Lecture 17. Lagrange multiplier method. Lagrangian function

Lagrange multiplier method for convex optimization.

Suppose $f: \mathbb{R}^n \to \mathbb{R}$ is a convex function. $g_1, g_2 \dots g_m : \mathbb{R}^n \to \mathbb{R}$ are affine functions. Then x^* is an optimal solution to $\begin{array}{l} \min f(x) \\ \text{s.t. } g_1(x) = \cdots = g_m(x) = 0 \end{array}$ if and only if x^* is feasible (i.e. $g_1(x^*) = \cdots = g_m(x^*) = 0$). and $\exists \lambda_1 \dots \lambda_m$. s.t. $\nabla f(x^*) + \lambda_1 \nabla g_1(x^*) + \cdots + \lambda_m \nabla g_m(x^*) = 0$.

For general case: hope level set of g looks like $\mathbb{R}^l$ (submanifolds).

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. $\nabla f(x^*)+\lambda^{*T}Dg(x^*)=0$.

Remark. Implicit function theorem is a local property.

Proof sketch: The key ingredients are parameterized curves.

We define tangent space: M is a (differentiable) submanifold, $p\in M$.

$$T_pM \triangleq \{\gamma'(0)\in\mathbb{R}^n : \gamma((-\varepsilon,\varepsilon))\subseteq M,\ \gamma(0)=p\}.$$

Theorem. Assume M is defined by the level set of $g_1(x)=\cdots=g_m(x)=0$, then $T_pM\subseteq \bigcap_{k=1}^m \ker \nabla g_k(p)^T$. If $\nabla g_1(p), \nabla g_2(p), \dots, \nabla g_m(p)$ are linearly independent (p is a regular point), then $T_pM=\bigcap_{k=1}^m \ker\nabla g_k(p)^T$

Let $g=\begin{pmatrix}g_1\\ \vdots\\ g_m\end{pmatrix}$. Then $Dg=\begin{pmatrix}\nabla g_1^T\\ \vdots\\ \nabla g_m^T\end{pmatrix}$, $T_pM=\ker Dg$.

Proof sketch. $\forall v \in T_pM$. $\exists \gamma: (-\varepsilon, \varepsilon) \to M$. such that $\begin{cases} \gamma(0) = p \\ \gamma'(0) = v \end{cases}$

So $\forall k$. $g_k(\gamma(t)) \equiv 0$. Taking derivatives of t at $t=0$

$\nabla g_k(p)^T(\gamma'(0)) = 0 \quad \Rightarrow T_pM \subseteq \ker \nabla g_k(p)^T$.

For equality. just compare dimensions.

Normal vector: $\forall v \in T_pM$. $\langle u, v \rangle = 0$. u is a normal vector

If $M: \{x : g_1(x) = \cdots = g_m(x) = 0\}$. $\nabla g_k(p)$ is normal of M at p.

Lemma. (First-order condition for optimality on submanifolds).

Suppose $M \subseteq \mathbb{R}^n$. is a submanifold. $f \in C^1(M)$ (continuously differentiable).

$p \in M$ is a local extreme point of f. Then $\nabla f(p)^T \in T_pM^\perp$.

Proof sketch. It suffices to show that $\forall v \in T_pM$. $\nabla f(p)^T v = 0$.

Since $v \in T_pM$. $\exists \gamma: (-\varepsilon, \varepsilon) \to M$. such that $\begin{cases} \gamma(0) = p \\ \gamma'(0) = v. \end{cases}$

p is a local extreme of $f \Rightarrow 0$ is a local extreme of $f(\gamma(t))$.

So $f'(\gamma(t)) = 0$. Thus $\nabla f(p)^T v = 0$.

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. $\nabla f(x^*) + \lambda^{*T} Dg(x^*) = 0$.

Proof. $\nabla f(x^*)^T \in T_pM^\perp$. $T_pM = \ker Dg$. $\Rightarrow \nabla f(x^*)^T \in \operatorname{im} Dg$.

$\nabla f(x^*) \in \operatorname{span}\{\nabla g_1(x^*) \ \dots \ \nabla g_m(x^*)\}$. $\square$

Recall the optimization problem $\begin{cases} \min & x_1 + x_2 \\ \text{s.t.} & x_1^2 + x_2^2 = 1 \end{cases}$.

$(x_1, x_2) = (-\frac{1}{2}, -\frac{1}{2})$ or $(x_1, x_2) = (\frac{1}{2}, \frac{1}{2})$?

Warning: No Taylor now.

Second-order necessary condition. If x^* is regular, and is an optimal solution to $\begin{cases} \min f(x) \\ \text{s.t. } g(x) = 0 \end{cases}$ then $\exists \lambda^*$. s.t. $\begin{cases} \nabla f(x^*) + \lambda_1^* \nabla g_1(x^*) + \dots = 0 \\ g_1(x^*) = \dots = g_m(x^*) = 0 \end{cases}$

and $\forall v \in T_{x^*}M$. $v^T \nabla^2 f(x^*) v + \sum_{i=1}^m \lambda_i^* \cdot v^T \nabla^2 g_i(x^*) v = 0$.

Proof sketch. By chain rule. $f(\gamma(t))$ has minimum point 0

$\Rightarrow f''(\gamma(0)) \geq 0 \Rightarrow \gamma'(0)^T \cdot \nabla^2 f(x^*) \cdot \gamma'(0) + \gamma''(0)^T \cdot \nabla f(x^*) \geq 0$.

note that $\nabla f(x^*) + \lambda_1^* \nabla g_1(x^*) + \dots + \lambda_m \nabla g_m(x^*) = 0$ and $\forall k$.

$g_k(\gamma(t)) \equiv 0 \Rightarrow g_k''(\gamma(t)) = \gamma'(0)^T \cdot \nabla^2 g_k(x^*) \cdot \gamma'(0) + \gamma''(0)^T \nabla g_k(x^*) = 0$

Thus $v^T \nabla^2 f(x^*) v + \sum_{i=1}^m \lambda_i^* \cdot v^T \nabla^2 g_i(x^*) v = 0$. $\forall v \in T_pM$. $\square$

Lagrangian function.

Let $L(x, \lambda) = f(x) + \lambda_1 g_1(x) + \dots + \lambda_m g_m(x)$ where $L: \mathbb{R}^{n+m} \to \mathbb{R}$

$x \in \mathbb{R}^n$. $\lambda \in \mathbb{R}^m$. Then $\begin{cases} \nabla_x L = \nabla f + \lambda_1 \nabla g_1 + \dots + \lambda_m \nabla g_m. \\ \nabla_\lambda L = (g_1(x), g_2(x), \dots, g_m(x))^T. \end{cases}$

So the Lagrange multiplier method can be written as:

if x^* is an optimal solution, then $\exists \lambda^*$, s.t. $\nabla L(x^*, \lambda^*) = 0$.