総合報告

微分不可能な関数に対する数理計画法 (2)

福 島 雅 夫

今回は，前回のつづきとして凸関数の準勾配を非凸関数に拡張した一般勾配の定義とその性質を述べるとともに，微分不可能な最適化問題に対する代表的な解法を解説する．

3.3. 局所リプシッツ関数の一般勾配

本節では，前節において考察した凸関数の準勾配の概念をより一般的な関数に拡張する．1970年代の初期に，Pshenichnyi [34] や Shor [38] などソビエトの研究者によって，通常の意味では微分できないようなあるクラスの連続関数に対して，微分係数または勾配を一般化した概念を定義し，最適性の必要条件や最小化手法を得ようとする試みがなされていた．ごく最近になって，Clarke [6, 7, 8]は，非常に広いクラスの関数に対して，一般勾配なるものが定義でき，しかもそれがすでに知られている微分可能な問題に対する諸々の結果にも広く応用できることを示した．

その基本的な考え方は，微分係数の定義式を一般化するというより，むしろ凸関数の準勾配の概念を非凸関数に対して一般化しようというものである．

つまり，前節に示したように，凸関数の方向微分係数が，準勾配を用いて評価できるという事実(3.4)，(3.5)と，方向微分係数の定義には関数の凸性は不要であることに着目し，方向微分係数をもつような任意の関数に対して，(3.4)，(3.5) と類似の関係を満たすようなものとして勾配ベクトルの一般化されたものを定義しようというのである．

さらに，Clarke は(3.3)によって定義された方向微分係数を用いるより，(3.3) をさらに一般化することによって，非常に広いクラスの関数に対して一般勾配が定義できることを指摘した．

以下では，主として Clarke に従って，一般勾配の概念を説明することにしよう．

R^n 上で定義された実数値関数 f を与え，f は局所リプシッツ連続であると仮定しよう．すなわち，R^n の任意の有界な部分集合 B に対して，ある定数 K が存在して，B 内のいかなる2点 x, y に対しても，

$$|f(x)-f(y)|\leq K\|x-y\| \quad (3.14)$$

が満たされるものとする．とくに，R^n 上の凸関数または凹関数はこの性質を有していることが知られている．([35, Th. 10.4])

関数 f の点 x における，方向 α に関する一般方向微分係数 (generalized directional derivative) は，

$$f^\circ(x;d)=\limsup_{\substack{y\to x\\ \lambda\downarrow 0}}[f(y+\lambda d)-f(y)]/\lambda \quad (3.15)$$

として定義される．(3.3) の方向微分係数の定義と比較して異なるのは，lim の代わりに lim sup となっている，すなわち極限の存在は仮定されていない点と，(3.15) では $y\to x$ に関しても lim sup をとっている点である．例として，

$$f(x)=-x \ (x\geq 0), \quad =2x \ (x<0)$$

なる R^1 上の関数 f を考えよう．定義より，

$$f'(0;d)=-d \ (d\geq 0), \quad =2d \ (d<0),$$

$$f^\circ(0;d)=2d \ (d\geq 0), \quad =-d \ (d<0)$$

となり，一般には $f'(x;d)$ と $f^\circ(x;d)$ は一致するとは限らない．$f'(x;d)$ と $f^\circ(x;d)$ がともに存在し，さらにすべての d に対して等しいとき，f は点 x において正則 (regular) であるといわれる．凸関数や連続的微分可能関数は任意の点で正則である．

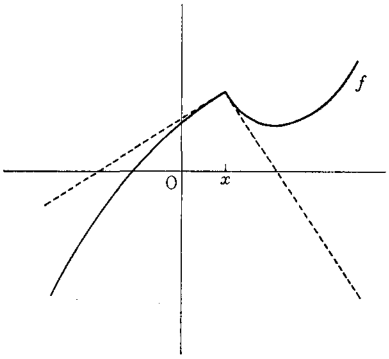
関数 f が局所リプシッツ連続ならば，(3.14) より $f^\circ(x;d)$ はつねに有限の値をとり，

$$f^\circ(x;d)\leq K\|d\| \quad (3.16)$$

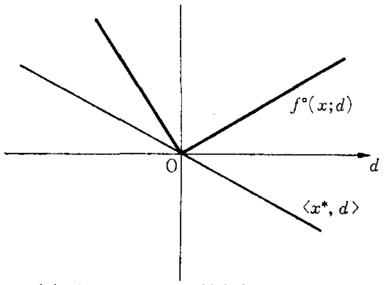
が成り立つ．さらに，任意の x において，$f^\circ(x;d)$ は d の関数として正斉次 (positively homogeneous) かつ劣加法的 (subadditive) である．すなわち，任意の実数 $\lambda>0$ に対して，

$$f^\circ(x;\lambda d)=\lambda f^\circ(x;d)$$

であり，任意のベクトル d, e に対して，



(a) 局所リプシッツ関数 f



(b) 点 x における一般方向微分係数

図 3.11 一般勾配 x^* の存在

$$f^\circ(x;d+e)\leqq f^\circ(x;d)+f^\circ(x;e)$$

が成り立つ.

証明：正斉次性は定義(3.15)より明らかであるので，劣加法的であることを示そう.

$$\begin{aligned} f^\circ(x;d+e)&=\limsup_{\substack{y\to x\\ \lambda\downarrow 0}}[f(y+\lambda d+\lambda e)-f(y)]/\lambda \\ &\leqq\limsup_{\substack{y\to x\\ \lambda\downarrow 0}}[f(y+\lambda d+\lambda e)-f(y+\lambda e)]/\lambda \\ &\quad+\limsup_{\substack{y\to x\\ \lambda\downarrow 0}}[f(y+\lambda e)-f(y)]/\lambda \\ &=f^\circ(x;d)+f^\circ(x;e). \end{aligned}$$ (証明終)

したがって，任意の点 x において，

$$f^\circ(x;d)\geqq\langle x^*, d\rangle \tag{3.17}$$

が，すべての方向 d に対して満たされるようなベクトル x^* が存在する．図 **3.11**(a)の関数 f の x における $f^\circ(x;d)$を d の関数としてあらわしたのが図3.11(b)であり，さらに.(3.17)を満たす x^* の存在も，図3.11(b)より明らかであろう．厳密には，(3.17)を満たす x^* が，少なくとも一つは存在するという事実は，関数解析におけるハーン・バナッハの定理より従う.

(3.17)を満たすベクトル x^* を，f の x における一般勾配(generalized gradient)とよび，そのような x^* 全体の集合を $\partial^*f(x)$でああらわすことにする.

そのとき，任意の点 x において，$\partial^*f(x)$は R^n の凸かつコンパクトな部分集合である.

証明：まず，$\partial^*f(x)$ が凸集合であることを示そう．x^*, y^* を任意の d に対して，

$$f^\circ(x;d)\geqq\langle x^*, d\rangle$$
$$f^\circ(x;d)\geqq\langle y^*, d\rangle$$

を満たすベクトルとし，λ を $0\leqq\lambda\leqq 1$ なる実数とすると，

$$\lambda f^\circ(x;d)\geqq\langle \lambda x^*, d\rangle$$
$$(1-\lambda)f^\circ(x;d)\geqq\langle (1-\lambda)y^*, d\rangle$$

が成り立つ．この二つの不等式を辺々加え合わせると，

$$f^\circ(x;d)\geqq\langle \lambda x^*+(1-\lambda)y^*, d\rangle$$

となり，$\lambda x^*+(1-\lambda)y^*\in\partial^*f(x)$ がいえる．よって $\partial^*f(x)$ は凸集合である．つぎに，$\partial^*f(x)$ がコンパクトであることを示そう．(3.17)より $\partial^*f(x)$ が閉集合であることは明らかであるので，それが有界であることをいえば充分である．ところが，(3.16)と(3.17)より，

$$K\|d\|\geqq f^\circ(x;d)\geqq\langle x^*, d\rangle,$$

が，すべての d について成り立つことから，$\|x^*\|\leqq K$ でなければならないのは明らかである．よって，$\partial^*f(x)$ は有界な集合であることが示された．(証明終)

ここで，凸関数の準勾配に対する関係(3.4)と，(3.17)を比較してみると，その類似性は明らかであろう．さらに，式(3.5)に対応して，一般勾配の場合にも，つぎの関係式が成立する.

$$f^\circ(x;d)=\max\{\langle x^*, d\rangle;\ x^*\in\partial^*f(x)\} \tag{3.18}$$

逆に，一般方向微分係数が与えられたとき，関係(3.18)を満たす集合 $\partial^*f(x)$を，一般勾配全体の集合として定義しても同じことである．(実際，Pshenichnyi [34] による一般化では，(3.18)の左辺を通常の方向微分係数 $f'(x;d)$ でおきかえた関係式が定義として用いられた．正則な関数に対しては，彼の定義とここでの定義は同一のものになる.)

さらに，式(3.17)による定義と等価な，もうひとつの定義が可能である．局所リプシッツ連続な関数 f は，ほとんどいたるところ微分可能であるという事実から，

$$\partial^*f(x)=\mathrm{co}\ \{\lim_{k\to\infty}\nabla f(x_k);\ x_k\to x\} \tag{3.19}$$

として，一般勾配の集合を定義する．ここで $\{x_k\}$ としては，f が x_k で微分可能であり，しかも点列 $\{\nabla f(x_k)\}$ の極限が存在し，かつ $\{x_k\}$ が点 x に収束するようなすべての点列を考えるものとする．(3.19)による定義が，(3.17)によるものと完全に一致することの証明は，ここでは省略する．([6] の Prop. (1.4)を参照されたい.)

実際問題に対しては，(3.19)による定義のほうが，(3.17)より直観的にわかりやすいように思われる．しかし，R^n より一般的な空間(たとえば，バナッハ空間など)で定義された関数に対しては，(3.19)は(3.17)と同値ではないことに注意しておこう.

先にも述べたように，凸関数は正則であるから，(3.17)は (3.4) と等価であり，一般勾配と準勾配の概念は一致する．つまり，f が凸関数ならば，$\partial f(x)=\partial^* f(x)$ が任意の点 x において成立する．

同様に，関数 f が x において，連続的に微分可能ならば，$\partial f(x)=\{\nabla f(x)\}$ が成立する．ここで，“連続的に”という仮定は本質的である．このことを見るために，

$$f(x)=x^2 \sin\ (1/x) \quad (x\neq 0), \ =0 \quad (x=0)$$

なる R^1 上の関数を考えてみよう．そのとき，関数 f はすべての点 x において微分可能であり，

$$\nabla f(x)=2x \sin(1/x)-\cos(1/x) \quad (x\neq 0),$$
$$=0 \quad (x=0)$$

である．(3.19)より，$\partial^* f(0)=\{x^*; -1\leqq x^*\leqq 1\}$ であるから，$\partial^* f(0)\neq\{\nabla f(0)\}$であることがわかる．ここで，$\partial^* f(0)$ が $\nabla f(0)$に一致しない理由は，$\nabla f(x)$ が0で連続でないことによることは明らかである．

つぎに，前節でも考察したマックス型関数を考える．

$$f(x)=\max_{i\in I} f_i(x) \tag{3.20}$$

(式(3.7)参照)ここで，Iは添字の集合 $\{1,2,\cdots,m\}$ であり，各 f_i は局所リプシッツ関数であるとする．そのとき，(3.20) で定義される関数 f も局所リプシッツ連続である．さらに，関数 f の一般勾配の集合 $\partial^* f(x)$は，

$$\partial^* f(x)\subset \mathrm{co}\ \{\partial^* f_i(x)\ ;\ i\in I(x)\} \tag{3.21}$$

を満足する．ただし，$I(x)$ は $\{i\in I; f(x)=f_i(x)\}$ で定義される I の部分集合である．(3.21)において，等号は一般には成り立たないが，各 f_i が正則ならば，

$$\partial^* f(x)=\mathrm{co}\ \{\partial^* f_i(x)\ ;\ i\in I(x)\} \tag{3.22}$$

がいえる．とくに，各 f_iが連続的に微分可能ならば，

$$\partial^* f(x)=\mathrm{co}\ \{\nabla f_i(x)\ ;\ i\in I(x)\} \tag{3.23}$$

が成立する．さらに，凸関数は正則であるから，(3.8)は (3.22) の特別な場合であることがわかる．(3.21)および (3.22)の証明は，ここでは省略する．([6,7] 参照)

いままでに見てきたように，一般勾配の考え方は凸関数の準勾配のそれと非常に密接な関連がある．とくに，正則な関数の一般勾配には，凸関数の準勾配がもっているほとんどの性質が保存されるように思われる．

最後に，つぎの非線形計画問題に対する最適性条件を考察しよう．

$$\begin{aligned} &\min \quad f(x) && (3.24)\\ &\text{subject to}\quad g_i(x)\leqq 0, \quad i=1,\cdots,m,\\ &\qquad\qquad h_j(x)=0, \quad j=1,\cdots,l. \end{aligned}$$

問題(3.24)が凸計画問題である場合，(3.24)が準凸または擬凸関数を含む場合，およびすべての関数が微分可能である場合には，さまざまな最適性の必要条件や充分条件が知られている．(たとえば，Mangasarian [24]を参照のこと．) ところで，関数 f, $g_i(i=1,\cdots,m)$ および $h_j(j=1,\cdots,l)$が局所的リプシッツ連続である場合には，つぎの結果が Clarke[7] によって与えられている．

点 x が問題(3.24) の局所的最小解ならば，つぎの条件を満足するような，すべてが0ではないような実数 $\lambda, \mu_i(i=1,\cdots,m)$, $\nu_j(j=1,\cdots,l)$が存在する．

$$\left.\begin{aligned} &\lambda\geqq 0, \quad \mu_i\geqq 0, \quad i=1,\cdots,m,\\ &\mu_i g_i(x)=0, \quad i=1,\cdots,m,\\ &0\in\lambda\partial^* f(x)+\sum_{i=1}^m \mu_i\partial^* g_i(x)+\sum_{j=1}^l \nu_j\partial^* h_j(x) \end{aligned}\right\} \tag{3.25}$$

条件(3.25)は，微分可能な問題に対する Fritz John の最適性の必要条件[24]に対応している．さらに(3.25)において，$\lambda>0$ としたものは，Kuhn-Tucker の最適性の必要条件に対応すると考えられる．そのためには，正規性の条件というある種の制約想定のようなものが必要であるが，詳細については Clarke [7] を参照されたい．

上に述べた方法とはやや異なったアプローチによる最適性条件の一般化は，Neustadt[28], Pshenichnyi[34], Craven と Mond [9], Pourciau [33], Mifflin [25] などによって与えられている．

4. 計算方法

この章では，必ずしも微分可能でない非線形関数の最小化のための計算方法について，現在までに発表されているものから代表的なものをいくつか選んで紹介する．4.1節から4.3節までは，凸関数の最小化法を考察し，引き続いて一般の非凸関数に対しても適用可能な計算方法について述べる．

以下では，つぎの最小化問題を取り扱う．

$$\min\ f(x) \qquad \text{s.t.}\ x\in C \tag{4.1}$$

ただし，f は R^n 上で定義された実数値関数，C は R^n の閉部分集合である．

ところで，(4.1)の目的関数 f の微分可能性を仮定しないならば，問題 (4.1) を制約のない問題，すなわち $C=R^n$とみなしてもほとんどの場合に一般性を失わないと考えられる．なぜなら，非常に特異な問題を除いて，

$$d_c(x)>0 \quad (x\notin C), \quad =0 \quad (x\in C)$$

なる関数 d_c を適当に選ぶことにより，制約のない問題

$$\min\ f(x)+Kd_c(x)$$

の解が，正の定数 K が充分大きいとき，問題(4.1)の解と一致するようにできるからである．とくに，

$$C=\{x\in R^n;\ g_i(x)\leqq 0\ (i=1,\cdots,\mathrm{m}),$$
$$h_j(x)=0\ (j=1,\cdots,l)\}$$

であらわされている場合には，

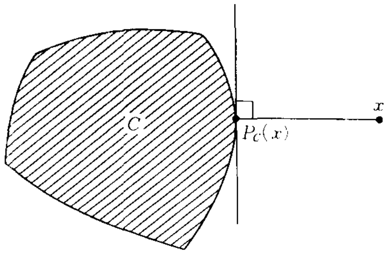


図 4.1　凸集合 C への射影作用素 P_c

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

とする方法（正確なペナルティ法，2.4節参照）が，よく知られている．

したがって，4.1節で述べる Polyak の準勾配法を除き，この章で解説する方法の大部分は制約のない問題

$$\min f(x) \quad \text{s.t.} \quad x\in R^n \tag{4.2}$$

のための方法であるが，それらは制約つきの問題(4.1)に対しても一般性を失うことなく適用できるのである．

4.1 準勾配法 (subgradient method)

問題(4.1)において，目的関数 f は凸関数，制約集合 C は凸集合であると仮定しよう．f が微分可能ならば，

$$x_{k+1}=P_c(x_k-\alpha_k\nabla f(x_k)) \tag{4.3}$$

なる反復公式によって生成される点列 $\{x_k\}$ は，ステップ長 α_k を適当に定めることにより，問題(4.1)のある最適解 $\bar{x}$ に収束することが知られている[4,16,23]．ただし，P_c は凸集合 C への射影作用素である．すなわち，R^n の任意の点 x に対して，$P_c(x)$ はつぎの条件

$$P_c(x)\in C$$
$$\text{かつ}\ \|P_c(x)-x\|=\min\{\|z-x\|;z\in C\}$$

を満たす点であり，一意的に存在する．（図 4.1 参照）

問題(4.2)に対しては，反復公式(4.3)は，

$$x_{k+1}=x_k-\alpha_k\nabla f(x_k)$$

となり，ステップ長 α_k を，

$$f(x_k-\alpha_k\nabla f(x_k))=\min\{f(x_k-\alpha\nabla f(x_k));\alpha\geqq 0\} \tag{4.4}$$

で定めるなら，これはよく知られた最大傾斜法による．

そこで，f の微分可能性が仮定されない一般の凸関数の場合にも，反復公式(4.3)の $\nabla f(x)$ を x における準勾配 x^* でおきかえることにより，点列 $\{x_k\}$ が最適解に収束するようにできるのではないかと考えるのは，ごく自然な発想であろう．この考え方にもとづいて，Polyak[31] は，つぎの反復公式(4.5)を提案し，その最適解への収束を証明した．

$$x_{k+1}=P_c(x_k-\alpha_k x_k^*) \tag{4.5}$$

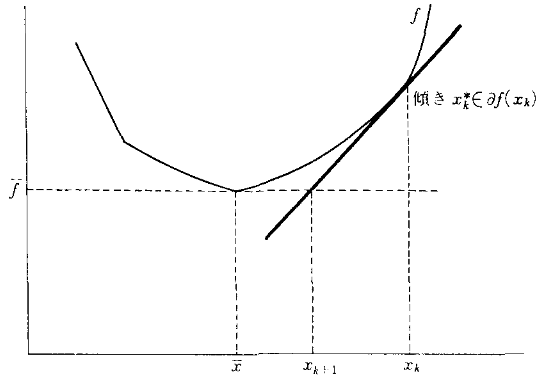


図 4.2　Polyak の方法 ($\lambda_k=1$)

ここで，x_k^* は $\partial f(x_k)$ の任意の要素であり，ステップ長 α_k は，

$$\lambda_k=\alpha_k\|x_k^*\| \tag{4.6a}$$

とおくと，

$$\lim_{k\to\infty}\lambda_k=0 \quad \text{かつ} \quad \sum_{k=1}^{\infty}\lambda_k=\infty \tag{4.6b}$$

を満足するように定められているものとする．(4.6)によってステップ長を決めるということは，余分な計算（たとえば(4.4)のような）をする必要がないという点で，たいへん魅力的ではあるが，残念ながら収束の速さという観点からはあまり有効ではないのである．

たとえば，$f(x)=\|x\|$，$C=R^n$ の場合を例にとってみよう．$x\neq 0$ なる点においては，つねに $\|x^*\|=1$ であるから，すべての k に対して $\alpha_k=1/k$ と選べば(4.6)は満足される．そのとき，ある有限な k で $x_k=0$ となってしまう特別な場合を除いて，定数 $C>0$ が存在して不等式 $\|x_k\|\geqq c/k$ が満たされる．これは，最も遅い収束の代表的なタイプの一つである．

そこで，Polyak[32] は(4.5)の収束の速さを改善するために，つぎの公式によってステップ長 α_k を定める方法を提案した．

$$\alpha_k=\lambda_k(f(x_k)-\bar{f})/\|x^*\|^2,\ 0<\varepsilon_1\leqq\lambda_k\leqq 2-\varepsilon_2,\ \varepsilon_2>0 \tag{4,7}$$

ここで，$\bar{f}=\min\{f(x);\ x\in C\}$ であり，$\varepsilon_1,\varepsilon_2$ は適当な定数である．(4.7)の意味は，つぎのように解釈できる．簡単のため，$C=R^1$ とし，すべての k に対して $\lambda_k=1$ と仮定しよう．そのとき，図4.2からわかるように，(4.5)によって定まる点 x_{k+1} は，関数 f のグラフの点 x_k における傾き x_k^* の接線と最適解 $\bar{x}$ における傾き0の接線の交点になっている．このことは，反復公式(4.5)，(4.7)が，方程式 $f(x)-\bar{f}=0$ を解くためのニュートン法とも非常に深い関連をもっていることを示唆している．さらに Polyak は，ある条件のもとで，方法(4.5)，

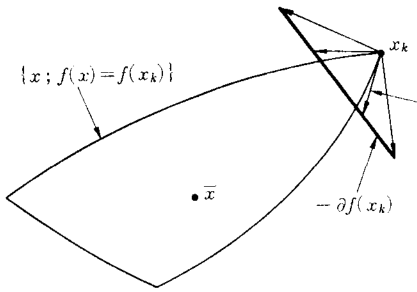


図 4.3 準勾配法

(4.7)によって生成される点列が最適解 $\bar{x}$ に1次収束をすること，すなわち，

$$\|x_k-\bar{x}\|\leqq q^k\|x_0-\bar{x}\|,\quad q<1$$

を満たす定数 q が存在することを示した．このことは，方法(4.5)，(4.7)が方法(4.5)，(4.6)より収束の速さの面ではかなりすぐれていることを意味している．

(4.7)では，関数 f の最小値 $\bar{f}$ が既知であると仮定されていた．実際，問題によっては $\bar{f}$ の値が，まえもってわかっている場合が少なくない．たとえば連立方程式

$$f_i(x)=0,\quad i=1,\cdots,n$$

の根を求める問題を，$f(x)=\max\{|f_i(x)|\,;\,i=1,\cdots,\mathrm{n}\}$ なる関数を導入することにより，$\min f(x)$ を求める問題に帰着することができるが，根が存在するときは明らかに $\bar{f}=0$ である．ところが一般の問題(4.1)の場合には $\bar{f}$ の値は未知であるので，そのときはつぎのような方法が用いられる．

まず，$\bar{f}$ の適当な推定値 $\tilde{f}$ をもちいて，反復(4.5)，(4.7)を実行する．そのとき，つぎの2つの場合が起こり得る．(1) ある．k において $f(x_k)\leqq\tilde{f}$ となるか，$\lim_{k\to\infty}f(x_k)=\tilde{f}$ となる．(2)ある実数 $\varepsilon>0$ が存在して，すべての k に対して $f(x_k)\geqq\tilde{f}+\varepsilon$ となる．(1)の場合には，$\tilde{f}>\bar{f}$ であるから，推定値 $\tilde{f}$ を小さくして反復(4.5)(4.7)を続行する．(2) の場合は，$\tilde{f}$ を大きくして反復(4.5)(4.7)を行なう．この手続きを繰返すことにより，推定値 $\tilde{f}$ は真の最小値 $\bar{f}$ に近づき，真の最適解 $\bar{x}$ に収束する点列 $\{x_k\}$ が得られる．

さて，ここで(4.5)において $x_k{}^*$ は点 x_k における任意の準勾配であったことを思い出そう．このことは，3.2節の終りに注意したように，方向 $-x_k{}^*$ は必ずしも関数 f の降下方向ではないことを意味している．したがって，数列 $\{f(x_k)\}$ は単調減少するとは限らないし，ステップ長 α_k を公式(4.4)のような方法で決定することも適当ではない．しかし，**図4.3** からもわかるように，点 x_k から方向 $-x_k{}^*$ に沿って移動するとき，ステップ長 α_k が小さければ，つぎの点 x_{k+1} は点 x_k よりも最適解 $\bar{x}$ に近づくことが示される．

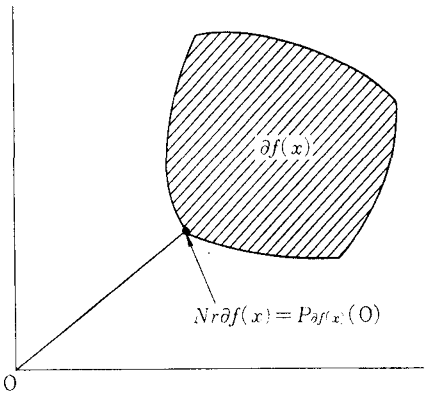


図 4.4 $\partial f(x)$におけるノルム最小の準勾配

この節で述べてきた準勾配法は，以下の節で解説する多くの方法の基礎となる非常に重要なものである．準勾配法の収束率に関するくわしい研究は Goffin [15] によって，OR においてあらわれる各種の問題に対する応用に関しては Held ら [19] の論文でそれぞれ述べられている．

4.2 ε 準勾配法 (ε-subgradient method)

前節で見たように，準勾配法は一般に降下法の性質をもたないが，このことは，たとえば，収束判定の困難化などをもたらし，実際の計算上あまり好ましいことではない．そこで準勾配法のそのような欠点を改良して降下法の性質をもつ方法がいくつか考案されているが，この章の残りはそれらの解説にあてることにする．

これからは，制約のない問題(4.2)のみを考えることにする．関数 f は凸であり，任意の点において集合 $\partial f(x)$が完全に計算できるものと仮定する．最後の仮定は，ある種の問題たとえば f がマックス型の関数(3.7)で添字集合 I の要素の数が比較的少ない場合などに対しては不自然なものではないことに注意しておこう．((3.8)，(3.12)参照)

そのとき，集合 $\partial f(x)$ の要素の中でノルムが最小になるようなものを $Nr\partial f(x)$ であらわすと ($Nr\partial f(x)=P_{\partial f(x)}(0)$ とも書ける．**図4.4**参照)，

$$-Nr\partial f(x)=f'(x;\tilde{d})\tilde{d} \tag{4.8}$$

が成り立つことが容易に確かめられる．ただし，$\tilde{d}$ は，

$$f'(x;\tilde{d})=\min\{f'(x;d)\,;\|d\|\leqq 1\},\ \|\tilde{d}\|\leqq 1 \tag{4.9}$$

を満足するベクトルである．(4.8)，(4.9) より，$-Nr\partial f(x)$ は x において関数 f の減少率が最も大きいような方向，すなわち最大降下方向 (direction of steepest descent) と見なすことができる．そこで(4.4)において $\nabla f(x)$ を $Nr\partial f(x)$ でおきかえることにより，最大傾斜

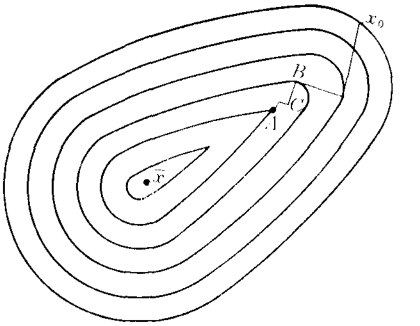


図 4.5 最大傾斜法が収束しない例

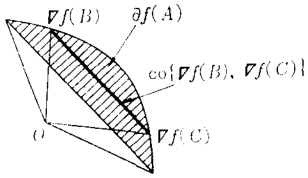


図 4.6
$\partial f(A)$の近似

法を一般化できると思われるかもしれない．ところがこの方法は，つぎの例に見るように，必ずしも最適解に収束しないのである．

図 4.5 で，実線はfの等高線をあらわし，太い実線は初期点を x_0 としたときの最大傾斜法によるトラジェクトリをあらわすものとする．そのとき，点列$\{x_k\}$は点Aに収束し，明らかに最適解$\bar{x}$へは収束しないことがわかる．この原因は，$Nr\partial f(x)$がxの関数として，点Aで連続ではないことにある．この欠点を改良した最大傾斜法の一般化が，以下に紹介する Bertsekas と Mitter [5] の ε 準勾配法である．

εを非負の実数とする．そのとき，つぎの不等式を満足するベクトル x^* を凸関数fの点xにおけるε準勾配とよび，そのようなx^*全体の集合を $\partial_\varepsilon f(x)$であらわす．

$$f(y)-f(x)\geqq\langle x^*, y-x\rangle-\varepsilon,\quad \forall y\in R^n \qquad (4.10)$$

明らかに，$\partial_0 f(x)=\partial f(x)$ であり，$0\leqq\varepsilon\leqq\varepsilon'$ ならば $\partial_\varepsilon f(x)\subset\partial_{\varepsilon'}f(x)$が成り立つ．さらに，任意の $\varepsilon\geqq0$ に対して，

$$0\in\partial_\varepsilon f(x)\Leftrightarrow f(x)\leqq\inf_{x\in R^n}f(x)+\varepsilon \qquad (4.11)$$

なる関係が成立することが，(4.10)より導かれる．((3.6)は(4.11)の特別な場合である．) つぎに，$0\notin\partial_\varepsilon f(x)$ のときは，

$$\sup\{\langle d, x^*\rangle;\ x^*\in\partial_\varepsilon f(x)\}<0 \qquad (4.12)$$

であるようなベクトル d が存在し，

$$f(x)-\varepsilon>\inf\{f(x+\lambda d);\ \lambda\geqq0\} \qquad (4.13)$$

なる関係が成立することが示される．とくに，$d=-Nr\partial_\varepsilon f(x)$ は (4.12) を満足することに注意しておく．Bertseka と Mitter によって提案された ε 準勾配法はつぎのように書ける．

ステップ 1　はじめに初期点 x_0 とパラメータ $\varepsilon_0>0$ および $0<a<1$ を設定する．$k=0$とおく．

ステップ 2　x_k および $\varepsilon_k>0$ に対して，$0\notin\partial_{a^p\varepsilon_k}f(x_k)$ となるような最小の整数 $p\geqq0$ を求め，$\varepsilon_{k+1}=a^p\varepsilon_k$ とおく．

ステップ 3　$\varepsilon=\varepsilon_{k+1}$ に対して(4.12)を満足する d をみつけ，$x_{k+1}=x_k+\lambda d$ とおく．ただし λ は $f(x_k)-\varepsilon_{k+1}>f(x_k+\lambda d)$ を満たす正数である．$k=k+1$ としてステップ 2 へもどる．

このアルゴリズムは常に well defined である．実際，(4.11)より x_k が最適解でなければステップ 2 において p は常に存在するし，(4.12), (4.13)よりステップ 3 の λ および d も常に存在する．さらに，$k\to\infty$のとき $\varepsilon_k\to0$ となり，点列 $\{x_k\}$ が問題 (4.2)の最適解に収束することも証明できる[5]．

この方法は理論的には大いに興味深いものであるが，残念ながら実用面ではかなり問題点が多い．そのおもな原因は，任意の点 x において $\partial_\varepsilon f(x)$ を計算しなければならないということである．この欠点を改良するために，Lemarechal [20] は点 x の周辺のいくつかの点で準勾配を一つずつ選び出し，それらの凸包で $\partial_\varepsilon f(x)$ を近似的にあらわそうと考えた．さらに彼は [21] において，この方法が共役勾配法の拡張になっていることを示した．ほとんど同じ頃，Wolfe [42] は Lemarechal の方法と非常によく似た方法を提案し，やはり共役勾配法との関連を指摘した．ここでは，それらを総称して共役準勾配法とよび次節においてその考え方を説明する．

4.3 共役準勾配法 (conjugate subgradient method)

Lemarechal の方法 [20, 21] と Wolfe の方法 [42] はその基本的な考え方においてほとんど同じなので，ここでは Wolfe の方法に的を絞って解説することにする．

さて，ここで図4.5を思い出してみよう．点 A における準勾配の集合 $\partial f(A)$ は**図4.6**のようにあらわされる．他方，点 B と点 C においてはfは微分可能であるから，$\nabla f(B)$および$\nabla f(C)$は図4.6に示されるようなベクトルである．そのとき，$\nabla f(B)$ や$\nabla f(C)$はそれぞれ単独では $\partial f(A)$ に関する情報をあまり与えてくれないが，その凸包 co$\{\nabla f(B), \nabla f(C)\}$ は $\partial f(A)$の比較的よい近似になっていることがわかるであろう．このような考え方をもとにして，ある点における探索方向を決定するときに，それ以前に通過してきた点における勾配または準勾配ベクトルから得られる情報を使用しようというのが，つぎに述べる Wolfe の方法のアルゴリズムである．

ステップ 1　はじめに，パラメータ $\varepsilon>0$, $\delta>0$, $m_2<$

$m_1 < 1/2$ を設定し，初期値として点 x_0，$\forall x_0^* \in \partial f(x_0)$ を選ぶ．$G_0=\{x_0^*\}$，$a_0=0$，$k=0$ としてステップ 2 へ進む．

ステップ 2　$d_k=-NrG_k$ を求める．もし$\|d_k\|<\varepsilon$ ならステップ 3 へ，そうでなければステップ 5 へ進む．

ステップ 3　$a_k \leqq \delta$ なら終了．$a_k > \delta$ ならステップ 4 へ．

ステップ 4　$x_{k+1}=x_k$, $x_{k+1}^*=x_k^*$,　$G_{k+1}=\{x_k^*\}$, $a_{k+1}=0$ としてステップ 2 へもどる．

ステップ 5　つぎの条件 (i)(ii)または(i)(iii)を満たす $t \geqq 0$ および $x_{k+1}^* \in \partial f(x_k+td_k)$を求める．

(i)　$\langle x_{k+1}^*, d_k \rangle \geqq -m_1\|d_k\|^2$,

(ii)　$f(x_k+td_k)-f(x_k) \leqq -m_2 t\|d_k\|^2$,

(iii)　$t\|d_k\| \leqq b$.

もし(ii)が満足されるなら，$x_{k+1}=x_k+td_k$ とおき，(iii)が満足されるときは $x_{k+1}=x_k$ とおいてステップ 6 へ．

ステップ 6　G の部分集合 H を適当に選び，$G_{k+1}=H \cup \{-d_k, x_{k+1}^*\}$ および $a_{k+1}=a_k+\|x_{k+1}-x_k\|$ とおいてステップ 2 へもどる．

ここで，パラメータ ε と δ は充分小さく選ぶことにする．$a_k \leqq \delta$ ならば G_k の各要素は $\partial f(x_k)$ のある要素に充分近いと期待され，さらに条件 $\|d_k\|=\|-NrG_k\|<\varepsilon$ は $\|Nr\,\partial f(x_k)\|$ が充分小さいことを意味する．そのとき，(3.6)より x_k は近似的に最適解とみなされるので，ステップ 3 の最適性の判定が妥当であることがわかる．またa_k が増加するとともに G_k には次第に古い情報が含まれてゆくことになり，G_k が $\partial f(x_k)$を近似しているとはいい難くなる．したがって，その時にはステップ 4 のリセットの操作が必要になる．ステップ 5 の条件(i)(ii)は通常の降下法においてよく知られたステップ長の決定方法であり[41]，x_k からx_{k+1} へ動くとき f の値が充分に減少することを示している．また，条件 (i) (iii)が成立するときは，方向 d_k は f の値を減少させるのに不適当であるので，そのときは x_k は変化させずに，ステップ 6 で G_k の $\partial f(x_k)$に対する近似度を改良するにとどめ，新たな探索方向を決定しようというのである．

なお，ステップ 5 を実行するために Wolfe は 2 分割法による一次元探索を提案している．なお，ステップ 2 で NrG_k を計算する問題は 2 次計画問題として定式化できるが，とくにこのような問題に対してすぐれた方法[27, 43]が考案されている．

Wolfe や Lemarechal が示したように，この方法は 2 次関数に対しては共役勾配法の性質を有しているので，一般的な関数に対しても非常によい収束性が期待される．

4.4　共役準勾配法の非凸関数への拡張

前節で述べた Wolfe の方法を一般化して，非凸関数に対しても適用できるアルゴリズムを開発する試みが Feuer [13] と Mifflin [25a] によってなされている．それらの方法は，基本的には Wolfe の方法において準勾配の演じている役割を一般勾配 (3.3節参照) で代行させようというものである．その意味で Feuer および Mifflin の方法は，Wolfe の方法と似かよっており，ここではそれらのアルゴリズムを詳述することは避ける．しかし，それらの方法においては Wolfe の方法に対して数々の改良が加えられており，とくに一次元探索法の巧妙さや収束の証明法など非常に見るべきものが多く，単に Wolfe の方法を形式的に一般化したもの以上の価値があると思われる．

さらに Mifflin [25, 25a] は，制約つきの問題 (4.1) (ただし，$C=\{x; g_i(x) \leqq 0, i=1, \cdots, m\}$) も以下のような方法で取り扱えることを指摘した．まず，

$$g(x)=\max\{g_i(x); i=1, \cdots, m\}$$

と定義することにより，問題(4.1)は，

$$\min f(x) \quad \text{s.t. } g(x) \leqq 0$$

なる制約式が一つの問題に書きかえられる．このとき，各 g_i が一般勾配をもつなら，g も一般勾配をもつから ((3.21)―(3.23) 参照)，任意の x に対してつぎの集合 $M(x)$を対応させることができる．

$$M(x)=\begin{cases} \partial^* f(x) & (g(x)<0\text{のとき}) \\ \text{co}\{\partial^* f(x) \cup \partial^* g(x)\} & (g(x)=0\text{のとき}) \\ \partial^* g(x) & (g(x)>0\text{のとき}) \end{cases} \tag{4.13}$$

そのとき最適性の条件(3.25)から点 x が最適解であるための必要条件は，$0 \in M(x)$ であることが容易にわかる．つまり，制約のない問題に対する方法で $\partial^* f(x)$ が使われているところを制約つきの問題では上の $M(x)$ で代用することにより，一般性を失うことなく Wolfe の方法や Feuer および Mifflin の方法が適用できるのである．なお，Levitin [22] にも(4.13)と同様なアイデアが使われている．

4.5　その他

以上の方法の他に数多くの微分不可能関数の最小化法が開発されている．大ざっぱにいって，それらはマックス型関数の最小化問題 (ミニマックス問題) に対するものと，もっと一般的な関数に対しても適用できる方法に分類できる．前者に属するものとしては[10, 11, 12]などがあり，後者に属するものとしては[1, 17, 18, 22, 26, 29, 30, 36, 37, 38, 39]などがある．

それらの方法はすべて，準勾配や一般勾配または方向

表 1　代表的な計算法の比較

	降下法	仮　定	探索方向	一次元探索	収束速度*	適用可能な関数
準勾配法 Polyak [31]	NO	$\partial f(x)$ の要素が一つ計算できる	$\forall x^* \in \partial f(x)$	不　要	$1/n$	凸関数
準勾配法 Polyak [32]	NO	$\partial f(x)$ の要素が一つ計算できる	$\forall x^* \in \partial f(x)$	不　要	1 次	凸関数
ε 準勾配法 Bertsekas & Mitter [5]	YES	$\partial_\varepsilon f(x)$ 全体が計算できる	部分問題を解く	要	? (最大傾斜法)	凸関数
共役準勾配法 Wolfe [42]	YES	$\partial f(x)$ の要素が一つ計算できる	部分問題を解く	要	? (共役傾斜法)	凸関数
Mifflin [25]	YES	$\partial^* f(x)$ の要素が一つ計算できる	部分問題を解く	要	? (共役傾斜法)	局所リプシッツ関数

*) カッコ内は微分可能な関数または 2 次関数に適用した場合.

微分係数を使って，微分不可能関数を直接取り扱おうとするものであるが，それ以外にも微分不可能関数を微分可能な関数で近似し，その近似関数を微分を用いる通常の方法（たとえば，勾配法，共役勾配法，DFP 法など）により（近似）最適解を計算しようとする試みがある．その目的に沿った近似法は [2, 3, 14, 40] で提案されている．

最後に本節で解説した方法の特徴を**表 1** にまとめておく．

5. む す び

以上，一般的な最適化問題に対して得られている結果を簡単に解説してきたが，その研究の現状は微分可能性の仮定のもとで得られている膨大な量の成果に比べていまだ充分とはいい難い．

とくに，理論面の研究は凸解析の理論や一般的な微分の理論などかなりのレベルに達しているのに対して，計算手法に関してはごく最近になってようやく活発な研究が行なわれるようになってきたといえるであろう．実際に微分不可能問題として定式化される問題が少なくないという事実を考えると今後とくに計算手法の開発やその実際問題への応用といった方向での研究の発展が望まれる．

本稿では第 2 章および第 3 章の記述に相当分の紙面をさいたため，第 4 章においていくつかの興味ある方法について触れることができなかったことを断わっておく．現在までに得られている成果については，[5, 13, 25, 42] などの序文において系統立てて述べられている．また，おもな論文は本稿でもできるだけ引用するよう努めたが，前記の論文の参考文献にも多くの重要な論文がリストアップされているのでそれらも合わせて参照されたい．

終りに，平素御指導いただき，また本稿をまとめるにあたっても貴重なご助言をくださった，京都大学工学部三根久教授に心から感謝の意を表します．

参 考 文 献

[1] L. G. Bazhenov, Convergence conditions of a minimization method of almost-differentiable functions, *Cybernetics* **8** (1972), pp. 607–609.

[2] D.P. Bertsekas, Nondifferentiable optimization via approximation, *Math. Programming Study* **3** (1975), pp. 1–25.

[3] ——, A general method for approximation based on the method of multipliers, *Proc. 13th Annual Allerton Conference on Circuit and System Theory*, Illinois, 1975, pp. 192–203.

[4] ——, On the Goldstein-Levitin-Polyak gradient projection method, *IEEE Trans. Auto. Control* AC-21 (1976), pp. 174–184.

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

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

[7] ——, A new approach to Lagrange multipliers, *Math. of Operations Res.* **1** (1976), pp 165–174.

[8] ———, Generalized gradients of Lipschitz functionals, *MRC Tech. Rep.* #1687, Math. Res. Center, Univ. of Wisonsin, Oct. 1976.

[9] B. D. Craven and B. Mond, Sufficient Fritz John optimality conditions for nondifferentiable convex programming, *J. Australian Math. Soc.* **19** (1976), pp. 462-468.

[10] V.F. Demyanov, Algorithms for some minimax problems, *J. Computer and System Sci.* **2** (1968), pp. 342-380.

[11] ———, On the maximization of a certain nondifferentiable function, *J. Opt. Theory and Appl.* **7** (1971), pp. 75-89.

[12] ——— and V. N. Malozemov, *Introduction to Minimax*, John Wiley, New York, 1974.

[13] A. Feuer, An extension of the method of conjugate subgradients to generalized minimax objectives, *School of Organization and Management*, Yale Univ., July 1976.

[14] A. M. Geoffrion, Objective function approximations in mathematical programming, *Math. Programming* **13** (1977), pp. 23-37.

[15] J. L. Goffin, On convergence rates of subgradient optimization method, *Math. Programming* **13** (1977), pp. 329-347.

[16] A. A. Goldstein, Convex programming in Hilbert space, *Bull. Amer. Math. Soc.* **70** (1964), pp. 709-710.

[17] ———, Optimization with corners,in *Nonlin ear Programming* 2, eds. O. L. Mangasarian et al., Academic Press, 1975, pp. 215-230.

[18] ———, Optimization of Lipschitz continuous functions, *Math. Programming* **13** (1977), pp. 14-22.

[19] M. Held, P. Wolfe and P. Crowder, Validation of subgradient Optimization, *Math. Programming* **6** (1974), pp. 62-88.

[20] C. Lemarechal, An algorithm for minimizing convex functions, *Information Processing 74*, North Holland, 1974, pp. 552-556.

[21] ———, An extension of Davidon methods to non differentiable problems, *Math. Programming Study* **3** (1975), pp. 95-109.

[22] E.S. Levitin, A general minimization method for unsmooth extremal problems, *USSR Camput. Math. Math. Phys.* **9 (4)** (1969), pp. 63-93.

[23] ——— and B.T. Polyak, Constrained minimization problems, *USSR Comput. Math. Math. Phys.* **6(5)** (1966), pp. 1-50.

[24] 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).

[25] R. Mifflin, Semismooth and Semiconvex functions in constrained optimization, *SIAM J. Control and Opt.* 15(1977), pp. 959-972.

[25a] ———, An algorithm for constrained optimization with semismooth functions, *Math. of Operations Res.* **2** (1977), pp. 191-207.

[26] H. Mine and M. Fukushima, A minimization method for a class of nondifferentiable nonconvex functions, *Dept. of Appl. Math. and Phys.*, Kyoto Univ., Dec. 1977.

[27] B. F. Mitchell, V. F. Demyanov and V. N. Malozemov, Finding the point of a polyhedron closest to the origin, *SIAM J. Control* **12**(1974), pp. 19-26.

[28] L. W. Neustadt, *Optimization : A Theory of Necessary Conditions*, Princeton Univ. Press, Princeton, N. J., 1976.

[29] E.A. Nurminskii, The quasigradient method for the solving of the nonlinear programming problems, *Cybernetics* **9** (1973), pp. 145-150.

[30] S. C. Parikh, Approximate cutting planes in nonlinear programming, *Math. Programming* **11** (1976), pp. 194-198.

[31] B. T. Polyak, A general method of solving extremum problems, *Soviet Math. Dokl.* **8** (1967), pp. 593-597.

[32] ———, Minimization of unsmooth functionals, *USSR Comput. Math. Math. Phys.* **9(3)** (1969), pp. 14-29.

[33] B. H. Pourciau, Analysis and optimization of Lipschitz continuous mappings, *J. Opt. Theory and Appl.* **22** (1977), pp. 311-351.

[34] B. N. Pshenichnyi, *Necessary Conditions for an Extremum*, Marcel Dekker, New York, 1971.

[35] R.T. Rockafellar, *Convex Analysis*, Princeton Univ. Press, Princeton, N. J., 1970.

[36] N. Z. Shor, Utilization of the operation of space dilatation in the minimization of convex functions, *Cybernetics* **6** (1970), pp. 7-15.

[37] ______, Convergence rate of the gradient descent method with dilatation of the space, *Cybernetics* **6** (1970), pp. 102-108.
[38] ______, A class of almost-differentiable functions and a minimization method for functions of this class, *Cybernetics* **8** (1972), pp. 599-606.
[39] ______, Convergence of a gradient method with space dilation in the direction of the difference between two successive gradients, *Cybernetics* **11** (1975), pp. 564-570.
[40] G.O. Wesolowsky and R.F. Love, A nonlinear approximation method for solving a generalized rectangular distance Weber problem, *Management Sci.* **18** (1972), pp. 656-664.
[41] P. Wolfe, Convergence conditions for ascent methods, *SIAM Review* **11** (1969), pp. 226-235.
[42] ______, A method of conjugate subgradients for minimizing nondifferentiable functions, *Math. Programming Study* **3** (1975), pp. 145-173.
[43] ______, Finding the nearest point in a polytope, *Math. Programming* **11** (1976), pp. 128-149.
(ふくしま・まさお　京都大学工学部)

SOLE――logistics

Sole という文字を見ると，食いしん坊の筆者などはすぐレストランのメニューにある“舌びらめ”のことかと思い，また，人によっては，feme sole（独身の女性）に胸をおどらせるかもしれません．しかし，これは，

Society Of Logistics Engineers

の頭文字をとったもので，れっきとした学会の名称です．そのジャーナルである Logistics Spectrum の最新号(1978年春)は Vol. 12, No. 1 ですから，学会発足以来，10年以上経過していることはたしかです．

筆者は寡聞にしてこの学会のことを知らず，昨冬 Virginia Polytechnic Institute & State University を訪れたとき，同大学工学系大学院の工学普及部長 Benjamin S. Blanchard 教授から“logistics”を連発されて大いに弱り，一体なんのことかといろいろ質問をあびせて，ようやくつぎの程度のことを知りました．

Logistics は辞書を引くと，後方業務，兵たん業などと書いてあります．IHIの宮内一郎氏によると「後方支援」と訳すのがよいそうです．これはまさに軍事用語ですが，その点ではORも同じことでしょう．軍事にかぎらず，あらゆる組織の活動は，strategy, tactics, logistics の三つにわけられる．strategy（戦略）は仕事を規定し，tactics（戦術）は仕事を行なうのに対し，logistics（後方支援）は，仕事を遂行できるように資材を補給する．ここでいう資材のなかには，物的資材のほかに設備・資金・人力・情報までも含んでいる．ということになりますと，logistics は非常に重要な概念になり，特定の機能を指すのではなくて，interdisciplinary なもの，いや，ORとまったく同じものと考えねばならないように見えます．

SOLE の機関紙には，logistics をつぎのように定義しています．

> “The art and science of management, engineering, and technical activities concerned with requirements, design, and supplying and maintaining resources to support objectives, plans, and operations.”

こうなると，いよいよORの一部の手法と区別がつかなくなったので，SOLE の国際関係副会長でもある前出のブランチャード博士がこの5月に来日されたとき，ぶしつけに聞いてみました．彼は「ORは手法であるけれども，logistics は management そのものである」といいます．しかし，これだけの説明では，トップ・マネージメントから技術スタッフ，現場の職組長までを勧誘してメンバーにしている SOLE の活動がどういうものか，じかに接したことのない者にはよくわかりません．そこで，米国ではORワーカーのどのくらいの人数が SOLE のメンバーになっているのかと恐る恐る伺いをたてますと，「それは調査していなくてよくわからないが，一部のORワーカーはたしかにメンバーになっている．米国のOR学会とは若干競争関係にある」と洩らしてくれました．SOLE の日本支部もやがてできるでしょうし（もうできているのかもしれません），そうなるとOR学会との関係もいささか微妙になることでしょう．たしかに SOLE のメンバーの層はOR学会よりも広く，あえて大胆な想像をすると，直接には戦争（企業の場合には生産）に加わらないすべての人々のためのものとなり，トップからQCサークルのメンバーまでも包含する学会であると考えられるでしょう．　(T.O.)