オペレーションズ・リサーチ基礎

非線形計画：二次の最適性条件とニュートン法

塩浦昭義
東京工業大学 経営工学系
shioura.a.aa@m.titech.ac.jp

関数のヘッセ行列

- 定義：n変数関数 f のヘッセ行列 $\mathrm{H}f(x)$

  ⇔ f の2次偏微分係数を要素とするｎｘｎ 行列

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

- f が1変数関数のときは，ヘッセ行列は2階微分と同じ

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

ヘッセ行列の例

例：

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

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

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

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

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

二次のテイラー展開

任意の関数 f はベクトル $a \in \mathbb{R}^n$ を使って
次の形に表現できる

関数$f(x)$の$x = a$における
二次のテイラー展開

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

関数$\psi(d)$は $d = (d_1, d_2, \dots, d_n)$ に関する
3次以上の項から構成される n 変数多項式関数
（定数項，一次の項，二次の項は含まれない）
➔任意のベクトル d に対して

$$\lim_{\varepsilon \to 0} \frac{\varphi(\varepsilon d)}{\varepsilon^2} = 0$$

二次のテイラー近似

関数$f(x)$の$x=a$における二次のテイラー展開

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

$x \simeq a$ のとき, $\psi(x-a)$の値は他の項に比べて
十分小さい（0に近い） ➔無視できる

$$\tilde{f}(x) = f(a) + \nabla f(a)^T(x-a) + \frac{1}{2}(x-a)^T \mathrm{H}f(a)(x-a)$$

関数$f(x)$の$x=a$における
二次のテイラー近似

- 二次関数
  $\nabla \tilde{f}(a) = \nabla f(a), \mathrm{H}\tilde{f}(a) = \mathrm{H}f(a)$
- $x \simeq a$のとき $\tilde{f}(x) \simeq f(x)$,
  とくに $\tilde{f}(a) = f(a)$

二次のテイラー近似の例

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

$$
\begin{aligned}
f_1(x) &= a^2 + (2a)(x-a) + \frac{1}{2}\cdot 2(x-a)^2 + \psi(x-a) \\
&= x^2 + \psi(x-a)
\end{aligned}
$$

つまり，$\psi(x=a)=0$であり，

f_1の二次のテイラー近似 $=$ f_1そのもの

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

の二次のテイラー近似は f に一致

V: $n \times n$行列
c: n 次元ベクトル
c_0: 定数

二次のテイラー近似の例

例2： $f_2(x) = \log x \Rightarrow f'(x) = \frac{1}{x}, f''(x) = -\frac{1}{x^2}$

f_2 の x=1 における二次のテイラー展開

$$f_2(x) = \log 1 + \frac{1}{1}(x-1) - \frac{1}{1^2}(x-1)^2 + \psi(x-a)$$

なお，$\psi(x-a) = \frac{(x-1)^3}{3} - \frac{(x-1)^4}{4} + \frac{(x-1)^5}{5} - \cdots$

二次のテイラー近似の例

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

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

a = -1 のとき

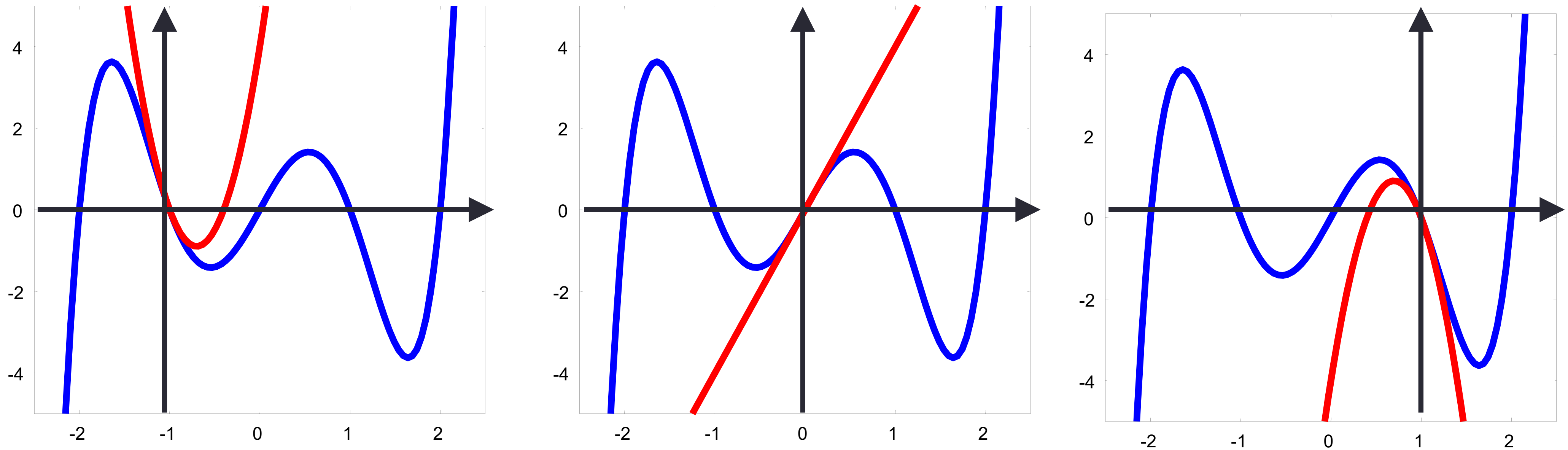
$$0 - 6(x+1) + 5(x+1)^2$$

a = 0 のとき

$$0 + 4x + 0x^2$$

a = 1 のとき

$$0 - 6(x-1) - 5(x-1)^2$$



極小解，極大解の判定方法

- 一変数関数 f の場場合
  - 極小，極大ならば傾き（一回微分）$f'(x)$ が 0
  - 極小，極大は二回微分 $f''(x)$ を使って判定
    - $f''(x) > 0$ ならば極小，$f''(x) < 0$ ならば極大
    - $f''(x) = 0$ の場合は不明
- 多変数関数の場合
  - 極小，極大ならば勾配ベクトル $f'(x)$ がゼロベクトル
  - 極小，極大はヘッセ行列 $\mathrm{H}f(x)$ を使って判定
    - どうやって判定？ --- 行列の（半）正定値性を使う

正定値（半正定値）・・・・行列が「正（非負）」

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 ⇔ $a_{11} > 0$

$\begin{pmatrix} 2 & -2 \\ -2 & 0 \end{pmatrix}$

※ A が2×2対称行列のとき、

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

半正定値
ではない

行列の正定値性、半正定値性

※ A が2×2対称行列のとき、

Aは正定値 ⇔ $a_{11} > 0$, $a_{22} > 0$, $a_{11}a_{22} - a_{12}{}^2 > 0$

$a_{11} \neq 0$ のとき　$y^T A y = a_{11}\left(y_1 + \frac{a_{12}}{a_{11}} y_2\right)^2 + \left(a_{22} - \frac{a_{12}^2}{a_{11}}\right) y_2^2$　(☆)

であることを使うと証明できる

[➔の証明] y^TAy に $y=(1,0), (0,1), (-a_{12}/a_{11}, 1)$ を代入すれば得られる

[🡐の証明] $y \neq 0$ より $\left(y_1 + \frac{a_{12}}{a_{11}} y_2\right) \neq 0$ または $y_2 \neq 0$

$\therefore \left(y_1 + \frac{a_{12}}{a_{11}} y_2\right)^2 > 0$ または $y_2^2 > 0$

上記の等式(☆)および仮定した不等式条件より, $y^TAy > 0$

2次の凸関数（再掲）

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：

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

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:

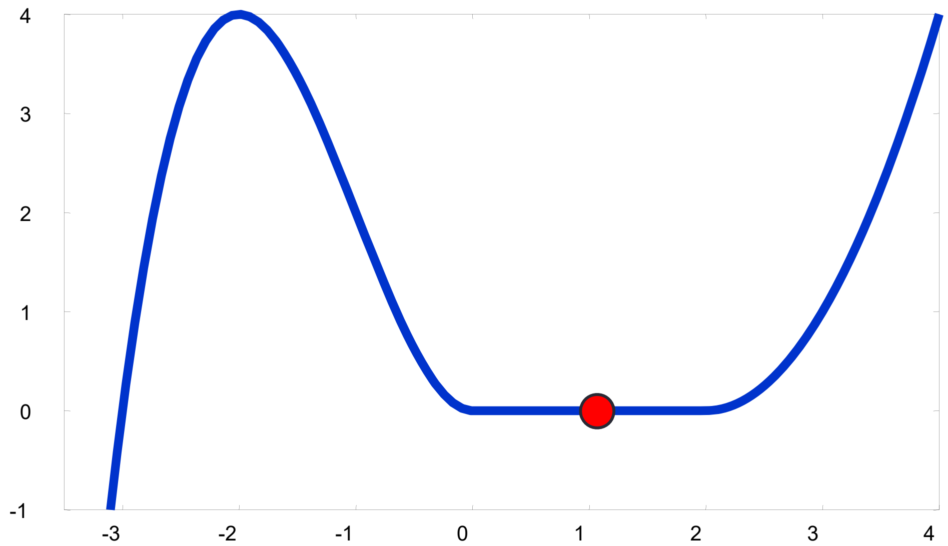
$x^* = 1$ は極小解

$0 \leqq x \leqq 2$ の範囲で $f(x) = 0$

⇒ $\nabla f(x^*) = f'(x^*) = 0$

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

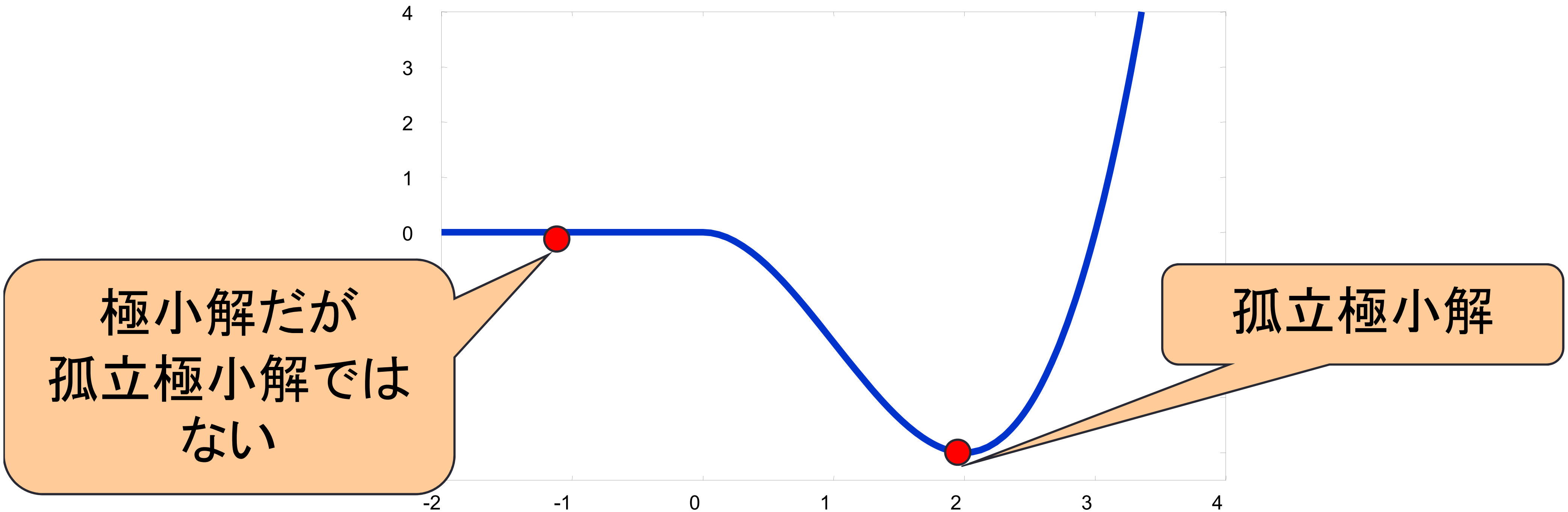
半正定値



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:Hf(x*) は正定値　⇒　(孤立)極小解

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

勾配を計算: $f'(x) = (x+2)x^2(x-2)(x-3)$

⇒ 停留点はx = -2, 0, 2, 3

2階微分を計算:

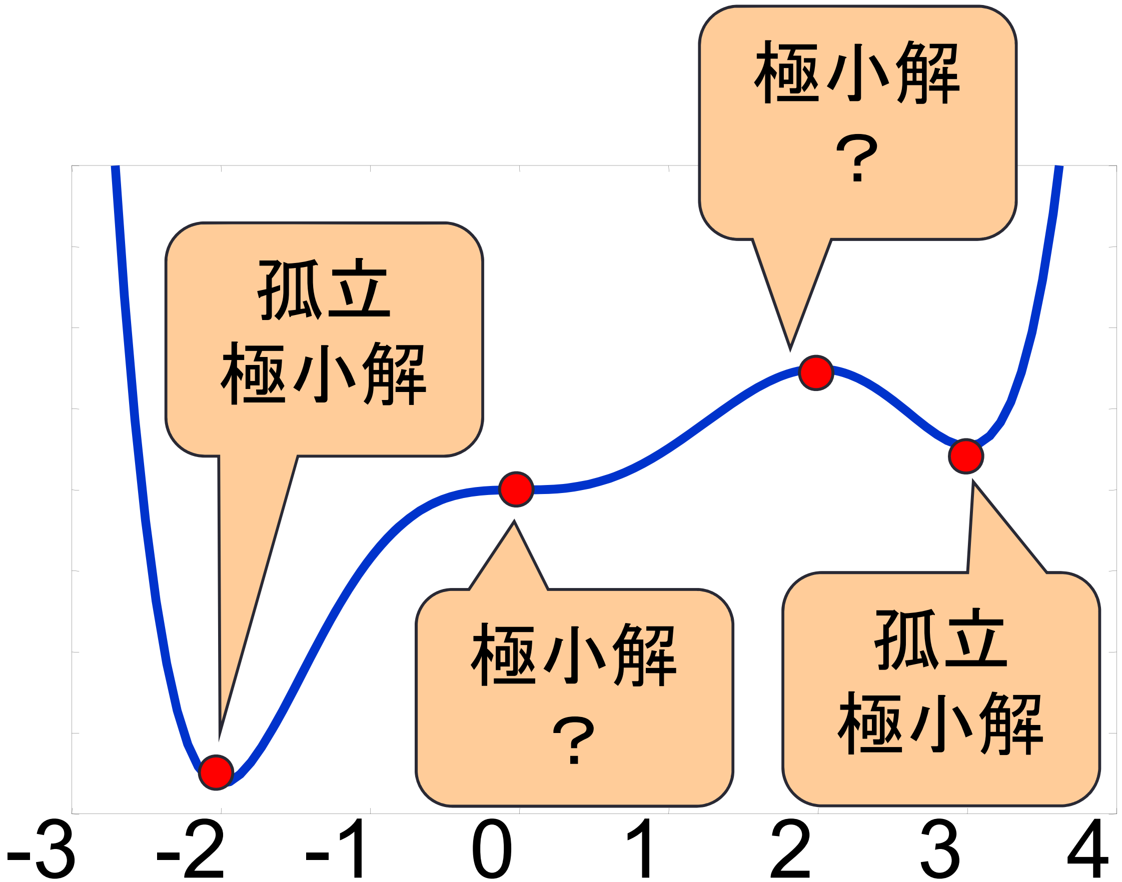
$$f''(x) = 5x^4 - 12x^3 - 12x^2 + 24x$$

⇒ 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)の例

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

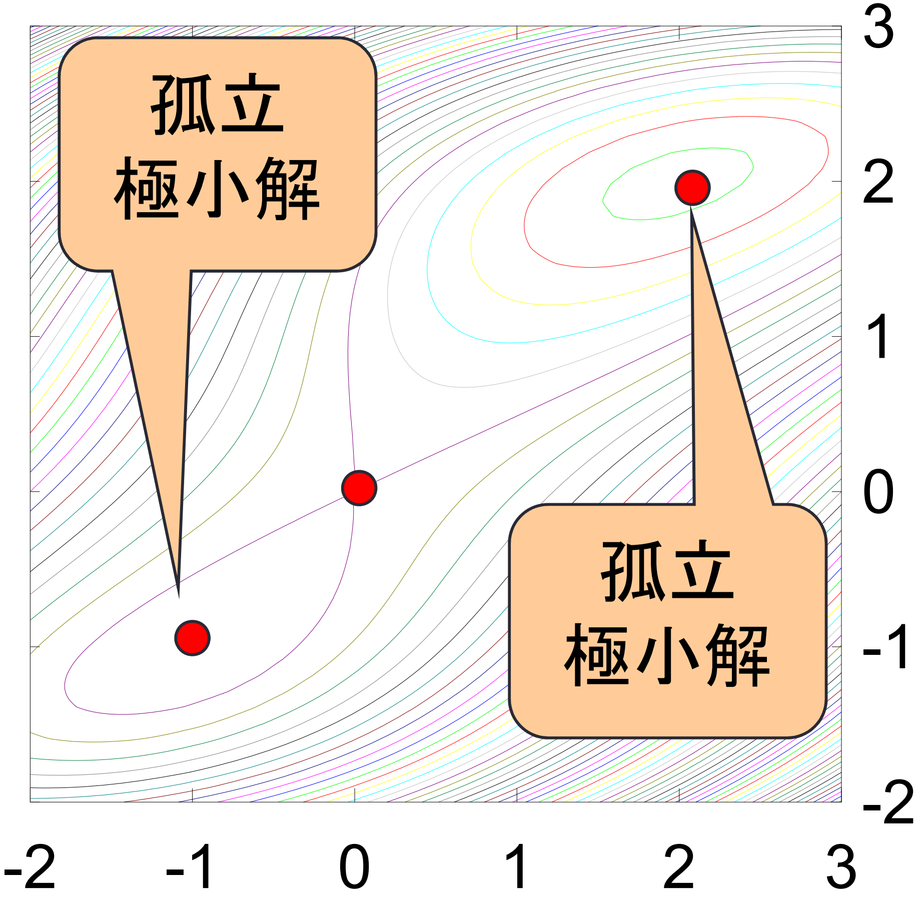
例2 $$f(\boldsymbol{x}) = x_1^2 - 2x_1x_2 + \frac{1}{4}x_2^4 - \frac{1}{3}x_2^3$$

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

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

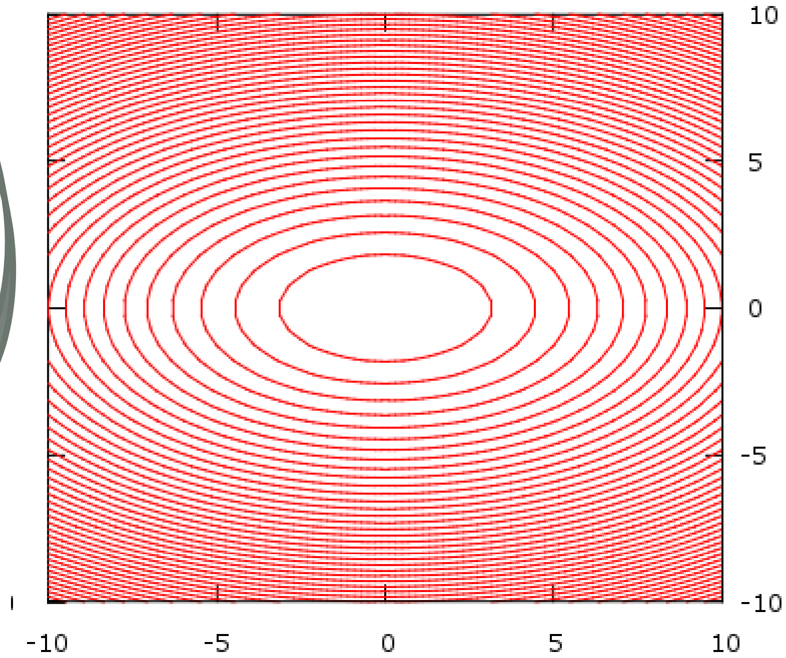
例3：$f(x_1, x_2) = x_1^2 + 3x_2^2$

- $\nabla f(x_1, x_2) = \begin{bmatrix} 2x_1 \\ 6x_2 \end{bmatrix}, \mathrm{H}f(x_1, x_2) = \begin{bmatrix} 2 & 0 \\ 0 & 6 \end{bmatrix}$
- $\nabla f(x_1, x_2)$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 : $f(x_1, x_2) = x_1^2 + x_2^4$

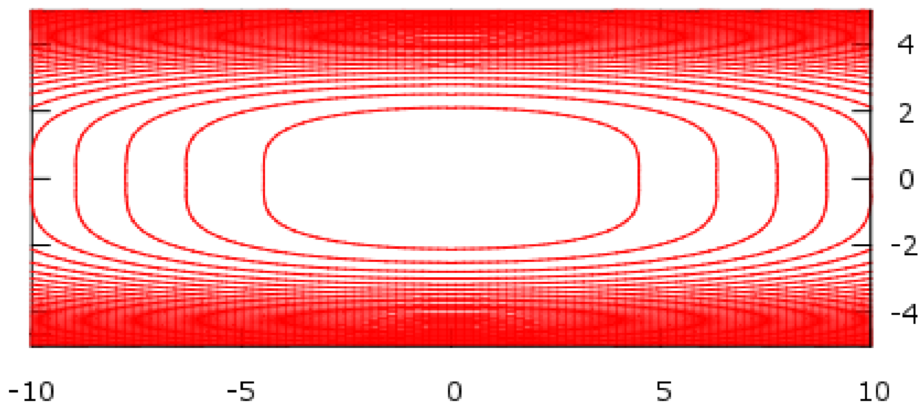
- $\nabla f(x_1, x_2) = \begin{bmatrix} 2x_1 \\ 4x_2^3 \end{bmatrix}, \nabla^2 f(x_1, x_2) = \begin{bmatrix} 2 & 0 \\ 0 & 12x_2^2 \end{bmatrix}$
- $\nabla f(x_1, x_2)$がゼロベクトルとなるのは (0,0) のみ（実は最適解）
- $\nabla^2 f(0,0) = \begin{bmatrix} 2 & 0 \\ 0 & 0 \end{bmatrix}$ は半正定値だが，正定値ではない

➔(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



2次の最適性条件(必要条件)証明

> 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

証明: A=H$f(x^*)$ とおく.
背理法: A は半正定値でないと仮定
➔ ある単位ベクトル y に対し, $y^TAy<0$
以下に示すように,
ある ε>0 に対して f(x*+ε'y)<f(x*) (0<∀ε'< ε) となり, 矛盾.
x = x* での2次のテイラー展開と∇$f(x^*)$ =0 を使うと,

$$f(x^*+\varepsilon y)=f(x^*)+\nabla f(x^*)^T(\varepsilon y)+\frac{1}{2}(\varepsilon y)^TA(\varepsilon y)+\psi(\varepsilon y)$$

$$=f(x^*)+\varepsilon^2\left(\frac{1}{2}y^TAy+\frac{\psi(\varepsilon y)}{\varepsilon^2}\right)$$

テイラー展開の性質より, ある ε>0 が存在して, 0<∀ε'< ε に対して

$$\frac{1}{2}y^TAy+\frac{\psi(\varepsilon' y)}{(\varepsilon')^2}<0$$ ∴ f(x*+ε'y)<f(x*)

2次の最適性条件(十分条件)証明

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

証明の概略:

x = x* での2次のテイラー展開 $\tilde{f}$ を考えると,

$\tilde{f}$ は凸関数, x* が最小解 ∴ x* は $\tilde{f}$ の極小解

x* のある近傍において, $\tilde{f}$ と f は十分に近い ∴ x* は f の極小解

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

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$f''(x) \geq 0$

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

2次関数$f(x)=\frac{1}{2}x^TVx+cx+c_0$の係数行列 V が

正定値行列のとき，最小解（最適解）は簡単に求められる！

- $\nabla f(x)=Vx+c$ ➔ 停留点は $x^*=-V^{-1}c$ のみ
- ヘッセ行列 $=V$, 正定値行列 ➔ 停留点は最小解

2次の十分条件より x* は最小解

※ 正定値行列は正則行列（逆行列をもつ）
半正定値行列は正則とは限らない

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：

V が正定値の2次関数に対して最適解は簡単に求められる！

ただし，一般の関数は2次とは限らない

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

$$\tilde{f}(x) = f(a) + \nabla f(a)^T (x - a) + \frac{1}{2}(x - a)^T \mathrm{H} f(a)(x - a)$$

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

⇒ $a - Hf(a)^{-1} \nabla f(a)$ 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 から $x - Hf(x)^{-1}\nabla f(x)$ への移動を繰り返す
（$-Hf(x)^{-1}\nabla f(x)$ を，xにおけるニュートン方向と呼ぶ）

入力 : 関数 f ，勾配ベクトル∇f，ヘッセ行列$\mathrm{H}f$
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

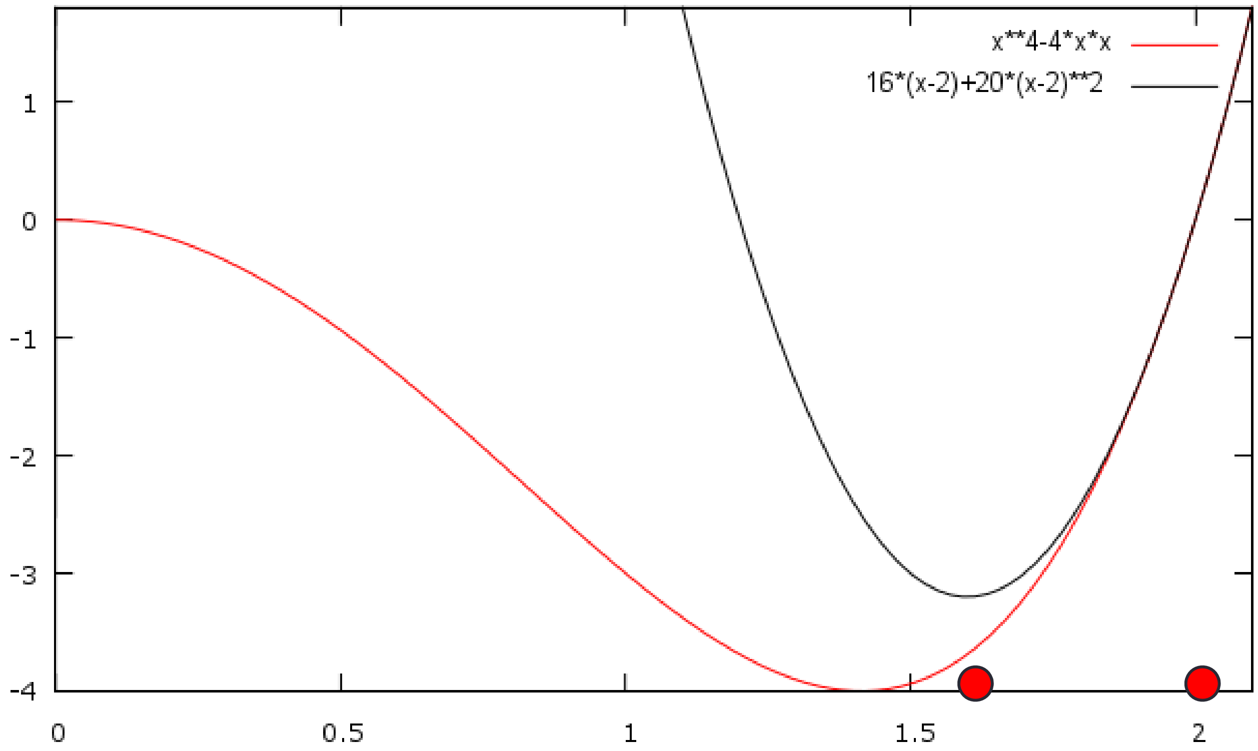
ステップ2 : ニュートン方向 $-Hf(x^k)^{-1}\nabla f(x^k)$ を計算

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

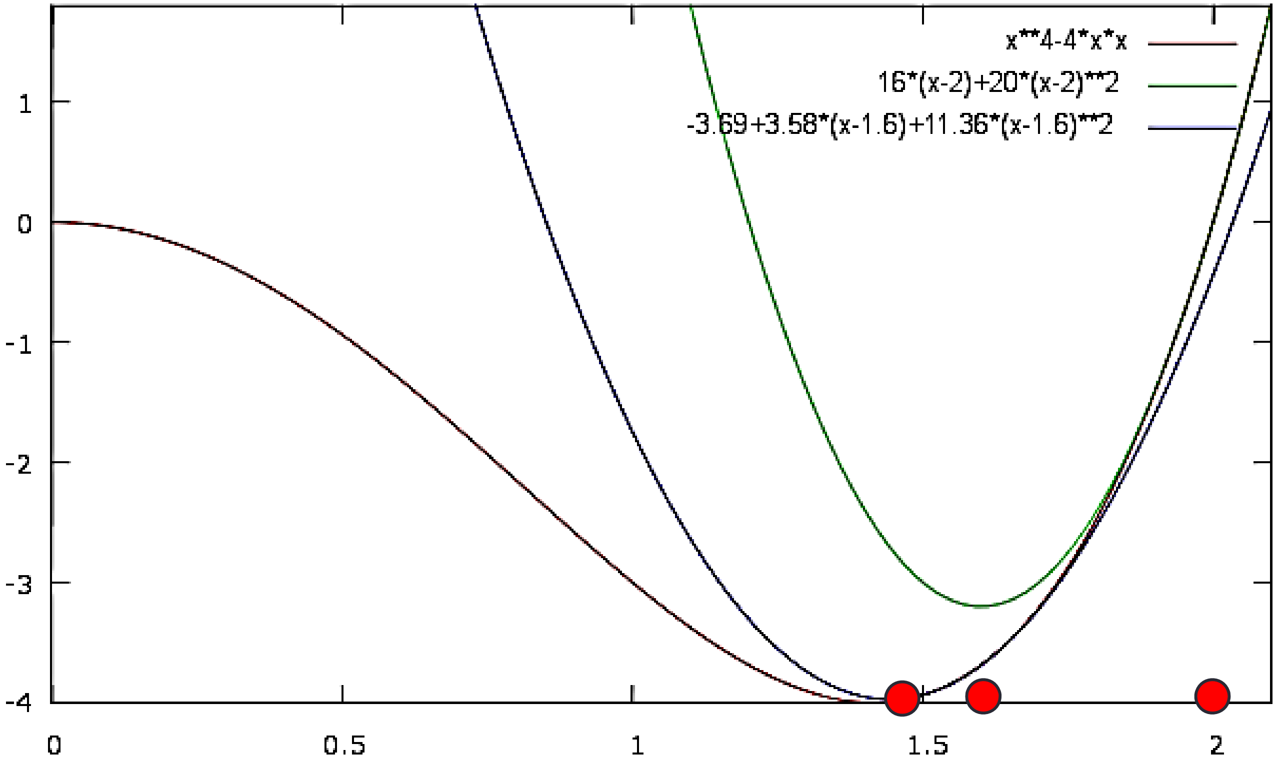
ニュートン法の実行例その1

- 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$x = 1.6 - 0.16 = 1.44$のとき
- $x^{(2)} := 1.44$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)

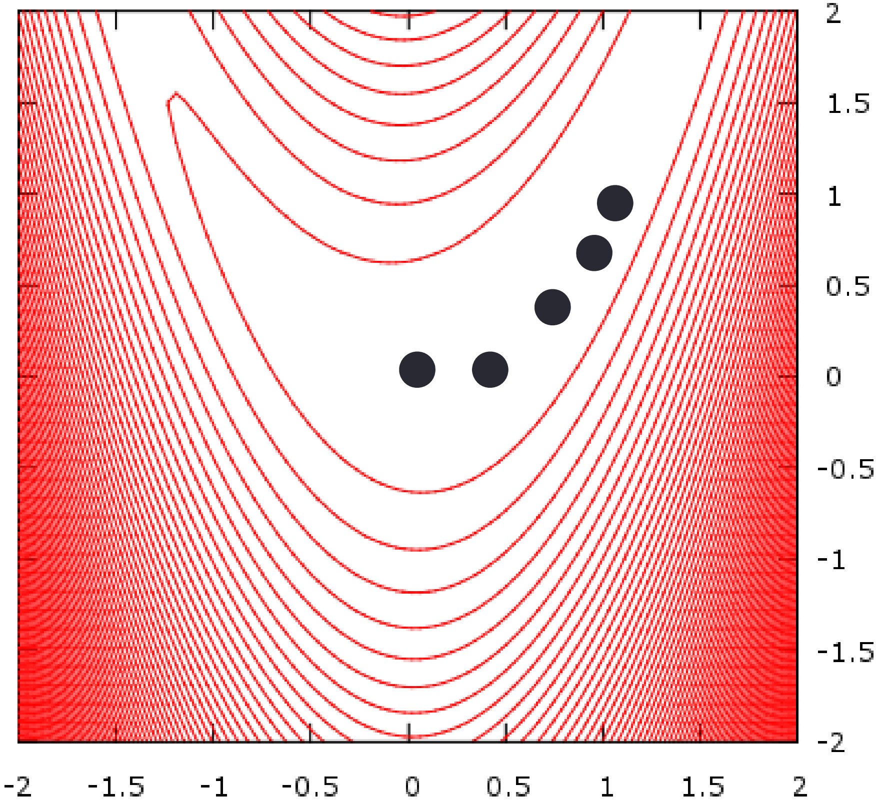


表 4.2　関数 (4.19) に対するニュートン法の計算結果

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

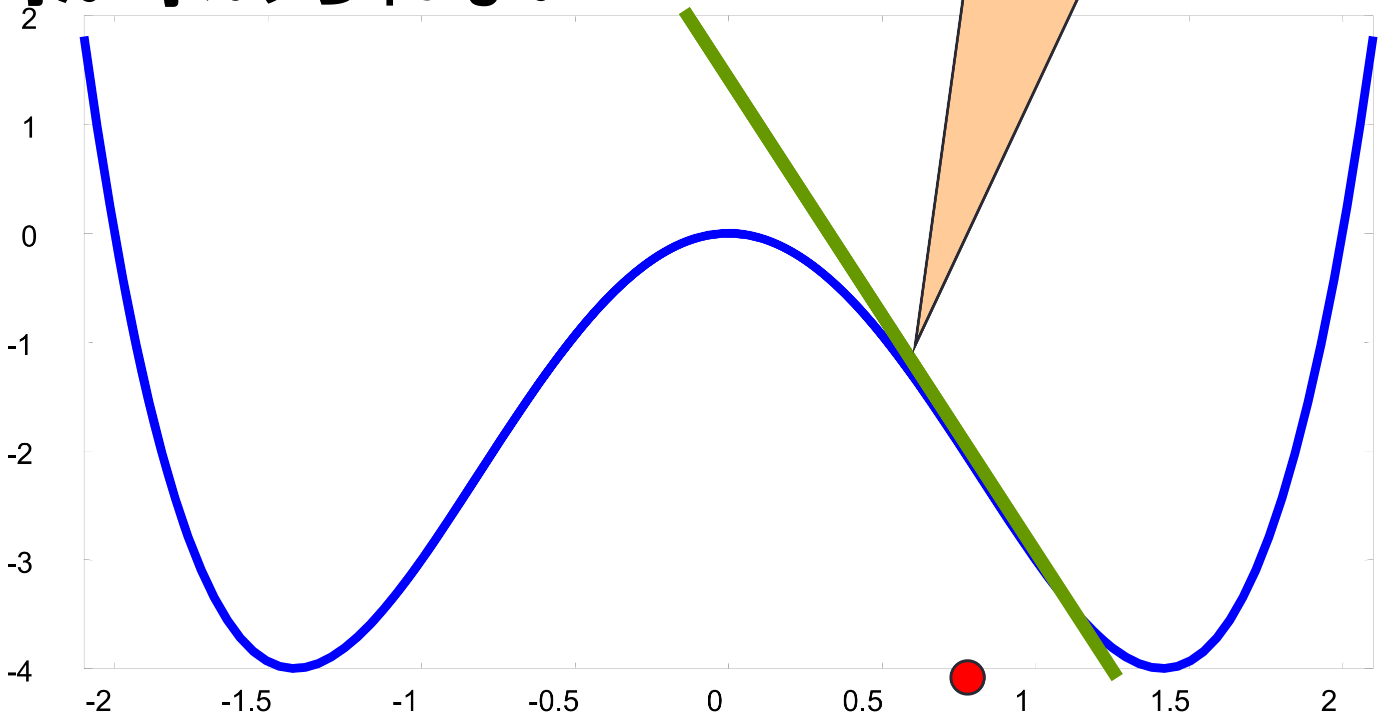
例1（続き）: 一変数関数 $f(x) = x^4 - 4x^2$

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

f を2次近似すると直線になる



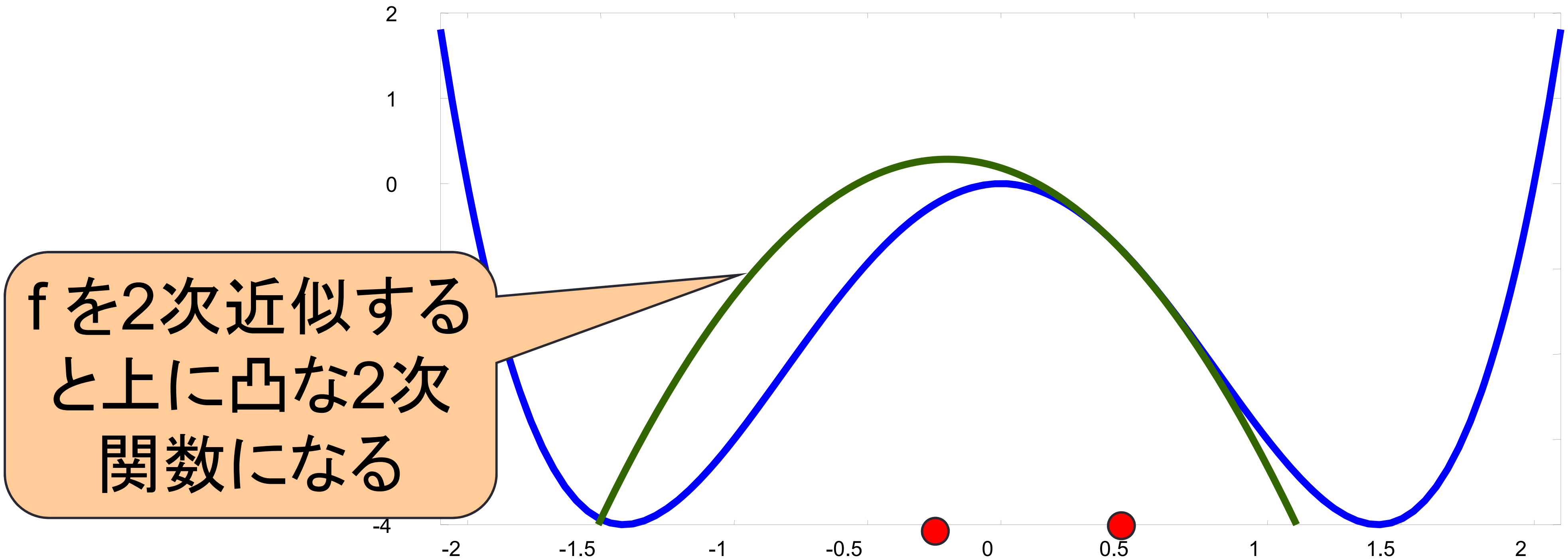
ニュートン法の問題点

- 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



レポート問題

問 1：次の3つの関数に対し，$(x_1, x_2) = (1,2)$における二次のテイラー近似を求めなさい（ log の底はeとする）

$$f_1(x_1, x_2) = x_1 \log x_2 - x_2 \log x_1 \qquad f_2(x_1, x_2) = x_1^2 + x_2^2 - 1$$

$$f_3(x, y) = x^2 - xy + 2y^2 - 2x + e^{x+y}$$

問 2 ：関数$f(x, y) = \frac{1}{4}x^4 - \frac{1}{2}x^2 + \frac{1}{3}y^3 + 2y^2$について考える

(a) 勾配ベクトルとヘッセ行列を計算せよ.

(b) すべての停留点（勾配ベクトルがゼロの点）を求めよ.

さらに，2次の最適性条件（十分条件）を用いて，極小解を求めよ.

問3：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 ⇔ $a_{11} \geqq 0$, $a_{22} \geqq 0$, $a_{11}a_{22} - a_{12}^2 \geqq 0$