最適化と凸関数

この資料は，室田一雄: 離散凸解析，共立出版, 2001 の 1.1 節に改訂を加えたものである．

関数 $f(x)$ が与えられたとき，その最小値を与える x を求めよというような問題を最適化問題と呼ぶ．例えば，2 次関数

$$f(x) = x^2 - 8x + 3 \tag{1}$$

の最小値を求めよという問題である．

上の 2 次関数の最適化問題は簡単で，$f(x)$ の微分 $f'(x) = 2x - 8$ がゼロになる点を計算して $x = 4$ が答えとなる．では，4 次関数

$$f(x) = 3x^4 - 4x^3 - 12x^2 + 3 \tag{2}$$

の場合はどうだろうか．同様に微分すると，$f'(x) = 12x^3 - 12x^2 - 24x = 12x(x+1)(x-2) = 0$ を満たす点は $x = 0, -1, 2$ の 3 個あり，これだけではどの点が最小値を与えるのか分らない．関数値 $f(0) = 3$, $f(-1) = -2$, $f(2) = -29$ を計算してみてはじめて $x = 2$ が答えと判明する．実は，$x = 0$ は極大値，$x = -1$ は最小値でない極小値，そして，$x = 2$ が最小値を与えている．

上の 4 次関数の例で，$x = -1$ は最小値でない極小値を与えていた．これを少し専門的な用語を使って言い表すと，$x = -1$ は局所最適性をもつが大域最適性をもたないということになる．この言葉を正確に定義しておこう．点 x が**大域 (的) 最適** (global optimal) であるとは，

$$\text{任意の } y \text{ に対して } f(x) \le f(y)$$

が成り立つことをいう．また，

$$\text{点 } x \text{ のある近傍 } U \text{ が存在して，} U \text{ の中の任意の } y \text{ に対して } f(x) \le f(y)$$

が成り立つとき，x は**局所 (的) 最適** (local optimal) であるという．大域最適は**最小**，局所最適は**極小**ということと同じである．明らかに最小値は極小値であるから，大域最適性から局所最適性が導かれる．しかし，上の例が示すように，この逆は成り立たない．

関数の局所的な状況は微分などを計算すれば分かるので，局所最適性は計算可能な性質と考えてよい．例えば，上の 4 次関数の例では，2 階導関数 $f''(x) = 36x^2 - 24x - 24$ の符号 $f''(0) = -24 < 0$, $f''(-1) = 36 > 0$, $f''(2) = 72 > 0$ から，$x = 0$ は局所最適でなく，$x = -1$ と $x = 2$ は局所最適であることが判明する．

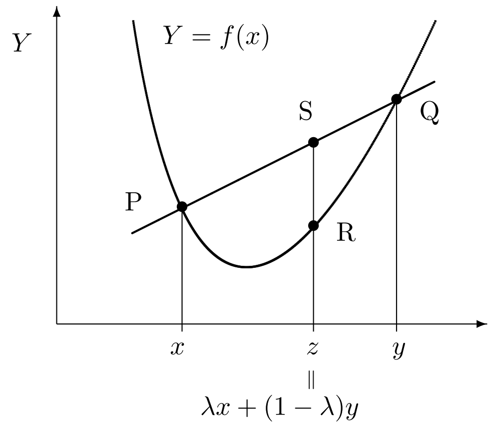

図 1: 凸関数の定義

これとは対照的に，大域的最適性を直接的な計算によって確認するのは難しい．一方，最適化問題において求められているのは大域的な最適解である．このジレンマは最適化問題において本質的な困難であって，これを一般的に解決する術 (すべ) はない．どんな最適化問題を与えられても大域最適解を出すようなアルゴリズムを作ることはまず不可能である．そこで，道は二つに分かれる．

第一の道は，この困難に正面から挑み，できるだけ広い範囲の最適化問題に対して有効なアルゴリズムを開発しようという方向である．最適化問題は現実社会で様々に利用されているので，この方向は応用の観点から極めて重要である．しかし，広い範囲の最適化問題を対象とするのであるから，厳密な意味での大域的最適解を出すことは諦めざるを得ない．理論的な保証はさておき，実際的な立場から見て有用な解が求められるかどうかが主な関心事である．

第二の道は，理論的に扱える美しい世界を広げようとする方向である．上の 2 次関数の例では，幸いなことに局所最適解が大域最適解であった．このような幸いなる状況をできるだけ一般的に捉えようとするとき，中心的な役割をするのが「凸関数」という概念である．これは図 1 のように定義される．すなわち，関数 $f(x)$ のグラフ上の任意の 2 点 $\mathrm{P} = (x, f(x))$, $\mathrm{Q} = (y, f(y))$ に対して，この 2 点に挟まれた範囲で $f(x)$ のグラフが線分 PQ より下にあるとき，関数 $f(x)$ は**凸**であるという (より丁寧に「**下に凸**」ということもある)．そして，凸である関数を**凸関数**という．式を使って書けば，任意の x, y, および $0 \le \lambda \le 1$ を満たす任意の実数 λ に対して不等式

$$\lambda f(x) + (1-\lambda) f(y) \ge f(\lambda x + (1-\lambda) y) \tag{3}$$

が成り立つような関数 f を凸関数と呼ぶということである．$z = \lambda x + (1-\lambda) y$

とおくと，図 1 において，

$$\mathrm{R} = (z, f(z)), \quad \mathrm{S} = (z, \lambda f(x) + (1-\lambda) f(y))$$

であり，不等式 (3) は S が R より上にあることを表している．多変数の場合 (x がベクトルの場合) にも，不等式 (3) によって凸関数の概念が定義される．例に用いた 2 次関数 (1) は凸関数であり，4 次関数 (2) は凸関数でない．

凸関数のいいところは，局所最適性から大域最適性が導かれることである．

定理 1: 凸関数の局所最適値 (極小値) は大域的にも最適 (最小) である．

(証明) 点 x を凸関数 f の局所最小値とすると，x のある近傍 U があって，任意の $z \in U$ に対して $f(z) \geq f(x)$ が成り立つ．任意の点 y を考える．不等式 (3) において $\lambda < 1$ を十分 1 に近くとると $z = \lambda x + (1-\lambda) y \in U$ となるので，

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

これより $f(y) \geq f(x)$ が得られる． ■

補足 2: 関数 f が滑らかな場合には，f の凸性は 2 階導関数の非負性 ($f''(x) \geq 0$) と同値である．多変数の場合には，ヘッセ行列 $\left(\dfrac{\partial^2 f}{\partial x_i \partial x_j} \right)$ が各点で半正定値であることと同値である．なお，一般に，対称行列 A が「任意のベクトル x に対して $x^\top A x \geq 0$」を満たすとき**半正定値**という．また,「任意のベクトル $x \neq \mathbf{0}$ に対して $x^\top A x > 0$」を満たすとき**正定値**という．

(半) 正定値性の判定条件を挙げておこう．対称行列 A の行番号集合 (=列番号集合) を $\{1, \ldots, n\}$ とする．A の部分行列 (小行列) で行番号集合が I，列番号集合が J であるものを $A[I, J]$ と表すとき，$I = J$ である小行列を A の**主小行列**と呼び，その行列式を**主小行列式** (principal minor) と呼ぶ．このとき，

$$A \text{ が半正定値} \iff A \text{ の任意の主小行列式} \geq 0, \tag{4}$$

$$A \text{ が正定値} \iff A \text{ の任意の主小行列式} > 0 \tag{5}$$

が成り立つことが知られている．さらに，行番号集合が $I = \{1, \ldots, k\}$ $(k \leq n)$ の形の主小行列を**首座小行列**と呼び，その行列式を**首座小行列式** (leading principal minor) と呼ぶ (首座小行列の概念は A の行番号のつけ方に依存することに注意)．このとき，

$$A \text{ が正定値} \iff A \text{ の任意の首座小行列式} > 0 \tag{6}$$

も成り立つ．主小行列式は 2^n 個，首座小行列式は n 個あることに注意されたい．なお，正定値性，半正定値性は Gauss の消去法と同様の掃出し演算により，$\mathrm{O}(n^3)$ 回の四則演算で判定できる． ■

以上