数理計画法 第10回
第3章　非線形計画
3.2　制約なし最適化

担当：　塩浦昭義
(情報科学研究科　助教授)
shioura@dais.is.tohoku.ac.jp

復習:最急降下法

最急降下法のアイディア：
勾配ベクトルと逆の方向に進むと関数値が減る

現在の点 x を x－ α ∇f(x) により更新
⇒　関数値 f(x) を減らしていく

ステップサイズ

ステップサイズの選び方：
次の一変数最適化問題を (近似的に) 解く
最小化　f(x－ α ∇f(x))　条件　α＞0

ヘッセ行列 [p.90]

関数 f の**ヘッセ行列**

$$
\mathrm{H}\, f(\boldsymbol{x}) = \begin{pmatrix} \dfrac{\partial^2 f}{\partial x_1 \partial x_1} & \dfrac{\partial^2 f}{\partial x_1 \partial x_2} & \cdots & \dfrac{\partial^2 f}{\partial x_1 \partial x_n} \\ \dfrac{\partial^2 f}{\partial x_2 \partial x_1} & \dfrac{\partial^2 f}{\partial x_2 \partial x_2} & \cdots & \dfrac{\partial^2 f}{\partial x_2 \partial x_n} \\ \vdots & \vdots & \ddots & \vdots \\ \dfrac{\partial^2 f}{\partial x_n \partial x_1} & \dfrac{\partial^2 f}{\partial x_n \partial x_2} & \cdots & \dfrac{\partial^2 f}{\partial x_n \partial x_n} \end{pmatrix}
$$

一変数関数の場合は2階導関数に一致

$$\mathrm{H}\, f(x) = f''(x)$$

ヘッセ行列 (続き) [p.89]

例：

$$f_1(x) = x^2 \Longrightarrow \nabla f_1(x) = 2x \Longrightarrow \mathrm{H}\, f_1(x) = 2$$

$$f_2(x_1, x_2) = \sin x_1 + \cos x_2 \Longrightarrow \nabla f_2(\boldsymbol{x}) = \begin{pmatrix} \cos x_1 \\ -\sin x_2 \end{pmatrix}$$

$$\Longrightarrow \mathrm{H}\, f_2(\boldsymbol{x}) = \begin{pmatrix} -\sin x_1 & 0 \\ 0 & -\cos x_2 \end{pmatrix}$$

$$f_3(x_1, x_2) = x_1 x_2 + \log x_1$$

$$\Longrightarrow \nabla f_3(\boldsymbol{x}) = \begin{pmatrix} x_2 + 1/x_1 \\ x_1 \end{pmatrix} \Longrightarrow \mathrm{H}\, f_3(\boldsymbol{x}) = \begin{pmatrix} -1/x_1^2 & 1 \\ 1 & 0 \end{pmatrix}$$

ヘッセ行列とテイラー展開[p.90]

関数 f は勾配ベクトルとヘッセ行列により表現される
2次関数により近似できる

$$f(\boldsymbol{x}+\boldsymbol{d}) = f(\boldsymbol{x}) + \nabla f(\boldsymbol{x})^T \boldsymbol{d} + \frac{1}{2}\boldsymbol{d}^T \mathrm{H} f(\boldsymbol{x}) \boldsymbol{d} + o(\|\boldsymbol{d}\|^2)$$

$$\|\boldsymbol{d}\| = \sqrt{d_1^2 + d_2^2 + \cdots + d_n^2}$$

関数 f の (x における)
2次のテイラー展開

関数 f を多項式で表現したときの
2次項までの情報を取り出した式

ヘッセ行列とテイラー展開 (続き) [p.90]

$$f(\boldsymbol{x}+\boldsymbol{d}) = f(\boldsymbol{x}) + \nabla f(\boldsymbol{x})^T \boldsymbol{d} + \frac{1}{2}\boldsymbol{d}^T \mathrm{H} f(\boldsymbol{x}) \boldsymbol{d} + o(\|\boldsymbol{d}\|^2)$$

例1： $f_1(x) = x^2$ $\quad \nabla f_1(x) = 2x$ $\quad \mathrm{H} f_1(x) = 2$

左辺 $= f_1(x+d) = (x+d)^2$

右辺 $= x^2 + (2x)d + d^2 = (x+d)^2$

∴2次関数 f_1 の2次のテイラー展開は f_1 に等しい

※一般に、2次関数 $f(\boldsymbol{x}) = \frac{1}{2}\boldsymbol{x}^T V \boldsymbol{x} + \boldsymbol{c}^T \boldsymbol{x} + c_0$

のテイラー展開は f に等しい

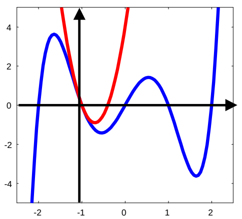
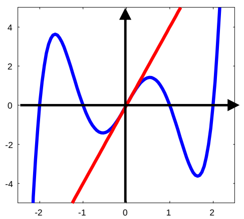
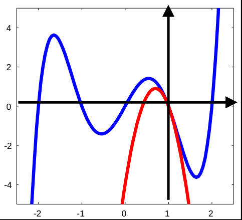
V: n×n行列
c: n 次元ベクトル
c_0: 定数

ヘッセ行列とテイラー展開 (続き) [p.90]

例2： $f(x) = (x-2)(x-1)x(x+1)(x+2) = x^5 - 5x^3 + 4x$

$\nabla f(x) = 5x^4 - 15x^2 + 4$ $\qquad \mathrm{H} f(x) = 20x^3 - 30x$

	x = -1 のとき	x = 0 のとき	x = 1 のとき
テイラー展開	$0 - 6d + 5d^2$	$0 + 4d + 0d^2$	$0 - 6d - 5d^2$

行列の正定値性、半正定値性[p.99]

正定値 (半正定値) ････行列が「正 (非負)」

定義： 正方行列 A は半正定値
⇔ 任意のベクトル y に対して $y^T A y \geqq 0$

定義： 正方行列 A は正定値
⇔ 任意の非零ベクトル y に対して $y^T A y > 0$

※ A が1×1行列のとき、
Aは半正定値 ⇔ $a_{11} \geqq 0$, Aは正定値 ⇔ $a_{11} > 0$

行列の正定値性、半正定値性[p.99]

定義：正方行列 A は**半正定値**
　⇔ 任意のベクトル y に対して $y^T A y \geqq 0$

定義：正方行列 A は**正定値**
　⇔ 任意の非零ベクトル y に対して $y^T A y > 0$

※ A が2×2行列のとき、
　Aは正定値 ⇔ $a_{11} > 0,\ a_{22} > 0,\ a_{11}a_{22} - a_{12}^2 > 0$
　Aは半正定値 ⇔ $a_{11} \geqq 0,\ a_{22} \geqq 0,\ a_{11}a_{22} - a_{12}^2 \geqq 0$

$\begin{pmatrix} 2 & -2 \\ -2 & 5 \end{pmatrix}$ 正定値　$\begin{pmatrix} 2 & -2 \\ -2 & 2 \end{pmatrix}$ 半正定値　$\begin{pmatrix} 2 & -2 \\ -2 & 0 \end{pmatrix}$ 半正定値ではない

2次の最適性条件（必要条件）[p.99]

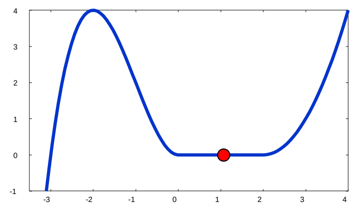
ヘッセ行列を用いた最適性条件
（極小解であるための条件）

定理（2次の必要条件）：
x^*: 制約なし問題の**極小解** ⇒ $Hf(x^*)$ は**半正定値**

※f が1変数関数の場合，「x^*:極小解 ⇒ $f''(x^*) \geqq 0$」

例：
x* = 1 は極小解
0≦x≦2 の範囲で f(x) = 0

⇒ ∇f(x*) = f '(x*) = 0
　Hf(x*) = f ''(x*) = 0
　　　半正定値

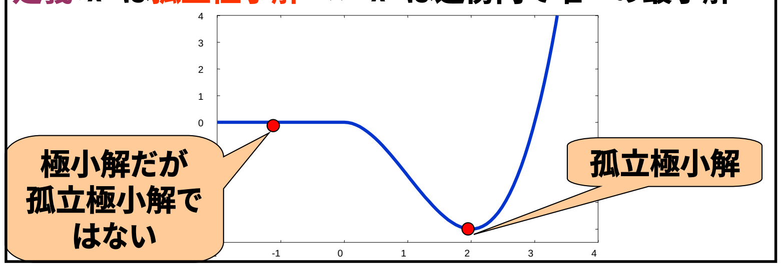


2次の最適性条件（十分条件）[p.100]

定理（2次の十分条件）：
x^* は**停留点**，$Hf(x^*)$ は**正定値** ⇒ x^*:（**孤立**）**極小解**

※f が1変数関数の場合，
　「$f'(x^*) = 0,\ f''(x^*) > 0$ ⇒ x^*: 孤立極小解」

定義：x^* は**孤立極小解** ⇔ x^* は近傍内で唯一の最小解



2次の最適性条件（十分条件）[p.100]

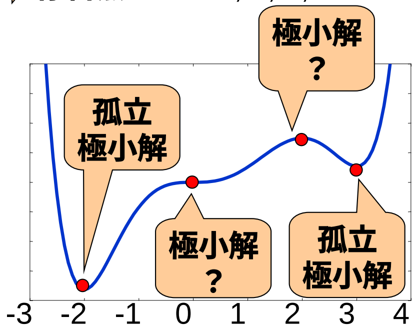
例1：$f(x) = \frac{1}{6}x^6 - \frac{3}{5}x^5 - x^4 + 4x^3$

$\nabla f(x) = (x+2)x^2(x-2)(x-3)$

⇒ 停留点はx = -2, 0, 2, 3

$Hf(x) = 5x^4 - 12x^3 - 12x^2 + 24x$

⇒ Hf(-2) = 80 > 0
Hf(0) = 0
Hf(2) = -16 < 0
Hf(3) = 45 > 0



2次の最適性条件（十分条件） [p.100]

例2（教科書の例3．4）：

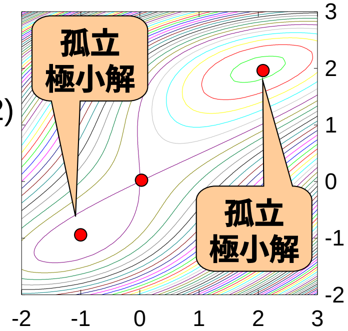
$$f(\boldsymbol{x})=x_1^2-2x_1x_2+\frac{1}{4}x_2^4-\frac{1}{3}x_2^3$$

$$\nabla f(\boldsymbol{x})=\begin{pmatrix}2x_1-2x_2\\ x_2^3-x_2^2-2x_1\end{pmatrix}$$

停留点は(0,0), (-1, -1), (2, 2)

$$\mathrm{H}f(\boldsymbol{x})=\begin{pmatrix}2 & -2\\ -2 & 3x_2^2-2_2\end{pmatrix}$$

(-1, -1), (2, 2) は
孤立極小解



極大解に関する性質

- x* は関数 f の（孤立）極大解
  ⇔ x* は関数 – f の（孤立）極小解
- x* における関数 – f のヘッセ行列は – Hf(x)

極大解であるための条件

定理：（必要条件）
x^*: 制約なし問題の極大解 ⇒ – Hf(x^*) は半正定値

定理：（十分条件）
x^* は停留点， – Hf(x^*) は正定値
⇒ x^*: 制約なし問題の（孤立）極大解

2次の必要条件の証明 [p.99]

定理　x^*: 極小解 ⇒ Hf(x^*) は半正定値

この値を
- δとおく

証明： ・Hf(x^*) が半正定値でないと仮定

ある u（||u|| = 1 と仮定）に対し u^T Hf(x^*)u ＜ 0

・x^* は極小解 ⇒ ∇f(x^*) ＝0

・2次のテイラー展開の式に d = εu を代入（ε＞0）

$$f(\boldsymbol{x}^*+\varepsilon\boldsymbol{u}) = f(\boldsymbol{x}^*)+\varepsilon\nabla f(\boldsymbol{x}^*)^T\boldsymbol{u}+\frac{1}{2}\varepsilon^2\boldsymbol{u}^T\mathrm{H}f(\boldsymbol{x}^*)\boldsymbol{u}+o(\varepsilon^2)$$
$$= f(\boldsymbol{x}^*)-\frac{1}{2}\varepsilon^2\delta+o(\varepsilon^2)= f(\boldsymbol{x}^*)-\varepsilon^2\left(\frac{1}{2}\delta-\frac{o(\varepsilon^2)}{\varepsilon^2}\right)$$

・ε を十分小さくする ⇒ δ/2 – o(ε²)/ ε² > 0
⇒ f(x^* + εu) < f(x^*)　　[矛盾]

制約なし問題の解法2：ニュートン法
[p.105]

定義：2次関数 $f(x)=\frac{1}{2}\boldsymbol{x}^TV\boldsymbol{x}+\boldsymbol{c}^T\boldsymbol{x}+c_0$
は狭義2次凸関数 ⇔ V は正定値行列

ニュートン法のアイディア：
狭義2次凸関数の最適解は簡単に求められる！

$$\nabla f(\boldsymbol{x})=V\boldsymbol{x}+\boldsymbol{c}\qquad \mathrm{H}\,f(\boldsymbol{x})=V$$

停留点は x^* = – V^{-1}c のみ, ヘッセ行列は V（正定値）

2次の十分条件より x^* は最適解

定理：x^*: 停留点, Hf(x^*)：正定値 ⇒ x^*: 孤立極小解

制約なし問題の解法2：ニュートン法 [p.105]

ニュートン法のアイディア：
狭義2次凸関数の最適解は簡単に求められる！

一般の関数 f は狭義2次凸関数とは限らない

⇨ **f を2次のテイラー展開により近似**

$$f(\boldsymbol{x}+\boldsymbol{d}) \cong f(\boldsymbol{x}) + \nabla f(\boldsymbol{x})^T \boldsymbol{d} + \frac{1}{2}\boldsymbol{d}^T \mathrm{H} f(\boldsymbol{x})\boldsymbol{d}$$

ヘッセ行列 Hf(x) が正定値のとき
最適解は　$d = -\mathrm{H}f(x)^{-1}\nabla f(x)$　　**ニュートン方向**

⇨ **x + d は f の最適解のより良い近似解と期待できる**

ニュートン法のアルゴリズム [p.106]

現在の点 x を繰り返しニュートン方向へ移動、最適解に近づける

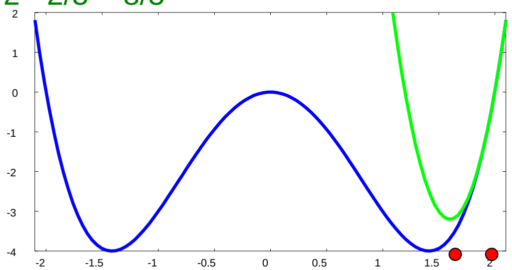
入力：関数 f とその勾配ベクトル∇f, ヘッセ行列Hf
初期点 x^0

ステップ0： k =0 とする
ステップ1： x^k が最適解に十分近ければ終了
ステップ2：ニュートン方向 $-\mathrm{H}f(x^k)^{-1}\nabla f(x^k)$ を計算
ステップ3： $x^{k+1}=x^k - \mathrm{H}f(x^k)^{-1}\nabla f(x^k)$ とおく
ステップ4： k=k+1として、ステップ1に戻る

ニュートン法の例 [p.106]

例1：一変数関数 $f(x) = x^4 - 4x^2$

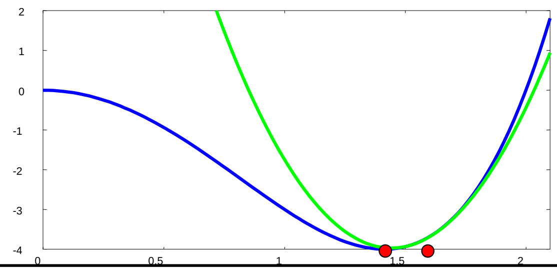
初期点 $x = 2$ **において f のテイラー近似を求める**
⇒ $f(2+d)$　　$0 + 16d + (40/2)d^2$
$d = -2/5$ **のとき最小**
⇒ **次の点は** $x = 2 - 2/5 = 8/5$



ニュートン法の例 [p.106]

例1(続き)：一変数関数 $f(x) = x^4 - 4x^2$

点 $x = 8/5$ **において f のテイラー近似を求める**
⇒ $f(8/5+d)$　　$-3.69 + 3.58d + 11.36d^2$
$d = -0.11$ **のとき最小**
⇒ **次の点は** $x = 1.6 - 0.11 = 1.49$



ニュートン法の特徴 [p.107]

長所：

- 最急降下法より**反復回数が少ない**
  - ■ 狭義2次凸関数に対しては**一反復**で終了
- 直線探索が不要

短所：

- **ヘッセ行列の逆行列の計算**が必要
  - ■ **ヘッセ行列の計算**ができないと破綻
  - ■ ヘッセ行列が**正則**でないと破綻
- ヘッセ行列が正定値でない場合には **目的関数値が増加する**可能性あり

ニュートン法の問題点 [p.107]

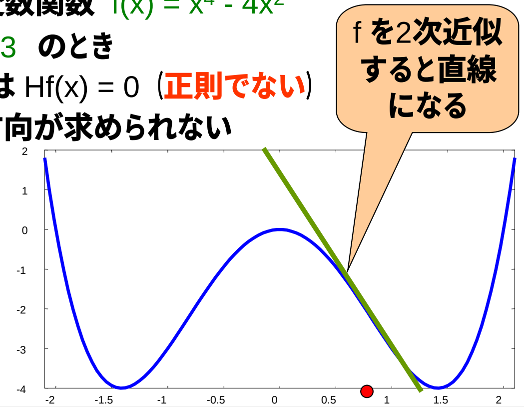
■ ヘッセ行列が**正則**でないと破綻

例1（続き）：一変数関数 $f(x) = x^4 - 4x^2$

初期点 $x = \sqrt{2/3}$ のとき

⇒ ヘッセ行列は $Hf(x) = 0$ （**正則でない**）

⇒ ニュートン方向が求められない



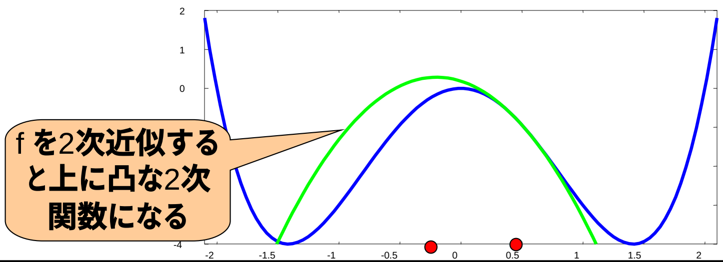
ニュートン法の問題点 [p.107]

- ヘッセ行列が正定値でない場合には **目的関数値が増加する**可能性あり

初期点 $x = 1/2$ のとき

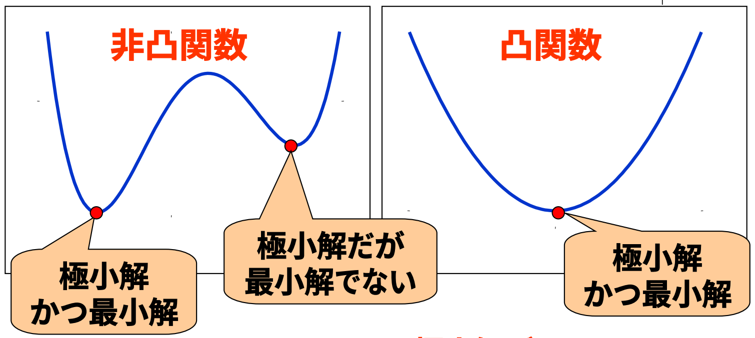
⇒ ヘッセ行列は $Hf(x) = -5$ （**正定値でない**）

⇒ ニュートン方向に進むと関数値が増加する



凸関数 [p.93]

最小化しやすい関数の形は？



最小解でない極小解がある
➔最小化が難しい

極小解が一つ
➔最小化しやすい

凸関数の定義 [p.94]

定義：関数 f は**凸関数**
⇔ 任意のベクトル x, y
および任意の 0≦t≦1 に対して
$(1-t)\,f(x) + t\,f(y) \geqq f((1-t)\,x + t\,y)$

注：教科書の定義と異なります

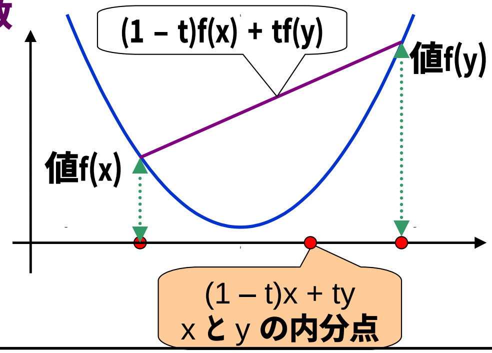
定義：関数 f は**狭義凸関数**
⇔ 任意の**異なる**ベクトル x, y
および任意の **0<t<1** に対して
$(1-t)\,f(x) + t\,f(y) > f((1-t)\,x + t\,y)$

凸関数の定義（続き） [p.94]

凸関数 ⇔ $(1-t)\,f(x) + t\,f(y) \geqq f((1-t)\,x + t\,y)$

狭義凸関数 ⇔ $(1-t)\,f(x) + t\,f(y) > f((1-t)\,x + t\,y)$

狭義凸関数の例



凸関数の定義（続き） [p.94]

凸関数 ⇔ $(1-t)\,f(x) + t\,f(y) \geqq f((1-t)\,x + t\,y)$

狭義凸関数 ⇔ $(1-t)\,f(x) + t\,f(y) > f((1-t)\,x + t\,y)$

狭義凸関数の例

2次関数 $f(\boldsymbol{x}) = \frac{1}{2}\boldsymbol{x}^T V \boldsymbol{x} + \boldsymbol{c}^T \boldsymbol{x} + c_0$

(V: n ×n 行列, c: n次元ベクトル, c_0: 定数)

V：正定値行列 ⇒ **狭義凸関数**

例えば

$$f(x_1, x_2) = \frac{1}{2}\begin{pmatrix} x_1 \\ x_2 \end{pmatrix}^T \begin{pmatrix} 2 & -2 \\ -2 & 5 \end{pmatrix}\begin{pmatrix} x_1 \\ x_2 \end{pmatrix}$$

凸関数の定義（続き） [p.94]

凸関数 ⇔ $(1-t)\,f(x) + t\,f(y) \geqq f((1-t)\,x + t\,y)$

狭義凸関数 ⇔ $(1-t)\,f(x) + t\,f(y) > f((1-t)\,x + t\,y)$

狭義凸関数の例

2次関数 $f(x) = a\,x^2$ $(a > 0)$ は**狭義凸関数**

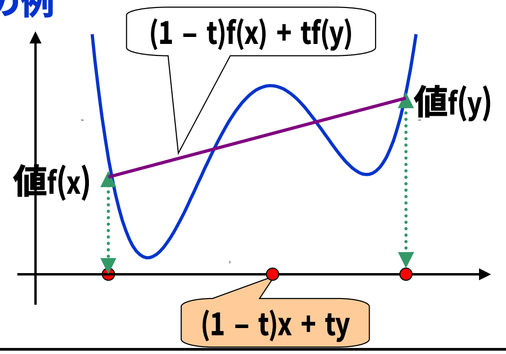
(証明) 任意の異なる x, y と $0 < t < 1$ に対して、

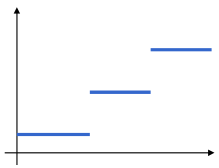
$(1\text{-}t)\,ax^2 + t\,a\,y^2 - a\,[(1\text{-}t)\,x + t\,y]^2$
$= (1\text{-}t)\,ax^2 + t\,a\,y^2 - a\,(1\text{-}t)^2\,x^2 - a\,t^2\,y^2 - 2a\,(1\text{-}t)\,t\,x\,y$
$= (t\text{-}t^2)\,ax^2 + (t\text{-}t^2)\,a\,y^2 - 2a\,(t\text{-}t^2)\,x\,y$
$= (t\text{-}t^2)\,a\,(x - y)^2$
> 0 （$0 < t < 1$, $x \neq y$ より）

凸関数の定義（続き） [p.94]

凸関数 ⇔ $(1-t)\,f(x) + t\,f(y) \geqq f((1-t)\,x + t\,y)$

非凸関数の例





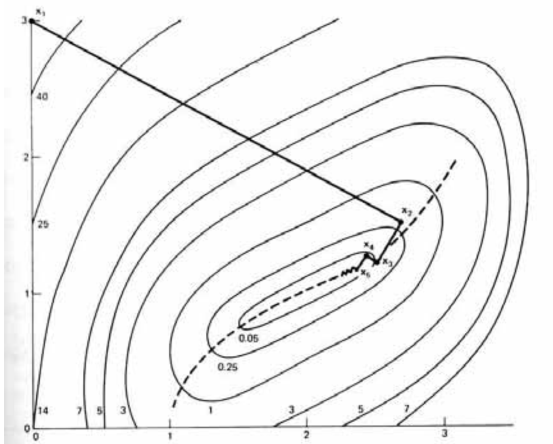
凸関数の最適解の性質 [p.101]

定理： f: 凸関数, x^*: f の極小解
⇒ x^*は制約なし問題の最適解

証明： x^* は極小解
⇒ ある $\varepsilon > 0$が存在して、
任意の x に対し $\|x - x^*\| < \varepsilon$ ならば $f(x) \geqq f(x^*)$

$f(y) < f(x^*)$ なる y が存在すると仮定
f は凸関数
⇒ $0 < t < 1$ なる任意の t に対して
$f((1 - t)\,y + t\,x^*) \leqq (1 - t)\,f(y) + t\,f(x^*) < f(x^*)$
t を1に近づけると
$(1 - t)\,y + t\,x^*$ と x^* の距離 $<\ \varepsilon$ （矛盾）

前回のレポートの解答： 最急降下法の問題



レポート問題

(1) 対称な2×2行列 A に対し、次の関係を証明せよ。
A は半正定値 ⇔ $a_{11} \geqq 0,\ a_{22} \geqq 0,\ a_{11}a_{22} - a_{12}^2 \geqq 0$
（ヒント：教科書の問題3.7の答えを参考にせよ）

(2) 関数 $f(x) = x^3 + 6x^2$ に対して
(a) 初期点を $x = 2$ としてニュートン法を適用せよ。
(b) 初期点を $x = 1$ としてニュートン法を適用せよ。
それぞれ、反復は2回行うこと。

(3) 狭義凸ではない凸関数の例をひとつ挙げなさい
（式で書く必要はなく，グラフを書けばよい）．