.

Let’s see a sufficient condition of interior local minimizer. Suppose f is twice differentiable.

Second order sufficient condition (SOSC): if x is an interior point and satisfies both $\nabla f(x^*) = 0$ and $\nabla^2 f(x^*) \succ 0$, then x is a strict local minimizer of f.

Example. Let $f(\boldsymbol{x}) = x_1^2 + x_2^2$, then the gradient is $\nabla f(\boldsymbol{x}) = (2x_1, 2x_2)^\top$ and the Hessian is

$$\nabla^2 f(\boldsymbol{x}) = \begin{bmatrix} 2 & 0 \\ 0 & 2 \end{bmatrix}$$

So $\boldsymbol{x} = (0, 0)^\top$ satisfies SOSC and hence is a strict local minimizer of f.

Let's take a look of the **Newton's method** for univariate functions:

$$\operatorname*{minimize}_{x \in \mathbb{R}} f(x)$$

Suppose we obtained $x^{(k)}$. Now approximate $f(x)$ nearby $x^{(k)}$ by a quadratic function $q(x)$:

$$q(x) = f(x^{(k)}) + f'(x^{(k)})(x - x^{(k)}) + \frac{1}{2} f''(x^{(k)})(x - x^{(k)})^2$$

Note that $q(x^{(k)}) = f(x^{(k)})$, $q'(x^{(k)}) = f'(x^{(k)})$ and $q''(x^{(k)}) = f''(x^{(k)})$.

If q approximates f well (we know it's true nearby $x^{(k)}$), then we can minimize $q(x)$ instead of $f(x)$. To this end, we compute

$$q'(x) = f'(x^{(k)}) + f''(x^{(k)})(x - x^{(k)}) = 0$$

and solve for x to use as our next iterate $x^{(k+1)}$, i.e.,

$$x^{(k+1)} = x^{(k)} - \frac{f'(x^{(k)})}{f''(x^{(k)})}$$

In practice, we give an initial guess $x^{(0)}$, and iterate the formula above to generate $\{x^{(k)} : k = 0, 1, \dots\}$. Then $x^{(k)} \to x^*$ quickly (under some conditions).

Example. Use Newton's method to find the minimizer of

$$f(x) = \frac{1}{2}x^2 - \sin x$$

with initial value $x^{(0)} = 0.5$. Stop when $|x^{(k+1)} - x^{(k)}| < 10^{-5}$.

Solution. We first compute

$$f'(x) = x - \cos x, \quad f''(x) = 1 + \sin x$$

So the Newton's method gives

$$x^{(k+1)} = x^{(k)} - \frac{x^{(k)} - \cos x^{(k)}}{1 + \sin x^{(k)}}$$

Apply this iteratively, we obtain $x^{(0)} = 0.5$, $x^{(1)} = 0.7552$, $x^{(2)} = 0.7391$, $x^{(3)} = 0.7390$, $x^{(4)} = 0.7390$ and terminate.

Note that $f''(x^{(4)}) = 1.673 > 0$, so we expect $x^{(4)}$ to be a strict local minimizer.

Newton's method can also be used for **root-finding**: if $g(x) = f'(x)$ for all x, then finding the roots of g is equivalent to finding the critical points of f.

So for root-finding, Newton's method iterates

$$x^{(k+1)} = x^{(k)} - \frac{g(x^{(k)})}{g'(x^{(k)})}$$

and $x^{(k)} \to x^*$ (under conditions) such that $g(x^*) = 0$.

A variation of the Newton's method is called the **secant method**, where $f''(x^{(k)})$ is replaced by $\frac{f'(x^{(k)})-f'(x^{(k-1)})}{x^{(k)}-x^{(k-1)}}$:

$$x^{(k+1)} = x^{(k)} - \frac{(x^{(k)} - x^{(k-1)})f'(x^{(k)})}{f'(x^{(k)}) - f'(x^{(k-1)})}$$
$$= \frac{x^{(k-1)}f'(x^{(k)}) - x^{(k)}f'(x^{(k-1)})}{f'(x^{(k)}) - f'(x^{(k-1)})}$$

Need initials $x^{(0)}$ and $x^{(1)}$ to start.

The secant method does not need computations of $f''(x)$ at the expense of more iterations.

Consider a general optimization problem

$$\underset{\boldsymbol{x}\in\Omega}{\text{minimize}}\ f(\boldsymbol{x})$$

If we have $\boldsymbol{x}^{(k)}$, then we want to find $\boldsymbol{d}^{(k)}$ (descent direction) and α_k (step size) to obtain

$$\boldsymbol{x}^{(k+1)} = \boldsymbol{x}^{(k)} + \alpha_k \boldsymbol{d}^{(k)}$$

such that $f(\boldsymbol{x}^{(k+1)})$ reduces from $f(\boldsymbol{x}^{(k)})$ as much as possible.

- How to find $\boldsymbol{d}^{(k)}$? Often times $\boldsymbol{d}^{(k)} = -\nabla f(\boldsymbol{x}^{(k)})$ but better choices may exist.

- Given $\boldsymbol{d}^{(k)}$, how to find α_k? This is an optimization with $\alpha_k \in \mathbb{R}_+$:

$$\underset{\alpha>0}{\text{minimize}}\, \phi(\alpha) := f(\boldsymbol{x}^{(k)} + \alpha \boldsymbol{d}^{(k)})$$

  which is called **line search**.

- We need $\phi'(\alpha) = \nabla f(\boldsymbol{x}^{(k)} + \alpha \boldsymbol{d}^{(k)})^\top \boldsymbol{d}^{(k)}$ and probably also $\phi''(\alpha) = (\boldsymbol{d}^{(k)})^\top \nabla^2 f(\boldsymbol{x}^{(k)} + \alpha \boldsymbol{d}^{(k)}) \boldsymbol{d}^{(k)}$. This can be computationally expensive.

- We may not need to find the best α every step. Sometimes we should allocate more cost to find a better $\boldsymbol{d}^{(k)}$.

These issues will be considered in depth later in the class.