第2章　凸関数

2.1　凸関数の性質

2.1.1　凸関数の定義

この本では，関数は十分な回数微分可能であるとする．2 次関数

$$f(x) = x^2 + bx + c$$

は最小値を簡単に求めることができる.．一般に 2 次関数のように下に凸な関数は，最小値を求めることが比較的楽である．

下に凸な関数は，最適化問題においては単に凸関数と呼ばれ，一般の n 変数関数に対して以下のように定義される．

定義．f を n 変数関数とする．$0 < \lambda < 1$ を満たすすべての λ とすべての $u, v \in \mathbb{R}^n$ に対して，

$$f((1-\lambda)u + \lambda v) \leq (1-\lambda)f(u) + \lambda f(v) \tag{2.1}$$

が成り立つとき，f を 凸関数 と呼ぶ．等号が成り立つのが $u = v$ のときに限るとき f を 狭義凸関数 と呼ぶ．

$(-1) \cdot f$ が凸関数のとき，f を 凹関数 と呼ぶ．狭義凹関数も同様である．

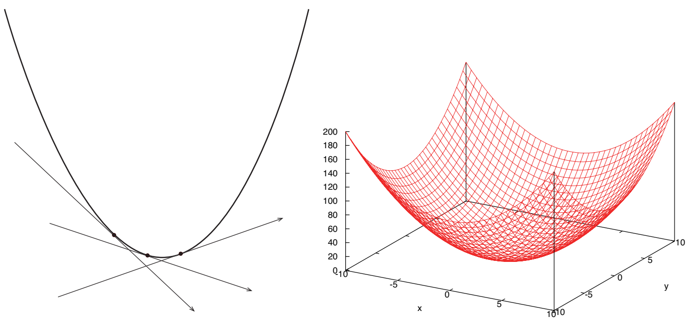


図 2.1: 凸関数の例 左が 1 変数凸関数，右が 2 変数凸関数

解説．$(1-\lambda)u + \lambda v$ は u, v の内分点，$(1-\lambda)f(u) + \lambda f(v)$ は $f(u), f(v)$ の内分点を表している．これより，不等式 (2.1) は，

$$
\begin{gathered}
\text{「}f\text{ が凸関数」} \\
\Updownarrow \\
\text{「点 }(u, f(u))\text{ と }(v, f(v))\text{ を結んだ線分が，} \\
f\text{ のグラフ上，又はグラフより上部にある」}
\end{gathered}
$$

ということを表す．

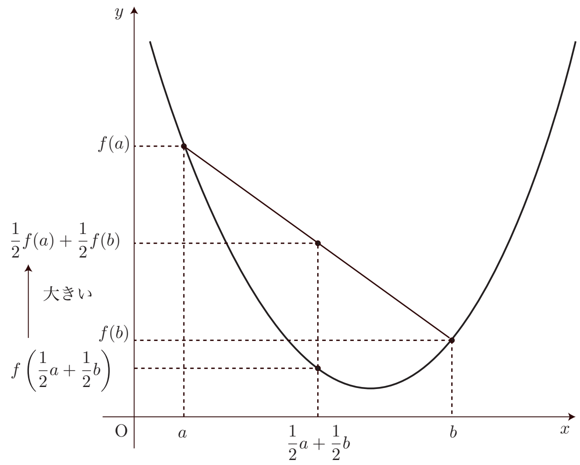


図 2.2: 1 変数凸関数に対して，定義式で $\lambda = \frac{1}{2}$ とした場合

一方，狭義凸関数の定義は

$$
\begin{gathered}
\text{「}f\text{ が狭義凸関数」} \\
\Updownarrow \\
\text{「点 }(u, f(u))\text{ と }(v, f(v))\text{ を結んだ線分が，}f\text{ グラフより上部にある」}
\end{gathered}
$$

ということを表している．

補足． 図 2.2 を参考にすると，一変数関数ならば，増減表を書いてグラフが下に突き出ていれば，凸関数の定義式が自然に成り立つ．

例 2. 凸関数だが，狭義凸関数でない関数の例をあげる．

$$f(x, y) = x + y^2$$

とすると，図 2.3 のグラフより f は凸関数である．ここで，$u = (1, 0)$，$v = (2, 0)$ に対して，

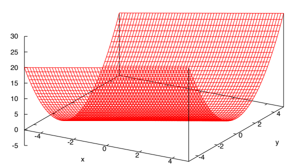


図 2.3: $z = x + y^2$

$$(1-\lambda)u + \lambda v = (1-\lambda)\begin{bmatrix} 1 \\ 0 \end{bmatrix} + \lambda \begin{bmatrix} 2 \\ 0 \end{bmatrix} = \begin{bmatrix} 1+\lambda \\ 0 \end{bmatrix}$$

なので，定義式の左辺は

$$f((1-\lambda)u+\lambda v)=f(1+\lambda,0)=1+\lambda$$

となり，右辺は

$$(1-\lambda)f(u)+\lambda f(u)=(1-\lambda)\cdot 1+\lambda\cdot 2=1+\lambda$$

となる．よって，$u\neq v$ かつ定義式で等号が成り立つので，f は狭義凸関数ではない．
図 2.3 にある f のグラフを見ると $u=(u_1,0),v=(v_1,0)$ となるような点に対して，$(u,f(u))$ と $(v,f(v))$ を結ぶ線分はちょうどグラフ上にある（グラフより上部にはない）．
形でいうと，狭義凸関数はグラフがすべての方向でちゃんと丸く突き出ているような関数である．

2.1.2　凸関数と接平面

次に，凸関数と接線の関係を考えてみよう．図 2.1 の左を見れば分かるように，一変数凸関数に接線を引くと，接線はグラフの下側に位置しグラフと交わることがない．同様の性質を 2 変数で考えてみる．2 変数関数 f の点 $(a,b,f(a,b))$ における f の接平面は，

$$z=f(a,b)+f_x(a,b)(x-a)+f_y(a,b)(y-b)$$

となる．凸関数のグラフに接している平面を想像すると，接平面は凸関数のグラフの常に下に来るので，これより以下を示すことができる：
2 変数関数 f が凸ならば, すべての $(a,b),(x,y)\in R^2$ に対して,

$$f(x,y)\geq f(a,b)+f_x(a,b)(x-a)+f_y(a,b)(y-b) \tag{2.2}$$

が成り立つ．
これは一般に n 変数関数に対しても成り立つので，その証明を以下に挙げる．

勾配ベクトル

まず次の記号を導入する．

定義. 2 変数関数 f と (x,y) に対して，

$$\nabla f(x,y)=\begin{bmatrix} f_x(x,y)\\ f_y(x,y)\end{bmatrix}$$

を f の (x,y) における 勾配ベクトル と呼ぶ．一般に，n 変数関数 f に関しては勾配ベクトルは，以下のように定義される．$u=(x_1,\ldots,x_n)$ に対して，

$$\nabla f(u)=\begin{bmatrix} f_{x_1}(u) \\ f_{x_2}(u) \\ \vdots \\ f_{x_n}(u) \end{bmatrix}$$

である．

例 **3.**

$$f(x,y)=x^2+3y^2$$

とすると，点 $(5,3)$ における勾配ベクトルは，$f_x=2x$, $f_y=6y$ より，

$$\nabla f(5,3)=\begin{bmatrix} f_x(5,3) \\ f_y(5,3) \end{bmatrix}=\begin{bmatrix} 2\cdot 5 \\ 6\cdot 3 \end{bmatrix}=\begin{bmatrix} 10 \\ 18 \end{bmatrix}$$

となる．

凸関数と接平面に関する不等式

この記号を用いて，式 (2.1.2) を示そう．

命題 **1.** n 変数関数 f が凸関数ならば，すべての $u,v\in\mathbb{R}^n$ に対して，

$$f(v)\geq f(u)+\nabla f(u)\cdot(v-u)$$

が成り立つ．ここで，$f(u)\cdot(v-u)$ はベクトル $f(u)$ と $(v-u)$ の内積を表す．

証明. $u,v\in\mathbb{R}^n$, $0<\lambda<1$ とする．関数 f のテーラー展開より，

$$f((1-\lambda)u+\lambda v)=f(u+\lambda(v-u))=f(u)+\nabla f(u)\cdot\{\lambda(u-v)\}+o(\lambda\|u-v\|)$$

となる．凸関数の定義より，

$$\begin{aligned}(1-\lambda)f(u)+\lambda f(v)&\geq f\left((1-\lambda)u+\lambda v\right)\\ &=f\left(u+\lambda(v-u)\right)\\ &=f(u)+\lambda\nabla f(u)\cdot(v-u)+o(\lambda\|v-u\|),\\ \lambda f(v)&\geq\lambda f(u)+\lambda\nabla f(u)\cdot(v-u)+o(\lambda\|v-u\|),\\ f(v)&\geq f(u)+\nabla f(u)\cdot(v-u)+\frac{o(\lambda\|v-u\|)}{\lambda}\end{aligned}$$

となる．$\lambda\to 0$ とすると，$\frac{o(\lambda\|u-v\|)}{\lambda}\to 0$ となるので，命題の不等式は示された．□

2.2　凸関数の判定

2.2.1　凸性と微分

関数 $f(x) = x^2$ はグラフが下に突き出ており，凸関数であることがわかる．それでは，関数

$$f(x) = \sqrt{1 + x^2}$$

は凸関数だろうか？

定義 2.1.1 を確認するのは困難なので，グラフの概形を調べよう．

$$f'(x) = \frac{x}{\sqrt{1 + x^2}}, \quad f''(x) = \frac{1}{(1 + x^2)^{\frac{3}{2}}}$$

より，すべての x に対して，$f''(x) > 0$ となる．よって，全区間で「下に凸」な関数であることがわかるので，$f(x)$ は凸関数である．

増減表を考えると，1 変数関数 に対しては，以下のような性質が成り立つ．

h が凸関数

（定義）$\Leftrightarrow$ $h'(t)$ が非減少関数 $(t_1 < t_2 \;\Rightarrow\; h'(t_1) \leq h'(t_2))$

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

例えば，関数 $h(x) = x^4$ はグラフが下に突き出ており，凸関数であることがわかる．いま，$h'(x) = 4x^3$，$h''(x) = 12x$ なので，上記の性質を満たす．

次に，2 変数関数

$$g(x, y) = x^2 - 2xy + 2y^2$$

は凸関数だろうか？

今度は定義 2.1.1 を直接確認するのも，グラフの概形を調べるのも大変である．しかし，実は 2 変数関数についても，微分を用いて関数が凸であるかどうかを調べることができる．

ヘッセ行列

まず新しい記号を導入する．多変数の場合, 1 変数関数の 1 階微分に勾配ベクトルが対応し, 2 階微分には以下のヘッセ行列が対応する．

定義．2 変数関数 f と (x, y) に対して，

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

を ヘッセ行列 と呼ぶ．この本では，扱う関数は十分な回数微分可能なので，常に $f_{xy} = f_{yx}$ が成り立つ．

例えば,

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

とすると, 勾配ベクトルは

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

で, ヘッセ行列は

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

となる. 一般の n 変数関数 f と $u = (x_1, \ldots, x_n)$ に対しては, ヘッセ行列は

$$\nabla^2 f(u) = \begin{bmatrix} \frac{\partial^2 f}{\partial x_1 \partial x_1}(u) & \frac{\partial^2 f}{\partial x_1 \partial x_2}(u) & \cdots & \frac{\partial^2 f}{\partial x_1 \partial x_n}(u) \\ \frac{\partial^2 f}{\partial x_2 \partial x_1}(u) & \ddots & & \vdots \\ \vdots & & & \\ \frac{\partial^2 f}{\partial x_n \partial x_1}(u) & \cdots & & \frac{\partial^2 f}{\partial x_n \partial x_n}(u) \end{bmatrix}$$

と定義される．

ヘッセ行列による条件

ヘッセ行列を用いると以下のような凸性の判定方法がある.

定理 2. n 変数関数 f に対して以下が成り立つ；

(1). f が凸関数 $\Longleftrightarrow$ 各点 $a \in \mathbb{R}^n$ で,

$${}^t u \nabla^2 f(a) u \geq 0 \ (\text{すべてのベクトル } u \in \mathbb{R}^n)$$

が成り立つ

(2). f が狭義凸関数 $\Longleftarrow$ 各点 $a \in \mathbb{R}^n$ で,

$${}^t u \nabla^2 f(a) u > 0 \ (u \neq 0 \text{ となるすべてのベクトル } u \in \mathbb{R}^n)$$

が成り立つ.

証明の概要. $b, c \in \mathbb{R}^n$ に対して,

$$h(t) = f\left((1-t)b + tc\right)$$

とおく. すると

「f が凸 $\Leftrightarrow$ すべての $b, c \in \mathbb{R}^n$ に対して $h(t)$ が凸」

が成り立つ．また，h は 1 変数関数なので，

$$「h(t) \text{ が凸} \Leftrightarrow h''(t) \geq 0」$$

が成り立つ．ここで，

$$\begin{cases} a = (1-t)b + tc = b + t(c-b) \\ u = c - b \end{cases}$$

に対して

$$h''(t) = {}^t u \nabla^2 f(a) u$$

となる．これより命題が導かれる． □

例 4.

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

とすると，勾配ベクトルは

$$\nabla f(x,y) = \begin{bmatrix} 2x - 2y \\ -2x + 4y \end{bmatrix}$$

で，ヘッセ行列は

$$\nabla^2 f(x,y) = \begin{bmatrix} 2 & -2 \\ -2 & 4 \end{bmatrix}$$

となる．定理 2 を用いて f が凸関数かどうか調べよう．$u = {}^t(u_1, u_2)$ とすると，各点 (x,y) で，

$${}^t u \nabla f(x,y) u = [\ u_1 \ \ u_2\ ] \begin{bmatrix} 2 & -2 \\ -2 & 4 \end{bmatrix} \begin{bmatrix} u_1 \\ u_2 \end{bmatrix} = 2u_1^2 - 4u_1u_2 + 4u_2^2$$

となる．さらに，すべてのベクトル u に対して

$$2u_1^2 - 4u_1u_2 + 4u_2^2 = 2(u_1^2 - 2u_1u_2 + 2u_2^2) = 2\left\{(u_1 - u_2)^2 + u_2^2\right\} \geq 0$$

となるので，定理 2 より f は凸関数である．

2.2.2 行列の正値生

2 次形式

定理 2 の条件（行列 A に対して，${}^t uAu \geq 0$ など）は線形代数でよく扱われるものである．この章の残りで，これら条件を行列の言葉を用いて詳しく調る．

定義. 行列 A とベクトル $u=(u_1,u_2,\ldots,u_n)$ に対して，$u_1,\ldots,u_n$ を変数に持つ多項式

$$p(u)={}^t uAu$$

を 2 次形式と呼ぶ.

例 **5.**

$$A=\begin{bmatrix}2&0\\0&1\end{bmatrix},$$

$u=(x,y)$ とすると，

$$^t uAu=[\ x\ \ y\ ]\begin{bmatrix}2&0\\0&1\end{bmatrix}\begin{bmatrix}x\\y\end{bmatrix}=[\ x\ \ y\ ]\begin{bmatrix}2x\\y\end{bmatrix}=2x^2+y^2$$

は 2 次形式である.
　また，

$$B=\begin{bmatrix}1&1\\1&-1\end{bmatrix}$$

とすると，

$$\begin{aligned}^t uBu=[\ x\ \ y\ ]\begin{bmatrix}1&1\\1&-1\end{bmatrix}\begin{bmatrix}x\\y\end{bmatrix}&=[\ x\ \ y\ ]\begin{bmatrix}x+y\\x-y\end{bmatrix}\\&=x^2+xy+yx-y^2=x^2+2xy-y^2\end{aligned}$$

も 2 次形式である.

正値性の定義

定義. 行列 A を持つ対称行列 とする.

- すべての $u\in R^n$ に対して，2 次形式が

  $$^t uAu\geq 0$$

  を満たすとき，A を 半正定値 と言う. また，逆の不等号が成り立つとき，A を 半負定値 と言う.

- $u\neq 0$ を満たすすべての $u\in R^2$ に対して，2 次形式が

  $$^t uAu>0$$

  を満たすとき，A を 正定値 と呼ぶ. 逆の不等号が成り立つとき，A を 負定値 と呼ぶ.

- $u \in R^2$ によって 2 次形式 ${}^t uAu$ の符号が正にも負にもなるとき，A を不定と言う．

例 6. 例 5 の行列

$$A = \begin{bmatrix} 2 & 0 \\ 0 & 1 \end{bmatrix}$$

の正値性を調べてみよう．$u = (x, y)$ とすると，2 次形式は

$${}^t uAv = 2x^2 + y^2$$

となる．ここで，$2x^2 + y^2 > 0$ $(u \neq 0)$ となるので，A は正定値である．

一方，行列

$$B = \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix}$$

に対して，2 次形式は

$${}^t uBu = x^2 + 2xy - y^2$$

となる．ここで，$u = (1, 0)$ とすると ${}^t uBu = 1 > 0$ となるが，$u = (0, 1)$ とすると ${}^t uBu = -1 < 0$ となるので，B は不定である．

凸関数の判定定理の再掲（正値性を用いて）

これらの言葉を用いると，定理 2 は以下のように書ける．

定理 3 (定理 2 の言い換え)**.** n 変数関数 f に対して以下が成り立つ：

- f が凸関数
  $\Longleftrightarrow$ 各点 $a \in \mathbb{R}^n$ において，ヘッセ行列 $\nabla^2 f(a)$ が半正定値である．
- f が狭義凸関数
  $\Longleftarrow$ 各点 $a \in \mathbb{R}^n$ において，ヘッセ行列 $\nabla^2 f(a)$ が正定値である．

正値性の調べ方

さて, 関数のヘッセ行列が各点で半正定値であれば, その関数は凸になることが分かった. それでは, 行列が半正定値であるとはどのように調べたらよいだろうか? 例 5 では，2 次形式に個別の工夫をすることで正値性を調べた．しかし，線形代数を用いると以下のような一般的な方法がある．

固有値を用いた判定法

定理 4. A を対称行列とする．以下の主張が成り立つ．

(1). A が半正定値 $\Leftrightarrow$ A の固有値 がすべて 0 以上

(2). A が正定値 $\Leftrightarrow$ A の固有値がすべて正

(3). A が不定 $\Leftrightarrow$ A が正と負の固有値を持つ

証明. A を $n \times n$ 行列とする．A は対称行列なので，ある直交行列 P が存在して,

$$P^{-1}AP = \Lambda$$

と対角化できる．ここで, Λ は A の固有値を対角要素に持つ対角行列である．いま，$P^{-1} = {}^tP$ に注意して，$v = {}^tPu$ と変数変換すると，$u = Pv$ より，

$${}^tuAu = {}^t(Pv)A(Pv) = {}^tv({}^tPAP)v = {}^tv\Lambda v = \lambda_1 v_1^2 + \cdots + \lambda_n v_n^2$$

を得る．ここで，$v = (v_1, \ldots, v_n)$ であり，$\lambda_1, \ldots, \lambda_n$ は A の固有値である．よって，固有値がすべて 0 以上ならば，${}^tuAu \geq 0$ （すべての u）となり，これより (1) が導かれる．他も同様である． □

例 7.

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

とすると，

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

となる．ヘッセ行列は定数行列であり，この固有値は $4, 8$ になる．よって，ヘッセ行列は正定値で，f は狭義凸関数である．

行列式を用いた判定法

正値性を調べたい行列が 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．

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

$$\begin{aligned} [\ x \quad y\ ]A \begin{bmatrix} x \\ y \end{bmatrix} &= ax^2 + 2bxy + dy^2 = a\left(x + \frac{by}{a}\right)^2 + \frac{y^2}{a}(ad - b^2) \\ &= a\left(x + \frac{by}{a}\right)^2 + \frac{1}{a}|A|\ y^2 \qquad (2.3) \end{aligned}$$

となる．ここで $|A|$ は A の行列式 $ad - b^2$ を表す．この式より，

定理 5.

(1). $|A| > 0$ かつ $a > 0 \Leftrightarrow A$ は正定値

(2). $|A| > 0$ かつ $a < 0 \Leftrightarrow A$ は負定値

(3). $|A| < 0 \Leftrightarrow A$ は不定

証明. (1) まず，$|A| > 0$ かつ $a > 0$ として，

$$(x, y) \neq (0, 0) \Rightarrow [\ x \quad y\ ] A \begin{bmatrix} x \\ y \end{bmatrix} > 0$$

を示す．いま，

$$a\left(x + \frac{by}{a}\right)^2 + \frac{1}{a}|A|\ y^2 \geq 0$$

となる．ここで，

$$a\left(x + \frac{by}{a}\right)^2 + \frac{1}{a}|A|\ y^2 = 0$$

とすると，各項は 0 以上なので，

$$\left(x + \frac{by}{a}\right) = 0,\ \text{かつ}\ y = 0$$

を得る．よって，$(x, y) = (0, 0)$ となるので，対偶をとると，式 (2.3) より上記が示される．逆に，

$$(x, y) \neq (0, 0) \Rightarrow [\ x \quad y\ ] A \begin{bmatrix} x \\ y \end{bmatrix} > 0$$

とする．上の 2 次形式に $(x, y) = (1, 0)$ を代入すると，$a > 0$ を得る．さらに，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．よって，定理 4 より $|A| > 0$ となる．(2) も同様に示される．

(3) 行列式の値は固有値の積と等しいことから，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．よって定理 4 より示される． □

例 8. 例 7 のヘッセ行列

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

の正値性を行列式を用いて調べる．

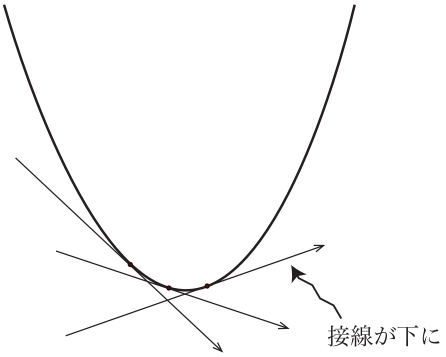
$$|\nabla^2 f(x, y)| = 32 > 0\ \text{かつ}\ f_{xx}(x, y) = 6 > 0$$

なので，ヘッセ行列は正定値である．

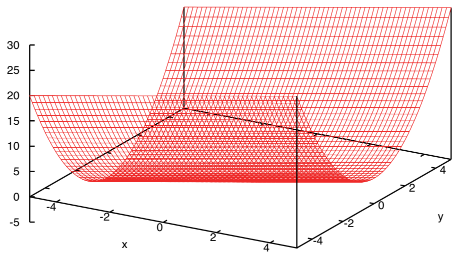
まとめ

凸関数

最適化問題で重要（3 章で説明）



$$\Longleftrightarrow \quad f(v) \geq \underbrace{f(u) + \nabla f(u) \cdot (u - v)}_{\text{接平面の式}}$$



$$\Longleftrightarrow \quad {}^t u \nabla^2 f(a) u \geq 0 \ (\text{すべての } u \in \mathsf{R}^n)$$

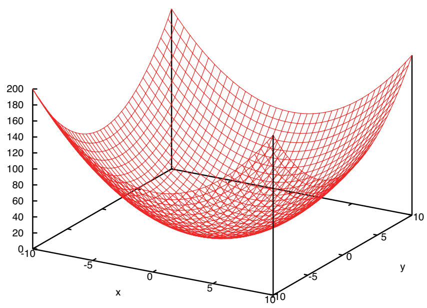
$\Updownarrow$ 2 次形式と見ると

$\nabla^2 f(a)$ が半正定値

$\Updownarrow$

$\nabla^2 f(a)$ のすべての固有値が ≥ 0

狭義凸関数



$$\Longleftarrow \quad {}^t u \nabla^2 f(a) u > 0$$

($u \neq 0$ となるすべてのベクトル $u \in \mathsf{R}^n$)

$\Updownarrow$

$\nabla^2 f(a)$ が正定値

$\Updownarrow$

$\nabla^2 f(a)$ のすべての固有値が正

登場した道具（2 変数の場合）

勾配ベクトル　$\nabla f(a) = \begin{bmatrix} f_x(a) \\ f_y(a) \end{bmatrix}$　ヘッセ行列　$\nabla^2 f(a) = \begin{bmatrix} f_{xx}(a) & f_{xy}(a) \\ f_{yx}(a) & f_{yy}(a) \end{bmatrix}$

2 次形式　${}^t u A u = [\ x \ \ y\ ] \begin{bmatrix} a & b \\ c & d \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} = ax^2 + (b+c)xy + dy^2$