数理計画法（数理最適化）第13回

非線形計画

ニュートン法，凸関数

担当：　塩浦昭義

(情報科学研究科　准教授)

shioura@dais.is.tohoku.ac.jp

行列の正定値性、半正定値性

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 ⇔ $a_{11} > 0$

2次の最適性条件（必要条件）

ヘッセ行列を用いた最適性条件

> 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**

例:

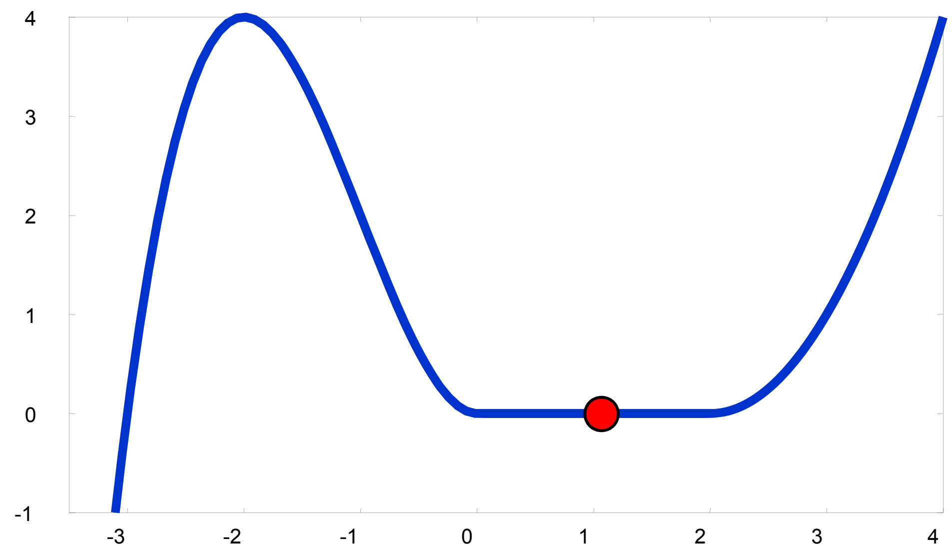
$x^* = 1$ は極小解

$0 \leqq x \leqq 2$ の範囲で $f(x) = 0$

⇒ $\nabla f(x^*) = f'(x^*) = 0$

$Hf(x^*) = f''(x^*) = 0$

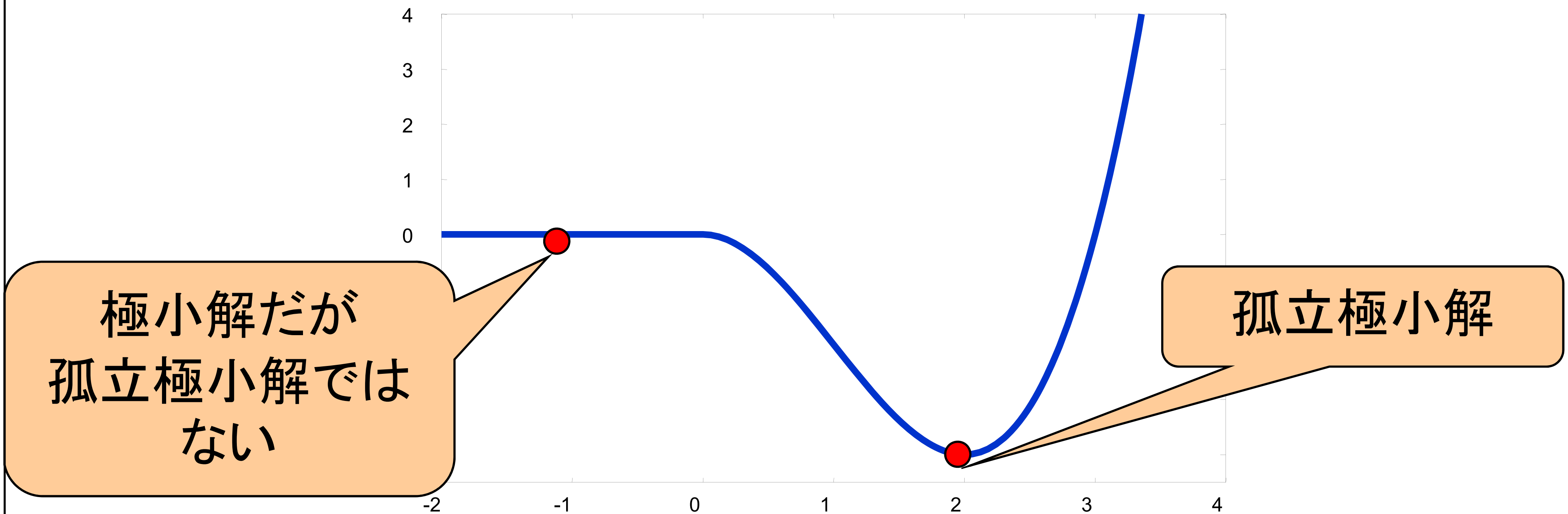
半正定値



2次の最適性条件(十分条件)

定理(2次の十分条件):
x* は停留点, Hf(x*) は正定値
⇒ x*: 制約なし問題の(孤立)極小解

定義:x* は孤立極小解
⇔ x* は極小、近傍内に同じ関数値をもつ点が存在しない



2次の最適性条件(十分条件)の例

定理: x* は停留点, Hf(x*) は正定値
⇒ x*: (孤立)極小解

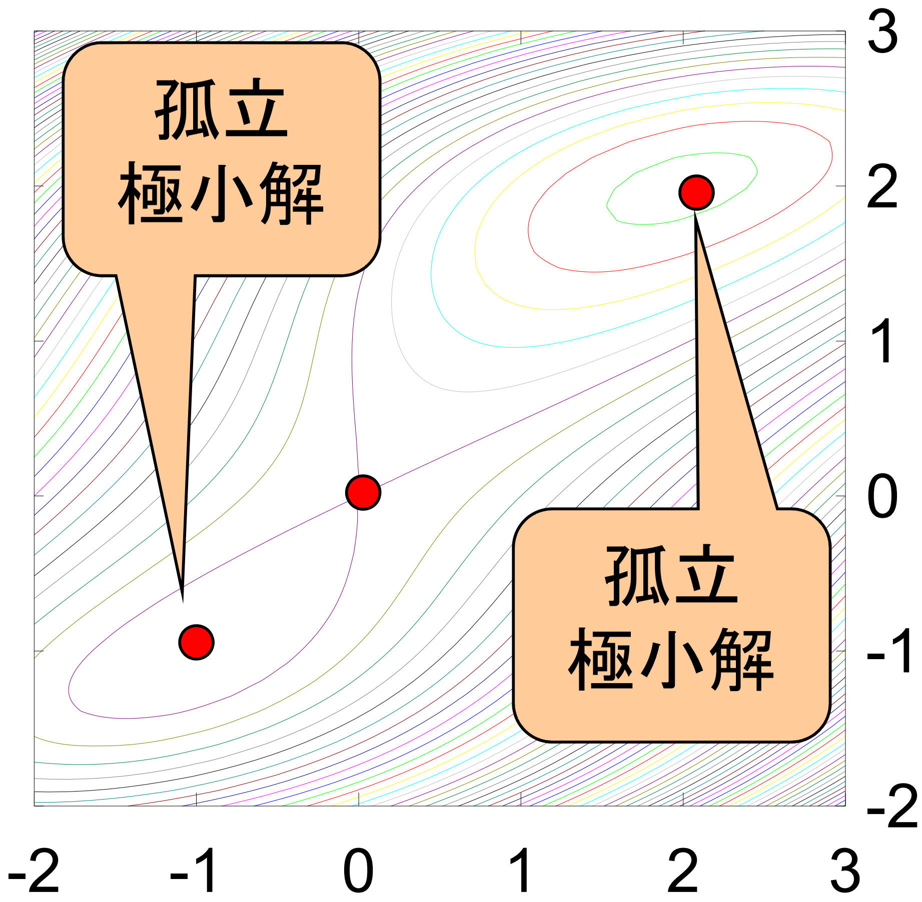
例2 $$f(\boldsymbol{x}) = x_1^2 - 2x_1x_2 + \frac{1}{4}x_2^4 - \frac{1}{3}x_2^3$$

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

⇒ 停留点は(0,0), (-1, -1), (2, 2)

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

⇒ (-1, -1), (2, 2) は孤立極小解

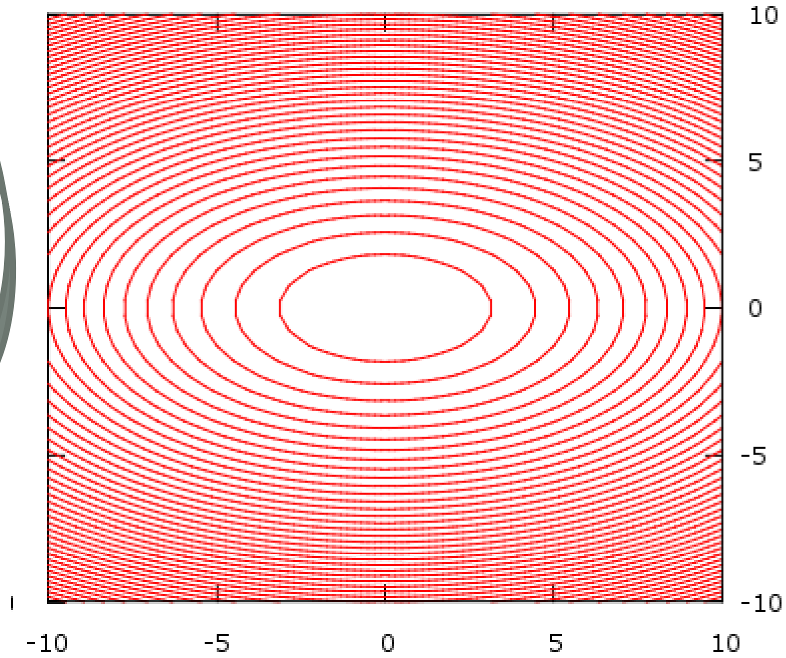


2次の最適性条件の例

例3 : $f(x_1, x_2) = x_1^2 + 3x_2^2$

- $\nabla f(x_1, x_2) = \begin{bmatrix} 2x_1 \\ 6x_2 \end{bmatrix}, \mathrm{H}f(x_1, x_2) = \begin{bmatrix} 2 & 0 \\ 0 & 6 \end{bmatrix}$
- $\nabla f(x_1, x_2)$がゼロベクトルとなるのは (0,0) のみ←停留点
- $\nabla^2 f(0,0) = \begin{bmatrix} 2 & 0 \\ 0 & 6 \end{bmatrix}$ は正定値行列

  ➔(0, 0) は孤立極小解



任意の非ゼロベクトル(y_1, y_2)に対して

$$\begin{bmatrix} y_1 & y_2 \end{bmatrix} \begin{bmatrix} 2 & 0 \\ 0 & 6 \end{bmatrix} \begin{bmatrix} y_1 \\ y_2 \end{bmatrix} = 2y_1^2 + 6y_2^2 > 0$$

2次の最適性条件の例

例4 : $f(x_1, x_2) = x_1^2 + x_2^4$

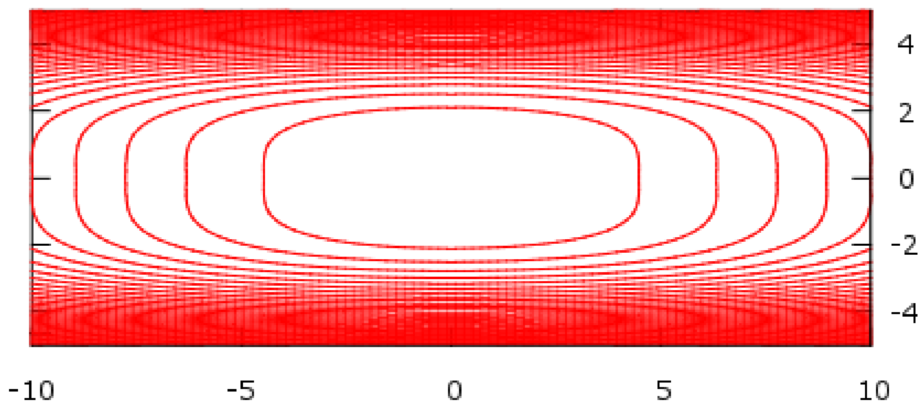
- $\nabla f(x_1, x_2) = \begin{bmatrix} 2x_1 \\ 4x_2^3 \end{bmatrix}, \nabla^2 f(x_1, x_2) = \begin{bmatrix} 2 & 0 \\ 0 & 12x_2^2 \end{bmatrix}$
- $\nabla f(x_1, x_2)$がゼロベクトルとなるのは (0,0) のみ（実は最適解）
- $\nabla^2 f(0,0) = \begin{bmatrix} 2 & 0 \\ 0 & 0 \end{bmatrix}$ は半正定値だが，正定値ではない

➔(0, 0) が極小解かどうかは，ヘッセ行列を使って判定できない（実際には極小解）

任意のベクトル(y_1, y_2)に対して

$$[y_1 \quad y_2] \begin{bmatrix} 2 & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} y_1 \\ y_2 \end{bmatrix} = 2y_1^2 \geq 0$$

$y_1 = 0$のときは$y_2 \neq 0$でも値は0



極大解に関する性質

- x^* は関数 f の（孤立）極大解
  $\Leftrightarrow$ 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:
x^*: 制約なし問題の極大解 $\Rightarrow$ $-Hf(x^*)$ は半正定値

定理:
x^* は停留点，$-Hf(x^*)$ は正定値
$\Rightarrow$ x^*: 制約なし問題の（孤立）極大解

制約なし問題の解法2：ニュートン法

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

制約なし問題の解法2:ニュートン法

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!

ただし，一般の関数は狭義2次凸とは限らない

⇒ 元の関数 f の代わりに，二次のテイラー近似 $\tilde{f}$ を使う

$$\tilde{f}(x) = f(a) + \nabla f(a)^T(x-a) + (x-a)^T \mathrm{H} f(a)(x-a)$$

- ヘッセ行列 Hf(x) が正定値のとき，
  $\tilde{f}$ の最適解は $x = a - Hf(x)^{-1}\nabla f(x)$
- $\tilde{f}$ は f の良い近似

⇒ $a - Hf(x)^{-1}\nabla f(x)$ 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

現在の点 x を $x - Hf(x)^{-1}\nabla f(x)$ へ移動させることを繰り返す
($-Hf(x)^{-1}\nabla f(x)$ を, xにおけるニュートン方向と呼ぶ)

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

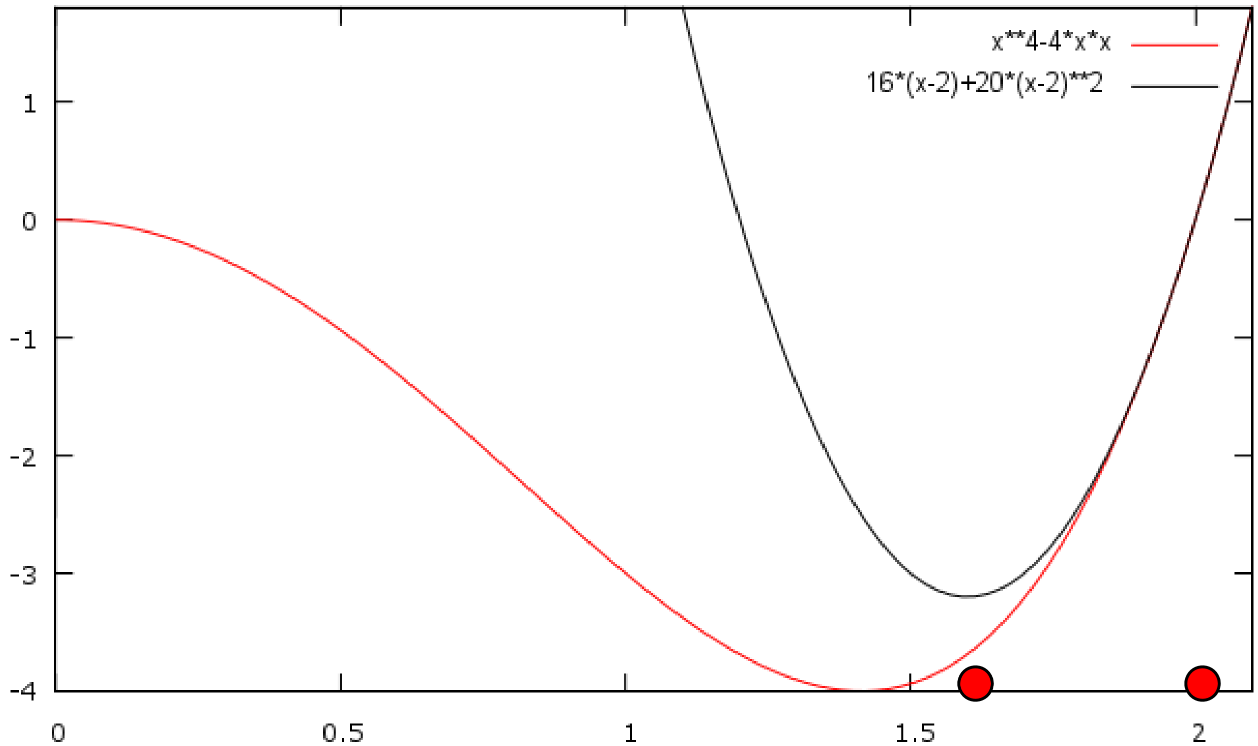
ステップ2：ニュートン方向 $-Hf(x^k)^{-1}\nabla f(x^k)$ を計算

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

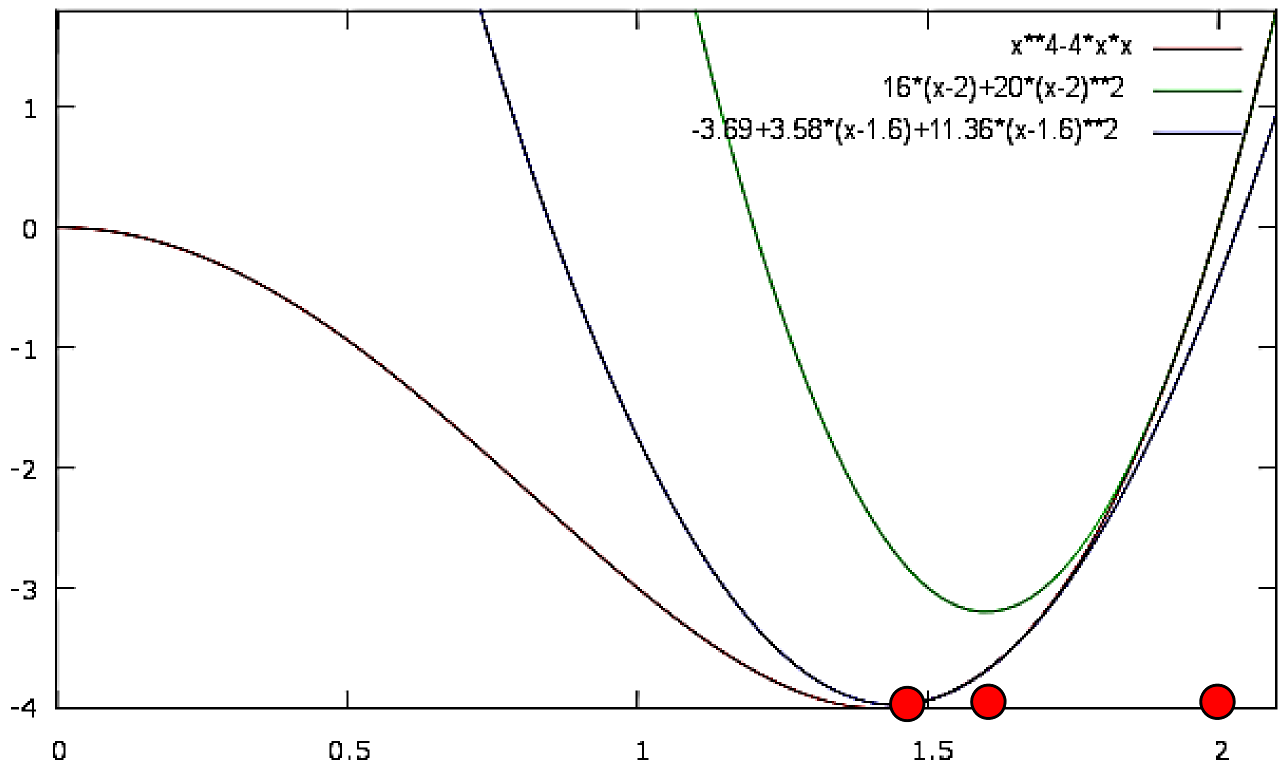
ニュートン法の実行例その1

- 一変数関数 $f(x) = x^4 - 4x^2$
- 初期点 $x^{(0)} = 2$
- テイラー近似は$\tilde{f}(x) = 16(x-2) + 20(x-2)^2$
- これが最小になるのは$x = 2 - 0.4 = 1.6$のとき
- $x^{(1)} := 1.6$とおく



ニュートン法の実行例その1

- 一変数関数 $f(x) = x^4 - 4x^2$
- 点 $x^{(1)} = 1.6$
- テイラー近似は$\tilde{f}(x) = -3.69 + 3.58(x - 1.6) + 11.36(x - 1.6)^2$
- これが最小になるのは$x = 1.6 - 0.11 = 1.49$のとき
- $x^{(2)} := 1.49$とおく



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

ニュートン法の例2

- 関数 $f(x) = (x_1 - 1)^2 + 10(x_1^2 - x_2)^2$ に適用
  - 初期解(0,0), 最適解は(1,1)
  - 6回の反復で最適解に到達
    - 最急降下法では100回反復後でも(0.91, 0.82)

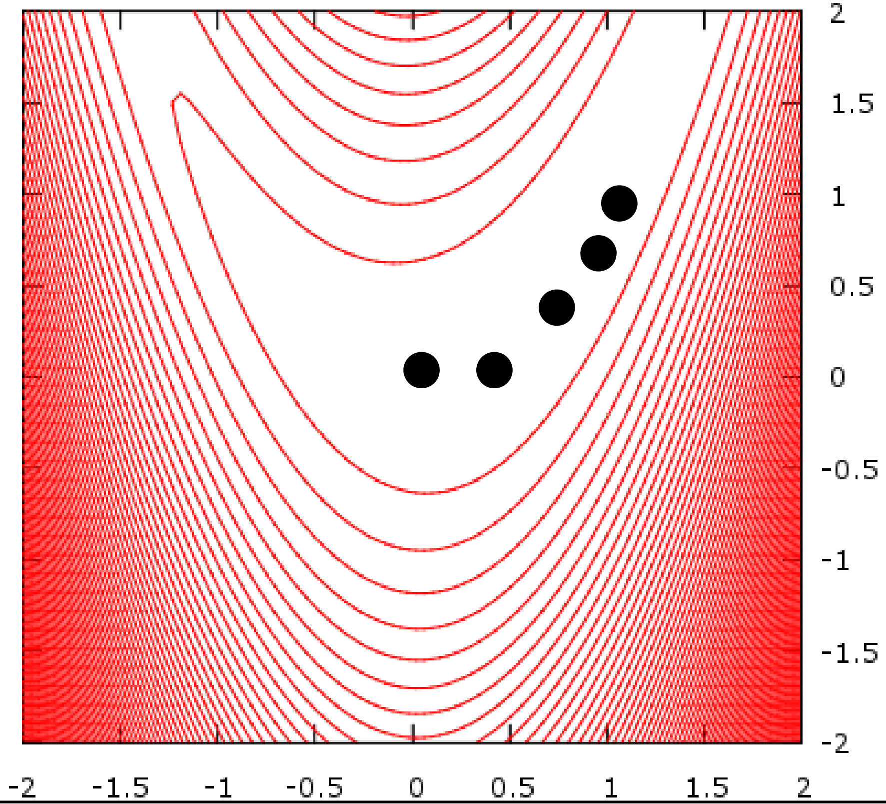


表 4.2　関数 (4.19) に対するニュートン法の計算結果

反復 k	$\boldsymbol{x}^{(k)}$	$f(\boldsymbol{x}^{(k)})$	$\|\nabla f(\boldsymbol{x}^{(k)})\|$
0	(0.00000, 0.00000)	0.10000×10^{1}	0.20000×10^{1}
1	(0.32341, 0.00000)	0.56717×10^{0}	0.20919×10^{1}
2	(0.73455, 0.46247)	0.12990×10^{0}	0.23209×10^{1}
3	(0.91297, 0.85632)	0.12775×10^{-1}	0.11054×10^{1}
4	(1.00450, 1.01041)	0.39429×10^{-4}	0.54177×10^{-1}
5	(0.99997, 0.99995)	0.16624×10^{-8}	0.46482×10^{-3}
6	(1.00000, 1.00000)	0.39340×10^{-17}	0.17062×10^{-7}

福島雅夫
「新版 数理計画入門」
（朝倉書店）より

ニュートン法の問題点

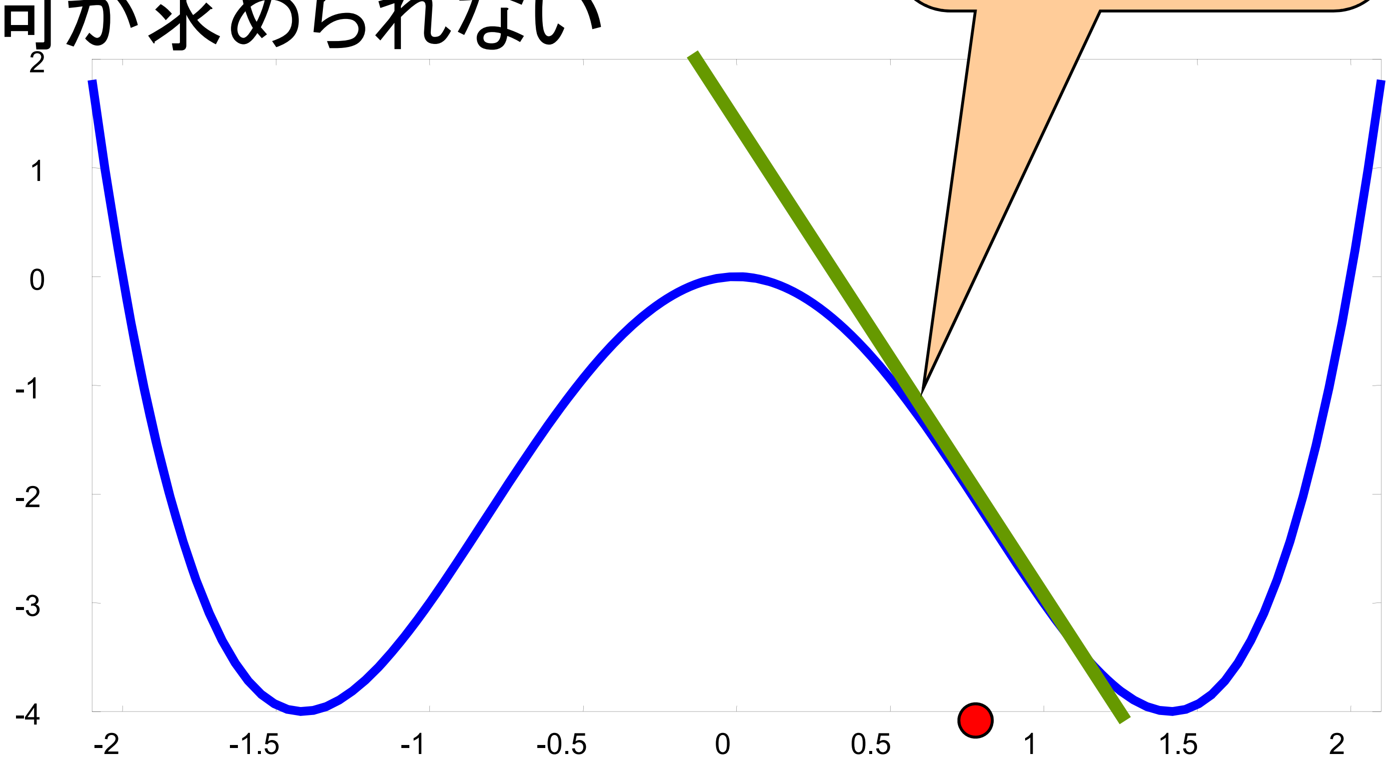
■ ヘッセ行列が正則でないと破綻

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



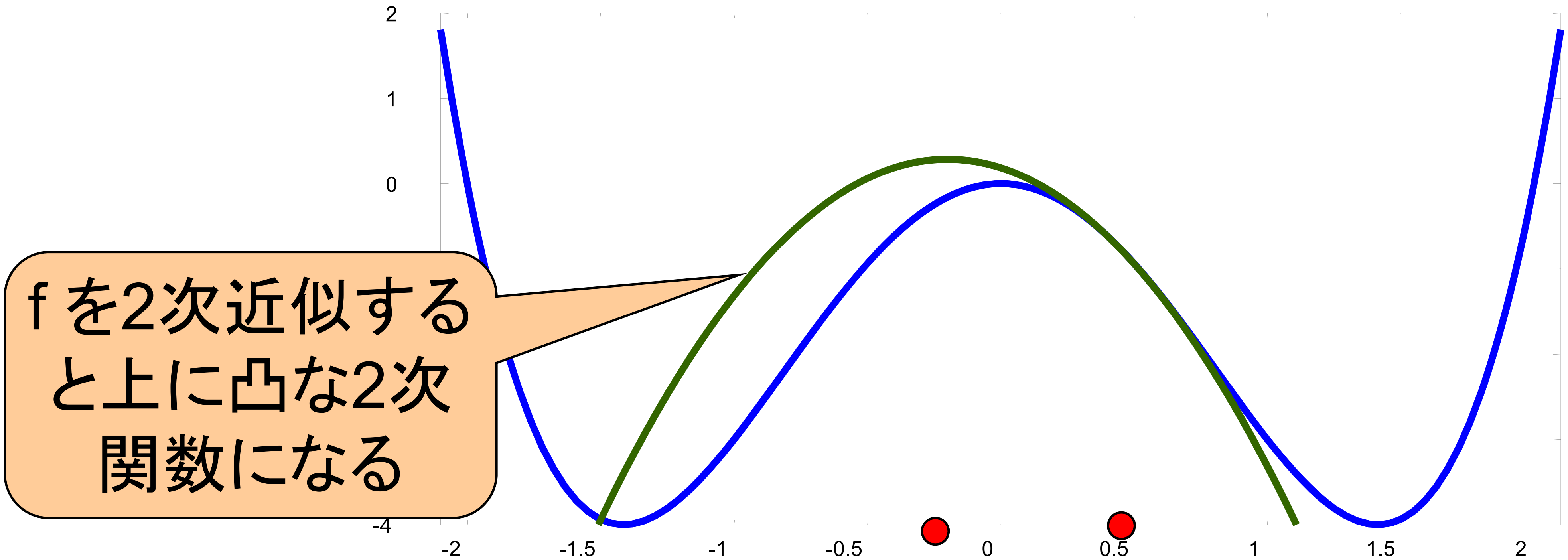
ニュートン法の問題点

- 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

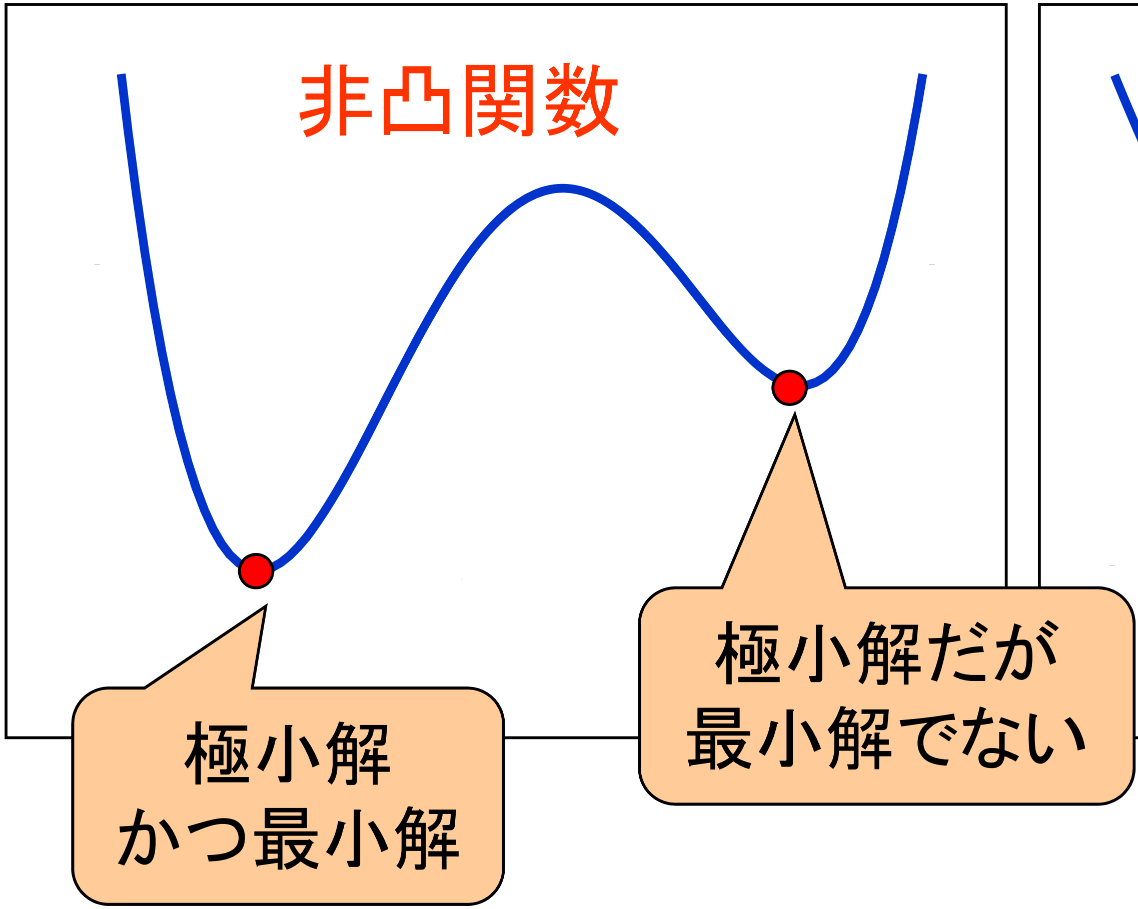
⇒ ヘッセ行列は $Hf(x) = -5$（正定値でない）

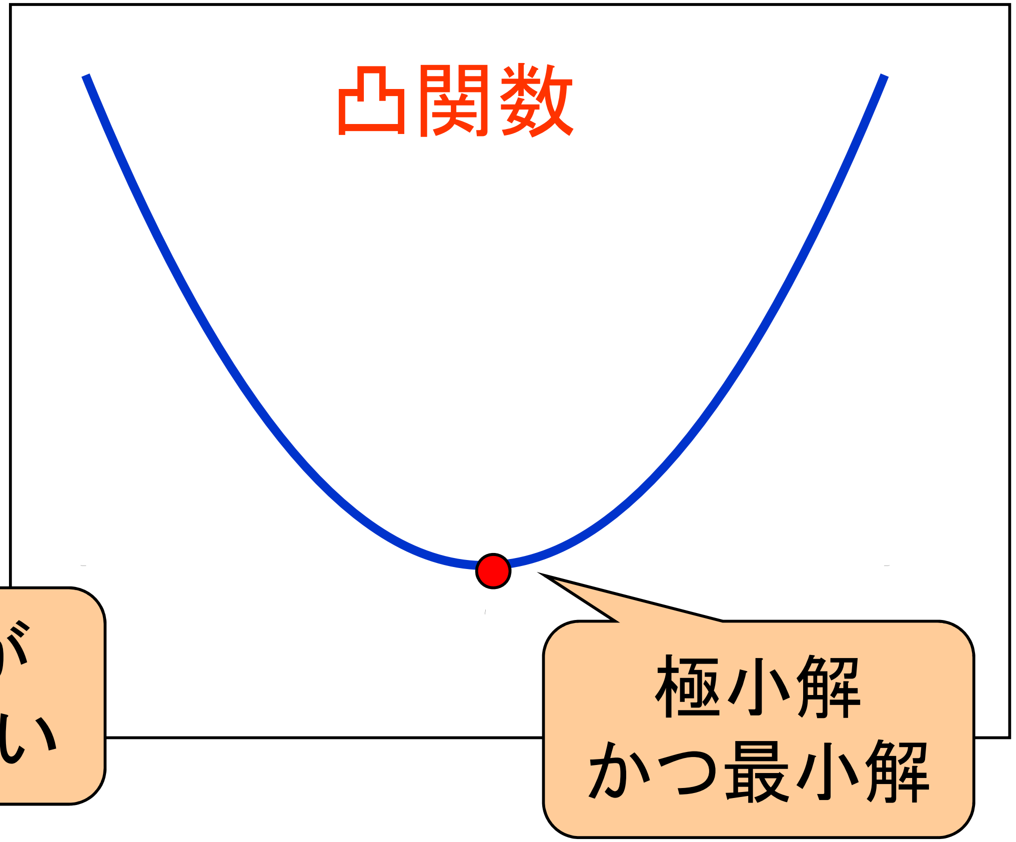
⇒ ニュートン方向に進むと関数値が増加する



凸関数

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

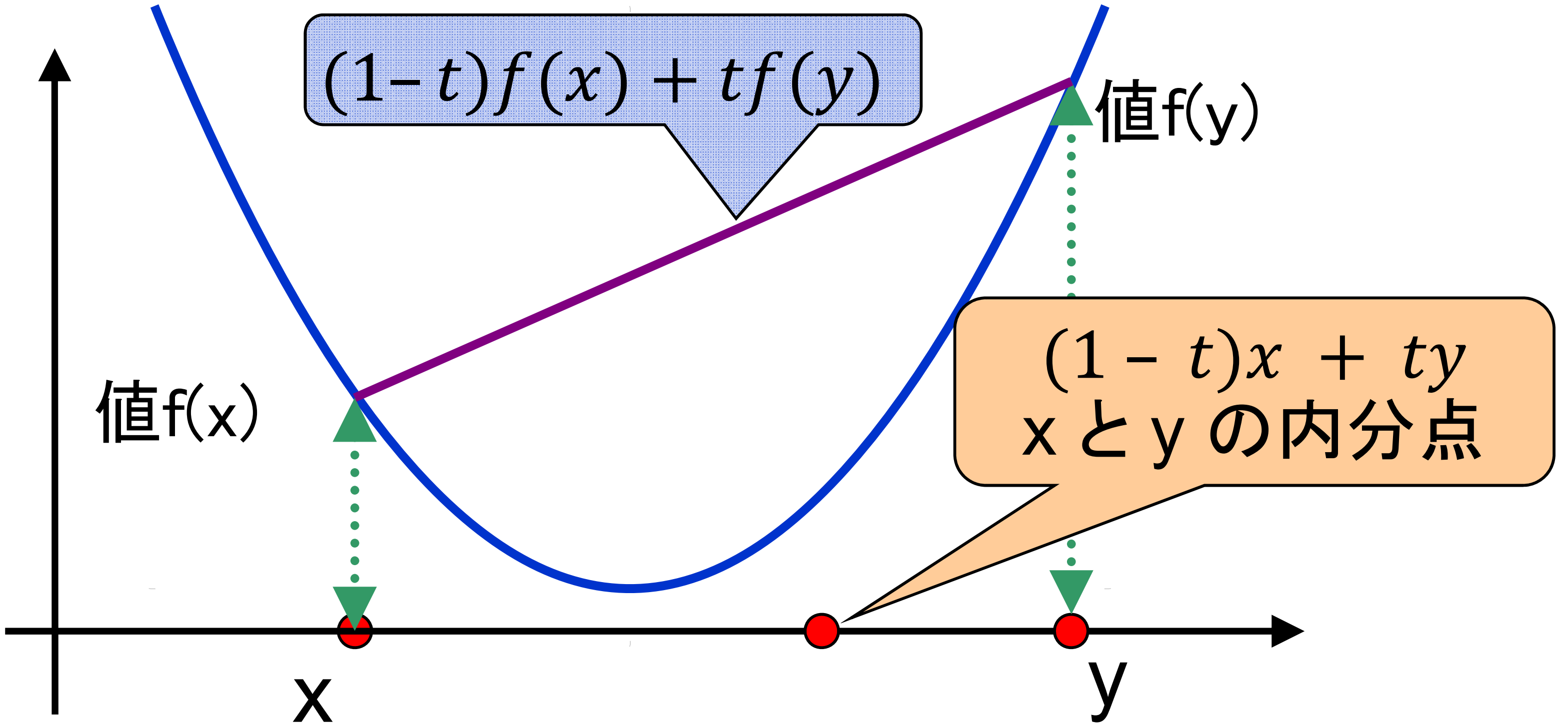
極小解が一つ
➔最小化しやすい

凸関数の定義

定義：関数 f は凸関数

⇔　任意の異なるベクトル x, y および任意の $0 < t < 1$ に対し

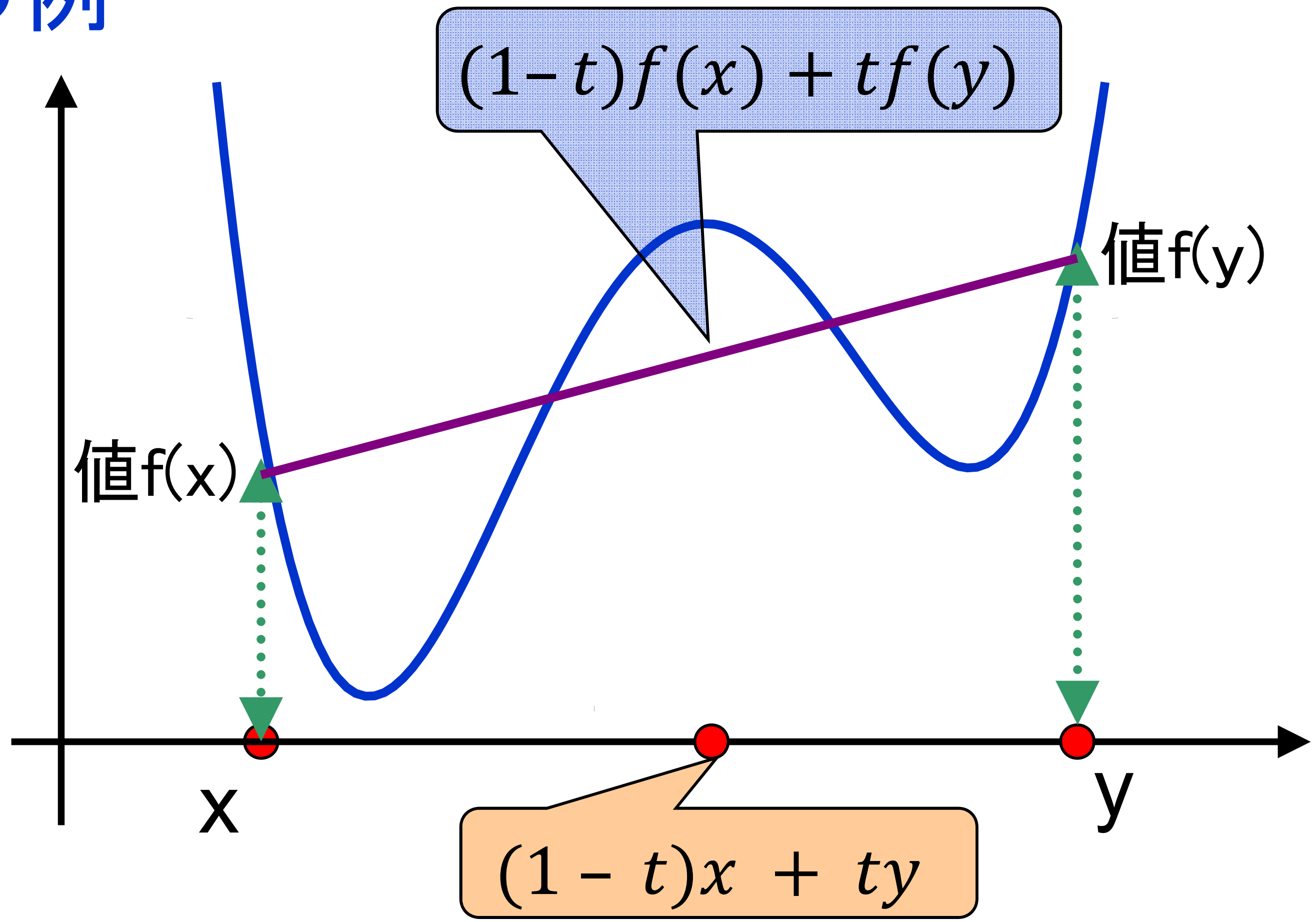
$$(1-t)f(x) + t\,f(y) \geq f((1-t)\,x + t\,y)$$

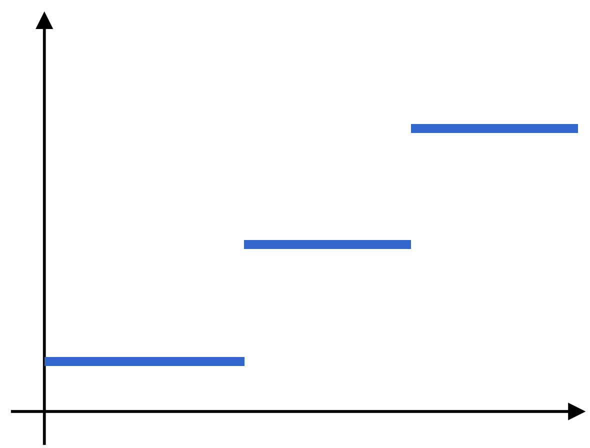


凸関数の定義(続き)

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

非凸関数の例





2次の凸関数

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

2次関数 $f(x) = a\,x^2 + bx + c\ (a > 0)$ は凸関数

（証明）任意の異なる x, y と $0 < t < 1$ に対して、

$(1-\mathrm{t})f(x) + tf(y) - f((1-t)x + ty)$

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

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

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

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

$> 0 \qquad (0 < t < 1, x \neq y$ より$)$

2次の凸関数(続き)

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

より一般に,

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

(V: n ×n 行列, c: n次元ベクトル, c_0: 定数)

は V が半正定値行列 ➔ 凸関数

例:

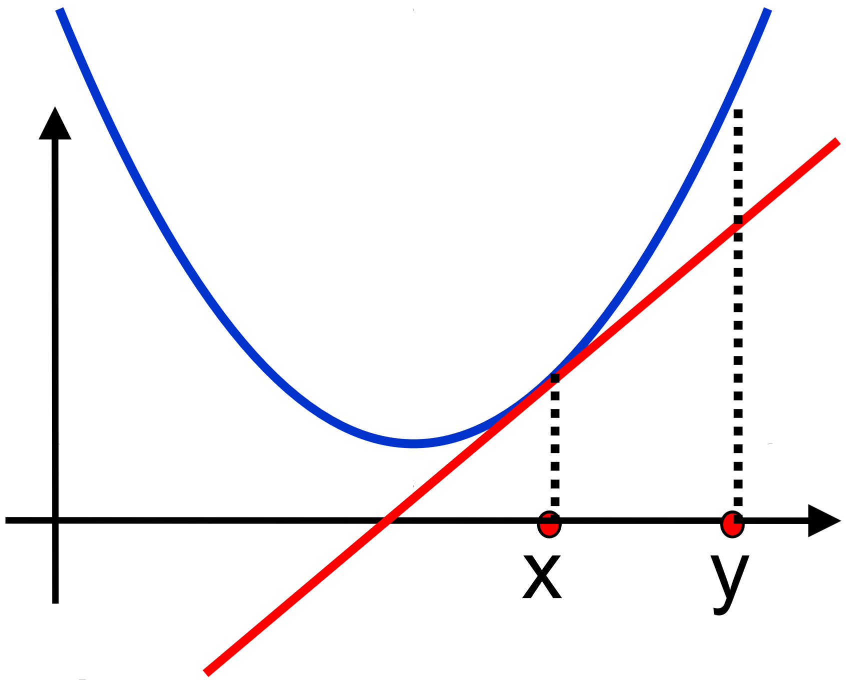
$$f(x_1, x_2) = \frac{1}{2}\begin{pmatrix} x_1 \\ x_2 \end{pmatrix}^T \begin{pmatrix} 2 & -2 \\ -2 & 5 \end{pmatrix}\begin{pmatrix} x_1 \\ x_2 \end{pmatrix}$$

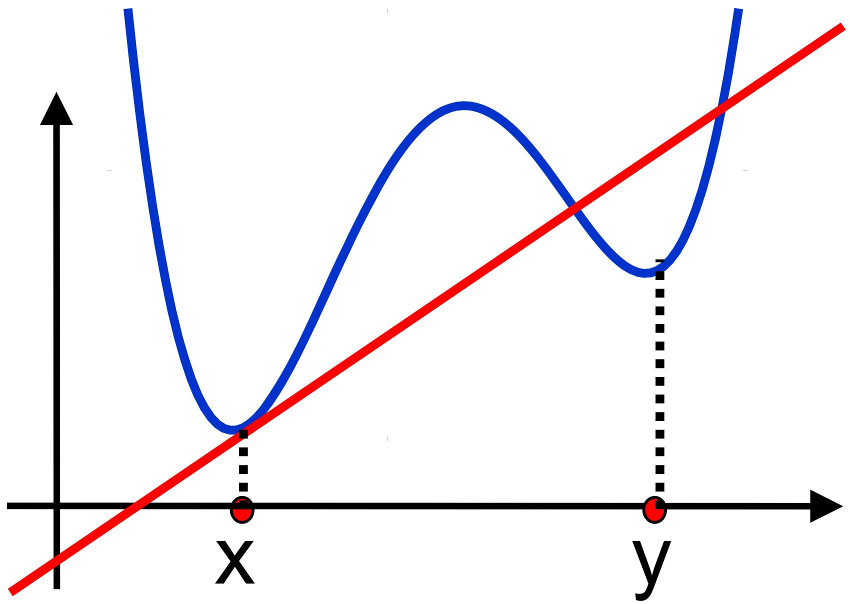
凸関数の特徴付け（その1）

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

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

証明は略





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

凸関数の特徴付け(その2)

定理: f: 凸関数, 微分可能(ヘッセ行列が定義可能)
←→ 任意のベクトル x に対して
ヘッセ行列 $Hf(x)$ が半正定値

証明は略

一変数凸関数の場合:
関数 f は凸関数←→任意のx に対して二階微分$f''(x) \geq 0$

凸関数の最適解の必要条件

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（$\nabla f(x^*)=0$）
⇒ 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，$\nabla f(x^*)=0$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

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）