Unconstrained Optimization

Defn: Let $S \subseteq \mathbb{R}^n$, $f: S \to \mathbb{R}$.

$x^* \in S$ is a (global) minimum point (global minimizer) if $f(x) \geq f(x^*)$ $\forall x \in S$.

(global) maximum point (global maximizer) if $f(x) \leq f(x^*)$ $\forall x \in S$.

strict (global) min. point if $f(x) > f(x^*)$ $\forall x \in S \setminus \{x^*\}$

strict (global) max point if $f(x) < f(x^*)$ $\forall x \in S \setminus \{x^*\}$.

Considering a minimization problem: $\left(\underset{x \in S}{\text{minimize}}\ f(x)\right)$ the value of the problem is the infimum:

$\inf\{f(x) \mid x \in S\}$ - greatest lower bound of $\{f(x) \mid x \in S\}$
- may or may not be attained.

- On the other hand, $\min\{f(x) \mid x \in S\}$ - minimum of $\{f(x) \mid x \in S\}$
  - has to be attained
  - may not always be well-defined.

e.g. let $f(x) = x$, $S = (0,1) \subseteq \mathbb{R}$, then $\min\{f(x) \mid x \in S\} = \min(0,1)$ is not well-defined as $0 \notin (0,1)$. But $\inf(0,1) = 0$.

Similarly for a maximization problem, the value is the supremum:

$\sup\{f(x) \mid x \in S\}$ - smallest upper bound of $\{f(x) \mid x \in S\}$.

⊛ In the book, always min / max notation is used - even if the min / max may not be attained.

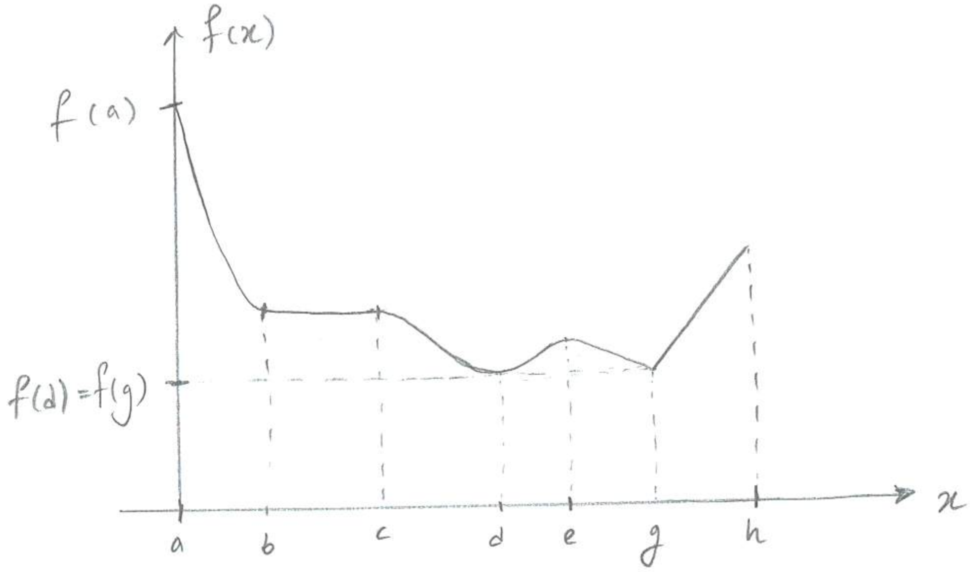
Defn: $S \subseteq \mathbb{R}^n$, $f: S \to \mathbb{R}$.

$x^* \in S$ is a local minimum point if $\exists\, r > 0$ s.t. $f(x^*) \leq f(x)$ $\forall x \in S \cap B(x^*, r)$

local maximum point " $f(x^*) \geq f(x)$ "

strict local minimum point " $f(x^*) < f(x)$ $\forall x \in S \cap B(x^*, r)$, $x \neq x^*$.

strict local maximum point " $f(x^*) > f(x)$ " .



Consider the function above defined on $S = [a, h]$.

a : strict global maximum point (also strict local max of course)

(b, c) : any point in this interval is non-strict local minimum & maximum point (not global max / min.)

d, g : strict local minimum points and (non-strict) global minimum points.

e, h : strict local maximum point

First Order Optimality Condition

<u>Fermat's Theorem:</u> $f : \mathbb{R} \to \mathbb{R}$, differentiable. If x^* is a local min. or max. then $f'(x^*) = 0$ holds.

<u>Theorem:</u> $U \subseteq \mathbb{R}^n$, $f : U \to \mathbb{R}$, differentiable. If $x^* \in \text{int}\, U$ is a local minimum / maximum point of f, then $\nabla f(x^*) = 0$.

Clearly, this is a "necessary" optimality condition, but not a sufficient one.

Consider $f(x) = x^3$. Clearly $x = 0$ is not a local max or min. $(f'(0) = 0)$

Hence, this result in not very practical by its own. If the existence of an (global) optimal solution is known, then it becomes practical.

<u>Weierstrass Thm:</u> let $C \subseteq \mathbb{R}^n$ be a nonempty compact set and f be continuous on C. Then, there exists a global min. & a global max. of f over C.

<u>Recall:</u> (C: compact) means (C: closed & bounded.) This may be too restrictive for many examples.

<u>Defn:</u> $f : \mathbb{R}^n \to \mathbb{R}$, continuous. f is called <u>coercive</u> (for minimization) if $\lim_{\|x\| \to \infty} f(x) = \infty$ holds. $\left(\text{for } \underline{\text{max}}: \lim_{\|x\| \to \infty} f(x) = -\infty \right)$

Theorem: let $f: \mathbb{R}^n \to \mathbb{R}$ be continuous and coercive, let $S \subseteq \mathbb{R}^n$ be a non-empty closed set. Then, f has a global minimum point over S.

- Note that $S = \mathbb{R}^n$ is also closed (also open).

Example: $f(x) = 3x^4 - 20x^3 + 42x^2 - 36x$.

Example: $f : \mathbb{R}^2 \to \mathbb{R}$, $f(x) = 3x_1^2 - x_1 x_2 + x_2^2 - 6x_1 + 2x_2 + 11$.

Classification of Matrices

Defn: Let $A \in \mathbb{R}^{n\times n}$ be a symmetric matrix. A is positive semidefinite $(A \succeq 0)$ if $x^TAx \geq 0$ for all $x \in \mathbb{R}^n$. A is positive definite $(A \succ 0)$ if $x^TAx > 0$ for all $x \in \mathbb{R}^n \setminus \{0\}$.

- Similarly, A is negative semidefinite if $x^TAx \leq 0$ for all $x \in \mathbb{R}^n$; $(A \preceq 0)$
  negative definite if $x^TAx < 0$ for all $x \in \mathbb{R}^n \setminus \{0\}$. $(A \prec 0)$
- A is indefinite if $\exists\, x, y \in \mathbb{R}^n$ s.t. $x^TAx > 0$ and $y^TAy < 0$.

A useful observation: If A is p.d. $(A \succ 0)$, then $a_{ii} > 0$ for all $i = 1, \dots, n$.

Note that $e_i^TAe_i = a_{ii} > 0$ holds. ✓

Similarly, if $A \succeq 0$, then $a_{ii} \geq 0$ $\forall i$.
if $A \preceq 0$, then $a_{ii} \leq 0$ $\forall i$
if $A \prec 0$, then $a_{i} < 0$ $\forall i$.

(The reverse of these implications are not true in general.)

By the above observations, we have the following simple result:
If $a_{ii} > 0$, $a_{jj} < 0$ for some $i, j \in \{1, \dots, n\}$, then A is an indefinite matrix.

It's, in general, difficult to check if a matrix is psd, pd, nsd, nd ....

Theorem: Let $A \in \mathbb{R}^{n\times n}$ be symmetric.

A is positive definite if and only if all its eigenvalues are positive.
" pos. semidefinite " " " " " " " non-negative.
negative definite " " " " " " " negative.
neg. semidefinite " " " " " " " non-positive.
indefinite " " " " there exists at least one positive and one negative eigenvalues.

Note that if one has a diagonal matrix $D = \begin{bmatrix} d_1 & & 0 \\ & \ddots & \\ 0 & & d_n \end{bmatrix}$, then the eigenvalues are simply the diagonal elements. Hence, the above theorem can be applied to the diagonal components directly.

<u>A simple observation for 2×2 matrices:</u> Let $A \in \mathbb{R}^{2\times 2}$, λ_1, λ_2 be the eigenvalues.

Recall that $\det A = \lambda_1 \cdot \lambda_2$ and $\operatorname{tr} A = \lambda_1 + \lambda_2$.

If $\det A \geq 0$, $\operatorname{tr} A \geq 0$, then $\lambda_1 \lambda_2 \geq 0$, $\lambda_1 + \lambda_2 \geq 0$ $\Rightarrow$ $\lambda_1, \lambda_2 \geq 0$ $\Rightarrow A$: p.s.d.

$\det A > 0$, $\operatorname{tr} A > 0$, " $\lambda_1 \lambda_2 > 0$, $\lambda_1 + \lambda_2 > 0$ $\Rightarrow$ $\lambda_1, \lambda_2 > 0$ $\Rightarrow A$: p.d.

$\det A < 0 \Rightarrow \lambda_1 \lambda_2 < 0 \Rightarrow \lambda_i < 0, \lambda_j > 0 \Rightarrow A$: indefinite.

$\det A > 0$, $\operatorname{tr} A < 0 \Rightarrow \lambda_1, \lambda_2 < 0 \Rightarrow A$: n.d.

$\vdots$

<u>Example:</u>

Second Order Optimality Conditions

Theorem (necessary cond.) $U \subseteq \mathbb{R}^n$ open, $f: U \to \mathbb{R}$ twice contsly differentiable.

Let $x^* \in U$ be a stationary point. Then,

- If x^* is a local min. point, then $\nabla^2 f(x^*) \succeq 0$ (psd).
- If x^* is a local max. point, then $\nabla^2 f(x^*) \preceq 0$ (psd.)

Proof: Let's show the first result, the second one also follows.

Let x^* be a local min. point. Then, there exists $r > 0$ s.t. $B(x^*, r) \subseteq U$

and $f(x) \geq f(x^*)$ $\forall x \in B(x^*, r)$. Let $d \in \mathbb{R}^n$, $d \neq 0$ and consider

$x^*_\alpha = x^* + \alpha d$. Note that for $0 < \alpha < \frac{r}{\|d\|}$, $x^*_\alpha \in B(x^*, r)$. Hence,

$f(x^*_\alpha) \geq f(x^*)$.

By linear approximation theorem, there exists $z_\alpha \in [x^*, x^*_\alpha]$ s.t.

$$\underbrace{f(x^*_\alpha) - f(x^*)}_{\geq 0} = \underbrace{\nabla f(x^*)^T}_{\substack{0 \\ \text{(stat. point)}}} (x^*_\alpha - x^*) + \frac{1}{2} \underbrace{(x^*_\alpha - x^*)^T}_{\alpha d} \nabla^2 f(z_\alpha) \underbrace{(x^*_\alpha - x^*)}_{\alpha d}.$$

$\Rightarrow \forall \alpha \in (0, \frac{r}{\|d\|})$, we obtain $d^T \nabla^2 f(z_\alpha) d \geq 0$.

Let's take limit as $\alpha \downarrow 0$: $d^T \nabla^2 f(x^*) d \geq 0$ (as $z_\alpha \to x^*$)

Since this is true for any $d \in \mathbb{R}^n$, $d \neq 0$, we have $\nabla^2 f(x^*) \succeq 0$.

∎

Theorem (suff. cond.) $U \subseteq \mathbb{R}^n$ open, $f: U \to \mathbb{R}$ twice cont'sly diff.ble

let $x^* \in U$ be a stationary point. Then,

- If $\nabla^2 f(x^*) \succ 0$, then x^* is a strict local min. point.
- If $\nabla^2 f(x^*) \prec 0$, then x^* is a strict local max. point.

Proof: Let's prove the first part. Let x^* satisfy $\nabla^2 f(x^*) \succ 0$.

Since Hessian is continuous, there exists $r > 0$ s.t. $B(x^*, r) \subseteq U$ and

$\nabla^2 f(x^*) \succ 0$ for all $x \in B(x^*, r)$.

By linear approx. theorem, for all $x \in B(x^*, r)$ there exist $z_x \in \underbrace{[x^*, x]}_{\subseteq B(x^*, r)}$ s.t.

$$f(x) - f(x^*) = \frac{1}{2}(x - x^*)^T \underbrace{\nabla^2 f(z_x)}_{\succ 0}(x - x^*) > 0 \quad (\text{for } x \neq x^*).$$

Hence, $f(x) > f(x^*)$ $\forall x \in B(x^*, r), x \neq x^*$.

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

Defn: $U \subseteq \mathbb{R}^n$ open, $f: U \to \mathbb{R}$ cont'ly differentiable. A stationary point x^* is called a <u>saddle point</u> of f over U if it's neither a local min. nor a local max. point.

Result: If $\nabla^2 f(x^*)$ is an indefinite matrix (for a stationary point x^*), then x^* is a saddle point.

Why? By necess. cond. thm, $\left.\begin{array}{l} \nabla^2 f(x^*) \not\succeq 0 \Rightarrow \text{not a local min.} \\ \nabla^2 f(x^*) \not\preceq 0 \Rightarrow \text{'' '' '' max.} \end{array}\right\}$ saddle point.

Theorem (global opt. cond) $f : \mathbb{R}^n \to \mathbb{R}$, twice continuously differentiable. Suppose that $\nabla^2 f(x) \succeq 0 \;\; \forall x \in \mathbb{R}^n$. If x^* is a stationary point of f, then it's a global minimum point of f.

Proof: By linear approximation theorem, for any $x \in \mathbb{R}^n$, there exists a vector $z_x \in [x^*, x]$ s.t.

$$f(x) - f(x^*) = \frac{1}{2}(x - x^*)^T \underbrace{\nabla f(z_x)}_{\succeq 0}(x - x^*) \geq 0 .$$

Hence $f(x) \geq f(x^*)$ for all $x \in \mathbb{R}^n$. Thus, x^* is a global min. point. ∎

(Similarly if $\nabla^2 f(x) \succ 0$, then x^* is strict global minimum)

Ex: Recall the example: $f(x) = 3x_1^2 - x_1 x_2 + x_2^2 - 6x_1 + 2x_2 + 11$.

Application: quadratic functions

Consider a function of the form: $f(x) = x^T A x + 2b^T x + c$

where $A \in \mathbb{R}^{n \times n}$ is a symmetric matrix, $b \in \mathbb{R}^n$, $c \in \mathbb{R}$.

e.g. Recall $f(x) = 3x_1^2 - x_1 x_2 + x_2^2 - 6x_1 + 2x_2 + 11$

Let $A = \begin{bmatrix} 3 & -1/2 \\ -1/2 & 1 \end{bmatrix}$, $b = \begin{bmatrix} -3 \\ 1 \end{bmatrix}$, $c = 11$.

$$[x_1 \ x_2] A \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} + 2b^T \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} + c = \left[3x_1 - \frac{1}{2}x_2, -\frac{1}{2}x_1 + x_2\right] \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} + 2b^T x + c$$

$$= 3x_1^2 - x_1 x_2 + x_2^2 - 6x_1 + 2x_2 + 11.$$

Gradient of a qudratic function: $\nabla f(x) = 2Ax + 2b$

Hessian of a " " : $\nabla^2 f(x) = 2A$.

Then, stationary points are solutions to $(Ax = -b)$. If A is invertible, then this has a unique solution $x^* = -A^{-1}b$. There may be infinitely many solutions or no solution...

If $A \succeq 0$, then any stationary point is a global minimum. ($\preceq \Rightarrow$ max.)

If $A \succ 0$, then A is invertible. Hence, there's a unique stationary point which is a strict global minimum point.