Optimality Conditions

SIE 449/549: Optimization for Machine Learning

Afrooz Jalilzadeh

Department of Systems and Industrial Engineering
University of Arizona

Unconstrained Optimization

- An unconstrained optimization problem is of the form

$$\begin{aligned} &\min \ f(x) \\ &s.t. \ \ x \in \mathbb{R}^n, \end{aligned} \tag{Opt}$$

  where $x \in \mathbb{R}^n$ is the **variable** and $f : \mathbb{R}^n \to \mathbb{R}$ is the **objective function**

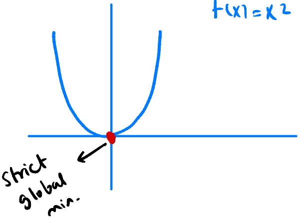
- x^* is the **solution** (or **global minimum**) of (Opt) if it satisfies

$$f(x^*) \leq f(x), \qquad \forall x \in \mathbb{R}^n$$

- x^* is the **strict global minimum** of (Opt) if it satisfies

$$f(x^*) < f(x), \qquad \forall x \in \mathbb{R}^n$$

$f(x) = x^2$

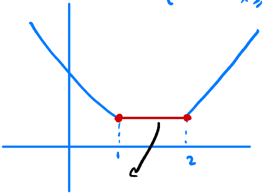


$\nabla f(x) = 2x$

$\nabla f(0) = 0$ ✓

$\nabla^2 f(x) = 2 > 0$

$f(x) = \begin{cases} -x+2 & x \leq 1 \\ 1 & 1 \leq x \leq 2 \\ x-1 & x \geq 2 \end{cases}$



all points in the line seg. are global min but not strict

Unconstrained Optimization

- Consider the following unconstrained optimization problem

$$\begin{aligned} &\min\ f(x) \\ &s.t.\ \ x \in \mathbb{R}^n, \end{aligned} \tag{Opt}$$

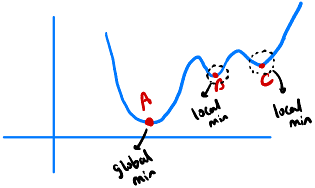
- it might not be easy to find a global minimum at all. However, even finding a point x^* such that

$$\exists \epsilon > 0 \ \text{ s.t. } \ f(x^*) \le f(x), \qquad \forall \|x - x^*\| \le \epsilon$$

  may still be valuable for us. Such an x^* is called a **local minimum** of f.

- x^* is the **strict local minimum** if it satisfies

$$\exists \epsilon > 0 \ \text{ s.t. } \ f(x^*) < f(x), \qquad \forall \|x - x^*\| \le \epsilon$$



Examples

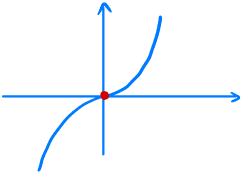
$f(x) = x^3$

$\nabla f(x) = 3x^2$

$\nabla f(0) = 0$

$\nabla^2 f(x) = 6x$

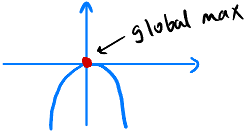
$\nabla^2 f(0) = 0$



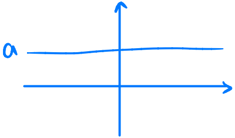
No min

No max.

$f(x) = -x^2$



$f(x) = a$



every point is a min and a max (not strict).

Optimality condition

Theorem 1 (First-order necessary condition)

Let $f : \mathbb{R}^n \to \mathbb{R}$ and suppose $x^* \in \mathbb{R}^n$ is a local minimum of f and that all the partial derivatives of f exist at x^*. Then $\nabla f(x^*) = 0$.

Theorem 2 (Second-order necessary condition)

Let $f : \mathbb{R}^n \to \mathbb{R}$ and suppose $x^* \in \mathbb{R}^n$ is a local minimum of f and f is twice continuously differentiable, then $\nabla^2 f(x^*) \succeq 0$.

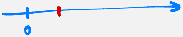
proof. $x(\alpha) = x^* + \alpha d$ $\quad x(\alpha) - x^* = \alpha d$

$$0 \leq \lim_{\alpha \to 0} \frac{f(x(\alpha)) - f(x^*)}{\|x(\alpha) - x^*\|^2} = \frac{\cancel{f(x^*)} + \overbrace{\nabla f(x^*)}^{=0}{}^T (x(\alpha) - x^*) + \frac{1}{2}(x(\alpha) - x^*)^T \nabla^2 f(x^*)(x(\alpha) - x^*) + O(\|x(\alpha) - x^*\|^3) - \cancel{f(x^*)}}{\|x(\alpha) - x^*\|^2}$$

$$= \lim_{\alpha \to 0} \frac{\frac{1}{2}\alpha^2 d^T \nabla^2 f(x^*) d}{\alpha^2 \|d\|^2}$$

$$\Longrightarrow d^T \nabla^2 f(x^*) d \geq 0 \quad \forall d \Longrightarrow \nabla^2 f(x^*) \succeq 0$$

Optimality condition



Theorem 3 (Sufficient optimality conditions)

Let $f : \mathbb{R}^n \to \mathbb{R}$ be twice continuously differentiable. Suppose x^* satisfies $\nabla f(x^*) = 0$ and $\nabla^2 f(x^*) \succ 0$. Then, x^* is a strict local minimum of (Opt).

proof. Since f is twice cont. diff. then $\exists \delta > 0$ s.t. $\nabla^2 f(y) \succ 0$ for any $y \in B(x^*, \delta)$

From mean-value theorem $\exists \eta$ between x^* and y s.t.

$$f(y) = f(x^*) + \overbrace{\langle \nabla f(x^*), y - x^* \rangle}^{=0} + \frac{1}{2} (y - x^*) \overbrace{\nabla^2 f(\eta)}^{\succ 0} (y - x^*)$$

$\Longrightarrow f(y) > f(x^*) \quad \forall y \in B(x^*, \delta)$

$\Longrightarrow x^*$ is a strict local min.

Optimality conditions

Remark 1

The optimality conditions are useful because:

- *they provide tractable conditions for optimality,*
- *they help narrow down the list of potential solutions,*
- *they are useful in the design and analysis of algorithms.*

Stationary Point and Saddle Point

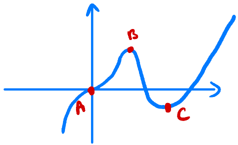
Definition 4 (Stationary Point)

min
max
saddle-point.

Let $f : \mathbb{R}^n \to \mathbb{R}$ and suppose that all the partial derivatives of f are defined at x^*. Then x^* is called a stationary point of f if $\nabla f(x^*) = 0$.

Definition 5 (Saddle Point)

Let $f : \mathbb{R}^n \to \mathbb{R}$ be a continuously differentiable function. A stationary point x^* is called a saddle point of f if it is neither a local minimum point nor a local maximum point of f.



A: saddle point
B: local max
C: local min

Theorem 6 (Sufficient Condition for a Saddle Point)

Let $f : \mathbb{R}^n \to \mathbb{R}$ be a twice continuously differentiable function and x^* is a stationary point. If $\nabla^2 f(x^*)$ is an indefinite matrix, then x^* is a saddle point.

Stationary Point and Saddle Point

Example 1

$\nabla f(x) = 0$

Consider the function $f(x) = 2x_1^3 + 3x_2^2 + 3x_1^2 x_2 - 24x_2$ *over* $\mathbb{R}^2$*. Find all the stationary points of* f *over* $\mathbb{R}^2$ *and classify them.*

$$\nabla f(x) = \begin{bmatrix} 6x_1^2 + 6x_1x_2 \\ 6x_2 + 3x_1^2 - 24 \end{bmatrix} = 0 \qquad \begin{cases} 6x_1(x_1 + x_2) = 0 \to x_1 = 0, \quad x_1 = -x_2 \\ 6x_2 + 3x_1^2 - 24 = 0 \to x_1 = 0 \quad x_2 = 4; \quad \xrightarrow{x_2 = -x_1} x_1 = 4 \text{ or } x_1 = -2 \end{cases}$$

Stationary points: $(0, 4), (4, -4), (-2, 2)$

$$\nabla^2 f(x) = \begin{bmatrix} 12x_1 + 6x_2 & 6x_1 \\ 6x_1 & 6 \end{bmatrix} \qquad \nabla^2 f(0,4) = \begin{bmatrix} 24 & 0 \\ 0 & 6 \end{bmatrix} \succ 0 \quad (0,4) \text{ is local min.}$$

$$\nabla^2 f(4,-4) = \begin{bmatrix} 24 & 24 \\ 24 & 6 \end{bmatrix} \qquad \det = (24 \times 6) - (24 \times 24) < 0 \quad \lambda_1 \cdot \lambda_2 < 0 \text{ indefinite.}$$

$(4,-4)$ is a saddle point.

$$\nabla^2 f(-2,2) = \begin{bmatrix} -12 & -12 \\ -12 & 6 \end{bmatrix} \text{ indef. so } (-2,2) \text{ is a saddle point.}$$

Global Optimality Conditions

Theorem 7 (Global Optimality Conditions)

Let $f : \mathbb{R}^n \to \mathbb{R}$ be twice continuously differentiable. Suppose that $\nabla^2 f(x) \succeq 0$ for any $x \in \mathbb{R}^n$. Let $x^* \in \mathbb{R}^n$ be a stationary point of f. Then x^* is a global minimum point of f.

Proof. mean value theorem: $\exists \eta$ between y and x^*:

$$f(y) = f(x^*) + \overbrace{\nabla f(x^*)^T}^{=0}(y-x^*) + \frac{1}{2}(y-x^*)^T \overbrace{\nabla^2 f(\eta)}^{\succeq 0}(y-x^*)$$

$$f(y) \geq f(x^*) \quad \forall y \implies x^* \text{ is a global min.}$$

Corollary: If $\nabla^2 f(x) \succ 0$ and $\nabla f(x^*) = 0 \implies x^*$ is a strict global min.

Stationary, Local minimum and Global minimum

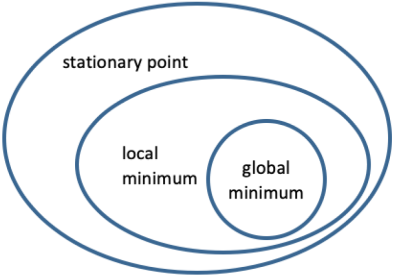


Figure 1: Stationary point, local minimum, and global minimum

Existence of optimal Solution

global min

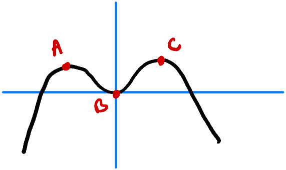
- Optimization Problem (Opt) does not always have a solution. For example:

$f(x) = x^2 - x^4$

$\nabla f(x) = 2x - 4x^3 = 2x(1-2x^2) = 0$

$\begin{cases} x = 0 \\ x = 1/\sqrt{2} \\ x = -1/\sqrt{2} \end{cases}$

$\nabla^2 f(x) = 2 - 12x^3$



- However, there are conditions which guarantee the existence of solution

Theorem 8 (Weierstrass)

closed and bounded.

Let f be a continuous function defined over a nonempty compact set $C \subseteq R^n$. Then there exists a global minimum point of f over C and a global maximum point of f over C.

- When the underlying set is not compact, Weierstrass theorem does not guarantee the existence of the solution, but certain properties of the function f can imply existence of the solution.

Example

Example 2

Consider the two-dimensional linear function $f(x, y) = x + y$ defined over the unit ball $B[0, 1] = \{(x, y)^T \mid x^2 + y^2 = 1\}$.

Because $B[0,1]$ is a compact set we have global min and global max.

Cauchy-Schwartz

$$x+y = \begin{pmatrix} x & y \end{pmatrix} \begin{pmatrix} 1 \\ 1 \end{pmatrix} \overset{\uparrow}{\leq} \sqrt{x^2+y^2}\,\sqrt{2} = \sqrt{2}$$

$\left(\frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}}\right)$ is a max point.

$\left(-1/\sqrt{2}, -1/\sqrt{2}\right)$ is a min point.

Coercive Function

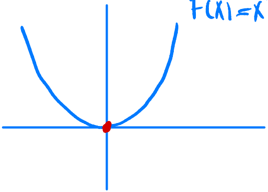
Definition 9 (Coercivity)

Let $f : \mathbb{R}^n \to \mathbb{R}$ be a continuous function. f is called **coercive** if

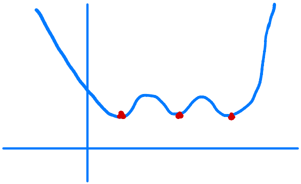
$$\lim_{\|x\| \to \infty} f(x) = \infty$$

Theorem 10

Let $f : \mathbb{R}^n \to \mathbb{R}$ be a continuous and coercive function. Then global minimum exists.



$f(x) = x^2$



$x^2 + y^2$ is Coercive, e^{x^2} Coercive, ...

Quadratic Functions

- Consider a quadratic function over $\mathbb{R}^n$ of the form $f(x) = \frac{1}{2}x^T Ax + b^T x + c$ where $A \in \mathbb{R}^{n\times n}$, is symmetric, $b \in \mathbb{R}^n$ and $c \in \mathbb{R}$.

$$\nabla f(x) = Ax + b \qquad \nabla^2 f(x) = A$$

Lemma 11

Let $f(x) = \frac{1}{2}x^T Ax + b^T x + c$ and $A \in \mathbb{R}^{n\times n}$, is symmetric, $b \in \mathbb{R}^n$ and $c \in \mathbb{R}$.

1. *x is a stationary point of f iff $Ax = -b$.*
2. *if $A \succeq 0$, then x is a global minimum point of f iff $Ax = -b$.*
3. *if $A \succ 0$, then $x = -A^{-1}b$ is a strict global minimum point of f.*

1. $\nabla f(x) = 0 \qquad Ax + b = 0 \qquad Ax = -b$

2. $A \succeq 0 \implies \nabla^2 f(x) = A \succeq 0 \iff Ax = -b$ is a global min.

Quadratic Functions

- Note that when $A \succ 0$, the global minimizer of f is $x^* = -A^{-1}b$, and consequently the minimal value of the function is:

$$f(x^*) = \frac{1}{2}(x^*)^T A x^* + b^T x^* + c$$

$$= \frac{1}{2}(-A^{-1}b)^T A(-A^{-1}b) - b^T A^{-1} b + c$$

$$= -\frac{1}{2} b^T A^{-1} b + c$$

- If A is not symmetric in $f(x) = \frac{1}{2}x^T A x + b^T x + c$, then one can consider $f(x) = \frac{1}{2}x^T \left(\frac{A+A^T}{2}\right) x + b^T x + c$, because $x^T A x = x^T \left(\frac{A+A^T}{2}\right) x$ and $\frac{A+A^T}{2}$ is a symmetric matrix.

$$x^T A x \stackrel{?}{=} x^T\left(\frac{A+A^T}{2}\right) x = \frac{1}{2}\left(x^T A x + \overbrace{x^T A^T x}^{\text{scalar}}\right) = \frac{1}{2}\left(x^T A x + x^T A x\right) = x^T A x$$