微分不可能な関数に対する数理計画法 (1)

福 島 雅 夫

1. まえがき

これまで数理計画法，とくに非線形計画法において目的関数あるいは制約関数の微分可能性の仮定は本質的な役割を演じてきた．実際，ラグランジュの未定乗数法や Kuhn-Tucker の定理など最適性条件に関する諸定理 [18] や，ニュートン法，共役勾配法，最大勾配法および勾配射影法，一般縮小勾配法などの代表的な最小化手法 [17] は，ほとんどすべてが関数の微分可能性にもとづいている．

しかし，つぎの章にみるように，応用上よくあらわれるいくつかの問題は，なめらかでない，すなわちすべての点では微分できない関数を用いて定式化されることが知られている．そのような実際上の必要性と，さらにそれ自身の理論的興味と相まって，近年，微分不可能な最適化問題を取扱った論文や報告が多く見られるようになっている．そこで，まず現在までの研究の流れを簡単にふりかえっておこう．

初期の頃（1960年代）の傾向は，微分可能性の仮定をとりはずすかわりに関数の凸性を仮定することであった．つまり凸関数は任意の点で準勾配（subgradient）をもつという事実（3.2節参照）に着目することにより，微分可能な場合において得られている多くの結果が拡張された．その成果は1970年に Rockafellar [28] によって集大成され，その後もこの方向に沿った研究が進められている．さらに，凸関数の準勾配の概念を非凸関数に対して拡張した一般勾配（generalized gradient）なるものが1975年に Clarke [3] によって定義された．それ以前にも Pshenichnyi [27] によって類似した概念が考えられているが Clarke のほうがより一般的であって，それ以後，かなり広いクラスの問題に対しても古典的な定理の多くが一般化できることがわかってきた．

他方，理論面での発展と平行して，準勾配を利用した Polyak [25, 26] Bertsekas と Mitter [2], Wolfe [32] などの凸関数の最小化法や，それらを非凸関数にも適用できるように改良した Mifflin [20], Feuer [8], Goldstein [11] などの方法が考案され，実際問題に対する数値解法の面でもすぐれた研究がなされつつある．さらに，特殊な構造をもつある種の問題に対しては，その特性を考慮した解法も開発されている．しかし，残念ながら，さきに述べた理論面での研究に比べ，応用面はまだ一歩遅れている感がぬぐえないように思える．

そのような現状をふまえたうえで，本稿では微分不可能な最適化問題に対して現在までに得られている主要な成果の概説を試みることにする．まず次章では，実際上よくあらわれる微分不可能な最適化問題のいくつかの例を示す．第3章では凸関数の準勾配および非凸関数の一般勾配の定義とその諸性質を述べる．第4章では代表的な最適化アルゴリズムを紹介する．なお，問題の本質に対する理解を深めるために第2章および第3章の記述をややくわしくした．そのため，第4章ではいくつかの基本的な方法を解説する，いわば少数重点主義をとらざるを得なかったことをご了承いただきたい．

2. 各種の微分不可能な最適化問題

この章では，微分不可能関数を含む問題として定式化される最適化問題の例を紹介しよう．

2.1. 大規模数理計画問題

つぎの資源配分問題（Resource Allocation Problem）を考えよう．ある製品を生産するシステムがあり，それは N 個の部分システムからなっており，生産に要する諸資源の量はシステム全体で $b=(b_1, \cdots, b_M)$ だけ所有しているとする．ただし，b_j $(j=1, \cdots, M)$ は第 j 番目の資源の所有量をあらわす．さらに，第 n 部分システムの生産水準をベクトル x_n であらわし，生産費用関数を $f_n(x_n)$ で，所要資源量をベクトル値関数 $g_n(x_n)$ で，システムの技術条件などによって定まる x_n の許容集合を

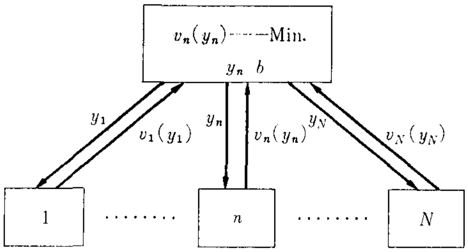

図 2.1 2-レベル・システム

X_n であらわすことにする．そのとき，与えられた資源量のもとで，システムの総生産費用を最小にする問題は，つぎのように書ける．

$$\min \sum_{n=1}^{N} f_n(x_n) \tag{2.1}$$

$$\text{s.t.} \quad \sum_{n=1}^{N} g_n(x_n) \leqq b, \quad x_n \in X_n (n=1, \cdots, N).$$

さて，このシステムを図 2.1 のように 2-レベル・システムと考え，上位レベルから各サブシステムへ資源量 y_n を配分するものとすると，各サブシステムは与えられた資源量のもとで生産費用を最小にしようと努めるであろう．その最小費用を y_n の関数として $v_n(y_n)$ とあらわすと，$v_n(y_n)$ は次式で定義される．

$$v_n(y_n) = \min\{f_n(x_n) \ ; \ g_n(x_n) \leqq y_n, \ x_n \in X_n\} \tag{2.2}$$

そのとき，問題(2.1)はつぎの問題と等価になる．

$$\min \sum_{n=1}^{N} v_n(y_n) \tag{2.3}$$

$$\text{s.t.} \quad \sum_{n=1}^{N} y_n \leqq b.$$

(2.1)で，f_n，g_n が凸関数で，X_n が凸集合なら，$v_n(y_n)$ は y_n に関して凸関数であり，さらに f_n，g_n が一次関数で X_n が凸多面体のときは，v_n は区分的一次関数になることが知られている．しかし，そのようなもっとも簡単な場合でも，$v_n(y_n)$ は y_n に関して微分可能であるとは限らないし，さらに，v_n の関数形を陽に得ることも一般には困難である．

それにもかかわらず (2.3) の形式化がしばしば採られるのは，個々の問題(2.2)は原問題(2.1)に比べて小規模であり，システム全体を詳細に記述することが非常に困難な場合でも，部分システムにおいてはモデルを比較的容易に記述でき，使用する計算機の容量や計算速度の面からもモデルの解析が実行可能になることによっている．

このような定式化による問題(2.1)へのアプローチは，たとえば Geoffrion [10]，Silverman [29]，Grinold [12]，Marsten et al. [19] らによって試みられている．

2.2. 割当て問題

N 個の仕事と N 人の従業員がいるとき，従業員 j が仕事 i をするときの費用が c_{ij} で与えられているものとする．そのとき，各従業員に一つの仕事を割当て，しかも各仕事には従業員のだれか 1 人が割当てられているようなすべての割当て方のなかで，総割当て費用が最小になるようなものを見出す問題を割当て問題 (Assignment Problem)という．この問題は，以下のように定式化できる．

$$\min \sum_{i=1}^{N} \sum_{j=1}^{N} c_{ij} x_{ij} \tag{2.4}$$

$$\begin{aligned} \text{s.t.} \quad & \sum_{j=1}^{N} x_{ij} = 1 \quad (i=1, \cdots, N) \\ & \sum_{i=1}^{N} x_{ij} = 1 \quad (j=1, \cdots, N) \\ & x_{ij} = 0 \text{ or } 1 \quad (i, j=1, \cdots, N) \end{aligned}$$

ここで，変数 x_{ij} はつぎのような意味をもっている．

$$x_{ij} = \begin{cases} 1 & \text{従業員 } j \text{ が仕事 } i \text{ をするとき,} \\ 0 & \text{その他のとき.} \end{cases}$$

問題(2.4)は整数計画問題であるが，制約条件から $x_{ij}=0$ または 1，という条件を取り除いたＬＰ問題の最適解は自動的に整数解になることが知られているから，問題(2.4)は条件，$x_{ij}=0$ または 1，を除いたＬＰ問題と等価になる [30]．よって，その双対ＬＰ問題

$$\max \sum_{i=1}^{N} v_i + \sum_{j=1}^{N} w_j \tag{2.5}$$

$$\text{s.t.} \quad v_i + w_j \leqq c_{ij} \quad (i, j=1, \cdots, N)$$

を解くことにより，割当て問題 (2.4) の最適解を得ることができる．

問題(2.5)から，変数 w_j を消去して，つぎの等価な問題を得る．

$$\max \sum_{i=1}^{N} v_i + \sum_{j=1}^{N} \min_{s} \, [c_{sj} - v_s] \tag{2.6}$$

問題(2.6)の目的関数は区分的に線形な凹関数である．

2.3. 設備配置問題

ある地域に N 個の倉庫を新しく設置することを考える．ところが，すでに M 個の倉庫が設けられており，新しい倉庫は全体的にみて，できるだけ均等に配置することが望ましい．具体的には，新しく設ける倉庫の位置を 2 次元ベクトル $x_i (i=1, \cdots, N)$ であらわし，すでに設けられている倉庫の位置を 2 次元ベクトル $a_j (j=1, \cdots, N)$ であらわすとき，つぎの評価関数を最小にするような x_n を見つける問題を一般多設備配置問題 (generalized multi-facility location problem)という．(図 2.2 参照)

$$\min \sum_{i=1}^{N} \sum_{j=1}^{M} w_{ij} \|x_i - a_j\| + \sum_{i=1}^{N} \sum_{k=i}^{N} v_{ik} \|x_i - x_k\| \tag{2.7}$$

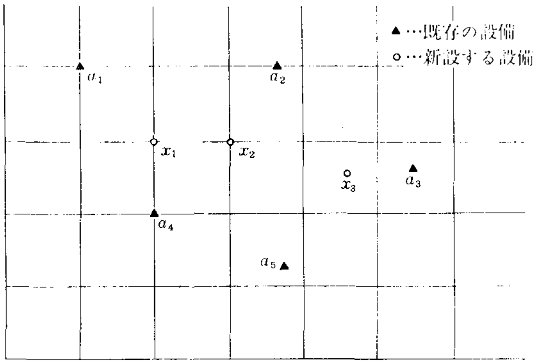

図 2.2 平面上の設備配置問題

ただし，w_{ij} および v_{ik} は，与えられた重み定数である．$\|\cdot\|$ は2点間の距離に相当する量をあらわすものでありノルムとよばれている．通常よく用いられるのはつぎの2種のノルムである．

$$\|x-y\|_1=|x_1-y_1|+|x_2-y_2|,$$
$$\|x-y\|_2=\sqrt{(x_1-y_1)^2+(x_2-y_2)^2}.$$

ここで，$x=(x_1, x_2)$，$y=(y_1, y_2)$ である．すなわち，$\|x-y\|_1$ は点 x と点 y の距離を座標軸に沿って測った l_1-ノルム（マンハッタン・ノルム）であり（図2.3(a)参照），$\|x-y\|_2$ は点 x と点 y を結ぶ線分の長さをあらわす l_2-ノルム（ユークリッド・ノルム）である．（図2.3(b)参照）

ノルムは一般に，正斉次凸関数であるから，問題(2.7)の目的関数は凸関数であるが，微分可能ではない．

設備配置問題は，(2.7)の他にも種々の定式化が試みられている．たとえば，

$$\min\ [\max_{i,j,k}(w_{ij}\|x_i-a_j\|,\ v_{ik}\|x_i-x_k\|)] \qquad (2.8)$$

のようなミニマックス型の定式化も考えられる．問題(2.8)の目的関数も，やはり微分不可能な凸関数である．

問題(2.7)や(2.8)に関するさまざまのアプローチについては，たとえば[13, 16, 24, 31]で述べられている．

2.4. ペナルティ関数

つぎに，一般的な非線形計画問題

$$\begin{aligned} \min\quad & f(x) & & (2.9)\\ \text{s.t.}\quad & g_i(x)\leqq 0 & & i=1,\cdots,m\\ & h_j(x)=0 & & j=1,\cdots,l \end{aligned}$$

の最適解を見つけることを考える．ただし，f, $g_i(i=1,\cdots,m)$, $h_j(j=1,\cdots,l)$が，問題(2.9)の局所的最適解のある近傍で連続的微分可能であると仮定する．そのとき，ある制約想定のもとで，つぎの制約のない問題の局所的最適解は，パラメータ $\mu>0$ が充分に小さければ，問題(2.9)のそれと一致することが，Pietrzykowski

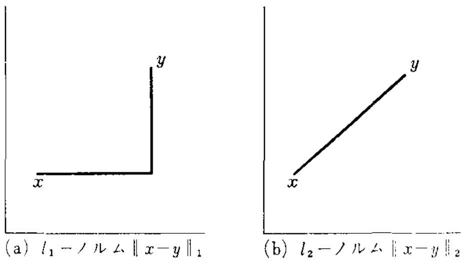

図 2.3 l_1-ノルムと l_2-ノルム

[23]によって証明されている．

$$\min\ \mu f(x)+\sum_{i=1}^{m}\max\ [0, g_i(x)]+\sum_{j=1}^{l}|h_j(x)| \qquad (2.10)$$

問題(2.10)の目的関数は一種のペナルティ関数であるが，一般に用いられている SUMT [9] と異なるのは，SUMT においてはペナルティ・パラメータを無限にゼロに近づけるか，または無限に大きくしなければならないのに対して，(2.10)ではパラメータを有限値として正確な解を得ることができる点である．その意味で，(2.10)は正確な(exact)ペナルティ関数とよばれている [4, 7, 15, 33]．

容易にわかるように，(2.10)は微分可能ではない．さらに，(2.10)の第2項および第3項をより一般的な関数でおきかえても，正確なペナルティ関数の望ましい性質が失われないようなものが存在することが示されている[1]．その場合も，ペナルティ関数は一般には微分可能ではない．

つぎに，問題(2.9)において等式制約が存在しない場合を考えよう．f, g_i は凸関数であると仮定すると，通常のペナルティ関数法は，

$$\min\quad f(x)-h_t(x) \qquad (2.11)$$

としてあらわされる．ただし，t は正のパラメータで，関数 $h_t(x)$ は x の凹関数で，

$$\lim_{t\to\infty} h_t(x)=\begin{cases}0 & g_i(x)\leqq 0(i=1,\cdots,m)\text{のとき},\\ -\infty & \text{それ以外のとき}.\end{cases}$$

なる性質をもつ（より厳密な理論は [9, 21] などを参照のこと）．ところが，ペナルティ関数(2.11)は，t が大きくなるにつれて，悪条件の(ill-conditioned)関数になる．この難点を克服するために，つぎのようなペナルティ関数が考えられている[22]．

$$\max\quad h_t{}^*(y)-f^*(y) \qquad (2.12)$$

ここで f^*, $h^*{}_t$ はそれぞれ f, h_t の共役関数([28]を参照)である．問題(2.12)が問題(2.11)と Fenchel の意味で双対であること [28] から，問題(2.12)を解くことにより，問題(2.11)の解を得ることができる．さらに，共役ペナルティ関数は，ある条件のもとで，t を大きく

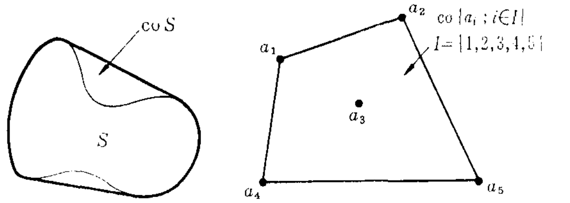

図 3.1　集合の凸包

していっても悪条件とはならないことがわかっている．ところが，共役ペナルティ関数(2.12)は，f, h_t が狭義に凸であるような特別の場合を除いて，微分可能ではない関数である．

2.5.　その他の例

以上，微分可能でない関数を含む最適化問題のいくつかの例を紹介してきたが，それ以外にも，応用上重要な問題のなかにも微分不可能な問題として定式化されるものが数多くある．そのいくつかをあげれば，チェビシェフ近似問題など各種ミニマックス問題［6］，グラフ理論に関連して起こる，n 次対称行列の固有値のうちで大きいほうから q 個($q \leqq n$)の和を，行列の対角要素の関数として最小にする問題［5］，巡回セールスマン問題の近似解を求める問題［14］，多品種最大流れ問題［14］など多岐にわたって存在している．

3.　準勾配と一般勾配の理論と応用

3.1.　凸集合と凸関数

n 次元ユークリッド空間 R^n の部分集合 C は，C 内の任意の２点を結ぶ線分がまた C に含まれるとき，凸集合(convex set)であるといわれる．そのことは，

$$(1-\lambda)x+\lambda y \in C, \quad \forall x \in C, \forall y \in C, \forall \lambda \in [0,1].$$

が満たされることと同値である．

さらに，任意の集合 S に対して，S を含む最小の凸集合を S の凸包(convex hull)といい，co S であらわす．とくに，S が有限個の点 $a_1, \cdots, a_m$ からなっているときは，S の凸包を co$\{a_i ; i \in I\}$，ただし $I=\{1, \cdots, m\}$，のようにあらわす．図 3.1 で，集合の凸包(斜線の部分)の例を示した．明らかに，任意の集合に対して凸包は一意的に定まる．

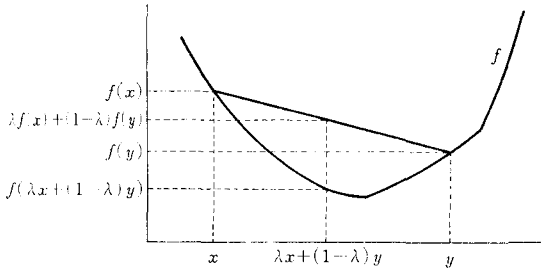

図 3.2　R^1 上の凸関数

つぎに，R^n 上で定義された実数値関数 f を考えよう．関数 f は，任意の $x \in R^n$, $y \in R^n$, および任意の実数 $\lambda \in [0,1]$ に対して，

$$f[\lambda x+(1-\lambda)y] \leqq \lambda f(x)+(1-\lambda)f(y)$$

を満たすとき，凸関数(convex function)であるといわれる．図 3.2 は，R 上の凸関数の一例を示している．凸関数はつねに連続関数であることが知られている．

任意の凸関数 f に対して，任意の等位面(等高線)で囲まれた集合

$$\{x \in R^n ; f(x) \leqq \mu\} \qquad \mu：実数$$

は凸集合になる．この集合は，しばしばレベル(level)集合とよばれる．図 3.3 は，R^2 上で定義された凸関数のレベル集合の一例である．ところがレベル集合が凸になることは，必ずしも関数の凸性を意味しないことに注意しておこう．

3.2.　凸関数の準勾配

さて，つぎに凸関数の解析的な性質について説明しよう．まず，凸関数 f が微分可能である場合を考えてみよう．そのとき，任意の点 x において，つぎの不等式

$$f(y)-f(x) \geqq \langle \nabla f(x), y-x \rangle, \quad \forall y \in R^n \quad (3.1)$$

が成立する．ここで $\nabla f(x) \in R^n$ は，f の x における勾配ベクトルであり，$\langle \cdot, \cdot \rangle$ は R^n における内積をあらわ

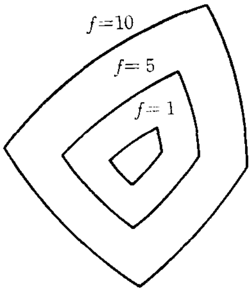

図 3.3　R^2 上の凸関数のレベル集合

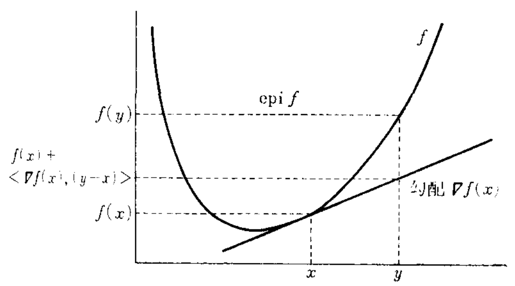

図 3.4　凸関数の勾配(微分可能な場合)

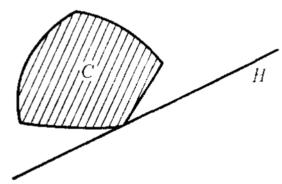

図 3.5　凸集合の支持超平面

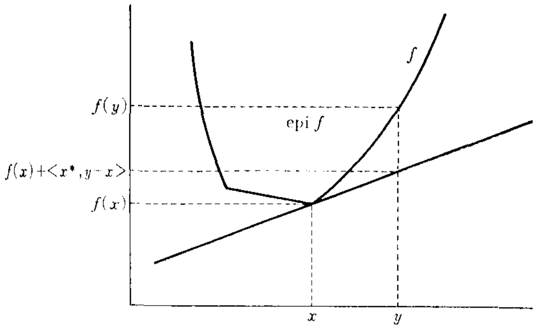

図 3.6 凸関数の準勾配(微分不可能な場合)

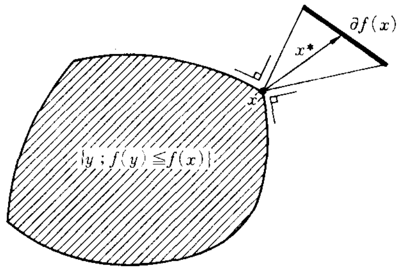

図 3.7 レベル集合と準勾配

す．(3.1)が常に満たされることは，**図 3.4** より直観的に明らかであろう．すなわち，点 x において f を一次式で近似したとき，その近似値は真の関数値を上まわることはない．

さて，上の事柄を幾何学的に厳密に述べておこう．あるベクトル $a\in R^n$ とある実数 α に対して定義される集合 $H=\{x\in R^n\,;\,\langle a, x\rangle=\alpha\}$ は，空間 R^n を二つの半空間に分離する．一般に，このような集合 H は超平面(hyperplane)とよばれるが，とくに全空間が R^3 のときは平面を，R^2 のときは直線をあらわしている．いま，R^n のある凸集合 C が，超平面 H の定める半空間の一方に含まれ，C の境界と H が共通点をもつとき，H を C の支持超平面(supporting hyperplane)という．(**図 3.5** 参照)

ここで話を凸関数 f にもどすと，不等式(3.1)は結局，空間 R^{n+1} において，点 x における f の一次近似式により定まる超平面

$$H=\{(y,\mu)\,;\,\mu=f(x)+\langle\nabla f(x),\ y-x\rangle,\ y\in R^n,\ \mu\in R\}$$

が，凸関数 f のエピグラフ(epigraph)

$$\text{epi } f=\{(y,\mu)\,;\,f(y)\geqq\mu,\ \ y\in R^n,\ \ \mu\in R\}$$

(epi f は R^{n+1} の凸集合になる)と点 $(x, f(x))$ において共通点をもつ支持超平面になっていることを意味している．この事実は，図 3.4 と図 3.5 の類似性からも推測できよう．

以上に述べてきた微分可能な凸関数に対する考察から，一般の凸関数に対しても，勾配ベクトルと類似の概念を考えることは容易である．実際，任意の凸集合はその境界上の任意の点において支持超平面をもつという事実から，R^n 上の凸関数 f に対して，R^{n+1} の点 $(x, f(x))$ において集合 epi f の支持超平面が存在することがわかる．(**図 3.6** 参照)つまり，つぎの関係を満足するベクトル $x^*\in R^n$ が存在する．

$$f(y)-f(x)\geqq\langle x^*, y-x\rangle,\qquad \forall y\in R^n. \qquad (3.2)$$

(3.1)，(3.2)から容易にわかるように，もし f が x において微分可能なら，(3.2)を満たす x^* は一意的に定まり，それが勾配ベクトル $\nabla f(x)$ にほかならない．一般的には，(3.2)を満たす x^* は一つだけとは限らないが，勾配の一種と考えられるので，凸関数 f の点 x における準勾配(subgradient)といわれる．f の x における準勾配全体の集合を $\partial f(x)$ であらわす．集合 $\partial f(x)$ は，任意の x において，有界で閉じている(すなわちコンパクトな)凸集合であることが知られている．

ここで，準勾配の性質を少し異なった角度から調べてみよう．点 x を固定して，f のレベル集合 $\{y\,;\,f(y)\leqq f(x)\}$ を考えよう．**図 3.7** は f が R^2 上で定義されている場合を示している．そのとき，x における f の準勾配 x^* は，x においてレベル集合の法ベクトル(normal vector)になっている．図 3.7 において，太い線分であらわした集合が，準勾配全体の集合 $\partial f(x)$ である．準勾配をこのように法ベクトルとして捉えることは，次章の最小化アルゴリズムの項で役に立つであろう．

いままでは，凸関数の準勾配の概念を，幾何学的な立場から解説してきたが，ここで少し解析学的な見地から眺めてみよう．R^n 上の関数 f の点 x における，方向 d に関する方向微分係数(one-sided directional derivative)は，

$$f'(x\,;\,d)=\lim_{\lambda\downarrow 0}\,[f(x+\lambda d)-f(x)]/\lambda \qquad (3.3)$$

で定義できる．一般の関数の場合は，方向微分係数は常に存在するとは限らないが，f が凸関数のときは(3.3)の極限が任意の x と d に対して常に存在することが知られている．さらに，凸関数の方向微分係数と準勾配のあいだには，つぎの重要な関係が成り立っている．

$$x^*\in\partial f(x)\Leftrightarrow \text{任意の } d \text{ に対して，}$$
$$f'(x\,;\,d)\geqq\langle x^*, d\rangle \qquad (3.4)$$

証明：式(3.2)において，$y=x+\lambda d\ (\lambda>0)$ とおくと，

$$[f(x+\lambda d)-f(x)]/\lambda\geqq\langle x^*, d\rangle$$

を得る．上の不等式の左辺は，容易に確かめられるように，λ の単調減少関数であるから，(3.4)が成立するこ

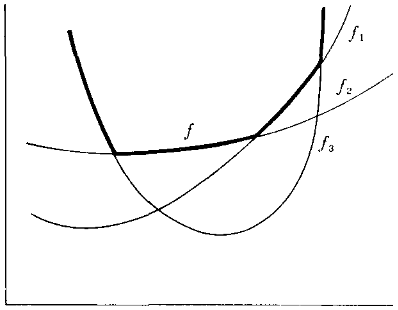

図 3.8 マックス型関数 $f=\max\{f_1, f_2, f_3\}$

とがわかる．(証明終)

関係(3.4)から直ちに，つぎの関係式が得られる．

$$f'(x\,;\,d)=\max\{\langle x^*, d\rangle\,;\,x^*\in\partial f(x)\} \tag{3.5}$$

もし，関数 f が点 x で微分可能ならば，

$$f'(x\,;\,d)=\langle\nabla f(x), d\rangle$$

となることは明らかであろう．

ところで，$f'(x\,;\,d)\geqq 0$ であることは，点 x から方向 d に沿って関数 f の値が減少しないことと等価であることに注意すると，すべての方向 d に対して $f'(x\,;\,d)\geqq 0$ が成立することは，x が f の局所的最小解であることと等価であるといえる．よく知られているように，凸関数の局所的最小解は，大域的最小解でもあるから，(3.4)の系として，つぎの結果が得られる．

$$0\in\partial f(x)\Leftrightarrow f(x)=\min_{z\in R^n} f(z) \tag{3.6}$$

つまり，関数 f が x において最小値をとるための必要充分条件は，0が点 x における f の準勾配になっていることである．これは，f が微分可能の場合には，f が x で最小になるとき，$\nabla f(x)=0$ となるという周知の事実の一般化になっている．

さて，ここで，応用上非常に有用な結果について述べておこう．つぎのマックス型の関数(**図 3.8** 参照)

$$f(x)=\max_{i\in I} f_i(x) \tag{3.7}$$

を考えよう．ただし，I は有限個の添字の集合，たとえば $I=\{1, 2, \cdots, m\}$，とする．第2章にあげた微分不可能関数のほとんどのものは，(3.7)の型の関数である．実際問題の中にあらわれる(3.7)型の関数においては，大抵の場合，各 f_i は非常に単純な関数(たとえば，一次または二次関数のような)であることが多いが，ここでは f_i は一般の凸関数であると仮定しておく．

そのとき，(3.7)で定義される関数 f もやはり R^n 上の凸関数になる．

証明：x_1, $x_2\in R^n$ とスカラー $\lambda\,(0\leqq\lambda\leqq 1)$ を任意に選ぶ．そのとき，f の定義と各 f_i の凸性より，

$$\begin{aligned}f[\lambda x_1+(1-\lambda)x_2]&=\max_{i\in I} f_i\,[\lambda x_1+(1-\lambda)x_2]\\&\leqq\max_{i\in I}\{\lambda f_i(x_1)+(1-\lambda)f_i(x_2)\}\\&\leqq\lambda\max_{i\in I} f_i(x_1)+(1-\lambda)\max_{i\in I} f_i(x_2)\\&=\lambda f(x_1)+(1-\lambda)f(x_2)\end{aligned}$$

であるから，f は凸関数である．(証明終)

つぎの公式は，(3.7)において各 f_i の準勾配が容易に計算できるときに，f の準勾配を計算するための手段を与えてくれる大変有用な公式である．

$$\partial f(x)=\mathrm{co}\,\{\partial f_i(x)\,;\,i\in I(x)\} \tag{3.8}$$

ただし，co は凸包をあらわし(3.1参照)，$I(x)$は I の部分集合 $I(x)=\{i\in I\,;\,f(x)=f_i(x)\}$ である．

証明：以下において，$A=\mathrm{co}\{\partial f_i(x)\,;\,i\in I(x)\}$ とおく．まず，

$$A=\{x^*\in R^n\,;\,x^*=\sum_{i\in I(x)}\mu_i x_i^*,\ \ x_i^*\in\partial f_i(x),\ \sum_{i\in I(x)}\mu_i=1,\ \ \mu_i\geqq 0\}$$

とあらわされることに注意しておく．

d を任意の(R^n の)ベクトルとする．任意の実数 $\lambda>0$ に対して，i_λ を添字集合 $I(x+\lambda d)$の任意の要素とする．そのとき，集合 I の要素の数は有限個だから，ある添字 $i^*\in I$ に対して，$i_{\lambda k}=i^*$ となるような0に収束する無限数列 $\{\lambda_k\}$ が存在する．$i^*\in I(x)$であることを示そう．

もし，$i^*\notin I(x)$ならば，$I(x)$の定義より $f_{i*}(x)<f(x)$ であり，$f(x+\lambda d)$と $f_i(x+\lambda d)$は λ に関して連続だから(λ に関して凸であることに注意)，充分小さい λ_k に対して，

$$f_{i*}(x+\lambda_k d)<f(x+\lambda_k d)$$

が成り立たなければならない．これは，$i^*=i_{\lambda k}\in I(x+\lambda_k d)$であることに反する．よって，$i^*\in I(x)$がいえる．

このことから，数列 $\{\lambda_k\}$ に対して，

$$[f(x+\lambda_k d)-f(x)]/\lambda_k=[f_{i*}(x+\lambda_k d)-f_{i*}(x)]/\lambda_k$$

が成り立ち，(3.3)より，

$$f'(x\,;\,d)=f_{i*}'(x\,;\,d)$$

を得る．さらに，$f(x)$および $I(x)$の定義より，

$$[f(x+\lambda d)-f(x)]/\lambda\geqq[f_i(x+\lambda d)-f_i(x)]/\lambda$$

が任意の $i\in I(x)$ に対して成り立つから，そのような i に対して，

$$f'(x\,;\,d)\geqq f_i'(x\,;\,d)$$

である．よって，(3.5)より，

$$\begin{aligned}f'(x\,;\,d)&=\max\{\langle x^*, d\rangle\,;\,x^*\in\partial f(x)\}\\&=\max_{i\in I(x)} f_i'(x\,;\,d)\end{aligned} \tag{3.9}$$

が成り立つ．ところが，

$$\begin{aligned}\max_{i\in I(x)} f_i'(x\,;\,d)&=\max\{\sum_{i\in I(x)}\mu_i f_i'(x\,;\,d)\,;\,\sum_{i\in I(x)}\mu_i=1, \mu_i\geqq 0\}\\&=\max\{\sum_{i\in I(x)}\mu_i\max\{\langle x_i^*, d\rangle\,;\,x_i^*\in\partial f_i(x)\}\,;\,\sum_{i\in I(x)}\mu_i=1, \mu_i\geqq 0\}\end{aligned}$$

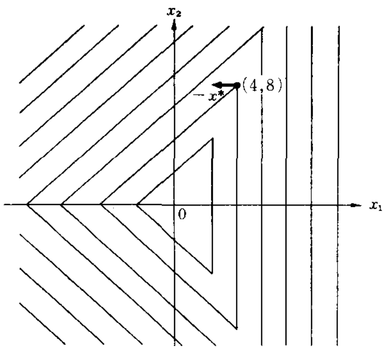

図 3.9 $f(x)=\max\{-(x_1+x_2), -x_1+x_2, x_1\}$

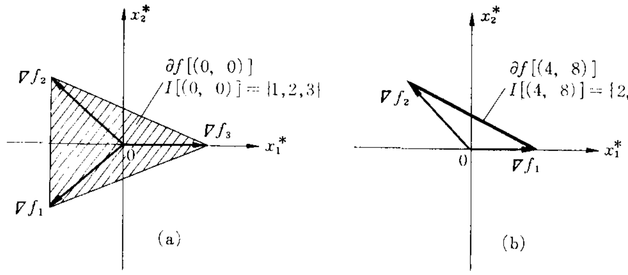

図 3.10 (a) $\partial f\ [(0,0)]$ および (b) $\partial f\ [(4,8)]$

$$=\max\{\sum_{i\in I(x)}\mu_i\langle x_i^*, d\rangle;$$
$$\sum_{i\in I(x)}\mu_i=1, \mu_i\geqq 0, x_i^*\in\partial f_i(x)\}$$
$$=\max\{\langle x^*, d\rangle ; x^*\in A\} \quad (3.10)$$

が導かれ，d は任意であったから，(3.9) および (3.10) より，すべての d に対して，

$$\max\{\langle x^*, d\rangle ; x^*\in\partial f(x)\}=\max\{\langle x^*, d\rangle ; x^*\in A\} \quad (3.11)$$

が得られる．さらに，$\partial f(x)$はコンパクト凸集合であり，すべての i に対して $\partial f_i(x)$がコンパクト凸集合であることから集合 A もまたコンパクト凸集合であるという事実を用いて，(3.11) から $\partial f(x)=A$ が示される．(このことを示すには，厳密にいうと，凸集合の分離定理を用いた議論が必要であるが，ここでは省略する．(3.11)から，$\partial f(x)=A$ が導けることは，直観的にも明らかであろう．)(証明終)

公式(3.8) の系として，(3.7) においてすべての f_i が微分可能なとき，つぎの公式が成立することがわかる．

$$\partial f(x)=\mathrm{co}\ \{\nabla f_i(x) ; i\in I(x)\} \quad (3.12)$$

簡単な例として，

$$f(x)=\max\{f_1(x), f_2(x), f_3(x)\}, \quad (3.13)$$
$$f_1(x)=-(x_1+x_2),\ f_2(x)=-x_1+x_2,\ f_3(x)=x_1$$

なる R^2 上の関数を考えてみよう．**図 3.9** は，この関数 f の等高線を示している．公式(3.12)を用いて，点(0,0) および点(4,8) における $\partial f(x)$ を計算すると，それぞれ $\partial f\ [(0,0)]=\mathrm{co}\ \{(-1,-1), (-1,1), (1,0)\}$ および $\partial f\ [(4,8)]=\mathrm{co}\{(-1,1),(1,0)\}$ となる．(**図 3.10** 参照)

ここで一つ注意すべきことは，関数 f が微分可能な場合は方向 $-\nabla f(x)$ は点 x における降下方向(その方向に微小移動したとき，関数値が減少するような方向) であるのに対して，f が微分不可能のときは方向 $-x^*$(ただし，$x^*\in\partial f(x)$)は必ずしも降下方向であるとは限らないということである．このことは，(3.13)で定義された関数 f は点 (4, 8) において，たとえば，方向 $-x^*=(-1,0)$ は関数値が増加する方向であることからも明らかであろう．(図 3.9 参照) この事実こそが，微分不可能な関数の最小化において，決定的な困難をもたらす最大の要因である．

以上，3.1 節および 3.2 節では，凸関数とその準勾配の諸性質を概説してきた．凸関数や凸集合の理論（いわゆる凸解析の理論）についての解説は多くの書物においてなされているが，基礎理論に関する教科書として，とくに Rockafellar [28] を参照されることをすすめる．

参 考 文 献

[1] D. P. Bertsekas, Necessary and sufficient conditions for a penalty method to be exact, *Math. Programming* **9** (1975), pp. 37-99.

[2] and S. K. Mitter, A descent numerical method for optimization problems with nondifferentiable cost functionals, *SIAM J. Control* **11** (1973), pp. 637-652.

[3] F. H. Clarke, Generalized gradients and applications, *Trans. Amer. Math. Soc.* **205** (1975), pp. 247-262.

[4] A. R. Conn, Constrained optimization using a nondifferentiable penalty function, *SIAM J. Numer. Anal.* **10** (1973), pp. 760-784.

[5] J. Cullum, W. E. Donath and P. Wolfe, The minimization of certain nondifferentiable sums of eigenvalues of symmetric matrices, *Math, Programming Study* **3** (1975), pp. 35-55.

[6] V. F. Demyanov and V. N. Malozemov, *Introduction to Minimax*, John Wiley, New York, 1974.

[7] J. P. Evans, F. J. Gould and J. W. Tolle, Exact penalty functions in nonlinear programming, *Math. Programming.* **4** (1973), pp. 72-97.

[8] A. Feuer, An extension of the method of conjugate subgradients to generalized minimax objectives, *School of Organization and Management*, Yale Univ., July 1976.
[9] A. V. Fiacco and G. P. McCormick, *Nonlinear Programming: Sequential Unconstrained Minimization Techniques*, John Wiley, New York, 1968.
[10] A. M. Geoffrion, Primal resource-directive approaches for optimizing nonlinear decomposable systems, *Operations Res.* **18** (1970), pp. 375-403.
[11] A. A. Goldstein, Optimization of Lipschitz continuous functions, *Math. Programming* **13** (1977), pp. 14-22.
[12] R. C. Grinold, Steepest ascent for large linear programs, *SIAM Review* **14** (1972), pp. 447-464.
[13] D. Hearn and T. J. Lowe, A subgradient procedure for the solution of minimax location problems, *Res. Rep.* No. 76-6, Indust. and Systems Engrg. Dept., University of Florida, March 1976.
[14] M. Held, P. Wolfe and P. Crowder, Validation of subgradient optimization, *Math. Programming* **6** (1974), pp. 62-88.
[15] S. Howe, New conditions for exactness of a simple penalty function, *SIAM J. Control* **11** (1973), pp. 378-381.
[16] R. F. Love, The dual of a hyperbolic approximation to the generalized constrained multi-facility location problem with lp distances, *Management Sci.* **21** (1974), pp. 22-33.
[17] D. G. Luenberger, *Introduction to Linear and Nonlinear Programming*, Addison-Wesley, Reading, Mass., 1973.
[18] O. L. Mangasarian, *Nonlinear Programming*, McGraw-Hill, New York, 1969. (非線形計画法, 関根智明訳, 培風館, 昭和47年)
[19] R. E. Marsten, W. W. Hogan and J. W. Blankenship, The boxstep method for large-scale optimization, *Operations Res.* **23** (1975), pp. 389-405.
[20] R. Mifflin, An algorithm for nonsmooth optimization, *School of Organization and Management*, Yale Univ., Dec. 1975.
[20a] ――――, An algorithm for constrained optimization with semismooth functions, *Math. of Operations Res.* **2** (1977), pp. 191-207.
[21] H. Mine and M. Fukushima, Penalty function theory for general convex programming problems, *J. Opt. Theory Appl.*, 24 (1978).
[22] ――――, K. Ohno and M. Fukushima, A "conjugate" interior penalty method for certain convex programs, *SIAM J. Control and Opt.* **15** (1977), pp. 747-755.
[23] T. Pietrzykowski, An exact potential method for constrained maxima, *SIAM J. Numer. Anal.* **6** (1969), pp. 299-304.
[24] A. Planchart and A. P. Hurter, Jr., An efficient algorithm for the solution of the Weber problem with mixed norms, *SIAM J. Control* **13** (1975), pp. 650-665.
[25] B. T. Polyak, A general method of solving extremum problems, *Soviet Math. Dokl.* **8** (1967), pp. 593-597.
[26] ――――, Minimization of unsmooth functionals, USSR Comput. Math. *Math. Phys.* **9** (3) (1969), pp. 14-29.
[27] B. N. Pshenichnyi, Necessary Conditions for an Extremum, Marcel Dekker, New York, 1971.
[28] R. T. Rockafellar, *Convex Analysis*, Princeton Univ. Press, Princeton, N. J., 1970.
[29] G. J. Silverman, Primal decomposition of mathematical programs by resource allocation: Parts I and II, *Operations Res.* **20** (1972), pp. 58-93.
[30] H. M. Wagner, *Principles of Operations Research*, Prentice-Hall, Englewood Cliffs, N. J., 1975. (オペレーションズ・リサーチ入門2, 真鍋竜太郎訳, 培風館, 昭和51年)
[31] G. O. Wesolowsky and R. F. Love, The optimal location of new facilities using rectangular distances, *Operations Res.* **19** (1971), pp. 124-130.
[32] P. Wolfe, A method of conjugate subgradients for minimizing nondifferentiable functions, *Math. Programming Study* **3** (1975), pp. 145-173.
[33] W. I. Zangwill, Non-linear programming via penalty functions, *Management Sci.* **13** (1967), pp. 344-358.