

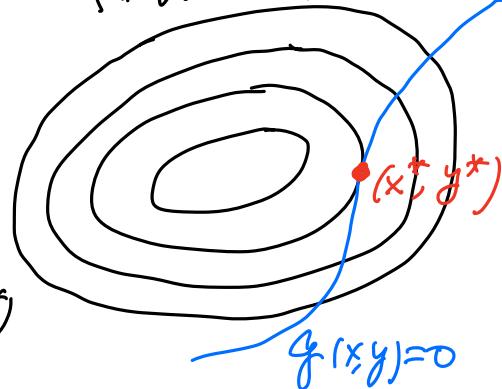
Lecture 5. 1/20/2023

- ① Finish proof of a convergence theorem for gradient descent. (See Lecture 4.)
- ② Methods of Lagrange multipliers / Penalty methods for constrained optimization

$$f(x, y) = \text{const.}$$

The method of Lagrange multipliers

Extremize $f(x, y)$ subject to $g(x, y) = 0$, $x, y \in \mathbb{R}$.



- ① At an extreme point (x^*, y^*) the curve $g(x, y) = 0$ and the level curve $f = f(x^*, y^*)$ are tangent to each other.

- ② Gradient \perp level surfaces: $\nabla f(x^*, y^*) = \lambda^* \nabla g(x^*, y^*)$, $\lambda^* \in \mathbb{R}$.

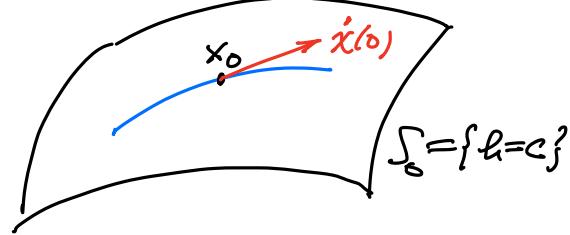
Lemma Let $E \subseteq \mathbb{R}^n$ be open, $h \in C^1(E)$, $x_0 \in E$, $c = h(x_0)$. and $S_0 = \{x \in E : h(x) = c\}$. If $x: (-\varepsilon, \varepsilon) \rightarrow S_0$ for some $\varepsilon > 0$ is a C^1 -curve on S_0 such that $x(0) = x_0$, then $\dot{x}(0) \cdot \nabla h(x_0) = 0$.

Notation: $\dot{x} = \frac{dx}{dt}$.

Proof $h(x(t)) = c \quad \forall t \in (-\varepsilon, \varepsilon)$. So, by the chain rule:

$$\frac{d}{dt}(h(x(t))) = \nabla h(x(t)) \cdot \dot{x}(t) = 0$$

Let $t=0$ and notice that $x(0) = x_0$. We get $\nabla h(x_0) \cdot \dot{x}(0) = 0$. QED



The method of Lagrange multipliers (for the above set up)

Define

$$L(x, y, \lambda) = f(x, y) - \lambda g(x, y).$$

Find a critical point (x^*, y^*, λ^*) of L .

$$\begin{cases} \nabla L(x^*, y^*, \lambda^*) = \nabla f(x^*, y^*) - \lambda^* \nabla g(x^*, y^*) = 0 \\ \partial_\lambda L(x^*, y^*, \lambda^*) = g(x^*, y^*) = 0 \end{cases}$$

① $L = L(x, y, \lambda)$: the Lagrange multipliers function, or the Lagrangian

② λ : a multiplier

The case that f, g are $n (\geq 2)$ variable functions is similar.

Example Let $a_i \in \mathbb{R} (i=1, \dots, n)$ and $\sigma > 0$. Maximize

$$f(x) = \sum_{i=1}^n a_i x_i \text{ subject to } \sum_{i=1}^n x_i^2 = \sigma^2.$$

Solution Let $L(x, \lambda) = \sum_{i=1}^n a_i x_i - \lambda (\sum_{i=1}^n x_i^2 - \sigma^2)$.

[2]

$$\partial_{x_k} L(x, \lambda) = a_k - 2\lambda x_k = 0 \Rightarrow x_k = \frac{a_k}{2\lambda} \quad (k=1, \dots, n).$$

$$\partial_\lambda L(x, \lambda) = \sum_{i=1}^n x_i^2 - \sigma^2 = 0 \Rightarrow \sum_{i=1}^n x_i^2 = \sigma^2.$$

$$\sum_{i=1}^n \left(\frac{a_i}{2\lambda}\right)^2 = \sigma^2 \Rightarrow \|a\|^2 \left(\sum_{i=1}^n a_i^2\right) = 4\lambda^2 \sigma^2 \Rightarrow \lambda = \pm \frac{\|a\|}{2\sigma}.$$

$$\boxed{\hat{x}_k = \frac{a_k}{2\lambda} = \pm \frac{\|a\|}{\|a\| + \sigma} \sigma \quad (k=1, 2, \dots, n).} \quad \boxed{f(\vec{x}) = \sum_{i=1}^n \frac{\sigma a_i^2}{\|a\| + \sigma} = \sigma \|a\|.}$$

Let $b = (b_1, \dots, b_n) \in \mathbb{R}^n$. Set $x_i = \frac{b_i}{\|b\|} \sigma$, $\|b\|^2 = \sum_{i=1}^n b_i^2$. Then,

$$\sum_{i=1}^n x_i^2 = \sigma^2. \text{ Thus, } f(x) \leq f(\vec{x}) \Rightarrow \sum_{i=1}^n \frac{a_i b_i}{\|b\|} \sigma \leq \sigma \|a\| \Rightarrow \sum_{i=1}^n a_i b_i \leq \|a\| \|b\|.$$

Replace b_i by $b_i \operatorname{sgn}(\sum_{k=1}^n a_k b_k)$ $\Rightarrow |a \cdot b| \leq \|a\| \|b\|$. Cauchy-Schwarz!

General set up Extremize $f(x)$ subject to $g(x) = 0$.

$f: E \subseteq \mathbb{R}^n \rightarrow \mathbb{R}$, $g: E \rightarrow \mathbb{R}^m$ ($m < n$). E : open, $g(x) = \begin{bmatrix} g_1(x) \\ \vdots \\ g_m(x) \end{bmatrix}$, $x = \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix}$

Define

$$\mathcal{L}(x, \lambda) = f(x) - \lambda \cdot g(x) = f(x) - \lambda_1 g_1(x) - \dots - \lambda_m g_m(x).$$

$$\partial_x \mathcal{L} = 0, \partial_\lambda \mathcal{L} = 0. \Rightarrow \text{Solve for } x, \lambda = (\lambda_1, \dots, \lambda_m).$$

$$\left\{ \begin{array}{l} \partial_{x_i} \mathcal{L} = \partial_x f - \lambda_1 \partial_{x_i} g_1 - \dots - \lambda_m \partial_{x_i} g_m = 0 \\ \partial_{x_n} \mathcal{L} = \partial_x f - \lambda_1 \partial_{x_n} g_1 - \dots - \lambda_m \partial_{x_n} g_m = 0 \end{array} \right. \left\{ \begin{array}{l} \nabla f(x)^T + \nabla g(x)^T \lambda = 0 \\ g(x) = 0 \end{array} \right.$$

$$\left\{ \begin{array}{l} \partial_{x_i} \mathcal{L} = g_i(x) = 0, \quad i=1, \dots, m \\ \partial_{x_n} \mathcal{L} = \partial_x f - \lambda_1 \partial_{x_n} g_1 - \dots - \lambda_m \partial_{x_n} g_m = 0 \end{array} \right. \left\{ \begin{array}{l} \text{consistent} \\ \text{with} \\ \nabla h(x) = \begin{bmatrix} \frac{\partial h_i}{\partial x_j} \end{bmatrix} \\ \text{for } h: \mathbb{R}^n \rightarrow \mathbb{R}^m \end{array} \right.$$

Notation: $\nabla f(x) = [\partial_{x_1} f, \dots, \partial_{x_n} f]^T$. So, $\nabla f(x)^T = \begin{bmatrix} \partial_x f(x) \\ \vdots \\ \partial_x f(x) \end{bmatrix} \in \mathbb{R}^n$.

Definition Call $x^* \in E$ a constrained local min. (or max) point, if $g(x^*) = 0$ and $f(x^*) \leq f(x)$ (or $f(x^*) \geq f(x)$) $\forall x \in B(x^*, \delta)$ with $g(x) = 0$ for some $\delta > 0$ with $B(x^*, \delta) \subseteq E$.

Constrained extreme point constrained max. or min. point.

Theorem Let $E \subseteq \mathbb{R}^n$ be open, $f \in C^1(E)$, and $g \in C^1(E, \mathbb{R}^m)$ with $n > m$.

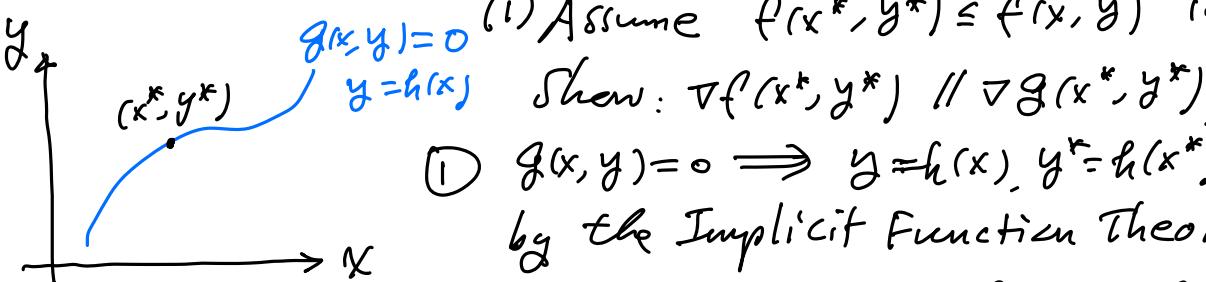
(1) If $x^* \in E$ is a constrained local extreme point of f subject to $g = 0$ and $\operatorname{rank}(\nabla g(x^*)) = m$, then there exist $\lambda^* = (\lambda_1^*, \dots, \lambda_m^*)^T \in \mathbb{R}^m$ such that (x^*, λ^*) is a critical point of $L(x, \lambda) = f(x) - \lambda \cdot g(x)$, $x \in E$.

(2) Conversely, assume $(x^*, \lambda^*) \in E \times \mathbb{R}^m$ is a critical point of L , $f \in C^2(B(x^*, \delta))$ and $g \in C^2(B(x^*, \delta), \mathbb{R}^m)$ for some $\delta > 0$ with $B(x^*, \delta) \subseteq E$,

rank $(\nabla g(x^*)) = m$, and $\tilde{x}^\top H_L(x^*, \lambda^*) \tilde{x} > 0$ for any $\tilde{x} \in \mathbb{R}^n$, $\tilde{x} \neq 0$, with [3]
 $\nabla g(x^*) \tilde{x} = 0$, where $H_L(x, \lambda) = [\frac{\partial^2 L}{\partial x_i \partial x_j}]$. Then, x^* is a strict local min pt. subject to $g=0$.

Ideas of proof. Special and simple case $n=2$, $m=1$.

y (1) Assume $f(x^*, y^*) \leq f(x, y)$ if $g(x, y) = 0$.



① $g(x, y) = 0 \Rightarrow y = h(x)$, $y^* = h(x^*)$. $g(x, h(x)) = 0, \forall x$
 by the Implicit Function Theorem (IFT).

② $f(x^*, y^*) \leq f(x, y)$ if $g(x, y) = 0 \Rightarrow f(x^*, h(x^*)) \leq f(x, h(x)) \quad \forall x$.

So, x^* is a local extremum point of $\varphi(x) \triangleq f(x, h(x)) \Rightarrow \varphi'(x^*) = 0$.

$$\varphi'(x) = \partial_x f(x^*, h(x^*)) + \partial_y f(x^*, h(x^*)) h'(x^*) = 0. \quad \underline{\partial_x f(x^*, y^*) + \partial_y f(x^*, y^*) h'(x^*) = 0}$$

③ $g(x, h(x)) = 0 \Rightarrow \partial_x g(x^*, h(x^*)) + \partial_y g(x^*, h(x^*)) h'(x^*) = 0$ (I)

$$\Rightarrow \underline{\partial_x g(x^*, y^*) + \partial_y g(x^*, y^*) h'(x^*) = 0}. \quad (I) + (II) \Rightarrow \frac{\partial_x f}{\partial_x g} = \frac{\partial_y f}{\partial_y g} \text{ at } x^*$$

(2) Need to show: the critical point (x^*, y^*, λ^*) of $L(x, y, \lambda)$ satisfies $g(x^*, y^*) = 0$, $f(x^*, y^*) \leq f(x, y)$ if $g(x, y) = 0$ and (x, y) is near (x^*, y^*) .

Since (x^*, y^*, λ^*) is a critical point of $L(x, y, \lambda) = f(x, y) - \lambda g(x, y)$,
 $\partial_\lambda L(x^*, y^*, \lambda^*) = -g(x^*, y^*) = 0$. Moreover, $\partial_x L = \partial_y L = \partial_\lambda L = 0$ at (x^*, y^*, λ^*)
 IFT $\Rightarrow y = h(x)$, $y^* = h(x^*)$. $\varphi(x) \triangleq f(x, y) = f(x, h(x))$. Show x^* is a local min. pt. of φ . $\varphi' = \partial_x f + \partial_y f h'$.

$$\varphi'(x^*) = \partial_x f(x^*, h(x^*)) + \partial_y f(x^*, h(x^*)) h'(x^*)$$

But $g(x, h(x)) = 0$. So, (note: $y^* = h(x^*)$)

$$\partial_x g(x^*, h(x^*)) + \partial_y g(x^*, h(x^*)) h'(x^*) = 0 \quad (\star)$$

Thus,

$$\begin{aligned} \varphi'(x^*) &= \partial_x f(x^*, y^*) - \lambda^* \partial_x g(x^*, y^*) + [\partial_y f(x^*, y^*) - \lambda^* \partial_y g(x^*, y^*)] h'(x^*) \\ &= \partial_x L(x^*, y^*, \lambda^*) + \partial_y L(x^*, y^*, \lambda^*) h'(x^*) = 0. \end{aligned}$$

Now, calculate $\varphi''(x^*)$.

$$\varphi'' = \partial_{xx} f + \partial_{xy} f h' + (\partial_{xy} f + \partial_{yy} f h') h' + \partial_y f h''$$

$$\partial_{xx} g + \partial_{xy} g h' + (\partial_{xy} g + \partial_{yy} g h') h' + \partial_{yy} g h'' = 0$$

Since $\partial_y g(x^*, y^*, \lambda^*) = 0$, we have by the 2nd eq by λ^* that
 $\varphi''(x^*) = \partial_{xx} f + 2\partial_{xy} f h' + \partial_{yy} f h'^2 - \lambda^* [\partial_{xx} g + 2\partial_{xy} g h' + \partial_{yy} g h'^2] = 0$ at (x^*, y^*) .

(4)

$$\partial_{xx} L(x^*, y^*, \lambda^*) + 2\partial_{xy} L(x^*, y^*, \lambda^*) h'(x^*) + \partial_{yy} L(x^*, y^*, \lambda^*) h'(x^*)^2 = 0 \quad (\star\star)$$

$$H_L = \begin{bmatrix} \partial_{xx} L & \partial_{xy} L \\ \partial_{xy} L & \partial_{yy} L \end{bmatrix}, \quad z := \begin{bmatrix} 1 \\ h'(x^*) \end{bmatrix}. \quad \text{Then } \nabla g(x^*, y^*) \cdot z = 0 \text{ by } (\star)$$

Thus, $\varphi''(x^*) = z^T H_L(x^*, y^*, \lambda^*) z > 0$, by assumption. So, x^* is a local min point of $\varphi(x) = f(x, h(x))$. QED