経済数学（法政用）: 第１１章

細矢祐誉

テーマ：最適化理論（ＫＫＴ定理）

・復習：ラグランジュ未定乗数法の原理

前章までの内容を思い出してみよう。まず、標準的な消費者の効用最大化問題

$$\begin{aligned} \max \quad & f(x) \\ \text{subject to.} \quad & x \in \mathbb{R}^n_+, \\ & p \cdot x \le m \end{aligned} \tag{1}$$

と、それに付随するラグランジュ関数

$$L(x, \lambda) = f(x) + \lambda(m - p \cdot x) \tag{2}$$

を考えよう。このとき、次の二つの関係を論じていたのだった。

1. x^* は問題 (1) の最大点である。
2. ある λ^* が存在して、$DL(x^*, \lambda^*) = 0^T$ である。

たとえば、1. から 2. を導出する証明を（概略だけ）振り返ってみよう。ここでは、$x^* \gg 0$ が (1) の最大点であるとし、$v \in \mathbb{R}^n$ で、ただし $p \cdot v = 0$ であるとする。このとき、$x(t) = x^* + tv$ として、$g(t) = f(x(t))$ とすると、$|t| > 0$ が十分 0 に近ければ $x(t) \ge 0$ であり、また

$$p \cdot x(t) = p \cdot x^* + tp \cdot v = p \cdot x^* \le m$$

である。よって、

$$g(t) = f(x(t)) \le f(x^*) = f(x(0)) = g(0)$$

が成り立つ。したがって g は 0 で極大になるので、$g'(0) = 0$ である。合成微分の公式から

$$0 = g'(0) = Df(x^*)v$$

となる。つまり、$p \cdot v = 0 \Rightarrow Df(x^*)v = 0$ がわかったのであり、ここから容易に、ある λ^* が存在して

$$Df(x^*) = \lambda^* p^T$$

が成り立つことがわかる。これは

$$D_x L(x^*, \lambda^*) = 0^T$$

を意味する。一方で f が増加的（つまり $x \gg y$ なら $f(x) > f(y)$）であれば $p \cdot x^* = m$ でしかあり得ないので、

$$0 = m - p \cdot x^* = D_\lambda L(x^*, \lambda^*)$$

となり、合わせて

$$DL(x^*, \lambda^*) = 0^T$$

を得る。ここまでで使っている仮定を整理すると、

- $x^* \gg 0$ である。
- f は x^* で全微分可能である。
- f は増加的である。

以上であり、これだけ成り立っていれば 1. から 2. が導出できることになる。

今度は 2. から 1. を導出する理屈を考えてみよう。いま、$D_\lambda L(x^*, \lambda^*) = 0$ なので、$m = p \cdot x^*$ である。一方で $D_x L(x^*, \lambda^*) = Df(x^*) - \lambda^* p^T = 0$ なので、

$$D_x L(x^*, \lambda^*) = \lambda^* p^T$$

であることがわかる。ここで背理法の仮定として、ある $x \geq 0$ について、$p \cdot x \leq m$ かつ $f(x) > f(x^*)$ であったと仮定してみよう。f が連続であるとすれば、x をほんの少しだけすべての座標を減らしても $f(x) > f(x^*)$ という関係を維持できるので、我々は $p \cdot x < m$ を仮定できる。ここで $x(t) = (1-t)x^* + tx$ と定義し、$g(t) = f(x(t))$ とする。f が準凹であれば、$0 \leq t \leq 1$ に対して

$$g(t) = f((1-t)x^* + tx) \geq \min\{f(x), f(x^*)\} = f(x^*) = g(0)$$

であるため、ここから

$$g'(0) = \lim_{t \downarrow 0} \frac{g(t) - g(0)}{t} \geq 0$$

がわかる。一方で合成微分の公式から

$$g'(0) = Df(x^*)(x - x^*) = \lambda^*[p \cdot x - p \cdot x^*] = \lambda^*[p \cdot x - m]$$

がわかる。よって、$p \cdot x < m$ より、$\lambda^* \leq 0$ がわかるが、もし f が増加的で $Df(x^*) \neq 0$ であればこれはあり得ない。以上をまとめると、

- f は連続で準凹、増加的である。
- f は x^* で微分可能で、$Df(x^*) \neq 0$ である。

この条件があれば、2. から 1. を導出できる。ここまでが標準のセットアップであった。

・非線形制約問題

経済学には、(1) 以外の形をした問題がたくさん出てくる。たとえば、(1) の双対問題として紹介される支出最小化問題

$$\begin{aligned} \min \quad & p \cdot y \\ \text{subject to.} \quad & y \in \mathbb{R}^n_+, \\ & f(y) \ge f(x) \end{aligned}$$

は、(1) と異なる点がふたつある。一つ目は max ではなく min である点だが、これは

$$\begin{aligned} \max \quad & -p \cdot y \\ \text{subject to.} \quad & y \in \mathbb{R}^n_+, \\ & f(y) \ge f(x) \end{aligned}$$

と書き換えてしまえばそれだけで解決する問題なので、重要ではない。むしろ問題なのは、制約条件が $p \cdot x \le m$ ほど単純ではないということである。上で書いた証明において、たとえば 1. から 2. を導出するところについては、$x(t) = x^* + tv$ としたときに、$p \cdot x(t) = p \cdot x^* \le m$ であることが非常に重要な役割を果たしていた。したがって、これが成り立たないかもしれない今回の状況では上のロジックは使えない。一方、2. から 1. を導出するロジックについては、$p \cdot x < m$ であるならば、$x(t) = (1-t)x^* + tx$ としたときに、$Df(x^*)(x - x^*) = \lambda^*(p \cdot x - m)$ と変形できて、$\lambda^* \le 0$ が出てくることが、本質的に重要であった。これもまた、今回のセットアップではできそうにない。ではどうすればよいか？

もう少し複雑な問題を考えよう。たとえば、次のような問題がある。ここで $p_H, p_L > 0$ かつ $p_H + p_L = 1$ とし、また $0 < c_L < c_H$ とする。

$$\begin{aligned} \max \quad & p_H[b(x_H) - w_H] + p_L[b(x_L) - w_L] \\ \text{subject to.} \quad & w_H \ge c_H(x_H),\ w_L \ge c_L(x_L), \\ & w_H - c_L(x_H) \le w_L - c_L(x_L),\ w_L - c_H(x_L) \le w_H - c_H(x_H). \end{aligned}$$

これは、契約理論における「真実表明的な直接顕示メカニズム」の設計問題で出てくる問題である。橋をかける企業と政府が交渉しているとし、x は橋の品質、$b(x)$ は品質によって保証される橋の社会的便益である。ただし企業が橋をかける費用は品質の関数であり、また p_H の確率で c_H で、p_L の確率で c_L である。c_H と c_L のどちらが真の費用関数かは、最初の段階では企業しか知らない。したがって政府は費用関数を聞き取り調査して、

その後で相手のコストに合わせて要求品質 x_H, x_L と、企業に与える報酬 w_H, w_L を支払う。ここで、企業は自分のコストについて嘘をつくことができるので、政府はそれを勘案して報酬体系を作らなければならない。最初の制約は**参加制約**と呼ばれ、これを下回ると企業はコスト割れするため、橋の建設を拒否する。二番目の制約は**誘因両立制約**と呼ばれ、これが成り立っていないと企業は自分のコストについて嘘をつきたくなる。このふたつの条件を元にして最適品質と報酬を決めるのが政府の目的である。

この主の問題では、条件が複数あることに注意しよう。したがって、いままでで使った問題よりもさらに高度になる。これらの問題を解決する一般的原理とはなにか？

・問題の定式化

以上のところから、問題を定式化しよう。

$$\begin{aligned} \max \quad & f(x) \\ \text{subject to.} \quad & x \in A, \\ & g(x) \geq 0 \end{aligned} \tag{3}$$

これが今回考える問題である。ただし、$A \subset \mathbb{R}^n$ で、$f : A \to \mathbb{R}$ とし、一方で $g : A \to \mathbb{R}^m$ とする。g がベクトル値関数であるというのが、上で述べた多数の制約を扱えるという問題に対応している。

これに対して今度は、

$$L(x, \lambda) = f(x) + \lambda^T g(x) \tag{4}$$

と定義しよう。転置記号から明らかなように、今回はもはや λ は数ではなくてベクトルである。ここで、また次のようなふたつの仮説を対比してみよう。

1. x^* は問題 (3) の最大点である。
2. ある $\lambda^* \geq 0$ について、$D_x L(x^*, \lambda^*) = 0^T$ となる。さらに、$\lambda_i^* g_i(x^*) = 0$ が常に成り立つ。

今回の目的は、1. と 2. の同値性を示すことである。まず 1. から 2. の方向を考えてみよう。

定理 1：x^* は A の内部に位置し、問題 (3) の最大点であるとする。また、f は x^* で微分可能、g は連続で x^* の付近で連続微分可能であるとし、$g_i(x^*) = 0$ となる座標 i の集

合を E とした時、ベクトル族 $(Dg_i(x^*))_{i\in E}$ は一次独立であるとする[*1]。このとき、ある $\lambda^*\geq 0$ が存在して、$D_xL(x^*,\lambda^*)=0^T$ が成り立ち、また $\lambda_i^*g_i(x^*)=0$ が常に成り立つ。

注意：定理 1 のうち、「x^* が A の内部にある」という条件は、以前の「$x^*\gg 0$」という条件に対応している。f の微分可能性が必要なのは当たり前であるが、一方で、増加的という条件に対応する条件はない。その代わり、ベクトル族の一次独立性という、見慣れない条件が追加されている。これは、以前には、$p\neq 0$ という条件に当たるもので、当たり前だったから仮定されていなかった。

証明：最初に、E が空集合であるときを考える。この場合、x^* は f の極大点であるから、$Df(x^*)=0^T$ である。したがって $\lambda^*=0$ とすれば主張が正しいことがわかる。

したがって以降、E は空集合でないとし、簡単のために番号を付け替えて、$E=\{1,2,...,k\}$ ということにしておこう。もちろん、$1\leq k\leq m$ である。ここで、陰関数定理と呼ばれる次の補題を紹介する（証明は省略。なお、本当は陰関数定理はもっと細かい主張を多数含む）。

補題 1：$f:B\to\mathbb{R}^m$ とし、ただし B は $\mathbb{R}^{n+m}$ の部分集合とする。また、(x^*,y^*) は B の内部に位置するとし、$f(x^*,y^*)=a$ であるとし、また f は (x^*,y^*) の周囲で連続微分可能で、$D_yf(x^*,y^*)$ は逆行列を持つ[*2]とする。このとき、x^* の近くで定義されたある関数 $y(x)$ が存在して、以下の性質を満たす。

1) $y(x^*)=y^*$ が成り立つ。
2) $f(x,y(x))=a$ が常に成り立つ。
3) y は連続微分可能である。

[*1] 一般に、ベクトル $v_1,...,v_k$ が一次独立であるとは、

$$a_1v_1+...+a_kv_k=0$$

となるような数の組 $a_1,...,a_k$ が、$a_1=...=a_k=0$ しか存在しないことを言う。

[*2] 一般にある行列 A に対して、

$$AB=BA=\begin{pmatrix}1&0&...&0\\0&1&...&0\\\vdots&\vdots&\ddots&\vdots\\0&0&...&1\end{pmatrix}$$

となる行列 B を A の逆行列と呼び、A^{-1} と書く。線形代数の教科書を参照。

さて、以降 $\hat{g}=(g_1,...,g_k)$ とし、$\bar{g}=(g_{k+1},...,g_m)$ としよう。このとき、x^* は問題 (3) の解であるから、制約をきつくした以下の問題

$$\begin{aligned} \max \quad & f(x) \\ \text{subject to.} \quad & x \in A, \\ & \hat{g}(x) = 0, \\ & \bar{g}(x) \gg 0 \end{aligned}$$

の解である。一方で、任意の $x \in A$ に対して $\hat{x}=(x_1,...,x_k)$ とし、$\bar{x}=(x_{k+1},...,x_n)$ としよう。$Dg_1(x^*),...,Dg_k(x^*)$ は一次独立なので、必要ならば座標の番号を付け替えることで、その最初の k 列を取り出してできる行列 $D_{\hat{x}}g(x^*)$ が逆行列を持つと仮定してよい[*3]。そこで陰関数定理からある連続微分可能な関数 $\hat{x}(\bar{x})$ が存在して、

$$\hat{x}(\bar{x}^*)=\hat{x}^*,\ \hat{g}(\hat{x}(\bar{x}),\bar{x})=\hat{g}(x^*)$$

を満たす。すると $\bar{x}$ が十分 $\bar{x}^*$ に近ければ $x(\bar{x})=(\hat{x}(\bar{x}),\bar{x})$ は上の問題の制約条件を満たすので、

$$f(x(\bar{x})) \le f(x^*) = f(x(\bar{x}^*))$$

がわかる。よって $\bar{x}^*$ は合成関数 $f \circ x$ の極大点なので、そこでの微分は 0 になり、

$$Df(x^*)Dx(\bar{x}^*) = D_{\hat{x}}f(x^*)D\hat{x}(\bar{x}^*) + D_{\bar{x}}f(x^*) = 0$$

がわかる。一方で、陰関数定理の帰結

$$\hat{g}(\hat{x}(\bar{x}),\bar{x}) = \hat{g}(x^*)$$

を $\bar{x}$ で微分して $\bar{x}=\bar{x}^*$ を代入すると、

$$D_{\hat{x}}\hat{g}(x^*)D\hat{x}(\bar{x}^*) + D_{\bar{x}}\hat{g}(x^*) = 0$$

がわかるので、整理すると

$$D\hat{x}(\bar{x}^*) = -(D_{\hat{x}}\hat{g}(x^*))^{-1}D_{\bar{x}}\hat{g}(x^*)$$

がわかる。これを上に代入して、

$$D_{\bar{x}}f(x^*) = D_{\hat{x}}f(x^*)(D_{\hat{x}}\hat{g}(x^*))^{-1}D_{\bar{x}}\hat{g}(x^*)$$

[*3] このあたりは線形代数の教科書で各自勉強して欲しい。

がわかる。一方で当然ながら

$$D_{\hat{x}}f(x^*) = D_{\hat{x}}f(x^*)(D_{\hat{x}}\hat{g}(x^*))^{-1}D_{\hat{x}}\hat{g}(x^*)$$

なので、これを合わせれば、

$$Df(x^*) = D_{\hat{x}}f(x^*)(D_{\hat{x}}\hat{g}(x^*))^{-1}D\hat{g}(x^*)$$

がわかるので、

$$\hat{\lambda}^* = -([D_{\hat{x}}f(x^*)(D_{\hat{x}}\hat{g}(x^*))^{-1}]^T)$$

と定義し、$\bar{\lambda}^* = 0$ とすれば、

$$Df(x^*) + (\lambda^*)^T Dg(x^*) = 0$$

となって、

$$D_x L(x^*, \lambda^*) = 0$$

がわかる。もちろん、$g_i(x^*) \neq 0$ ならば $\lambda_i^* = 0$ なので、$\lambda_i^* g_i(x^*) = 0$ も成り立つ。

残っている問題は $\lambda_i^* \geq 0$ を示すことだけである。議論はすべて同じなので、$\lambda_1^* \geq 0$ だけを示す。まず $1 \leq i \leq k$ に対して

$$Dg_i(x^*)y = \begin{cases} 1 & \text{if } i = 1, \\ 0 & \text{otherwise.} \end{cases}$$

となるベクトル y を取る（このようなベクトルが存在することは一次独立性から示せるが、今回はその証明は省略する）。そして、ふたたび $\hat{g} = (g_1, ..., g_k)$ として、

$$h(z, \eta) = \hat{g}(x^* + \eta y + (D\hat{g}(x^*))^T z) - \eta D\hat{g}(x^*)y$$

と定義する。このとき、

$$h(0, 0) = \hat{g}(x^*) = 0,\ D_z h(0, 0) = D\hat{g}(x^*)(D\hat{g}(x^*))^T$$

となる。この $D_z h(0, 0)$ は逆行列を持つことが簡単に証明できる[*4]。よって 0 の付近で定義されたある連続微分可能な関数 ξ が存在して、

$$\xi(0) = 0,\ \hat{g}(x^* + \eta y + (D\hat{g}(x^*))^T \xi(\eta)) = \eta D\hat{g}(x^*)y$$

[*4] 一般にある正方行列 A が逆行列を持つ必要十分条件は、$Aa = 0$ となるベクトル a がゼロベクトルのみであることであることが知られている。今回の場合、$A = D\hat{g}(x^*)(D\hat{g}(x^*))^T$ として、$Aa = 0$ とすると、

$$0 = a^T Aa = a^T D\hat{g}(x^*)(D\hat{g}(x^*))a = \|a_1 Dg_1(x^*) + ... + a_k Dg_k(x^*)\|^2$$

が示せるので、

$$a_1 Dg_1(x^*) + ... + a_k Dg_k(x^*) = 0$$

であるが、一次独立性の仮定によって $a_1 = ... = a_k = 0$ であり、よって A は逆行列を持つ。

が常に成り立つ。これを η で微分して $\eta = 0$ と置くと、

$$D\hat{g}(x^*)(y + (D\hat{g}(x^*))^T \xi'(0)) = D\hat{g}(x^*)y$$

がわかるので、整理すると

$$D\hat{g}(x^*)(D\hat{g}(x^*))^T \xi'(0) = 0$$

がわかる。両辺に $D\hat{g}(x^*)(D\hat{g}(x^*))^T$ の逆行列を左から掛けることで、

$$\xi'(0) = 0$$

がわかる。そこで今度は

$$x(\eta) = x^* + \eta y + (D\hat{g}(x^*))^T \xi(\eta)$$

と定義すると、ξ の定義から

$$\hat{g}(x(\eta)) = \eta D\hat{g}(x^*)y = (\eta, 0, ..., 0)$$

である。よって $\eta \geq 0$ が十分 0 に近いならば、$x(\eta)$ は問題 (3) の制約条件を満たす。x^* は問題 (3) の最大点なので、そのような η に対して

$$f(x(\eta)) \leq f(x^*) = f(x(0))$$

であり、よって、

$$Df(x^*)y = Df(x^*)x'(0) = \lim_{\eta \downarrow 0} \frac{f(x(\eta)) - f(x(0))}{\eta} \leq 0$$

がわかる。一方で、

$$Df(x^*) + (\lambda^*)^T Dg(x^*) = 0$$

なので、この両辺に y を掛けると、

$$Df(x^*)y + \lambda_1^* = 0$$

がわかり、よって

$$\lambda_1^* = -Df(x^*)y \geq 0$$

である。以上で証明が完成した。 ■

続いて 2. から 1. への証明を行う。こちらは上の複雑な証明よりは、少しばかり簡単である。

定理２：A は凸集合で、x^* は A の内部に属しているとし、f, g は共に準凹で連続[*5]、x^* で微分可能であるとし、また (4) 式の L が、ある $\lambda^* \ge 0$ に対して、$D_x L(x^*, \lambda^*) = 0^T$ であり、かつ $\lambda_i^* g_i(x^*) = 0$ がすべての $i = 1, ..., m$ について成り立つとする。さらに $Df(x^*) \neq 0$ であるとする。このとき、x^* は問題 (3) の最大点である。

注意：A が凸集合であるという条件があるが、これは以前には A に当たるものが $\mathbb{R}^n_+$ という、当たり前のように凸集合であるものだったから必要がなかった。一方、$Df(x^*) \neq 0$ が必要なのは相変わらずである。また x^* が A の内部に属しているという条件が、どうしても必要であった。これは後に出てくる z が A に属していないと証明が破綻するからであるが、以前は x^* よりほんの少しだけ少ない数を取れば z の用を足せたので、やはり必要なかった。

証明：いま $x \in A, g(x) \ge 0$ であるとする。このとき、$x(t) = (1-t)x^* + tx$ と定義すると、準凹性の条件から $0 \le t \le 1$ のときに

$$g_i(x(t)) \ge \min\{g_i(x), g_i(x^*)\} \ge 0$$

となるので、$x(t)$ は (3) の制約条件を満たす。ここで $g_i(x^*) = 0$ ならば $g_i(x(t)) \ge g_i(x^*)$ なので、

$$Dg_i(x^*)(x - x^*) = \lim_{t \downarrow 0} \frac{g_i(x(t)) - g_i(x(0))}{t} \ge 0$$

である。よって特に、

$$\sum_{i=1}^m \lambda_i^* Dg_i(x^*)(x - x^*) \ge 0$$

がわかる。一方で、$D_x L(x^*, \lambda^*) = Df(x^*) + (\lambda^*)^T Dg(x^*) = 0$ なので、ここから

$$Df(x^*)(x - x^*) = -\sum_{i=1}^m \lambda_i^* Dg_i(x^*)(x - x^*) \le 0$$

がわかることになる。

さて、$c > 0$ を十分小さく取り、$w = -cDf(x^*)$ として、$z = x^* + w$ と定義しよう。c が十分小さければ $z \in A$ が成り立つ。明らかに

$$Df(x^*)(z - x^*) = Df(x^*)w = -c\|w\|^2 < 0$$

[*5] g が準凹というのは、g_i がすべて準凹という意味である。

であることに注意が必要である。ここで、

$$y(t) = (1-t)x + tz,\ x(t) = (1-t)x^* + tz$$

と定義する。もし $0 < t < 1$ であれば、

$$Df(x^*)(x(t) - x^*) = tDf(x^*)(z - x^*) < 0,$$

$$Df(x^*)(y(t) - x(t)) = (1-t)Df(x^*)(x - x^*) \leq 0$$

となるので、まとめると

$$Df(x^*)(y(t) - x^*) < 0$$

がわかる。

ここでもし $f(y(t)) > f(x^*)$ であったとすれば、$z(s) = (1-s)x^* + sy(t)$ とすると、

$$f(z(s)) \geq \min\{f(y(t)), f(x^*)\} = f(x^*)$$

なので、

$$0 > Df(x^*)(y(t) - x^*) = \lim_{s \downarrow 0} \frac{f(z(s)) - f(z(0))}{s} \geq 0$$

となって矛盾が生ずる。よって、$f(y(t)) \leq f(x^*)$ である。$t \to 0$ とすれば f の連続性から

$$f(x) \leq f(x^*)$$

を得ることになるが、x は (3) の制約条件を満たす任意の点であったのだから、x^* は問題 (3) の最大点である。以上で証明が完成した。 ■