

目次

第 2 章 最適化理論の基礎	3
2.1 最適化問題の定義	3
2.2 最適化問題の分類	6
2.3 最小点の存在	9
2.4 微分法と凸関数	10
2.4.1 Taylor の定理	11
2.4.2 凸関数	12
2.4.3 微分法と凸関数の例題	16
2.5 制約なし最適化問題	17
2.5.1 極小点の必要条件	18
2.5.2 極小点の十分条件	19
2.5.3 最小点の十分条件	20
2.5.4 制約なし最適化問題の例題	20
2.5.5 制約なし最適化問題の解法に関する考察	21
2.6 等式制約つき最適化問題	21
2.6.1 極小点の必要条件	22
2.6.2 Lagrange 乗数法	24
2.6.3 極小点の十分条件	28
2.6.4 等式制約つき最適化問題の例題	29
2.6.5 直接微分法と随伴変数法	30
2.6.6 等式制約つき最適化問題の解法に関する考察	34
2.7 不等式制約つき最適化問題	35
2.7.1 極小点の必要条件	36
2.7.2 最小点の必要十分条件	37
2.7.3 KKT 条件	38
2.7.4 極小点の十分条件	44
2.7.5 KKT 条件を用いた最小点の十分条件	44
2.7.6 不等式制約つき最適化問題の例題	45
2.7.7 不等式制約つき最適化問題の解法に関する考察	46

2.8	等式と不等式制約つき最適化問題	47
2.8.1	等式と不等式制約つき最適化問題に対する Lagrange 乗数法	48
2.8.2	等式と不等式制約つき最適化問題の解法に関する考察	50
2.9	双対定理	50
2.9.1	双対定理の例題	51
2.10	第 2 章のまとめ	54
2.11	第 2 章の演習問題	55

第2章 最適化理論の基礎

第1章では、最適設計問題の具体例を挙げて最適性の条件を求めるまでをみてきた。その過程で最適化理論で使われる用語や結果が使われた。本章では、最適化理論を体系的に解説しておきたい。

本章の舞台は有限次元ベクトル空間である。すなわち、設計変数の線形空間と状態変数の線形空間が有限次元であると仮定する。本書では、第7章以降で、関数空間上で最適設計問題を構成してその解法を考える。しかし、本章で示される概念や結果の多くはそのまま関数空間でも使うことができる。その意味で、本章の内容は本書の基盤をなすといえる。いいかえれば、本書がどれだけ信頼に値するかは本章の内容にかかっているとんでもない。そのために、やや抽象的な内容になるが、定義と定理の形式でまとめておきたい。ここでは、これらの抽象的な内容を具体的なイメージにむすびつけることができるように、読者になじみのある簡単なばね系を用いた例題をはさみながら説明を進めていくことにする。

2.1 最適化問題の定義

最初に、第1章でみてきた問題を念頭において、本章で考察の対象とする最適化問題を定義しよう。問題 1.2.2 では、断面積を $\mathbf{a} \in X = \mathbb{R}^n$ 、変位を $\mathbf{u} \in U = \mathbb{R}^n$ とおき、第1章ではそれらを設計変数および状態変数とよんだ。ここで、 X と U はそれらの線形空間を表していた。まず、この問題が含まれるようないくとおりかの最適化問題を定義しよう。

最初に、線形空間に対する記号法について、他の章とは異なる使い方をすることを断っておく。第1章では、 X と U は設計変数と状態変数が入る線形空間を表していた。第2章では、設計変数と状態変数をあわせて設計変数とよび、その線形空間を X と表すことにする。すなわち、第1章の変数に対して $\mathbf{x} = (\mathbf{a}^\top, \mathbf{u}^\top)^\top \in X$ のように定義しようということである。このような定義に変更しなければならない理由は、標準的な最適化問題では、設計変数と状態変数が区別されずに、すべての変数が設計変数として扱われるためである。

次に、第1章でみてきた問題との関係を考えながら最適化問題を定義しよう。問題 1.2.2 では、 $n \in \mathbb{N}$ に対して $\mathbf{x} = (\mathbf{a}^\top, \mathbf{u}^\top)^\top \in \mathbb{R}^{2n}$ とおいたとき、 $\mathbf{h}(\mathbf{x}) =$

$K(\mathbf{a})\mathbf{u} - \mathbf{p} = \mathbf{0}_{\mathbb{R}^n}$ を等式制約においていた。 $2n$ を $d \in \mathbb{N}$ におきかえれば、次のような最適化問題になる。

問題 2.1.1 (最適化問題) $X = \mathbb{R}^d$ とする。 $f_0, f_1, \dots, f_m : X \rightarrow \mathbb{R}$ は与えられているとする。 また、 $n < d$ に対して $\mathbf{h} = (h_1, \dots, h_n)^\top : X \rightarrow \mathbb{R}^n$ は与えられているとする。 このとき、

$$\min_{\mathbf{x} \in X} \{f_0(\mathbf{x}) \mid f_1(\mathbf{x}) \leq 0, \dots, f_m(\mathbf{x}) \leq 0, \mathbf{h}(\mathbf{x}) = \mathbf{0}_{\mathbb{R}^n}\}$$

を満たす \mathbf{x} を求めよ。 □

また、問題 1.2.2 において $\mathbf{a} \in X = \mathbb{R}^n$ を設計変数とみなし、評価関数を $\tilde{f}_i(\mathbf{a}) = f_i(\mathbf{a}, \mathbf{u}(\mathbf{a}))$ とみなせば、等式制約を除いた形式で $\mathbf{a} \in X$ に対する最適化問題をかくことができる。 ここで、 $\mathbf{a} \in X = \mathbb{R}^n$ を $\mathbf{x} \in X = \mathbb{R}^d$ にかきかえて、 $\tilde{f}_i(\mathbf{a})$ を $f_i(\mathbf{x})$ にかきかえれば、問題 1.2.2 は次のような最適化問題になる。

問題 2.1.2 (最適化問題) $X = \mathbb{R}^d$ とする。 $f_0, f_1, \dots, f_m : X \rightarrow \mathbb{R}$ が与えられたとする。 このとき、

$$\min_{\mathbf{x} \in X} \{f_0(\mathbf{x}) \mid f_1(\mathbf{x}) \leq 0, \dots, f_m(\mathbf{x}) \leq 0\}$$

を満たす \mathbf{x} を求めよ。 □

さらに、**設計変数の許容集合**あるいは**実行可能集合**を

$$S = \{\mathbf{x} \in X \mid f_1(\mathbf{x}) \leq 0, \dots, f_m(\mathbf{x}) \leq 0\} \tag{2.1.1}$$

とおくことにする。 このとき、問題 2.1.2 は次の問題と同値である。

問題 2.1.3 (最適化問題) $X = \mathbb{R}^d$, $f_0, f_1, \dots, f_m : X \rightarrow \mathbb{R}$ は与えられたとする。 S を式 (2.1.1) とおく。 このとき、

$$\min_{\mathbf{x} \in S} f_0(\mathbf{x})$$

を満たす \mathbf{x} を求めよ。 □

本章では、問題 2.1.1 から問題 2.1.3 の最小点について成り立つ関係をみていきたい。 本題に入る前に、これから使われる用語の定義といくつかの注意を示しておくことにする。 本章では f_0 を**目的関数**、 \mathbf{h} を**等式制約関数**、 f_1, \dots, f_m を**不等式制約関数**とよぶことにする。 ここで、二つのことを注意しておきたい。

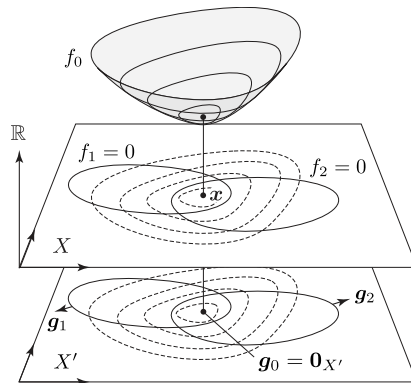


図 2.1: すべての制約が無効なときの最小点

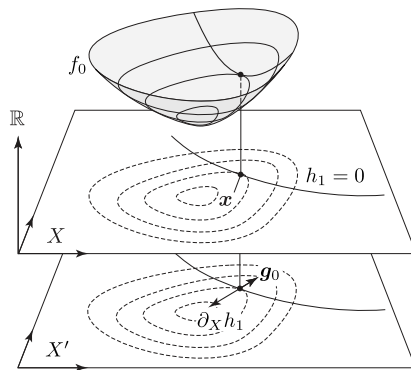


図 2.2: 等式制約が有効なときの最小点

一つは、目的関数の最大化と最小化の関係と不等式制約における不等号の向きに関する注意である。 f_0 の最大化は $-f_0$ の最小化と同値である。そこで、 f_0 の最小化に限定することにする。さらに f_1, \dots, f_m を非正に限定しても一般性を失わないことになる。

もう一つは、 $i \in \{1, \dots, n\}$ に対して、等式制約 $h_i = 0$ は二つの不等式制約 $h_i \leq 0$ と $-h_i \leq 0$ を同時に課すことと同値であることに関する注意である。そこで、問題 2.1.1 における $h_i = 0$ を $h_i \leq 0$ と $-h_i \leq 0$ におきかえることによって、問題 2.1.1 を問題 2.1.2 にかきかえることができる。しかし、本章では、あえて等式制約が区別してかかれた問題 2.1.1 の最小点において成り立つ関係を詳しくみていくことにする。その理由は次のようである。最適設計問題では必ず状態方程式が等式制約として現れる。将来、それが偏微分方程式の境界値問題になる。そのときに、本章で示される等式制約の扱い (Lagrange 乗数法あるいは随伴変数法) は、関数空間

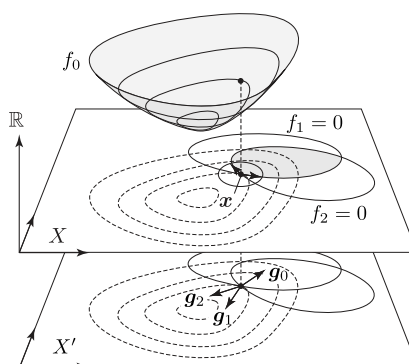


図 2.3: 不等式制約が有効なときの最小点

上での等式制約 (偏微分方程式の境界値問題) の扱いを考える際の指導原理となる。その詳細は第 7 章で示される。

本章で考える最適化問題を図に示しておこう。図 2.1 から図 2.3 に $X = \mathbb{R}^2$ の場合の最適化問題における最小点の例を示す。 g_0, g_1, g_2 および $\partial_X h_1$ は $x \in X$ における f_0, f_1, f_2 および h_1 の x の任意変動に対する勾配 (定義 2.4.1) を表す。 X' はそれらの勾配が属する空間で X の双対空間とよばれる (定義 4.4.5)。しかし、有限次元ベクトル空間では $X' = X$ となることから、ここでは $X' = \mathbb{R}^2$ であるとみてよい。第 7 章でこの図を使うときは、 X' は X とは別のベクトル空間として扱われる。今後、最適化問題が定義されたときは、そこに登場する変数や関数をこれらの図に当てはめて状況を把握してほしい。

2.2 最適化問題の分類

次に、問題 2.1.1 から問題 2.1.3 で使われている関数の性質に注目して、最適化問題の分類方法について示しておこう。

問題 2.1.1 において f_0, \dots, f_m および h_1, \dots, h_n がすべて 1 次関数、あるいは問題 2.1.2 と問題 2.1.3 において f_0, \dots, f_m がすべて 1 次関数のとき、これらの問題を線形最適化問題あるいは線形計画問題という。図 2.4 に $X = \mathbb{R}^2$ のときの線形最適化問題のようすを示す。線形最適化問題に対しては、1 次関数の性質を利用した単体法や主双対内点法などの解法が知られている。非線形最適化問題も逐次線形近似により線形最適化問題に変換して解かれることもある。しかし、本書では直接利用することがないことから詳細を省くことにする。

一方、問題 2.1.1 において f_0, \dots, f_m および h_1, \dots, h_n のいずれかが 1 次関数ではない、あるいは問題 2.1.2 と問題 2.1.3 において f_0, \dots, f_m のいずれかが 1 次関

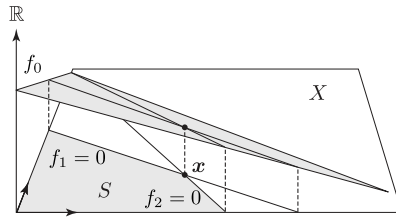


図 2.4: 線形最適化問題における最小点

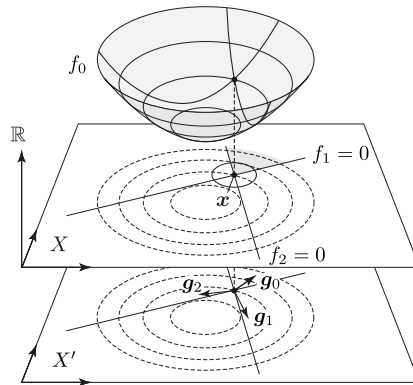


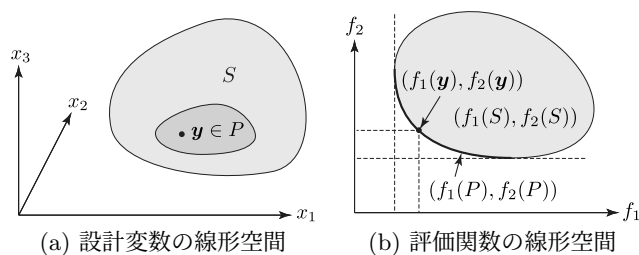
図 2.5: 2次最適化問題における最小点

数ではないとき、これらの問題を**非線形最適化問題**あるいは**非線形計画問題**という。

また、 f_0 が2次関数で、 h_1, \dots, h_n および f_1, \dots, f_m のすべてが1次関数のとき、あるいは問題 2.1.2 と問題 2.1.3 において f_0 が2次関数、 f_1, \dots, f_m のすべてが1次関数のとき、これらの問題を**2次最適化問題**あるいは**2次計画問題**という。図 2.5 に $X = X' = \mathbb{R}^2$ のときの2次最適化問題のイメージを示す。

さらに、問題 2.1.1 において f_0, \dots, f_m が凸関数 (定義 2.4.3) で、 h_1, \dots, h_n が1次関数のとき、あるいは問題 2.1.2 と問題 2.1.3 において f_0, \dots, f_m が凸関数のとき、これらの問題を**凸最適化問題**あるいは**凸計画問題**という。

ここで、混乱する例をあげておきたい。問題 1.1.4 は、1.1.7 項でみたように、凸最適化問題であった。設計変数に対する \tilde{f}_0 の Hesse 行列が正定値 (定理 2.4.6) で、かつ f_1 が1次関数 (定理 2.4.4) になっていたことから、問題 1.1.4 を問題 2.1.2 とみなしたときに、凸最適化問題と判定された。しかし、問題 1.1.4 を問題 2.1.1 にかきかえたときに、凸最適化問題ではないようにみえる。実際、問題 2.1.1 において $\mathbf{x} = (\mathbf{a}^\top, \mathbf{u}^\top)^\top$ とおき、 $\mathbf{h}(\mathbf{x}) = \mathbf{K}(\mathbf{a})\mathbf{u} - \mathbf{p} = \mathbf{0}_{\mathbb{R}^n}$ とおいたとき、 $\mathbf{h}(\mathbf{x})$ は \mathbf{x} の1次関数になっていないからである。このように、等式制約を含む問題では、等式制約を含まない問題にかきかえたときに凸最適化問題になっていることがあること

図 2.6: Pareto 解の集合 P

に注意されたい.

なお、問題 2.1.1 から問題 2.1.3 の形式とは異なるが、目的関数を複数設けた最適化問題は**多目的最適化問題**とよばれる. たとえば、 S を $X = \mathbb{R}^d$ の部分集合として、 $f_1, \dots, f_m : X \rightarrow \mathbb{R}$ を評価関数とすると、

$$\min_{\mathbf{x} \in S} f_1(\mathbf{x}), \dots, \min_{\mathbf{x} \in S} f_m(\mathbf{x})$$

を同時に満たす \mathbf{x} を求めるような問題である. このような問題の最小点は存在するとは限らない. 最小点が存在しないとき、次善の策として **Pareto 解**の集合が使われる. Pareto 解の集合とは、次の条件を満たす集合 $P \subset S$ として定義される (図 2.6). P の要素 \mathbf{y} を決めたと、すべての $i \in \{1, \dots, m\}$ に対して

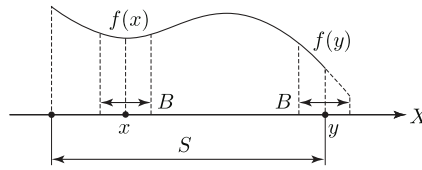
$$f_i(\mathbf{x}) \leq f_i(\mathbf{y})$$

が成り立つような $\mathbf{x} \in S$ はない. また、 $i \in \{1, \dots, m\}$ を決めたと、すべての $\mathbf{y} \in P$ に対して

$$f_i(\mathbf{x}) < f_i(\mathbf{y})$$

が成り立つような $\mathbf{x} \in S$ はない. このような Pareto 解の集合から 1 つの点を選ぶためには、何らかの価値観に基づく選考規準が必要となる. 本書では、選考規準に関する議論を避けて、多目的最適化問題は扱わないことにする.

さらに、問題 2.1.3 において、設計変数の許容集合 S が整数のように離散的な点の集合となるとき、そのような問題を**整数最適化問題**あるいは**整数計画問題**という. この問題は NP 困難とよばれる性質をもち、厳密解は容易にはみつけれられない. また、近似解を求めるために特別の工夫が必要となる. このような問題も本書では扱わないことにする.

図 2.7: 関数 f の極小点 x と最小点 y

2.3 最小点の存在

最適化問題が定義されたので、次に、それらの問題において最小点が存在するための条件について考えてみることにしよう。以下は当然のように思われるが、関数最適化問題ではそれを意識しないと最小点のない問題を定義することになる。

まず、極小点と最小点の違いに注意しよう。

定義 2.3.1 (極小点と最小点) $X = \mathbb{R}^d$ として、 S を X の空でない部分集合とする。 $f: S \rightarrow \mathbb{R}$ とする。 $x \in S$ のある近傍 (凸開集合) B が存在して、任意の $y \in B$ に対して

$$f(x) \leq f(y)$$

が成り立つとき、 $f(x)$ を**極小値**、 x を**極小点**という。さらに、任意の $y \in S$ に対して、上の不等式が成り立つとき、 $f(x)$ を**最小値**、 x を**最小点**という。 \square

図 2.7 は $X = \mathbb{R}$ で $S \subset X$ が有界閉集合のときの関数 f の極小点 x と最小点 y を示す。近傍や開集合などの用語の定義は A.1.1 項に示されている。最小点を定義 2.3.1 のように定義するとき、最小点が存在するための十分条件について、次のような **Weierstrass の定理**が知られている (たとえば, [10, 定理 13, p. 27], [1, 22.6 Maximum and Minimum Value Theorem, p. 154], [9, 4.16 Theorem, p. 89])。

定理 2.3.2 (Weierstrass の定理) 問題 2.1.3 において、 S は $X = \mathbb{R}^d$ の有界閉部分集合、かつ $f_0: S \rightarrow \mathbb{R}$ は連続とする。このとき、最小点 $x \in S$ が存在する。 \square

なお、関数の連続性に関する定義は A.1.2 項に示されている。定理 2.3.2 は、連続を**下半連続** (A.1.2 項) に拡張しても成り立つ。しかし、今後の展開で下半連続を必要としないことから省略することにする。

問題 2.1.3 において、 S が有界閉集合でない場合に最小点が存在しない場合があることをみてみよう。たとえば、問題 1.1.4 では、

$$S = \{ \mathbf{a} \in X = \mathbb{R}^2 \mid \mathbf{a} \geq \mathbf{a}_0, f_1(\mathbf{a}) = l(a_1 + a_2) - c_1 \leq 0 \}$$

が仮定された. この S は有界閉集合である. しかし, これを

$$S = \{ \mathbf{a} \in X = \mathbb{R}^2 \mid \mathbf{a} > \mathbf{0}_{\mathbb{R}^2}, f_1(\mathbf{a}) = l(a_1 + a_2) - c_1 \leq 0 \}$$

におきかえれば, この S は有界閉集合ではなくなる. たとえば $p_1 \neq 0$ および $p_2 = 0$ とおいたとき, $a_1 = c_1/2l$ および $a_2 = 0$ が f_0 の極小点になるが, この \mathbf{a} は S に含まれない. そこで, S の中に最小点は存在しないことになる.

また, 関数が連続でない場合, あるいは値が一意に定まらない関数になっている場合も, 最小値が存在する保証はないことになる. これまでみてきたような関数ではそのようなものはなかった. しかし, 関数最適化問題では注意していないと値が一意に定まらない関数 (汎関数) を扱っていることになりかねない. 第4章では, 設計変数が定義された線形空間は連続 (完備) で, その上で定義された評価関数も連続であるような条件を考えていくことになる.

2.4 微分法と凸関数

最適化問題に取り組む前に, 最適化理論で使われる微分法の基礎と凸関数の定義についてみておこう. 関数 f の微分について, 今後, 次の定義を用いることにする.

定義 2.4.1 (勾配と Hesse 行列) $X = \mathbb{R}^d$ とする. $\mathbf{x} \in X$ の近傍 $B \subset X$ 上で関数 $f: B \rightarrow \mathbb{R}$ が定義されているとする. 任意の $\mathbf{y} = (y_1, \dots, y_d)^\top \in X$ に対して,

$$\begin{aligned} \partial_X f(\mathbf{x}) &= \begin{pmatrix} \lim_{y_1 \rightarrow 0} (f(\mathbf{x} + (y_1, 0, \dots, 0)^\top) - f(\mathbf{x})) / y_1 \\ \vdots \\ \lim_{y_d \rightarrow 0} (f(\mathbf{x} + (0, \dots, y_d)^\top) - f(\mathbf{x})) / y_d \end{pmatrix} \\ &= \begin{pmatrix} \partial f / \partial x_1 \\ \vdots \\ \partial f / \partial x_d \end{pmatrix}(\mathbf{x}) = \mathbf{g}(\mathbf{x}) \end{aligned}$$

が $X' = \mathbb{R}^d$ の要素として定まるとき, f は \mathbf{x} において微分可能といい,

$$f'(\mathbf{x})[\mathbf{y}] = \mathbf{g}(\mathbf{x}) \cdot \mathbf{y}$$

を f の \mathbf{x} における微分あるいは全微分, $\mathbf{g}(\mathbf{x})$ を f の \mathbf{x} における勾配という. さらに,

$$\partial_X \partial_X^\top f(\mathbf{x}) = \begin{pmatrix} \partial^2 f / (\partial x_1 \partial x_1) & \dots & \partial^2 f / (\partial x_1 \partial x_d) \\ \vdots & \ddots & \vdots \\ \partial^2 f / (\partial x_d \partial x_1) & \dots & \partial^2 f / (\partial x_d \partial x_d) \end{pmatrix}(\mathbf{x}) = \mathbf{H}(\mathbf{x})$$

が $\mathbb{R}^{d \times d}$ の要素として定まるとき、 f は \mathbf{x} において 2 階微分可能といい、 \mathbf{x} からの任意の変動 $\mathbf{y}_1, \mathbf{y}_2 \in X$ に対して、

$$f''(\mathbf{x})[\mathbf{y}_1, \mathbf{y}_2] = \mathbf{y}_2 \cdot (\mathbf{H}(\mathbf{x}) \mathbf{y}_1)$$

を f の \mathbf{x} における 2 階微分、 $\mathbf{H}(\mathbf{x})$ を Hesse 行列という。□

本書では、すべての $\mathbf{x} \in X$ において $k \in \{0, 1, 2, \dots\}$ 階微分まで連続な関数 $f: X \rightarrow \mathbb{R}$ の全体からなる集合を $C^k(X; \mathbb{R})$ とかくことにする (定義 4.2.2)。また、表記の簡素化のために、特にことわらずに $\partial_X f, \partial_X f_0, \dots, \partial_X f_m$ をそれぞれ $\mathbf{g}, \mathbf{g}_0, \dots, \mathbf{g}_m$ 、および $\partial_X \partial_X^\top f, \partial_X \partial_X^\top f_0, \dots, \partial_X \partial_X^\top f_m$ をそれぞれ $\mathbf{H}, \mathbf{H}_0, \dots, \mathbf{H}_m$ のようにかくことにする。また、 X の要素を \mathbf{x} とおいているときには $\partial_X f$ を $f_{\mathbf{x}}$ とかくことにする。さらに、偏微分方程式などでは $\partial_X f$ は ∇f のように表されることになる。今後、いろいろな微分が定義されるため、多様な表現方法が必要となる。それぞれの場面で定義が示されるので混乱のないようにしてほしい。

2.4.1 Taylor の定理

次に示される Taylor の定理は、今後、いろいろな場面で利用される。

定理 2.4.2 (Taylor の定理) $X = \mathbb{R}^d$ とする。 $\mathbf{a} \in X$ の近傍 B 上で関数 $f \in C^2(B; \mathbb{R})$ が定義されているとする。任意の $\mathbf{b} \in B$ に対して、 $\mathbf{y} = \mathbf{b} - \mathbf{a}$ とおけば、

$$f(\mathbf{b}) = f(\mathbf{a}) + \mathbf{g}(\mathbf{a}) \cdot \mathbf{y} + \frac{1}{2!} \mathbf{y} \cdot (\mathbf{H}(\mathbf{a} + \theta \mathbf{y}) \mathbf{y}) \quad (2.4.1)$$

を満たす $\theta \in (0, 1)$ が存在する。□

証明 まず、 $X = \mathbb{R}$ のときを考える。このときの \mathbf{a}, \mathbf{b} および \mathbf{y} をそれぞれ a, b および y とかくことにする。任意の $x \in B$ に対して

$$h(x) = f(b) - \{f(x) + f'(x)(b-x) + k(b-x)^2\}$$

とおく。ただし、 df/dx を f' とかいた。また、定数 k は

$$h(a) = h(b) = 0$$

を満たすように決定されるものとする。このとき、

$$\begin{aligned} h'(x) &= -f'(x) - f''(x)(b-x) + f'(x) + 2k(b-x) \\ &= -f''(x)(b-x) + 2k(b-x) \end{aligned}$$

となる。Rolle の定理 ($h(a) = h(b)$ のときの平均値の定理) より、

$$h'(c) = 0$$

を満たす c が (a, b) 上に存在する. そこで, $c = a + \theta y$ とかけて,

$$k = \frac{1}{2} f''(a + \theta y)$$

となる. $h(a) = 0$ にこの結果を代入すれば, 定理の結果が得られる.

次に, $X = \mathbb{R}^2$ とする. 任意の $\mathbf{a} = (a_1, a_2)^\top$ と $\mathbf{y} = (y_1, y_2)^\top$ に対して, $t \in \mathbb{R}$ の関数を

$$\begin{aligned} \phi(t) &= f(\mathbf{a} + t\mathbf{y}) \\ &= f(\mathbf{a}) + t \left(y_1 \frac{\partial}{\partial x_1} + y_2 \frac{\partial}{\partial x_2} \right) f(\mathbf{a}) + \frac{t^2}{2} \left(y_1 \frac{\partial}{\partial x_1} + y_2 \frac{\partial}{\partial x_2} \right)^2 f(\mathbf{a} + \theta\mathbf{y}) \end{aligned}$$

とおく. $\phi(1)$ の $t=0$ 周りの Taylor 展開 ($X = \mathbb{R}$ のときの Taylor 展開) を求めると

$$\begin{aligned} \phi(1) &= f(\mathbf{b}) \\ &= f(\mathbf{a}) + \left(y_1 \frac{\partial}{\partial x_1} + y_2 \frac{\partial}{\partial x_2} \right) f(\mathbf{a}) + \frac{1}{2} \left(y_1 \frac{\partial}{\partial x_1} + y_2 \frac{\partial}{\partial x_2} \right)^2 f(\mathbf{a} + \theta\mathbf{y}) \\ &= f(\mathbf{a}) + \left\{ \left(\frac{\partial}{\partial x_1} \quad \frac{\partial}{\partial x_2} \right) f(\mathbf{a}) \right\} \begin{pmatrix} y_1 \\ y_2 \end{pmatrix} \\ &\quad + \frac{1}{2} (y_1 \quad y_2) \left\{ \begin{pmatrix} \frac{\partial}{\partial x_1} \\ \frac{\partial}{\partial x_2} \end{pmatrix} \left(\frac{\partial}{\partial x_1} \quad \frac{\partial}{\partial x_2} \right) f(\mathbf{a} + \theta\mathbf{y}) \right\} \begin{pmatrix} y_1 \\ y_2 \end{pmatrix} \end{aligned}$$

となる. $X = \mathbb{R}^d$ のときは式 (2.4.1) となる. □

Taylor の定理は任意の微分階数に対して示される. しかし, ここでは2階導関数までを用いた表記にとどめた. その理由は, \mathbb{R}^d 上で定義された関数の高階微分に対する表記法が定義されていないためである. この表記法は定義 4.2.2 で示される. また, $\mathbf{y} \cdot (\mathbf{H}(\mathbf{a} + \theta\mathbf{y}) \mathbf{y}) / 2!$ のような θ を含む項は剰余項とよばれる. 本書では, Bachmann-Landau の small- o 記号を用いて, 式 (2.4.1) を

$$f(\mathbf{a} + \mathbf{y}) = f(\mathbf{a}) + \mathbf{g}(\mathbf{a}) \cdot \mathbf{y} + \frac{1}{2!} \mathbf{y} \cdot (\mathbf{H}(\mathbf{a}) \mathbf{y}) + o(\|\mathbf{y}\|_{\mathbb{R}^d}^2) \quad (2.4.2)$$

ともかくことにする. ここで, $o(\epsilon)$ は $\lim_{\epsilon \rightarrow 0} o(\epsilon) / \epsilon = 0$ が成り立つ関数として定義されている. この式を \mathbf{a} まわりの f の Taylor 展開という.

2.4.2 凸関数

次に凸関数の定義と基本的な結果についてみておこう. あとで示されるように, 凸最適化問題では極小点は同時に最小点になる (定理 2.5.6). そのために, 関数の凸性は最適化理論では役に立つ重要な性質の一つである. まず, 凸関数の定義を示そう.

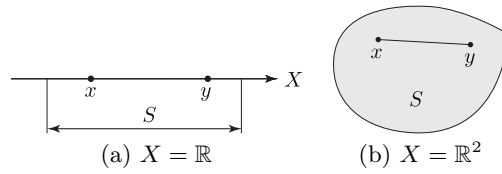


図 2.8: 線形空間 X の部分集合 S が凸集合の例

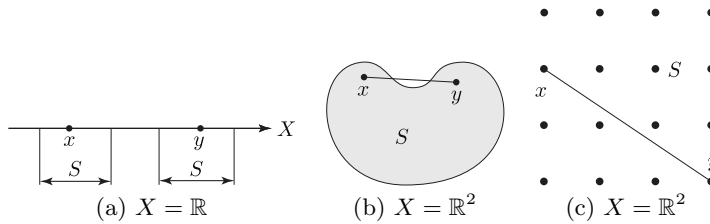


図 2.9: 線形空間 X の部分集合 S が凸集合ではない例

定義 2.4.3 (凸関数) $X = \mathbb{R}^d$ とする. S を X の空ではない部分集合とする. 関数 $f : S \rightarrow \mathbb{R}$ が任意の $x \in S, y \in S$ および $\theta \in (0, 1)$ に対して

$$(1 - \theta)x + \theta y \in S \tag{2.4.3}$$

$$f((1 - \theta)x + \theta y) \leq (1 - \theta)f(x) + \theta f(y) \tag{2.4.4}$$

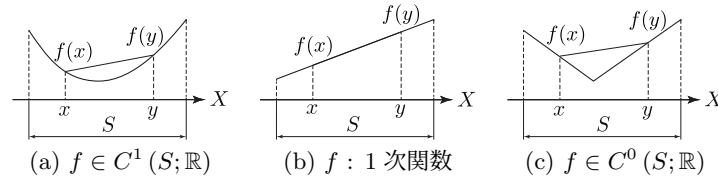
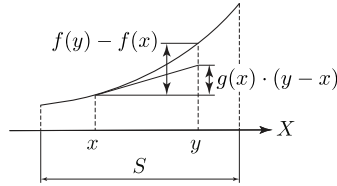
を満たすとき, f を**凸関数**という. また, 式 (2.4.4) の不等号が \geq のとき, **凹関数**という. □

式 (2.4.3) と式 (2.4.4) は, それぞれ集合の凸性と関数の凸性を定義する条件になっている. 図 2.8 と図 2.9 にはそれぞれ S が凸集合であるときと凸集合ではないときの例が示されている. 図 2.9 (c) のような S が整数値ベクトルの集合の場合には**整数計画問題**となる. 許容集合がこのような集合の場合には, 評価関数の勾配が $\mathbf{0}_X$ となるような点が見つかったとしても, その点はその許容集合に含まれるとは限らない. 整数計画問題の難しさはこのような集合の性質にあると考えられる. 図 2.10 には f が凸関数であるときの例が示されている. 導関数が不連続でも凸性は成り立つことに注意されたい.

凸関数が 1 階微分可能ならば次の結果が得られる.

定理 2.4.4 (凸関数と 1 階微分) $X = \mathbb{R}^d, S \subseteq X$ を開凸集合, および $f \in C^1(S; \mathbb{R})$ とする. f が凸関数であるための必要十分条件は, 任意の $x, y \in S$ および $y \in S$ に対して

$$g(x) \cdot (y - x) \leq f(y) - f(x) \tag{2.4.5}$$

図 2.10: f が凸関数の例 ($X = \mathbb{R}$)図 2.11: 凸関数と1階微分 ($X = \mathbb{R}$)

が成り立つことである。 □

証明 まず、必要性 (f が凸関数ならば式 (2.4.5) が成り立つこと) を示す。 f は凸関数なので、任意の $x \in S$ および任意の $\theta \in (0, 1)$ に対して

$$f((1-\theta)x + \theta y) = f(x + \theta(y-x)) \leq f(x) + \theta(f(y) - f(x))$$

が成り立つ。そこで、

$$\frac{f(x + \theta(y-x)) - f(x)}{\theta} \leq f(y) - f(x) \quad (2.4.6)$$

が得られる。 $\theta \rightarrow 0$ のとき、式 (2.4.5) が得られる。

次に、十分性 (式 (2.4.5) が成り立つならば f が凸関数であることを) を示す。 $x = (1-\theta)z + \theta w$ および $y = z$ とおけば、式 (2.4.5) は

$$f(z) - f(x) \geq g(x) \cdot (z - x) \quad (2.4.7)$$

となる。一方、 $x = (1-\theta)z + \theta w$ および $y = w$ とおけば、式 (2.4.5) は

$$f(w) - f(x) \geq g(x) \cdot (w - x) \quad (2.4.8)$$

となる。式 (2.4.7) を $(1-\theta)$ 倍、式 (2.4.8) を θ 倍して両辺の和をとると、

$$(1-\theta)f(z) + \theta f(w) - f(x) \geq g(x) \cdot \{(1-\theta)z + \theta w - x\} = 0$$

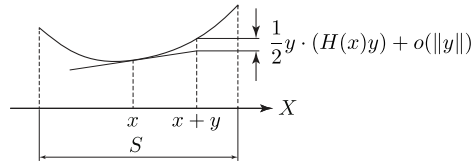
が得られる。すなわち、 S は凸集合で、 f の凸性

$$(1-\theta)f(z) + \theta f(w) \geq f((1-\theta)z + \theta w)$$

が成り立つことから f は凸関数である。 □

図 2.11 は定理 2.4.4 が成り立つようすを示している。

さらに、凸関数が2階微分可能ならば Hesse 行列を使った条件が得られる。それを示すために、**正定値実対称行列**の定義を示しておこう。

図 2.12: 凸関数と 2 階微分 ($X = \mathbb{R}$)

定義 2.4.5 (正定値実対称行列) $A = A^\top \in \mathbb{R}^{d \times d}$ とする. 任意の $\mathbf{x} \in \mathbb{R}^d$ に対して

$$\mathbf{x} \cdot (A\mathbf{x}) \geq \alpha \|\mathbf{x}\|_{\mathbb{R}^d}^2$$

を満たす $\alpha > 0$ が存在するとき, A を**正定値**であるという. $\alpha \geq 0$ が存在するとき, A を**半正定値**であるという. また,

$$\mathbf{x} \cdot (A\mathbf{x}) \leq -\alpha \|\mathbf{x}\|_{\mathbb{R}^d}^2$$

を満たす $\alpha > 0$ が存在するとき, A を**負定値**であるという. $\alpha \geq 0$ が存在するとき, A を**半負定値**であるという. \square

定義 2.4.5 において, A が正定値ならばすべての固有値は正となり, α はそれらの最小値と一致する. また, A が負定値ならばすべての固有値は負となり, $-\alpha$ はそれらの最大値と一致する. これらのことを演習問題 2.1 で確認してほしい.

これらの定義を用いて, 凸関数が 2 階微分可能ならば次の結果が得られる. 図 2.12 はそのようすを示している.

定理 2.4.6 (凸関数と 2 階微分) $X = \mathbb{R}^d$ とおき, $S \subseteq X$ を開凸集合, および $f \in C^2(S; \mathbb{R})$ とする. f が凸関数であるための必要十分条件は, 任意の $\mathbf{x} \in S$ に対して Hesse 行列 $H(\mathbf{x})$ が半正定値であることである. \square

証明 まず, 必要性を示そう. f は凸関数なので, 任意の $\mathbf{x} \in S$, $\mathbf{y} \in S$ および任意の $\theta \in (0, 1)$ に対して, 式 (2.4.6) が成り立つ. 式 (2.4.6) の右辺は, $f \in C^2(S; \mathbb{R})$ なので, 任意の $\mathbf{x} \in S$ および $\mathbf{y} \in S$ に対して, ある $\vartheta \in (0, 1)$ が存在して

$$f(\mathbf{y}) - f(\mathbf{x}) = \mathbf{g}(\mathbf{x}) \cdot (\mathbf{y} - \mathbf{x}) + \frac{1}{2} (\mathbf{y} - \mathbf{x}) \cdot \{H((1 - \vartheta)\mathbf{x} + \vartheta\mathbf{y})(\mathbf{y} - \mathbf{x})\} \quad (2.4.9)$$

とかける. 式 (2.4.6) の左辺は $\theta = \vartheta$ のとき,

$$\begin{aligned} & \frac{f(\mathbf{x} + \vartheta(\mathbf{y} - \mathbf{x})) - f(\mathbf{x})}{\vartheta} \\ &= \mathbf{g}(\mathbf{x}) \cdot (\mathbf{y} - \mathbf{x}) + \frac{1}{2} \vartheta (\mathbf{y} - \mathbf{x}) \cdot \{H((1 - \vartheta)\mathbf{x} + \vartheta\mathbf{y})(\mathbf{y} - \mathbf{x})\} \end{aligned} \quad (2.4.10)$$

となる。したがって、 $\theta = \vartheta$ のときの式 (2.4.6) に式 (2.4.9) と式 (2.4.10) を代入すれば、

$$(1 - \vartheta)(\mathbf{y} - \mathbf{x}) \cdot \{\mathbf{H}((1 - \vartheta)\mathbf{x} + \vartheta\mathbf{y})(\mathbf{y} - \mathbf{x})\} \geq 0$$

が得られる。 $\vartheta \rightarrow 0$ のとき、 $\mathbf{H}(\mathbf{x})$ は半正定値となる。

次に十分性を示そう。任意の $\mathbf{x} \in S$ および $\mathbf{y} \in S$ に対してある $\vartheta \in (0, 1)$ が存在して、式 (2.4.9) が満たされる。 $\mathbf{H}((1 - \vartheta)\mathbf{x} + \vartheta\mathbf{y})$ は半正定値であることより、式 (2.4.9) の右辺第2項は非零である。このとき、

$$f(\mathbf{y}) - f(\mathbf{x}) \geq \mathbf{g}(\mathbf{x}) \cdot (\mathbf{y} - \mathbf{x})$$

が成り立つ。したがって、定理 2.4.4 より f は凸関数である。□

2.4.3 微分法と凸関数の例題

微分法と凸関数に関する定理を簡単な力学の問題に対して使ってみよう。まず、例題 1.1.1 で考えた1自由度ばね系のポテンシャルエネルギーが正定値であることを確かめよう。

例題 2.4.7 (1自由度ばね系のポテンシャルエネルギー) 図 1.2 のような1自由度ばね系のポテンシャルエネルギーは

$$\pi(u) = \int_0^u (kv - p) dv = \frac{1}{2}ku^2 - pu$$

であった。 π は凸関数であることを示せ。□

解答 π に対して

$$\frac{d^2\pi}{du^2} = k > 0$$

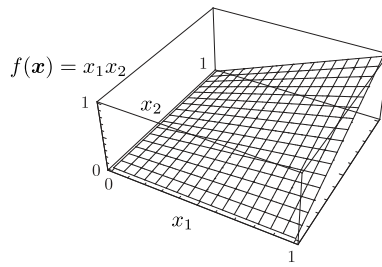
となる。定理 2.4.6 より、 π は凸関数である。□

さらに、2自由度ばね系のポテンシャルエネルギーも凸関数になることを確認しておこう。

例題 2.4.8 (2自由度ばね系のポテンシャルエネルギー) 図 1.3 のような2自由度ばね系のポテンシャルエネルギーは

$$\pi(\mathbf{u}) = \frac{1}{2}k_1u_1^2 + \frac{1}{2}k_2(u_2 - u_1)^2 - (p_1u_1 + p_2u_2)$$

であった。 π は凸関数であることを示せ。□

図 2.13: 関数 $f(\mathbf{x}) = x_1x_2$

解答 π の Hesse 行列は

$$\mathbf{H} = \begin{pmatrix} k_1 + k_2 & -k_2 \\ -k_2 & k_2 \end{pmatrix}$$

となる. \mathbf{H} の固有値 λ は

$$\det \begin{vmatrix} k_1 + k_2 - \lambda & -k_2 \\ -k_2 & k_2 - \lambda \end{vmatrix} = 0$$

を満たすことから,

$$\lambda_1, \lambda_2 = \frac{k_1 + 2k_2 \pm \sqrt{(k_1 + 2k_2)^2 - 4k_1k_2}}{2}$$

となる. $k_1, k_2 > 0$ のとき $\lambda_1, \lambda_2 > 0$ となる. すべての固有値が正なので, \mathbf{H} は正定値行列である (定理 A.2.1). 定理 2.4.6 より, π は凸関数である. \square

この節の最後に, 身近な関数であるが凸関数にならない例をみておこう.

例題 2.4.9 (長方形の面積) $x_1 \in \mathbb{R}$ および $x_2 \in \mathbb{R}$ を長方形の縦と横の長さとする. このとき, 面積 $f(\mathbf{x}) = x_1x_2$ は凸関数ではないことを示せ. \square

解答 $\mathbf{x} = (1, 0)^\top$ および $\mathbf{y} = (0, 1)^\top$ を式 (2.4.4) に代入すれば,

$$\begin{aligned} f((1-\theta)\mathbf{x} + \theta\mathbf{y}) &= \{(1-\theta)x_1 + \theta y_1\} \{(1-\theta)x_2 + \theta y_2\} = (1-\theta)\theta \\ &\geq (1-\theta)f(\mathbf{x}) + \theta f(\mathbf{y}) = 0 \end{aligned}$$

となる. したがって, f は凸関数ではない (図 2.13). \square

2.5 制約なし最適化問題

これより問題 2.1.1 から問題 2.1.3 のいろいろな場合を想定して, 最適化理論を構成する定理をみていこう. まずは, 制約がない場合を考える. あるいは, すべて

の制約が無効である場合とを考えてもよい。すなわち、問題 2.1.3 において、最小点は $S \subseteq X = \mathbb{R}^d$ の内点にある場合である。そこで、この節では S は開集合であるとみなし、 f_0 を f とかくことにして、次の問題を考えることにしよう。

問題 2.5.1 (制約なし最適化問題) $X = \mathbb{R}^d$ とする。 S を X の開部分集合として、 $f: S \rightarrow \mathbb{R}$ が与えられたとき、

$$\min_{x \in S} f(x)$$

を満たす x を求めよ。 □

2.5.1 極小点の必要条件

$x \in S$ が極小点のときに成り立つ条件を極小点の必要条件という。 f が 1 階微分可能ならば微分が定義されて、極小点の必要条件について、次の結果が得られる。

定理 2.5.2 (極小点 1 次の必要条件) 問題 2.5.1 において、 $f \in C^1(S; \mathbb{R})$ とする。このとき、 $x \in S$ が極小点ならば、

$$\mathbf{g}(x) = \mathbf{0}_{\mathbb{R}^d} \tag{2.5.1}$$

が成り立つ。あるいは、任意の $\mathbf{y} \in X$ に対して

$$\mathbf{g}(x) \cdot \mathbf{y} = 0 \tag{2.5.2}$$

が成り立つ。 □

証明 $\mathbf{g}(x) \neq \mathbf{0}_{\mathbb{R}^d}$ において矛盾を導こう。 $\mathbf{y} = -\mathbf{g}(x)$ とおくと、 $\mathbf{g}(x) \cdot \mathbf{y} = -\|\mathbf{g}(x)\|_{\mathbb{R}^d}^2 < 0$ となる。 \mathbf{g} は連続であるので、ある t_1 が存在して、 $t \in [0, t_1]$ に対して

$$\mathbf{g}(x + t\mathbf{y}) \cdot \mathbf{y} < 0$$

が成り立つ。また、平均値の定理より、任意の $t \in (0, t_1]$ に対して

$$f(x + t\mathbf{y}) = f(x) + t\mathbf{g}(x + \theta t\mathbf{y}) \cdot \mathbf{y}$$

となる $\theta \in (0, 1)$ が存在する。ここで、 $\theta t \in (0, t_1)$ より、 $\mathbf{g}(x + \theta t\mathbf{y}) \cdot \mathbf{y} < 0$ である。この関係を上式に代入すれば、 $f(x + t\mathbf{y}) < f(x)$ の矛盾が得られる。 □

定理 2.5.2 では、式 (2.5.1) のようなベクトルどおしの等号を用いた条件と、式 (2.5.2) のような任意のベクトルとの内積がゼロとなる条件が同値として扱われている。両者が同値であるのは、式 (2.5.2) は任意の $\mathbf{y} \in X$ に対して成り立つことを要請しているためである。このような任意のベクトルを用いた表現は、今後、頻繁に現れる。ここでそれらの同値性について了解しておいてほしい。

f が 2 階微分可能ならば、Hesse 行列が定義できて次の結果が得られる。

定理 2.5.3 (極小点 2 次の必要条件) 問題 2.5.1 において, $f \in C^2(S; \mathbb{R})$ とする. このとき, $x \in S$ が極小点ならば, Hesse 行列 $\mathbf{H}(x)$ は半正定値である. \square

証明 x が極小点ならば, 定義 2.3.1 より x の近傍 $B \subset S$ をとり, 任意の $y \in B$ に対して

$$f(y) - f(x) \geq 0$$

とできる. 一方, x は極小点なので, Taylor の定理より

$$f(y) - f(x) = \frac{1}{2}(\mathbf{y} - \mathbf{x}) \cdot \{\mathbf{H}(x)(\mathbf{y} - \mathbf{x})\} + o(\|\mathbf{y} - \mathbf{x}\|_{\mathbb{R}^d}^2)$$

が成り立つ. $z \in X$ を任意として, 両辺に $2z/\|\mathbf{y} - \mathbf{x}\|_{\mathbb{R}^d}^2$ をかけて, $\mathbf{y} \rightarrow \mathbf{x}$ とすれば,

$$z \cdot (\mathbf{H}(x)z) \geq 0$$

が成り立つ. \square

2.5.2 極小点の十分条件

次に, $x \in S$ が極小点になるための条件をみてみよう. それを示すために, 停留点を次のように定義する.

定義 2.5.4 (停留点) $S \subseteq X$ を開集合とする. $x \in S$ において

$$\mathbf{g}(x) = \mathbf{0}_{\mathbb{R}^d}$$

のとき, x を **停留点** という. \square

f が 2 階微分可能ならば極小点の十分条件について, 次の結果が得られる.

定理 2.5.5 (極小点 2 次の十分条件) 問題 2.5.1 において, $f \in C^2(S; \mathbb{R})$ とする. このとき, $x \in S$ が停留点で, かつ Hesse 行列 $\mathbf{H}(x)$ が正定値ならば, x は極小点である. \square

証明 x が停留点ならば, x の近傍 $B \subset S$ をとり, 任意の $x + \mathbf{y} \in B$ に対して, $\theta \in (0, 1)$ が存在して

$$f(x + \mathbf{y}) - f(x) = \frac{1}{2}\mathbf{y} \cdot (\mathbf{H}(x + \theta\mathbf{y})\mathbf{y})$$

が得られる. $\mathbf{H}(x + \theta\mathbf{y})$ は正定値である. したがって,

$$f(x + \mathbf{y}) > f(x)$$

が成り立つ. \square

2.5.3 最小点の十分条件

前項までは極小点についての条件であった。ここでは、最小点になるための条件についてみてみよう。問題 2.5.1 が凸最適化問題ならば、次の結果が得られる。

定理 2.5.6 (最小点の十分条件) 問題 2.5.1 において、 $S \subseteq X$ を空でない開凸集合、 $f: S \rightarrow \mathbb{R}$ を凸関数とする。このとき、 $x \in S$ が極小点ならば、 x は S 上の最小点である。□

証明 x が極小点ならば、 x の近傍 $B \subset S$ をとり、 x は B において最小となる。もしも、 x とは別に最小点 $y \in S$ があったと仮定すると、 $\theta \in (0, 1)$ に対して

$$(1 - \theta)f(x) + \theta f(y) \geq f((1 - \theta)x + \theta y) = f(z)$$

が成り立つ。ここで、 θ を十分小さくとれば $z \in B$ とできる。これは x が極小点であることに反する。そこで、 x は唯一の極小点 (最小点) である。□

2.5.4 制約なし最適化問題の例題

前節と同様、ここでもばね系を用いて制約なし問題に対するこれまでの結果を確認してみよう。まず、1 自由度ばね系について次のことを確認しよう。

例題 2.5.7 (1 自由度ばね系の力のつり合い方程式) 図 1.2 の 1 自由度ばね系における力のつり合い方程式を満たす u は、例題 2.4.7 のポテンシャルエネルギー π の最小点であることを示せ。□

解答 例題 2.4.7 より、 π は凸関数である。そこで、力のつり合い方程式を満たす u は定理 2.5.5 より極小点となる。さらに、定理 2.5.6 より最小点となる。□

多自由度ばね系に対しては次のようになる。

例題 2.5.8 (2 自由度ばね系の力のつり合い方程式) 図 1.3 の 2 自由度ばね系における力のつり合い方程式を満たす u は、例題 2.4.8 のポテンシャルエネルギー π の最小点であることを示せ。□

解答 例題 2.4.8 より、 π は凸関数である。そこで、力のつり合い方程式を満たす u は定理 2.5.5 より極小点となる。さらに、定理 2.5.6 より最小点となる。□

2.5.5 制約なし最適化問題の解法に関する考察

この節で得られた結果とあとで示される結果をあわせて、制約なし最適化問題 (問題 2.5.1) の解法に関して次のことがいえる.

- (1) 定理 2.5.2 より, 極小点であれば停留点となる. そこで, 停留点を求めればその点は極小点の候補となる.
- (2) 停留点 \boldsymbol{x} が求められた上で, Hesse 行列 $\boldsymbol{H}(\boldsymbol{x})$ が正定値であることが確認されれば, 定理 2.5.5 より, \boldsymbol{x} は極小点と判定される.
- (3) f が凸関数であれば, 極小点と判定された停留点 \boldsymbol{x} は, 定理 2.5.6 より, 最小点と判定される.
- (4) f の凸性が不明な場合には, さまざまな試行点から第 3 章で示される最適化法によって極小点を求め, それらの中から最小点を見つける方法がとられる.

2.6 等式制約つき最適化問題

次に, 問題 2.1.1 において等式制約 $\boldsymbol{h}(\boldsymbol{x}) = \mathbf{0}_{\mathbb{R}^n}$ が有効で, かつすべての不等式制約は無効である場合を考えよう. この問題は, 2.1 節の冒頭でも説明されたように, 問題 1.2.2 で使われた断面積 \boldsymbol{a} と変位 \boldsymbol{u} を $\boldsymbol{x} = (\boldsymbol{a}^\top, \boldsymbol{u}^\top)^\top$ とみなし, 目的関数 $f_0(\boldsymbol{u}) = \boldsymbol{p} \cdot \boldsymbol{u}$ を $f_0(\boldsymbol{x})$ のようにかき, 状態方程式 $\boldsymbol{K}(\boldsymbol{a})\boldsymbol{u} - \boldsymbol{p} = \mathbf{0}_{\mathbb{R}^n}$ を等式制約 $\boldsymbol{h}(\boldsymbol{x}) = \mathbf{0}_{\mathbb{R}^n}$ とかいた場合に対応する.

このように, $n < d$ 個の等式制約が課されている場合には, 変数 $\boldsymbol{x} \in X = \mathbb{R}^d$ の要素はすべて独立ではなく, その内の n 個の要素は従属変数になると考えられる. そこで, $\boldsymbol{u} \in U = \mathbb{R}^n$ とおいて, 残りの要素を $\boldsymbol{\xi} \in \Xi = \mathbb{R}^{d-n}$ とおいたときには, $\boldsymbol{x} = (\boldsymbol{\xi}^\top, \boldsymbol{u}^\top)^\top \in X = \Xi \times U$ とかけることになる. 最適設計問題では $\boldsymbol{\xi}$ が設計変数にあたる. この節では, $X = \Xi \times U$ であることに注意しながら, 次の問題を考えることにしよう. なお, この節でも不等式制約関数 f_1, \dots, f_m が現れないことから, f_0 を f とかくことにしよう.

問題 2.6.1 (等式制約つき最適化問題) $X = \mathbb{R}^d$ とする. $n < d$ に対して, $f: X \rightarrow \mathbb{R}$ および $\boldsymbol{h} = (h_1, \dots, h_n)^\top: X \rightarrow \mathbb{R}^n$ が与えられたとき,

$$\min_{\boldsymbol{x} \in X} \{ f(\boldsymbol{x}) \mid \boldsymbol{h}(\boldsymbol{x}) = \mathbf{0}_{\mathbb{R}^n} \}$$

を満たす \boldsymbol{x} を求めよ.

□

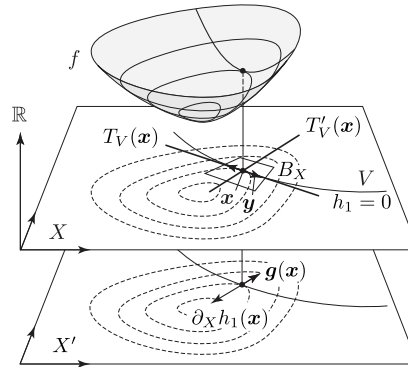


図 2.14: 等式制約つき最適化問題の極小点 ($X = \mathbb{R}^2$, $n = 1$)

2.6.1 極小点の必要条件

x が問題 2.6.1 の極小点ならば成り立つ関係について考えてみよう. 図 2.14 に $X = \mathbb{R}^2$ および $n = 1$ の場合の極小点のようすを示す. この図をみながら, 極小点ならば成り立つ関係についてみてから, 一般の場合について考えることにしよう.

図 2.14 の x を極小点とする. すなわち, B_X を x の近傍とすれば, x から等式制約 $h_1(x + y) = 0$ を満たしながら $x + y \in B_X$ に移動したとき, $f(x) \leq f(x + y)$ が成り立つことになる. ここで, f と h_1 が $C^1(B_X; \mathbb{R})$ の要素であると仮定する. このとき, $\partial_X f = g$ と $\partial_X h_1$ が定義できて, 次の関係が成り立つ.

(1) y と $\partial_X h_1$ は直交する.

(2) y と g は直交する.

(1) は, 制約を満たす y の方向に移動しても制約が満たされる関係を表している. 実際,

$$h_1(x + y) = h_1(x) + \partial_X h_1 \cdot y + o(\|y\|_{\mathbb{R}^d})$$

とかけるので, x の近傍 $x + y \in B_X$ で $h_1(x + y) = h_1(x)$ であることは $\partial_X h_1 \cdot y = 0$ と同値である. 一方, (2) は, 制約を満たす方向に移動しても目的関数に変動しないことを表している. この関係は定理 2.5.2 において目的関数の変動方向が y に限定されたのみで, 式 (2.5.2) と同じ関係を表している. これらの関係を一般化して, 等式制約つき最適化問題の極小点に対する必要条件について考えてみよう.

まず, (1) の関係を一般化しよう. 許容集合を

$$V = \{x \in X \mid h(x) = \mathbf{0}_{\mathbb{R}^n}\} \quad (2.6.1)$$

とおく. 図 2.14 では $h_1 = 0$ を満たす曲線上の点の集合となる. ここで, $\mathbf{x} \in V$ の近傍を $B_X \subset X$ とかいて, $\mathbf{h} \in C^1(B_X; \mathbb{R}^n)$ のとき,

$$T_V(\mathbf{x}) = \{\mathbf{y} \in X \mid \mathbf{h}_{\mathbf{x}^\top}(\mathbf{x})\mathbf{y} = \mathbf{0}_{\mathbb{R}^n}\} \quad (2.6.2)$$

を \mathbf{x} における許容方向集合あるいは接面という. ただし, $\mathbf{h}_{\mathbf{x}^\top} = (\partial h_i / \partial x_j) = (\partial_X h_1, \dots, \partial_X h_n)^\top \in \mathbb{R}^{n \times d}$ は \mathbf{h} の \mathbf{x} の変動に対する Jacobi 行列に相当し, 行列の階数は n とする. すなわち, $\partial_X h_1, \dots, \partial_X h_n$ は 1 次独立であることと同値である. 図 2.14 では曲線 $h_1 = 0$ の \mathbf{x} における接線を $T_V(\mathbf{x})$ とおいたことになる.

$X = \mathbb{R}^3$ のときは $n = 1$ と $n = 2$ の場合が考えられる. 図 2.15 は, それらの場合の $T_V(\mathbf{x})$ を示す. $n = 1$ の場合, $h_1 = 0$ を満たす点の集合 V は曲面になり, $T_V(\mathbf{x})$ は, 文字通り, \mathbf{x} におけるその曲面の接面になる. $n = 2$ の場合, $h_1 = 0$ と $h_2 = 0$ を同時に満たす点の集合 V は曲線になり, $T_V(\mathbf{x})$ は \mathbf{x} におけるその曲線の接線になる. $n = 2$ のときに, $\mathbf{h}_{\mathbf{x}^\top}(\mathbf{x}) = (\partial_X h_1(\mathbf{x}), \partial_X h_2(\mathbf{x}))^\top$ の階数が $n = 2$ であることは, $\partial_X h_1(\mathbf{x})$ と $\partial_X h_2(\mathbf{x})$ が異なる方向を向いていることを意味する. 式 (2.6.2) による $T_V(\mathbf{x})$ の定義は, 零空間と像空間あるいは核空間と値空間の定義 (A.3 節) に従えば,

$$T_V(\mathbf{x}) = \text{Ker } \mathbf{h}_{\mathbf{x}^\top}(\mathbf{x})$$

とかける.

それに対して, 上記 (2) の条件を一般化すると次のようになる. 上記 (2) の条件は, \mathbf{g} は $T_V(\mathbf{x})$ に含まれるすべてのベクトルと直交することを表していた. そこで,

$$T'_V(\mathbf{x}) = (T_V(\mathbf{x}))' = \{\mathbf{z} \in X' \mid \mathbf{z} \cdot \mathbf{y} = 0 \text{ for all } \mathbf{y} \in T_V(\mathbf{x})\} \quad (2.6.3)$$

とおき, $T'_V(\mathbf{x})$ を $T_V(\mathbf{x})$ の双対集合あるいは双対面とよぶことにする. このとき, f と \mathbf{h} が 1 階微分可能で, T'_V が評価可能ならば, 次の結果が得られる.

定理 2.6.2 (極小点 1 次の必要条件) 問題 2.6.1 において, $f \in C^1(X; \mathbb{R})$ および $\mathbf{h} \in C^1(X; \mathbb{R}^n)$ で, $\mathbf{x} \in V$ において $\partial_X h_1(\mathbf{x}), \dots, \partial_X h_n(\mathbf{x})$ が 1 次独立とする. このとき, \mathbf{x} が極小点ならば, 任意の $\mathbf{y} \in T_V(\mathbf{x})$ に対して,

$$\mathbf{g}(\mathbf{x}) \cdot \mathbf{y} = 0 \quad (2.6.4)$$

が成り立つ. また,

$$\mathbf{g}(\mathbf{x}) \in T'_V(\mathbf{x}) \quad (2.6.5)$$

が成り立つ. □

証明 任意の $\mathbf{y} \in T_V(\mathbf{x})$ に対して, $\mathbf{g}(\mathbf{x}) \cdot \mathbf{y} \neq 0$ とすると, $\mathbf{g} \cdot \mathbf{y} < 0$ となるような \mathbf{y} がとれる. このとき, 定理 2.5.2 と同様の矛盾が得られる. 式 (2.6.5) は, $T'_V(\mathbf{x})$ の定義より, 式 (2.6.4) と同値である. □

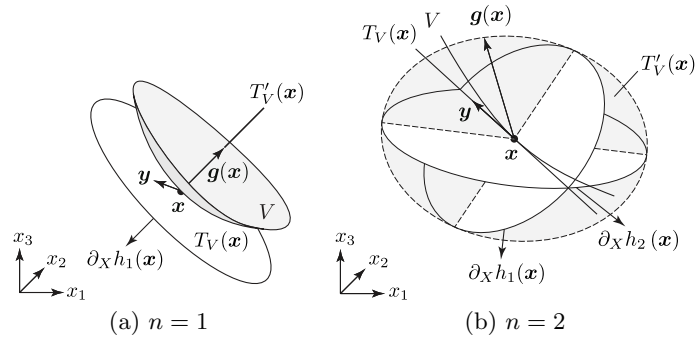


図 2.15: 等式制約つき最適化問題の極小点 ($X = \mathbb{R}^3$ と $X' = \mathbb{R}^3$ を重ねている)

2.6.2 Lagrange 乗数法

定理 2.6.2 は, $x \in V$ が極小点ならば成り立つ条件を任意の $y \in T_V(x)$ あるいは $T'_V(x)$ を用いて与えていた. その意味は分かり易くても, その評価は容易ではない. そこで, 任意の $y \in T_V(x)$ や $T'_V(x)$ を用いない表現方法について考えてみよう.

最初に, 図 2.14 において基本的な関係を確認してから一般化していくことにしよう. ここでも, f と h_1 が $C^1(B_X; \mathbb{R})$ の要素であると仮定する. ここで, g と $y \in T_V(x)$ は直交し, $y \in T_V(x)$ と $\partial_X h_1$ が直交することに注目する. この関係は, g と $\partial_X h_1$ が同じ方向を向いていることと同値である. さらに, その関係は,

$$g + \lambda_1 \partial_X h_1 = \mathbf{0}_{\mathbb{R}^2} \quad (2.6.6)$$

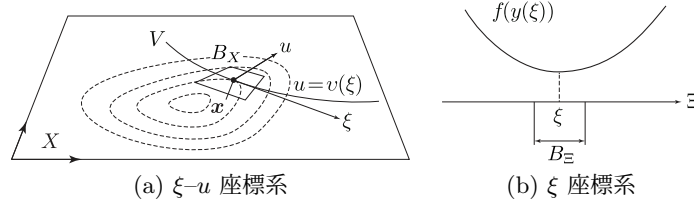
が成り立つようなある $\lambda_1 \in \mathbb{R}$ が存在することと同値である. なぜならば, もしも, g と $\partial_X h_1$ が同じ方向ではない非零のベクトルならば式 (2.6.6) が成り立つような λ_1 は存在しないためである. 実際, 方向が異なる二つのベクトル $a, b \in \mathbb{R}^d$ が固定されたとき, $a + \lambda b = \mathbf{0}_{\mathbb{R}^d}$ を満たす $\lambda \in \mathbb{R}$ が存在しないことを確認してみよ.

式 (2.6.6) を一般化しよう. 最初に, 次のことを一般化する.

- $y \in T_V(x)$ と g は直交する.

この条件は, 変数 x が等式制約が満たされる方向に移動したときの評価関数 f の勾配がゼロになる条件から得られる. V を式 (2.6.1) の許容集合とする. ここで, 次の条件が満たされていると仮定する.

仮定 2.6.3 (陰関数定理の仮定) $d > n$ のもとで $X = \mathbb{R}^d = \Xi \times \mathbb{R}^n$ とする. $h : X \rightarrow \mathbb{R}^n$ は, $x = \left(\xi_0^\top, u_0^\top \right)^\top \in \Xi \times \mathbb{R}^n$ の近傍 $B_X = B_\Xi \times B_{\mathbb{R}^n}$ において,

図 2.16: 等式制約を満たす集合と ξ - u 局所座標系

- (1) $\mathbf{h}(\mathbf{x}) = \mathbf{0}_{\mathbb{R}^n}$ (すなわち $\mathbf{x} \in V$),
- (2) $\mathbf{h} \in C^0(B_X; \mathbb{R}^n)$,
- (3) 任意の $\tilde{\mathbf{x}} = (\boldsymbol{\xi}^\top, \mathbf{u}^\top)^\top \in B_\Xi \times B_{\mathbb{R}^n}$ に対して $\mathbf{h}(\boldsymbol{\xi}, \cdot) \in C^1(B_{\mathbb{R}^n}; \mathbb{R}^n)$,
- (4) \mathbf{x} における Jacobi 行列 $\mathbf{h}_{\mathbf{u}^\top}(\mathbf{x})$ は正則

を満たしている. □

このとき, [陰関数定理](#) (定理 A.4.1) より, ある近傍 (凸開集合) $U_\Xi \times U_{\mathbb{R}^n} \subset B_\Xi \times B_{\mathbb{R}^n}$ と連続な関数 $\mathbf{v}: U_\Xi \rightarrow U_{\mathbb{R}^n}$ (\mathbf{v} はギリシャ文字 ϵ の太文字) が存在して, $\mathbf{h}(\mathbf{x}) = \mathbf{0}_{\mathbb{R}^n}$ は

$$\mathbf{u} = \mathbf{v}(\boldsymbol{\xi}) \quad (2.6.7)$$

と同値となる. 図 2.16 は, $X = \mathbb{R}^2$ および $n = 1$ のときの等式制約を満たす集合と ξ - u 局所座標系を示す.

そこで,

$$\tilde{\mathbf{x}}(\boldsymbol{\xi}) = (\boldsymbol{\xi}^\top, \mathbf{v}^\top(\boldsymbol{\xi}))^\top$$

とおけば, $\tilde{\mathbf{x}}(\boldsymbol{\xi}) \in V$ となる. $\boldsymbol{\xi}$ を V の局所座標という.

さらに, $\tilde{f}(\boldsymbol{\xi}) = f(\tilde{\mathbf{x}}(\boldsymbol{\xi}))$ とかく. $f \in C^1(B_X; \mathbb{R}^n)$ ならば, $\boldsymbol{\xi} \in B_\Xi$ および $\mathbf{x} = \tilde{\mathbf{x}}(\boldsymbol{\xi}) \in B_X$ が極小点のとき,

$$\partial_\Xi \tilde{f}(\boldsymbol{\xi}) = \left(\frac{\partial \tilde{\mathbf{x}}}{\partial \boldsymbol{\xi}^\top}(\boldsymbol{\xi}) \right)^\top \frac{\partial f}{\partial \mathbf{x}}(\mathbf{x}) = (\tilde{\mathbf{x}}_{\boldsymbol{\xi}^\top}(\mathbf{x}))^\top \mathbf{g}(\mathbf{x}) = \mathbf{0}_{\mathbb{R}^{d-n}} \quad (2.6.8)$$

が成り立つ. この関係は, 零空間と像空間の定義に従えば,

$$\mathbf{g}(\mathbf{x}) \in T'_V(\mathbf{x}) = \text{Ker}(\tilde{\mathbf{x}}_{\boldsymbol{\xi}^\top}(\mathbf{x}))^\top \quad (2.6.9)$$

とかける.

一方,

- $\mathbf{y} \in T_V(\mathbf{x})$ と $\partial_X h_1$ は直交する.

の一般化は、変数 \mathbf{x} が等式制約を満たす条件から得られる. $\mathbf{h}(\mathbf{x}) = \mathbf{h}(\tilde{\mathbf{x}}(\boldsymbol{\xi})) = \mathbf{0}_{\mathbb{R}^n}$ の両辺を $\boldsymbol{\xi}$ で微分すると,

$$\frac{\partial \mathbf{h}}{\partial \mathbf{x}^\top}(\mathbf{x}) \frac{\partial \tilde{\mathbf{x}}}{\partial \boldsymbol{\xi}^\top}(\boldsymbol{\xi}) = \mathbf{h}_{\mathbf{x}^\top}(\mathbf{x}) \tilde{\mathbf{x}}_{\boldsymbol{\xi}^\top}(\mathbf{x}) = \mathbf{0}_{\mathbb{R}^n \times (d-n)} \quad (2.6.10)$$

が成り立つ. この関係は、零空間と像空間の定義に従えば、 $\tilde{\mathbf{x}}_{\boldsymbol{\xi}^\top}(\mathbf{x})$ の像空間が $\mathbf{h}_{\mathbf{x}^\top}(\mathbf{x})$ の零空間になっていることから,

$$T_V(\mathbf{x}) = \text{Ker } \mathbf{h}_{\mathbf{x}^\top}(\mathbf{x}) = \text{Im } \tilde{\mathbf{x}}_{\boldsymbol{\xi}^\top}(\mathbf{x}) \quad (2.6.11)$$

とかける.

これらの関係を用いれば、 $T_V(\mathbf{x})$ あるいは $T'_V(\mathbf{x})$ を用いない次のような極小点の必要条件が得られる.

定理 2.6.4 (Lagrange 関数を用いた極小点 1 次の必要条件) 問題 2.6.1 において、 $f \in C^1(X; \mathbb{R})$ および $\mathbf{h} \in C^1(X; \mathbb{R}^n)$ とする. $\mathbf{x} \in X$ において $\partial_X h_1(\mathbf{x}), \dots, \partial_X h_n(\mathbf{x})$ が 1 次独立とする. このとき、 \mathbf{x} が極小点ならば、

$$\mathbf{g}(\mathbf{x}) + \partial_X \mathbf{h}^\top(\mathbf{x}) \boldsymbol{\lambda} = \mathbf{0}_{\mathbb{R}^d}, \quad (2.6.12)$$

$$\mathbf{h}(\mathbf{x}) = \mathbf{0}_{\mathbb{R}^n} \quad (2.6.13)$$

を満たす $\boldsymbol{\lambda} \in \mathbb{R}^n$ が存在する. \square

証明 定理の仮定より、 \mathbf{x} において仮定 2.6.3 が成り立つ. \mathbf{x} が問題 2.6.1 の極小点ならば、式 (2.6.9) が成り立つ. それに、式 (2.6.11) を用いれば、

$$\mathbf{g}(\mathbf{x}) \in T'_V(\mathbf{x}) = (T_V(\mathbf{x}))^\perp = (\text{Im } \tilde{\mathbf{x}}_{\boldsymbol{\xi}^\top}(\mathbf{x}))^\perp = (\text{Ker } \mathbf{h}_{\mathbf{x}^\top}(\mathbf{x}))^\perp$$

が得られる. さらに、零空間と像空間の直交補空間に関する補題 A.3.1 より、

$$\mathbf{g}(\mathbf{x}) \in (\text{Ker } \mathbf{h}_{\mathbf{x}^\top}(\mathbf{x}))^\perp = \text{Im } (\mathbf{h}_{\mathbf{x}^\top}(\mathbf{x}))^\top = \text{Im } (\partial_X h_1, \dots, \partial_X h_n)$$

が得られる. この関係は、式 (2.6.12) と同値である. また、式 (2.6.13) は \mathbf{x} が問題 2.6.1 の極小点ならば成り立つ. \square

定理 2.6.4 の証明で現れた最後の関係は、式 (2.6.6) の一般化になっている. すなわち、 $\mathbf{g}(\mathbf{x})$ は $\partial_X h_1, \dots, \partial_X h_n$ の線形結合で表されることを意味しているからである. 定理 2.6.4 の中に現れた、 $\boldsymbol{\lambda} = (\lambda_1, \dots, \lambda_n)^\top \in \mathbb{R}^n$ は等式制約 $\mathbf{h}(\mathbf{x}) = \mathbf{0}_{\mathbb{R}^n}$ に対する Lagrange 乗数とよばれる. さらに、定理 2.6.4 の式 (2.6.12) と式 (2.6.13) は Lagrange 関数を用いた極小点 1 次の必要条件とよばれる. その理由は、次の関係が成り立つためである. 問題 2.6.1 の最適化問題に対する Lagrange 関数を

$$\mathcal{L}(\mathbf{x}, \boldsymbol{\lambda}) = f(\mathbf{x}) + \boldsymbol{\lambda} \cdot \mathbf{h}(\mathbf{x}) \quad (2.6.14)$$

と定義する. $(\mathbf{x}, \boldsymbol{\lambda})$ の任意変動 $(\mathbf{y}, \hat{\boldsymbol{\lambda}}) \in X \times \mathbb{R}^n$ に対する \mathcal{L} の微分は

$$\begin{aligned}\mathcal{L}'(\mathbf{x}, \boldsymbol{\lambda})[\mathbf{y}, \hat{\boldsymbol{\lambda}}] &= f'(\mathbf{x})[\mathbf{y}] + \boldsymbol{\lambda} \cdot (\partial_{\mathbf{x}^\top} \mathbf{h}(\mathbf{x}) \mathbf{y}) + \hat{\boldsymbol{\lambda}} \cdot \mathbf{h}(\mathbf{x}) \\ &= \mathbf{g}(\mathbf{x}) \cdot \mathbf{y} + (\partial_X \mathbf{h}^\top(\mathbf{x}) \boldsymbol{\lambda}) \cdot \mathbf{y} + \hat{\boldsymbol{\lambda}} \cdot \mathbf{h}(\mathbf{x})\end{aligned}\quad (2.6.15)$$

となる. このとき, 任意変動 $(\mathbf{y}, \hat{\boldsymbol{\lambda}}) \in X \times \mathbb{R}^n$ に対する Lagrange 関数 \mathcal{L} の極小点 1 次の必要条件 (停留条件) $\mathcal{L}'(\mathbf{x}, \boldsymbol{\lambda})[\mathbf{y}, \hat{\boldsymbol{\lambda}}] = 0$ は, 定理 2.6.4 の式 (2.6.12) と式 (2.6.13) と同値となる.

そこで, 次の問題の解を問題 2.6.1 の解の候補とする方法を考えることができる. その方法を **等式制約つき最適化問題に対する Lagrange 乗数法** という.

問題 2.6.5 (等式制約つき最適化問題に対する Lagrange 乗数法) 問題 2.6.1 に対して, $\mathcal{L}(\mathbf{x}, \boldsymbol{\lambda})$ を式 (2.6.14) とおく. $\mathcal{L}(\mathbf{x}, \boldsymbol{\lambda})$ の停留条件

$$\partial_X \mathcal{L}(\mathbf{x}, \boldsymbol{\lambda}) = \mathbf{g}(\mathbf{x}) + \partial_X \mathbf{h}^\top(\mathbf{x}) \boldsymbol{\lambda} = \mathbf{0}_{\mathbb{R}^d}, \quad (2.6.16)$$

$$\frac{\partial \mathcal{L}}{\partial \boldsymbol{\lambda}}(\mathbf{x}, \boldsymbol{\lambda}) = \mathbf{h}(\mathbf{x}) = \mathbf{0}_{\mathbb{R}^n} \quad (2.6.17)$$

を満たす $(\mathbf{x}, \boldsymbol{\lambda})$ を求めよ. □

Lagrange 乗数法は, 今後, 様々な場面で使われる. Lagrange 乗数法は, 問題 2.6.1 の解法を意味しているわけではなく, 問題 2.6.1 の極小点における条件を表現する方法という意味で理解してほしい. ここでは, Lagrange 乗数の物理的意味についてみておこう. 式 (2.6.16) は, 任意の $\mathbf{y} \in X$ に対して

$$\lambda_i = - \frac{(\mathbf{g}(\mathbf{x}) + \sum_{j \in \{1, \dots, n\}, j \neq i} \lambda_j \partial_X h_j(\mathbf{x})) \cdot \mathbf{y}}{\partial_X h_i(\mathbf{x}) \cdot \mathbf{y}} \quad (2.6.18)$$

のようにかかる. f および h_1, \dots, h_n が力学量するとき, λ_i は f/h_i の単位をもつ力学量となる. 実際, 問題 1.1.4 では, 等式制約 $\mathbf{K}(\mathbf{a}) \mathbf{u} = \mathbf{p}$ は力 [N] の単位をもち, $f_0 = \mathbf{p} \cdot \mathbf{u}$ は仕事 [Nm] の単位をもち, 状態方程式に対する Lagrange 乗数 (随伴変数) として導入された \mathbf{v}_0 は変位 [m = Nm/N] の単位をもっていた. 物理的意味については, 不等式制約つき最適化問題に対する Lagrange 乗数法についても同様である. 問題 1.1.4 では, f_1 は体積 [m³] の単位をもち, f_0 は仕事 [Nm] の単位をもっていた. そこで, Lagrange 乗数 λ_1 はエネルギー密度 [N/m² = Nm/m³] の単位をもっていたわけである.

定理 2.6.4 では評価関数が 1 階微分可能であることが仮定されていた. さらに 2 階微分可能ならば次の結果が得られる. 以後, Lagrange 関数の \mathbf{x} の変動に対する Hesse 行列を $\partial_X \partial_X^\top \mathcal{L}(\mathbf{x}, \boldsymbol{\lambda}) = \mathbf{H}_{\mathcal{L}}(\mathbf{x}, \boldsymbol{\lambda}) \in \mathbb{R}^{d \times d}$ とかくことにしよう.

定理 2.6.6 (Lagrange 関数を用いた極小点 2 次の必要条件) 問題 2.6.1 において, $f \in C^2(X; \mathbb{R})$ および $\mathbf{h} \in C^2(X; U)$ とする. $\mathbf{x} \in V$ において $\partial_X h_1(\mathbf{x}), \dots, \partial_X h_n(\mathbf{x})$ が 1 次独立とする. このとき, \mathbf{x} が極小点ならば, 任意の $\mathbf{y} \in T_V(\mathbf{x})$ に対して,

$$\mathbf{y} \cdot (\mathbf{H}_{\mathcal{L}}(\mathbf{x}, \boldsymbol{\lambda}) \mathbf{y}) \geq 0$$

が成り立つ. □

証明 式 (2.6.8) と同様の計算により

$$\begin{aligned} \frac{\partial^2 \tilde{f}}{\partial \boldsymbol{\xi} \partial \boldsymbol{\xi}^\top}(\boldsymbol{\xi}) &= \left(\frac{\partial \mathbf{y}}{\partial \boldsymbol{\xi}^\top}(\boldsymbol{\xi}) \right)^\top \frac{\partial^2 f}{\partial \mathbf{x} \partial \mathbf{x}^\top}(\mathbf{x}) \frac{\partial \mathbf{y}}{\partial \boldsymbol{\xi}^\top}(\boldsymbol{\xi}) \\ &= (\mathbf{y}_{\boldsymbol{\xi}^\top}(\mathbf{x}))^\top \partial_X \partial_X^\top f(\mathbf{x}) \mathbf{y}_{\boldsymbol{\xi}^\top}(\mathbf{x}) \in \mathbb{R}^{(d-n) \times (d-n)} \end{aligned}$$

が成り立つ. \mathbf{x} が極小点ならば, $\partial^2 \tilde{f} / \partial \boldsymbol{\xi} \partial \boldsymbol{\xi}^\top(\boldsymbol{\xi})$ は正定値行列である. また, 式 (2.6.10) から

$$\begin{aligned} \left(\frac{\partial \mathbf{y}}{\partial \boldsymbol{\xi}^\top}(\boldsymbol{\xi}) \right)^\top \frac{\partial h_i}{\partial \mathbf{x} \partial \mathbf{x}^\top}(\mathbf{x}) \frac{\partial \mathbf{y}}{\partial \boldsymbol{\xi}^\top}(\boldsymbol{\xi}) \\ = (\mathbf{y}_{\boldsymbol{\xi}^\top}(\mathbf{x}))^\top \partial_X \partial_X^\top h_i(\mathbf{x}) \mathbf{y}_{\boldsymbol{\xi}^\top}(\mathbf{x}) = \mathbf{0}_{\mathbb{R}^{(d-n) \times (d-n)}} \end{aligned}$$

が成り立つ. $\mathcal{L}(\mathbf{x}, \boldsymbol{\lambda}) = f(\mathbf{x}) + \boldsymbol{\lambda} \cdot \mathbf{h}(\mathbf{x})$ より, 定理の式が成り立つ. □

定理 2.6.6 に基づけば, 極小点 \mathbf{x} における Lagrange 関数 $\mathcal{L}(\mathbf{x}, \boldsymbol{\lambda})$ は, 接面 $T_V(\mathbf{x})$ に対して $\tilde{f}(\boldsymbol{\xi})$ を 2 次関数で近似した関数になっていると解釈される. 実際, 極小点 \mathbf{x} では $\partial \tilde{f} / \partial \boldsymbol{\xi} = \mathbf{0}_{\mathbb{R}^{d-n}}$ となり, 定理 2.6.6 の証明より,

$$\begin{aligned} \frac{\partial^2 \tilde{f}}{\partial \boldsymbol{\xi} \partial \boldsymbol{\xi}^\top}(\boldsymbol{\xi}) &= (\mathbf{y}_{\boldsymbol{\xi}^\top}(\mathbf{x}))^\top \partial_X \partial_X^\top \left(f(\mathbf{x}) + \sum_{i \in \{1, \dots, n\}} \lambda_i h_i \right) \mathbf{y}_{\boldsymbol{\xi}^\top}(\mathbf{x}) \\ &= (\mathbf{y}_{\boldsymbol{\xi}^\top}(\mathbf{x}))^\top \partial_X \partial_X^\top \mathcal{L}(\mathbf{x}, \boldsymbol{\lambda}) \mathbf{y}_{\boldsymbol{\xi}^\top}(\mathbf{x}) \end{aligned}$$

が成り立つ. そこで, 任意の $\boldsymbol{\eta} \in \mathbb{R}^{d-n}$ に対して $\mathbf{y} = \mathbf{y}_{\boldsymbol{\xi}^\top}(\mathbf{x}) \boldsymbol{\eta} \in T_V(\mathbf{x})$ とおけば,

$$\tilde{f}(\boldsymbol{\xi} + \boldsymbol{\eta}) = \mathcal{L}(\mathbf{x}, \boldsymbol{\lambda}) + \mathbf{y}^\top \partial_X \partial_X^\top \mathcal{L}(\mathbf{x}, \boldsymbol{\lambda}) \mathbf{y} + o(\|\mathbf{y}\|_X^2)$$

が成り立つからである.

2.6.3 極小点の十分条件

さらに, 極小点であるための十分条件について, 次の結果が得られる.

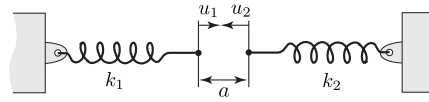


図 2.17: ばねの結合問題

定理 2.6.7 (Lagrange 関数を用いた極小点 2 次の十分条件) 問題 2.6.5 において, $f \in C^2(X; \mathbb{R})$ および $h \in C^2(X; U)$ とする. $\mathbf{x} \in X$ において $\partial_X h_1(\mathbf{x}), \dots, \partial_X h_n(\mathbf{x})$ が 1 次独立とする. このとき, \mathbf{x} が問題 2.6.5 の解で, 任意の $\mathbf{y} \in T_V(\mathbf{x})$ に対して

$$\mathbf{y} \cdot (H_{\mathcal{L}}(\mathbf{x}, \boldsymbol{\lambda}) \mathbf{y}) > 0$$

を満たす $(\mathbf{x}, \boldsymbol{\lambda})$ が存在するならば, \mathbf{x} は問題 2.6.1 の極小点である. □

証明 \bar{f} に対して定理 2.5.5 の証明を用いる. □

2.6.4 等式制約つき最適化問題の例題

ばね系の等式制約つき最適化問題を Lagrange 乗数法で解いてみよう.

例題 2.6.8 (ばねの結合問題) 図 2.17 の 2 自由度ばね系に対して, k_1 と k_2 をばね定数を表す正の実定数, a をばねのすきま (長さ) を表す正の実定数とする. これらのばねが結合されたときにポテンシャルエネルギーが最小となる変位 $\mathbf{u} = (u_1, u_2)^T \in \mathbb{R}^2$ を求めよ. すなわち,

$$\min_{\mathbf{u} \in \mathbb{R}^2} \left\{ f(\mathbf{u}) = \frac{1}{2} k_1 u_1^2 + \frac{1}{2} k_2 u_2^2 \mid h_1(\mathbf{u}) = a - (u_1 + u_2) = 0 \right\}$$

を満たす \mathbf{u} を求めよ. □

解答 最初に, 代入法で解いてみよう. $u_2 = a - u_1$ とおき,

$$f(\mathbf{u}) = \bar{f}(u_1) = \frac{1}{2} k_1 u_1^2 + \frac{1}{2} k_2 (a - u_1)^2$$

とかく. このとき,

$$\frac{d\bar{f}}{du_1}(u_1) = k_1 u_1 - k_2 (a - u_1) = (k_1 + k_2) u_1 - k_2 a = 0$$

より, \bar{f} の停留点は

$$u_1 = \frac{k_2}{k_1 + k_2} a, \quad u_2 = a - u_1 = \frac{k_1}{k_1 + k_2} a$$

となる。さらに,

$$\frac{d^2 \bar{f}}{du_1^2}(u_1) = k_1 + k_2 > 0$$

より, $(u_1, u_2)^\top$ は最小点である。

次に, Lagrange 乗数法で解いてみよう。Lagrange 関数を

$$\mathcal{L}(\mathbf{u}, \lambda) = \frac{1}{2}k_1 u_1^2 + \frac{1}{2}k_2 u_2^2 + \lambda(a - u_1 - u_2)$$

とおく。 $\mathcal{L}(\mathbf{u}, \lambda)$ の停留条件は

$$\begin{pmatrix} \mathcal{L}_{u_1} \\ \mathcal{L}_{u_2} \\ \mathcal{L}_\lambda \end{pmatrix} = \begin{pmatrix} k_1 u_1 - \lambda \\ k_2 u_2 - \lambda \\ a - u_1 - u_2 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}$$

となる。すなわち,

$$\begin{pmatrix} k_1 & 0 & -1 \\ 0 & k_2 & -1 \\ 1 & 1 & 0 \end{pmatrix} \begin{pmatrix} u_1 \\ u_2 \\ \lambda \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ a \end{pmatrix}$$

とかける。この式より

$$\begin{pmatrix} u_1 \\ u_2 \\ \lambda \end{pmatrix} = \frac{1}{k_1 + k_2} \begin{pmatrix} 1 & -1 & k_2 \\ -1 & 1 & k_1 \\ -k_2 & -k_1 & k_1 k_2 \end{pmatrix} \begin{pmatrix} 0 \\ 0 \\ a \end{pmatrix} = \begin{pmatrix} \frac{k_2}{k_1 + k_2} a \\ \frac{k_1}{k_1 + k_2} a \\ \frac{k_1 k_2}{k_1 + k_2} a \end{pmatrix}$$

が得られる。 \mathbf{u} は代入法の結果と一致する。また, $\lambda = k_1 u_1 = k_2 u_2$ は内力の意味を持つ。

さらに, Lagrange 関数の $\mathbf{u} \in U = \mathbb{R}^2$ の変動に対する Hesse 行列

$$\partial_U \partial_U^\top \mathcal{L}(\mathbf{u}, \lambda) = \mathbf{H}_{\mathcal{L}}(\mathbf{u}, \lambda) = \begin{pmatrix} k_1 & 0 \\ 0 & k_2 \end{pmatrix}$$

は \mathbf{u} および λ によらず正定値である。そこで, 定理 2.6.7 より, \mathbf{u} は極小点である。さらに, あとで示される定理 2.7.9 の系 (系 2.7.10) より, \mathbf{u} は最小点であることが示される。 \square

2.6.5 直接微分法と随伴変数法

これまで, 等式制約つき最適化問題 (問題 2.6.1) について極小点の必要条件と十分条件についてみてきた。ここでは, 問題 2.6.1 を第1章でみてきたような最適設計問題の形式におきかえて, 設計変数の変動に対する評価関数の微分を計算する方法についてみておきたい。第1章ではそれらの方法を直接微分法と随伴変数法とよんで手順だけを見てきた。ここでは, それらの定義を示し, 随伴変数法は Lagrange 乗数法と同値であることをみておこう。

第1章でとりあげた最適設計問題における状態決定問題は等式制約に相当する。ここでは状態決定問題を次のように定義する。

問題 2.6.9 (線形システム問題) $\Xi = \mathbb{R}^{d-n}$ および $U = \mathbb{R}^n$ とおく. $\mathbf{K} : \Xi \rightarrow \mathbb{R}^{n \times n}$ および $\mathbf{b} : \Xi \rightarrow \mathbb{R}^n$ は与えられたとする. $\xi \in \Xi$ に対して

$$\mathbf{K}(\xi) \mathbf{u} = \mathbf{b}(\xi) \quad (2.6.19)$$

を満たす $\mathbf{u} \in U$ を求めよ. \square

このような状態決定問題を等式制約とする等式制約つき最適化問題を次のように定義する.

問題 2.6.10 (等式制約つき最適化問題) 問題 2.6.9 において $\mathbf{K} \in C^1(\Xi; \mathbb{R}^{n \times n})$ および $\mathbf{b} \in C^1(\Xi; \mathbb{R}^n)$ とする. $f \in C^1(\Xi \times U; \mathbb{R})$ が与えられたとき,

$$\min_{(\xi, \mathbf{u}) \in \Xi \times U} \{f(\xi, \mathbf{u}) \mid \text{問題 2.6.9}\}$$

を満たす (ξ, \mathbf{u}) を求めよ. \square

解説を始める前に, 第1章において断面積微分とよんだ設計変数の変動に対する評価関数の微分は, 2.6節においてこれまで使われてきた $\mathbf{g}(\mathbf{x})$ とは定義が異なることを注意しておこう. 実際, 式 (2.6.8) では, $\tilde{f}(\xi) = f(\xi, \mathbf{v}(\xi))$ の $\xi \in \Xi$ に対する微分を $\partial_{\Xi} \tilde{f}(\xi) = (\mathbf{y}_{\xi^{\top}}(\mathbf{x}))^{\top} \mathbf{g}(\mathbf{x})$ とかいた. このときの $\mathbf{g}(\mathbf{x})$ は $\partial_{\mathbf{x}} f \in \mathbb{R}^d$ の意味で使われた. 一方, 第1章では $\partial_{\Xi} \tilde{f}_0 \in \mathbb{R}^{d-n}$ を \mathbf{g}_0 とかいていた. ここでは, 第2章の表記法にあわせて, $\tilde{\mathbf{g}} = \partial_{\Xi} \tilde{f}$ とかくことにする.

まず, 直接微分法の定義を示そう. f, \mathbf{K} および \mathbf{b} が1階微分可能ならば,

$$\tilde{\mathbf{g}} = \partial_{\Xi} \tilde{f}(\xi) = \frac{\partial f}{\partial \xi}(\xi, \mathbf{u}) + \left(\frac{\partial \mathbf{u}}{\partial \xi^{\top}}(\xi) \right)^{\top} \frac{\partial f}{\partial \mathbf{u}}(\xi) \quad (2.6.20)$$

が成り立つ. 一方, 式 (2.6.19) を ξ_1, \dots, ξ_{d-n} で偏微分して, 得られた列ベクトルを横に並べた行列を

$$\frac{\partial \mathbf{K}}{\partial \xi^{\top}} \mathbf{u} + \mathbf{K} \frac{\partial \mathbf{u}}{\partial \xi^{\top}} = \frac{\partial \mathbf{b}}{\partial \xi^{\top}}$$

とかく. すなわち,

$$\frac{\partial \mathbf{K}}{\partial \xi^{\top}} \mathbf{u} = \left(\frac{\partial \mathbf{K}}{\partial \xi_1} \mathbf{u} \quad \dots \quad \frac{\partial \mathbf{K}}{\partial \xi_{d-n}} \mathbf{u} \right) \in \mathbb{R}^{n \times (d-n)} \quad (2.6.21)$$

とおいた. そこで,

$$\frac{\partial \mathbf{u}}{\partial \xi^{\top}} = \mathbf{K}^{-1} \left(\frac{\partial \mathbf{b}}{\partial \xi^{\top}} - \frac{\partial \mathbf{K}}{\partial \xi^{\top}} \mathbf{u} \right) \quad (2.6.22)$$

の右辺を計算し、その結果を式 (2.6.20) に代入して \tilde{g} を求める方法を直接微分法とよぶことにする。ここで、 $\partial f/\partial \xi$, $\partial f/\partial u$, $\partial K/\partial \xi^\top$ および $\partial b/\partial \xi^\top$ は解析的に計算可能とみなしている。

それに対して、随伴変数法は次のように定義される。まず、 f に対する随伴問題を次のように定義する。

問題 2.6.11 (f に対する随伴問題) 問題 2.6.10 において、 $\xi \in \Xi$ に対して K および f は与えられたとする。このとき、

$$K^\top v = \frac{\partial f}{\partial u} \quad (2.6.23)$$

を満たす関数 $v \in U$ を求めよ。 \square

問題 2.6.11 の解 v を随伴変数とよぶ。ここで、式 (2.6.22) と式 (2.6.23) を組み合わせれば、

$$\left(\frac{\partial u}{\partial \xi^\top} \right)^\top \frac{\partial f}{\partial u} = \left(\frac{\partial b}{\partial \xi^\top} - \frac{\partial K}{\partial \xi^\top} u \right)^\top K^{-\top} K^\top v = \left(\frac{\partial b}{\partial \xi^\top} - \frac{\partial K}{\partial \xi^\top} u \right)^\top v \quad (2.6.24)$$

が得られる。式 (2.6.24) を式 (2.6.20) に代入すれば、

$$\tilde{g} = \frac{\partial f}{\partial \xi} + \left(\frac{\partial b}{\partial \xi^\top} - \frac{\partial K}{\partial \xi^\top} u \right)^\top v \in \mathbb{R}^{d-n} \quad (2.6.25)$$

が得られる。そこで、問題 2.6.11 を解いて v を求め、式 (2.6.25) で \tilde{g} を計算する方法を随伴変数法とよぶことにする。この定義では Lagrange 関数の定義は不要であった。

一方、Lagrange 乗数法を用いても随伴変数法と同じ結果が得られることをみておこう。まず、 $v \in \mathbb{R}^n$ を等式制約 (状態方程式) に対する Lagrange 乗数とにおいて、Lagrange 関数を

$$\mathcal{L}(\xi, u, v) = f(\xi, u) + v \cdot (b(\xi) - K(\xi)u)$$

と定義する。 $\mathcal{L}(\xi, u, v)$ の u および v に対する停留条件はそれぞれ

$$\frac{\partial \mathcal{L}}{\partial u}(\xi, u, v) = \frac{\partial f}{\partial u} - K^\top v = \mathbf{0}_{\mathbb{R}^n}, \quad \frac{\partial \mathcal{L}}{\partial v}(\xi, u, v) = b - Ku = \mathbf{0}_{\mathbb{R}^n}$$

となる。これらは式 (2.6.23) および式 (2.6.19) と一致する。さらに、 \mathcal{L} の ξ に対する偏微分は、

$$\frac{\partial \mathcal{L}}{\partial \xi}(\xi, u, v) = \frac{\partial f}{\partial \xi} + \left(\frac{\partial b}{\partial \xi^\top} - \frac{\partial K}{\partial \xi^\top} u \right)^\top v = \tilde{g}$$

となり，式 (2.6.25) と一致する．この一致から，Lagrange 乗数法と随伴変数法は同値となる．また，随伴変数は Lagrange 乗数と同値であることになる．

さらに， $\tilde{f}(\boldsymbol{\xi})$ の Hesse 行列 $\partial_{\Xi}\partial_{\Xi}^{\top}\tilde{f}$ を $\tilde{\mathbf{H}}$ とかくことにして，Lagrange 乗数法を用いて $\tilde{\mathbf{H}}$ を求める方法を考えてみよう．ここでは，Fréchet 微分の定義 (定義 4.5.4) に従い，任意の $\boldsymbol{\eta}_1 \in \Xi$ を用いた $\tilde{f}'(\boldsymbol{\xi})[\boldsymbol{\eta}_1] = \tilde{\mathbf{g}} \cdot \boldsymbol{\eta}_1$ に対する Lagrange 関数を

$$\begin{aligned} \mathcal{L}_1(\boldsymbol{\xi}, \mathbf{u}, \mathbf{v}, \mathbf{w}, \mathbf{z}) &= \tilde{\mathbf{g}}(\boldsymbol{\xi}, \mathbf{u}, \mathbf{v}) \cdot \boldsymbol{\eta}_1 + \mathbf{w} \cdot (\mathbf{b}(\boldsymbol{\xi}) - \mathbf{K}(\boldsymbol{\xi})\mathbf{u}) + \mathbf{z} \cdot \left(\frac{\partial f}{\partial \mathbf{u}} - \mathbf{K}^{\top}\mathbf{v} \right) \end{aligned} \quad (2.6.26)$$

とおく．ただし， $\tilde{\mathbf{g}}(\boldsymbol{\xi}, \mathbf{u}, \mathbf{v})$ は式 (2.6.25) で与えられるものとする． $\mathbf{w} \in U$ と $\mathbf{z} \in U$ は， $\tilde{\mathbf{g}}(\boldsymbol{\xi}, \mathbf{u}, \mathbf{v})$ における \mathbf{u} と \mathbf{v} が満たすべき等式制約のために用意された Lagrange 乗数である． $\boldsymbol{\eta}_1 \in \Xi$ は \mathcal{L}_1 においては定ベクトルとみなす．式 (2.6.26) の \mathcal{L}_1 は，第 1 章の平均コンプライアンスの場合，式 (1.1.43) の \mathcal{L}_{10} に対応する．第 1 章で行ったことを一般化すれば，次のようになる．

$(\boldsymbol{\xi}, \mathbf{u}, \mathbf{v}, \mathbf{w}, \mathbf{z})$ の任意変動 $(\boldsymbol{\eta}_2, \hat{\mathbf{u}}, \hat{\mathbf{v}}, \hat{\mathbf{w}}, \hat{\mathbf{z}}) \in \Xi \times U^4$ に対する \mathcal{L}_1 の微分は

$$\begin{aligned} \mathcal{L}'_1(\boldsymbol{\xi}, \mathbf{u}, \mathbf{v}, \mathbf{w}, \mathbf{z})[\boldsymbol{\eta}_2, \hat{\mathbf{u}}, \hat{\mathbf{v}}, \hat{\mathbf{w}}, \hat{\mathbf{z}}] &= \mathcal{L}'_{1\boldsymbol{\xi}}(\boldsymbol{\xi}, \mathbf{u}, \mathbf{v}, \mathbf{w}, \mathbf{z})[\boldsymbol{\eta}_2] + \mathcal{L}'_{1\mathbf{u}}(\boldsymbol{\xi}, \mathbf{u}, \mathbf{v}, \mathbf{w}, \mathbf{z})[\hat{\mathbf{u}}] \\ &\quad + \mathcal{L}'_{1\mathbf{v}}(\boldsymbol{\xi}, \mathbf{u}, \mathbf{v}, \mathbf{w}, \mathbf{z})[\hat{\mathbf{v}}] + \mathcal{L}'_{1\mathbf{w}}(\boldsymbol{\xi}, \mathbf{u}, \mathbf{v}, \mathbf{w}, \mathbf{z})[\hat{\mathbf{w}}] \\ &\quad + \mathcal{L}'_{1\mathbf{z}}(\boldsymbol{\xi}, \mathbf{u}, \mathbf{v}, \mathbf{w}, \mathbf{z})[\hat{\mathbf{z}}] \end{aligned} \quad (2.6.27)$$

となる．式 (2.6.27) の右辺第 4 項は， \mathbf{u} が状態決定問題の解ならばゼロとなる．式 (2.6.27) の右辺第 5 項は， \mathbf{v} が随伴問題の解ならばゼロとなる．また，式 (2.6.27) の右辺第 2 項は，

$$\begin{aligned} \mathcal{L}'_{1\mathbf{u}}(\boldsymbol{\xi}, \mathbf{u}, \mathbf{v}, \mathbf{w}, \mathbf{z})[\hat{\mathbf{u}}] &= \tilde{\mathbf{g}}_{\mathbf{u}^{\top}}(\boldsymbol{\xi}, \mathbf{u}, \mathbf{v})[\hat{\mathbf{u}}] \cdot \boldsymbol{\eta}_1 - \mathbf{w} \cdot (\mathbf{K}(\boldsymbol{\xi})\hat{\mathbf{u}}) \\ &= \hat{\mathbf{u}} \cdot \left(\tilde{\mathbf{g}}_{\mathbf{u}}^{\top}(\boldsymbol{\xi}, \mathbf{u}, \mathbf{v})[\boldsymbol{\eta}_1] - \mathbf{K}^{\top}(\boldsymbol{\xi})\mathbf{w} \right) \end{aligned} \quad (2.6.28)$$

となる．ここで，任意の $\hat{\mathbf{u}} \in U$ に対して式 (2.6.28) がゼロとなる条件は， \mathbf{w} を決定する随伴問題となる．式 (2.6.27) の右辺第 3 項は，

$$\begin{aligned} \mathcal{L}'_{1\mathbf{v}}(\boldsymbol{\xi}, \mathbf{u}, \mathbf{v}, \mathbf{w}, \mathbf{z})[\hat{\mathbf{v}}] &= \tilde{\mathbf{g}}_{\mathbf{v}^{\top}}(\boldsymbol{\xi}, \mathbf{u}, \mathbf{v})[\hat{\mathbf{v}}] \cdot \boldsymbol{\eta}_1 - \mathbf{z} \cdot (\mathbf{K}^{\top}(\boldsymbol{\xi})\hat{\mathbf{v}}) \\ &= \hat{\mathbf{v}} \cdot \left(\tilde{\mathbf{g}}_{\mathbf{v}}^{\top}(\boldsymbol{\xi}, \mathbf{u}, \mathbf{v})[\boldsymbol{\eta}_1] - \mathbf{K}(\boldsymbol{\xi})\mathbf{z} \right) \end{aligned} \quad (2.6.29)$$

となる. ここで, 任意の $\hat{v} \in U$ に対して式 (2.6.29) がゼロとなる条件は, z を決定する随伴問題となる.

最後に, 式 (2.6.27) の右辺第1項は,

$$\begin{aligned} & \mathcal{L}_{1\xi}(\xi, u, v, w, z) [\eta_2] \\ &= - \left\{ w^\top \left(\frac{\partial \mathbf{K}(\xi)}{\partial \xi_1} u \quad \dots \quad \frac{\partial \mathbf{K}(\xi)}{\partial \xi_{d-n}} u \right) \right. \\ & \quad \left. + z^\top \left(\frac{\partial \mathbf{K}^\top(\xi)}{\partial \xi_1} v \quad \dots \quad \frac{\partial \mathbf{K}^\top(\xi)}{\partial \xi_{d-n}} v \right) \right\} \eta_2 \end{aligned}$$

となる.

ここで, $u, v, w(\eta_1)$ と $z(\eta_1)$ はそれぞれ上記の条件を満たすように決定されたとする. このときの $f(\xi, u)$ を $\tilde{f}(\xi)$ とかくことにすれば,

$$\mathcal{L}_{1\xi}(\xi, u, v, w, z) [\eta_2] = \tilde{f}''(\xi) [\eta_1, \eta_2] = \tilde{g}_H(\xi, \eta_1) \cdot \eta_2 \quad (2.6.30)$$

のようにかける. ここで, \tilde{f} の Hesse 勾配 \tilde{g}_H は

$$\begin{aligned} \tilde{g}_H(\xi, \eta_1) = & - \left\{ w^\top(\eta_1) \left(\frac{\partial \mathbf{K}(\xi)}{\partial \xi_1} u \quad \dots \quad \frac{\partial \mathbf{K}(\xi)}{\partial \xi_{d-n}} u \right) \right. \\ & \left. + z^\top(\eta_1) \left(\frac{\partial \mathbf{K}^\top(\xi)}{\partial \xi_1} v \quad \dots \quad \frac{\partial \mathbf{K}^\top(\xi)}{\partial \xi_{d-n}} v \right) \right\}^\top \end{aligned} \quad (2.6.31)$$

のように与えられる.

2.6.6 等式制約つき最適化問題の解法に関する考察

この節で得られた結果とあとで示される結果をあわせて, 等式制約つき最適化問題 (問題 2.6.1) の解法について次のことがいえる.

- (1) 定理 2.6.4 より, Lagrange 乗数法 (問題 2.6.5) の解 (x, λ) は極小点の必要条件を満たす. そのような x は極小点の候補となる.
- (2) Lagrange 乗数法 (問題 2.6.5) の解 (x, λ) が得られたとき, Lagrange 関数の x の変動に対する Hesse 行列 $\partial_X \partial_X^\top \mathcal{L}(x, \lambda) = \mathbf{H}_{\mathcal{L}}(x, \lambda)$ が等式制約を満たす任意の変動 $y \in T_V(x)$ に対して $y \cdot (\mathbf{H}_{\mathcal{L}}(x, \lambda) y) > 0$ が確認されたならば, 定理 2.6.7 より, x は極小値と判定される.
- (3) 等式制約つき最適化問題 (問題 2.6.1) が凸最適化問題のとき, あとで示される定理 2.7.9 の系 (系 2.7.10) より, Lagrange 乗数法の解 x は最小点と判定される.

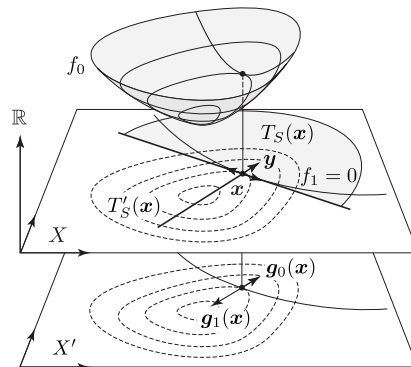


図 2.18: $X = \mathbb{R}^2$ および $m = 1$ のときの不等式制約つき最適化問題の極小点

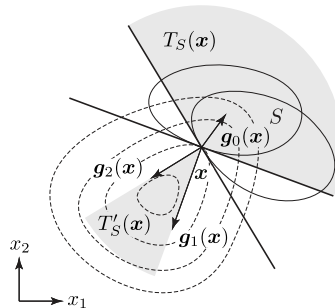


図 2.19: $X = \mathbb{R}^2$ および $m = 2$ のときの不等式制約つき最適化問題における極小点

- (4) 等式制約つき最適化問題 (問題 2.6.1) が凸最適化問題ではなくても, \tilde{f} が凸関数のときには, あとで示される定理 2.7.2 の系 (系 2.7.3) より, \tilde{f} の停留点 ($\tilde{g} = \mathbf{0}_{\mathbb{R}^{d-n}}$ が成り立つ \mathbf{x}) は, 最小点と判定される.

2.7 不等式制約つき最適化問題

制約条件を等式から不等式に変更してみよう. ここでは, 問題 2.1.2 あるいは問題 2.1.3 のような不等式制約のみが仮定されている場合について考えよう. 図 2.18 は, $X = \mathbb{R}^2$ および $m = 1$ の場合の極小点のようすを示す. $X = \mathbb{R}^2$ および $m = 2$ のときは 図 2.19 のようになる. まずは, これらの図を使って極小点ならば成り立つ条件についてみてから, 一般の場合を考えていくことにしよう.

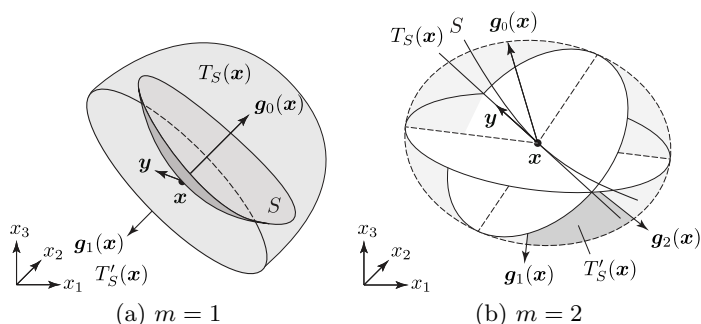


図 2.20: 不等式制約つき最適化問題における極小点 ($X = \mathbb{R}^3$ と $X' = \mathbb{R}^3$ を重ねている)

2.7.1 極小点の必要条件

最初に, 図 2.18 のように極小点 x では不等式制約の 1 つが有効である場合について考えてみよう. このとき, 次のことがいえる.

- (1) x からの変動が許されている方向は $\mathbf{g}_1 \cdot \mathbf{y} \leq 0$ を満たすすべての \mathbf{y} の方向である. このようなすべての方向は, 図中 $T_S(x)$ とかかれた半円の領域になる.
- (2) x が極小点であれば, すべての $\mathbf{y} \in T_S(x)$ 方向への変動に対して f_0 は変動しないかあるいは増加するはずである. この関係は, すべての $\mathbf{y} \in T_S(x)$ に対して $\mathbf{g}_0 \cdot \mathbf{y} \geq 0$ となることを意味する. いいかえれば, すべての $\mathbf{y} \in T_S(x)$ に対して $\mathbf{z} \cdot \mathbf{y} \leq 0$ が満たされるような \mathbf{z} の向きは, 図中 $T'_S(x)$ とかかれた半直線の領域に入る.
- (3) x が極小点であれば, すべての $\mathbf{y} \in T_S(x)$ に対して $\mathbf{g}_0 \cdot \mathbf{y} \geq 0$ となることは, $-\mathbf{g}_0$ が $T'_S(x)$ に含まれていることと同値である.

不等式制約の二つが有効になった場合は, 図 2.19 のようになる. x から変動が許される方向の集合 $T_S(x)$ は二つの不等式制約が同時に満たされる方向に制限されるために扇形となる. それに対して $T'_S(x)$ は, 不等式制約が 1 つの場合よりも広がって, 扇形になる. x が極小点であれば, すべての $\mathbf{y} \in T_S(x)$ に対して $\mathbf{g}_0 \cdot \mathbf{y} \geq 0$ となる関係や $-\mathbf{g}_0$ が $T'_S(x)$ に含まれる関係はここでも成り立つ. さらに, 図 2.20 は $X = \mathbb{R}^3$ のときの極小点のようすを示す.

以上のことを一般化しよう. 図 2.18 から 2.20 では, \mathbf{g}_1 や \mathbf{g}_2 を使って $T_S(x)$ と $T'_S(x)$ が定義された. ここでは, $C_S(x)$ を $T_S(x)$ を含むような許容方向集合として定義して, $C_S(x)$ と $C'_S(x)$ を使って同様の議論をする. なお, $T_S(x)$ と

$C_S(\mathbf{x})$ の関係は、あとで示される命題 2.7.4 の条件が満たされれば同一視することができる。

まず、不等式制約を満たす設計変数の許容集合 S を式 (2.1.1) のように定義する。また、 $\mathbf{x} \in S$ に対して

$$I_A(\mathbf{x}) = \{i \in \{1, \dots, m\} \mid f_i(\mathbf{x}) = 0\} = \{i_1, \dots, i_{|I_A(\mathbf{x})|}\} \quad (2.7.1)$$

を有効な制約に対する添え字の集合とする。さらに、方向を表すベクトル \mathbf{y} が定まるような $\mathbf{x} \in S$ に収束するすべての無限点列 $\{\mathbf{y}_k\}_{k \in \mathbb{N}} \in S$ に対して

$$C_S(\mathbf{x}) = \left\{ \mathbf{y} \in X \mid \frac{\mathbf{y}}{\|\mathbf{y}\|} = \lim_{k \rightarrow \infty} \frac{\mathbf{y}_k - \mathbf{x}}{\|\mathbf{y}_k - \mathbf{x}\|} \text{ for } \mathbf{y} \neq \mathbf{0}_X \right\}$$

を S の許容方向集合あるいは接錐という。 $C_S(\mathbf{x})$ に対して

$$C'_S(\mathbf{x}) = \{z \in X' \mid z \cdot \mathbf{y} \leq 0 \text{ for all } \mathbf{y} \in C_S(\mathbf{x})\}$$

を $C_S(\mathbf{x})$ の双対錐という。

f_0 が 1 階微分可能で、 C'_S が評価可能ならば、上で示された考察のように、次の結果は分かりやすい結果である。

定理 2.7.1 (極小点 1 次の必要条件) 問題 2.1.2 において、 $f_0 \in C^1(X; \mathbb{R})$ とする。 \mathbf{x} が極小点ならば、任意の $\mathbf{y} \in C_S(\mathbf{x})$ に対して、

$$\mathbf{g}_0(\mathbf{x}) \cdot \mathbf{y} \geq 0 \quad (2.7.2)$$

が成り立つ。また、

$$-\mathbf{g}_0(\mathbf{x}) \in C'_S(\mathbf{x}) \quad (2.7.3)$$

が成り立つ。 □

証明 任意の $\mathbf{y} \in C_S(\mathbf{x})$ に対して、 $\mathbf{g}_0(\mathbf{x}) \cdot \mathbf{y} \neq 0$ とすると、 $\mathbf{g}_0(\mathbf{x}) \cdot \mathbf{y} < 0$ となるような \mathbf{y} がとれる。このとき、定理 2.5.2 と同様の矛盾が得られる。式 (2.7.3) は式 (2.7.2) と同値である。 □

2.7.2 最小点の必要十分条件

問題 2.1.2 が凸最適化問題で、 C'_S が評価可能ならば、次のような最小点の必要十分条件が得られる。

定理 2.7.2 (最小点 1 次の必要十分条件) 問題 2.1.2 において, f_0 は $C^1(X; \mathbb{R})$ の要素で, f_1, \dots, f_m は $C^0(X; \mathbb{R})$ の要素で, かつ f_0, \dots, f_m は凸関数とする. S を式 (2.1.1) とおく. このとき, $\mathbf{x} \in S$ が最小点であるための必要十分条件は

$$-\mathbf{g}_0(\mathbf{x}) \in C'_S(\mathbf{x})$$

が成り立つことである. □

証明 必要条件は定理 2.7.1 より明らかである. 十分条件を示す. 任意の $\mathbf{y} \in S$ に対して, $\beta_k \in (0, 1)$ かつ $\beta_k \rightarrow 0$ なる $\{\beta_k\}_{k \in \mathbb{N}}$ を用いて $\{\mathbf{y}_k\}_{k \in \mathbb{N}}$ を $\mathbf{y}_k = (1 - \beta_k)\mathbf{x} + \beta_k\mathbf{y}$ とする. S は凸集合であるので, $\{\mathbf{y}_k\}_{k \in \mathbb{N}} \subseteq S$ となり, $C'_S(\mathbf{x})$ の定義より, $\mathbf{y} - \mathbf{x} \in C'_S(\mathbf{x})$ となる. そこで, $C'_S(\mathbf{x})$ の定義と $-\mathbf{g}_0(\mathbf{x}) \in C'_S(\mathbf{x})$ より,

$$-\mathbf{g}_0(\mathbf{x}) \cdot (\mathbf{y} - \mathbf{x}) \leq 0$$

が成り立つ. f_0 は凸関数なので, 定理 2.4.4 より,

$$\mathbf{g}_0(\mathbf{x}) \cdot (\mathbf{y} - \mathbf{x}) \leq f_0(\mathbf{y}) - f_0(\mathbf{x})$$

が成り立つ. よって, $f_0(\mathbf{x}) \leq f_0(\mathbf{y})$ が成り立つ. □

最小点が S の内点にあるならば, 定理 2.7.2 は次のようになる. この結果は定理 2.5.6 と同値となる.

系 2.7.3 (最小点 1 次の必要十分条件) 問題 2.1.2 において, f_0 は $C^1(X; \mathbb{R})$ の要素で, f_1, \dots, f_m は $C^0(X; \mathbb{R})$ で, かつ f_0, \dots, f_m は凸関数とする. S を式 (2.1.1) とおく. このとき, S の内点 \mathbf{x} が問題 2.1.3 の最小点であるための必要十分条件は

$$\mathbf{g}_0(\mathbf{x}) = \mathbf{0}_{\mathbb{R}^d}$$

が成り立つことである. □

証明 \mathbf{x} が S の内点であれば, $C'_S(\mathbf{x}) = X$ となる. よって, $C'_S(\mathbf{x}) = \{\mathbf{0}_{\mathbb{R}^d}\}$ となる. 定理 2.7.2 より, $\mathbf{g}_0(\mathbf{x}) = \mathbf{0}_{\mathbb{R}^d}$ が得られる. □

2.7.3 KKT 条件

これまでみてきた問題 2.1.2 の極小点あるいは最小点についての必要条件や十分条件には, すべての $\mathbf{y} \in C'_S(\mathbf{x})$ や C'_S に対する条件が含まれていた. しかし, これらの評価は容易ではない. そこで, 等式制約つき最適化問題でみてきたように, ここの Lagrange 関数を用いた表現を考えてみることにしよう.

まず, 不等式制約の 1 つが有効となっている図 2.18 の場合について考えてみよう. 不等式制約条件は有効であるから, 等式制約が課されていたと考えても \mathbf{x} は極小点となる. このとき, 等式制約のときに用いた式 (2.6.6) は

$$\mathbf{g}_0 + \lambda_1 \mathbf{g}_1 = \mathbf{0}_{\mathbb{R}^2} \tag{2.7.4}$$

にかきかえられる。しかし、不等式制約では有効な変動方向が広がる。そのことを詳しくみていこう。任意の $\mathbf{y} \in \mathbb{R}^2$ と式 (2.7.4) の内積をとれば、

$$\mathbf{g}_0 \cdot \mathbf{y} + \lambda_1 \mathbf{g}_1 \cdot \mathbf{y} = 0 \quad (2.7.5)$$

となる。式 (2.7.5) において、 \mathbf{y} が不等式制約を満たす方向であれば、 $\mathbf{g}_1 \cdot \mathbf{y} \leq 0$ となる。また、 \mathbf{x} が極小点であれば、そのような \mathbf{y} のときに目的関数が変動しないかあるいは増加することから、 $\mathbf{g}_0 \cdot \mathbf{y} \geq 0$ となる。この二つの条件を同時に満たすためには、

$$\lambda_1 \geq 0 \quad (2.7.6)$$

でなければならない。また、極小点では本来の不等式制約

$$f_1 \leq 0 \quad (2.7.7)$$

は満たされる。さらに、不等式制約が無効 ($f_1(\mathbf{x}) < 0$) の場合、不等式制約はなかった場合と等しくなることから、 $\lambda_1 = 0$ となる。一方、不等式制約が有効 ($f_1(\mathbf{x}) = 0$) の場合は、式 (2.7.6) が成り立つ。これら二つの関係は

$$\lambda_1 f_1 = 0 \quad (2.7.8)$$

ならば満たされる。そこで、式 (2.7.4)、式 (2.7.6)、式 (2.7.7) および式 (2.7.8) が1つの不等式制約が有効な場合の極小点で成り立つ条件となる。これらの条件は、あとで示される KKT 条件の $m = 1$ の場合になっている。

次に、図 2.19 のような極小点で二つの不等式制約が有効になっている場合について考えてみよう。ここでも、等式制約が課されていたとみなせば、

$$\mathbf{g}_0 + \lambda_1 \mathbf{g}_1 + \lambda_2 \mathbf{g}_2 = \mathbf{0}_{\mathbb{R}^2} \quad (2.7.9)$$

が成り立つようなある $\lambda_1, \lambda_2 \in \mathbb{R}$ が存在することと同値である。式 (2.7.9) を

$$-\mathbf{g}_0 = \lambda_1 \mathbf{g}_1 + \lambda_2 \mathbf{g}_2 \quad (2.7.10)$$

のようにかきかえよう。ここで、 \mathbf{g}_1 と \mathbf{g}_2 を固定して、

$$\lambda_1 \geq 0, \quad \lambda_2 \geq 0 \quad (2.7.11)$$

によってつくられる式 (2.7.10) 右辺のベクトルは、図 2.19 の $T'_S(\mathbf{x})$ とかかれた扇形になる。 $T'_S(\mathbf{x})$ の定義はのちに式 (2.7.15) で定義される。したがって、式 (2.7.10) は定理 2.7.1 の $C'_S(\mathbf{x})$ を $T'_S(\mathbf{x})$ にかきかえられた条件になっている。また、極小点であれば、本来の不等式制約

$$f_1 \leq 0, \quad f_2 \leq 0 \quad (2.7.12)$$

は満たされる．さらに，上で説明された理由により，極小点では

$$\lambda_1 f_1 = 0, \quad \lambda_2 f_2 = 0 \quad (2.7.13)$$

が成り立つ．そこで，式 (2.7.9), 式 (2.7.11), 式 (2.7.12) および式 (2.7.13) が二つの不等式制約が有効な場合の極小点で成り立つ条件となる．これらの条件は，あとで示される KKT 条件の $m = 2$ の場合になっている．

以上の結果を一般化しよう．まず，結果を示すために必要な定義と仮定を示そう． $\mathbf{x} \in S$ の近傍を $B_X \subset X$ とかく．ここで， $i \in I_A(\mathbf{x})$ に対して $f_i \in C^1(B_X; \mathbb{R})$ で，任意の $\mathbf{y} \in B_X$ に対して $\mathbf{g}_i(\mathbf{y})$ は1次独立であるとする．このとき，

$$T_S(\mathbf{x}) = \{\mathbf{y} \in X \mid \mathbf{g}_i(\mathbf{x}) \cdot \mathbf{y} \leq 0 \text{ for all } i \in I_A(\mathbf{x})\}$$

を \mathbf{x} における線形化許容方向集合という． $T_S(\mathbf{x})$ は，等式制約つき最適化問題における零空間に対応させて，

$$T_S(\mathbf{x}) = \text{Kco}(\mathbf{g}_{i_1}(\mathbf{x}), \dots, \mathbf{g}_{i_k}(\mathbf{x}))^\top \quad (2.7.14)$$

とかくことにする．また，

$$T'_S(\mathbf{x}) = \{\mathbf{z} \in X' \mid \mathbf{z} \cdot \mathbf{y} \leq 0 \text{ for all } \mathbf{y} \in T_S(\mathbf{x})\} \quad (2.7.15)$$

を $T_S(\mathbf{x})$ の双対錐という．

ここで， $T_S(\mathbf{x})$ と $C_S(\mathbf{x})$ の違いについて調べておこう． $T_S(\mathbf{x})$ は閉凸多面錐となるが， $C_S(\mathbf{x})$ は閉凸多面錐になるとは限らない [4]．たとえば，

$$S = \{\mathbf{y} \in \mathbb{R}^2 \mid f_1 = -y_1^2 + y_1^3 + y_2^2 \leq 0, f_2 = -y_1 - y_2 \leq 0\} \quad (2.7.16)$$

のとき，

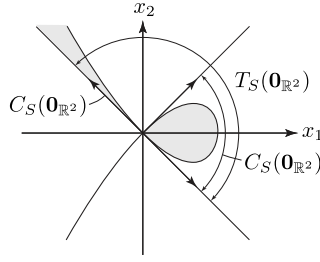
$$\begin{aligned} C_S(\mathbf{0}_{\mathbb{R}^2}) &= \{\mathbf{y} \in \mathbb{R}^2 \mid y_1 + y_2 \geq 0, y_1 - y_2 \geq 0\} \cup \{\mathbf{y} \in \mathbb{R}^2 \mid y_1 + y_2 = 0\}, \\ C'_S(\mathbf{0}_{\mathbb{R}^2}) &= \{\alpha(1, 1)^\top \in \mathbb{R}^2 \mid \alpha \leq 0\} \end{aligned}$$

となる．図 2.21 に $C_S(\mathbf{0}_{\mathbb{R}^2})$ を示す．図より， $C_S(\mathbf{0}_{\mathbb{R}^2})$ は閉凸多面錐ではないことがわかる．また，一般に，任意の $\mathbf{x} \in S$ に対して

$$C_S(\mathbf{x}) \subseteq T_S(\mathbf{x})$$

が成り立つ．たとえば，式 (2.7.16) の S のとき，

$$T_S(\mathbf{0}_{\mathbb{R}^2}) = \{\mathbf{y} \in \mathbb{R}^2 \mid y_1 + y_2 \geq 0\},$$

図 2.21: C_S が閉凸多面錐ではない例: $C_S(\mathbf{0}_{\mathbb{R}^2})$

$$T'_S(\mathbf{0}_{\mathbb{R}^2}) = \left\{ \alpha(-1, -1)^\top \in \mathbb{R}^2 \mid \alpha \geq 0 \right\}$$

となる.

$T_S(\mathbf{x}) = C_S(\mathbf{x})$ となるための十分条件は 1 次の制約想定とよばれる. 次に示される Cottle の制約想定はその一つである [5].

命題 2.7.4 (Cottle の制約想定) 問題 2.1.2 に対して, S を式 (2.1.1) とおく. $\mathbf{x} \in S$ に対して, $I_A(\mathbf{x})$ を式 (2.7.1) とおく. このとき, すべての $i \in I_A(\mathbf{x})$ に対して, $\mathbf{g}_i(\mathbf{x})$ が 1 次式のときは $\mathbf{g}_i(\mathbf{x}) \cdot \mathbf{y} \leq 0$ となり, 1 次式以外のときは $\mathbf{g}_i(\mathbf{x}) \cdot \mathbf{y} < 0$ となるような $\mathbf{y} \in X$ が存在するとき, $T_S(\mathbf{x}) = C_S(\mathbf{x})$ となる. \square

このような 1 次の制約想定を用いれば, 問題 2.1.2 の極小点ならば成り立つ条件を次のようにかくことができる.

定理 2.7.5 (KKT 条件) 問題 2.1.2 において, f_0, \dots, f_m は $C^1(X; \mathbb{R})$ の要素とする. $\mathbf{x} \in S$ において 1 次の制約想定を満たすとす. \mathbf{x} が極小点ならば,

$$\mathbf{g}_0(\mathbf{x}) + \sum_{i \in \{1, \dots, m\}} \lambda_i \mathbf{g}_i(\mathbf{x}) = \mathbf{0}_{\mathbb{R}^d}, \quad (2.7.17)$$

$$f_i(\mathbf{x}) \leq 0 \quad \text{for } i \in \{1, \dots, m\}, \quad (2.7.18)$$

$$\lambda_i f_i(\mathbf{x}) = 0 \quad \text{for } i \in \{1, \dots, m\}, \quad (2.7.19)$$

$$\lambda_i \geq 0 \quad \text{for } i \in \{1, \dots, m\} \quad (2.7.20)$$

を満たす $\boldsymbol{\lambda} = (\lambda_1, \dots, \lambda_m)^\top \in \mathbb{R}^m$ が存在する. \square

証明 問題 2.1.2 の不等式制約は, 任意の $\mathbf{t} = (t_1, \dots, t_m)^\top \in \mathbb{R}^m$ に対して

$$h_i(\mathbf{x}, t_i) = f_i(\mathbf{x}) + t_i^2 = 0 \quad \text{for } i \in \{1, \dots, m\} \quad (2.7.21)$$

のようにかかれる. そこで, \mathbf{t} を設計変数に加え, 問題 2.1.2 の Lagrange 関数を

$$\mathcal{L}(\mathbf{x}, \mathbf{t}, \boldsymbol{\lambda}) = f_0(\mathbf{x}) + \sum_{i \in \{1, \dots, m\}} \lambda_i h_i(\mathbf{x}, t_i) \quad (2.7.22)$$

とおく. ここで, (\mathbf{x}, \mathbf{t}) が式 (2.7.21) の等式制約を満たすときの f_0 の極小点ならば, 定理 2.6.4 より

$$\mathcal{L}_{\mathbf{x}}(\mathbf{x}, \mathbf{t}, \boldsymbol{\lambda}) = \mathbf{g}_0(\mathbf{x}) + \sum_{i \in \{1, \dots, m\}} \lambda_i \mathbf{g}_i(\mathbf{x}) = \mathbf{0}_{\mathbb{R}^d}, \quad (2.7.23)$$

$$\mathcal{L}_{t_i}(\mathbf{x}, \mathbf{t}, \boldsymbol{\lambda}) = 2\lambda_i t_i = 0 \quad \text{for } i \in \{1, \dots, m\}, \quad (2.7.24)$$

$$\mathcal{L}_{\lambda_i}(\mathbf{x}, \mathbf{t}, \boldsymbol{\lambda}) = f_i(\mathbf{x}) + t_i^2 = 0 \quad \text{for } i \in \{1, \dots, m\} \quad (2.7.25)$$

が成り立つ. 式 (2.7.23) と式 (2.7.25) はそれぞれ式 (2.7.17) と式 (2.7.18) と同値である. また, 式 (2.7.24) の両辺に t_i をかけて, 式 (2.7.21) を用いれば, 式 (2.7.19) が得られる.

さらに, 式 (2.7.20) が成り立つことは次のようにして確かめられる. \mathbf{x} が問題 2.1.2 の極小点で, 1 次の制約想定が満たされていれば, 定理 2.7.1 と式 (2.7.14) および Farkas の補題 (補題 A.3.2) より, $i_1, \dots, i_k \in I_A(\mathbf{x})$ に対して

$$\begin{aligned} -\mathbf{g}_0(\mathbf{x}) \in C'_S(\mathbf{x}) = T'_S(\mathbf{x}) = (T_S(\mathbf{x}))' &= \left(\text{Kco}(\mathbf{g}_{i_1}(\mathbf{x}), \dots, \mathbf{g}_{i_k}(\mathbf{x}))^\top \right)' \\ &= \text{Ico}(\mathbf{g}_{i_1}, \dots, \mathbf{g}_{i_k}) \end{aligned}$$

が成り立つ. ただし, $\boldsymbol{\lambda} = (\lambda_{i_1}, \dots, \lambda_{i_k})^\top$ に対して

$$\text{Ico}(\mathbf{g}_{i_1}, \dots, \mathbf{g}_{i_k}) = \{ (\mathbf{g}_{i_1}, \dots, \mathbf{g}_{i_k}) \boldsymbol{\lambda} \in X' \mid \boldsymbol{\lambda} \geq \mathbf{0}_{\mathbb{R}^k} \}$$

とおいた. また, $i \notin I_A(\mathbf{x})$ に対して $\lambda_i = 0$ とおく. この関係は, 式 (2.7.20) が成り立つことを示している. \square

式 (2.7.17) から式 (2.7.20) を KKT (Karush-Kuhn-Tucker) 条件という. 式 (2.7.18) は \mathbf{x} が不等式制約を満たしていることを示している. 式 (2.7.19) は, 相補性条件とよばれ, 有効でない制約 $f_i(\mathbf{x}) < 0$ に対して $\lambda_i = 0$ となることで式 (2.7.17) から \mathbf{g}_i を削除する作用をする. 最後に式 (2.7.17) と式 (2.7.20) をあわせた条件は, 図 2.19 を用いた考察で解説されたように, 図 2.19 の $T'_S(\mathbf{x})$ とかかれた扇形に $-\mathbf{g}_0$ が入る条件を与えており, 定理 2.7.1 の $C'_S(\mathbf{x})$ を $T'_S(\mathbf{x})$ にかきかえた条件になっている. 証明の中で使われた \mathbf{t} はスラック変数とよばれる.

不等式制約つき最適化問題に対しても Lagrange 関数を定義しよう. ここでは, あとで示される双対定理との関係を考えて

$$\mathcal{L}(\mathbf{x}, \boldsymbol{\lambda}) = \begin{cases} f_0(\mathbf{x}) + \sum_{i \in \{1, \dots, m\}} \lambda_i f_i(\mathbf{x}) & (\boldsymbol{\lambda} \geq \mathbf{0}_{\mathbb{R}^m}) \\ -\infty & (\boldsymbol{\lambda} \not\geq \mathbf{0}_{\mathbb{R}^m}) \end{cases} \quad (2.7.26)$$

とおく. ここで, $\boldsymbol{\lambda} = (\lambda_1, \dots, \lambda_m)^\top \in \mathbb{R}^m$ は Lagrange 乗数である. このとき, 式 (2.7.17) と式 (2.7.18) は $\mathcal{L}(\mathbf{x}, \boldsymbol{\lambda})$ を用いた条件にかきかえられる. そこで, 次の問題の解を問題 2.1.2 の解の候補とする方法を不等式制約つき最適化問題に対する Lagrange 乗数法という.

問題 2.7.6 (不等式制約つき最適化問題に対する Lagrange 乗数法) 問題 2.1.2 に対して, $\mathcal{L}(\mathbf{x}, \boldsymbol{\lambda})$ を式 (2.7.26) とおく. このとき, KKT 条件

$$\mathcal{L}_{\mathbf{x}}(\mathbf{x}, \boldsymbol{\lambda}) = \mathbf{g}_0(\mathbf{x}) + \sum_{i \in \{1, \dots, m\}} \lambda_i \mathbf{g}_i(\mathbf{x}) = \mathbf{0}_{\mathbb{R}^d}, \quad (2.7.27)$$

$$\mathcal{L}_{\lambda_i}(\mathbf{x}, \boldsymbol{\lambda}) = f_i(\mathbf{x}) \leq 0 \quad \text{for } i \in \{1, \dots, m\}, \quad (2.7.28)$$

$$\lambda_i f_i(\mathbf{x}) = 0 \quad \text{for } i \in \{1, \dots, m\}, \quad (2.7.29)$$

$$\lambda_i \geq 0 \quad \text{for } i \in \{1, \dots, m\} \quad (2.7.30)$$

を満たす $(\mathbf{x}, \boldsymbol{\lambda})$ を求めよ. \square

さらに, f_0, \dots, f_m が 2 階微分可能のとき, 問題 2.1.2 の極小点ならば, 次の条件が得られる. KKT 条件を満たす $(\mathbf{x}, \boldsymbol{\lambda}) \in X \times \mathbb{R}^m$ に対して, 式 (2.7.1) と式 (2.1.1) の $I_A(\mathbf{x})$ と S のほかに

$$\bar{I}_A(\boldsymbol{\lambda}) = \{i \in \{1, \dots, m\} \mid \lambda_i > 0\}, \quad (2.7.31)$$

$$\bar{S} = S \cap \{\mathbf{x} \in X \mid f_i(\mathbf{x}) = 0, i \in \bar{I}_A(\boldsymbol{\lambda})\} \quad (2.7.32)$$

を定義する. \bar{S} の \mathbf{x} における許容方向集合を $T_{\bar{S}}(\mathbf{x})$ とかく. また, Lagrange 関数 $\mathcal{L}(\mathbf{x}, \boldsymbol{\lambda})$ の \mathbf{x} の変動に対する Hesse 行列を $\mathbf{H}_{\mathcal{L}}(\mathbf{x}, \boldsymbol{\lambda}) = \partial_X \partial_X^\top \mathcal{L}(\mathbf{x}, \boldsymbol{\lambda})$ とかく.

定理 2.7.7 (Lagrange 関数を用いた極小点 2 次の必要条件) 問題 2.1.2 において, f_0, \dots, f_m は $C^2(X; \mathbb{R})$ の要素とする. $\mathbf{x} \in X$ において $i \in I_A(\mathbf{x})$ に対する \mathbf{g}_i が 1 次独立で, かつ 1 次の制約想定を満たすとす. このとき, \mathbf{x} が問題 2.1.2 の極小点ならば, 任意の接ベクトル $\mathbf{t} \in T_{\bar{S}}(\mathbf{x})$ に対して,

$$\mathbf{t} \cdot (\mathbf{H}_{\mathcal{L}}(\mathbf{x}, \boldsymbol{\lambda}) \mathbf{t}) \geq 0$$

が成り立つ. \square

証明 $\{\mathbf{y}_k\}_{k \in \mathbb{N}} \in \bar{S}$ を $i \in \{1, \dots, m\}$ に対して $\lambda_i f_i(\mathbf{y}_k) = 0$ を満たし, $\mathbf{t} / \|\mathbf{t}\|_{\mathbb{R}^d} = \lim_{k \rightarrow \infty} (\mathbf{y}_k - \mathbf{x}) / \|\mathbf{y}_k - \mathbf{x}\|_{\mathbb{R}^d}$ となるようにとる. \mathbf{x} は極小点であることから, \mathbf{x} の近傍 B をとり, 任意の $\mathbf{y}_k \in B$ に対して, $\theta \in (0, 1)$ が存在して

$$\begin{aligned} f_0(\mathbf{y}_k) - f_0(\mathbf{x}) &= \mathcal{L}(\mathbf{y}_k, \boldsymbol{\lambda}) - \mathcal{L}(\mathbf{x}, \boldsymbol{\lambda}) \\ &= \frac{1}{2} (\mathbf{y}_k - \mathbf{x}) \cdot \{\mathbf{H}_{\mathcal{L}}(\mathbf{x} + \theta(\mathbf{y}_k - \mathbf{x}), \boldsymbol{\lambda})(\mathbf{y}_k - \mathbf{x})\} \geq 0 \end{aligned} \quad (2.7.33)$$

が成り立つ. 両辺に $2\mathbf{t} / \|\mathbf{y}_k - \mathbf{x}\|_{\mathbb{R}^d}^2$ をかけて, $k \rightarrow \infty$ とすれば, 定理の式が得られる. \square

2.7.4 極小点の十分条件

問題 2.1.2 の極小点であるための十分条件について、次の条件が得られる。 $T_{\bar{S}}(\mathbf{x})$ は式 (2.7.32) で定義された \bar{S} の \mathbf{x} における線形化許容方向集合とする。

定理 2.7.8 (極小点 2 次の十分条件) 問題 2.1.2 において、 f_0, \dots, f_m は $C^2(X; \mathbb{R})$ の要素とする。 $\mathbf{x} \in X$ において KKT 条件を満たし、任意の $\mathbf{y} \in T_{\bar{S}}(\mathbf{x})$ に対して

$$\mathbf{y} \cdot (\mathbf{H}_{\mathcal{L}}(\mathbf{x}, \boldsymbol{\lambda}) \mathbf{y}) > 0$$

を満たす $(\mathbf{x}, \boldsymbol{\lambda})$ が存在するならば、 \mathbf{x} は極小点である。□

証明 $(\mathbf{x}, \boldsymbol{\lambda})$ が KKT 条件を満たすならば、式 (2.7.33) が得られる。ここで、 $\mathbf{H}_{\mathcal{L}}(\mathbf{x}, \boldsymbol{\lambda})$ は正定値である。したがって、 \mathbf{x} は極小点となる。□

2.7.5 KKT 条件を用いた最小点の十分条件

さらに、問題 2.1.2 が凸最適化問題ならば、最小点の十分条件が得られる。

定理 2.7.9 (KKT 条件を用いた最小点 1 次の十分条件) 問題 2.1.2 において、 f_0, \dots, f_m は $C^1(X; \mathbb{R})$ の要素で、かつ凸関数とする。 $\mathbf{x} \in X$ において $i \in I_A(\mathbf{x})$ に対する \mathbf{g}_i が 1 次独立で、かつ 1 次の制約想定を満たし、さらに $(\mathbf{x}, \boldsymbol{\lambda})$ は KKT 条件を満たすとする。このとき、 \mathbf{x} は最小点である。□

証明 KKT 条件を満たす $\lambda_1, \dots, \lambda_m$ を固定して、

$$\mathcal{L}(\mathbf{y}) = f_0(\mathbf{y}) + \sum_{i \in \{1, \dots, m\}} \lambda_i f_i(\mathbf{y})$$

とおく。KKT 条件より、 $\partial_X \mathcal{L}(\mathbf{x}) = \mathbf{0}_{\mathbb{R}^d}$ が成り立つ。 \mathcal{L} は凸関数なので、系 2.7.3 より、 \mathbf{x} は \mathcal{L} の最小点となる。すなわち、任意の $\mathbf{y} \in S$ に対して

$$f_0(\mathbf{x}) + \sum_{i \in \{1, \dots, m\}} \lambda_i f_i(\mathbf{x}) \leq f_0(\mathbf{y}) + \sum_{i \in \{1, \dots, m\}} \lambda_i f_i(\mathbf{y})$$

が成り立つ。KKT 条件より、 $i \in \{1, \dots, m\}$ に対して $\lambda_i f_i(\mathbf{x}) = 0$ および $\lambda_i \geq 0$ なので、任意の $\mathbf{y} \in S$ に対して

$$f_0(\mathbf{x}) \leq f_0(\mathbf{y})$$

が成り立つ。□

さらに、定理 2.7.9 を応用すれば、等式制約つき最適化問題 (問題 2.6.1) に対して、次のような最小点の十分条件が得られる。

系 2.7.10 (等式制約に対する最小点 1 次の十分条件) f_0 は $C^1(X; \mathbb{R})$ の要素で、かつ凸関数とする。 $\mathbf{h} \in C^1(X; \mathbb{R}^n)$ は 1 次関数で、かつ $\mathbf{x} \in X$ において $\partial_X h_1(\mathbf{x}), \dots, \partial_X h_n(\mathbf{x})$ が 1 次独立で、1 次の制約想定を満たすとする。 $(\mathbf{x}, \boldsymbol{\lambda})$ は問題 2.6.5 の解とする。このとき、 \mathbf{x} は問題 2.6.1 の最小点である。 \square

証明 等式制約は

$$\mathbf{h}(\mathbf{x}) \leq \mathbf{0}_{\mathbb{R}^n}, \quad -\mathbf{h}(\mathbf{x}) \leq \mathbf{0}_{\mathbb{R}^n}$$

と同値である。これらに対する Lagrange 乗数を $\boldsymbol{\lambda}_+ = (\lambda_{+1}, \dots, \lambda_{+n})^\top \in \mathbb{R}^n$, $\boldsymbol{\lambda}_- = (\lambda_{-1}, \dots, \lambda_{-n})^\top \in \mathbb{R}^n$ とおく。 h_1, \dots, h_n は 1 次関数なので凸関数である。なぜならば、任意の $\mathbf{x} \in X$ および $\mathbf{y} \in X$ に対して

$$\partial_X h_i(\mathbf{x}) \cdot (\mathbf{y} - \mathbf{x}) = h_i(\mathbf{y}) - h_i(\mathbf{x})$$

が成り立つ (定理 2.4.4) ためである。したがって、この問題は不等式制約の凸最適化問題となっている。このときの KKT 条件は、

$$\begin{aligned} \mathcal{L}_{\mathbf{x}}(\mathbf{x}, \boldsymbol{\lambda}_+, \boldsymbol{\lambda}_-) &= \mathbf{g}_0(\mathbf{x}) + \partial_X \mathbf{h}^\top(\mathbf{x})(\boldsymbol{\lambda}_+ - \boldsymbol{\lambda}_-) = \mathbf{0}_{\mathbb{R}^d}, \\ \mathcal{L}_{\lambda_{+i}}(\mathbf{x}, \boldsymbol{\lambda}_+, \boldsymbol{\lambda}_-) &= h_i(\mathbf{x}) \leq 0 \quad \text{for } i \in \{1, \dots, n\}, \\ \mathcal{L}_{\lambda_{-i}}(\mathbf{x}, \boldsymbol{\lambda}_+, \boldsymbol{\lambda}_-) &= -h_i(\mathbf{x}) \leq 0 \quad \text{for } i \in \{1, \dots, n\}, \\ \lambda_{+i} h_i(\mathbf{x}) &= 0 \quad \text{for } i \in \{1, \dots, n\}, \quad \lambda_{-i} h_i(\mathbf{x}) = 0 \quad \text{for } i \in \{1, \dots, n\}, \\ \lambda_{+i} &\geq 0 \quad \text{for } i \in \{1, \dots, n\}, \quad \lambda_{-i} \geq 0 \quad \text{for } i \in \{1, \dots, n\} \end{aligned}$$

とかけると、 $\boldsymbol{\lambda}_+ - \boldsymbol{\lambda}_- = \boldsymbol{\lambda}$ とかければ、 $(\mathbf{x}, \boldsymbol{\lambda})$ は問題 2.6.5 の解と同値になる。 \square

2.7.6 不等式制約つき最適化問題の例題

KKT 条件を、ばねの結合問題に対して使ってみよう。

例題 2.7.11 (ばねの結合問題) 例題 2.6.8 において、ばねの結合条件が不等式に変更されたときの最小点を求めよ。すなわち、

$$\min_{\mathbf{u} \in \mathbb{R}^2} \left\{ f_0(\mathbf{u}) = \frac{1}{2}k_1 u_1^2 + \frac{1}{2}k_2 u_2^2 \mid f_1(\mathbf{u}) = a - (u_1 + u_2) \leq 0 \right\}$$

を満たす \mathbf{u} を求めよ。また、

$$\min_{\mathbf{u} \in \mathbb{R}^2} \left\{ f_0(\mathbf{u}) = \frac{1}{2}k_1 u_1^2 + \frac{1}{2}k_2 u_2^2 \mid f_1(\mathbf{u}) = (u_1 + u_2) - a \leq 0 \right\}$$

を満たす \mathbf{u} を求めよ。 \square

解答 $f_1 = a - (u_1 + u_2) \leq 0$ のとき、Lagrange 関数は、 $\lambda \in \mathbb{R}$ を Lagrange 乗数として、

$$\mathcal{L}(\mathbf{u}, \lambda) = \frac{1}{2}k_1 u_1^2 + \frac{1}{2}k_2 u_2^2 + \lambda(a - u_1 - u_2)$$

となる. $\mathcal{L}(\mathbf{u}, \lambda)$ の停留条件は例題 2.6.8 の結果と等しく, $k_1 > 0, k_2 > 0$ および $a > 0$ のとき

$$u_1 = \frac{k_2}{k_1 + k_2}a > 0, \quad u_2 = \frac{k_1}{k_1 + k_2}a > 0, \quad \lambda = \frac{k_1 k_2}{k_1 + k_2}a > 0$$

となる. この結果は KKT 条件を満たす. この問題は, 例題 2.6.8 で調べたように, 凸最適化問題なので, 定理 2.7.9 により \mathbf{u} は最小点となる.

一方, $f_1 = (u_1 + u_2) - a \leq 0$ のとき, Lagrange 関数は

$$\mathcal{L}(\mathbf{u}, \lambda) = \frac{1}{2}k_1 u_1^2 + \frac{1}{2}k_2 u_2^2 + \lambda(u_1 + u_2 - a)$$

となる. $\mathcal{L}(\mathbf{u}, \lambda)$ の停留条件は, $k_1 > 0, k_2 > 0$ および $a > 0$ のとき

$$u_1 = \frac{k_2}{k_1 + k_2}a > 0, \quad u_2 = \frac{k_1}{k_1 + k_2}a > 0, \quad \lambda = -\frac{k_1 k_2}{k_1 + k_2}a < 0$$

となる. $\lambda < 0$ より, この結果は KKT 条件を満たさない. したがって, 結合制約を無効とみなし, $\lambda = 0$ とおき, 問題を

$$\min_{\mathbf{u} \in \mathbb{R}^2} \left\{ f_0(\mathbf{u}) = \frac{1}{2}k_1 u_1^2 + \frac{1}{2}k_2 u_2^2 \right\}$$

にかきかえる. この問題に対して

$$\mathbf{g}_0(\mathbf{u}) = \begin{pmatrix} k_1 & 0 \\ 0 & k_2 \end{pmatrix} \begin{pmatrix} u_1 \\ u_2 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$$

より, $\mathbf{u} = \mathbf{0}_{\mathbb{R}^2}$ が得られる. □

2.7.7 不等式制約つき最適化問題の解法に関する考察

この節の内容から, 不等式制約つき最適化問題 (問題 2.1.2) の解法について次のことがいえる.

- (1) 定理 2.7.5 より, Lagrange 乗数法 (問題 2.7.6) の解 $(\mathbf{x}, \boldsymbol{\lambda})$ は極小点の必要条件を満たす. そのような \mathbf{x} は極小点の候補となる.
- (2) Lagrange 乗数法 (問題 2.7.6) の解 $(\mathbf{x}, \boldsymbol{\lambda})$ が得られたとき, Lagrange 関数の \mathbf{x} の変動に対する Hesse 行列 $\partial_{\mathbf{x}} \partial_{\mathbf{x}}^{\top} \mathcal{L}(\mathbf{x}, \boldsymbol{\lambda}) = \mathbf{H}_{\mathcal{L}}(\mathbf{x}, \boldsymbol{\lambda})$ が不等式制約を等式で満たす任意の変動 $\mathbf{y} \in T_{\bar{S}}(\mathbf{x})$ に対して $\mathbf{y} \cdot (\mathbf{H}_{\mathcal{L}}(\mathbf{x}, \boldsymbol{\lambda}) \mathbf{y}) > 0$ が確認されたならば, 定理 2.7.8 より, \mathbf{x} は極小点と判定される.
- (3) 不等式制約つき最適化問題 (問題 2.1.2) が凸最適化問題のとき, 定理 2.7.9 より, Lagrange 乗数法の解 \mathbf{x} は最小点と判定される.

2.8 等式と不等式制約つき最適化問題

最適設計問題では、等式制約として状態方程式が設定され、不等式制約として評価関数についての制約が使われる。2.1節の冒頭で定義された最適化問題 (問題 2.1.1) は、そのようなことを意識して定義された。ここでは、最適設計問題との対応を意識して、問題 2.1.1 を次のようにかくことにしよう。

問題 2.8.1 (等式と不等式制約つき最適化問題) $d > n$ として、 $\Xi = \mathbb{R}^{d-n}$ および $U = \mathbb{R}^n$ とおく。 $\mathbf{K} : \Xi \rightarrow \mathbb{R}^{n \times n}$ と $\mathbf{b} : \Xi \rightarrow \mathbb{R}^n$ が与えられ、 $f_0, f_1, \dots, f_m : \Xi \times U \rightarrow \mathbb{R}$ が与えられたとき、

$$\min_{(\boldsymbol{\xi}, \mathbf{u}) \in \Xi \times U} \{ f_0(\boldsymbol{\xi}, \mathbf{u}) \mid \mathbf{h}(\boldsymbol{\xi}, \mathbf{u}) = -\mathbf{K}(\boldsymbol{\xi})\mathbf{u} + \mathbf{b}(\boldsymbol{\xi}) = \mathbf{0}_{\mathbb{R}^n}, \\ f_1(\boldsymbol{\xi}, \mathbf{u}) \leq 0, \dots, f_m(\boldsymbol{\xi}, \mathbf{u}) \leq 0 \}$$

を満たす $(\boldsymbol{\xi}, \mathbf{u})$ を求めよ。 □

問題 2.8.1 に対して、等式制約を満たす $(\boldsymbol{\xi}, \mathbf{u})$ の集合を

$$V = \{(\boldsymbol{\xi}, \mathbf{u}) \in \Xi \times U \mid \mathbf{h}(\boldsymbol{\xi}, \mathbf{u}) = -\mathbf{K}(\boldsymbol{\xi})\mathbf{u} + \mathbf{b}(\boldsymbol{\xi}) = \mathbf{0}_{\mathbb{R}^n}\}$$

とおく。さらに、 $i \in \{0, 1, \dots, m\}$ に対して

$$\tilde{f}_i(\boldsymbol{\xi}) = \{f_i(\boldsymbol{\xi}, \mathbf{u}) \mid (\boldsymbol{\xi}, \mathbf{u}) \in V\} \quad (2.8.1)$$

とかくことにする。このとき、 $\tilde{f}_i(\boldsymbol{\xi})$ の $\boldsymbol{\xi}$ に対する微分 $\tilde{\mathbf{g}}_i$ は、2.6.5 項でみてきたように、式 (2.6.25) と同様の

$$\tilde{\mathbf{g}}_i = \frac{\partial f_i}{\partial \boldsymbol{\xi}} + \left(\frac{\partial \mathbf{b}}{\partial \boldsymbol{\xi}^\top} - \frac{\partial \mathbf{K}}{\partial \boldsymbol{\xi}^\top} \mathbf{u} \right)^\top \mathbf{v}_i \in \mathbb{R}^{d-n} \quad (2.8.2)$$

によって求められる。ただし、 $\left(\partial \mathbf{K} / \partial \boldsymbol{\xi}^\top \right) \mathbf{u}$ は式 (2.6.21) で定義されているものとする。ここで、 $\mathbf{v}_i \in U$ は、式 (2.6.23) と同様の随伴方程式

$$\mathbf{K}^\top \mathbf{v}_i = \frac{\partial f_i}{\partial \mathbf{u}} \quad (2.8.3)$$

の解とする。

このようにして得られた $\tilde{\mathbf{g}}_0, \dots, \tilde{\mathbf{g}}_m$ は評価関数が $\tilde{f}_0(\boldsymbol{\xi}), \dots, \tilde{f}_m(\boldsymbol{\xi})$ とみなされたときの $\boldsymbol{\xi}$ に対する微分になっている。そこで、これらを用いれば、問題 2.8.1 は次のようにかきかえられる。

問題 2.8.2 (不等式制約つき最適化問題) $\Xi = \mathbb{R}^{d-n}$ とおく. $\tilde{f}_0, \tilde{f}_1, \dots, \tilde{f}_m : \Xi \rightarrow \mathbb{R}$ が与えられたとき,

$$\min_{\xi \in \Xi} \left\{ \tilde{f}_0(\xi) \mid \tilde{f}_1(\xi) \leq 0, \dots, \tilde{f}_m(\xi) \leq 0 \right\}$$

を満たす ξ を求めよ. □

問題 2.8.2 に対する Lagrange 関数を

$$\tilde{\mathcal{L}}(\xi, \lambda) = \tilde{f}_0(\xi) + \sum_{i \in \{1, \dots, m\}} \lambda_i \tilde{f}_i(\xi) \quad (2.8.4)$$

とおく. このとき, 問題 2.8.2 に対する KKT 条件は

$$\tilde{\mathcal{L}}_{\xi}(\xi, \lambda) = \tilde{g}_0 + \sum_{i \in \{1, \dots, m\}} \lambda_i \tilde{g}_i = \mathbf{0}_{\mathbb{R}^{d-n}}, \quad (2.8.5)$$

$$\tilde{\mathcal{L}}_{\lambda_i}(\xi, \lambda) = \tilde{f}_i(\xi) \leq 0 \quad \text{for } i \in \{1, \dots, m\}, \quad (2.8.6)$$

$$\lambda_i \tilde{f}_i(\xi) = 0 \quad \text{for } i \in \{1, \dots, m\}, \quad (2.8.7)$$

$$\lambda_i \geq 0 \quad \text{for } i \in \{1, \dots, m\} \quad (2.8.8)$$

となる. そこで, 問題 2.8.2 に対する極小点の候補を求めるための Lagrange 乗数法は次のようになる.

問題 2.8.3 (不等式制約つき最適化問題に対する Lagrange 乗数法) 問題 2.8.2 に対して, $\tilde{\mathcal{L}}(\xi, \lambda)$ を式 (2.8.4) とおく. このとき, KKT 条件 (式 (2.8.5) から式 (2.8.8)) を満たす (ξ, λ) を求めよ. □

そこで, 問題 2.8.3 の解 (ξ, λ) に対して, 定理 2.7.5 から定理 2.7.8 の結果が得られることになる. これらの結果から等式と不等式制約つき最適化問題 (問題 2.8.1) の解法に関していえることは, 2.8.2 項にまわすことにしよう.

2.8.1 等式と不等式制約つき最適化問題に対する Lagrange 乗数法

問題 2.8.1 の最小点満たす条件については上記のとおりである. ここでは, 問題 2.8.1 に対する Lagrange 関数を定義して, それらを用いて問題 2.8.3 の Lagrange 乗数法につなげることを考えよう. ここで示される内容は, 第1章でも示された内容である. 第7章以降の最適設計問題においても評価関数の設計変数に対する微分を求める際に使われる考え方である. 第2章の内容とそれらとの関係を明確にするために, あえてこの項を設けることにした.

問題 2.8.1 に対する Lagrange 関数を

$$\mathcal{L}(\boldsymbol{\xi}, \mathbf{u}, \mathbf{v}_0, \dots, \mathbf{v}_m, \boldsymbol{\lambda}) = \mathcal{L}_0(\boldsymbol{\xi}, \mathbf{u}, \mathbf{v}_0) + \sum_{i \in \{1, \dots, m\}} \lambda_i \mathcal{L}_i(\boldsymbol{\xi}, \mathbf{u}, \mathbf{v}_i) \quad (2.8.9)$$

とおく。ただし、 $\boldsymbol{\lambda} = (\lambda_1, \dots, \lambda_m)^\top \in \mathbb{R}^m$ を $f_1 \leq 0, \dots, f_m \leq 0$ に対する Lagrange 乗数とする。また、

$$\mathcal{L}_i(\boldsymbol{\xi}, \mathbf{u}, \mathbf{v}_i) = f_i(\boldsymbol{\xi}, \mathbf{u}) + \mathcal{L}_S(\boldsymbol{\xi}, \mathbf{u}, \mathbf{v}_i) \quad (2.8.10)$$

を $f_i(\boldsymbol{\xi}, \mathbf{u})$ に対する Lagrange 関数とする。さらに、

$$\mathcal{L}_S(\boldsymbol{\xi}, \mathbf{u}, \mathbf{v}_i) = -\mathbf{v}_i \cdot (\mathbf{K}(\boldsymbol{\xi})\mathbf{u} - \mathbf{b}(\boldsymbol{\xi})) \quad (2.8.11)$$

を等式制約に対する Lagrange 関数、 $\mathbf{v}_0, \dots, \mathbf{v}_m$ を $\mathbf{h}(\boldsymbol{\xi}, \mathbf{u}) = \mathbf{0}_{\mathbb{R}^n}$ に対する Lagrange 乗数 (随伴変数) であり、それぞれ f_0, \dots, f_m のために定義されたものであるとする。

式 (2.8.2) の $\tilde{\mathbf{g}}_i$ は次のようにして得られる。 $(\boldsymbol{\xi}, \mathbf{u}, \mathbf{v}_i)$ の任意の変動 $(\boldsymbol{\eta}, \hat{\mathbf{u}}, \hat{\mathbf{v}}_i) \in \Xi \times U \times U$ に対する \mathcal{L}_i の微分 (全微分) は

$$\begin{aligned} \mathcal{L}'_i(\boldsymbol{\xi}, \mathbf{u}, \mathbf{v}_i)[\boldsymbol{\eta}, \hat{\mathbf{u}}, \hat{\mathbf{v}}_i] \\ = \mathcal{L}_{i\xi}(\boldsymbol{\xi}, \mathbf{u}, \mathbf{v}_i)[\boldsymbol{\eta}] + \mathcal{L}_{iu}(\boldsymbol{\xi}, \mathbf{u}, \mathbf{v}_i)[\hat{\mathbf{u}}] + \mathcal{L}_{iv_i}(\boldsymbol{\xi}, \mathbf{u}, \mathbf{v}_i)[\hat{\mathbf{v}}_i] \end{aligned} \quad (2.8.12)$$

となる。式 (2.8.12) の右辺第 3 項は、

$$\mathcal{L}_{iv_i}(\boldsymbol{\xi}, \mathbf{u}, \mathbf{v}_i)[\hat{\mathbf{v}}_i] = -\hat{\mathbf{v}}_i \cdot (\mathbf{K}(\boldsymbol{\xi})\mathbf{u} - \mathbf{b}(\boldsymbol{\xi})) = \mathcal{L}_S(\boldsymbol{\xi}, \mathbf{u}, \hat{\mathbf{v}}_i) \quad (2.8.13)$$

となる。式 (2.8.13) は \mathbf{u} が等式制約を満たすときにゼロとなる。式 (2.8.12) の右辺第 2 項は、

$$\mathcal{L}_{iu}(\boldsymbol{\xi}, \mathbf{u}, \mathbf{v}_i)[\hat{\mathbf{u}}] = -\hat{\mathbf{u}} \cdot \left(\mathbf{K}^\top(\boldsymbol{\xi})\mathbf{v}_i - \frac{\partial f_i}{\partial \mathbf{u}} \right) \quad (2.8.14)$$

となる。式 (2.8.14) は \mathbf{v}_i が式 (2.8.3) を満たすときにゼロとなる。さらに、式 (2.8.12) の右辺第 1 項は、

$$\mathcal{L}_{i\xi}(\boldsymbol{\xi}, \mathbf{u}, \mathbf{v}_i)[\boldsymbol{\eta}] = \left\{ \frac{\partial f_i}{\partial \boldsymbol{\xi}} + \left(\frac{\partial \mathbf{b}}{\partial \boldsymbol{\xi}^\top} - \frac{\partial \mathbf{K}}{\partial \boldsymbol{\xi}^\top} \mathbf{u} \right)^\top \mathbf{v}_i \right\} \cdot \boldsymbol{\eta} \quad (2.8.15)$$

となる。

以上の結果から、 \mathbf{u} を求める等式制約は任意の $\hat{\mathbf{v}}_i \in U$ に対して $\mathcal{L}_{iv_i}(\boldsymbol{\xi}, \mathbf{u}, \mathbf{v}_i)[\hat{\mathbf{v}}_i] = 0$ が成り立つ条件と同値であり、 \mathbf{v}_i を求める随伴方程式は任意の $\hat{\mathbf{u}} \in U$ に対して $\mathcal{L}_{iu}(\boldsymbol{\xi}, \mathbf{u}, \mathbf{v}_i)[\hat{\mathbf{u}}] = 0$ が成り立つ条件と同値であることがわかる。さらに、 \tilde{f}_i の微分 $\tilde{f}'_i(\boldsymbol{\xi})[\boldsymbol{\eta}]$ は、それらの \mathbf{u} と \mathbf{v}_i を用いたときの $\mathcal{L}_{i\xi}(\boldsymbol{\xi}, \mathbf{u}, \mathbf{v}_i)[\boldsymbol{\eta}] = \tilde{\mathbf{g}}_i \cdot \boldsymbol{\eta}$ によってえられることが示されたことになる。

2.8.2 等式と不等式制約つき最適化問題の解法に関する考察

これまでの結果より, 等式と不等式制約つき最適化問題 (問題 2.8.1) の解法について次のことがいえる.

- (1) $\tilde{f}_0, \dots, \tilde{f}_m$ を式 (2.8.1) のようにおく. このとき, 等式と不等式制約つき最適化問題 (問題 2.8.1) は不等式制約つき最適化問題 (問題 2.8.2) にかきかえられる.
- (2) $i \in \{0, 1, \dots, m\}$ に対して \tilde{f}_i の微分 $\tilde{f}'_i(\boldsymbol{\xi})[\boldsymbol{\eta}]$ は, 問題 2.8.1 の Lagrange 関数を式 (2.8.10) の \mathcal{L}_i とおいたとき, 等式制約を満たす \mathbf{u} と随伴方程式 (あるいは任意の $\hat{\mathbf{u}} \in U$ に対して $\mathcal{L}_{i\mathbf{u}}(\boldsymbol{\xi}, \mathbf{u}, \mathbf{v}_i)[\hat{\mathbf{u}}] = 0$) を満たす \mathbf{v}_i を使って, $\mathcal{L}_{i\boldsymbol{\xi}}(\boldsymbol{\xi}, \mathbf{u}, \mathbf{v}_i)[\boldsymbol{\eta}] = \tilde{\mathbf{g}}_i \cdot \boldsymbol{\eta}$ によってえられる.
- (3) 不等式制約つき最適化問題 (問題 2.8.2) に対する Lagrange 乗数法 (問題 2.8.3) の解 $(\boldsymbol{\xi}, \boldsymbol{\lambda})$ が得られたとする (具体的な方法は第 3 章で示される). このとき, Lagrange 関数 $\tilde{\mathcal{L}}(\boldsymbol{\xi}, \boldsymbol{\lambda})$ の $\boldsymbol{\xi}$ の変動に対する Hesse 行列 $\partial_{\Xi} \partial_{\Xi}^T \tilde{\mathcal{L}}(\boldsymbol{\xi}, \boldsymbol{\lambda}) = \mathbf{H}_{\tilde{\mathcal{L}}}(\boldsymbol{\xi}, \boldsymbol{\lambda})$ が不等式制約を等式で満たす任意の変動

$$\boldsymbol{\eta} \in T_{\bar{S}}(\boldsymbol{\xi}) = \{\boldsymbol{\eta} \in \Xi \mid \tilde{\mathbf{g}}_i(\boldsymbol{\xi}) \cdot \boldsymbol{\eta} = 0 \text{ for all } i \in I_A(\boldsymbol{\xi})\}$$

に対して

$$\boldsymbol{\eta} \cdot (\mathbf{H}_{\tilde{\mathcal{L}}}(\boldsymbol{\xi}, \boldsymbol{\lambda}) \boldsymbol{\eta}) > 0$$

が確認されたならば, 定理 2.7.8 より, $\boldsymbol{\xi}$ は極小値と判定される.

- (4) 不等式制約つき最適化問題 (問題 2.8.2) が凸最適化問題のとき, 定理 2.7.9 より, Lagrange 乗数法の解 $\boldsymbol{\xi}$ とそのときの $\mathbf{h}(\boldsymbol{\xi}, \mathbf{u}) = \mathbf{0}_{\mathbb{R}^n}$ を満たす \mathbf{u} は最小点と判定される.

第 1 章の 1 次元線形弾性体と 1 次元定常 Stokes 流れ場の形状最適化問題は, 等式と不等式制約つき最適化問題の例題となっていた. それらの問題は凸最適化問題であった. そこで, KKT 条件を満たす \mathbf{a} がみつければ, それらは最小点であると判定される.

2.9 双対定理

2.7 節および 2.8 節で用いた KKT 条件は, f_0, \dots, f_m が 1 階微分可能であることを必要とした. 次に示される **双対定理** を用いれば, この 1 階微分可能性を凸性に

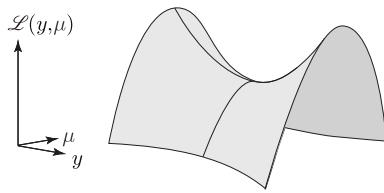


図 2.22: 鞍点

変えることができる．本書ではこの定理を直接使わないため，証明を省略して結果だけを示すことにする．

次のような制約想定を定義しよう．

定義 2.9.1 (Slater 制約想定) 問題 2.1.2 において， $f(y) < \mathbf{0}_{\mathbb{R}^m}$ が成り立つようなある $y \in S$ が存在するとき，Slater 制約想定が満たされているという． \square

双対定理は次のように表される [4-6]．

定理 2.9.2 (双対定理) 問題 2.1.2 が凸最適化問題で，かつ Slater 制約想定が満たされるとする． $\mathcal{L}(x, \lambda)$ を式 (2.7.26) とする．このとき， $x \in X$ が最小点であるための必要十分条件は，任意の $y \in X$ および $\mu \geq \mathbf{0}_{\mathbb{R}^m}$ に対して

$$\mathcal{L}(x, \mu) \leq \mathcal{L}(x, \lambda) \leq \mathcal{L}(y, \lambda) \quad (2.9.1)$$

が成り立つような $\lambda \geq \mathbf{0}_{\mathbb{R}^m}$ が存在することである． \square

式 (2.9.1) を満たす (x, λ) は図 2.22 のような鞍点となる．そのために，双対定理は鞍点定理ともよばれる．

2.9.1 双対定理の例題

ここでも，ばねの結合問題に対して双対定理を使ってみよう．

例題 2.9.3 (ばねの結合問題) 例題 2.7.11 において，

$$\min_{\mathbf{u} \in \mathbb{R}^2} \left\{ f_0(\mathbf{u}) = \frac{1}{2}k_1u_1^2 + \frac{1}{2}k_2u_2^2 \mid f_1(\mathbf{u}) = a - (u_1 + u_2) \leq 0 \right\}$$

を満たす \mathbf{u} は Lagrange 関数の鞍点になることを示せ． \square

解答 \mathbf{u} をこの問題の最小点とみなし, $\mathbf{v} \in \mathbb{R}^2$ を任意の変数とおく. このとき, $f_0(\mathbf{v})$ と $f_1(\mathbf{v})$ は凸関数なので, この問題は凸最適化問題である. また, $(v_1, v_2) = (a/4, a/4)$ のとき $f_1(\mathbf{v}) < 0$ となることから, Slater 制約想定は満たされている. この問題の Lagrange 関数は

$$\mathcal{L}(\mathbf{v}, \mu) = f_0(\mathbf{v}) + \mu f_1(\mathbf{v}) = \frac{1}{2}k_1 v_1^2 + \frac{1}{2}k_2 v_2^2 + \mu(a - v_1 - v_2)$$

で定義される. ここで, $\mu \in \mathbb{R}$ は $f_1 \leq 0$ に対する Lagrange 乗数である. このとき, $\mu > 0$ に対して

$$\begin{aligned} \mathcal{L}(\mathbf{u}, \mu) &= \inf_{\mathbf{v} \in \mathbb{R}^2} \mathcal{L}(\mathbf{v}, \mu) = \mathcal{L}\left(\frac{\mu}{k_1}, \frac{\mu}{k_2}, \mu\right) \\ &= \frac{1}{2} \frac{\mu^2}{k_1} + \frac{1}{2} \frac{\mu^2}{k_2} + \mu \left(a - \frac{\mu}{k_1} - \frac{\mu}{k_2}\right) = -\frac{1}{2} \left(\frac{1}{k_1} + \frac{1}{k_2}\right) \mu^2 + a\mu \end{aligned}$$

となる. さらに, $\tilde{\mathcal{L}}(\mu) = \mathcal{L}(\mathbf{u}(\mu), \mu)$ とおき,

$$\frac{d\tilde{\mathcal{L}}}{d\mu} = -\left(\frac{1}{k_1} + \frac{1}{k_2}\right)\mu + a = 0$$

を満たす μ を λ とおく. また,

$$\frac{d^2\tilde{\mathcal{L}}}{d\mu^2} = -\left(\frac{1}{k_1} + \frac{1}{k_2}\right) < 0$$

となる. したがって,

$$\mathcal{L}(\mathbf{u}, \mu) \leq \mathcal{L}(\mathbf{u}, \lambda)$$

が成り立つ. 一方, $\mathcal{L}(\mathbf{u}, \lambda) \leq \mathcal{L}(\mathbf{v}, \lambda)$ は明らかである. □

例題 2.9.3 の最小点が Lagrange 関数の鞍点になっていることを図で確認してみよう. この問題の変数は $(u_1, u_2, \lambda)^\top \in X = \mathbb{R}^3$ であった. このままでは図にすることは難しい. そこで, $u_2 = 0$ とおく. あるいは $k_2 \rightarrow \infty$ を仮定する. また, $k_1 = 1$ および $a = 1$ とおく. このとき,

$$\mathcal{L}(u_1, \lambda) = \frac{1}{2}u_1^2 + \lambda(1 - u_1)$$

となり, 鞍点は $(u_1, \lambda) = (1, 1)$ となる. 図 2.23 にそのようすを示す. 鞍点の条件が成り立っていることが確かめられる.

なお, λ は内力の意味をもつ. また, $-\mathcal{L}(\mathbf{u}, \lambda)$ はコンプリメンタリエネルギーとよばれる. 工学では, 変位を既知として内力を求めるときに, コンプリメンタリエネルギーの最小化問題が使われる.

この章の最後に, 第1章でみてきたような最適設計問題に対して双対定理を使ってみよう. 1.1 節や 2.2 節で解説されたように, 本書で考えるような最適設計問題は凸最適化問題になっていた. それならば双対定理が適用されて, 最適設計問題の

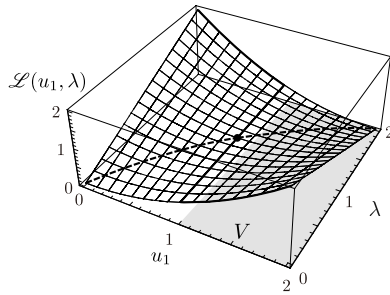


図 2.23: 例題 2.9.3 の鞍点

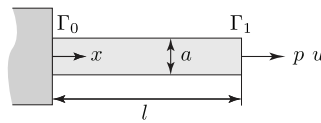


図 2.24: 1 個の断面積をもつ 1 次元線形弾性体

最小点では設計変数と Lagrange 乗数に対して Lagrange 関数は鞍点を形成しているはずである。そのことを図で確かめてみよう。

図にするために、変数を二つに限定することにしよう。1つを Lagrange 乗数にとれば、設計変数を 1つに限定する必要がある。そこで、図 2.24 のような 1つだけの断面積をもつ 1 次元線形弾性体を考えることにしよう。

例題 2.9.4 (平均コンプライアンス最小化問題) $e_Y = 1, l = 1, c_1 = 1$ および $p = 1$ とする。このとき、

$$\min_{(a,u) \in \mathbb{R}^2} \left\{ f_0(u) = pu \mid f_1(a) = la - c_1 \leq 0, \frac{e_Y}{l} au = p \right\}$$

を満たす (a, u) を求めよ。また、この点における Lagrange 関数を図示せよ。 □

解答 $\tilde{f}_0(a) = f_0(u(a)) = f_0(1/a) = 1/a$ となる。 $\tilde{f}_0(a)$ と $f_1(a)$ は凸関数なので、

$$\min_{a \in \mathbb{R}} \left\{ \tilde{f}_0(a) \mid f_1(a) \leq 0 \right\}$$

は凸最適化問題である。Slater 制約想定が満たされることは自明である。この問題の Lagrange 関数は、 $\lambda \in \mathbb{R}$ を $f_1(a) \leq 0$ に対する Lagrange 乗数とにおいて

$$\mathcal{L}(a, \lambda) = \tilde{f}_0(a) + \lambda f_1(a) = \frac{1}{a} + \lambda(a - 1)$$

で与えられる。このとき、

$$\mathcal{L}_a = -\frac{1}{a^2} + \lambda = 0, \quad \mathcal{L}_\lambda = a - 1 = 0$$

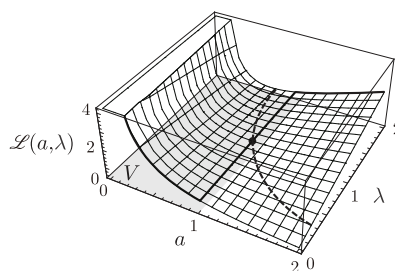


図 2.25: 例題 2.9.4 の鞍点

より, $(a, \lambda) = (1, 1)$ が $\mathcal{L}(a, \lambda)$ の停留点となる. 図 2.25 は, この点の近傍における \mathcal{L} を示す. 停留点の近傍が鞍点になっていることが確かめられる. \square

2.10 第2章のまとめ

第2章では, 有限次元ベクトル空間上の最適化問題に関する理論をみてきた. 要点は以下のようである.

- (1) 最適化問題は, 一般に, 設計変数の集合の中から評価関数が最小となる要素を求める問題として定義される. 一方, 最適設計問題では, 第1章でみてきたように, 設計変数のほかに状態変数が定義され, 評価関数は設計変数と状態変数の関数として定義された. そのとき, 状態変数は状態方程式によって一意に決定された. そこで, 最適設計問題を一般的な最適化問題の枠組みに当てはめるためには, 状態変数を設計変数に含めて, 状態方程式を等式制約とみなせばよい (2.1 節).
- (2) 制約なし最適化問題の極小点では目的関数の勾配が 0 になる (定理 2.5.2). また, 目的関数の停留点 (勾配が 0) で, Hesse 行列が正定値ならば, 極小点である (定理 2.5.5). さらに, 目的関数が凸関数ならば, 極小点は最小点である (定理 2.5.6 あるいは系 2.7.3).
- (3) 等式制約つき最適化問題の極小点では Lagrange 関数が停留する (定理 2.6.4). また, Lagrange 関数の停留点で, 等式制約が満たされるような変数の変動に対して Lagrange 関数の Hesse 行列が正定値ならば, その停留点は極小点である (定理 2.6.7). さらに, 凸最適化問題ならば Lagrange 関数の停留点は最小点である (系 2.7.10).

- (4) 不等式制約つき最適化問題の極小点では KKT 条件が成り立つ (定理 2.7.5). また, KKT 条件を満たす点で, 不等式制約を満たす変数の任意変動に対して Lagrange 関数の Hesse 行列が正定値ならば, その点は極小点である (定理 2.7.8). さらに, 凸最適化問題では, KKT 条件を満たす点は最小点である (定理 2.7.9).
- (5) 等式と不等式制約つき最適化問題の極小点では, 等式制約が満たされるもとのみに変動できる変数が変動したときの評価関数の微分を用いた KKT 条件が成り立つ (式 (2.8.5) から式 (2.8.8)).
- (6) 不等式制約つき最適化問題が凸最適化問題のとき, 最小点は Lagrange 関数の鞍点となる (定理 2.9.2).

最適化理論に関する文献もたくさん存在する. 本章で引用された文献のほかにくつかをあげると [2, 3, 7, 8, 11, 12] である.

2.11 第2章の演習問題

2.1 定義 2.4.5 において, \mathbf{A} が正定値ならば α は \mathbf{A} の固有値の中の最小値と一致することを示せ. また, \mathbf{A} が負定値ならば $-\alpha$ は \mathbf{A} の固有値の中の最大値と一致することを示せ. (ヒント: 定理 A.2.1 を参考にせよ.)

2.2 関数 $f: \mathbb{R}^2 \mapsto \mathbb{R}$ を

$$f(x_1, x_2) = \frac{1}{2}(ax_1^2 + 2bx_1x_2 + cx_2^2) + dx_1 + ex_2$$

とする. ただし, $a, b, c, d \in \mathbb{R}$ は定数とする. f が最小値をとるための必要条件を示せ. また, 十分条件が $a > 0$ および $ac - b^2 > 0$ であることを示せ. ただし, Sylvester 判定法 (定理 A.2.2) を用いてよい.

2.3 周の長さが一定値以下である長方形のうちで, 面積が最大となるものは正方形であることを, 次の順に示せ.

- 長方形の隣り合う2辺の長さを $\mathbf{x} = (x_1, x_2)^\top \in \mathbb{R}^2$ とおき, 周の長さを制限する正の実定数を c_1 とおいて問題を構成せよ.
- Lagrange 関数を定義して, KKT 条件を求めよ.
- KKT 条件を満たす解が最小点であることを示せ. (ヒント: この問題は凸最適化問題であるかを考え, そうでない場合には, 目的関数が制約を満たす変数の集合に対する関数 (本文では \tilde{f}_0 とかいた) で再構成された問題が凸最適化問題になっていることを示せばよい.)

参考文献

- [1] Bartle, R. G. *The Elements of Real Analysis, 2nd Ed.* Wiley, New York, 1976.
- [2] Bazaraa, M. S. and Shetty, C. M. *Nonlinear Programming: Theory and Algorithms.* Wiley, New York, 1979.
- [3] Chong, E. K. P. and Żak, S. H. *An Introduction to Optimization.* Wiley, New York, 2008.
- [4] 藤田宏, 今野浩, 田邊國士. 最適化法. 岩波書店, 東京, 1994.
- [5] 福島雅夫. 非線形最適化の基礎. 朝倉書店, 東京, 2001.
- [6] Güler, O. *Foundations of Optimization.* Springer, New York, 2010.
- [7] Jeter, M. W. *Mathematical Programming: An Introduction to Optimization.* M. Dekker, Inc., New York, 1986.
- [8] 久志本茂. 最適化問題の基礎. 森北出版, 東京, 1979.
- [9] Rudin, W. *Principles of Mathematical Analysis.* McGraw-Hill, New York, 1976.
- [10] 高木貞治. 解析概論, 改訂第3版. 岩波書店, 東京, 1961.
- [11] 田村明久, 村松正和. 最適化法. 共立出版, 東京, 2002.
- [12] 矢部博. 工学基礎: 最適化とその応用. 数理工学社, 東京, 2006.