

最適化数学 講義ノート 2 (担当: 関口 良行)

1 復習と一次の最適性条件の意味

前回, 以下の問題を考えた:

$$(P) \quad \text{最小化 } J(x, y)$$

これに対し, 次のような定理を得た.

定理 1 (一次の最適性必要条件). (\bar{x}, \bar{y}) が局所最小解, または局所最大解のとき, $\nabla J(\bar{x}, \bar{y}) = 0$ となる.

$\nabla J(x, y) = (J_x(x, y), J_y(x, y))$ は勾配ベクトルと呼ばれる. $(\bar{x}, \bar{y}) \in R^2$ で, $\nabla J(\bar{x}, \bar{y}) = (0, 0)$ を満たすものを停留点と呼ぶ. まず, 上の定理の直感的な意味を考えよう.

1.1 1 次最適性必要条件の直感的意味

内積を用いて, $J_x(\bar{x}, \bar{y})u + J_y(\bar{x}, \bar{y})v = \nabla J(\bar{x}, \bar{y}) \cdot (u, v)$ と書ける. これを使うとテーラー展開より,

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

と書ける. ここで, 充分小さな数 $\delta > 0$ を使って, 特に $(\bar{u}, \bar{v}) = -\delta(J_x, J_y)$ とおくと,

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

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, $\nabla J(\bar{x}, \bar{y}) \neq 0$ なら右辺第二項は負になり, 右辺の値は $J(\bar{x}, \bar{y})$ より小さい. よって, 点 (\bar{x}, \bar{y}) を (\bar{u}, \bar{v}) 方向に少し動かせば, J の値が $J(\bar{x}, \bar{y})$ 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, $\nabla J(\bar{x}, \bar{y}) = (0, 0)$ でないと駄目そうだと予想ができる. また, $\nabla J(\bar{x}, \bar{y}) = (0, 0)$ のとき, (\bar{x}, \bar{y}) では $z = J(x, y)$ という関数は”平ら” になっていそうだ.

2 2 次の情報

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 $J(x) = x^3$ 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.

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

定義.

$$\nabla^2 J(\bar{x}, \bar{y}) = \begin{bmatrix} J_{xx}(\bar{x}, \bar{y}) & J_{yx}(\bar{x}, \bar{y}) \\ J_{xy}(\bar{x}, \bar{y}) & J_{yy}(\bar{x}, \bar{y}) \end{bmatrix}$$

を J のヘッセ行列と言う. ここで, J が充分滑らかな時, $J_{xy} = J_{yx}$ なので, ヘッセ行列は対称行列になる.

練習問題 1. 以下の関数の勾配ベクトル, ヘッセ行列を求めよ.

$$(1) f(x, y) = x^4 + y^4 - 2x^2 - 2y^2 + 4xy \quad (2) f(x, y) = e^{-(x^2+y^2)}$$

2.1 正值性

ここで, 2 次の情報を用いた最適性条件のために, 行列の言葉を用意する.

定義. A を実数成分を持つ対称行列とする.

- すべての $(x, y) \in R^2$ に対して,

$$\begin{bmatrix} x & y \end{bmatrix} A \begin{bmatrix} x \\ y \end{bmatrix} \geq 0$$

のとき, A を半正定値行列と呼ぶ. また, 逆の不等号が成り立つとき, A を半負定値行列と呼ぶ.

- $(x, y) \neq (0, 0)$ を満たすすべての $(x, y) \in R^2$ に対して,

$$\begin{bmatrix} x & y \end{bmatrix} A \begin{bmatrix} x \\ y \end{bmatrix} > 0$$

のとき, A を正定値行列と呼ぶ. 逆の不等号が成り立つとき, A を負定値行列と呼ぶ.

- $(x, y) \in R^2$ によって

$$\begin{bmatrix} x & y \end{bmatrix} A \begin{bmatrix} x \\ y \end{bmatrix}$$

の符号が正にも負にもなるとき, A を不定値行列と呼ぶ.

これらの言葉を用いると以下のような関係が成り立つ.

定理 2 (2 次の最適性条件).

(必要性) (\bar{x}, \bar{y}) が問題 (P) の局所最小解である

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

(十分性) $\nabla J(\bar{x}, \bar{y}) = (0, 0)$ かつ $\nabla^2 J(\bar{x}, \bar{y})$ が正定値

$\Rightarrow (\bar{x}, \bar{y})$ は問題 (P) の局所最小解

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) $\nabla^2 J(\bar{x}, \bar{y})$ が不定値のとき, (\bar{x}, \bar{y}) は局所最適解ではない.

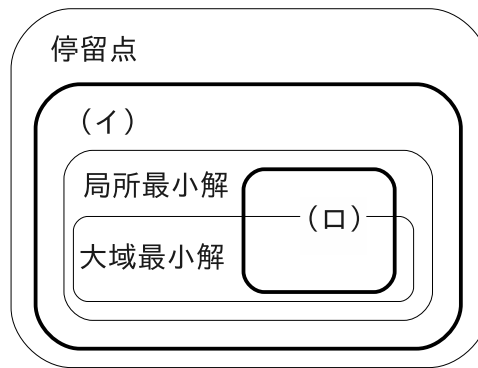
解説. 半正定値行列の定義を見たときに非常に都合の良い条件に見えるかもしれないが, 実は関数が局所最小値をとっていれば, 自然に現れる性質である.

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

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

図の包含関係が示すように, (\bar{x}, \bar{y}) 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, $J(x, y) = x^4 + y^4$ 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 $J(x, y) \geq 0$ なので, $(0, 0)$ は大域最小解になっている.

2.2 幾何的イメージ

2 次の最適性条件に対して幾何的イメージをつけるために, 凸関数という概念を導入する. 数ベクトルに対して, $a = (x, y)$ などとおく.

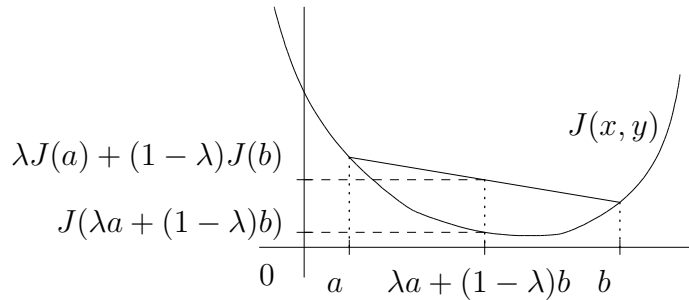
2.2.1 凸関数

定義 (凸関数). $0 \leq \lambda \leq 1$ を満たす任意の λ と任意の $a, b \in R^2$ に対して,

$$J(\lambda a + (1 - \lambda)b) \leq \lambda J(a) + (1 - \lambda)J(b)$$

が成り立つとき, J を凸関数と呼ぶ.

定義中, $\lambda a + (1 - \lambda)b$ は a, b を結ぶ線分の内分点を表している. 下に凸関数の概観を示す. 関数 J のグラフを a, b を結ぶ直線で切って, その直線を横軸にとってやる. すると, 上の不等式が成り立っているということは, a, b を結ぶ線分の内分点 $\lambda a + (1 - \lambda)b$ での関数値が, 同じ割合で $J(a)$ と $J(b)$ を内分した値よりも常に小さいか等しくなっていることを表す. または, 下図のように, 点 $(a, J(a))$ と $(b, J(b))$ を結んだ線分が常に J のグラフの上にあることを表す.



例 1. 1 変数の場合だが, $J(x) = x^2$ は凸関数である. グラフから直ちに分かるが, 定義に戻ればきちんと示すことができる.

定義によると, J が凸関数であることを示すには, $\lambda J(a) + (1 - \lambda)J(b) - \{J(\lambda a + (1 - \lambda)b)\} \geq 0$ を示せばよい. 式変形をすると

$$\begin{aligned} & \{\lambda a^2 + (1 - \lambda)b^2\} - \{\lambda a + (1 - \lambda)b\}^2 \\ &= \lambda a^2 + (1 - \lambda)b^2 - \lambda^2 a^2 - 2\lambda(1 - \lambda)ab - (1 - \lambda)^2 b^2 \\ &= \lambda a^2(1 - \lambda) + (1 - \lambda)b^2\{1 - (1 - \lambda)\} - 2\lambda(1 - \lambda)ab \\ &= \lambda(1 - \lambda)(a^2 + b^2 - 2ab) = \lambda(1 - \lambda)(a - b)^2 \geq 0. \end{aligned}$$

よって, J は凸関数である. この証明から $J(x, y) = x^2 + y^2$ も凸関数であることが類推できるであろう.

2.2.2 1 変数凸関数の性質

証明は省略するが, グラフの概形から 1 変数の場合以下のような性質が成り立つことがわかるだろう. h を 1 変数関数とする.

h が凸関数

$$\Leftrightarrow \text{すべての } t \in R \text{ で, } h'(t) \text{ が非減少 } (t_1 < t_2 \Rightarrow h'(t_1) \leq h'(t_2))$$

$$\Leftrightarrow \text{すべての } t \in R \text{ で, } h''(t) \geq 0$$

例 1 の関数 $J(x) = x^2$ では実際に $J'(x) = 2x$, $J''(x) = 2$ なので, 上記の性質を満たすことがわかる.

2.2.3 2 次の最適性条件のイメージ

凸関数の概念を使うと、2 次の最適性条件に以下のようなイメージをつけられる。 $(u, v) \in R^2$ に対して、 $h(t) = J((\bar{x}, \bar{y}) + t(u, v))$ とおくと

$$\begin{aligned}h'(t) &= J_x(\bar{x} + tu, \bar{y} + tv)u + J_y(\bar{x} + tu, \bar{y} + tv)v \\h''(0) &= J_{xx}uu + J_{xy}uv + J_{yx}vu + J_{yy}vv \\&= \begin{bmatrix} u & v \end{bmatrix} \begin{bmatrix} J_{xx}(\bar{x}, \bar{y}) & J_{yx}(\bar{x}, \bar{y}) \\ J_{xy}(\bar{x}, \bar{y}) & J_{yy}(\bar{x}, \bar{y}) \end{bmatrix} \begin{bmatrix} u \\ v \end{bmatrix}.\end{aligned}$$

となる。よって、

$\nabla^2 J(\bar{x}, \bar{y})$ が正定値

$\Leftrightarrow (u, v) \neq (0, 0)$ を満たす任意の $(u, v) \in R^2$ に対して、 $h''(0) > 0$

$\Rightarrow (u, v) \neq (0, 0)$ を満たす任意の $(u, v) \in R^2$ に対して、 h が「0 の近くで凸」

$\Rightarrow J$ が「 (\bar{x}, \bar{y}) の近くで凸」

従って、 $\nabla^2 J(\bar{x}, \bar{y})$ が正定値であり、点 (\bar{x}, \bar{y}) が停留点ならば、その点は関数 J 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。

2.3 正值性の調べ方

さて、2 次の最適性条件より、ある点が関数の停留点であり、そこでのヘッセ行列が正定値であれば、その点は関数の局所最小解になることがわかった。それでは、行列が正定値であるとはどのように調べたらよいただろうか？

2.3.1 固有値を用いた判定法

それには行列の理論による以下のような定理が有用である。ある数 a が非負であるとは、 $a \geq 0$ であることを表し、ある数が正であるとは $a > 0$ であることを表す。

定理 3. A を実対称行列とする。以下の主張が成り立つ：

- (1). A が半正定値 (半負定値) $\Leftrightarrow A$ の固有値がすべて非負 (非正)
- (2). A が正定値 (負定値) $\Leftrightarrow A$ の固有値がすべて正 (負)
- (3). A が不定値 $\Leftrightarrow A$ が正と負の固有値を持つ

解説. λ を A の固有値, (u, v) をその固有ベクトルで $\|(u, v)\| = 1$ となるものとする. すると

$$\begin{bmatrix} u & v \end{bmatrix} A \begin{bmatrix} u \\ v \end{bmatrix} = \begin{bmatrix} u & v \end{bmatrix} \lambda \begin{bmatrix} u \\ v \end{bmatrix} = \lambda(u^2 + v^2) = \lambda$$

となる. よって $\begin{bmatrix} x & y \end{bmatrix} A \begin{bmatrix} x \\ y \end{bmatrix}$ の値は, 少なくとも A の固有値と同じ値をとることがわかる. これは定理の主張 (3) と (1), (2) の右向きの矢印「 \Rightarrow 」を証明している.

(1), (2) の左向きの矢印「 \Leftarrow 」はその値の範囲に関する問題になる. これは以下の定理によって示される. \square

定理 4. 実対称行列 A に対して, 最適化問題

$$\begin{aligned} & \text{最小化 } \begin{bmatrix} x & y \end{bmatrix} A \begin{bmatrix} x \\ y \end{bmatrix} \\ & \text{制約 } \|(x, y)\| = 1 \end{aligned}$$

の最小値は, 行列 A の固有値で最小のものに等しい. また, 最大値は A の固有値で最大ものに等しい.

解説. 後に最適化数学を用いて示す. \square

2.3.2 行列式を用いた判定法

2×2 行列であれば, 以下のように行列式を用いて判定できることがわかる.

実対称行列を $A = \begin{bmatrix} a & b \\ b & d \end{bmatrix}$ とし, $a \neq 0$ とする. 任意の $(x, y) \in \mathbb{R}^2$ に対して,

$$\begin{bmatrix} x & y \end{bmatrix} A \begin{bmatrix} x \\ y \end{bmatrix} = ax^2 + 2bxy + dy^2 = a \left(x + \frac{by}{a} \right)^2 + \frac{y^2}{a} (ad - b^2) = a \left(x + \frac{by}{a} \right)^2 + \frac{y^2}{a} \begin{vmatrix} a & b \\ b & d \end{vmatrix}$$

となる ($\begin{vmatrix} a & b \\ b & d \end{vmatrix}$ は A の行列式 $|A|$ を表す). この式より,

命題 5.

(1). $a > 0, |A| > 0 \Rightarrow A$ は正定値

(2). $a < 0, |A| > 0 \Rightarrow A$ は負定値

(3). $|A| < 0 \Rightarrow A$ は不定値

解説. (1), (2) は命題の前の式から証明できる. (3) については, 線形代数で学んだように, 行列式の値は固有値の積と等しい (λ_1, λ_2 を固有値とすると, $|A| = \lambda_1 \lambda_2$). よって, 2×2 行列の場合, 行列式が負ということは, 二つある固有値の符号が異なるということである. \square

2.4 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.

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

の局所最適解を探し, そこでの値を求める.

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

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

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

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

となる.

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

関数 J のヘッセ行列は

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

となる. 命題 5 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, $J(1, 1) = J(-1, -1) = -2$

補足. $\nabla^2 J(1, 1)$ 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 J のヘッセ行列は,

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

となり, その行列式は 0 になり命題 5 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 J の値の変化を調べる。
 $J(x, 0) = x^4 - x^2 = x^2(x^2 - 1)$ より充分小さい x で J は負の値をとる。一方、
 $y = -x$ という関係を満たす点での J の値の変化を調べると、 $J(x, -x) = x^4$
 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、 J 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。 \square

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

このヒントから J の概形を想像してみてほしい。正解のグラフは最後に載せておく。

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。

$$(1) J(x, y) = x^3 - 3xy + y^3 \quad (2) J(x, y) = 3x^2 + 2xy + y^2 \quad (3) J(x, y) = x^3 + y^3 - 9xy + 1$$

3 付録

3.1 定理 2 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、 $h(t) = J((\bar{x}, \bar{y}) + (u, v))$ 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、 $h'(0) = \nabla J(\bar{x}, \bar{y}) \cdot (u, v) = 0$ 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、 $h''(0) = [u \ v] \nabla^2 J(\bar{x}, \bar{y}) \begin{bmatrix} u \\ v \end{bmatrix}$ であり、 $h''(0) \geq 0$ は任意の (u, v) に対して成り立つので、 $\nabla^2 J(\bar{x}, \bar{y})$ は半正定値になる。

(十分性) (\bar{x}, \bar{y}) が停留点で, $\nabla^2 J(\bar{x}, \bar{y})$ 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,

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

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

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

を得る. ここで, $t = \|(u, v)\|$ とおくと, $\|t^{-1}(u, v)\| = 1$ なので, 定理 4 を適用すると,

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

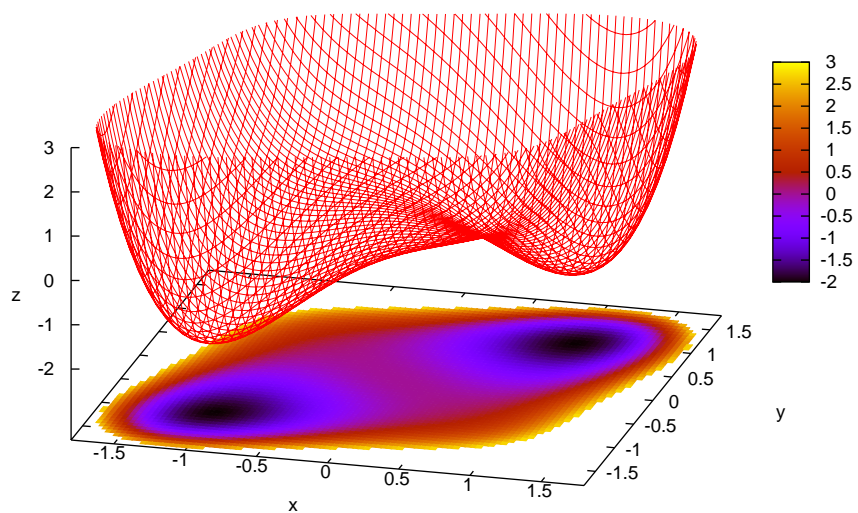
が成り立つことがわかる. よって上式と ε の選び方より,

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

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 $\nabla^2 J(\bar{x}, \bar{y})$ 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 2 の図も参照). □

3.2 $x^4 + y^4 - (x + y)^2$ のグラフ (gnuplot 使用)



節 2.2.3 で解説したように, 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.