数理計画法　第12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　准教授)
shioura@dais.is.tohoku.ac.jp

期末試験について

- 日時：２月５日（木）午後１時より
- 受験資格者：ネットワーク計画，非線形計画に関するレポートを一回以上提出した学生のみ
- 教科書等の持込は不可
- 座席は指定
- 試験内容：ネットワーク計画，非線形計画の範囲
  （今回の内容まで）
  （詳しくはWeb上の過去問を参考にしてください）

呼び出し：A6TB2004 安久津誠君
A6TB2229 松川祐己君，A6TB2243 村山竜太君
A5TB2065 翁長修一君，A6TB2167 田村直樹君

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

関数 f(x) の x=a における二次のテイラー展開
（a は定数ベクトル）

$$f(x) = \underbrace{f(a) + \nabla f(a)^T (x-a) + \frac{1}{2}(x-a)^T \mathsf{H} f(a)(x-a)} + \psi(x-a)$$

この部分が二次関数, f(x) を近似

関数 $\psi(x-a) = \varphi(x_1 - a_1, x_2 - a_2, \ldots, x_n - a_n)$ は
$x_1 - a_1, x_2 - a_2, \ldots, x_n - a_n$ に関する
３次以上の項からなる n 変数多項式関数
(定数項, 一次, 二次の項は全く含まれない)

制約なし問題の解法2:ニュートン法 [p.105]

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

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 $d = -Hf(x)^{-1} \nabla f(x)$

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 $-Hf(x^k)^{-1}\nabla f(x^k)$ を計算

ステップ3: $x^{k+1}=x^k - Hf(x^k)^{-1}\nabla f(x^k)$ 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]

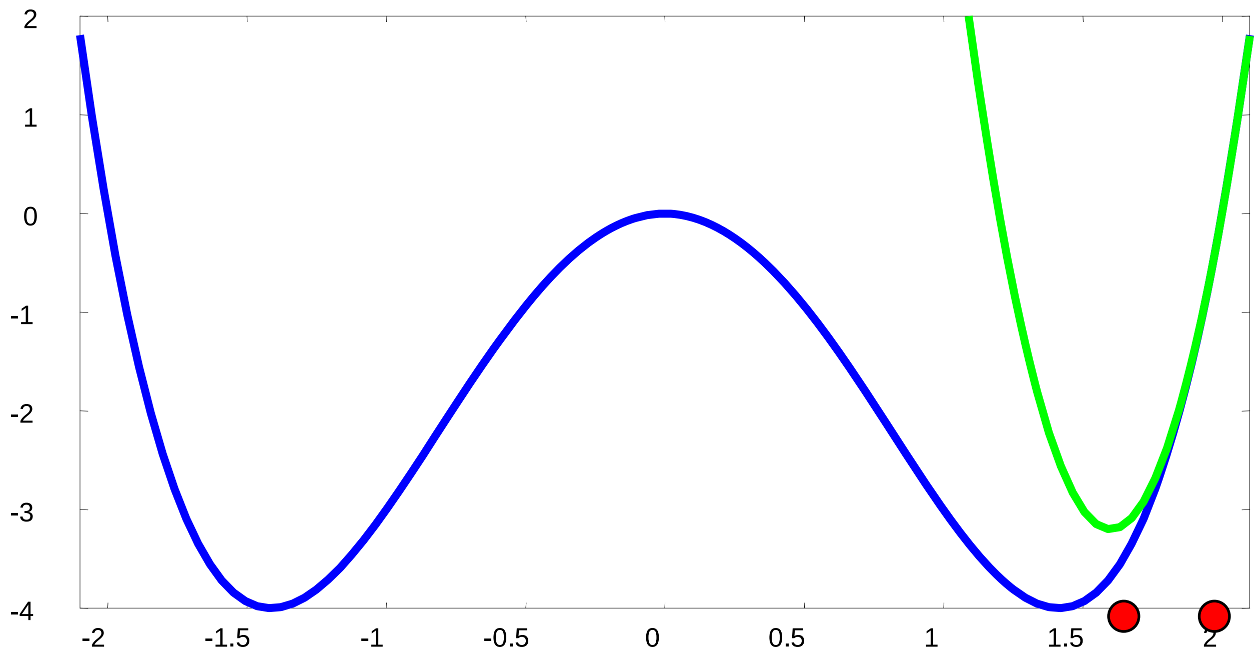
例1：一変数関数 $f(x) = x^4 - 4x^2$

初期点 $x = 2$ においてfのテイラー近似を求める

⇒ $f(2+d) \fallingdotseq 0 + 16d + (40/2)d^2$

$d = -2/5$ 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]

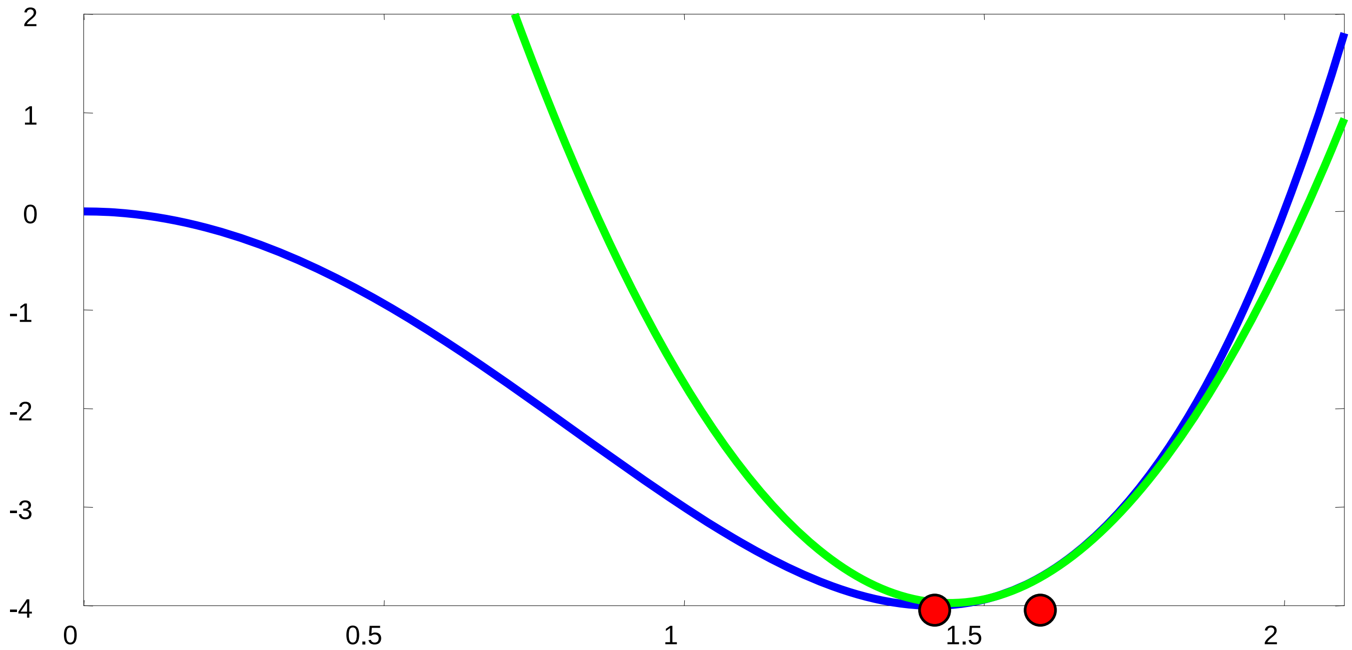
例1(続き)：一変数関数 $f(x) = x^4 - 4x^2$

点 $x = 8/5$ においてfのテイラー近似を求める

⇒ $f(8/5+d) \fallingdotseq -3.69 + 3.58d + 11.36d^2$

$d = -0.11$ 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]

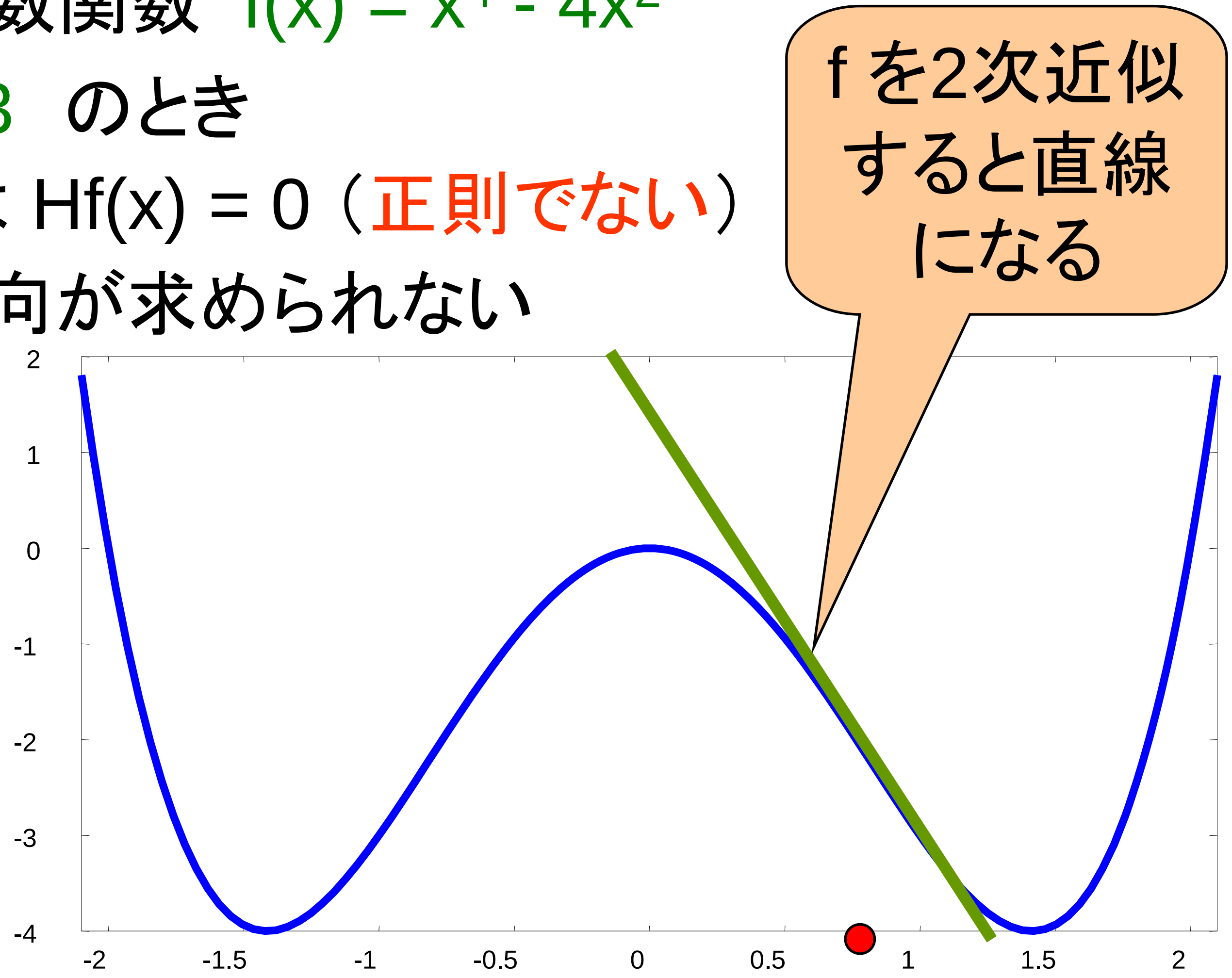
■ ヘッセ行列が正則でないと破綻

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



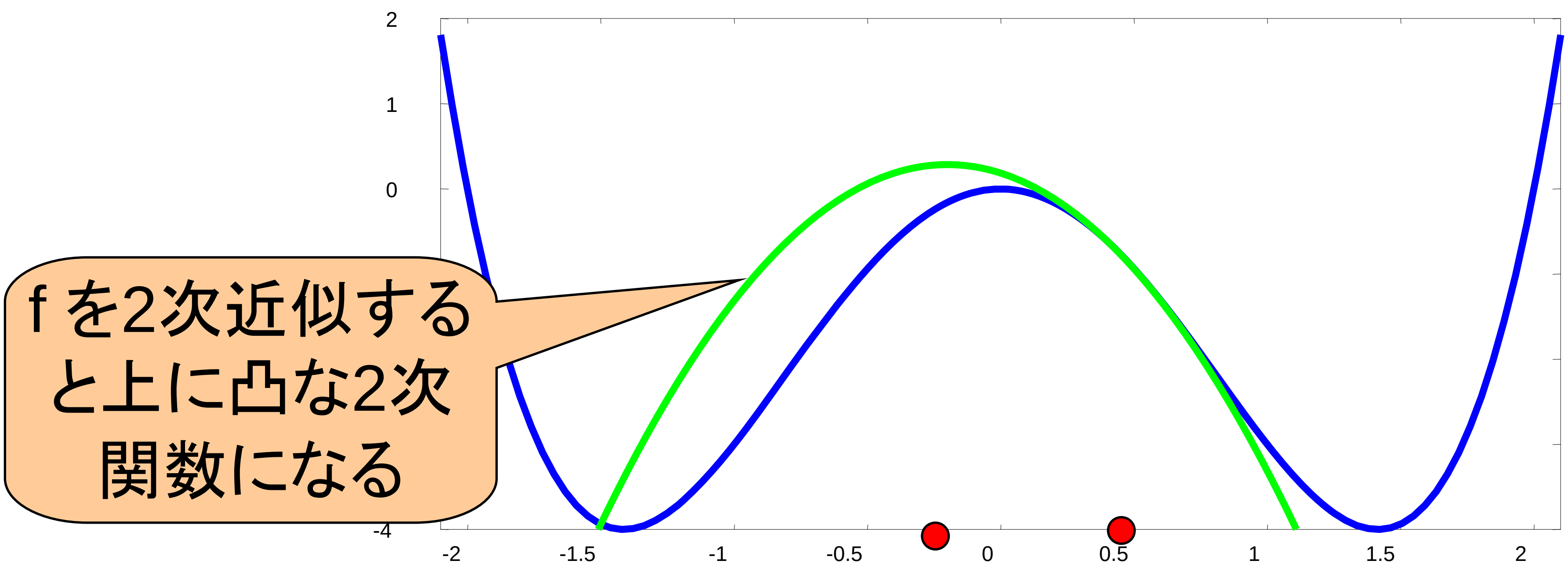
ニュートン法の問題点 [p.107]

- 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

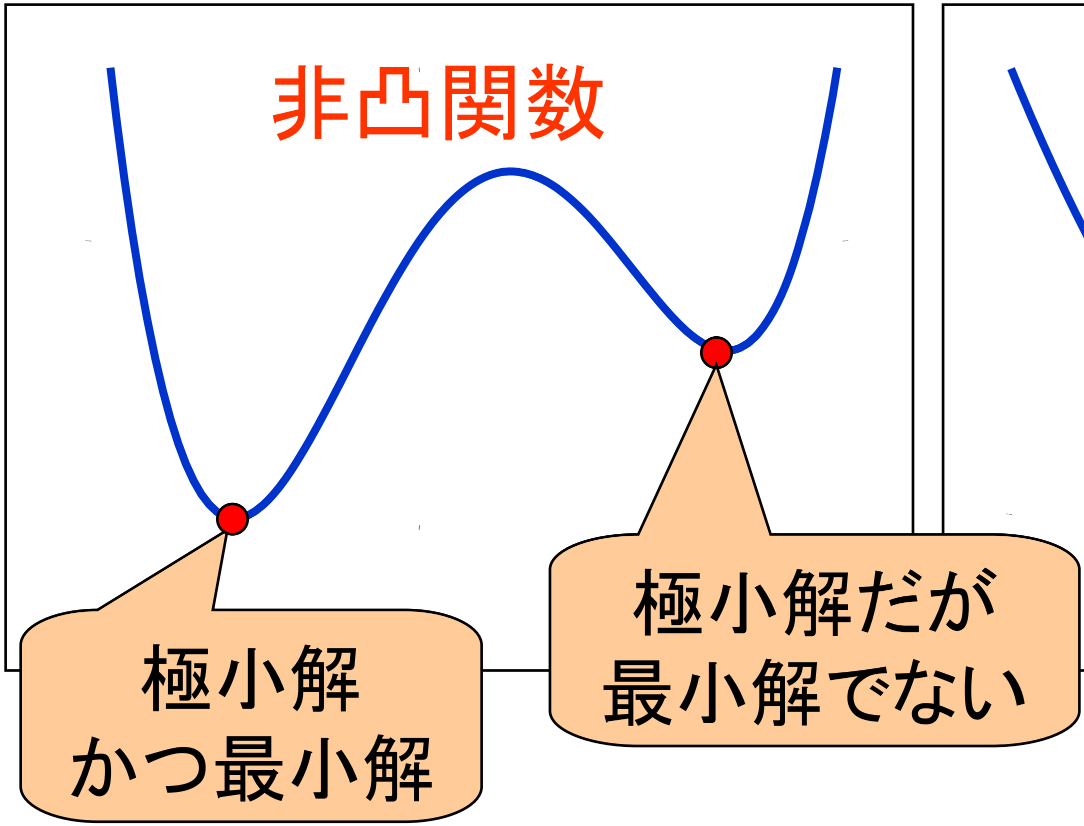
⇒ ヘッセ行列は $Hf(x) = -5$（正定値でない）

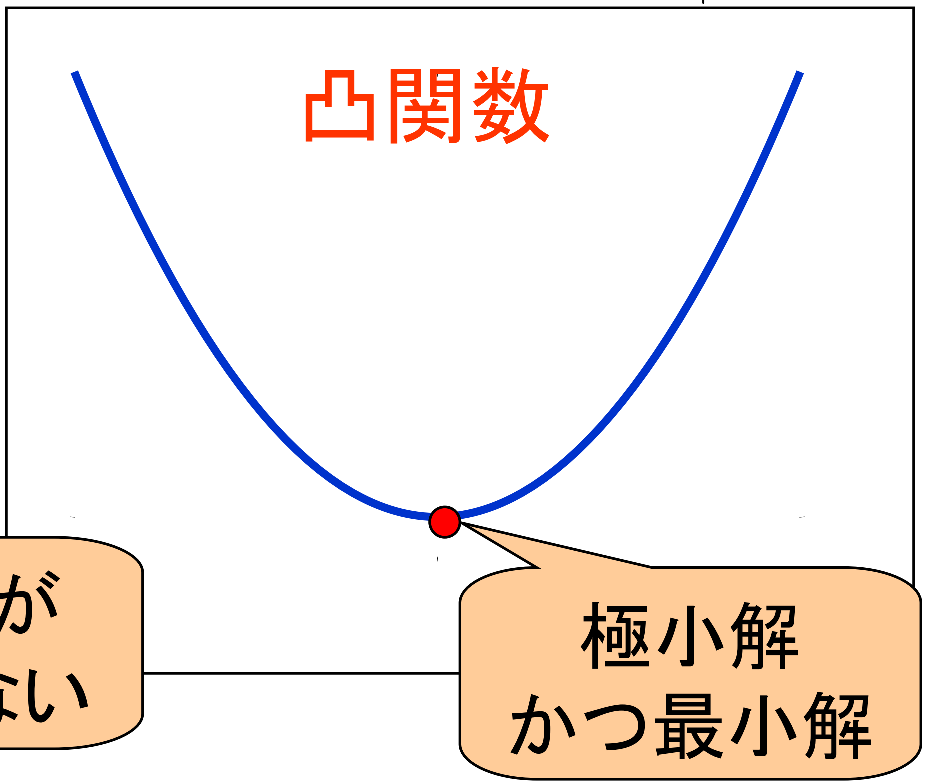
⇒ ニュートン方向に進むと関数値が増加する



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 $0 \leqq t \leqq 1$ 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

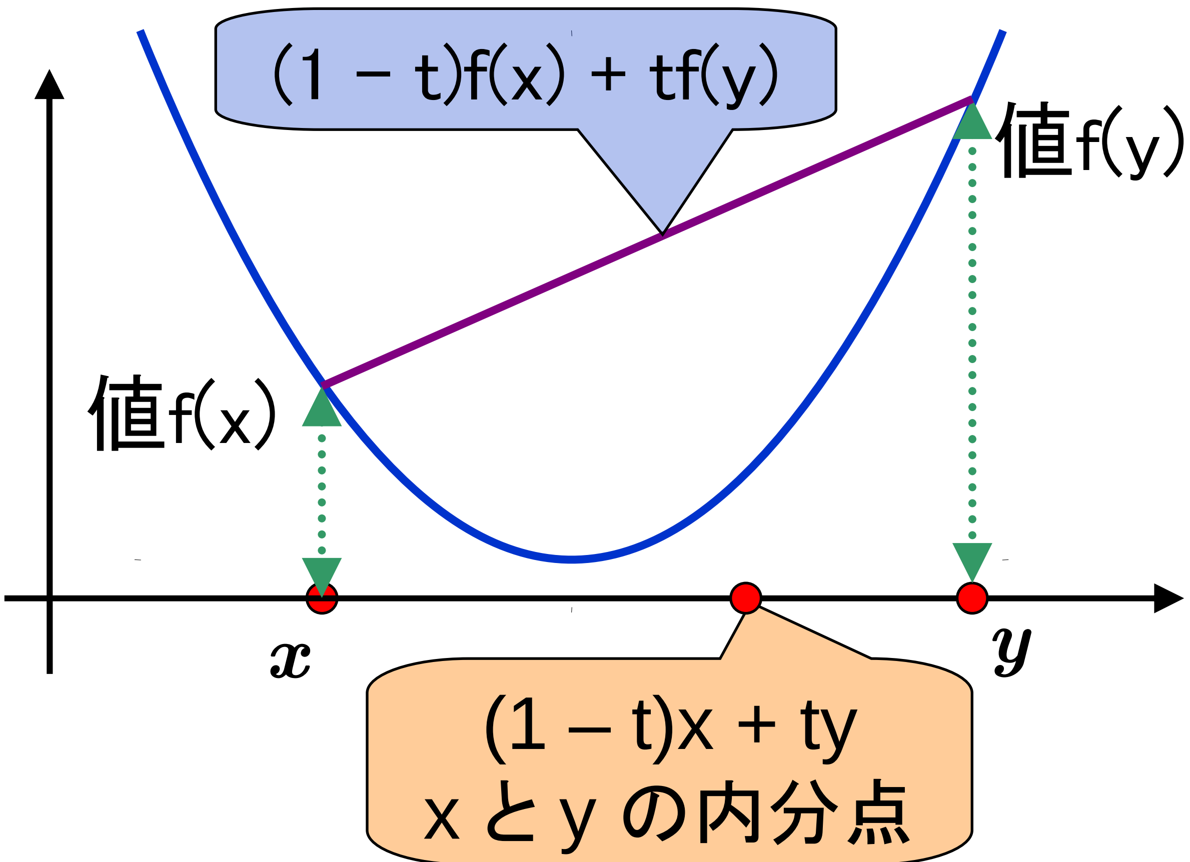
$$(1-t)\,f(x) + t\,f(y) > f((1-t)\,x + t\,y)$$

凸関数の定義(続き) [p.94]

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

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

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) ≧ 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 $\Leftrightarrow$ $(1-t)\, f(x) + t\, f(y) \geqq f((1-t)\, x + t\, y)$

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

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、

$$(1-t)\, ax^2 + t\, a\, y^2 - a\, [(1-t)\, x + t\, y]^2$$

$$= (1-t)\, ax^2 + t\, a\, y^2 - a\, (1-t)^2\, x^2 - a\, t^2\, y^2 - 2a\, (1-t)\, t\, x\, y$$

$$= (t-t^2)\, ax^2 + (t-t^2)\, a\, y^2 - 2a\, (t-t^2)\, x\, y$$

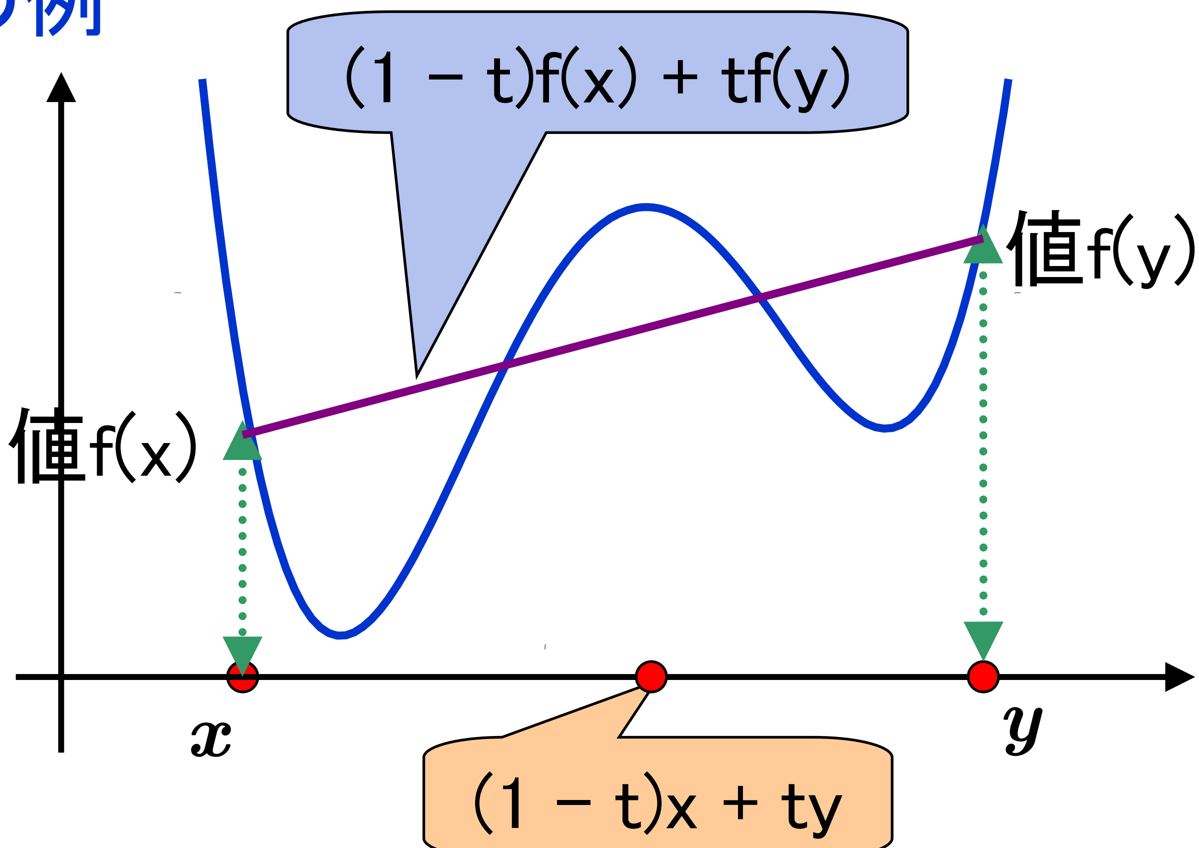
$$= (t-t^2)\, a\, (x-y)^2$$

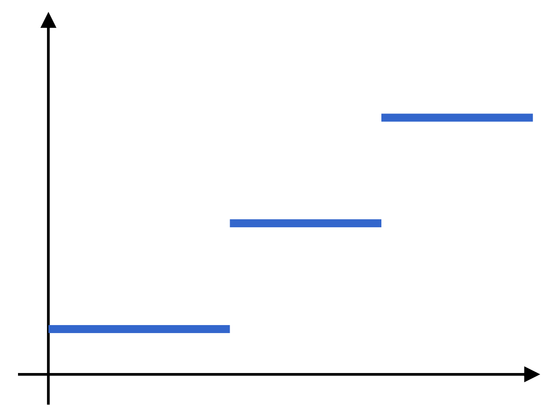
> 0 （$0 < t < 1$, $x \neq y$ より）

凸関数の定義(続き) [p.94]

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

非凸関数の例

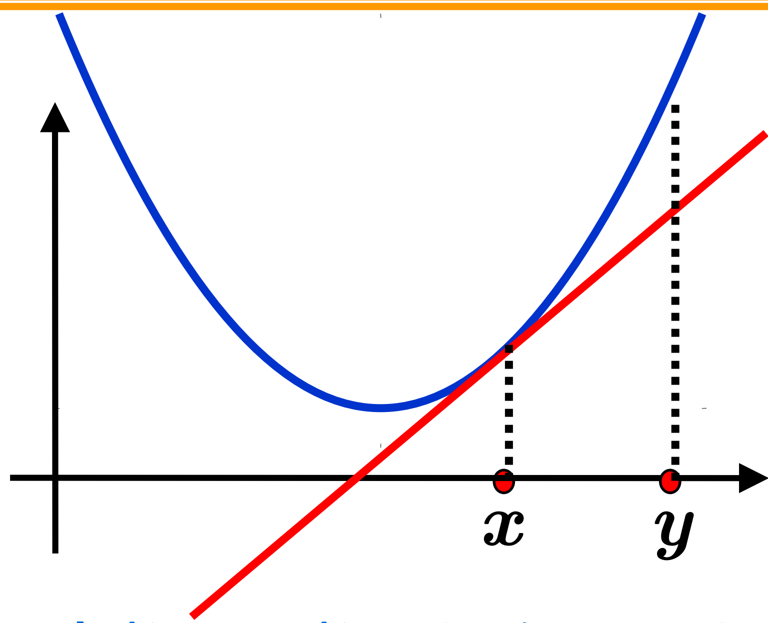




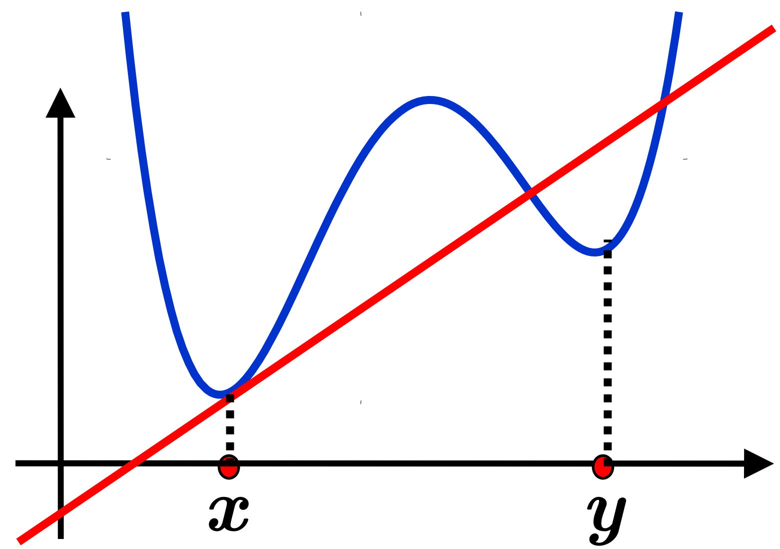
凸関数の性質 [p.95]

定理: f: 凸関数, 微分可能（勾配ベクトルが定義可能）
⇒ 任意のベクトル x, y に対して次の不等式が成立

$$f(y) \geq f(x) + \nabla f(x)(y - x)$$



一変数凸関数の場合: x における
接線 $y = f(x) + \nabla f(x)(y - x)$
より f(y) は上にある



一変数非凸関数の場合は
成り立たない

凸関数の最適解の必要条件 [p.101]

定理: f: 凸関数, 微分可能（勾配ベクトルが定義可能）
x*: f の停留点（∇f(x*)=0）
⇒ x*は制約なし問題の最適解

証明: f は凸関数なので，任意のx, y に対して次が成り立つ

$$f(y) \geq f(x) + \nabla f(x)(y - x)$$

x = x* を代入すると，∇f(x*)=0なので

$$f(y) \geq f(x^*) + \nabla f(x^*)(y - x^*) = f(x^*)$$

すなわち，任意のベクトル y の関数値より，x* の関数値は少ない（または等しい）

∴ x*は最適解

凸関数の最適解の必要条件 [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）

レポート問題
（提出は任意，締切２月５日）

問題１：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。

問題２：関数 $f(x) = |x|$ は凸関数である．これを証明せよ．また，この関数は狭義凸ではない．理由を説明せよ．

前回のレポート問題の答え

問題2:

関数 $f(x_1, x_2) = x_1^3 + 2x_1x_2^2 - 4x_2^2 + x_2^3$ について考える．

(2) 任意のベクトル (a_1, a_2) に対し，$\nabla f(a_1, a_2) = (d_1, d_2)$ とする．
このとき，$(d_1, d_2) \neq (0, 0)$ ならば，十分小さい $\delta > 0$ に対して
$f(a_1 - \delta d_1, a_2 - \delta d_2) < f(a_1, a_2)$ が成り立つことを証明せよ．

証明:

定義より $(d_1, d_2) = (3a_1^2 + 2a_2^2, 4a_1a_2 - 8a_2 + 3a_2^2)$

$f(a_1 - \delta d_1, a_2 - \delta d_2)$ を計算すると

前回のレポート問題の答え

証明の続き:

$$
\begin{aligned}
& f(a_1 - \delta d_1, a_2 - \delta d_2) \\
& = (a_1 - \delta d_1)^3 + 2(a_1 - \delta d_1)(a_2 - \delta d_2)^2 - 4(a_2 - \delta d_2)^2 + (a_2 - \delta d_2)^3 \\
& = (a_1^3 + 2a_1 a_2^2 - 4a_2^2 + a_2^3) \\
& \quad + \delta(-3a_1^2 d_1 - 2d_1 a_2^2 - 4a_1 a_2 d_2 + 8a_2 d_2 - 3a_2^2 d_2) \\
& + \delta^2(3a_1 d_1^2 + 2a_1 d_1^2 + 4d_1 a_2 d_2 - 4d_2^2 + 3a_2 d_2^2) \\
& \quad + \delta^3(-d_1^3 - 2d_1 d_2^2 - d_2^3) \\
& = f(a_1, a_2) - \delta(d_1^2 + d_2^2) \\
& \quad + \delta^2(3a_1 d_1^2 + 2a_1 d_1^2 + 4d_1 a_2 d_2 - 4d_2^2 + 3a_2 d_2^2) \\
& \quad + \delta^3(-d_1^3 - 2d_1 d_2^2 - d_2^3)
\end{aligned}
$$

δ が十分小さいと仮定すると

$$
\begin{aligned}
(d_1^2 + d_2^2)/2 > \delta(3a_1 d_1^2 + 2a_1 d_1^2 + 4d_1 a_2 d_2 - 4d_2^2 + 3a_2 d_2^2) \\
+ \delta^2(-d_1^3 - 2d_1 d_2^2 - d_2^3)
\end{aligned}
$$

よって $f(a_1 - \delta d_1, a_2 - \delta d_2) - f(a_1, a_2) < -\delta(d_1^2 + d_2^2)/2 < 0$