Lecture 16. Lagrange multiplier method

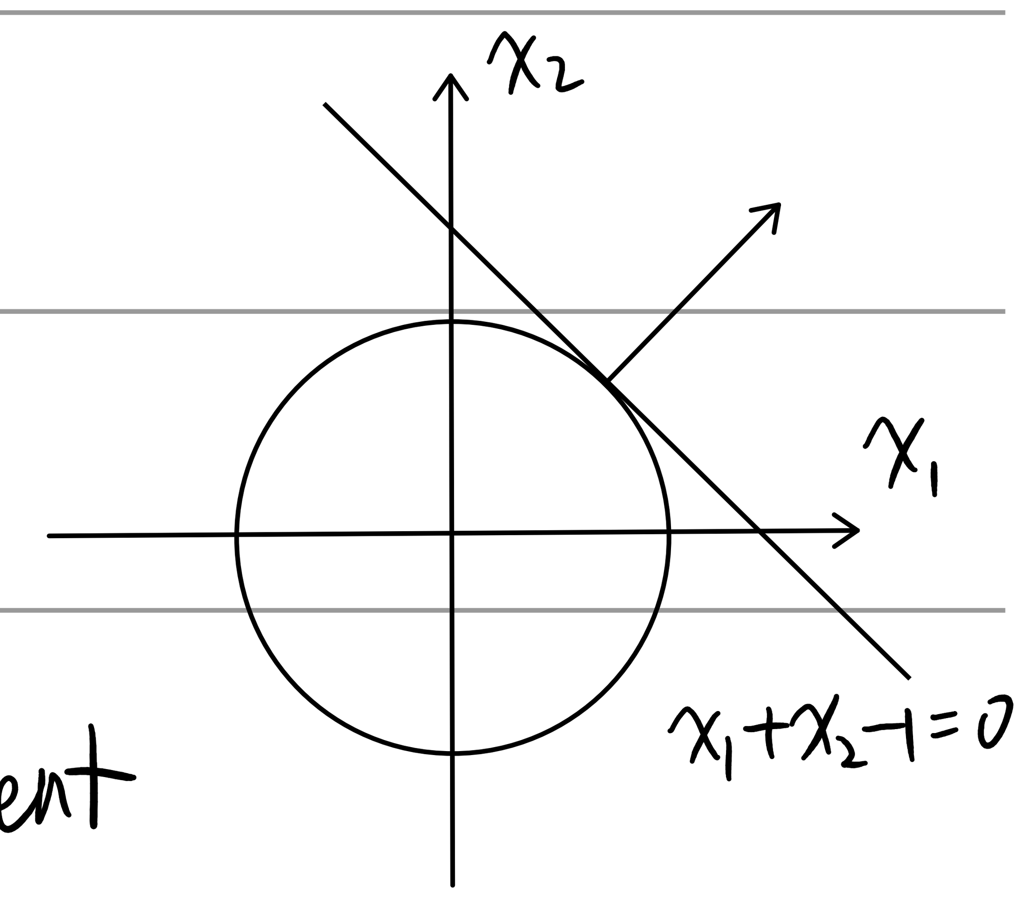
Consider a convex optimization with equality constraints.

$$\min\ f(x_1, x_2) = x_1^2 + x_2^2. \quad \text{subject to} \quad g(x_1, x_2) = x_1 + x_2 - 1 = 0$$

First question: how to certify minimum value? $\begin{cases} x_1 = 1/3 \\ x_2 = 2/3 \end{cases}$ or $\begin{cases} x_1 = 1/2 \\ x_2 = 1/2 \end{cases}$

Already known: for linear programs, $\exists$ an optimal solution at a vertex.

Obviously does not hold if f is not affine.



Recall the intuition for LASSO: if x^* is an optimal point. then the two level set should be tangent

$S_1: \{x \in \mathbb{R}^2 : f(x) = f(x^*)\}$. $S_2: \{x = (x_1, x_2) \in \mathbb{R}^2 : x_1 + x_2 - 1 = 0\}$.

Otherwise suppose $x^* \neq x' \in S_1 \cap S_2$. Let $\theta \in (0,1)$. consider $\theta x^* + \bar{\theta} x'$.

S_2 is convex (affine). So $\theta x^* + \bar{\theta} x' \in S_2$. is a feasible solution.

S_1 is not convex. but sublevel set $\{x : f(x) \le f(x^*)\}$ is convex.

So $f(\theta x^* + \bar{\theta} x') < f(x^*)$. contradicts x^* is an optimal point.

Now we need a formal statement for tangency (not singleton intersection).

Also. it is not a good description if S_1 overlaps S_2. see $f = |x_1| + |x_2|$.

We say S_1 is tangent with S_2 at x^*. if they share a common tangent line at x^*. The slope of the tangent line is the derivative.

For S_2: slope $= -1$. For S_1: $x_2 = \sqrt{c - x_1^2}$. $\frac{\partial x_2}{\partial x_1} = -\frac{2x_1}{2\sqrt{c-x_1^2}}$

$\frac{\partial x_2}{\partial x_1} = -1 \Rightarrow x_1 = \sqrt{c-x_1^2} = x_2$ however it is hard to generalize

A better viewpoint: consider normal vector

$S_2: \{(x_1, x_2): x_1 + x_2 - 1 = 0\}$ normal vector $(1, 1) = \nabla g(x^*)$.

$S_1: \{x: f(x) = f(x^*)\}$. normal vector ? $\nabla f(x^*)$

Require implicit function theorem. Only consider the simple example.

$S_1: \{x: f(x) = c\}$. Assume $\exists\, \varphi$. $(x_1, x_2) \in S_1$ iff $x_2 = \varphi(x_1)$.

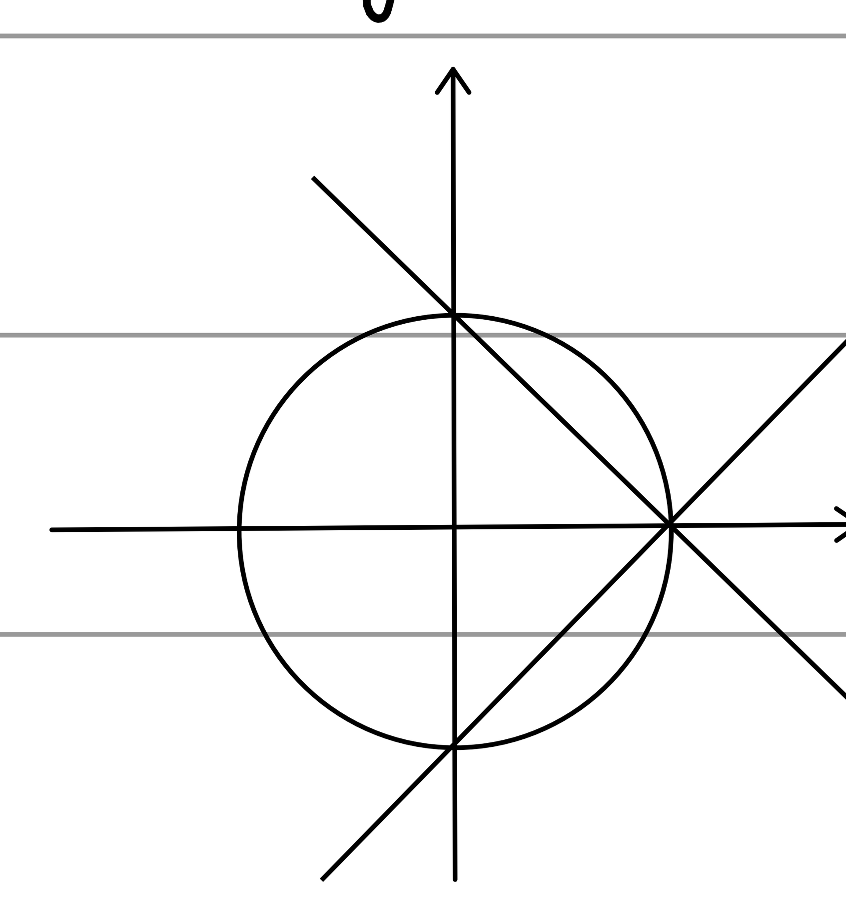
Recall slope of tangent line $= \varphi'$. Since $f(x_1, \varphi(x_1)) = c$.

taking derivatives of x_1 on both sides: $\frac{\partial f}{\partial x_1} + \frac{\partial f}{\partial x_2} \cdot \frac{\partial x_2}{\partial x_1} = 0$

$\Rightarrow \varphi'(x_1) = \frac{\partial x_2}{\partial x_1} = -\frac{\partial f}{\partial x_1} \Big/ \frac{\partial f}{\partial x_2}$ if $\frac{\partial f}{\partial x_2} \neq 0$. so normal vector $= \nabla f$.

Overall. x^* is a minimum point if $\exists\, \lambda \neq 0$. $\nabla f(x^*) + \nabla g(x^*) \cdot \lambda = 0$.

How about more constraints?



$\min f(x_1, x_2) = x_1^2 + x_2^2$. $\nabla f(x^*) = 2x^*$.

s.t. $g_1(x_1, x_2) = x_1 + x_2 - 1 = 0$ $\quad \nabla g_1 = (1, 1)$

$\quad\; g_2(x_1, x_2) = x_1 - x_2 - 1 = 0$. $\quad \nabla g_2 = (1, -1)$

$x^* = (1, 0)$. $\Rightarrow \nabla f(x^*) = (2, 0)$. $\exists\, \lambda_1, \lambda_2$, $\nabla f(x^*) + \lambda_1 \nabla g_1 + \lambda_2 \nabla g_2 = 0$.

Lagrange multiplier method: Suppose f is a convex function. $g_1, \dots g_m$ are affine functions. Then x^* is an optimal solution to $\min f(x)$ s.t. $g_1(x) = \cdots = g_m(x) = 0$. iff there exists $\lambda_1 \dots \lambda_m$ such that

$$\nabla f(x^*) + \lambda_1 \nabla g_1(x^*) + \cdots + \lambda_m \nabla g_m(x^*) = 0$$

Proof. Assume $g_i(x) = a_i^T x + b_i$. $A = \begin{pmatrix} a_1^T \\ \vdots \\ a_m^T \end{pmatrix} \in \mathbb{R}^{m \times n}$. $\operatorname{rank}(A) = m < n$.

Let $\Omega = \{x : g_1(x) = \cdots = g_m(x) = 0\}$ be the feasible set.

Then $\forall x \in \Omega$, $\nabla f(x^*)^T (x - x^*) \geq 0$ since x^* is an optimal solution.

If $x^* + v \in \Omega$, so is $x^* - v$. Thus $\nabla f(x^*)^T (x - x^*) = 0$ $\forall x \in \Omega$.

Also, $g_i(x) = g_i(x^*) = 0$, so $\forall i$, $a_i^T (x - x^*) = 0 \Rightarrow A(x - x^*) = 0$.

$\Rightarrow \Omega - x^* \in \ker(A)$. On the other hand, $\forall x$ satisfying $A(x - x^*) = 0$ has $x \in \Omega$. So $\Omega - x^* = \ker(A)$. Since $\nabla f(x^*)^T v = 0$ for all $v \in \ker(A)$. $\nabla f(x^*) \in \operatorname{span}\{a_1^T, \dots a_m^T\} = \operatorname{span}\{\nabla g_1, \dots \nabla g_m\}$.

"$\Leftarrow$". If $\exists$ such $\lambda_1 \dots \lambda_m$. $\nabla f(x^*)^T (x - x^*) = \sum \lambda_i a_i^T (x - x^*) = 0$.

So $f(x) \geq f(x^*) + \nabla f(x^*)^T (x - x^*) = f(x^*)$ by convexity. $\square$

Recall linear program: $\min c^T x$. s.t. $Ax = b$.

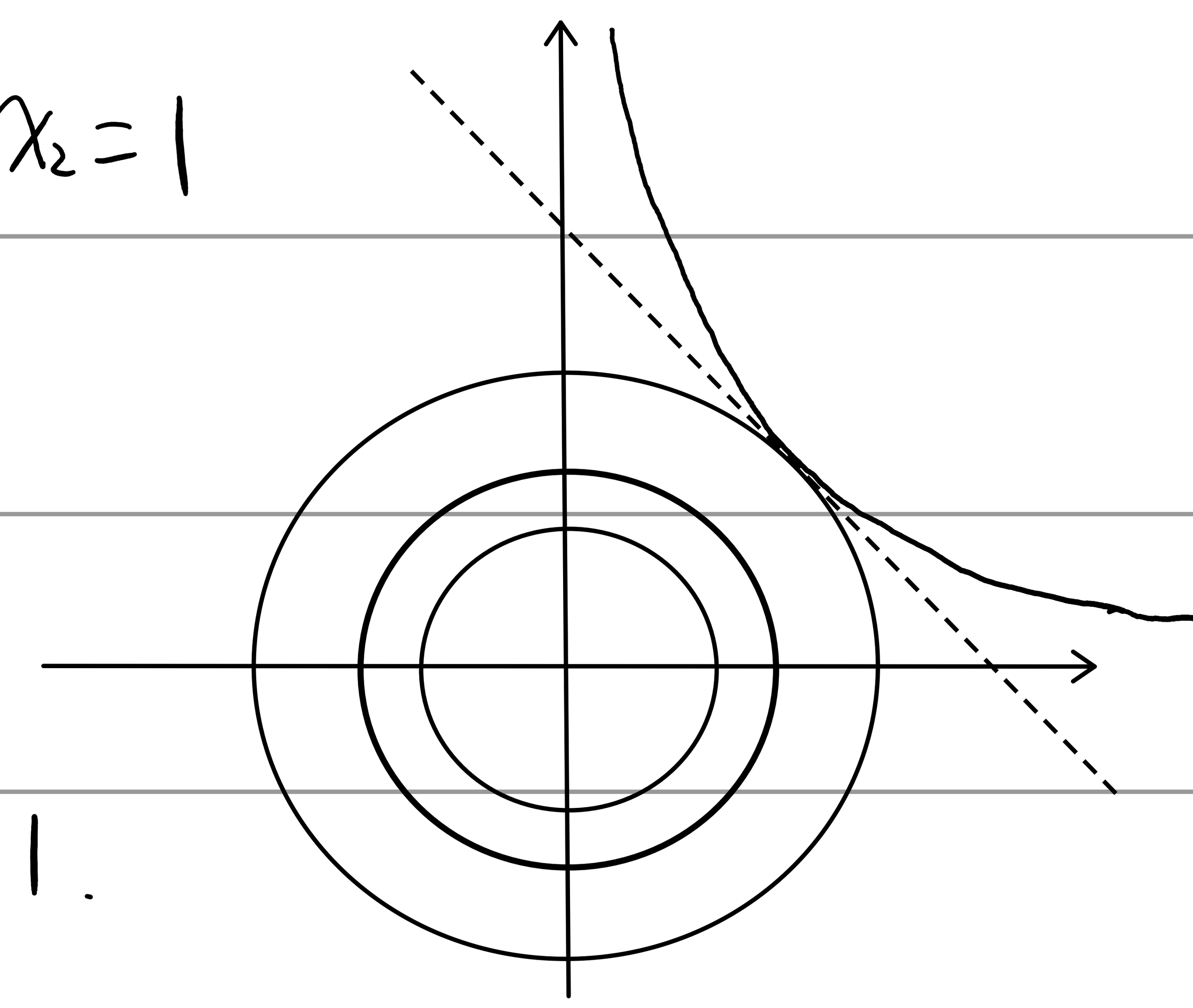
Lagrange multiplier: x^* is optimal if ① $Ax = b$ ② $\exists \lambda$. $c^T = \lambda^T A$.

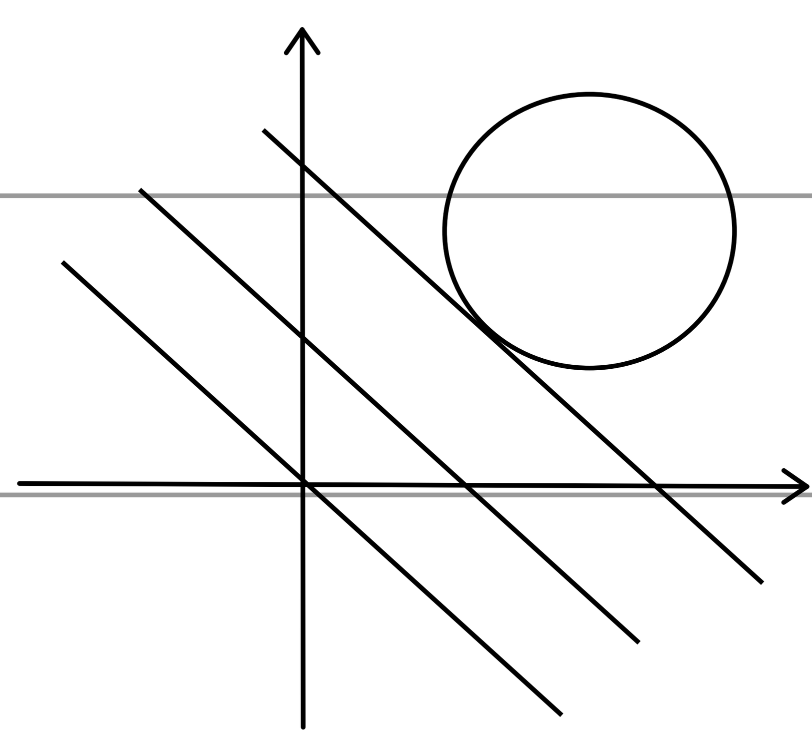
Strong duality: dual problem $\max y^T b$. s.t. $y^T A = c^T$.

x^* is optimal for (P) iff $\exists\, y^*$ optimal for (D). $c^T x = y^T A x = y^T b$.

A natural question: how about general objective and constraints?

$\min\ x_1^2 + x_2^2$ subject to $x_1 x_2 = 1$

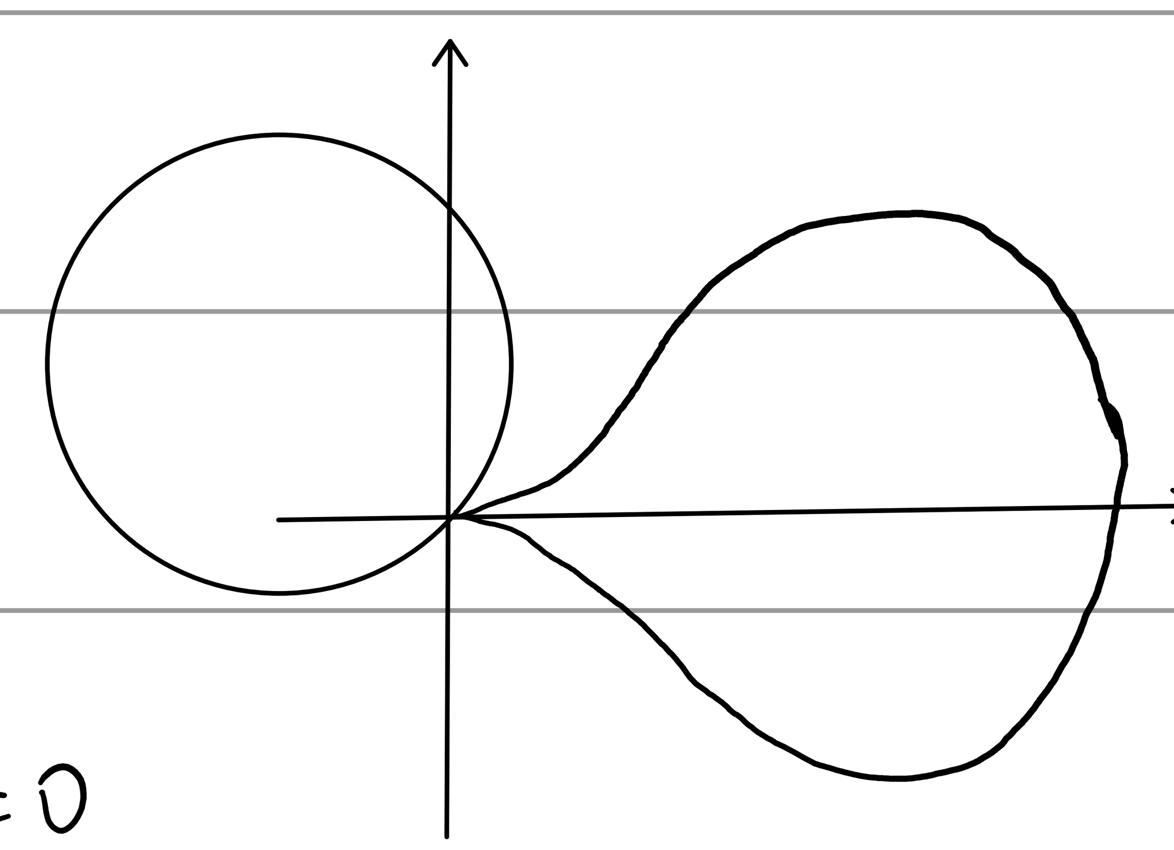




$\min\ x_1 + x_2$

s.t. $(x_1-2)^2 + (x_2-2)^2 = 1$.

However. consider the example



$\min f(x_1, x_2) = (x_1+1)^2 + (x_2-1)^2$

s.t. $g(x_1, x_2) = (x_1^2 + x_2^2)^2 - 2x_1(x_1^2 + x_2^2) + 3x_2^2 = 0$.

Why? Because the level set of $x_1 x_2$. $(x_1-2)^2 + (x_2-2)^2$ can be approximated by $\mathbb{R}^2$ locally. (differentiable) manifolds. 流形

Homeomorphism 同胚 $\exists\, \varphi : \Omega_1 \to \Omega_2$ φ invertible. φ. φ^{-1} continuous

Diffeomorphism 微分同胚. φ. $\varphi^{-1} \in C^\infty$ smooth.

(Sub)manifolds: neighborhood of each point is similar (homeomorphism) to $\mathbb{R}^m$

Proposition: The graphs of continuous differentiable functions are submanifolds.

$\{(x, f(x)) \in \mathbb{R}^{n+1}\}$. $\varphi: \mathbb{R}^{n+1} \to \mathbb{R}^{n+1}$ $(x, f(x)) \mapsto (x, 0)$.

Parameterized curve: $\gamma: (-\varepsilon, \varepsilon) \to \mathbb{R}^n$. $\gamma'(t) = 0$ $\forall t \in (-\varepsilon, \varepsilon)$.

$\varphi: \mathbb{R}^n \to \mathbb{R}^n$. $(x_1, \dots x_n) \mapsto \gamma^{-1}(x_1, \dots x_n) \cdot \gamma'(0)$ tangent line.

Now we consider level set $\{x \in \mathbb{R}^n : g(x) = 0\}$. (sub) manifolds ?

If $g(x_1, \dots, x_n) = 0 \Rightarrow x_n = h(x_1, \dots x_{n-1})$. then the level set is the graph of h. and thus is a submanifold if h is "good".

Implicit function theorem.

Let $g: \mathbb{R}^{n+m} \to \mathbb{R}^m$ continuously differentiable. $(x, y) = (x_1, \dots, x_n, y_1, \dots, y_m)$. and $g(x, y) = 0$ at $(x, y) = (a, b)$. If $D_y g(a, b)$. i.e. the Jacobian matrix $\left(\frac{\partial g_i}{\partial y_j}\right)_{1 \le i, j \le m}$ is invertible at (a, b). then there exists open sets $U \subseteq \mathbb{R}^n$. $V \subseteq \mathbb{R}^m$ and continuously differentiable function $\varphi: U \to V$. s.t. $a \in U$. $b \in V$. $\varphi(a) = b$ and $\forall (x, y) \in U \times V$. $g(x, y) = 0$ iff $y = \varphi(x)$.

Moreover. $\forall x \in U$. $D\varphi(x) = -\left(D_y g(x, y)\right)^{-1}_{m\times m} D_x g(x, y)_{m\times n}$ where $y = \varphi(x)$.

Lagrange multiplier. Let $f: \mathbb{R}^n \to \mathbb{R}$ and $g: \mathbb{R}^n \to \mathbb{R}^m$ continuously differentiable. x^* is an optimal solution to $\begin{cases} \min f(x) \\ \text{s.t. } g(x) = 0 \end{cases}$. If $\operatorname{rank} Dg(x^*) = m$ $(< n)$. (x^* is regular). then $\exists$ unique λ^*. s.t. $\nabla f(x^*) + \lambda^{*T} Dg(x^*) = 0$.