

- Proof of the theorem for the method of Lagrange multipliers.
- Penalty methods

The method of Lagrange multipliers

Extremize (min. or max.) $f(x, y)$

subject to $g(x, y) = 0$.

Call (x^*, y^*) a constrained local min (or max) point, if $f(x^*, y^*) \leq f(x, y)$ ($f(x, y) : g(x, y) = 0$).

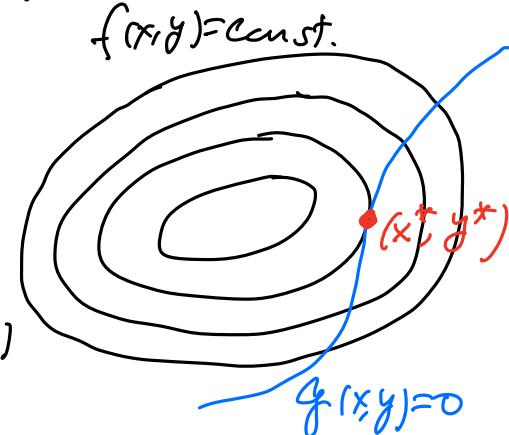
A constrained local extreme point (x^*, y^*) is a constrained local min. or max. point.

Observation At the constrained extreme point (x^*, y^*) , $\nabla f \parallel \nabla g$.

Define $L(x, y, \lambda) = f(x, y) - \lambda g(x, y)$, λ : Lagrange function

Solve $\nabla L(x, y, \lambda) = \nabla f(x, y) - \lambda \nabla g(x, y) = 0 \Rightarrow (x^*, y^*, \lambda^*)$.

$$\partial_\lambda L(x, y, \lambda) = -g(x, y) = 0$$



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

$$L(x, \lambda) = f(x) - \lambda \cdot g(x) = f(x) - \lambda_1 g_1(x) - \dots - \lambda_m g_m(x), \quad \lambda = [\lambda_1 \dots \lambda_m]^T \in \mathbb{R}^m.$$

$\nabla_x L = 0$, $\nabla_\lambda L = 0$. \Rightarrow Solve for $x = (x_1, \dots, x_n)$ and $\lambda = (\lambda_1, \dots, \lambda_m)$

$$\begin{cases} \partial_{x_i} L = \partial_{x_i} f - \lambda_1 \partial_{x_i} g_1 - \dots - \lambda_m \partial_{x_i} g_m \quad (i=1, \dots, n) \\ \partial_{\lambda_j} L = -g_j \quad (j=1, \dots, m). \end{cases} \quad \begin{cases} \nabla f(x)^T - \nabla g(x)^T \lambda = 0 \\ g(x) = 0 \end{cases}$$

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 $\text{rank } (\nabla g(x^*)) = m$, 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 , [2] $f \in C^2(B(x^*, \delta))$ and $g \in C^2(B(x^*, \delta), \mathbb{R}^m)$ for some $\delta > 0$ with $B(x^*, \delta) \subseteq E$, $\text{rank } (\nabla g(x^*)) = m$, and $\tilde{x}^T H_L(x^*, \lambda^*) \tilde{x} > 0$ for any $\tilde{x} \in \mathbb{R}^n$, $\tilde{x} \neq 0$, with $\nabla g(x^*) \tilde{x} = 0$, 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 $g=0$.

(May skip the proof for the general case in class.)

Proof (1) Let $g = [\begin{matrix} g_1 \\ \vdots \\ g_m \end{matrix}]$. Need to show $\nabla f(x^*)^T \in \text{span}\{\nabla g_1(x^*), \dots, \nabla g_m(x^*)\}^T$. i.e., $\exists \lambda^* = [\lambda_1^*, \dots, \lambda_m^*]^T \in \mathbb{R}^m$ such that $\nabla f(x^*)^T = \sum_{j=1}^m \lambda_j^* \nabla g_j(x^*)^T = [\nabla g(x^*)]^T \lambda^*$. This implies that (x^*, λ^*) is a local critical point of $L(x, \lambda)$.

① Assume W.L.O.G. that the last m columns of ∇g , $\partial_{n-m+1} g, \dots, \partial_n g$, are (nearly) independent at x^* . By the Implicit Function Theorem, \exists open $U \subseteq \mathbb{R}^n$, open $V \subseteq \mathbb{R}^{n-m}$, and $h \in C^1(V, \mathbb{R}^m)$ s.t. $x^* = (y^*, z^*) \in U \cap \{x \in E : g(x) = 0\} \cap V = \{(y, h(y)) : y \in V\}$. i.e., $g(x) = g(y, z) = 0 \iff y = h(x), z^* = h(y^*)$. Here, for $x \in \mathbb{R}^n$, we write $x = [\begin{matrix} y \\ z \end{matrix}] \in \mathbb{R}^{n-m} \times \mathbb{R}^m$.

Notation:

$$h(y) = \begin{bmatrix} h_1(y) \\ \vdots \\ h_m(y) \end{bmatrix} \in \mathbb{R}^m, \quad y = [x_1, \dots, x_{n-m}]^T \in \mathbb{R}^{n-m}, \quad z = [x_{n-m+1}, \dots, x_n]^T = h(y) \in \mathbb{R}^m.$$

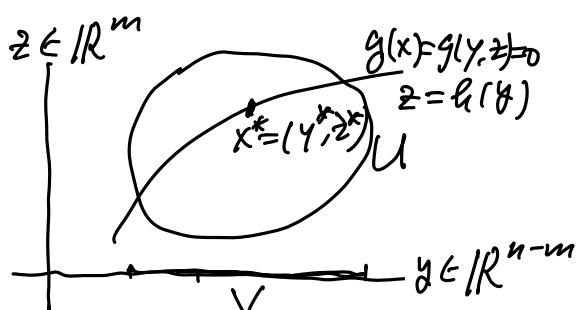
$$x = [\begin{matrix} y \\ z \end{matrix}] \in \mathbb{R}^n, \quad f(x) = f(y, z) \in \mathbb{R}^1, \quad g(x) = \begin{bmatrix} g_1(x) \\ \vdots \\ g_m(x) \end{bmatrix} \in \mathbb{R}^m.$$

$$\nabla f \in \mathbb{R}^n = \mathbb{R}^{n \times 1}, \quad \nabla_y f \in \mathbb{R}^{n-m} = \mathbb{R}^{(n-m) \times 1}, \quad \nabla_z f \in \mathbb{R}^m, \quad \nabla g \in \mathbb{R}^{m \times n}.$$

$$\nabla g \in \mathbb{R}^{m \times n}, \quad \nabla_y g \in \mathbb{R}^{m \times (n-m)}, \quad \nabla_z g \in \mathbb{R}^{m \times m}, \quad \nabla h \in \mathbb{R}^{m \times (n-m)}$$

② $x^* = (y^*, z^*)$ is a constrained local extreme pt. $\Rightarrow y^*$ is a local extreme pt. (unconstrained) for $\varphi(y) = f(y, h(y))$ (since $z = h(y)$). $\Rightarrow \nabla \varphi(y^*) = 0$. Write $f(x) = f(y, z)$, $\nabla f(x) = [\nabla_y f(x), \nabla_z f(x)]$. By the chain rule: $\nabla \varphi(y)^T = \nabla_y f(y, h(y))^T + \nabla h(y)^T \nabla_z f(y, h(y))$. [Verify it!] Hence, $\nabla \varphi(y^*) = 0 \Rightarrow \nabla_y f(x^*)^T + \nabla h(y^*)^T \nabla_z f(x^*)^T = 0$.

③ Compute $\nabla_h(y^*)$ from $g(y, h(y)) = 0$. Again by the chain rule, $\nabla_y g(y, h(y)) + \nabla_z g(y, h(y)) \nabla_h(y) = 0$. So, $\nabla_y g(x^*) + \nabla_z g(x^*) \nabla_h(y^*) = 0$.



So, $\nabla h(y^*) = -[\nabla_2 g(x^*)]^{-1} \nabla_y g(x^*)$ [Note: last m columns of ∇g are linearly indep. at x^* .] Therefore, combining the two equations. [3]

$$\nabla_y f(x^*)^T - [\nabla_y g(x^*)]^T [\nabla_2 g(x^*)]^{-T} \nabla_2 f(x^*)^T = 0$$

i.e., $\nabla_y f(x^*)^T = [\nabla_y g(x^*)]^T [(\nabla_2 g(x^*))^{-T} \nabla_2 f(x^*)] = [\nabla_y g(x^*)]^T \lambda^*$

where $\lambda^* = (\nabla_2 g(x^*))^{-T} \nabla_2 f(x^*)^T \in \mathbb{R}^{m \times 1}$

$$④ \nabla_2 f(x^*)^T = I \cdot \nabla_2 f(x^*)^T = [\nabla_2 g(x^*)]^T [\nabla_2 g(x^*)]^{-T} \nabla_2 f(x^*)^T = [\nabla_2 g(x^*)]^T \lambda^*.$$

Hence,

$$\nabla f(x^*)^T = \begin{bmatrix} \nabla_y f(x^*)^T \\ \nabla_2 f(x^*)^T \end{bmatrix} = \begin{bmatrix} [\nabla_y g(x^*)]^T \lambda^* \\ [\nabla_2 g(x^*)]^T \lambda^* \end{bmatrix} = \nabla g(x^*)^T \lambda^*.$$

(2) As in (1), we assume the last m columns of ∇g are linearly indep. at x^* . Hence, by the IFT, locally, $g(x) = 0 \Rightarrow z = h(y)$, where $x = (y, z)$, $g(y, h(y)) = 0$ and $x^* = (y^*, z^*) = (y^*, h(y^*))$. Let $\varphi(y) = f(y, h(y))$. We show that y^* is a local min pt. of φ .

Since $(x^*, \lambda^*) = (y^*, h(y^*), \lambda^*)$ is a critical pt of L , we have $\nabla L = 0$ at (x^*, λ^*) . So, $g(x^*) = g(y^*, z^*) = 0$, i.e., $z^* = h(y^*)$. By the chain rule,

$$\nabla \varphi(y)^T = \nabla_y f(y, h(y))^T + \nabla h(y)^T \nabla_2 f(y, h(y))^T.$$

Hence, $\nabla \varphi(y^*)^T = \nabla_y f(x^*)^T + \nabla h(y^*)^T \nabla_2 f(x^*)^T$. (I)

Since $g(y, h(y)) = 0$, we have by the chain rule that

$$\nabla_y g^T + \nabla h^T \nabla_2 g^T = 0. \text{ Hence,}$$

$$\nabla_y g^T(x^*) + \nabla h^T(y^*) \nabla_2 g^T(x^*) = 0.$$
 (II)

Now, (I) and (II) imply that

$$\nabla \varphi(y^*)^T = \nabla_y L(x^*, \lambda^*)^T + \nabla h^T(y^*) \nabla_2 L(x^*, \lambda^*)^T = 0$$

Since (x^*, λ^*) is a critical point of L .

Now, calculate $\nabla^2 \varphi(x^*)$. Write $h = \begin{bmatrix} h_1 \\ \vdots \\ h_m \end{bmatrix}$, $y = [y_1, \dots, y_{n-m}]^T$.

$$z = [z_1, \dots, z_m]^T. \quad \varphi(y) = f(y, h(y)) = f(y_1, \dots, y_{n-m}, h_1(y), \dots, h_m(y)).$$

$$\partial_{y_i} \varphi = \partial_{y_i} f + \sum_{k=1}^m \partial_{z_k} f \partial_{y_i} h_k, \quad i = 1, \dots, n-m.$$

$$\partial_{y_i y_j} \varphi = \partial_{y_i y_j} f + \sum_{k=1}^m \partial_{y_i z_k}^2 f \partial_{y_j} h_k$$

$$+ \sum_{k=1}^m (\partial_{y_i z_k}^2 f + \sum_{l=1}^m \partial_{z_k z_l}^2 f \partial_{y_j} h_l) \partial_{y_i} h_k + \sum_{k=1}^m \partial_{z_k} f \partial_{y_i y_j}^2 h_k$$
 (III)

$i, j = 1, 2, \dots, n.$

Now, $g(y, h(y)) = 0$. So, $\partial_\alpha g(y, h(y)) = 0$ ($\alpha = 1, \dots, m$).

$\forall \alpha \in \{1, \dots, m\}$:

$$\partial_{y_i} g_\alpha + \sum_{k=1}^m \partial_{z_k} g_\alpha \partial_{y_i} h_k = 0 \quad (i = 1, \dots, n)$$

$$\partial_{y_i y_j}^2 g_\alpha + \sum_{k=1}^m \partial_{z_k z_k}^2 g_\alpha \partial_{y_i} h_k$$

$$+ \sum_{k=1}^m (\partial_{y_j z_k}^2 g_\alpha + \sum_{l=1}^m \partial_{z_k z_l}^2 g_\alpha \partial_{y_j} h_l) \partial_{y_i} h_k + \sum_{k=1}^m \partial_{z_k} g_\alpha \partial_{y_i y_j}^2 h_k \quad (IV)$$

$$= 0 \quad (i, j = 1, \dots, n, \alpha = 1, \dots, m).$$

Note: $(x^*, \lambda^*) = (y^*, h(y^*), \lambda^*) = (y^*, z^*, \lambda^*)$ is a critical point of L . Hence, $\partial_{z_k} L = \partial_{z_k} f - \sum_{\alpha=1}^m \lambda_\alpha^* \partial_{z_k} g_\alpha = 0$ at (x^*, λ^*) . We can now multiply (IV) by $-\lambda_\alpha^*$, then sum over α . This, and (III), implies at (x^*, λ^*) that

$$\begin{aligned} \partial_{y_i y_j}^2 \phi &= \partial_{y_i y_j}^2 L + \sum_{k=1}^m \partial_{y_i z_k}^2 L \partial_{y_j} h_k \\ &\quad + \sum_{k=1}^m \partial_{y_j z_k}^2 L \partial_{y_i} h_k + \sum_{k=1}^m \sum_{l=1}^m \partial_{z_k z_l}^2 L \partial_{y_j} h_l \partial_{y_i} h_k. \end{aligned} \quad (V)$$

Let $\hat{y} = (\hat{y}_1, \dots, \hat{y}_{n-m})^\top \in \mathbb{R}^{n-m}$. Assume $\hat{y} \neq 0$. We show that $\hat{y}^\top \nabla^2 \phi(x^*) \hat{y} > 0$. This will imply that y^* is a local min. pt. of $\phi(y) = f(y, h(y))$, and hence will complete the proof.

Let $\hat{x} = \begin{bmatrix} \hat{y} \\ \nabla h(y^*) \hat{y} \end{bmatrix} \in \mathbb{R}^n$. Clearly, $\hat{x} \neq 0$. We show $\underbrace{\nabla g(x^*) \hat{x}}_{m \times n} \in \underbrace{\mathbb{R}^m}_{n \times 1} \in \underbrace{\mathbb{R}^{m \times 1}}_{m \times 1} = 0$.

Start with $g(y, h(y)) = 0$. We have by the chain rule

$$\partial_{y_i} g_\alpha + \sum_{k=1}^m \partial_{z_k} g_\alpha \partial_{y_i} h_k = 0 \quad \forall \alpha \in \{1, \dots, m\} \quad \forall i \in \{1, \dots, n-m\}.$$

Multiplying the eq. by \hat{y}_i and summing over i , we get

$$\sum_{i=1}^{n-m} \partial_{y_i} g_\alpha(x^*) \hat{y}_i + \sum_{i=1}^{n-m} \sum_{k=1}^m \partial_{z_k} g_\alpha(x^*) \partial_{y_i} h_k(x^*) \hat{y}_i = 0 \quad \forall \alpha.$$

This is the same as

$$\nabla g(x^*) \hat{x} = [\nabla g(x^*) \quad \nabla h(x^*)] \begin{bmatrix} \hat{y} \\ \nabla h(x^*) \hat{y} \end{bmatrix} = 0 \in \mathbb{R}^{m \times 1}$$

Now, by the assumption, we have $\hat{x}^\top H_L(x^*, \lambda^*) \hat{x} = \hat{x}^\top \nabla^2 L(x^*, \lambda^*) \hat{x} > 0$. But, by direct verifications from (V), we have

$$\hat{y}^\top \nabla^2 \phi(x^*) \hat{y} = \hat{x}^\top H_L(x^*, \lambda^*) \hat{x} > 0.$$

QED

The penalty method.

Minimize $f(x)$ subject to $g(x) = 0$. $f: \mathbb{R}^n \rightarrow \mathbb{R}$, $g: \mathbb{R}^n \rightarrow \mathbb{R}^m$, $m < n$.
 Let $M >> 1$. Minimize $F(x) = f(x) + \frac{M}{2} \|g(x)\|^2$.

Theorem Let $E \subseteq \mathbb{R}^n$ be open and bounded, $f \in C(\bar{E})$, and $g \in C(E, \mathbb{R}^m)$ with $m < n$. Assume $\{x \in E : g(x) = 0\} \neq \emptyset$. Let $0 < M_k \uparrow \infty$. Assume for each k , $x_k = \arg \min_{\bar{E}} F_k$, where $F_k(x) = f(x) + \frac{M_k}{2} \|g(x)\|^2$. If $x_k \rightarrow \bar{x} \in \bar{E}$, then \bar{x} is a min. point of f constrained by $g=0$, i.e., $f(\bar{x}) \leq f(x)$ for any $x \in E$ such that $g(x)=0$.

Idea of proof (see Prob. 7 of HW #2).

① $\exists \hat{x} \in E : g(\hat{x}) = 0$. $F_k(x_k) \leq F_k(\hat{x}) = f(\hat{x})$ ($k=1, 2, \dots$). $\Rightarrow \{F_k(x_k)\}$ is bounded. Since f is bounded on \bar{E} ,

$$\min_{\bar{E}} f(x) + \frac{M_k}{2} \|g(x_k)\|^2 \leq F_k(x_k) \leq f(\hat{x}), \quad k=1, 2, \dots$$

Thus, $g(x_k) \rightarrow 0$. But $g(x_k) \rightarrow g(\bar{x})$. So, $g(\bar{x}) = 0$.

② Now, $x_k \rightarrow \bar{x} \Rightarrow f(x_k) \rightarrow f(\bar{x})$.

$$\begin{aligned} f(\bar{x}) &= f(\hat{x}) + \frac{M_k}{2} \|g(\hat{x})\|^2 \geq F_k(x_k) = f(x_k) + \frac{M_k}{2} \|g(x_k)\|^2 \\ &\geq f(x_k) \rightarrow f(\bar{x}) \text{ as } k \rightarrow \infty. \end{aligned}$$

So, $M_k \|g(x_k)\| \rightarrow 0$.

③ $\forall x \in E$. Assume $g(x) = 0$. Since $F_k(x_k) \leq F_k(x)$, we have

$$f(x_k) + \frac{M_k}{2} \|g(x_k)\|^2 \leq f(x) + \frac{M_k}{2} \|g(x)\|^2 = f(x).$$

Sending $k \rightarrow \infty$, we get by ② $f(\bar{x}) = \liminf_{k \rightarrow \infty} f(x_k) \leq f(x)$. QED