Lecture 18. Lagrange multiplier method; Newton's method.

Lagrange multiplier: Let $f: \mathbb{R}^n \to \mathbb{R}$. $g: \mathbb{R}^n \to \mathbb{R}^m$. opt. $\min f(x)$ s.t. $g(x)=0$.

Suppose $f, g \in C^1$ and x^* is an optimal solution. If x^* is regular (i.e. rank $Dg(x^*) = \underset{<n}{m}$). Then $\exists$ unique λ^* s.t. $\nabla L(x^*, \lambda^*) = 0$. Lagrangian.

Proof. $\forall v$. $df(x^*)(v) = \nabla_v f(x^*) = 0$. $\Rightarrow T_pM \subseteq \ker df(x^*)$. $T_pM = \ker dg$.

$\Rightarrow \ker df(x^*) \supseteq \ker dg(x^*)$. i.e. $\exists \lambda^*$. $df(x^*) = \lambda_1^* dg_1(x^*) + \cdots + \lambda_m^* dg_m(x^*)$. $\square$

We start from the explicit proof for linear constraints (affine g).

where the key ingredients are $\nabla f(x^*)^T(x-x^*) = 0$ and $Dg^T(x-x^*) = 0$.

However. if M defined by zero points is a general set. they're wrong.

Roughly. $\nabla f(x^*)^T v = 0$ if v is a tangent vector. so we should develop a kind of differential on a particular structer. The idea is to relax $\mathbb{R}^n$ to keep only local properties. which is submanifolds.

Then we formally define (differentiable) submanifolds and develop differential df. A key example is that the graph is a submanifold.

Implicit function theorem tells us when we can obtain $y = \Phi(x)$ given $g(x, y) = 0$. That's the modern formulation of Lagrange multiplier.

Assume x^* is a regular point. the necessary conditions are:

$\exists \lambda^* = (\lambda_1, \dots, \lambda_m)^T \in \mathbb{R}^m$. $\nabla f(x^*) = (\lambda^*)^T Dg(x^*)$ and $g(x^*) = 0$.

Define Lagrangian $L(x, \lambda) = f(x) - \lambda^T g(x^*)$. so we have

$\nabla_x L(x, \lambda) = \nabla f(x) - \lambda^T Dg(x)$. $\nabla_\lambda L(x, \lambda) = g(x)$.

Finally. x^* optimal and regular $\Rightarrow \nabla L(x^*, \lambda^*) = 0$. $(\min_{x,\lambda} L)$.

Remark: we should remind the dual of LP here. where each constraint is given a multiplier. Consider the strong duality.

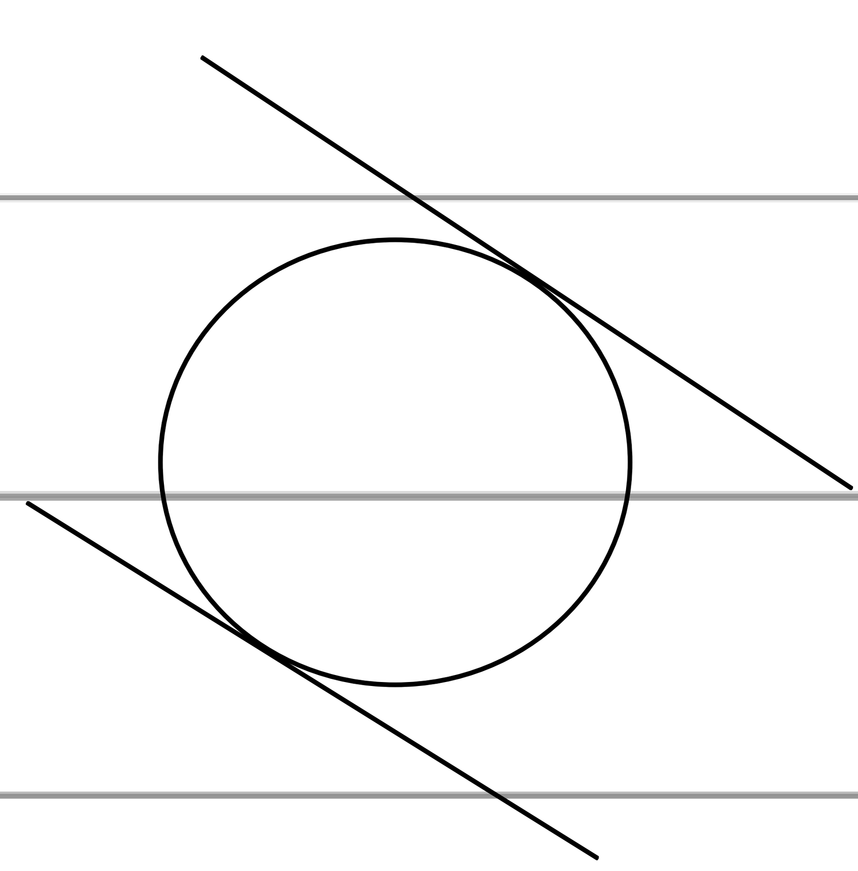
Primal. $\min c^T x$. s.t. $Ax = b$. Dual: $\max y^T b$. s.t. $y^T A = c^T$.

Lagrange multiplier: $L(x, y) = c^T x - y^T (Ax - b)$.

$\nabla_x L(x, y) = c^T - y^T A$. $\nabla_y L(x, y) = Ax - b$. $\nabla L = 0$ if $\begin{cases} y^T A = c^T \\ Ax = b \end{cases}$.

strong duality of LP: x^* optimal if $\exists y^*$. s.t. $y^T A = c^T$.

Lagrange multiplier is a necessary condition. Is it sufficient?

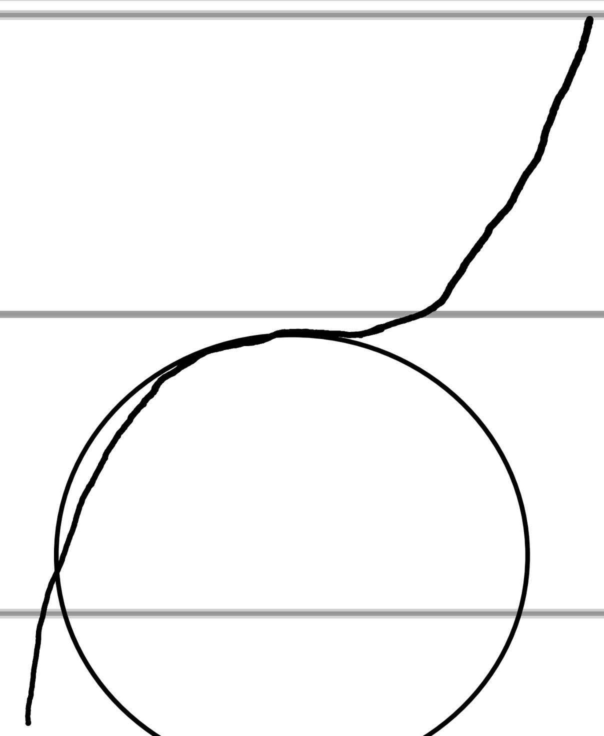


$$\min x_1^2 + x_2^2. \quad \text{s.t. } x_1 + x_2 = 1.$$

$$L(x, \lambda) = x_1^2 + x_2^2 - \lambda(x_1 + x_2 - 1).$$

$$\nabla L = 0 \iff \begin{cases} 2x_1 - \lambda = 0 \\ 2x_2 - \lambda = 0 \\ x_1 + x_2 - 1 = 0 \end{cases} \iff x_1 = x_2 = \tfrac{1}{2}.\ \lambda = 1.$$

$\min x_1 + x_2$. s.t. $x_1^2 + x_2^2 = \frac{1}{4}$ $\quad \nabla L = 0 \iff \begin{cases} 1 - 2\lambda x_1 = 0 \\ 1 - 2\lambda x_2 = 0 \\ x_1^2 + x_2^2 = \frac{1}{4} \end{cases}$



$$\min \; x_1^4 + x_2^4 \quad \text{s.t.} \quad x_1^3 + 1 - x_2 = 0$$

$$L(x, \lambda) = x_1^4 + x_2^4 - \lambda(x_1^3 + 1 - x_2).$$

$$\nabla L = 0 \iff \begin{cases} 4x_1^3 - 3\lambda x_1^2 = 0 \\ 4x_2^3 + \lambda = 0 \\ x_1^3 + 1 - x_2 = 0 \end{cases} \Leftarrow \begin{cases} x_1 = 0 \\ x_2 = 1 \\ \lambda = 2 \end{cases}.$$

Is it a local minimum or maximum? consider $\tilde{f} = x_1^4 + (x_1^3+1)^4$

$\tilde{f}'(x_1) = 4x_1^2(x_1 + 3(x_1^3+1)^3)$. $x_1 \in (-\varepsilon, \varepsilon) \Rightarrow \tilde{f}(x_1) \geq 0$.

$\min f(x)$. s.t. $g(x) = 0$. equivalent to $\min_{x, \lambda} L(x, \lambda)$? assume nondegenerate

Assume $\min L(x, \lambda)$ exists. $(x^*, \lambda^*) = \operatorname{argmin} L(x, \lambda)$

$\nabla L(x^*, \lambda^*) = 0 \Rightarrow g(x^*) = 0$ feasible. $\forall$ feasible x. $\forall \lambda \in \mathbb{R}^m$

$f(x) = L(x, \lambda)$. $\min L(x, \lambda)$ exists $\Rightarrow f$ cannot be unbounded below

So. $\exists \min L(x, \lambda) \Rightarrow \exists \min_{g=0} f$. Furthermore. $\min_{g=0} f = \min_{x, \lambda} L(x, \lambda)$.

Proof: suppose not. let $y, \theta = \operatorname{argmin} L(x, \lambda)$. $x^* = \operatorname{argmin} f(x)$.

$y \neq x^*$ by assumption. However $\nabla L(y, \theta) = 0 \Rightarrow f(y) = L(y, \theta)$.

$\exists! \lambda^*$. $\nabla L(x^*, \lambda^*) = 0 \Rightarrow L(x^*, \lambda^*) = f(x^*)$. contradiction $\square$.

Consider f convex and g affine. Assume $\exists (x^*, \lambda^*)$. $\nabla L(x^*, \lambda^*) = 0$.

Define $\hat{L}(x) = L(x, \lambda^*) = f(x) - \lambda^* g(x)$. convex. note $\nabla \hat{L}(x^*) = 0$.

$\Rightarrow x^* = \operatorname{argmin}_x \hat{L}(x)$. $\forall$ feasible x. $f(x) = \hat{L}(x)$. so $x^* = \operatorname{argmin}_{g(x)=0} f(x)$.

$g(x)$ affine $\Rightarrow$ linearly independent. otherwise redundant or infeasible.

Remark: In fact. if $f(x)$ convex and $g(x)$ affine. regularity is not necessary. If $g(x)=0$. infeasible. $\nabla L(x,\lambda)=0$ also infeasible.

If g linearly independent. $\exists$ unique λ^*. if redundant, $\exists$ more λ^*.

First-order optimality condition for ECP (equality constrained convex prob.):

$x^* = \arg\min f(x)$. s.t. $g(x)=0$. iff $\exists \lambda^*$. $\nabla L(x^*, \lambda^*)=0$.

Second-order optimality condition: $\forall v \in T_pM$. $v^T \nabla^2 f(p) v \geq 0$?

consider $f = x_1^2 + x_2^2$. s.t. $(x_1-2)^2 + x_2^2 = 1$. $x^* = (1,0)^T$. $\nabla^2 f = \begin{pmatrix} 2 & 0 \\ 0 & 2 \end{pmatrix}$

$T_{x^*}M = t(-2,0)^T$ $t \in \mathbb{R}$. $v^T \nabla^2 f \cdot v = 8t^2$. But if $f = -(x_1^2 + x_2^2)$?

Why? note that if $v \in T_pM$. $p+v$ may not in M. So Taylor is not an approximation. but an (approximate) upper/lower bound. (revisit when KKT).

Let $\gamma: (-\varepsilon, \varepsilon) \to M \in C^2$ be a curve on M. $\gamma(0)=p$. $\gamma'(0)=v$.

If p is a local minimum of f. then 0 is a local min of $f\circ\gamma$.

$\varphi = f\circ\gamma: (-\varepsilon,\varepsilon) \to \mathbb{R}$. $\Rightarrow \varphi''(0) \geq 0$. By chain rule.

$\frac{d}{dt}\varphi = (\gamma_1'(t), \dots, \gamma_n'(t)) \cdot \nabla f(\gamma(t))$. $\frac{d}{dt}(u^T v) = v^T \cdot \frac{d}{dt} u + u^T \cdot \frac{d}{dt} v$.

$\Rightarrow \frac{d^2}{dt^2}\varphi = \frac{d\gamma}{dt}^T \cdot \nabla^2 f(\gamma(t)) \cdot \frac{d\gamma}{dt} + (\gamma_1''(t), \dots, \gamma_n''(t)) \cdot \nabla f(\gamma(t)) \geq 0.$

Note that by the first-order condition $\nabla f(x^*) = \sum_{i=1}^{m} \lambda_i^* \nabla g_i(x^*)$

$\frac{d^2}{dt^2}(g_i \circ \gamma) \equiv 0$. so $v^T \nabla^2 f(x^*) v - \sum_{i=1}^{m} \lambda_i^* v^T \nabla^2 g_i(x^*) v \geq 0$.

Second-order necessary condition: $\min f(x)$. s.t. $g(x) = 0$. (M).

If x^* is a regular point, then $\exists \lambda^*$. s.t. $\begin{cases} \nabla L(x^*, \lambda^*) = 0. \\ v^T \nabla_x L \cdot v \geq 0, \ \forall v \in T_{x^*}M. \end{cases}$

Sufficient condition: if x^* is a regular point, $\exists \lambda^*$. s.t.

$\nabla L(x^*, \lambda^*) = 0$. $\forall v \in T_{x^*}M$, $v \neq 0$. $v^T \nabla_x L(x^*, \lambda^*) v > 0$. ?

Example. $\min x^T Q x$. s.t. $x^T x = 1$. $Q = \text{diag}\{2, 1\}$. (hw).

Equality constrained (convex) quadratic optimization problem.

$\min \frac{1}{2} x^T Q x + w^T x$. $Q \geq 0$. s.t. $Ax = b$.

Lagrangian $L(x, \lambda) = \frac{1}{2} x^T Q x + w^T x - \lambda^T (Ax - b)$.

Lagrange condition: $\begin{cases} \nabla_x L(x, \lambda) = Qx + w - A^T \lambda = 0 \\ \nabla_\lambda L(x, \lambda) = Ax - b = 0 \end{cases}$ assume $Q > 0$.

$\begin{pmatrix} Q & A^T \\ A & 0 \end{pmatrix} \begin{pmatrix} x \\ -\lambda \end{pmatrix} = \begin{pmatrix} -w \\ b \end{pmatrix}$ KKT Matrix. Block Gaussian elimination.

row 2 $\leftarrow AQ^{-1}$ row 1 $-$ row 2. $(0, AQ^{-1}A^T) \begin{pmatrix} x \\ -\lambda \end{pmatrix} = b + AQ^{-1}w$

$\Rightarrow \lambda = (AQ^{-1}A^T)^{-1}(b + AQ^{-1}w)$. $Qx - A^T\lambda = -w$

$\Rightarrow x = -Q^{-1}w + Q^{-1}A^T (AQ^{-1}A^T)^{-1}(b + AQ^{-1}w)$ provided $\exists (AQ^{-1}A^T)^{-1}$

If Q invertible ($Q > 0$). and $\underbrace{\text{coloumns of } A^T \text{ linearly independent}}_{\text{rank}(A) = m}$. done.

If KKT matrix nonsigular $\overset{\text{rank}(A)=m}{\Rightarrow}$ $\exists$ unique solution (x^*, λ^*).

① $\circ$ $\ker(Q) \cap \ker(A) = \{0\}$. Q and A have no nontrivial common kernel (nullspace).

② $\circ$ $Ax = 0$, $x \neq 0 \Rightarrow x^T Q x > 0$. $Q > 0$ on $\ker(A)$.

③ $\circ$ $F^T Q F > 0$ for $\forall F \in \mathbb{R}^{n\times(n-m)}$ s.t. $\mathrm{Im}(F) = \{Fv : v \in \mathbb{R}^{n-m}\} = \ker(A)$.

Proof. "$\Rightarrow$". if $0 \neq x \in \ker(Q) + \ker(A)$ $\begin{pmatrix} Q & A^T \\ A & 0 \end{pmatrix} \begin{pmatrix} x \\ 0 \end{pmatrix} = 0$.

"$1 \Rightarrow 2$". $Ax = 0$, $x \neq 0 \Rightarrow Qx \neq 0$. lemma: $Q \succeq 0$, $Qx = 0$ iff $x^T Q x = 0$.

$Q = U \Lambda U^T$. $Q = \sum_{i=1}^n \xi_i u_i u_i^T$. $\xi_i \geq 0$. $\Rightarrow x^T Q x = \sum_{i=1}^n \xi_i \|u_i^T x\|^2$

$x^T Q x = 0 \Rightarrow u_i^T x = 0$ if $\xi_i > 0$. $\Rightarrow Qx = \sum_{i=1}^n u_i (\xi_i u_i^T x) = 0$.

"$\Leftarrow$". Assume $\begin{pmatrix} Q & A^T \\ A & 0 \end{pmatrix} \begin{pmatrix} v \\ w \end{pmatrix} = 0$. need to show $v = w = 0$.

$v^T Q v = v^T(-A^T w) = -(Av)^T w = 0$. contradicts if $Av = 0$, $v \neq 0$.

$Qv + A^T w = 0 \Rightarrow A^T w = 0$. contradicts $\mathrm{rank}(A) = m$ if $w \neq 0$.

"$2 \Leftrightarrow 3$". $\mathrm{Im}(F) = \ker(A)$. $0 \neq y \in \ker(A) \Longleftrightarrow \exists x \neq 0$. $y = Fx$. $\square$

If KKT system unsolvable $\Rightarrow$ QP is infeasible or unbounded below.