最適化数学　講義ノート 4　(担当: 関口 良行)

1　法線錐の公式

1.1　凸集合

最小化問題において, 目的関数が凸関数で実行可能領域が閉かつ凸な集合であれば, 実行可能領域に関する停留点は, 大域最適解になる. これを証明しよう.

例 1.

凸関数の例; $J(x,y)=x^2+y^2$, $J(\boldsymbol{x})=A\boldsymbol{x}$ (任意の実行列 A).
実対称行列 A が半正定値の時, $J(\boldsymbol{x})=\boldsymbol{x}^TA\boldsymbol{x}+\langle \boldsymbol{a},\boldsymbol{x}\rangle+b$

凸集合の例; $\{(x,y)\in\mathbb{R}^n \mid a\le x\le b, c\le y\le c\}$, $\{(x,y)\in\mathbb{R}^2 \mid x^2+y^2\le 1\}$, $\{\boldsymbol{x}\in\mathbb{R}^n \mid A\boldsymbol{x}=\boldsymbol{b}\}$ (A: 行列, $\boldsymbol{b}$: 数ベクトル)

命題 1 (定義その 2)**.** *$J\colon\mathbb{R}^n\to\mathbb{R}$ を滑らかな関数とする. すると,*

$$
\begin{aligned}
&J \text{ が凸関数}\\
&\qquad\Longleftrightarrow \text{任意の } \boldsymbol{x},\boldsymbol{y}\in\mathbb{R}^n \text{ に対して, } J(\boldsymbol{y})\ge J(\boldsymbol{x})+\langle\nabla J(\boldsymbol{x}),\boldsymbol{y}-\boldsymbol{x}\rangle.
\end{aligned}
$$

この講義では充分滑らかな関数のみを扱うので, この命題の条件で凸関数を定義して良い.

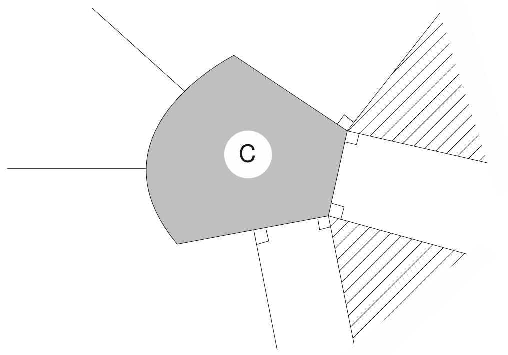
また, 凸集合に対しては法線錐の簡単な公式がある.

命題 2. *$C\in\mathbb{R}^n$ が閉凸集合であるとき, C の $\bar{\boldsymbol{x}}\in C$ における法線錐 $N_C(\bar{\boldsymbol{x}})$ は*

$$N_C(\bar{\boldsymbol{x}})=\{\boldsymbol{v}\in\mathbb{R}^n \mid \langle\boldsymbol{v},\boldsymbol{x}-\bar{\boldsymbol{x}}\rangle\le 0, \boldsymbol{x}\in C\}$$

と書ける.

凸集合の場合, 命題 2 からわかるように, 任意の $\boldsymbol{x}\in C$ に対して, 法線ベクトル $\boldsymbol{v}$ は $\boldsymbol{x}-\bar{\boldsymbol{x}}$ と鈍角をなすベクトルになる.



凸集合 C に対する法線錐 $N_C(\bar{x})$ の例. C の外向きに直角に出ているベクトルの集合が法線錐になる. 図ではイメージしやすいように $\bar{x}$ に平行移動した $N_C(\bar{x})+\bar{x}$ を書いている.

定理 3. 最小化問題

$$(P)\ \text{最小化}\ J(\boldsymbol{x})$$
$$\text{制約}\ \boldsymbol{x} \in C$$

で, J を凸関数, C を凸集合とする. $\bar{\boldsymbol{x}} \in C$ を 問題 (P) の C に関する停留点とすると, $\bar{\boldsymbol{x}}$ は大域最小解になる.

証明. 点 $\bar{\boldsymbol{x}} \in C$ を C に関する停留点とする. 定義より, この点は $-\nabla J(\bar{\boldsymbol{x}}) \in N_C(\bar{\boldsymbol{x}})$ を満たす. ここで命題 2 の凸集合に対する法線錘の公式を用いると, 任意の $\boldsymbol{x} \in C$ に対して,

$$\langle -\nabla J(\bar{\boldsymbol{x}}), \boldsymbol{x} - \bar{\boldsymbol{x}} \rangle \leq 0$$

が成り立つ.

一方, 命題 1 より, 任意の $\boldsymbol{x} \in C$ に対して,

$$J(\boldsymbol{x}) \geq J(\bar{\boldsymbol{x}}) + \langle \nabla J(\bar{\boldsymbol{x}}), \boldsymbol{x} - \bar{\boldsymbol{x}} \rangle$$

が成り立つ. 従って一つ目の不等式より, $J(\boldsymbol{x}) \geq J(\bar{\boldsymbol{x}})$ が成り立ち, これは $\bar{\boldsymbol{x}}$ が問題 (P) の大域最小解であることを表す. □

1.2 凸計画の例題

実行可能領域に関する停留点の求め方を, 復習のため再度例題を用いて学ぶ.

例 2.

$$\text{最小化}\ J(x, y) = 3x^2 + 4y^2 + 4xy - 42x - 44y$$
$$\text{制約}\ (x, y) \in C := \{(x, y) \in \mathbb{R}^2 \mid x \in [0, 4], y \in [0, 4]\}$$

勾配ベクトルとヘッセ行列を計算すると, それぞれ $\nabla J(x, y) = (6x + 4y - 42, 4x + 8y - 44)$, $\nabla^2 J(x, y) = \begin{bmatrix} 6 & 4 \\ 4 & 8 \end{bmatrix}$ となる. $J_x x = 6 > 0$ と $|\nabla^2 J(x, y)| = 32$ なので, $\nabla^2 J(x, y)$ は任意の (x, y) に対して正定値になる. よって, J は凸関数である. また, 実行可能領域 C は凸集合であることがわかる. 従って, 上記の最小化問題は凸計画になるので, C に関する停留点は, 大域最小解になる.

C に関する停留点を求める. $-\nabla J(x, y) \in N_C(x, y)$ を書き下すと

$$-J_x(x, y) = -6x - 4y + 42 \in N_{[0,4]}(x)$$
$$-J_y(x, y) = -4x - 8y + 44 \in N_{[0,4]}(y)$$

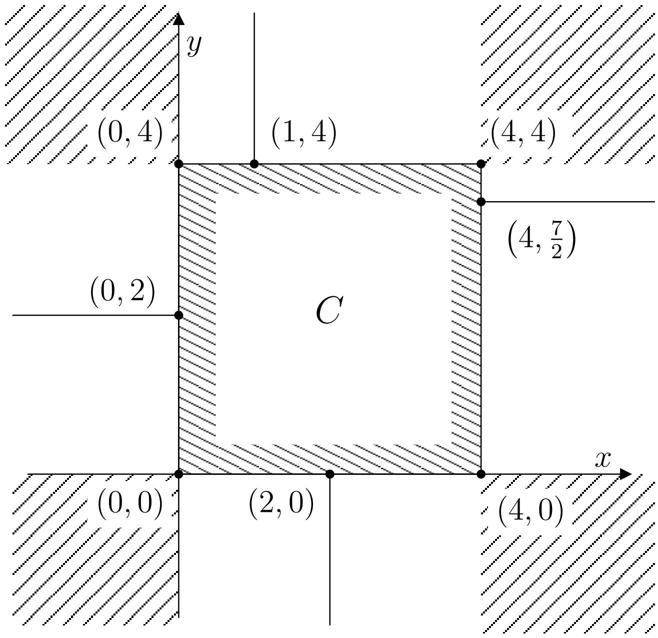
となる. $(x, y) \in C$ ならば,

$$-J_x(4, 4) = 2 \leq -J_x(x, y) \leq 42 = -J_x(x, y)$$
$$-J_y(4, 4) = -4 \leq -J_y(x, y) \leq 44 = -J_y(x, y)$$

となる．ここで，任意の $(x,y)\in C$ に対して，特に $-J_x(x,y)>0$ となるので，$-J_x(x,y)\in N_{[0,4]}(x)$ が満たされるのは，$x=4$ のときのみになる．(必要があれば下図参照)

次に，$x=4$ を代入して，$-J_y(4,y)=-8y+28\in N_{[0,4]}(y)$ を満たす y を探す．y の候補として区間 $[0,4]$ の端点を考えると，それぞれ $-J_y(4,4)<0$, $N_{[0,4]}(4)=[0,\infty)$ と，$-J_y(0,4)>0$, $N_{[0,4]}(0)=(-\infty,0]$ となり，上式を満たさない．y の候補として，$0<y<4$ を満たす点を考える．この場合 $N_{[0,4]}(y)=\{0\}$ となるので，$-J_y(4,y)=0$ かつ $0<y<4$ を満たす点を探すと，$y=28/8=7/2$ が求まる．実際に $(4,7/2)$ に対して，C に関する停留点の式を確認すると，$-\nabla J(4,7/2)=(4,0)$, $N_{[0,4]}(4)=[0,\infty)$, $N_{[0,4]}(7/2)=\{0\}$ となり，$-\nabla J(4,7/2)\in N_C(4,7/2)$ を満たす．

従って，C に関する停留点は $(x,y)=(4,7/2)$ となり，そこでの目的関数値は $J(4,7/2)=-169$ となる．この問題は凸計画なので $(4,7/2)$ は大域最小解になる．



C に対して，図上の各点における $N_C(x,y)+(x,y)$ を図示した様子

1.3　等式制約

充分滑らかな関数 $f\colon\mathbb{R}^n\to R$ に対して，$C=\{\bar{\boldsymbol{x}}\in\mathbb{R}^n\mid f(\bar{\boldsymbol{x}})=0\}$ とする．この場合，接錘と法線錘がきちんと計算できる．

補題 4. $C=\{\boldsymbol{x}\in\mathbb{R}^n\mid f(\boldsymbol{x})=0\}$, $\bar{\boldsymbol{x}}\in C$ とする．このとき $\nabla f(\bar{\boldsymbol{x}})\neq(0,\ldots,0)$ ならば，

$$T_C(\bar{\boldsymbol{x}})=\{\boldsymbol{u}\in\mathbb{R}^n\mid\langle\nabla f(\bar{\boldsymbol{x}}),\boldsymbol{u}\rangle=0\}$$
$$N_C(\bar{\boldsymbol{x}})=\{\lambda\nabla f(\bar{\boldsymbol{x}})\mid\lambda\in\mathbb{R}\}.$$

証明. 証明は少し長いので付録で行う.

$\square$

定理 5 (ラグランジュの乗数法)**.** 最小化問題

$$\begin{aligned} \text{最小化}\ & J(\bar{\boldsymbol{x}}) \\ \text{制約}\ & f(\bar{\boldsymbol{x}}) = 0 \end{aligned}$$

を考える. $\bar{\boldsymbol{x}}$ を局所最小解とする. $\nabla f(\bar{\boldsymbol{x}}) \neq (0,\ldots,0)$ ならば, ある数 λ が存在して,

$$\nabla J(\bar{\boldsymbol{x}}) = \lambda \nabla f(\bar{\boldsymbol{x}})$$

が成り立つ.

証明. 基本最適性条件より, $\bar{\boldsymbol{x}}$ は $-\nabla J(\bar{\boldsymbol{x}}) \in N_C(\bar{\boldsymbol{x}})$ を満たす. $\nabla f(\bar{x}) \neq (0,\ldots,0)$ なので, 補題 4 の法線錘の公式より, ある $\lambda_0 \in \mathbb{R}$ が存在して, $-\nabla J(\bar{\boldsymbol{x}}) = \lambda_0 \nabla f(\bar{\boldsymbol{x}})$ を満たす. よって, $\lambda = -\lambda_0$ とおけば良い.

$\square$

補足. この定理は $C = \{\boldsymbol{x} \in \mathbb{R}^n \mid f(\boldsymbol{x}) = 0\}$ とすると,

$$\bar{\boldsymbol{x}}\ \text{が}\ C\ \text{に関する停留点} \Leftrightarrow \text{ある数}\ \lambda\ \text{が存在して},\ J(\bar{\boldsymbol{x}}) = \lambda \nabla f(\bar{\boldsymbol{x}})$$

ということを表している.

補足. $C = \{\bar{\boldsymbol{x}} \in \mathbb{R}^n \mid f(\bar{\boldsymbol{x}}) = 0\}$ のように与えられる集合は凸でない場合が多いので, C に関する停留点は大域最適解になるとは限らない. 例えば, 円上の点の集合 $C = \{(x,y) \in R^n \mid x^2 + y^2 = 1\}$ は凸でない. 一方, $C = \{(x,y) \in R^n \mid ax + by = 1\}$ (a, b は定数) は凸である.

証明はしないが (実数の性質を使う) 最適解の存在定理を一つ挙げておく.

定理 6. 目的関数が連続で, 実行可能領域が有界閉集合ならば, 最適化問題は最小解と最大解を持つ.

制約付き最適化問題で特に等式制約を持つ問題は, 実行可能領域が有界閉集合になる場合が多い. 最適解の存在があらかじめ分かっていると議論が楽になり, 1 階微分の情報 (基本最適性条件) のみで, 最適解を探せることも多い.

1.4 等式制約付き最小化問題の解き方

例 3.

$$\begin{aligned} \text{最小化}\ & J(x,y) := x - y \\ \text{制約}\ & (x,y) \in C := \{(x,y) \in \mathbb{R}^2 \mid 2x^2 + 3y^2 - 1 = 0\} \end{aligned}$$

基本最適性条件より, C に関する停留点を求める．定理 5 (直後の補足にも注意) の適用を目指す．まず $f(x,y) = 2x^2 + 3y^2 - 1$ とおく．すると勾配ベクトルは $\nabla f(x,y) = (4x, 6y)$ となる．ここで, C 上には $\nabla f(x,y)$ をゼロベクトルにする点はないので, 任意の $(x,y) \in C$ に対して定理 5 が適用できる．C に関する停留点を見つけるには,

$$
\begin{aligned}
J_x(x,y) = 1 = \lambda 4x \\
J_y(x,y) = -1 = \lambda 6y
\end{aligned}
$$

を満たす $(x,y) \in C, \lambda$ を見つければ, その (x,y) が求める点になる (λ は存在すれば良く, 必ずしも計算する必要はない).

第一式より $4x \neq 0$ となるので, $\lambda = 1/(4x)$ として J_y の式に代入すると, $y = -2x/3$ を得る．これを $f(x,y) = 0$ に代入すると, $x = \pm\sqrt{3/10}$ が求まる．このとき $y = \mp\sqrt{2/15}$ となる．よって C に関する停留点は $(x,y) = (\pm\sqrt{3/10}, \mp\sqrt{2/15})$ になる．

さて, C は楕円を表すので有界閉集合になり, 定理 6 よりこの問題は最小解を持つ．最小解が存在することが分かれば, それは C に関する停留点の内のどれかになる (講義ノート 2 の包含関係を表す図を参照). よって, $(x,y) = (\pm\sqrt{3/10}, \mp\sqrt{2/15})$ での目的関数 J の値を調べれば, $(-\sqrt{3/10}, \sqrt{2/15})$ が最小解であることが分かる．

例 4 (2 次形式の最大値, 最小値)**.** A を実対称行列として以下の最小化問題を考える．

$$
\begin{aligned}
(P)\ \text{最小化}\ & J(x,y) := \begin{bmatrix} x & y \end{bmatrix} A \begin{bmatrix} x \\ y \end{bmatrix} \\
\text{制約}\ & (x,y) \in C := \{(x,y) \in \mathbb{R}^n \mid x^2 + y^2 = 1\}
\end{aligned}
$$

最小値は A の最小固有値, 最小解はその固有ベクトルになる．

まず C は有界閉集合なので, 定理 6 より最小解が存在することが分かる．それを $(\bar{x}, \bar{y})$ とおく．$f(x,y) = x^2 + y^2 - 1$ とすると, 任意の $(x,y) \in C$ に対して, $\nabla f(x,y)$ はゼロベクトルにならない．よって, 定理 5 より, ある $\lambda \in \mathbb{R}$ が存在して, $\nabla J(\bar{x}, \bar{y}) = \lambda \nabla f(\bar{x}, \bar{y})$ が成り立つ．この式より,

$$
A \begin{bmatrix} \bar{x} \\ \bar{y} \end{bmatrix} = \lambda \begin{bmatrix} \bar{x} \\ \bar{y} \end{bmatrix}
$$

を得る．これは λ が A の固有値で, $(\bar{x}, \bar{y})$ がその固有ベクトルであることを示す．よって, C に関する停留点の集合と A の固有ベクトルの集合は等しい．

ここで, $(\bar{x}, \bar{y})$ が A の固有ベクトルのとき, $J(\bar{x}, \bar{y}) = \lambda(\bar{x}^2 + \bar{y}^2) = \lambda$ となるので, C に関する停留点の中で J を最小にするのは, $(\bar{x}, \bar{y})$ が A の最小固有値に対する固有ベクトルのときであり, そのとき J の値は A の最小固有値と等しくなる．

2　付録

2.1　定理 4 の証明

まず $\boldsymbol{u} \in T_C(\bar{\boldsymbol{x}}) \Rightarrow \langle \nabla f(\bar{\boldsymbol{x}}), \boldsymbol{u} \rangle = 0$ を示す．$\boldsymbol{u} \in T_C(\bar{\boldsymbol{x}})$ とする．すると, ある関数 $\boldsymbol{x}\colon [0, \varepsilon] \to C$ で, $\boldsymbol{x}(0) = \bar{\boldsymbol{x}}$ かつ $\boldsymbol{x}'_+(0) = \boldsymbol{u}$ となるものが存在する．ここで, $\boldsymbol{w}(t) = t^{-1}(\boldsymbol{x}(t) - \boldsymbol{x}(0))$ とおくと, $\lim_{t\to+0} \boldsymbol{w}(t) = \boldsymbol{u}$ が成り立つ．いま, $\boldsymbol{x}(t) \in C$ より, $f(\boldsymbol{x}(t)) = 0$ なので, テーラー展開を適用すると,

$$0 = f(\boldsymbol{x}(t)) = f(\bar{\boldsymbol{x}} + t\boldsymbol{w}(t)) = f(\bar{\boldsymbol{x}}) + \langle \nabla f(\bar{\boldsymbol{x}}), t\boldsymbol{w}(t) \rangle + o(\|t\boldsymbol{w}(t)\|)$$

を得る．これより $\langle \nabla f(\bar{\boldsymbol{x}}), t\boldsymbol{w}(t) \rangle + o(\|t\boldsymbol{w}(t)\|) = 0$ が成り立つので, 両辺を t で割り, $t \to +0$ とすると, $\langle \nabla f(\bar{\boldsymbol{x}}), \boldsymbol{u} \rangle = 0$ を得る．

逆に $\langle \nabla f(\bar{\boldsymbol{x}}), \boldsymbol{u} \rangle = 0 \Rightarrow \boldsymbol{u} \in T_C(\bar{\boldsymbol{x}})$ を示す．仮定より $\nabla f(\bar{\boldsymbol{x}}) \neq 0$ なので, $\boldsymbol{v} \in \mathbb{R}^n$ で $\langle f(\bar{\boldsymbol{x}}), \boldsymbol{v} \rangle \neq 0$ を満たすものが存在する．ここで, $h(s, t) = f(\bar{\boldsymbol{x}} + s\boldsymbol{v} + t\boldsymbol{u})$ とおく．すると, h は 2 変数関数で, $h(0, 0) = 0$, $h_s(0, 0) = \langle \nabla f(\bar{\boldsymbol{x}}), \boldsymbol{v} \rangle \neq 0$ となるので, 陰関数定理より, ある微分可能な関数 $\varphi(t)$ が存在して, $\varphi(0) = 0$ かつ 0 に充分近い t に対して $h(\varphi(t), t) = f(\bar{\boldsymbol{x}} + \varphi(t)\boldsymbol{v} + t\boldsymbol{u}) = 0$ が成り立つ．ここで, $\boldsymbol{x}(t) = \bar{\boldsymbol{x}} + \varphi(t)\boldsymbol{v} + t\boldsymbol{u}$ とおくと, $\boldsymbol{x}(0) = \bar{\boldsymbol{x}}$ で $\boldsymbol{x}(t) \in C$ となる．さらに, $\boldsymbol{x}'(0) = \boldsymbol{u}$ となることを示す．$h(\varphi(t), t) = 0$ の両辺を t で微分してから $t = 0$ を代入すると,

$$\left.\frac{d}{dt}h(\varphi(t), t)\right|_{t=0} = h_s(\varphi(0), 0)\varphi'(0) + h_t(\varphi(0), 0) = \langle \nabla f(\bar{\boldsymbol{x}}), \varphi'(0)\boldsymbol{v} \rangle + \langle \nabla f(\bar{\boldsymbol{x}}), \boldsymbol{u} \rangle = 0$$

よって, $\varphi'(0)\langle \nabla f(\bar{\boldsymbol{x}}), \boldsymbol{v} \rangle = 0$ となり, $\boldsymbol{v}$ の選び方から, $\varphi'(0) = 0$ を得る．よって $\boldsymbol{x}'(0) = \boldsymbol{u}$ となるので, $\boldsymbol{u} \in T_C(\bar{\boldsymbol{x}})$ が示された．

次に法線錘の公式を示す．ベクトル $\nabla f(\bar{\boldsymbol{x}})$ によって, 張られる $\mathbb{R}^n$ の部分空間を $\langle \nabla f(\bar{\boldsymbol{x}}) \rangle$, その直交補空間を $\langle \nabla f(\bar{\boldsymbol{x}}) \rangle^\perp$ とする．接錘の公式より, $T_C(\bar{\boldsymbol{x}}) = \langle \nabla f(\bar{\boldsymbol{x}}) \rangle^\perp$ となる．いま $T_C(\bar{\boldsymbol{x}})$ は部分空間なので $\boldsymbol{u} \in T_C(\bar{\boldsymbol{x}})$ ならば $-\boldsymbol{u} \in T_C(\bar{\boldsymbol{x}})$ となることに注意すると, 法線錘の定義より, $N_C(\bar{\boldsymbol{x}}) = \{\boldsymbol{v} \in \mathbb{R}^n \mid \langle \boldsymbol{v}, \boldsymbol{u} \rangle = 0, \boldsymbol{u} \in T_C(\bar{\boldsymbol{x}})\}$ となる．これより, $N_C(\bar{\boldsymbol{x}}) = T_C(\bar{\boldsymbol{x}})^\perp = (\langle \nabla f(\bar{\boldsymbol{x}}) \rangle^\perp)^\perp = \langle \nabla f(\bar{\boldsymbol{x}}) \rangle = \{\lambda \nabla f(\bar{\boldsymbol{x}}) \mid \lambda \in \mathbb{R}\}$ を得る．