

2 最適解

以下, 問題を取り上げるときは, 最小化問題を扱うが最大化問題も同様なことが言える.

f を数ベクトル空間 \mathbb{R}^n 上の関数, C をその部分集合とする.

最小化問題とは以下のような問題 (P) を指す:

$$(P) \quad \begin{array}{ll} \text{最小化} & f(x) \\ \text{制約} & x \in C \end{array}$$

ここで, 「 $x \in C$ 」とは「 x が C に含まれる」ことを意味する. また, 最小化する $f(x)$ を**目的関数**と呼ぶ.

例 5.

$$\begin{array}{ll} \text{最小化} & f(x) = x^2 + ax + b \\ \text{制約} & x \in \mathbb{R} \end{array}$$

目的関数は f である. 最小値を取る点を見つけないので, 式変形をすると

$$f(x) = x^2 + ax + b = \left(x + \frac{a}{2}\right)^2 - \frac{a^2}{4} + b \geq -\frac{a^2}{4} + b$$

となるので, 最小値は $-a^2/4 + b$ で最小解は $x = -a/2$ と分かる.

ここで, 最小解というものを改めて定義しておく.

定義. $\bar{x} \in C$ が任意の $x \in C$ に対して

$$f(x) \geq f(\bar{x})$$

のとき $f(\bar{x})$ を (P) の**大域最小値**, \bar{x} を**大域最小解**と呼ぶ.

大域最小解を見つけられれば最も良いが, それは一般的にに難しい. そこで, より弱い解を探すことを目標とする.

定義. $\bar{x} \in C$ が \bar{x} に**充分近い**任意の $x \in C$ に対して

$$f(x) \geq f(\bar{x})$$

のとき $f(\bar{x})$ を (P) の**局所最小値**, \bar{x} を**局所最小解**と呼ぶ.

最大化問題の場合は不等式が逆向きになり, それぞれ『最小』を『最大』に置き換える. 最小化と最大化をひとまとめにして『最適化』, 『最適値』, 『最適解』と呼ぶ.

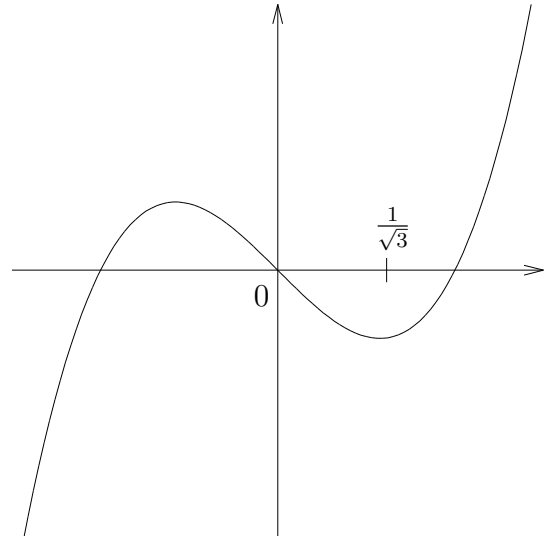
例 6. 例 5 の場合は $x = -a/2$ は大域最小解となる.

一方, 次の最小化問題

$$\text{最小化 } f(x) = x^3 - x$$

$$\text{制約 } x \in \mathbb{R}$$

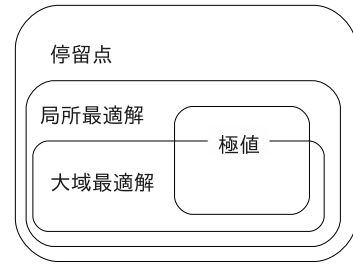
については 大域最小解は存在しないが, 右のグラフから分かるように $x = 1/\sqrt{3}$ に充分近い x ではグラフは下に凸になっているので, そこは局所最小解になっている.



注意

微分積分でつぎのようなことを学ぶ.

\bar{x} に充分近い $x \neq \bar{x}$ に対して, $f(x) > f(\bar{x})$ が成り立つとき, f は \bar{x} で**極小値**をとると言う. また, $f(x) < f(\bar{x})$ が成り立つとき, f は \bar{x} で**極大値**をとると言う. 二つを総称して, **極値**と呼ぶ.



局所最小値は等号も入れた『 $f(x) \geq f(\bar{x})$ 』で定義されるので, **広義の極小値**と同じである. 同様に局所最適値は**広義の極値**と同じである.

局所最小値と極小値は別物であるが, それらを求める手法は結果的にあまり差が無い. 講義では主に局所最小値を扱うが, 必要に応じて極値という言葉も使用する.

例 7 (極小値にならない最小解). $f(x, y) = 4x^2 + 4xy + y^2$ を最小化する問題を考える. $\nabla f(x, y) = (8x + 4y, 4x + 2y)$ なので, 停留点は $2x + y = 0$ を満たす. これより, 停留点はパラメータ t を用いて $x = t$ とすると, $(t, -2t)$ と書ける. この点上で f の値は $f(t, -2t) = 4t^2 - 8t^2 + 4t^2 = 0$ となる. よって, どんなに $(0, 0)$ の近くを考えても, そこに $f(0, 0)$ と同じ値をとる点があるので極小値にはならない. 一方で, $f(x, y) = (2x + y)^2$ なので,

$$f(x, y) \geq 0 = f(t, -2t)$$

が成り立つ. よって, $(t, -2t)$ は大域最小解になる.

3 制約無し最適化問題

一般に最小化問題は制約があるものが多いが、まずは制約の無い最小化問題のから始める:

$$\text{最小化 } f(x)$$

制約無し最適化問題の解析はこれからの理論の基礎になる.

また制約の無い最小化でも重要な応用例は多くある. 例えば, 観測データを一次関数で近似する最小二乗法は制約無し最適化問題である.

3.1 一次最適性必要条件

定理 5 (一次最適性必要条件). \bar{x} が局所最適解ならば,

$$\nabla f(\bar{x}) = 0 \quad (\text{ゼロベクトル})$$

が成り立つ.

定義. $\nabla f(\bar{x}) = 0$ の時, \bar{x} を f の **停留点** と呼ぶ.

3.1.1 1 次最適性必要条件の直感的意味

2 変数関数の場合を説明する. テーラー展開より,

$$f((\bar{x}, \bar{y}) + (u, v)) = f(\bar{x}, \bar{y}) + \nabla f(\bar{x}, \bar{y})(u, v) + o(\|(u, v)\|)$$

となる. ここで, $\nabla f(\bar{x}, \bar{y})(u, v) = J_x u + J_y v$ であった. 充分小さな数 $\delta > 0$ を使って, 特に $(u, v) = (-\delta J_x, -\delta J_y)$ とおくと,

$$f((\bar{x}, \bar{y}) + \delta(u, v)) \approx f(\bar{x}, \bar{y}) - \delta(J_x^2 + J_y^2)$$

となる. (ここで \approx は”左辺は右辺に近い” と言うことを表す)

これより, もし, $\nabla f(\bar{x}, \bar{y}) \neq 0$ なら右辺第二項は負になり, 右辺の値は $f(\bar{x}, \bar{y})$ より小さい. よって, 点 (\bar{x}, \bar{y}) を $(-\delta J_x, -\delta J_y)$ 方向に少し動かせば, J の値が $f(\bar{x}, \bar{y})$ より小さい点が見つかりそうである.

従って, (\bar{x}, \bar{y}) が局所最適解になるには, $\nabla f(\bar{x}, \bar{y}) = (0, 0)$ でないと駄目そうだと予想ができる. また, $\nabla f(\bar{x}, \bar{y}) = (0, 0)$ のとき, (\bar{x}, \bar{y}) の周りでは $f(x, y)$ のグラフは”谷底”, “山の頂点” あるいは”平ら” になっていそうだ (下図参照).

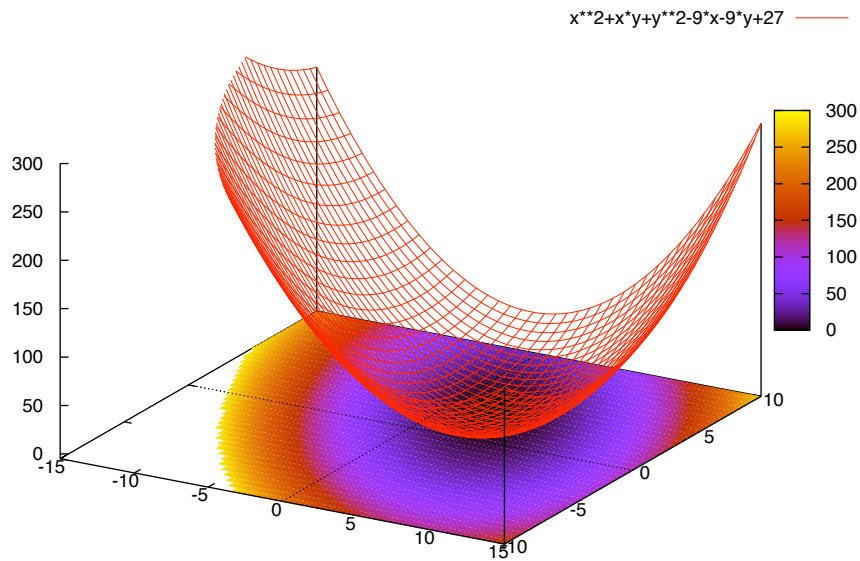


図 1: $x^2 + xy + y^2 - 9x - 9y + 27$ のグラフ. $(x, y) = (3, 3)$ が停留点

3.1.2 定理 5 の証明

簡単のため, 2 変数の場合に定理 5 を証明する. (x, y) のノルムを $\|(x, y)\| = \sqrt{x^2 + y^2}$ とおく. (\bar{x}, \bar{y}) を局所最小解とする. 定義より, (\bar{x}, \bar{y}) に充分近い (x, y) に対して

$$f(x, y) \geq f(\bar{x}, \bar{y})$$

が成り立つ. $\|(u, v)\| = 1$ となるような任意の (u, v) に対して,

$$h(t) = f((\bar{x}, \bar{y}) + t(u, v))$$

とおく. 充分小さい $t > 0$ に対して, $(\bar{x}, \bar{y}) + t(u, v)$ は (\bar{x}, \bar{y}) に近くなるので,

$$h(t) \geq h(0)$$

が成り立つ. ここで, h を 0 でテーラー展開すると,

$$h(t) = h(0) + h'(0)t + o(t)$$

となるので, 上の不等式より $h'(0)t + o(t) \geq 0$ となり, 両辺を t で割ると

$$h'(0) + \frac{o(t)}{t} \geq 0$$

を得る. そこで, t を 0 に近づけると $o(t)/t$ も 0 に近づくので, $h'(0) \geq 0$ となる. 必要ならば t をさらに小さくすれば $h(-t) \geq h(0)$ も成り立つので, $h'(0) \leq 0$ も同様に得る. よって $h'(0) = 0$ が成り立つ.

今, $h'(t) = \frac{d}{dt}f(\bar{x} + tu, \bar{y} + tv) = J_x(\bar{x} + tu, \bar{y} + tv)u + J_y(\bar{x} + tu, \bar{y} + tv)v$ より,

$$h'(0) = J_x(\bar{x}, \bar{y})u + J_y(\bar{x}, \bar{y})v = 0$$

を得る. この等式は任意の $\|(u, v)\| = 1$ を満たす (u, v) に対して成り立つので, $\nabla f(\bar{x}, \bar{y}) = (0, 0)$ が成り立つ.

練習問題 1. 次の関数の停留点を求めよ.

$$(1) f(x, y) = x^2 - xy + 2y^2 - x - 2y \quad (2) f(x, y) = x^3 + y^3 - 3xy$$

3.1.3 停留点が局所最適解か?

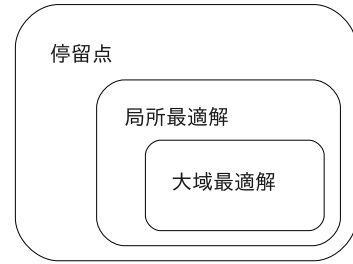
停留点は局所最適解とは限らない.

例 8.

$$\text{最小化 } f(x, y) = x^2 - y^2$$

では $\nabla f(0, 0) = (0, 0)$ となるが, 局所最適解ではない.

しかし, 局所最適解ならば常に停留点になるので最適解の候補となる点を得ることができる.



もちろん最適解が存在しない場合もある.

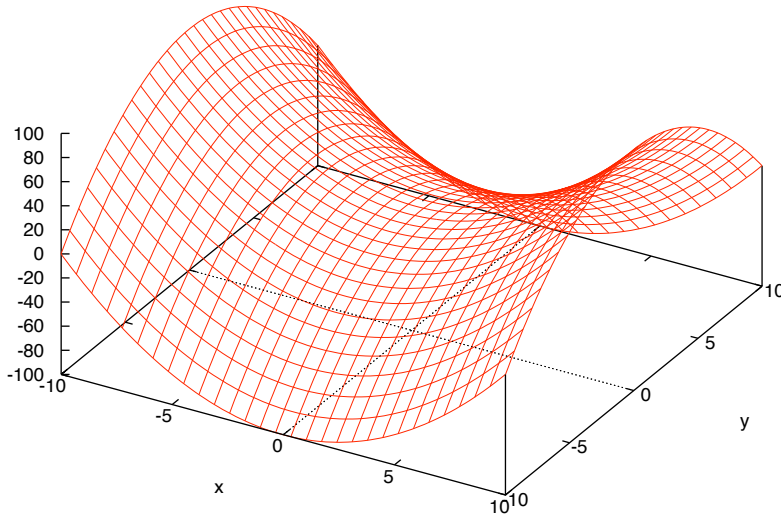


図 2: $x^2 - y^2$ のグラフ

(\bar{x}, \bar{y}) を停留点とする. このとき, (\bar{x}, \bar{y}) を少し動かして (x', y') にしたとき,

- (1). すべての動かし方で, $f(x', y') \geq f(\bar{x}, \bar{y})$ ならば, 局所最小解である
- (2). すべての動かし方で, $f(x', y') \leq f(\bar{x}, \bar{y})$ ならば, 局所最大解である
- (3). 動かし方により, $f(x', y')$ が $f(\bar{x}, \bar{y})$ よりも大きくなったり, 小さくなったりする場合はどちらでもない.

図 3.1.3 の $f(x, y) = x^2 - y^2$ の場合, 停留点 $(0, 0)$ から x だけ動かすと, J の値が大きくなり, y だけ動かすと f の値が小さくなるので, (3) の場合に当てはまる.

停留点の中から局所最適解を探す際に, 地道に $f(x', y')$ の値を調べなければならぬこともあるが, 2 階の導関数を調べることで, 局所最適解を判定できたり, その候補をさらに絞り込むことができる.

3.2 2 次の最適性条件

我々の目標は局所最適解を求めることにある。しかし、1 階微分の情報を使って求まるのは停留点だけで、そのなかには局所最適解になっていないものもある。(例 $f(x) = x^3$ で、0 は停留点だが局所最適ではない) このままでは情報が足りないので、2 階微分の情報を使う。 n 変数関数 f を最小化する問題を考える。

局所最適解を求めたい
 → 1 階微分を使って停留点を求める
 まだわからない
 → 2 階微分を使う

定理 6 (2 次の最適性条件).

(必要性) \bar{x} が局所最小解である $\Rightarrow \nabla f(\bar{x}) = 0$ かつ $\nabla^2 f(\bar{x})$ が半正定値

(十分性) $\nabla f(\bar{x}) = 0$ かつ $\nabla^2 f(\bar{x})$ が正定値 $\Rightarrow \bar{x}$ は局所最小解

局所最大解についても、それぞれ対応する箇所を半負定値、負定値に置き換えたものが成り立つ。

(否定) $\nabla^2 f(\bar{x})$ が不定値のとき、 \bar{x} は局所最適解ではない。

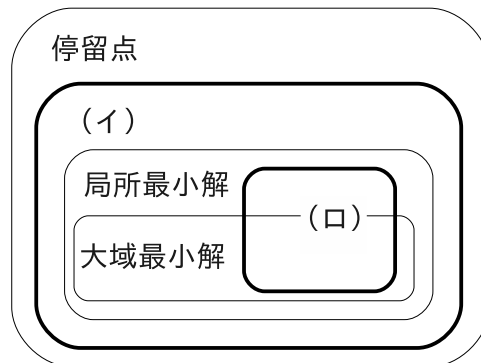
解説. 半正定値行列の定義を見たときに非常に都合の良い条件に見えるかもしれないが、実は関数が局所最小値をとっていれば、自然に現れる性質である。

なので、十分性にある正定値行列の条件も非常に強い条件に見えるが、あながちそうとは言えないのだ。

定理の証明は後ろの節で行う。

□

最適性と 2 次の条件の関係



(イ) $\nabla f(\bar{x}, \bar{y}) = (0, 0)$ かつ $\nabla^2 f(\bar{x}, \bar{y})$ が半正定値

(ロ) $\nabla f(\bar{x}, \bar{y}) = (0, 0)$ かつ $\nabla^2 f(\bar{x}, \bar{y})$ が正定値

図の包含関係が示すように、 \bar{x} が条件 (ロ) を満たさなくても最適解になることがある。実際、 $f(x, y) = x^4 + y^4$ の $(0, 0)$ でのヘッセ行列はゼロ行列になるので条件 (ロ) は満たさない。しかし、任意の (x, y) で $f(x, y) \geq 0$ なので、 $(0, 0)$ は大域最小解になっている。

3.2.1 2 次の最適性条件の幾何的イメージ

凸関数の概念を使うと、2 次の最適性条件に幾何的なイメージをつけられる。すべての (x, y) に対して、 $\nabla^2 f(x, y)$ が正定値であると f は凸関数になるのであった。

いま、 $\nabla^2 f(\bar{x}, \bar{y})$ が正定値で、 (x, y) が (\bar{x}, \bar{y}) に十分近い点であると仮定する。すると $f_{xx}(x, y), f_{xy}(x, y), f_{yy}(x, y)$ の値も $f_{xx}(\bar{x}, \bar{y}), f_{xy}(\bar{x}, \bar{y}), f_{yy}(\bar{x}, \bar{y})$ に近い。したがって、

$$\nabla^2 f(\bar{x}, \bar{y}) \text{ が正定値} \Rightarrow \nabla^2 f(x, y) \text{ が正定値} \Rightarrow f \text{ が } \llbracket (\bar{x}, \bar{y}) \text{ の近くで凸} \rrbracket$$

従って、 $\nabla^2 f(\bar{x}, \bar{y})$ が正定値であり、さらに点 (\bar{x}, \bar{y}) が停留点ならば、その点は関数 f の「局所的に凸」な部分の底にあるということを表している。

3.3 最適性と凸性

2 階微分の情報を使って 2 次の最適性条件を調べても、ある点 (\bar{x}, \bar{y}) が局所最適解であることしかわからない。それでは大域最適解を見つけるにはどうすれば良いだろうか？ 実は関数が凸であるときは大域最適解を簡単に見つけられる。

定理 7. $f: \mathbb{R}^n \rightarrow \mathbb{R}$ を凸関数とする。すると f の停留点は大域最小解になる。

Proof. 停留点を \bar{x} とする。 f は凸なので、任意の $y \in \mathbb{R}^n$ に対して、 $f(y) \geq f(\bar{x}) + \nabla f(\bar{x})(y - \bar{x})$ が成り立つ。いま $\nabla f(\bar{x}) = 0$ なので、 $f(y) \geq f(\bar{x})$ が成り立つ。これは \bar{x} が f の大域最小解であることを意味する。

□

3.4 局所最適解の求め方

それでは 2 次の最適性条件を用いて例題を解いてみよう。

例 9.

$$\text{最小化 } f(x, y) = x^2 + xy + y^2 - 9x - 9y + 27$$

の局所最適値を求めよ。

まず、局所最適解であれば停留点になっているので、方程式

$$\nabla f(x, y) = (2x + y - 9, x + 2y - 9) = (0, 0)$$

を解く。すると停留点は $(x, y) = (3, 3)$ だと分かる。この点に対して 2 次の最適性条件を調べてみよう。

ヘッセ行列を計算すると

$$\nabla^2 f(x, y) = \begin{bmatrix} 2 & 1 \\ 1 & 2 \end{bmatrix}$$

となる。固有値を求めると、1, 3 となるので、ヘッセ行列は正定値である。よって、 $(3, 3)$ は局所最小解であり、局所最小値は 0 となる。グラフの概形は図 3.1.1 を参照。

例 10. $f(x, y) = x^3 - 3xy + y^3$

$$\nabla f(x, y) = (3x^2 - 3y, 3y^2 - 3x), \nabla^2 f(x, y) = \begin{bmatrix} 6x & -3 \\ -3 & 6y \end{bmatrix}$$

まず停留点を求める. $\nabla f(x, y) = (0, 0)$ を書き下すと, 連立方程式

$$\begin{cases} 3x^2 - 3y = 0 \\ 3y^2 - 3x = 0 \end{cases}$$

をえる. この解は $(x, y) = (0, 0), (1, 1)$ になり, これが停留点になる.

次に, 2 次の最適性条件を調べる.

$(0, 0)$ ではヘッセ行列の行列式 $|\nabla^2 f(0, 0)|$ は $|\nabla^2 f(0, 0)| < 0$ を満たすので, ヘッセ行列は不定値. よってこの点では局所最適ではない.

$(1, 1)$ では $f_{xx}(1, 1) = 6 > 0$ かつ $|\nabla^2 f(1, 1)| = 27 > 0$ なので, ヘッセ行列は正定値. よってこの点は局所最小解で, 局所最小値は -1 である.

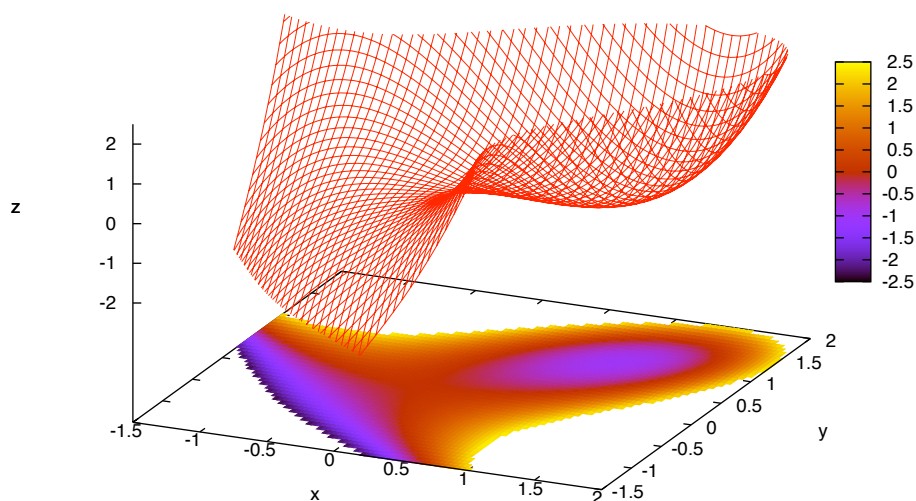


図 3: $z = x^3 - 3xy + y^3$ のグラフ

例 11.

$$\text{最小化 } f(x, y) = x^4 + y^4 - (x + y)^2$$

の局所最適解を探し, そこの値を求める.

まず, 局所最適解であれば停留点になっているので, 方程式

$$\nabla f(x, y) = (4x^3 - 2(x + y), 4y^3 - 2(x + y)) = (0, 0)$$

を解く. すると停留点は $(x, y) = (0, 0), (1, 1), (-1, -1)$ だと分かる. この点に対して 2 次の最適性条件を調べてみよう.

ヘッセ行列を計算すると

$$\nabla^2 f(x, y) = \begin{bmatrix} 12x^2 - 2 & -2 \\ -2 & 12y^2 - 2 \end{bmatrix}$$

となる.

(i) 点 $(1, 1), (-1, -1)$ で 2 次の条件を調べる.

関数 f のヘッセ行列は

$$\nabla^2 f(1, 1) = \nabla^2 f(-1, -1) = \begin{bmatrix} 10 & -2 \\ -2 & 10 \end{bmatrix}$$

となる. 命題 4 を適用すると, この行列の 1 行 1 列の要素は 10 なので正. さらに行列式は $10^2 - (-2)^2 > 0$ なので, ヘッセ行列は $(1, 1), (-1, -1)$ で正定値になる. よって, 2 次の最適性条件より, この二つの点は局所最小解になり, その点での値は, $f(1, 1) = f(-1, -1) = -2$

補足. $\nabla^2 f(1, 1)$ の固有値は $\begin{vmatrix} 10 - \lambda & -2 \\ -2 & 10 - \lambda \end{vmatrix} = 0$ を満たす λ であり, $\lambda = 8, 12$ となる. よって固有値はすべて正であることが確かめられる.

(ii) 点 $(0, 0)$ で 2 次の条件を調べる.

関数 f のヘッセ行列は,

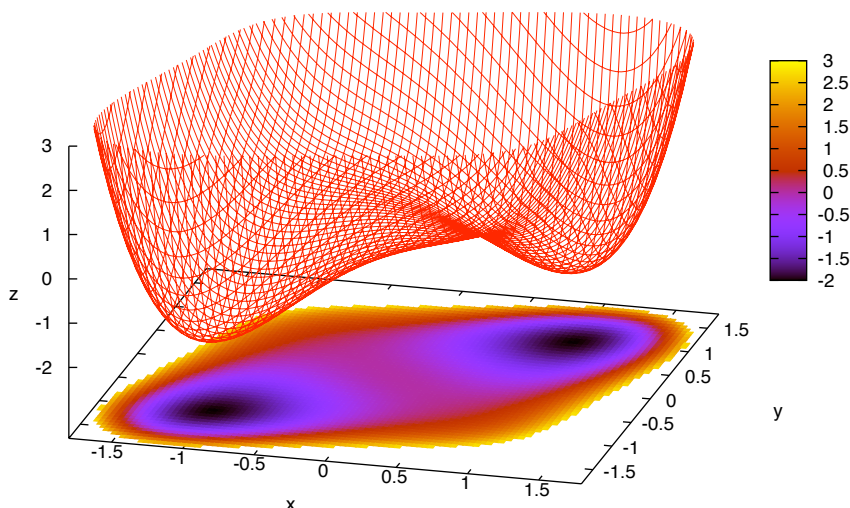
$$\nabla^2 f(0, 0) = \begin{bmatrix} -2 & -2 \\ -2 & -2 \end{bmatrix}$$

となり, その行列式は 0 になり命題 4 では最適性の判定ができない. 固有値を計算してみると 0, 4 となる. よってヘッセ行列は点 $(0, 0)$ で半正定値になる. この点では 2 次の最適性必要条件が成り立っているが, まだ局所最適解がそうでないか判断がつかない.

そこで地道に調べてみる. $y = 0$ として x 軸上のみで f の値の変化を調べる. $f(x, 0) = x^4 - x^2 = x^2(x^2 - 1)$ より充分小さい x で f は負の値をとる. 一方, $y = -x$ という関係を満たす点での f の値の変化を調べると, $f(x, -x) = x^4$ となり原点以外の点で正である. よって局所最小でも最大でもない.

まとめると, f は $(1, 1), (-1, -1)$ で局所最小値 -2 をとる. なお $(0, 0)$ は停留点だが局所最適解ではない. □

さて、いま解析した $f(x, y) = x^4 + y^4 - (x + y)^2$ の概形は下の図のようになる。 x, y どちらに関しても、どんどん大きな数を代入すれば $f(x, y)$ の値がどんどん大きくなるのがわかる。また $f(x, 0)$ と同様に $f(0, y)$ を調べると、 y が 0 に近いとき、 f の値が負になることがわかる。一方、上で求めたように、 f の停留点は 3 つしかなく、 $(1, 1), (-1, -1)$ では f は -2 まで窪んでいることもわかっている。



グラフを見ると確かに $(1, 1), (-1, -1)$ は関数の局所的に凸な部分の底にあることがわかる。

練習問題 2. 以下の関数の局所最適解と局所最適値を求めよ。

(1) $f(x, y) = x^3 - 3xy + y^3$ (2) $f(x, y) = 3x^2 + 2xy + y^2$ (3) $f(x, y) = x^3 + y^3 - 9xy + 1$

4 付録

4.1 定理 6 の証明

証明. (必要性) (\bar{x}, \bar{y}) を局所最小解とする. 任意の $(u, v) \in \mathbb{R}^2$ に対して、 $h(t) = f((\bar{x}, \bar{y}) + (u, v))$ とおく. すると充分小さな数 $t > 0$ に対して、 $h(t) \geq h(0)$ が成り立つ. いま、0 に関するテーラー展開より、

$$h(t) = h(0) + h'(0)t + h''(0)t^2 + o(t^2)$$

となる. さらに、1 次の最適性条件より、 $h'(0) = \nabla f(\bar{x}, \bar{y}) \cdot (u, v) = 0$ なので、

$$h''(0)t^2 + o(t^2) \geq 0$$

を得る. ここで, 両辺を t^2 で割り, $t \rightarrow 0$ とすると, $o(t^2)/t^2 \rightarrow 0$ となるので, $h''(0) \geq 0$ が成り立つ. いま, $h''(0) = [u \ v] \nabla^2 f(\bar{x}, \bar{y}) \begin{bmatrix} u \\ v \end{bmatrix}$ であり, $h''(0) \geq 0$ は任意の (u, v) に対して成り立つので, $\nabla^2 f(\bar{x}, \bar{y})$ は半正定値になる.

(十分性) (\bar{x}, \bar{y}) が停留点で, $\nabla^2 f(\bar{x}, \bar{y})$ を正定値とし, その最小固有値を λ とする. ここで, 定理 3 より, $\lambda > 0$ となる. 2 変数に対するテーラー展開より, 小さな数 ε で $0 < \varepsilon < \lambda/2$ を満たすものに対しても, $(0, 0)$ に充分近い (u, v) ならば,

$$\left| f((\bar{x}, \bar{y}) + (u, v)) - \left\{ f(\bar{x}, \bar{y}) + \nabla f(\bar{x}, \bar{y}) \cdot (u, v) + \frac{1}{2} [u \ v] \nabla^2 f(\bar{x}, \bar{y}) \begin{bmatrix} u \\ v \end{bmatrix} \right\} \right| < \varepsilon \|(u, v)\|^2$$

が成り立つ (左辺は絶対値). さらに (\bar{x}, \bar{y}) が停留点であることから, 特に

$$f((\bar{x}, \bar{y}) + (u, v)) > f(\bar{x}, \bar{y}) + \frac{1}{2} [u \ v] \nabla^2 f(\bar{x}, \bar{y}) \begin{bmatrix} u \\ v \end{bmatrix} - \varepsilon \|(u, v)\|^2$$

を得る. ここで, $t = \|(u, v)\|$ とおくと, $\|t^{-1}(u, v)\| = 1$ なので, 定理 ?? を適用すると,

$$[u \ v] \nabla^2 f(\bar{x}, \bar{y}) \begin{bmatrix} u \\ v \end{bmatrix} \geq \lambda \|(u, v)\|^2$$

が成り立つことがわかる. よって上式と ε の選び方より,

$$f((\bar{x}, \bar{y}) + (u, v)) > f(\bar{x}, \bar{y}) + \left(\frac{1}{2} \lambda - \varepsilon \right) \|(u, v)\|^2 > f(\bar{x}, \bar{y})$$

を得る. いま (u, v) は $(0, 0)$ に充分近い任意の点なので, これは (\bar{x}, \bar{y}) が局所最小解であることを表す.

(否定) (\bar{x}, \bar{y}) が局所最適解か局所最大解ならば, ヘッセ行列 $\nabla^2 f(\bar{x}, \bar{y})$ は半正定値か半負定値になるので, ヘッセ行列が不定値の場合は, そのどちらにもなっていない (定理 6 の図も参照). □