

6 非線形計画問題の例

自然科学, 実社会における決定問題を最適化問題として定式化する場合, 制約条件や目的関数は線形関数で表されるとは限らない. 本節以降しばらく, そのような非線形関数によって表される最適化問題について理解を深めていく.¹

【応用例 3.1】最小 2 乗法 m 個の観測データ $(x_1, y_1), \dots, (x_m, y_m) \in \mathbb{R}^{n+1}$ が与えられたとき, x と y の関係をなるべくうまく説明する線形関数 $y = a^\top x + b$ を求めたい.

もし $y_i = a^\top x_i + b, i = 1, \dots, m$ が成り立っていれば, $y = a^\top x + b$ が求める線形関数となるが, 一般にはこれは成り立たない (すなわち, 少なくともある i に対して $y_i \neq a^\top x_i + b$). そこで

$$y_i \approx a^\top x_i + b, \quad i = 1, \dots, m$$

となるように直線のパラメータ a, b を定める. この $\approx$ の意味をどう定義するかでいろいろな可能性が考えられるが, 広く用いられている手法は次の最適化問題を解くことである.

$$\text{最小化} \quad \sum_{i=1}^m (y_i - a^\top x_i - b)^2 \quad (1)$$

これは誤差の二乗の和を最小にする方法であることから, 最小二乗法と呼ばれている. 最適化問題としては, 無制約の凸 2 次計画問題である.

【応用例 3.1b】最小 2 乗法 (2) m 個の観測データ $(x_1, y_1), \dots, (x_m, y_m) \in \mathbb{R}^2$ が与えられたとき, x と y の関係をなるべくうまく説明する 3 次関数 $y = b_0 + b_1x + b_2x^2 + b_3x^3$ を求めたい.

Exercise 6.1 この問題を最小 2 乗法の形で定式化せよ.

【応用例 3.1c】最小 2 乗法 (3) m 個の観測データ $(x_1, y_1), \dots, (x_m, y_m) \in \mathbb{R}^2$ が与えられたとき, x と y の関係をなるべくうまく説明する関数 $y = b_0 + b_1 \exp(b_2x)$ を求めたい.

Exercise 6.2 この問題を最小 2 乗法の形で定式化せよ.

【応用例 3.2】施設配置問題 平面上に相異なる $m (\geq 2)$ 個の需要点 $(x_1, y_1), \dots, (x_m, y_m)$ が散らばっているとす. このとき, 共通に利用できる施設 (例えば, 病院) を 1 つだけ設置する.

- (1) 各需要点からのユークリッド距離の総和が最小になる配置を求める問題を定式化せよ.
- (2) 各需要点からのユークリッド距離の 2 乗の総和が最小になる配置を求める問題を定式化せよ.
- (3) 施設から最も遠い需要点のユークリッド距離が最小になる配置を求める問題を定式化せよ.

Exercise 6.3 (1), (2), (3) の定式化をそれぞれ記せ.

【応用例 3.3】最小包囲円問題 平面上に相異なる $m (\geq 2)$ 個の点 $(x_1, y_1), \dots, (x_m, y_m)$ が散らばっているとす. このとき, この m 個の点を含み, しかも面積が最小になる円を求める問題を最小包囲円問題と呼ぶ.

Exercise 6.4 最小包囲円問題を定式化せよ.

¹ここに紹介する議論は LP の一般化でもあるので, 「OR 第 2」の LP の議論との結びつきを考えていくとよい.

Exercise 6.5 (最小包囲球問題) 最小包囲円問題を n 次元空間の問題に拡張する. $\mathbb{R}^n$ 上に相異なる m 個の点 $\mathbf{x}_1, \dots, \mathbf{x}_m$ が散らばっているとす ($m \geq n$). このとき, この m 個の点を含み, しかも体積 (に相当するもの) が最小になる (超) 球を求める問題を定式化せよ.

【応用例 3.4】最尤推定問題 I 独立な試行に基づき, 1 次元正規分布に従う m 個の観測値 $x_1, \dots, x_m$ が得られた状況において, 正規分布のパラメータ μ と σ を推定したい. このとき, 密度関数 $f(x|\mu, \sigma) = \frac{1}{\sqrt{2\pi}\sigma} \exp(-\frac{(x-\mu)^2}{2\sigma^2})$ に対して, μ, σ を決定変数と考えて, 以下で定義される尤度 L を最大にする μ, σ をパラメータの推定値とする方法を _____ と呼ぶ^a.

$L(\mu, \sigma) = \underline{\hspace{2cm}}$: 尤度関数.

^aここでは正規分布を仮定しているが, 一般に最尤法は正規分布に限らない.

すなわち, $\max L(\mu, \sigma)$ を行うが, 密度 f が常に正であること, 対数関数 $\ln z$ が単調増加であることから,

$$\max \ln L(\mu, \sigma) = \max \sum_{i=1}^m \underline{\hspace{2cm}}$$

を行うことと等価である (対数尤度最大化) .

Exercise 6.6 上記の 1 次元正規分布の対数尤度最大化問題を定式化せよ.

Exercise 6.7 n 次元正規分布 $N(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ の対数尤度最大化問題を定式化せよ.

【応用例 3.5】最尤推定問題 II 「ある通勤客が 2 つの交通手段 (たとえば, バスと電車) のいずれを使うか」「ある企業が倒産するか存続するか」「顧客がリピーターとなるか一見さんで終わるか」というように, 2 つの事象のうちのいずれの 1 つが起こるかの確率をその対象 (通勤客, 企業, 顧客) の属性 $\mathbf{x} := (x_1, \dots, x_n)^\top$ から求めたい. その 1 つの方法が 2 項ロジット・モデル (ロジスティック回帰モデル) である. これは 2 つの選択肢のラベル y を “1” (バス, 倒産, リピーター) と “0” (電車, 存続, 一見さん) で表したとして,

$$\mathbb{P}\{y = 1|\mathbf{x}\} = \frac{\exp(b + \mathbf{a}^\top \mathbf{x})}{1 + \exp(b + \mathbf{a}^\top \mathbf{x})}, \quad \mathbb{P}\{y = 0|\mathbf{x}\} = \frac{1}{1 + \exp(b + \mathbf{a}^\top \mathbf{x})}$$

として表すことを仮定する.

独立な試行に基づき m 個の観測値 $\mathbf{x}_1, \dots, \mathbf{x}_m \in \mathbb{R}^n$ とそれらのラベル $y_1, \dots, y_m \in \{0, 1\}$ が得られた状況において, 最尤推定法により $(b, \mathbf{a})$ を推定したい. このとき尤度関数 L は

$$L(\mathbf{a}, b) := \prod_{i=1}^m p_i^{y_i} (1 - p_i)^{1-y_i}$$

ただし,

$$p_i := \mathbb{P}\{y = 1|\mathbf{x}_i\}$$

である. 通常, 尤度関数 L を最大化する代わりに対数尤度 $\ln L = \underline{\hspace{2cm}}$ の最大化を考える.

【応用例 3.6】非線形費用付き生産・輸送問題 既述の輸送問題に以下の仮定を追加した、非線形費用付き生産・輸送問題を考える：

- 輸送問題においては1単位あたり輸送コストは一定 $c_{ij} > 0$ であり, x_{ij} 単位の輸送コストは $c_{ij}x_{ij}$ であった. ここでは, 輸送コストが線形でなく, $c_{ij}(x_{ij}) = \gamma_{ij} \ln(1 + x_{ij})$ で表されるとする ($\gamma_{ij} > 0$) .
- 輸送問題においては工場の供給量 $a_i > 0$ が与えられていた. ここでは, 供給量も決定変数とし, y_i 単位の生産に対して $g_i(y_i) = d_i(\exp(y_i) - 1)$ コストがかかるものとする ($d_i > 0$) .
- 目的は, 輸送コストと生産コストの合計の最小化.
- 生産量 y_i は u_i を超えない.

Exercise 6.8 この非線形費用付き生産・輸送計画問題を定式化せよ.

この問題は _____ (nonconvex) の最適化問題になるため, 最適性の保証付き解法については, 本講義の範囲を超える.

【応用例 3.7】ポートフォリオ選択問題 3種類の証券(1単位あたり)の現在の価格が P_1, P_2, P_3 , 1年後の価格が確率変数 F_1, F_2, F_3 で与えられているとする. 証券 i を x_i 単位購入すると, 資金の1年後の価値は

$$W(\mathbf{x}) := \underline{\hspace{2cm}}$$

なる確率変数で表現される. 投資に際しては1年後の資金の期待値(リターン)がなるべく大きく, しかもなるべくリスク(不確実性)を抑えられるよう購入単位数 (x_1, x_2, x_3) を決定したい.

典型的にはリスクは資金 $W(\mathbf{x})$ の分散で測られ,

$$f(\mathbf{x}) = \mathbb{V}[W(\mathbf{x})] = \underline{\hspace{2cm}}$$

で表される. また, 資金の期待値は

$$g(\mathbf{x}) := \mathbb{E}[W(\mathbf{x})] = \underline{\hspace{2cm}}$$

と表される. ここで, $\mu_i := \mathbb{E}[F_i]$, $\sigma_{ij} := \mathbb{E}[(F_i - \mu_i)(F_j - \mu_j)]$. 手持ちの資金 $M > 0$ の範囲内で, なるべく大きなリターンを目指しつつ, なるべくリスクを小さくするよう購入単位数を求める問題は以下のような最適化問題として定式化される:

最大化	$g(\mathbf{x}) - Cf(\mathbf{x}) = \underline{\hspace{2cm}}$
条件	$\underline{\hspace{2cm}}$
	$x_i \geq 0, (i = 1, 2, 3).$

ただし, $C > 0$ はリターンとリスクのトレードオフ・パラメータ. このようなモデルを _____ **モデル** (あるいは**平均・分散モデル**) という. 最適化問題としては, (線形制約付き) 凸2次計画問題である.

7 凸集合と凸関数

【用語集】 勾配ベクトル, ヘッセ行列, 凸集合, 凸関数

この節では非線形計画問題を扱う上で基礎となる事項(勾配ベクトル, ヘッセ行列, 凸関数等)をまとめる. これらは非線形計画が“うまく”解けるための条件を理解するのに必要な概念である.

7.1 [準備] 勾配ベクトルとヘッセ行列

1 変数関数 f に対して微分 f' を考えるのと同様に, n 変数関数に対しても微分 (すなわち, 傾き) の概念を拡張したものが勾配ベクトルである:

定義 (勾配ベクトル gradient vector) $f: \mathbb{R}^n \rightarrow [-\infty, \infty]$ が $\mathbf{x} \in \mathbb{R}^n$ の各要素について偏微分可能なとき, $\mathbf{x}$ における f の勾配ベクトル $\nabla f(\mathbf{x})$ を以下のように定義する:

$$\nabla f(\mathbf{x}) := \left(\frac{\partial f(\mathbf{x})}{\partial x_1}, \frac{\partial f(\mathbf{x})}{\partial x_2}, \dots, \frac{\partial f(\mathbf{x})}{\partial x_n} \right)^\top \in \mathbb{R}^n.$$

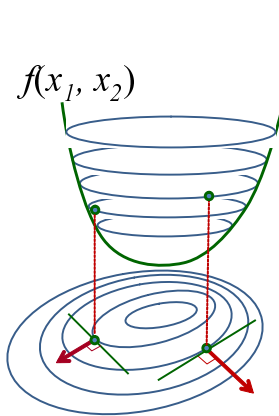
ただし,

$$\frac{\partial f(\mathbf{x})}{\partial x_i} := \lim_{h \rightarrow 0} \frac{f(x_1, \dots, x_i + h, \dots, x_n) - f(x_1, \dots, x_i, \dots, x_n)}{h}$$

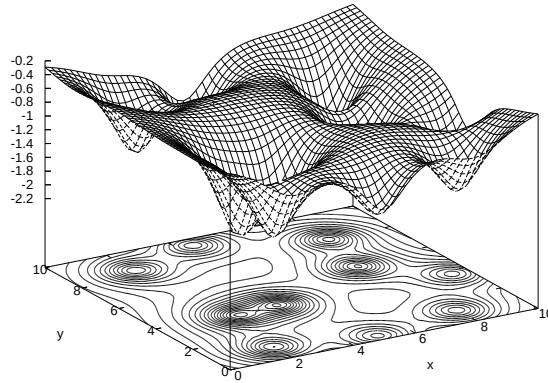
$f(\mathbf{x}, \mathbf{y})$ のように変数を分けて考える場合, 変数 $\mathbf{x} \in \mathbb{R}^n, \mathbf{y} \in \mathbb{R}^\ell$ それぞれについて勾配を考えることがある:

$$\nabla_{\mathbf{x}} f(\mathbf{x}, \mathbf{y}) := \left(\frac{\partial f(\mathbf{x}, \mathbf{y})}{\partial x_1}, \dots, \frac{\partial f(\mathbf{x}, \mathbf{y})}{\partial x_n} \right)^\top \in \mathbb{R}^n; \quad \nabla_{\mathbf{y}} f(\mathbf{x}, \mathbf{y}) := \left(\frac{\partial f(\mathbf{x}, \mathbf{y})}{\partial y_1}, \dots, \frac{\partial f(\mathbf{x}, \mathbf{y})}{\partial y_\ell} \right)^\top \in \mathbb{R}^\ell.$$

Exercise 7.1 $f(x, y) = x^2 + xy + 2y^2$ の $\nabla f(x, y)$ を求めよ. また, (x, y) 平面上に f の等高線を何本か描き, その上のいくつかの点における ∇f を求めよ.



(a) 凸な関数と勾配ベクトル



(b) 高度に非凸な関数の等高線

図 1: $\mathbb{R}^2$ 上の関数と等高線

また, 関数 f が (1 回) 微分可能であるということを勾配ベクトル $\nabla f(\mathbf{x})$ を用いて以下のように定義する。

定義 (f の 1 回微分可能性)

$$\lim_{h \rightarrow 0+} \frac{o(h)}{h} = 0$$

を満たす関数 $o(\cdot): [0, \infty) \rightarrow \mathbb{R}$ が存在して, 任意の $\mathbf{d}$ について,

$$f(\mathbf{x} + \mathbf{d}) = f(\mathbf{x}) + \nabla f(\mathbf{x})^\top \mathbf{d} + o(\|\mathbf{d}\|)$$

が成り立つとき, $\mathbf{x}$ において f が微分可能であると言う。ただし, $\|\mathbf{d}\|$ は $\mathbf{d}$ の (ユークリッド) ノルム。また, $\nabla f(\mathbf{x})$ が $\mathbf{x}$ で連続ならば, 連続的微分可能と言う。

1 変数関数 f に対して 2 階微分 f'' を考えるのと同様に, n 変数関数に対しても 2 階微分 (すなわち, 曲率) の概念を拡張したものがヘッセ行列である:

定義 (ヘッセ行列 Hessian matrix) x において $f: \mathbb{R}^n \rightarrow [-\infty, \infty]$ が 2 次の偏微係数を持つとき、 x における f のヘッセ行列 $\nabla^2 f(x)$ を以下のように定義する:

$$\nabla^2 f(x) := \begin{pmatrix} \frac{\partial^2 f(x)}{\partial x_1^2} & \cdots & \frac{\partial^2 f(x)}{\partial x_1 \partial x_n} \\ \vdots & \frac{\partial^2 f(x)}{\partial x_i \partial x_j} & \vdots \\ \frac{\partial^2 f(x)}{\partial x_n \partial x_1} & \cdots & \frac{\partial^2 f(x)}{\partial x_n^2} \end{pmatrix} \in \mathbb{R}^{n \times n}$$

Exercise 7.2 $f(x, y) = x^2 + xy + 2y^2$ の $\nabla^2 f(x, y)$ を求めよ.

Exercise 7.3 $c \in \mathbb{R}^n$, $Q \in \mathbb{R}^{n \times n}$ とする. このとき, $x \in \mathbb{R}^n$ に関する勾配ベクトル, ヘッセ行列について以下が成り立つことを示せ.

$$(1) \nabla(c^\top x) = c \quad (2) \nabla(x^\top Qx) = (Q + Q^\top)x \quad (3) \nabla^2(x^\top Qx) = Q + Q^\top$$

これより, Q が対称 (i.e., $Q = Q^\top$) な場合, $\nabla(x^\top Qx) = 2Qx$, $\nabla^2(x^\top Qx) = 2Q$.

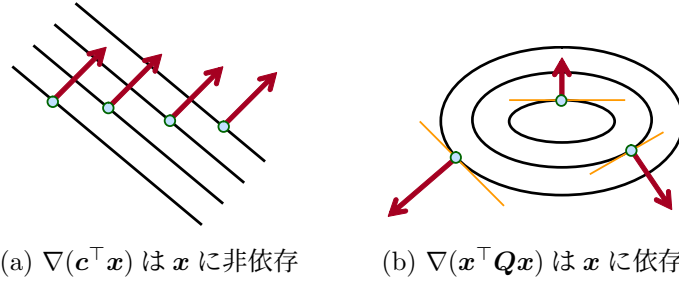


図 2: 等高線と勾配

1 階の微分可能性と同じ要領で、関数 f が 2 回微分可能であることをヘッセ行列と勾配ベクトル $\nabla f(x)$ を用いて以下のように定義する。

定義 (f の 2 回微分可能性) ある $o(\cdot)$ が存在して、任意の d について、

$$f(x + d) = f(x) + \nabla f(x)^\top d + \frac{1}{2} d^\top \nabla^2 f(x) d + o(\|d\|^2)$$

が成り立つとき、 x において f は 2 回微分可能であると言う。

また、 $\nabla^2 f(x)$ が x で連続ならば、連続的 2 回微分可能と言う。

平均値の定理やテイラーの定理なども n 次元に拡張される

平均値の定理 関数 $f: \mathbb{R}^n \rightarrow \mathbb{R}$ が凸集合 $S \subset \mathbb{R}^n$ 上で微分可能とする. このとき、

$$\forall x, y \in S, \exists \tau \in (0, 1): f(x) - f(y) = \nabla f((1 - \tau)x + \tau y)^\top (x - y).$$

テイラーの定理 関数 $f: \mathbb{R}^n \rightarrow \mathbb{R}$ が凸集合 $S \subset \mathbb{R}^n$ 上で連続的 2 回微分可能とする. このとき、

$$\forall x, y \in S, \exists \tau \in (0, 1): f(x) - f(y) = \nabla f(y)^\top (x - y) + \frac{1}{2} (x - y)^\top \nabla^2 f((1 - \tau)x + \tau y) (x - y).$$

実対称行列の正定値性と半正定値性 n 次実数値対称行列 S に対して、以下の 3 つはすべて等価な条件である:

- (1) $\forall y \in \mathbb{R}^n, y^\top S y \geq 0$.
- (2) S の固有値がすべて非負.
- (3) $S = D^2$ をみたす実対称行列 $D \in \mathbb{R}^{n \times n}$ が存在.

これらのいずれかが満たされるとき, S は**半正定値** (positive semidefinite) (非負定値 nonnegative definite とも言う) であるという. また, $(1') \mathbf{y}^\top S \mathbf{y} > 0$ for all $\mathbf{y} \neq \mathbf{0}$ が成り立つとき, **正定値** (positive definite) という. $n = 1$ のときを考えると, S はスカラー s で, $s \geq 0$ が半正定値 ($s > 0$ が正定値) に対応する. 行列の (半) 正定値性はシステムの安定性を記述するのによく用いられる.

Exercise 7.4 (1) 行列 $\begin{pmatrix} 3 & 2 \\ 2 & 2 \end{pmatrix}$ は正定値であることを示せ.

(2) 2×2 の実対称行列で半正定値でない行列を作り, そうでない理由を述べよ.

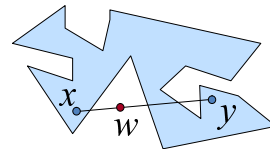
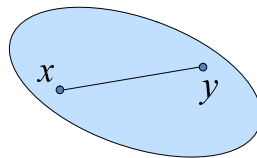
(3) 2×2 の実対称行列で半正定値であるが, 正定値でない行列を 1 つ作り, その理由を述べよ.

7.2 凸集合

まず $\mathbb{R}^n$ 上の集合の凸性から始めよう.

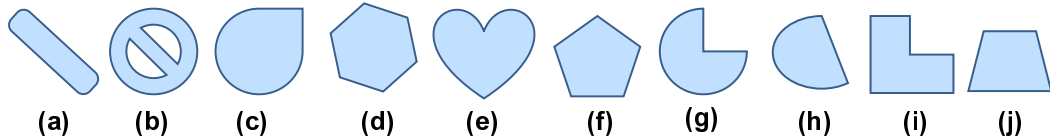
定義 (凸集合 convex set) $X \subset \mathbb{R}^n$ が**凸集合** $\stackrel{\text{def}}{\iff} \mathbf{x}_1, \mathbf{x}_2 \in X, \lambda \in [0, 1] \Rightarrow (1 - \lambda)\mathbf{x}_1 + \lambda\mathbf{x}_2 \in X$

ある $\lambda \in [0, 1]$ に対し, $(1 - \lambda)\mathbf{x} + \lambda\mathbf{y}$ で表される点を点 $\mathbf{x}$ と点 $\mathbf{y}$ の _____ (convex combination) とよぶ. 直感的には, 凸集合はへこみのない集合のこと. 空集合は凸集合.



(a) 凸集合は任意の凸結合が含まれる (b) 凸集合でないと含まれない凸結合が存在

Exercise 7.5 次のように $\mathbb{R}^2$ 上の領域で示される集合 (a)~(j) の中で凸集合はどれか?



Exercise 7.6 次の集合は凸集合かどうか?

- (a) $\{x \in \mathbb{R} : x(x - 1)(x - 2) \leq 0\}$
- (b) $\{\mathbf{x} \in \mathbb{R}^n : \mathbf{a}^\top \mathbf{x} = b\}$ for some $\mathbf{a} \in \mathbb{R}^n, b \in \mathbb{R}$
- (c) $\{(x_1, x_2) \in \mathbb{R}^2 : x_1^2 + x_2^2 \geq 1\}$
- (d) $\{(x_1, x_2) \in \mathbb{R}^2 : x_1 x_2 \geq 1, x_1 \geq 0, x_2 \geq 0\}$
- (e) $\{(x_1, x_2, x_3) \in \mathbb{R}^3 : \sqrt{x_1^2 + x_2^2} \leq x_3\}$

Exercise 7.7 $f : \mathbb{R}^n \rightarrow [-\infty, \infty]$ が凸関数であるとき, $S := \{\mathbf{x} \in \mathbb{R}^n : f(\mathbf{x}) \leq 0\}$ は凸集合となることを示せ.

定理 7.1 凸集合と凸集合の積集合 (交わり) は凸集合.

Proof. 2つの凸集合を F_1, F_2 とする. もし両者の交わりが空集合ならば, 定義よりそれは自明に凸集合であるので, 交わりは空でないと仮定する. $F_1 \cap F_2$ から 2 点 $\mathbf{x}, \mathbf{y}$ をとると, $\mathbf{x}$ と $\mathbf{y}$ を結ぶ線分は (F_1, F_2 はともに凸集合なので) F_1 にも F_2 にも属している. これは $F_1 \cap F_2$ が凸集合であることを表す. $\square$

7.3 凸関数

関数が“でこぼこ”した形をしていると、一般に最小化が難しい．一方、関数の形が下に膨らんでいると、最小化は比較的容易である．この下に膨らんだ形をする関数のことを凸関数とよぶ．より正確に言うと、凸関数とは関数のグラフの「上側」の部分が凸集合になっている関数のことである．

定義 (エピグラフ epigraph) 実数値関数 f が与えられたとき、以下の集合を f の**エピグラフ**とよぶ:
 $\text{epi } f := \{(x, y) \in \mathbb{R}^n \times \mathbb{R} : y \geq f(x)\}.$

定義 (凸関数 convex function) 実数値関数 f のエピグラフが凸集合のとき、 f は**凸関数**.

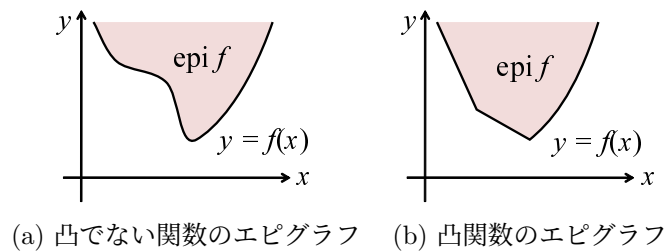


図 3: エピグラフの図示

定理 7.2 $\mathbb{R}^n$ 上の関数 f が凸関数であることは、次の条件と同値である：

任意の点 $x_1, x_2 \in \mathbb{R}^n$ 、任意の $\lambda \in [0, 1]$ に対し、

$$(1 - \lambda)f(x_1) + \lambda f(x_2) \geq f((1 - \lambda)x_1 + \lambda x_2).$$

Exercise 7.8 上の定理を証明せよ．

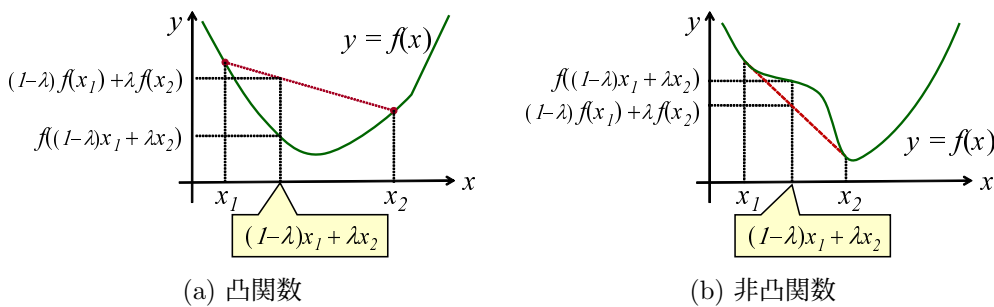


図 4: 「関数値の凸結合」 $\geq$ 「凸結合の関数値」

Exercise 7.9 次の関数は凸関数かどうか？

$$(a) f(x) = x(x - 1)(x - 2) \quad (b) g(x) = -|x| \quad (c) h(x_1, x_2) = x_1 x_2 \quad (d) h(x_1, x_2) = \sqrt{x_1 x_2}$$

定理 7.2 の条件を凸関数の定義としてもよい．下に、凸関数、狭義凸関数、凹関数の定義を挙げておく．

定義 (凸関数 convex function) $f: \mathbb{R}^n \rightarrow [-\infty, \infty]$ が凸関数

$$\stackrel{\text{def}}{\iff} x_1, x_2 \in \mathbb{R}^n, \lambda \in [0, 1] \Rightarrow (1 - \lambda)f(x_1) + \lambda f(x_2) \geq f((1 - \lambda)x_1 + \lambda x_2)$$

定義 (狭義凸関数 strictly convex function) $f: \mathbb{R}^n \rightarrow [-\infty, \infty]$ が狭義凸関数

$$\stackrel{\text{def}}{\iff} x_1, x_2 \in \mathbb{R}^n, \lambda \in (0, 1) \Rightarrow (1 - \lambda)f(x_1) + \lambda f(x_2) > f((1 - \lambda)x_1 + \lambda x_2)$$

定義 (凹関数 concave function) $f: \mathbb{R}^n \rightarrow [-\infty, \infty]$ が凹関数

$$\stackrel{\text{def}}{\iff} -f \text{ が凸関数, i.e., } x_1, x_2 \in \mathbb{R}^n, \lambda \in [0, 1] \Rightarrow (1 - \lambda)f(x_1) + \lambda f(x_2) \leq f((1 - \lambda)x_1 + \lambda x_2)$$

7.4 微分可能な凸関数

微分可能な関数 $f: \mathbb{R} \rightarrow [-\infty, \infty]$ が凸関数であることの必要十分条件は、任意の $x, y \in \mathbb{R}$ について、

$$f(y) \geq f(x) + f'(x)(y - x)$$

が成り立つことである。一般に、 $f: \mathbb{R}^n \rightarrow [-\infty, \infty]$ の微分可能な凸関数ならば、任意の $x, y \in \mathbb{R}^n$ について、

$$f(y) \geq f(x) + \nabla f(x)^\top (y - x) = f(x) + \sum_{j=1}^n \frac{\partial f(x)}{\partial x_j} (y_j - x_j) \quad \dots \text{【勾配不等式】} \quad (2)$$

が成り立つ。

定理 7.3 $f: \mathbb{R}^n \rightarrow [-\infty, \infty]$ が微分可能であるとする。このとき、 f が凸関数であることの必要十分条件は、任意の 2 つのベクトル $x, y \in \mathbb{R}^n$ に対して勾配不等式 (2) が成り立つことである。

Proof. $(\Rightarrow)$ f が凸関数であるとする。このとき任意の $x, y \in \mathbb{R}^n, \lambda \in (0, 1)$ に対して、 $(1 - \lambda)f(x) + \lambda f(y) \geq f((1 - \lambda)x + \lambda y)$ が成り立つので、

$$\begin{aligned} f(y) - f(x) &\geq \frac{1}{\lambda} \{f(x + \lambda(y - x)) - f(x)\} \\ &= \frac{1}{\lambda} \{f(x) + \lambda \nabla f(x)^\top (y - x) + o(\lambda) - f(x)\} \quad (\text{テイラーの定理}) \\ &= \nabla f(x)^\top (y - x) + \frac{1}{\lambda} o(\lambda). \end{aligned}$$

$\lambda \rightarrow +0$ より、勾配不等式を得る。

$(\Leftarrow)$ 勾配不等式が成り立つとする。この勾配不等式において、 x の代わりに $(1 - \lambda)x + \lambda y$ を入れると、

$$f(y) \geq f((1 - \lambda)x + \lambda y) + \nabla f((1 - \lambda)x + \lambda y)^\top [y - \{(1 - \lambda)x + \lambda y\}].$$

同様に、 x と y の役割を入れても成り立つので、

$$f(x) \geq f((1 - \lambda)x + \lambda y) + \nabla f((1 - \lambda)x + \lambda y)^\top [x - \{(1 - \lambda)x + \lambda y\}].$$

これらの式それぞれに λ と $1 - \lambda$ を掛けて辺々加えて整理すると、

$$(1 - \lambda)f(x) + \lambda f(y) \geq f((1 - \lambda)x + \lambda y)$$

となり、 f は凸関数となる。 □

Exercise 7.10 関数 $f(x, y) = x^2 + y^2$ が $\mathbb{R}^2$ 上で凸関数であることを示し、点 $(\bar{x}, \bar{y})$ における接平面の方程式、および、勾配不等式を求めよ。

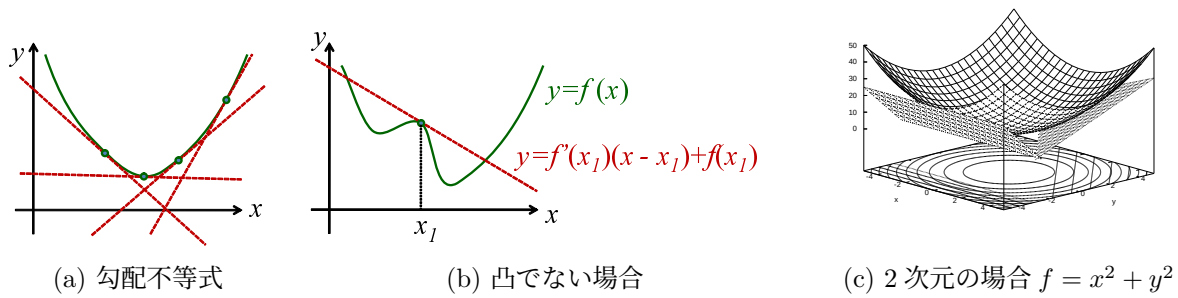


図 5: 勾配不等式

7.5 2階微分可能な凸関数

warm-up 2次関数 $f(x) = ax^2 + bx + c$ が凸関数であるための条件は何か?

$f: \mathbb{R}^n \rightarrow [-\infty, \infty]$ を2回連続微分可能とする. f が凸関数であることは, 任意の $x \in \mathbb{R}$ について, 以下が成り立つことと等価である.

$$f''(x) \geq 0$$

一般に, 任意の $\mathbf{x} \in \mathbb{R}^n$ に対して, ヘッセ行列 $\nabla^2 f(\mathbf{x})$ が, **半正定値**であることが条件である.

定理 7.4 $f: \mathbb{R}^n \rightarrow [-\infty, \infty]$ が連続的 2回微分可能であるとする. このとき, f が凸関数であることの必要十分条件は, 任意の $\mathbf{x} \in \mathbb{R}^n$ に対してヘッセ行列 $\nabla^2 f(\mathbf{x})$ が半正定値行列であることである.

Proof. $(\Rightarrow)$ f が凸関数であるとする. f を2次テイラー近似すると, 任意の $\lambda > 0$, $\mathbf{d} \in \mathbb{R}^n$ に対して,

$$f(\mathbf{x} + \lambda \mathbf{d}) = f(\mathbf{x}) + \lambda \nabla f(\mathbf{x})^\top \mathbf{d} + \frac{\lambda^2}{2} \mathbf{d}^\top \nabla^2 f(\mathbf{x}) \mathbf{d} + o(\lambda^2)$$

が成り立つ. f は凸なので, 先に証明した勾配不等式

$$f(\mathbf{x} + \lambda \mathbf{d}) \geq f(\mathbf{x}) + \lambda \nabla f(\mathbf{x})^\top \mathbf{d}$$

が成り立つ. これら2式より,

$$\frac{1}{2} \lambda^2 \mathbf{d}^\top \nabla^2 f(\mathbf{x}) \mathbf{d} + o(\lambda^2) \geq 0.$$

ここで, 両辺 λ^2 で割って $\lambda \rightarrow +0$ とすると,

$$\mathbf{v}^\top \nabla^2 f(\mathbf{x}) \mathbf{v} \geq 0.$$

$(\Leftarrow)$ ヘッセ行列が半正定値であるとする. テイラーの定理より,

$$\forall \mathbf{x}, \mathbf{y} \in \mathbb{R}^n, f(\mathbf{y}) = f(\mathbf{x}) + \nabla f(\mathbf{x})^\top (\mathbf{y} - \mathbf{x}) + \frac{1}{2} (\mathbf{y} - \mathbf{x})^\top \nabla^2 f(\mathbf{x} + \tau(\mathbf{y} - \mathbf{x})) (\mathbf{y} - \mathbf{x})$$

なる $\tau \in (0, 1)$ が存在する. f の半正定値性より,

$$f(\mathbf{y}) - \{f(\mathbf{x}) + \nabla f(\mathbf{x})^\top (\mathbf{y} - \mathbf{x})\} = \frac{1}{2} (\mathbf{y} - \mathbf{x})^\top \nabla^2 f(\mathbf{x} + \tau(\mathbf{y} - \mathbf{x})) (\mathbf{y} - \mathbf{x}) \geq 0$$

が成り立ち, 勾配不等式を得る. よって前の命題より f は凸. □

Exercise 7.11 任意の $a, b \in \mathbb{R}$ に対して, 関数 $f(x_1, x_2) = \ln(e^{ax_1} + e^{bx_2})$ が凸関数であることを示せ.

Exercise 7.12 関数 f が行列 $\mathbf{V} \in \mathbb{R}^{n \times n}$, ベクトル $\mathbf{c} \in \mathbb{R}^n$, 実数 γ を使って

$$f(\mathbf{x}) = \frac{1}{2} \mathbf{x}^\top \mathbf{V} \mathbf{x} + \mathbf{c}^\top \mathbf{x} + \gamma \tag{3}$$

で定義されているとする. 行列 $\mathbf{V}$ が対称正定値行列のとき, 関数 f は狭義凸2次関数であることを示せ.

8 制約無し最適化問題

【用語集】 (大域的) 最適解, 局所的最適解, 最急降下法, ニュートン法

実数値関数 $f(x)$ に対して, $f(x)$ を最小にする x を求めることを考える. ただし, x には特に制約がなく, $\mathbb{R}^n$ のすべての値を取りうるものとする. このような最適化問題:

$$(P) \begin{cases} \text{最小化} & f(x) \\ \text{条件} & x \in \mathbb{R}^n. \end{cases}$$

を**制約無し最適化問題**とよぶ. この場合, x の動く範囲が自明なので, 単に

$$\text{最小化 } f(x) \quad \text{あるいは} \quad \min f(x)$$

と書いてもよい.

定義 ((大域的) 最適解 (global) optimal solution) x^* が (P) の**大域的最適解 (最小解)**

$$\stackrel{\text{def}}{\iff} \forall x, f(x^*) \leq f(x)$$

定義 (局所的最適解 local optimal solution) x^* が (P) の**局所的最適解 (極小解)**

$$\stackrel{\text{def}}{\iff} \exists \varepsilon > 0 : \forall x \in B(x^*, \varepsilon), f(x^*) \leq f(x). \quad \text{ただし, } B(x^*, \varepsilon) := \{x \in \mathbb{R}^n : \|x - x^*\| < \varepsilon\}$$

局所的最適解 (極小解) の定義において不等式が厳密に成り立つとき, つまり「 x^* の近傍の点 $x \neq x^*$ において $f(x^*) < f(x)$ 」が成り立つとき, x^* は**孤立局所的最適解 (孤立極小解)**である, という.

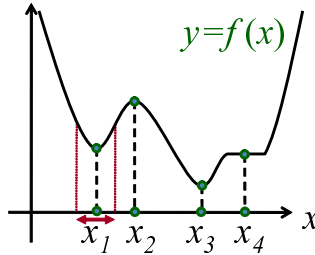


図 6: 局所的最適解と大域的最適解

Exercise 8.1 $\mathbb{R}$ 上で $f(x) = x^3$ の最小化を考えたとき, $x = 0$ は局所的最適解 (極小解) かどうか判定せよ.

8.1 制約無し最小化問題の最適性条件

一般に, ベクトル x が最適化問題の最適解であるための必要条件と十分条件を**最適性条件** (optimality condition) と呼ぶ. ここでは制約無し最適化問題 (P) の最適性条件について考える.

定理 8.1 (1 次の必要条件) x^* が無制約最小化問題 (P) の局所的最適解で, $f: \mathbb{R}^n \rightarrow [-\infty, \infty]$ が x^* の近傍で微分可能であるならば, $\nabla f(x^*) = \mathbf{0}$ が成り立つ.

Proof. 背理法で示す. $\nabla f(x^*) \neq \mathbf{0}$ と仮定する. $d := -\nabla f(x^*)$ とおくと, $d^\top \nabla f(x^*) = -\|\nabla f(x^*)\|^2 < 0$. $\nabla f(x)$ が x^* で連続なので, $d^\top \nabla f(x)$ も連続, i.e.,

$$\exists \bar{t} > 0 : \forall t \in [0, \bar{t}], d^\top \nabla f(x^* + td) < 0.$$

また, 平均値の定理より,

$$\forall t \in (0, \bar{t}), \exists \xi \in (0, 1) : f(x^* + td) = f(x^*) + td^\top \nabla f(x^* + \xi td)$$

であるから, $f(x^* + td) < f(x^*)$. これは x^* の局所的最適性に矛盾.

□

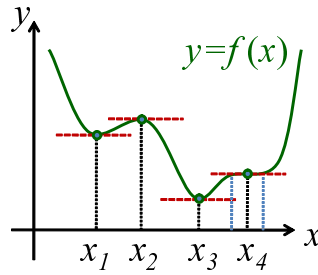


図 7: 停留点

$\nabla f(\mathbf{x}) = \mathbf{0}$ を満たす点のことを _____ (stationary point) と呼ぶ. 停留点は必ずしも局所的最適解ではないが, 制約無し最適化問題の局所的最適解は常に停留点である.

また, 関数のヘッセ行列 (2 次微分) の情報を使うと, 最適性に関してより詳しい情報が得られる. 以下にその 2 つの定理を紹介する. これらの定理は 2 次微分を使うため, 2 次の条件に関する定理と呼ばれる.

定理 8.2 (2 次の必要条件) 連続的 2 回微分可能な関数 f の局所的最小解 $\mathbf{x}^*$ において, $\nabla^2 f(\mathbf{x}^*)$ は半正定値行列である.

Proof. 背理法で示す. $\nabla^2 f(\mathbf{x}^*)$ が半正定値行列でないと仮定する, i.e.,

$$\exists \mathbf{d} \in \mathbb{R}^n : \mathbf{d}^\top \nabla^2 f(\mathbf{x}^*) \mathbf{d} < 0.$$

また, $\nabla^2 f(\mathbf{x})$ が $\mathbf{x}^*$ で連続であるから, $\mathbf{d}^\top \nabla^2 f(\mathbf{x}) \mathbf{d}$ も連続, i.e.,

$$\exists \bar{t} > 0 : \forall t \in [0, \bar{t}], \mathbf{d}^\top \nabla^2 f(\mathbf{x}^* + t\mathbf{d}) \mathbf{d} < 0.$$

また, テイラーの定理より,

$$\forall t \in (0, \bar{t}], \exists \xi \in (0, 1) : f(\mathbf{x}^* + t\mathbf{d}) = f(\mathbf{x}^*) + t\mathbf{d}^\top \nabla f(\mathbf{x}^*) + \frac{1}{2}t^2 \mathbf{d}^\top \nabla^2 f(\mathbf{x}^* + \xi t\mathbf{d}) \mathbf{d}.$$

$t\xi \in (0, \bar{t})$ なので, $\mathbf{d}^\top \nabla^2 f(\mathbf{x}^* + \xi t\mathbf{d}) \mathbf{d} < 0$ であり, 1 次の必要条件から $\nabla f(\mathbf{x}^*) = \mathbf{0}$ であるから, $f(\mathbf{x}^* + t\mathbf{d}) < f(\mathbf{x}^*)$. これは $\mathbf{x}^*$ の局所的最適性に矛盾. $\square$

定理 8.3 (2 次の十分条件) 2 回連続微分可能な関数 f の停留点 $\mathbf{x}^*$ において, $\nabla^2 f(\mathbf{x}^*)$ が正定値行列ならば, $\mathbf{x}^*$ は孤立極小解である.

Proof. $\nabla^2 f(\mathbf{x})$ が $\mathbf{x}^*$ で連続であるから, $\exists \varepsilon > 0 : \forall \mathbf{y} \in B(\mathbf{x}^*, \varepsilon), \forall \mathbf{d} \in \mathbb{R}^n \setminus \{\mathbf{0}\}, \mathbf{d}^\top \nabla^2 f(\mathbf{y}) \mathbf{d} > 0$. 今, $0 < \|\mathbf{d}\| < \varepsilon$ なる $\mathbf{d} \in \mathbb{R}^n$ についてテイラーの定理を適用すると,

$$\exists \xi \in (0, 1) : f(\mathbf{x}^* + \mathbf{d}) = f(\mathbf{x}^*) + \nabla f(\mathbf{x}^*)^\top \mathbf{d} + \frac{1}{2} \mathbf{d}^\top \nabla^2 f(\mathbf{x}^* + \xi \mathbf{d}) \mathbf{d}.$$

$\nabla f(\mathbf{x}^*)^\top \mathbf{d} = 0, \mathbf{d}^\top \nabla^2 f(\mathbf{x}^* + \xi \mathbf{d}) \mathbf{d} > 0$ より, $f(\mathbf{x}^* + \mathbf{d}) > f(\mathbf{x}^*)$ となり, $\mathbf{x}^*$ の孤立局所的最適性が言える. $\square$

Exercise 8.2 $f(x) = ax^2 + bx + c$ として (P) を考えた場合, 上で述べている最適性の条件 (1, 2 次の必要条件, 2 次の十分条件) は何か?

Exercise 8.3 $f(x) = \frac{x^4}{4} - \frac{x^3}{3} - x^2 + 3$ として, この関数の $\mathbb{R}$ 上の最小化を考えると, 停留点 (1 つとは限らない) で最適性の条件 (1, 2 次の必要条件, 2 次の十分条件) をチェックしなさい.

Exercise 8.4 $f(\mathbf{x}) = \frac{1}{2} \mathbf{x}^\top \mathbf{V} \mathbf{x} + \mathbf{c}^\top \mathbf{x} + \gamma$ とする.

(1) $\nabla f(\mathbf{x}), \nabla^2 f(\mathbf{x})$ を求めよ.

(2) (P) を考えた場合, 上で述べている最適性の条件 (1, 2 次の必要条件, 2 次の十分条件) は何か?

定理 8.4 微分可能な関数 $f: \mathbb{R}^n \rightarrow [-\infty, \infty]$ が凸関数であるとき, f の停留点 ($\nabla f(\mathbf{x}^*) = \mathbf{0}$ を満たす点 $\mathbf{x}^*$) は f の大域的最適解である.

Exercise 8.5 定理 8.4 を示せ.

凸関数が微分可能でない場合は, 停留点は存在しないかもしれないが, 局所解は存在するかもしれない. その場合には次の定理が成り立つ.

定理 8.5 関数 $f: \mathbb{R}^n \rightarrow [-\infty, \infty]$ が凸関数であるとき, 極小解は f の大域的最適解である.

Proof. $\mathbf{x}^*$ を極小解とする, i.e.,

$$\exists \varepsilon > 0: \forall \mathbf{x} \in B(\mathbf{x}^*, \varepsilon), f(\mathbf{x}^*) \leq f(\mathbf{x}).$$

同時に $\mathbf{x}^*$ は (P) の大域的最適解でないとする, i.e.,

$$\exists \hat{\mathbf{x}}: f(\hat{\mathbf{x}}) < f(\mathbf{x}^*).$$

今, $\lambda \in (0, 1)$ に対して

$$\tilde{\mathbf{x}}(\lambda) := (1 - \lambda)\mathbf{x}^* + \lambda\hat{\mathbf{x}}$$

とする. f は凸関数であるから,

$$f(\tilde{\mathbf{x}}(\lambda)) \leq (1 - \lambda)f(\mathbf{x}^*) + \lambda f(\hat{\mathbf{x}}) < f(\mathbf{x}^*).$$

λ を十分 0 に近く取る (具体的には $0 < \lambda < \varepsilon / \|\hat{\mathbf{x}} - \mathbf{x}^*\|$ を満たすように取る) と, $\tilde{\mathbf{x}}(\lambda) \in B(\mathbf{x}^*, \varepsilon)$ となり, $\mathbf{x}^*$ が局所的最適解であることに矛盾する. $\square$

よって, 凸関数の最小化問題においては, 極小解を求めれば大域的最適解が得られることが分かる.

例 8.1 Exercise 7.12 に示した, 狭義凸 2 次関数 (3) の最小化問題を考える. 1 次の必要条件 (Exercise 8.4 (2)) より, 停留点は $\mathbf{x}^* = -\mathbf{V}^{-1}\mathbf{c}$ の 1 点のみである. $\mathbf{V}$ が正定値だから正則であり, 逆行列が存在することに注意. 停留点 $\mathbf{x}^*$ におけるヘッセ行列は $\mathbf{V}$ で正定値だから, 2 次の十分条件が満たされ, $\mathbf{x}^*$ は極小解であることがわかる. これが 大域的最適解であることは, 定理 8.4 や 8.5 により保証される. $\square$

8.2 制約無し最適化問題の例題

ここで, 最小 2 乗法の問題 (1) を考えてみる. 問題 (1) をベクトル・行列表現すると次のようになる.

$$\min_{\mathbf{d} \in \mathbb{R}^{n+1}} \|\mathbf{X}\mathbf{d} - \mathbf{y}\|^2 = (\mathbf{X}\mathbf{d} - \mathbf{y})^\top (\mathbf{X}\mathbf{d} - \mathbf{y}).$$

ただし,

$$\mathbf{X} = \begin{bmatrix} \mathbf{x}_1^\top & 1 \\ \vdots & \vdots \\ \mathbf{x}_m^\top & 1 \end{bmatrix} \in \mathbb{R}^{m \times (n+1)}, \quad \mathbf{d} = \begin{bmatrix} a_1 \\ \vdots \\ a_n \\ b \end{bmatrix} \in \mathbb{R}^{n+1}, \quad \mathbf{y} = \begin{bmatrix} y_1 \\ \vdots \\ y_m \end{bmatrix} \in \mathbb{R}^m.$$

関数 $f(\mathbf{d}) = \|\mathbf{X}\mathbf{d} - \mathbf{y}\|^2$ を変数 $\mathbf{d}$ で微分して “= 0” とした式を解けばよい. それは,

$$(\mathbf{X}^\top \mathbf{X})\mathbf{d} = \mathbf{X}^\top \mathbf{y} \tag{4}$$

と表される (正規方程式).

例 8.2 ある地域の 20 世帯の 1 年間の家計について調べたところ, 可処分所得 x [百万円] と消費支出 y [百万円] の間に次の関係があった.

x	5.8	5.7	4.3	3.5	5.6	5.8	6.0	6.2	6.4	3.8	5.3	2.3	6.8	7.8	8.8	6.2	7.4	6.6	10.1	9.3
y	2.5	2.4	2.1	1.7	2.4	2.4	2.4	2.6	2.7	1.8	2.3	1.2	3.0	3.2	3.2	2.5	2.8	2.9	3.8	3.5

これから可処分所得が 350[万円] のとき、消費支出（の期待値）はいくらであると推定されるか。

（解） $\mathbf{x} \in \mathbb{R}^n$ の次元が $n = 1$ の時には、(4) から次の連立一次方程式（正規方程式）が得られる。

$$\begin{pmatrix} \frac{\sum_{i=1}^m x_i}{\sum_{i=1}^m x_i} & \frac{\sum_{i=1}^m x_i^2}{n} \\ \frac{\sum_{i=1}^m x_i^2}{n} & \frac{\sum_{i=1}^m x_i^3}{n} \end{pmatrix} \begin{pmatrix} a \\ b \end{pmatrix} = \begin{pmatrix} \frac{\sum_{i=1}^m y_i}{\sum_{i=1}^m x_i} \\ \frac{\sum_{i=1}^m x_i y_i}{n} \end{pmatrix}$$

値を入れると、

$$\begin{pmatrix} 834.87 & 123.7 \\ 123.7 & 20 \end{pmatrix} \begin{pmatrix} a \\ b \end{pmatrix} = \begin{pmatrix} 185.37 \\ 29.4 \end{pmatrix}$$

となり、これを解くと $b = 0.62, a = 0.32$ が得られる。従って、 x と y の関係は $y = 0.62 + 0.32x$ で近似できる。 $x = 3.5$ を代入すると

$$y = 0.62 + 0.32 \times 3.5 = 1.74$$

であるから 174 万円と推定される。

□

Exercise 8.6 【応用例 3.4】の最尤推定問題を解き、 μ と σ の最尤推定量を求めよ。

8.3 制約無し最適化の解法

例 8.1 を通して、関数が狭義凸 2 次関数ならば、最適解が直接求められることが分かった。しかし、一般の制約無し最適化問題 (P) ではこのようなことは期待できず、最適解を求めるには反復解法を用いるのが一般的である。各反復で“現在の点よりも最適解に近い点を求める”ことにより、徐々に最適解に近付いていくことを考える。

8.3.1 最急降下法

勾配ベクトルの方向に少し進めば目的関数値は上がる。反対に、勾配ベクトルの逆方向に進めば、目的関数値は下がる。勾配ベクトルのこの性質を利用するのが**最急降下法**である。最急降下法の具体的な手順は以下の通りである。

最急降下法 (steepest descent method)

(Step1) 適当に初期解 $\mathbf{x}^{(0)}$ を選ぶ。 $t = 0$ とする。

(Step2) $\nabla f(\mathbf{x}^{(t)}) = \mathbf{0}$ であれば、 $\mathbf{x}^{(t)}$ が停留点であるので終わり。そうでなければ、 $\mathbf{d}^{(t)} = -\nabla f(\mathbf{x}^{(t)})$ とし、 $f(\mathbf{x}^{(t)} + \alpha \mathbf{d}^{(t)})$ が最小となるような正の数 α を選ぶ。

(Step3) $\mathbf{x}^{(t+1)} = \mathbf{x}^{(t)} + \alpha \mathbf{d}^{(t)}$ を次の近似解とし、 $t \rightarrow t + 1$ として step2 に戻る。

ここで $\alpha > 0$ は移動する量を決めるパラメータであり、ステップサイズと呼ばれる。ステップサイズ α は、 $\mathbf{d}^{(t)} = -\nabla f(\mathbf{x}^{(t)})$ として次の問題を構成し、その解として決められる。

$$\begin{array}{ll} \text{最小化} & f(\mathbf{x}^{(t)} + \alpha \mathbf{d}^{(t)}) \\ \text{条件} & \alpha \geq 0. \end{array}$$

このような 1 次元変数の最適化問題を解いてステップサイズを決定することを**直線探索** (line search) と呼ぶ。直線探索問題を厳密に解くことは一般に困難であるが、近似的には効率よく解くことができることが知られている。

単純なアルゴリズムであるため、効率は後述のニュートン法に比べて劣ると言われている。一方、利点としては、1 回の反復にかかる計算量が小さいこと、勾配ベクトルさえあれば計算できることが挙げられる。

8.3.2 ニュートン法

各反復で“現在の点よりも最適解に近い点を求める”ために、目的関数 f を 2 次近似してみる。点 $\mathbf{x}$ のまわりで f をテイラー展開し、3 次以上の項を無視すると、次のようになる。

$$f(\mathbf{x} + \mathbf{d}) \approx f(\mathbf{x}) + \nabla f(\mathbf{x})^\top \mathbf{d} + \frac{1}{2} \mathbf{d}^\top \nabla^2 f(\mathbf{x}) \mathbf{d}. \quad (5)$$

ここで、 $\mathbf{x}$ は現在の点であり、 $\mathbf{d}$ は次の点 $\mathbf{x} + \mathbf{d}$ を決めるための変数ベクトルである。左辺の最小化は一般に困難であるが、右辺はもし $\nabla^2 f(\mathbf{x})$ が正定値ならば $\mathbf{d}$ に関する狭義凸 2 次関数であり、例 8.1 で見たように $\mathbf{d} = -\nabla^2 f(\mathbf{x})^{-1} \nabla f(\mathbf{x})$ で最小化される。もし、(5) の右辺の 2 次近似が左辺を充分よく近似しているのであれば、現在の点 $\mathbf{x}$ よりも $\mathbf{x} + \mathbf{d}$ の方が最適解に近いだろう。ニュートン法では、それを期待して、 $\mathbf{x}$ を $\mathbf{x} - \nabla^2 f(\mathbf{x})^{-1} \nabla f(\mathbf{x})$ で置き換える反復を繰り返す。

ニュートン法 Newton method

(Step1) 適当に初期解 $\mathbf{x}^{(0)}$ を選ぶ。 $t = 0$ とする。

(Step2) $\nabla f(\mathbf{x}^{(t)}) = \mathbf{0}$ であれば、 $\mathbf{x}^{(t)}$ を出力して終わり。

(Step3) $\nabla^2 f(\mathbf{x})$ が正定値であれば

$$\nabla f(\mathbf{x}) + \nabla^2 f(\mathbf{x}) \mathbf{d}^{(t)} = \mathbf{0} \Leftrightarrow \mathbf{d}^{(t)} = -\{\nabla^2 f(\mathbf{x})\}^{-1} \nabla f(\mathbf{x})$$

を満たす $\mathbf{d}^{(t)}$ (ニュートン方向とよばれる) を求める。 $\mathbf{x}^{(t+1)} = \mathbf{x}^{(t)} + \mathbf{d}^{(t)}$ を次の近似解とし、 $t \rightarrow t+1$ として step2 に戻る。

一般に、ニュートン法は最急降下法より効率が良いといわれている。また、基本的には直線探索する必要がないことも特徴である。しかし、ニュートン方向を計算するには正則なヘッセ行列が必要となることが、短所として挙げられる。 $\mathbf{x}^{(t)}$ が充分、局所的最適解に近ければ、 $\nabla^2 f(\mathbf{x})$ は正定値となるが、任意の点について言えるわけではない。その場合は、ヘッセ行列に単位ベクトルを足すなど、工夫する必要がある。

9 制約付き非線形最適化

[用語集] ラグランジュの未定乗数法, 凸計画 (凸最適化), カルーシュ・キューン・タッカー条件 (KKT 条件)

この節では、制約無し最適化問題に対する最適性条件 (1 次の必要条件) を、制約付き最適化の場合に拡張する。定理 8.1 の拡張版である最適性条件 (KKT 条件) を導くのが目的である。等式および不等式を同時に含む一般的な定式化に対する KKT 条件を理解するために、やや冗長であるが、まず等式制約のみの定式化、不等式制約のみの定式化それぞれのケースについて紹介し、最後に一般的な定式化に対する KKT 条件の結果を示す。

制約付き最適化問題とは、決定変数 $\mathbf{x}$ が F (F は $\mathbb{R}^n$ の集合) に限定された最適化問題である:

$$(\text{CP}) \quad \begin{cases} \text{最小化} & f(\mathbf{x}) \\ \text{条件} & \mathbf{x} \in F \subset \mathbb{R}^n. \end{cases} \quad (6)$$

前節では制約無し最適化問題に対して“大域的最適解 (最小解)”や“局所的最適解 (極小解)”を定義したが、制約付き最適化問題 (6) に対しては次のように定義される。

定義 ((大域的) 最適解) $\mathbf{x}^*$ が (6) の大域的最適解 (最小解) $\stackrel{\text{def}}{\iff} \forall \mathbf{x} \in F, f(\mathbf{x}^*) \leq f(\mathbf{x})$

定義 (局所的最適解) $\mathbf{x}^*$ が (6) の局所的最適解 (極小解)

$$\stackrel{\text{def}}{\iff} \exists \varepsilon > 0 : \forall \mathbf{x} \in F \cap B(\mathbf{x}^*, \varepsilon), f(\mathbf{x}^*) \leq f(\mathbf{x})$$

9.1 等式制約のみをもつ非線形計画問題

まず, $F = \{x \in \mathbb{R}^n : g_i(x) = 0, i = 1, 2, \dots, m\}$, すなわち, 制約式が等式のみからなる場合を考える:

$$\begin{cases} \text{最小化} & f(x) \\ \text{条件} & g_i(x) = 0, \quad i = 1, 2, \dots, m. \end{cases} \quad (7)$$

この問題に対して, 以下で定義される関数 $L: \mathbb{R}^n \times \mathbb{R}^m \rightarrow [-\infty, \infty]$ を _____ と呼ぶ.

$$L(x, \lambda) := f(x) + \lambda^\top g(x) = f(x) + \sum_{i=1}^m \lambda_i g_i(x).$$

また, 変数 $\lambda_i, i = 1, \dots, m$, を **ラグランジュ乗数** とよぶ.

このとき, 問題 (7) の最適性の必要条件 (KKT 条件) は以下のように記述される:

定理 9.1 (ラグランジュの未定乗数法 ((7) の必要条件)) $f, g_i (i = 1, \dots, m)$ が微分可能で, かつ, x^* が等式制約付き最適化問題 (7) の局所的最適解とする. このとき, $\lambda^* \in \mathbb{R}^m$ が存在して,

$$\begin{cases} \nabla_x L(x^*, \lambda^*) = 0 \\ \nabla_\lambda L(x^*, \lambda^*) = 0 \end{cases}$$

が成り立つ. ここで, $\nabla_x L(x, \lambda), \nabla_\lambda L(x, \lambda)$ はラグランジュ関数 L を x および λ でそれぞれ微分したもの. すなわち, この方程式系は以下のようにも書ける.

$$\begin{cases} \frac{\partial}{\partial x_j} L(x^*, \lambda^*) = \text{_____} + \sum_{i=1}^m \lambda_i^* \frac{\partial g_i(x^*)}{\partial x_j} = 0, & j = 1, \dots, n \\ \frac{\partial}{\partial \lambda_i} L(x^*, \lambda^*) = \text{_____} = 0, & i = 1, \dots, m. \end{cases}$$

上記のシステムは $n + m$ 個の変数に関する $n + m$ 本の非線形方程式である. 方程式を解くだけで停留点 (に相当するもの) が得られるという意味で, 等式制約のみをもつ最適化問題は制約無し最適化問題に近く, 不等式を含んだ一般の最適化問題よりも扱いやすい. 具体的には, 非線形方程式の解を求めるためのニュートン法を用いて数値的に解くことが考えられる. しかし, 特に規模の小さな問題や構造を持った問題などでは, この方程式を直接解いて解を求めることができる場合がある. そのような方法を _____ の _____ とよぶ. 以下にいくつかその例を見てみよう.

例 9.1 次の非線形計画問題の最適解をラグランジュの未定乗数法を使って求めよう.

$$\begin{cases} \text{最小化} & x^2 + y^2 \\ \text{条件} & x + y = 1. \end{cases}$$

1 本の等式制約に対応して双対変数 $\lambda \in \mathbb{R}$ を導入した上で, _____ 関数 $L(x, y, \lambda)$ を

$$L(x, y, \lambda) = x^2 + y^2 + \lambda(x + y - 1)$$

とすると, (x, y) に関する必要条件は

$$\begin{pmatrix} 2x \\ 2y \end{pmatrix} + \lambda \begin{pmatrix} 1 \\ 1 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$$

となる. したがって, $x = y = -\frac{1}{2}\lambda$. これを制約式 (λ に関する _____ 条件) に代入すれば, $\lambda = \text{_____}$ が得られる. これより, 最適解は (_____), 最適値は _____ となる. $\square$

Exercise 9.1 【応用例 3.7】で 3 つの資産価格 F_1, F_2, F_3 が互いに独立とする. 制約条件を $\sum_{i=1}^3 P_i x_i = M$ に変更し, ポートフォリオ選択問題を解き, 最適ポートフォリオ (x_1, x_2, x_3) を求めよ.

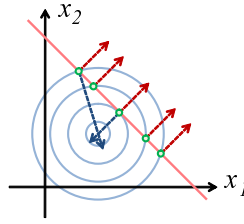


図 8: 等式制約の必要条件の含意

Exercise 9.2 制約式が線形等式制約のみで、目的関数が狭義凸 2 次関数の問題

$$\begin{array}{ll} \text{最小化} & \frac{1}{2} \mathbf{x}^\top \mathbf{V} \mathbf{x} + \mathbf{c}^\top \mathbf{x} + \gamma \\ \text{条件} & \mathbf{A} \mathbf{x} = \mathbf{b} \end{array}$$

を考える。ただし、 $n > m$, $\mathbf{A} \in \mathbb{R}^{m \times n}$, $\mathbf{V} \in \mathbb{R}^{n \times n}$, $\mathbf{c} \in \mathbb{R}^n$, $\mathbf{b} \in \mathbb{R}^m$ であり、 $\mathbf{V}$ は正定値対称行列とする。この問題の最適解を、ラグランジュの未定乗数法を用いて求めよ。

9.2 不等式制約のみをもつ非線形計画問題

次に制約が不等式で与えられている最小化問題について考える。

$$\begin{array}{ll} \text{最小化} & f(\mathbf{x}) \\ \text{条件} & g_i(\mathbf{x}) \leq 0, \quad i = 1, 2, \dots, m. \end{array} \quad (8)$$

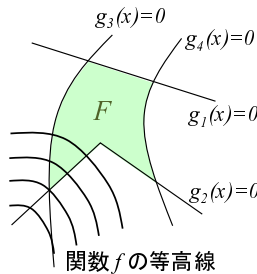


図 9: 不等式制約のみの最適化問題

制約付き最適化問題 (8) の目的関数と全ての制約関数が微分可能であるとする。制約付き問題における最適性は、最適解が制約式の境界上にあるかないかによって状況が変わってくる。最適解 $\mathbf{x}^*$ が境界上にない（内点である）場合、制約無し問題の場合と同様に $\nabla f(\mathbf{x}^*) = \mathbf{0}$ が 1 次の必要条件となる。しかし、最適解 $\mathbf{x}^*$ が境界上にある場合、上式が成り立つとは限らない。問題 (8) のラグランジュ関数を

$$L(\mathbf{x}, \boldsymbol{\lambda}) := f(\mathbf{x}) + \boldsymbol{\lambda}^\top \mathbf{g}(\mathbf{x}) = f(\mathbf{x}) + \sum_{i=1}^m \lambda_i g_i(\mathbf{x}).$$

とすれば、最適性の必要条件は、以下のように表される：

定理 9.2 (1 次の必要条件) 点 $\mathbf{x}^*$ を不等式制約付き問題 (8) の局所最小解とし, f, g_i ($i = 1, \dots, m$) は $\mathbf{x}^*$ において微分可能とする. 適当な _____ が満たされれば, 次の等式・不等式システム (_____ 条件, KKT 条件) を満たす $\lambda^* \in \mathbb{R}^\ell$ が存在する.

$$\begin{cases} \nabla_{\mathbf{x}} L(\mathbf{x}, \lambda) &= \mathbf{0} \\ \nabla_{\lambda} L(\mathbf{x}, \lambda) &\leq \mathbf{0}, \quad \lambda \geq \mathbf{0} \\ \nabla_{\lambda} L(\mathbf{x}, \lambda)^\top \lambda &= 0, \end{cases}$$

あるいは,

$$\frac{\partial f(\mathbf{x}^*)}{\partial x_j} + \sum_{i=1}^m \lambda_i^* \frac{\partial g_i(\mathbf{x}^*)}{\partial x_j} = 0, \quad j = 1, \dots, n. \quad (9)$$

$$\text{_____}, \quad i = 1, \dots, m. \quad (10)$$

$$\lambda_i^* \geq 0, \quad i = 1, \dots, m.$$

$$\text{_____}, \quad i = 1, \dots, m. \quad (11)$$

上の定理は制約がある条件を満たしていれば「KKT 条件が局所最適性の必要条件となる」と言っている. この“制約条件が満たしているべき条件”のことを _____ (constraint qualification) という. 直観的に言えば, 制約想定は制約条件の性質の良さを表す条件と言える. 制約想定についてはいくつか知られており, 以下に 2 つ紹介しておく.

(_____ の制約想定') g_i ($i = 1, \dots, m$) が凸関数で, $g_i(\mathbf{x}^0) < 0$ ($i = 1, \dots, m$) なる $\mathbf{x}^0$ が存在する.

(線形独立制約想定) $i \in I(\mathbf{x}^*)$ なる全ての i に対して, $\nabla g_i(\mathbf{x}^*)$ が線形独立. ここで, $I(\mathbf{x}) := \{i : g_i(\mathbf{x}) = 0, i = 1, \dots, m\}$.

これらが成り立たない問題に対しては, KKT 条件は必要条件にさえならないかもしれない.

Exercise 9.3 どれくらい緩い条件か吟味せよ (具体的な応用問題を定式化し, これが成り立つかどうかどうかが観察せよ). また, これらが成り立たない例を作成せよ.

KKT 条件は局所最適性の必要条件であるから, 満たしているからといって $\mathbf{x}^*$ が局所最適解になっているとは限らない. また, $\mathbf{x}^*$ が局所最適解であっても, 制約想定がみたされていないならば, KKT 条件を満たす $\lambda^* \in \mathbb{R}^\ell$ が存在しない場合もある.

例 9.2 KKT 条件と制約想定の関係を見るために, 次の例を考える.

$$\begin{array}{ll} \text{最小化} & f(\mathbf{x}) = x_1 - x_2 \\ \text{条件} & g_1(\mathbf{x}) = -x_1^2 + x_2 \leq 0 \\ & g_2(\mathbf{x}) = -x_2 \leq 0 \\ & g_3(\mathbf{x}) = x_1 - 1 \leq 0 \\ & g_4(\mathbf{x}) = -x_1^3 \leq 0. \end{array}$$

この問題の最適解は $(1, 1)^\top \in \mathbb{R}^2$ と $(0, 0)^\top \in \mathbb{R}^2$ の 2 つである (図を描いて確認せよ). $\mathbf{x}^* = (1, 1)^\top$ のとき KKT 条件が成立するが, $\mathbf{x}^* = (0, 0)^\top$ のときには, KKT 条件をみたす λ^* は存在しない. $\square$

KKT 条件はどんな条件だろうか. (10) は, $\mathbf{x}^*$ が実行可能解であることを表している. (11) は, $g_i(\mathbf{x}^*) < 0$ ならば $\lambda_i^* = 0$ であることを示している. すると, $g_i(\mathbf{x}^*) = 0$ である制約 ($\mathbf{x}^*$ で有効な制約という) の添字集合 $I(\mathbf{x}^*)$ を用いれば, (9) を

$$\nabla f(\mathbf{x}^*) + \sum_{i \in I(\mathbf{x}^*)} \lambda_i^* \nabla g_i(\mathbf{x}^*) = \mathbf{0}$$

と書くことができる. 有効な制約がない場合や, そもそも無制約な場合には, $\nabla f(\mathbf{x}^*) = 0$ という定理 8.1 の場合に帰着される. また, 有効な制約がない場合, $\boldsymbol{\lambda}^* = \mathbf{0}$ である.

KKT 条件は次のように解釈できる. 解 $\mathbf{x}^*$ が制約不等式 $g_i(\mathbf{x}) \leq 0$ の定義する領域の “内部”, すなわち不等式 $g_i(\mathbf{x}^*) < 0$ の成り立つところであれば, $\lambda_i^* g_i(\mathbf{x}^*) = 0$ より $\lambda_i^* = 0$ でなければならない. このことは, 制約不等式に対するラグランジュ乗数が線形計画における “_____ 変数” の役割を果たしていることを意味する. すなわち, 制約 $g_i(\mathbf{x}^*) \leq 0$ がある資源 i の限度を表しているとするれば, λ_i^* がその資源の価値 (効用) を表している. まだ資源が残っているときは $g_i(\mathbf{x}^*) < 0$ であり, 価値は $\lambda_i^* = 0$ である. KKT 条件から, 解においては

$$-\frac{\partial f(\mathbf{x}^*)}{\partial x_j} = \sum_{i=1}^m \lambda_i^* \frac{\partial g_i(\mathbf{x}^*)}{\partial x_j}$$

が成り立つ. したがって, $\mathbf{x}^*$ を $\mathbf{x}^* + \Delta \mathbf{x}$ だけ変化させると, $f(\mathbf{x})$ の減少量 Δf は 1 階近似として次のように書ける.

$$\Delta f = -\frac{\partial f(\mathbf{x}^*)}{\partial x_j} \Delta x_j = \sum_{i=1}^m \lambda_i^* \frac{\partial g_i(\mathbf{x}^*)}{\partial x_j} \Delta x_j = \sum_{i=1}^m \lambda_i^* \Delta g_i$$

$\lambda_i^* > 0$ なら $g_i(\mathbf{x}^*)$ が限界に達しているから, $g(\mathbf{x})$ の単位の増分に対して $f(\mathbf{x})$ が λ_i^* だけ増加する. しかし, $\lambda_i^* = 0$ なら $g_i(\mathbf{x}^*)$ はまだ限界に達していないから, $g_i(\mathbf{x}^*)$ を増やしても f は変化しないことが確認できる.

Exercise 9.4 $\mathbf{A} \in \mathbb{R}^{m \times n}$, $\mathbf{b} \in \mathbb{R}^m$, $\mathbf{c} \in \mathbb{R}^n$ とする. このとき不等式標準形の線形計画問題:

$$(LP) \begin{cases} \text{最大化} & \mathbf{c}^\top \mathbf{x} \\ \text{条件} & \mathbf{A}\mathbf{x} \leq \mathbf{b}, \mathbf{x} \geq \mathbf{0}. \end{cases}$$

の KKT 条件を求め, 既に学んだ LP の最適性条件と一致していることを確認せよ.

9.3 等式制約と不等式制約の両方をもつ場合

ここでは, 等式制約と不等式制約の両方をもつ非線形計画問題を考えてみよう.

$$\begin{cases} \text{最小化} & f(\mathbf{x}) \\ \text{条件} & g_i(\mathbf{x}) \leq 0, \quad i = 1, \dots, m \\ & h_j(\mathbf{x}) = 0, \quad j = 1, \dots, \ell. \end{cases} \quad (12)$$

この問題のラグランジュ関数を以下のように定義する:

$$L(\mathbf{x}, \boldsymbol{\lambda}, \boldsymbol{\mu}) = f(\mathbf{x}) + \boldsymbol{\lambda}^\top \mathbf{g}(\mathbf{x}) + \boldsymbol{\mu}^\top \mathbf{h}(\mathbf{x}) = f(\mathbf{x}) + \sum_{i=1}^m \lambda_i g_i(\mathbf{x}) + \sum_{j=1}^{\ell} \mu_j h_j(\mathbf{x}).$$

ただし, $\boldsymbol{\lambda} \geq \mathbf{0}$. 等式制約を 2 つの不等式制約に書き換えて, 不等式条件のみをもつ問題で行った議論と同様に考えることにより, 次の KKT 条件が求められる.

$$\begin{cases} \nabla_{\mathbf{x}} L(\mathbf{x}, \boldsymbol{\lambda}, \boldsymbol{\mu}) = \mathbf{0} & \Leftrightarrow \frac{\partial f(\mathbf{x})}{\partial x_k} + \sum_{i=1}^m \lambda_i \frac{\partial g_i(\mathbf{x})}{\partial x_k} + \sum_{j=1}^{\ell} \mu_j \frac{\partial h_j(\mathbf{x})}{\partial x_k} = 0, \quad k = 1, \dots, n \\ \nabla_{\boldsymbol{\lambda}} L(\mathbf{x}, \boldsymbol{\lambda}, \boldsymbol{\mu}) \leq \mathbf{0} & \Leftrightarrow g_i(\mathbf{x}) \leq 0, \quad i = 1, \dots, m \\ \nabla_{\boldsymbol{\mu}} L(\mathbf{x}, \boldsymbol{\lambda}, \boldsymbol{\mu}) = \mathbf{0} & \Leftrightarrow h_j(\mathbf{x}) = 0, \quad j = 1, \dots, \ell \\ \boldsymbol{\lambda} \geq \mathbf{0} & \Leftrightarrow \lambda_i \geq 0, \quad i = 1, \dots, m \\ \boldsymbol{\lambda}^\top \nabla_{\boldsymbol{\lambda}} L(\mathbf{x}, \boldsymbol{\lambda}, \boldsymbol{\mu}) = 0 & \Leftrightarrow \lambda_i g_i(\mathbf{x}) = 0, \quad i = 1, \dots, m. \end{cases} \quad (13)$$

ラグランジュ乗数 λ_i の非負性は不等式制約 $g_i(\mathbf{x}) \leq 0$ に対応していることに注意.

定理 9.3 (_____ (KKT) 条件) 点 $\mathbf{x}^*$ を制約付き問題 (12) の局所的最小解とし, f, g_i ($i = 1, \dots, m$), h_j ($j = 1, \dots, \ell$) は $\mathbf{x}^*$ において微分可能とする. 適当な 制約想定 が満たされれば, KKT 条件 (13) を満たす $\boldsymbol{\lambda}^* \in \mathbb{R}^m$, $\boldsymbol{\mu}^* \in \mathbb{R}^\ell$ が存在する.

ここで, 先に紹介した制約想定は以下のように書き換えておく:

(スレイターの _____) g_i ($i = 1, \dots, m$) が凸関数で, h_j ($j = 1, \dots, \ell$) が線形関数であり, $g_i(\mathbf{x}^0) < 0$ ($i = 1, \dots, m$) かつ $h_j(\mathbf{x}^0) = 0$ ($j = 1, \dots, \ell$) なる $\mathbf{x}^0$ が存在する.

(線形独立制約想定) $i \in I(\mathbf{x}^*)$ なる全ての i に対して, $\nabla g_i(\mathbf{x}^*)$ と全ての $j = 1, \dots, \ell$ に対して $\nabla h_j(\mathbf{x}^*)$ が線形独立. ここで, $I(\mathbf{x}) := \{i : g_i(\mathbf{x}) = 0, i = 1, \dots, m\}$.

Exercise 9.5 $A \in \mathbb{R}^{m \times n}$, $\mathbf{b} \in \mathbb{R}^m$, $\mathbf{c} \in \mathbb{R}^n$ とする. このとき線形計画問題:

$$(\text{LP}) \quad \begin{cases} \text{最小化} & \mathbf{c}^\top \mathbf{x} \\ \text{条件} & A\mathbf{x} = \mathbf{b}, \mathbf{x} \geq \mathbf{0}. \end{cases}$$

の KKT 条件を求め, LP の最適性条件と一致していることを確認せよ.

Exercise 9.6 2 次計画問題:

$$(\text{QP}) \quad \begin{cases} \text{最小化} & \mathbf{c}^\top \mathbf{x} + \frac{1}{2} \mathbf{x}^\top Q \mathbf{x} \\ \text{条件} & A\mathbf{x} \geq \mathbf{b}, \mathbf{x} \geq \mathbf{0}. \end{cases}$$

の KKT 条件を求めよ. ただし, $A \in \mathbb{R}^{m \times n}$, $\mathbf{b} \in \mathbb{R}^m$, $\mathbf{c} \in \mathbb{R}^n$, $Q \in \mathbb{R}^{n \times n}$.

9.4 凸計画問題

定義 (凸計画問題 convex programming) 問題 (12) は, 関数 f , h_i , $i = 1, 2, \dots, m$ が凸関数, g_j , $j = 1, \dots, \ell$ が線形関数であるとき, **凸計画問題**であるという.

凸性は大域的最適性を調べる上で重要な役割を果たしている.

定理 9.4 (凸計画問題の KKT 条件) 問題 (12) は凸計画問題とする. 関数 f , h_i , $i = 1, 2, \dots, m$, g_j , $j = 1, \dots, \ell$ が微分可能であるとき, ある解 $(\mathbf{x}^*, \boldsymbol{\lambda}^*, \boldsymbol{\mu}^*)$ が KKT 条件 (13) を満たせば, $\mathbf{x}^*$ が問題 (12) の最適解となる.

よって, 定理 9.2 と定理 9.4 より, 凸計画問題に対しては制約想定の下で, KKT 条件が大域的最適性の必要十分条件となる.

Exercise 9.7 (1) が凸関数であるとする ($i=1, \dots, m$). が凸関数であるとする ($i = 1, \dots, m$). $g_i : \mathbb{R}^n \rightarrow [-\infty, \infty]$ が凸関数であるとする ($i = 1, \dots, m$). このとき, 任意の $a_i \geq 0$, $i = 1, \dots, m$, に対して, $\eta(\mathbf{x}) := \sum_{i=1}^m a_i h_i(\mathbf{x})$ も凸関数となることを示せ.

(2) (1) の性質および定理 7.3 を利用して, 定理 9.4 を示せ.

参考文献

- [1] 山下・福島 (2008) 『数値計画法』(コロナ社)
- [2] 矢部 (2006) 『最適化とその応用』(数理工学社)