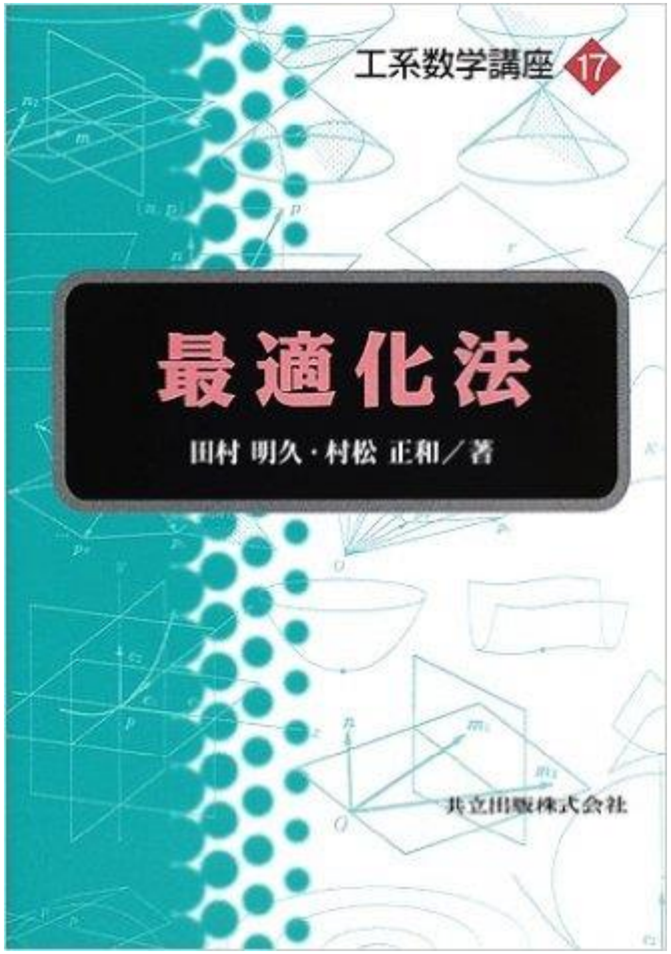
連続最適化(1): 制約なし最適化

佐藤一誠
sato@k.u-tokyo.ac.jp
http://www.ms.k.u-tokyo.ac.jp

参考書

- 田村，村松：最適化法，共立出版，2002



最適化問題と最適化アルゴリズム

- 最適化問題(optimization problem):
  - $\mathcal{X} \subset \mathbb{R}^d$ 上に定義される関数 $f : \mathcal{X} \to \mathbb{R}$ の最小値を求めよ

$$\min_{\boldsymbol{x} \in \mathcal{X}} f(\boldsymbol{x}) \qquad \boldsymbol{x} = (x^{(1)}, \ldots, x^{(d)})^\top$$

- 最適化アルゴリズム(optimization algorithm):
  - 収束する点列の生成アルゴリズム
  - 最適性条件: 最適化問題の最適解であるための必要条件

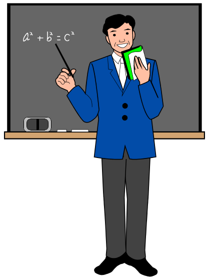
制約なし最適化問題

- 制約なし最適化問題(unconstrained optimization problem):
  - $\mathcal{X} \subset \mathbb{R}^d$ 上に定義される関数 $f : \mathcal{X} \to \mathbb{R}$ の最小値を求めよ

$$\min_{\boldsymbol{x} \in \mathcal{X}} f(\boldsymbol{x}) \qquad \boldsymbol{x} = (x^{(1)}, \ldots, x^{(d)})^\top$$

  - f を目的関数(objective function)とよぶ
  - 目的関数は微分可能(differentiable)と仮定する
  - 最適性条件: x^*が最適解のとき

$$\nabla f(\boldsymbol{x}^*) = \mathbf{0} \qquad [\nabla f(\boldsymbol{x})]_j = \frac{\partial f}{\partial x^{(j)}}$$



講義の流れ

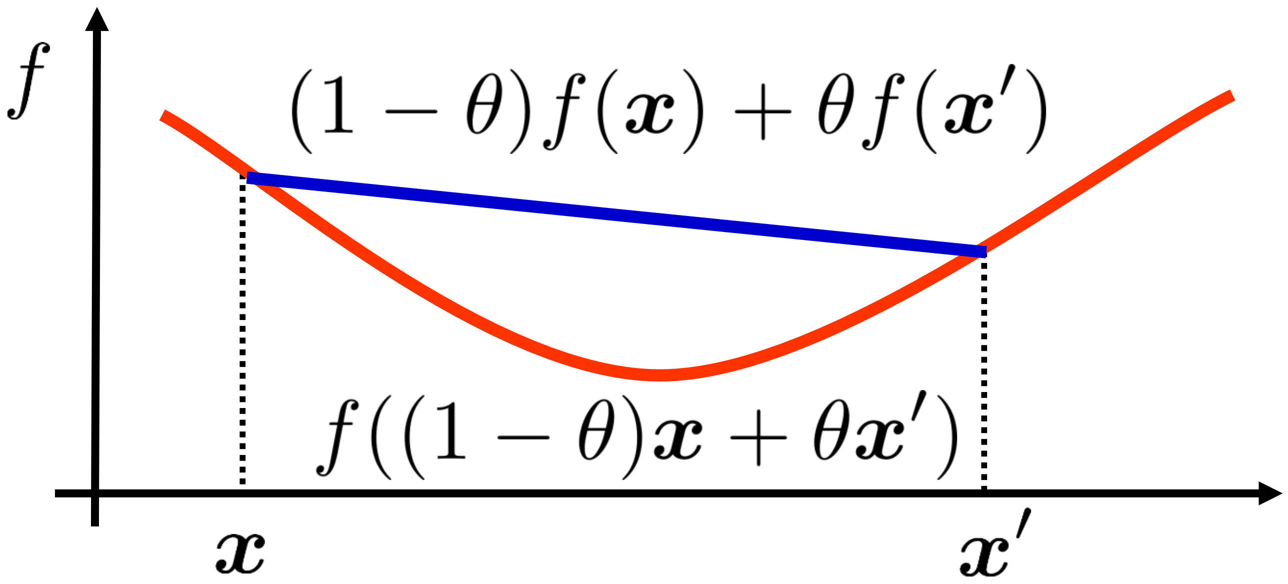
1. 凸最適化問題
2. 最急降下法
3. ニュートン法
4. 準ニュートン法

凸関数

- 任意の $\boldsymbol{x}, \boldsymbol{x}' \in \mathcal{X}$，任意の $\theta \in [0,1]$ に対して

$$f((1-\theta)\boldsymbol{x} + \theta\boldsymbol{x}') \leq (1-\theta)f(\boldsymbol{x}) + \theta f(\boldsymbol{x}')$$

ならば, f を凸関数(convex function)とよぶ



凸関数

$$f((1-\theta)\boldsymbol{x}+\theta\boldsymbol{x}') \leq (1-\theta)f(\boldsymbol{x})+\theta f(\boldsymbol{x}')$$

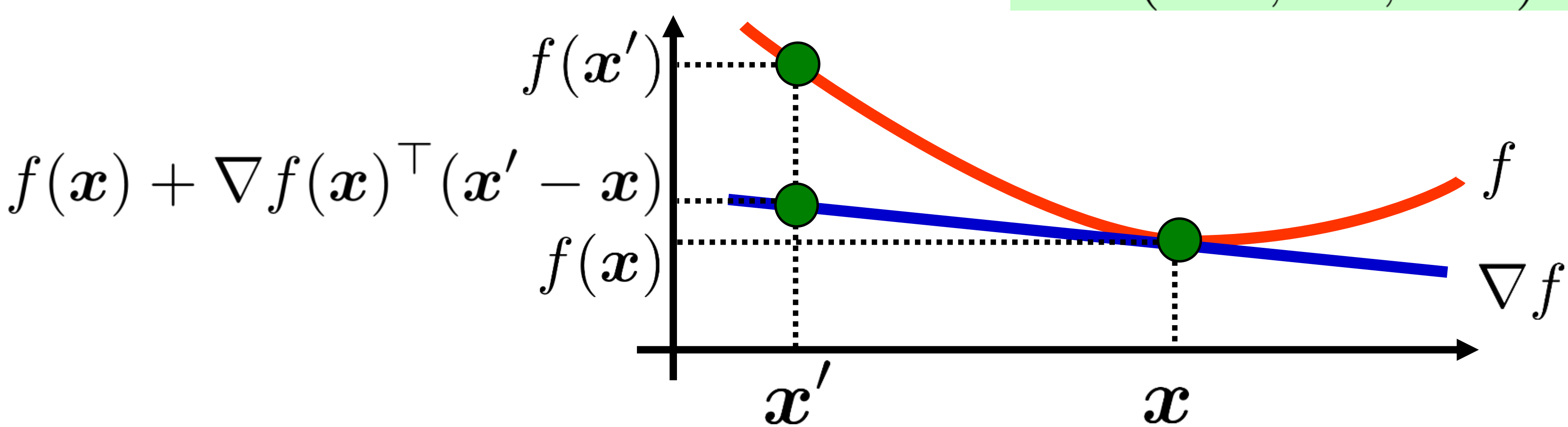
- f が一階微分可能のとき，f が凸関数となるための必要十分条件は

$$\forall \boldsymbol{x}, \boldsymbol{x}' \in \mathcal{X}, \quad f(\boldsymbol{x}') \geq f(\boldsymbol{x}) + \nabla f(\boldsymbol{x})^\top (\boldsymbol{x}' - \boldsymbol{x})$$

  - 接線が常に関数の下にくる

$$[\nabla f(\boldsymbol{x})]_j = \frac{\partial f(\boldsymbol{x})}{\partial x^{(j)}}$$

$$\boldsymbol{x} = (x^{(1)}, \ldots, x^{(d)})^\top$$



凸関数

$$f((1-\theta)\boldsymbol{x}+\theta\boldsymbol{x}') \leq (1-\theta)f(\boldsymbol{x})+\theta f(\boldsymbol{x}')$$

- 行列が半正定値(positive semi-definite)
  ⇔固有値(eigenvalue)が全て非負
- fが二階微分可能のとき，fが凸関数となるための必要十分条件はヘッセ行列(Hessian matrix) $\nabla^2 f(\boldsymbol{x})$ が半正定値:

$$\forall \boldsymbol{x} \in \mathcal{X}, \quad \nabla^2 f(\boldsymbol{x}) \geq \boldsymbol{O}$$

$$\boldsymbol{x} = (x^{(1)}, \ldots, x^{(d)})^\top$$

$$[\nabla^2 f(\boldsymbol{x})]_{j,j'} = \frac{\partial^2 f(\boldsymbol{x})}{\partial x^{(j)} \partial x^{(j')}}$$

演習

■ 以下の関数の凸性を示せ

1. $f(x) = x^2, \ \ x \in \mathbb{R}$ f'(x)=2x, f''(x)=2 > 0
2. $f(x) = e^x, \ \ x \in \mathbb{R}$ f''(x) = e^x > 0
3. $f(x) = -\log x, \ \ x > 0$ f''(x) = 1/x^2
4. $f(x) = x \log x, \ \ x > 0$ f''(x) = 1/x

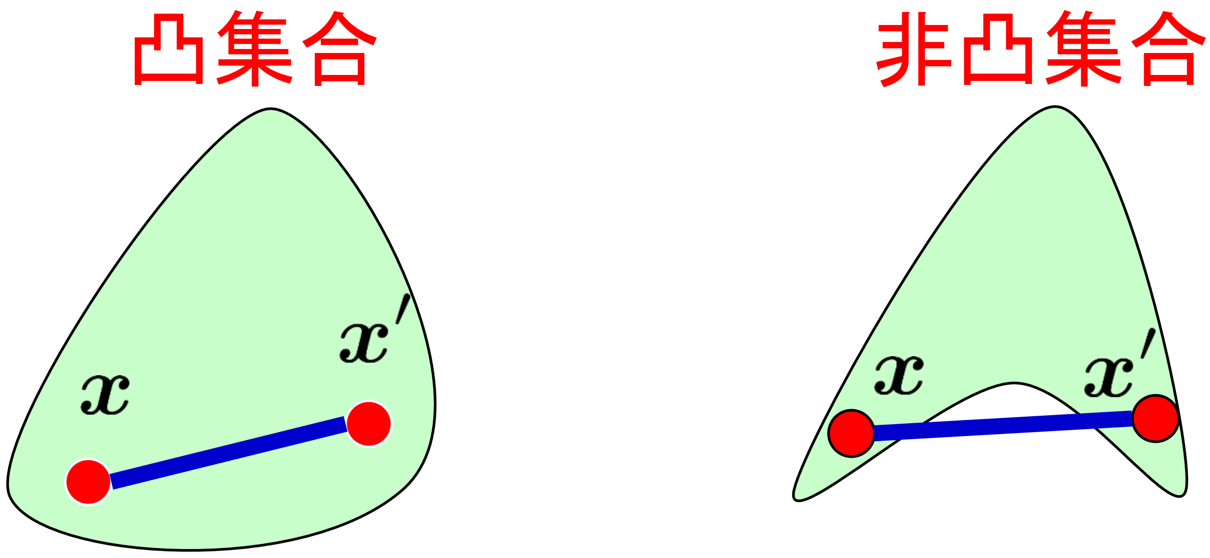
解答例

■二階微分を求める

1. $(x^2)'' = (2x)' = 2 > 0$
2. $(e^x)'' = (e^x)' = e^x > 0$
3. $(-\log x)'' = (-x^{-1}) = x^{-2} > 0$
4. $(x \log x)'' = (\log x + 1)' = x^{-1} > 0$

凸集合

- 任意の $\boldsymbol{x}, \boldsymbol{x}' \in \mathcal{X}$, 任意の $\theta \in [0, 1]$ に対して, $(1-\theta)\boldsymbol{x} + \theta\boldsymbol{x}'$ も $\mathcal{X}$ に属するとき, $\mathcal{X}$ を凸集合 (convex set)とよぶ
- $\boldsymbol{x}$ と $\boldsymbol{x}'$ を結ぶ線分上の全ての点が $\mathcal{X}$ に属する



凸最適化問題

$$\min_{\boldsymbol{x} \in \mathcal{X}} f(\boldsymbol{x})$$

- 凸最適化問題(convex optimization problem):
  関数 f が凸で集合 $\mathcal{X}$ も凸
  - 最適解が一意に定まる:

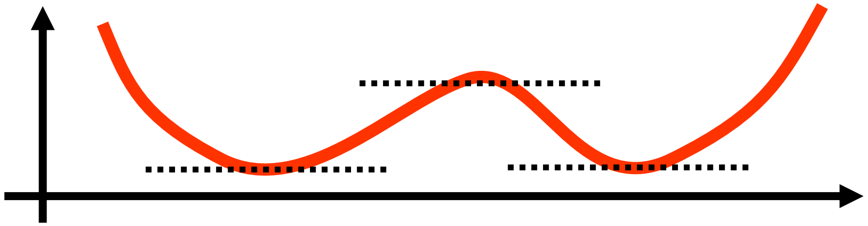
$$\boldsymbol{x}^* = \operatorname*{argmin}_{\boldsymbol{x} \in \mathcal{X}} f(\boldsymbol{x})$$

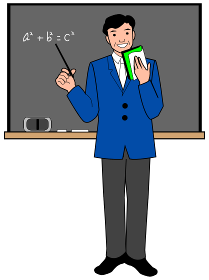
  - 最適解の必要十分条件:
    - 勾配(gradient)がゼロ

$$[\nabla f(\boldsymbol{x})]_j = \frac{\partial f}{\partial x^{(j)}}$$

$$\nabla f(\boldsymbol{x}^*) = \mathbf{0} \qquad \boldsymbol{x} = (x^{(1)}, \ldots, x^{(d)})^\top$$

    - 凸でない場合は必要条件





講義の流れ

1. 凸最適化問題
2. 最急降下法
   A) ステップ幅の選択
   B) 収束性
   C) 微分不可能な場合
3. ニュートン法
4. 準ニュートン法

最急降下法

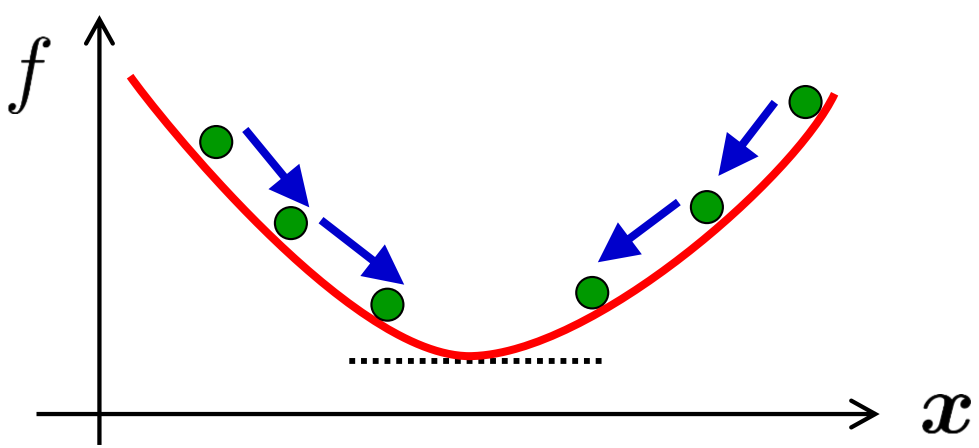
- 最急降下法(steepest descent method):
  1. 適当に初期値 $\boldsymbol{x}_0$ を定める.
  2. $k = 0, 1, 2, \ldots$ に対して,
     収束するまで以下を繰り返す

$$\boldsymbol{x}_{k+1} = \boldsymbol{x}_k - \varepsilon_k \nabla f(\boldsymbol{x}_k)$$

$\varepsilon_k > 0$: ステップ幅(step size)

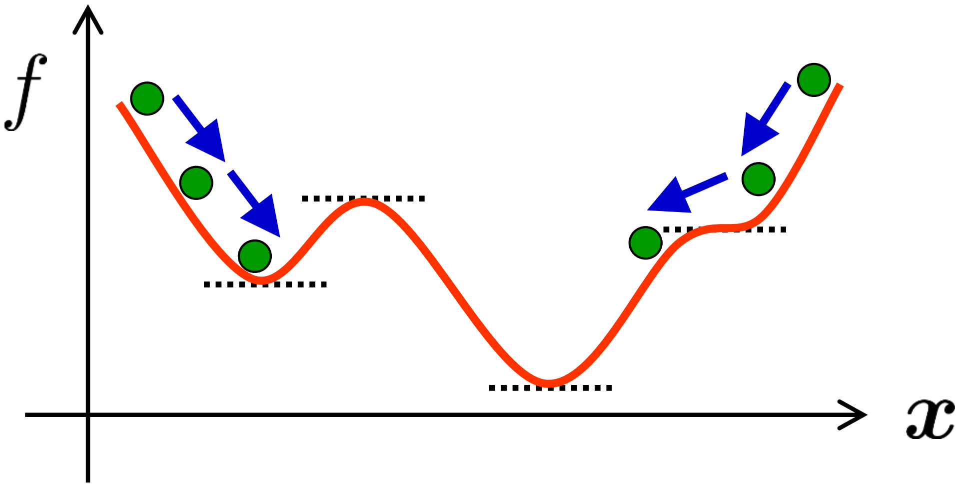
勾配を下るように
パラメータを更新

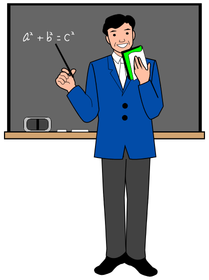
- 勾配法(gradient method)ともよぶ



最急降下法(続き)

- 凸最適化問題の場合:
  - 大域的最適解(global optimal solution)が求まる
- 非凸最適化問題の場合:
  - 一般に局所最適解(local optimal solution)しか求められない
  - 様々な初期値から何度か学習し,
一番小さい値を採用する





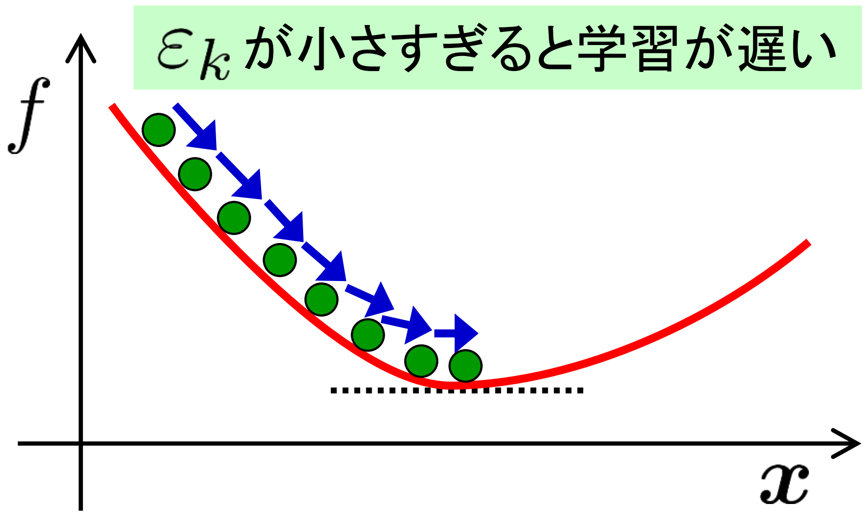
講義の流れ

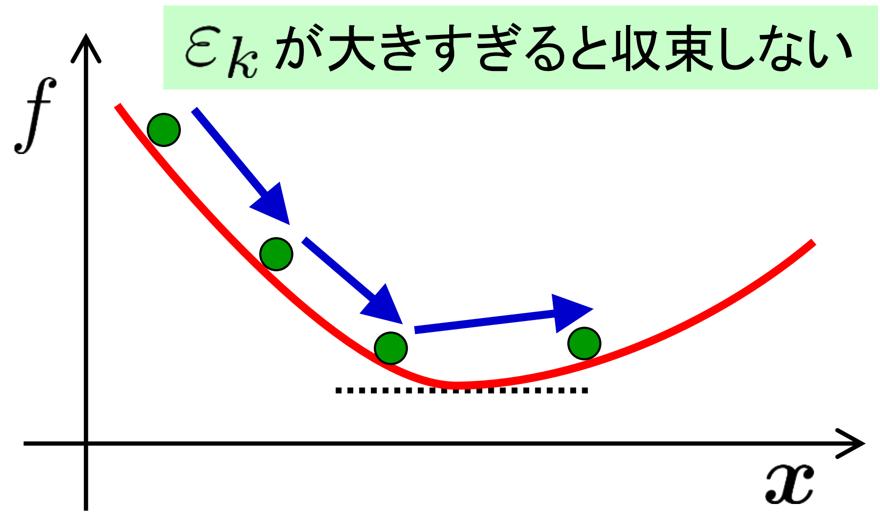
1. 凸最適化問題
2. 最急降下法
   - A) ステップ幅の選択
     - i. 厳密直線探索
     - ii. バックトラック直線探索
   - B) 収束性
   - C) 微分不可能な場合
3. ニュートン法
4. 準ニュートン法

ステップ幅の選択

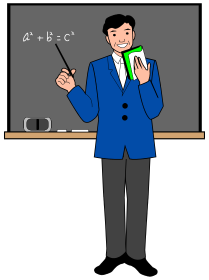
$$\boldsymbol{x}_{k+1} = \boldsymbol{x}_k - \varepsilon_k \nabla f(\boldsymbol{x}_k)$$

■ ステップ幅 ε_k の選び方が難しい：





- 焼きなまし(annealing)：「最初は大きく，徐々に小さく」．しかし，適切に実装することは容易でない
- 正規化(normalization) $\varepsilon_k = \varepsilon'_k / \|\nabla f(\boldsymbol{x}_k)\|$ ：勾配が大きいとき安定するが，勾配が小さいときは不安定



講義の流れ

1. 凸最適化問題
2. 最急降下法
   - A) ステップ幅の選択
     - i. 厳密直線探索
     - ii. バックトラック直線探索
   - B) 収束性
   - C) 微分不可能な場合
3. ニュートン法
4. 準ニュートン法

厳密直線探索

$$\boldsymbol{x}_{k+1} = \boldsymbol{x}_k - \varepsilon_k \nabla f(\boldsymbol{x}_k)$$

■ 厳密直線探索(exact line search):

- 目的関数の値を最小にするステップ幅を求める

$$\min_{\varepsilon_k > 0} f(\boldsymbol{x}_k - \varepsilon_k \nabla f(\boldsymbol{x}_k))$$

- 一般に，一次元の非線形最小化問題を解く必要があるため，最適なステップ幅の探索に時間がかかる

演習

$$\min_{\varepsilon_k > 0} f(\boldsymbol{x}_k - \varepsilon_k \nabla f(\boldsymbol{x}_k))$$

- $f(\boldsymbol{x}) = \frac{1}{2}\boldsymbol{x}^\top \boldsymbol{A}\boldsymbol{x}$ に対する厳密直線探索の解が

$\boldsymbol{A}$: 正定値対称行列

$$\varepsilon_k = \frac{\|\nabla f(\boldsymbol{x}_k)\|^2}{\nabla f(\boldsymbol{x}_k)^\top \boldsymbol{A} \nabla f(\boldsymbol{x}_k)}$$

$$\nabla f(\boldsymbol{x}) = \boldsymbol{A}\boldsymbol{x}$$

で与えられることを示せ

テキスト

解答例

$$\min_{\varepsilon_k>0} f(\boldsymbol{x}_k - \varepsilon_k \nabla f(\boldsymbol{x}_k)) \quad f(\boldsymbol{x}) = \frac{1}{2}\boldsymbol{x}^\top \boldsymbol{A}\boldsymbol{x}$$

- $f(\boldsymbol{x}_k - \varepsilon_k \nabla f(\boldsymbol{x}_k))$

$$= \frac{1}{2}(\boldsymbol{x}_k - \varepsilon_k \nabla f(\boldsymbol{x}_k))^\top \boldsymbol{A}(\boldsymbol{x}_k - \varepsilon_k \nabla f(\boldsymbol{x}_k))$$

- これをε_kに関して微分しゼロとおけば,

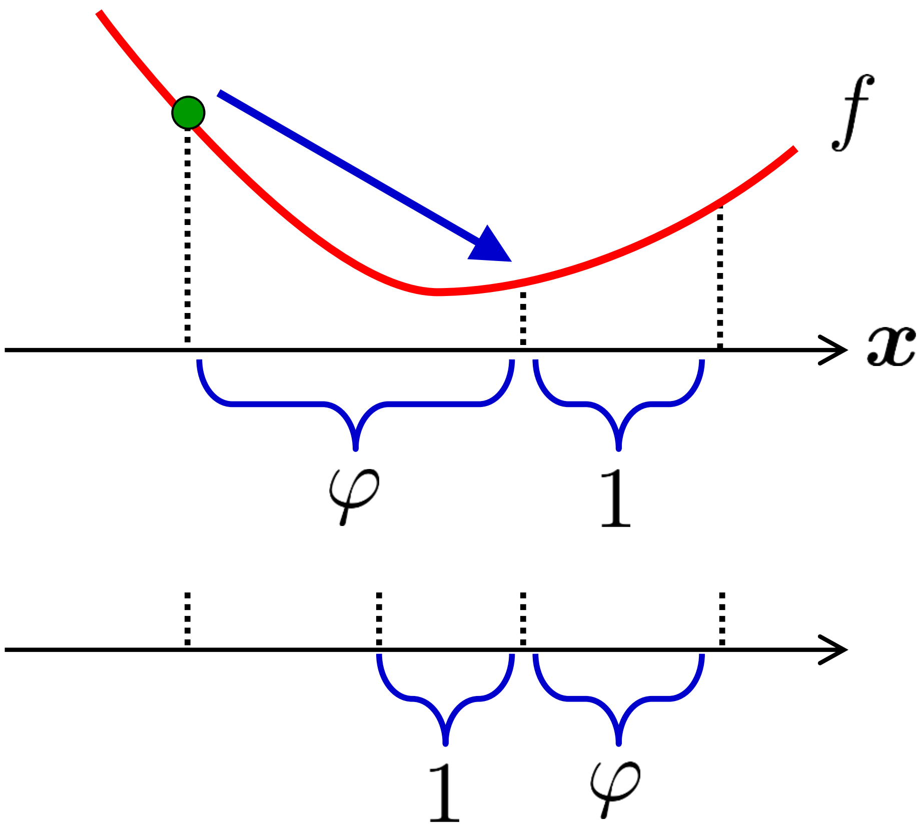
$$\varepsilon_k \nabla f(\boldsymbol{x}_k)^\top \boldsymbol{A} \nabla f(\boldsymbol{x}_k) - \nabla f(\boldsymbol{x}_k)^\top \boldsymbol{A}\boldsymbol{x}_k = 0$$

- これをε_kについて解けば

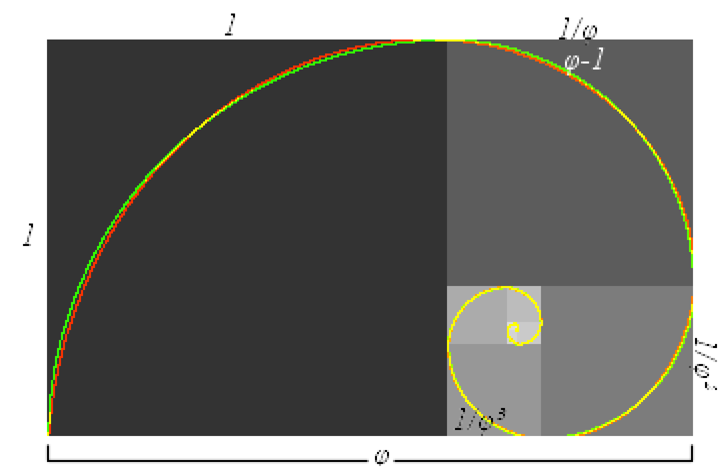
$$\varepsilon_k = \frac{\|\nabla f(\boldsymbol{x}_k)\|^2}{\nabla f(\boldsymbol{x}_k)^\top \boldsymbol{A} \nabla f(\boldsymbol{x}_k)} \quad \nabla f(\boldsymbol{x}) = \boldsymbol{A}\boldsymbol{x}$$

厳密直線探索

- f が単峰関数(unimodal function)の場合は，探索によって最小値を探す
  - 二分探索法(binary search)：1対1で分割
  - 黄金分割探索(golden section search)：1対φで分割



$$\varphi = \frac{1+\sqrt{5}}{2} \approx 1.62$$

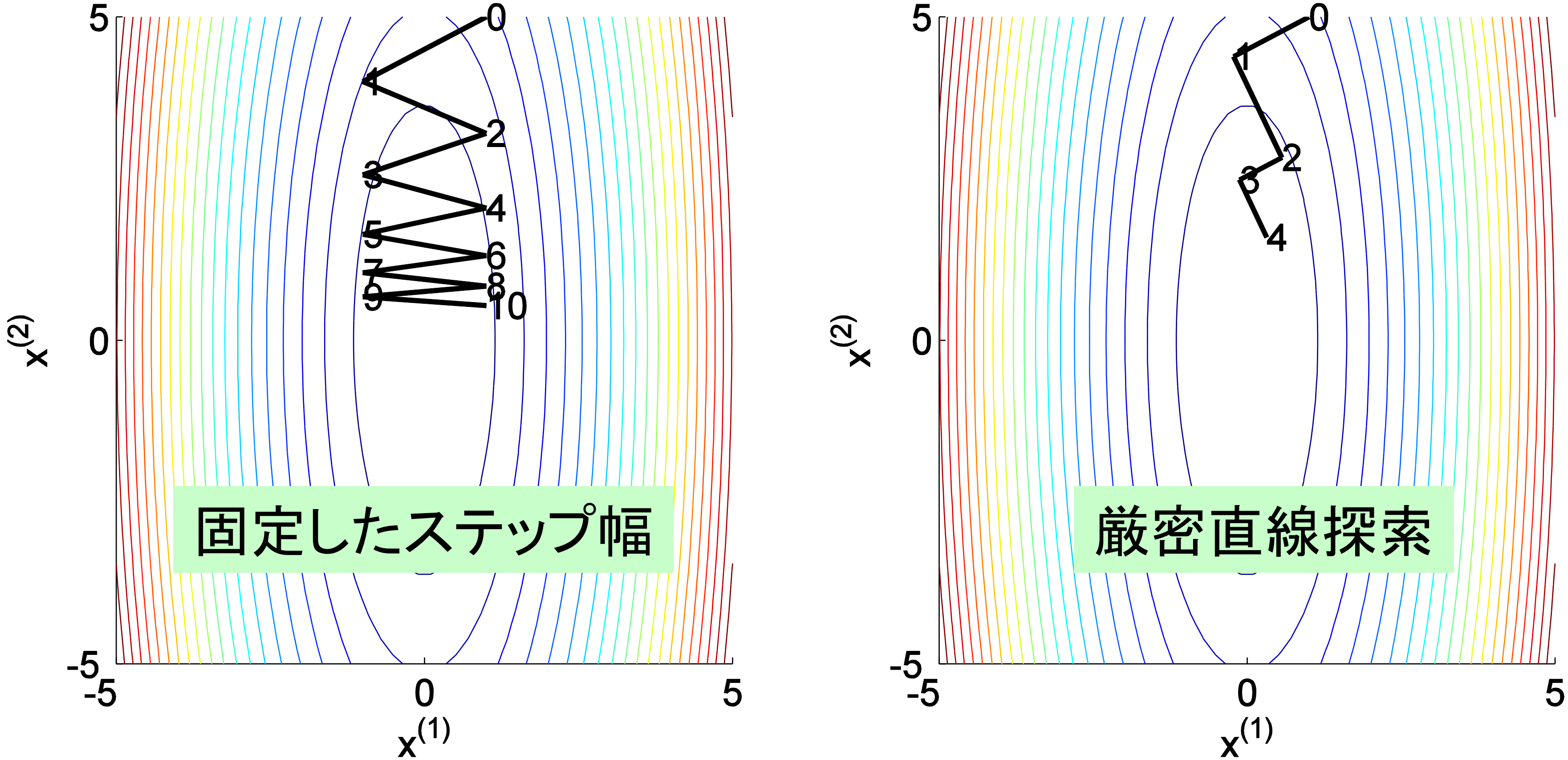


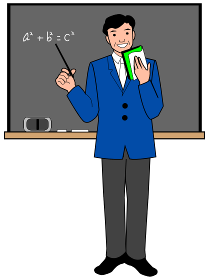
Wikipediaより

実行例

$$f(x^{(1)}, x^{(2)}) = 10(x^{(1)})^2 + (x^{(2)})^2$$

- 厳密直線探索により反復回数が軽減
- ただし，一般に各反復の計算に時間がかかる





講義の流れ

1. 凸最適化問題
2. 最急降下法
   - A) ステップ幅の選択
     - i. 厳密直線探索
     - ii. バックトラック直線探索
   - B) 収束性
   - C) 微分不可能な場合
3. ニュートン法
4. 準ニュートン法

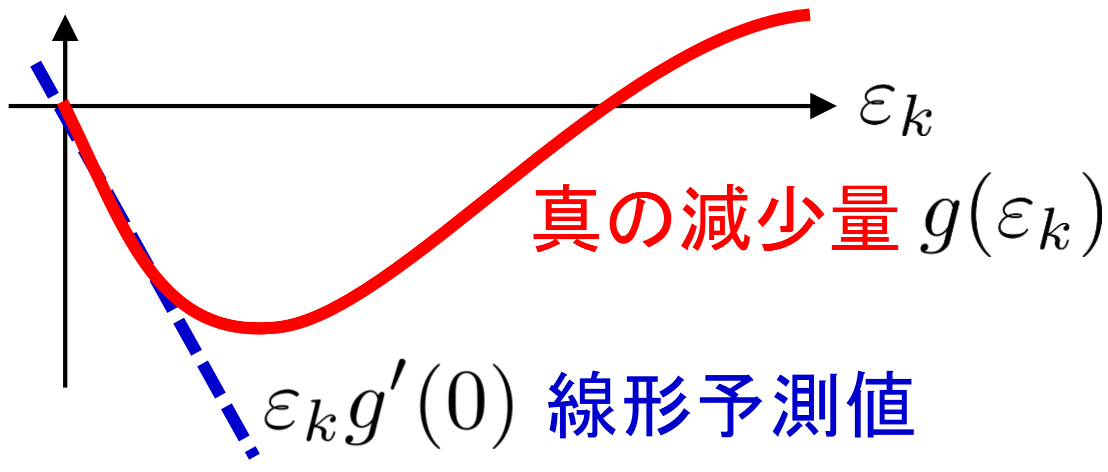
最小値の近似探索

$$\boldsymbol{x}_{k+1} = \boldsymbol{x}_k - \varepsilon_k \nabla f(\boldsymbol{x}_k)$$

- 最小値の探索を近似的に高速に実行
- f の真の減少量:

$$g(\varepsilon_k) = f(\boldsymbol{x}_k - \varepsilon_k \nabla f(\boldsymbol{x}_k)) - f(\boldsymbol{x}_k)$$

- 数学演習: f の減少量の線形予測値 $\varepsilon_k g'(0)$ を求めよ



解答例

$$g(\varepsilon_k) = f(\boldsymbol{x}_k - \varepsilon_k \nabla f(\boldsymbol{x}_k)) - f(\boldsymbol{x}_k)$$

- $g'(\varepsilon_k) = -\nabla f(\boldsymbol{x}_k)^\top \nabla f(\boldsymbol{x}_k - \varepsilon_k \nabla f(\boldsymbol{x}_k))$

より,

$$g'(0) = -\nabla f(\boldsymbol{x}_k)^\top \nabla f(\boldsymbol{x}_k) = -\|\nabla f(\boldsymbol{x}_k)\|^2$$

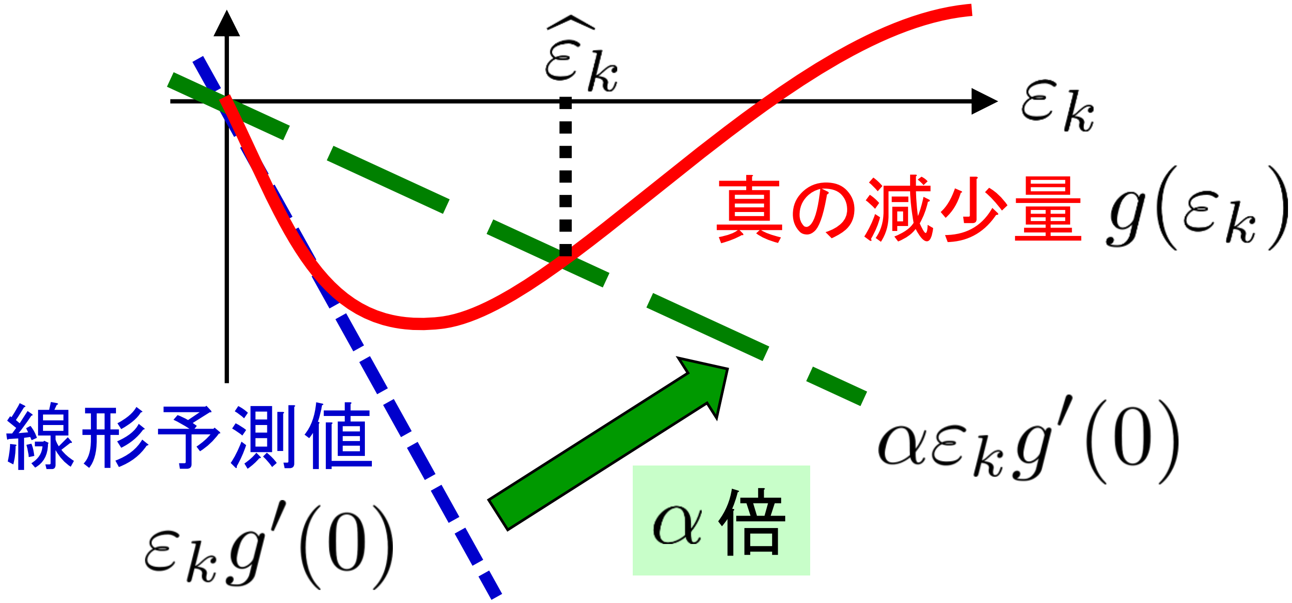
従って,

$$\varepsilon_k g'(0) = -\varepsilon_k \|\nabla f(\boldsymbol{x}_k)\|^2$$

アルミホ規準

■ アルミホ規準(Armijo rule): 目的関数の減少量の線形予測値の α 倍の減少を保証　$0 < \alpha < 1$

$$\underbrace{f(\boldsymbol{x}_k - \varepsilon_k \nabla f(\boldsymbol{x}_k)) - f(\boldsymbol{x}_k)}_{f\text{ の真の減少量}} \leq \underbrace{-\alpha \varepsilon_k \|\nabla f(\boldsymbol{x}_k)\|^2}_{\substack{f\text{ の減少量の線形予測値} \\ -\varepsilon_k \|\nabla f(\boldsymbol{x}_k)\|^2 \text{ の}\alpha\text{倍}}}$$



アルミホ規準

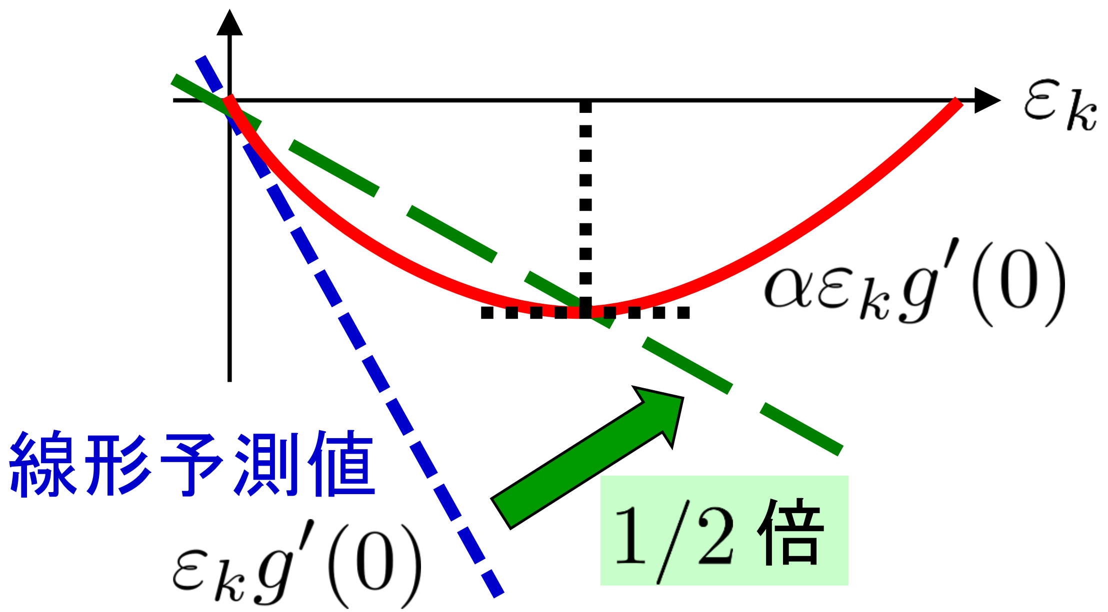
$$g(\varepsilon_k) \leq \alpha \varepsilon_k g'(0)$$

- $f(\boldsymbol{x}) = \frac{1}{2}\boldsymbol{x}^\top \boldsymbol{A}\boldsymbol{x}$ のとき，$\alpha = 1/2$ のアルミホ規準

を満たす最大の ε_k は，最適なステップ幅

$$\frac{\boldsymbol{x}_k^\top \boldsymbol{A}^2 \boldsymbol{x}_k}{\boldsymbol{x}_k^\top \boldsymbol{A}^3 \boldsymbol{x}_k}$$

と一致する



証明は宿題

バックトラック直線探索

$$\boldsymbol{x}_{k+1} = \boldsymbol{x}_k - \varepsilon_k \nabla f(\boldsymbol{x}_k)$$

- バックトラック直線探索(backtracking line search)
  - $\varepsilon_k = 1$ に初期化
  - アルミホ規準

$$f(\boldsymbol{x}_k - \varepsilon_k \nabla f(\boldsymbol{x}_k)) - f(\boldsymbol{x}_k) \le -\alpha \varepsilon_k \|\nabla f(\boldsymbol{x}_k)\|^2$$

が成り立つまで ε_k を

$$\varepsilon_k \leftarrow \beta \varepsilon_k$$

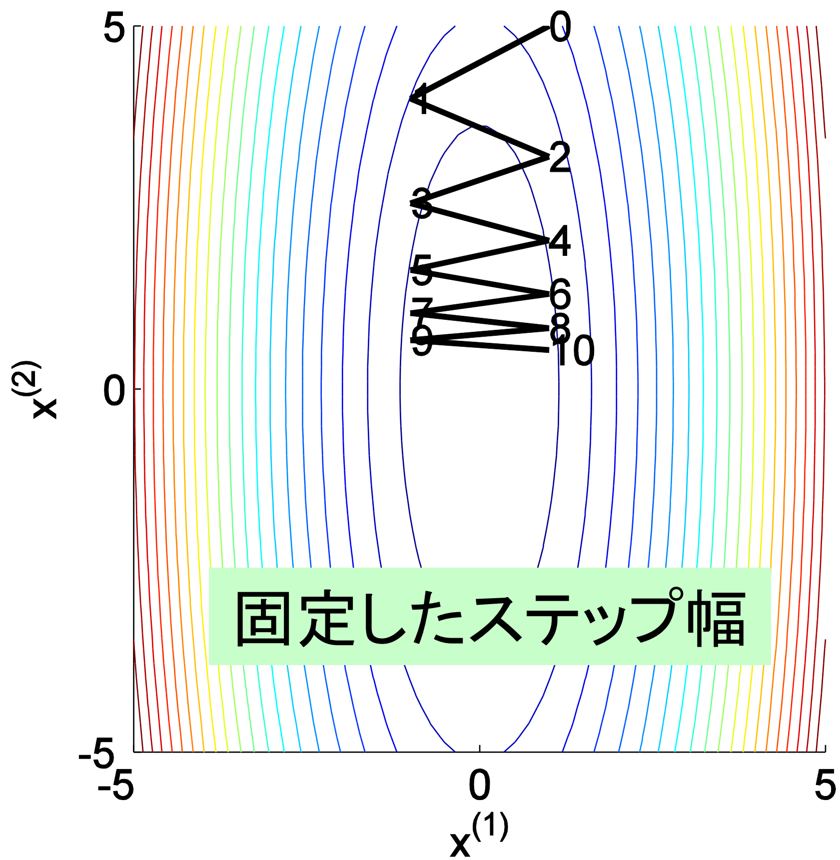
と減衰させる

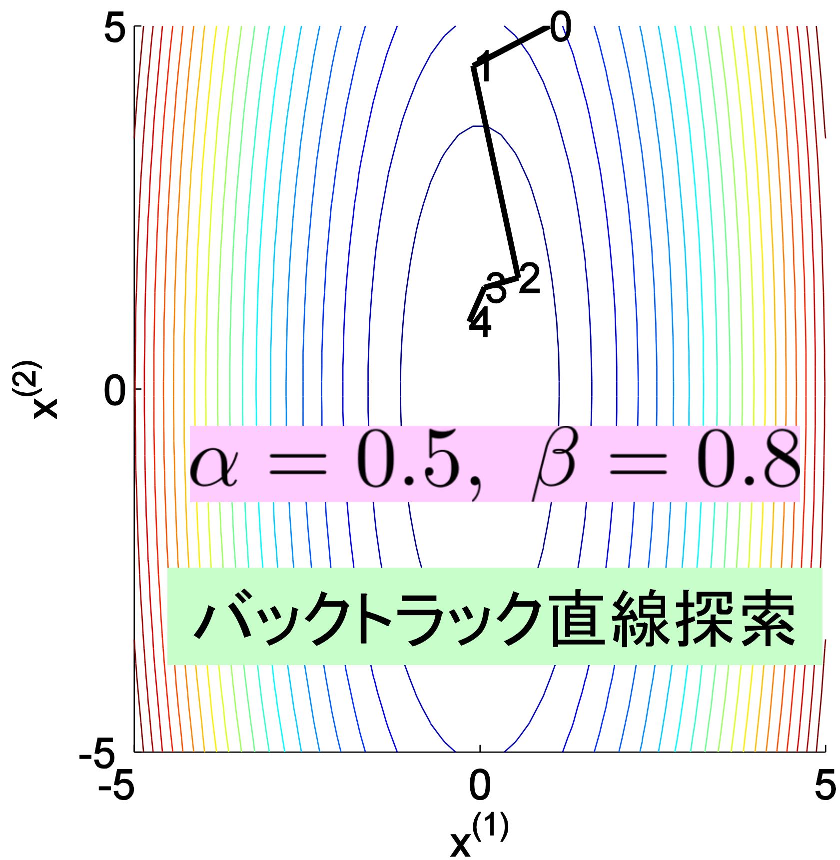
$$0 < \alpha, \beta < 1$$

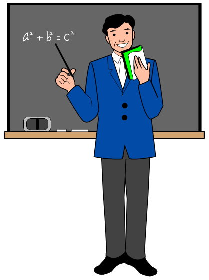
実行例

$$f(x^{(1)}, x^{(2)}) = 10(x^{(1)})^2 + (x^{(2)})^2$$

- バックトラック直線探索により，反復回数が軽減







講義の流れ

1. 凸最適化問題
2. 最急降下法
   - A) ステップ幅の選択
   - B) 収束性
   - C) 微分不可能な場合
3. ニュートン法
4. 準ニュートン法

最急降下法の収束性

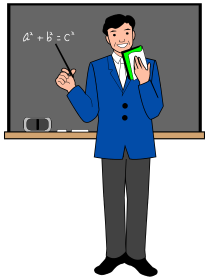
- k が十分大きいとき，

$$\|\boldsymbol{x}_{k+1} - \boldsymbol{x}^*\| \leq c\|\boldsymbol{x}_k - \boldsymbol{x}^*\|$$

$$\boldsymbol{x}^* = \operatorname*{argmin}_{\boldsymbol{x}\in\mathcal{X}} f(\boldsymbol{x}) \qquad 0 < c < 1$$

- 厳密直線探索を用いた最急降下法は，
一次収束(linear convergence)する

$$n\text{次収束}:\ \|\boldsymbol{x}_{k+1} - \boldsymbol{x}^*\| \leq c\|\boldsymbol{x}_k - \boldsymbol{x}^*\|^n$$



講義の流れ

1. 凸最適化問題
2. 最急降下法
   A) ステップ幅の選択
   B) 収束性
   C) 微分不可能な場合
3. ニュートン法
4. 準ニュートン法

微分不可能な制約なし最適化問題

$$\min_{\boldsymbol{x} \in \mathcal{X}} f(\boldsymbol{x})$$

- 目的関数 f が微分不可能(non-differentiable)の場合はどうするか？
- 微分の概念を一般化した劣微分を用いる

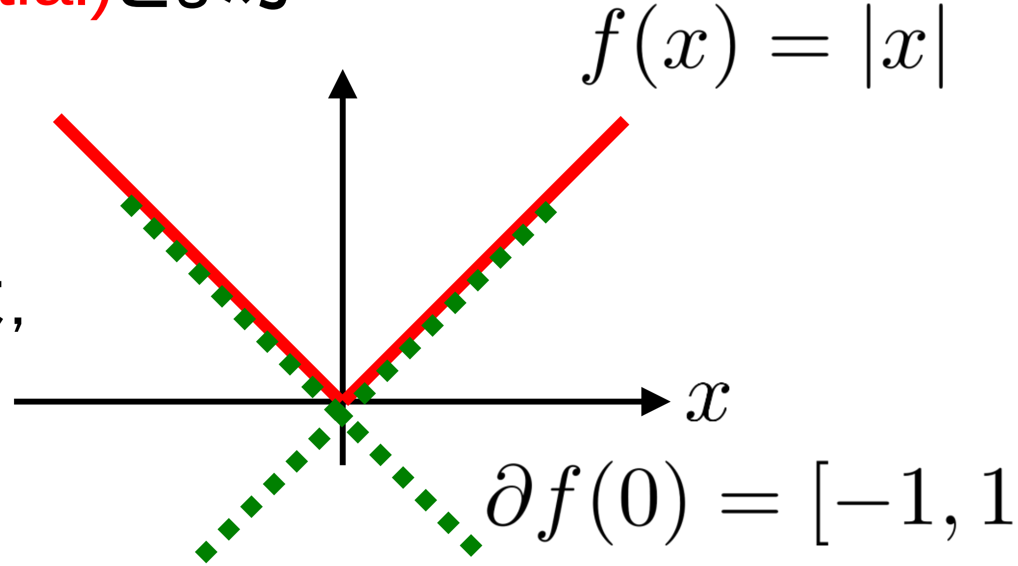
劣勾配法

- 凸関数 f の $\boldsymbol{x}'$ での劣勾配(sub-gradient)とは，全ての $\boldsymbol{x} \in \mathcal{X}$ に対して次式を満たす$\boldsymbol{\xi}$:

$$f(\boldsymbol{x}) \geq f(\boldsymbol{x}') + \boldsymbol{\xi}^\top(\boldsymbol{x} - \boldsymbol{x}')$$

  - f が微分可能なとき，$\boldsymbol{\xi} = \nabla f(\boldsymbol{x}')$
  - 上式を満たす$\boldsymbol{\xi}$全体を$\partial f(\boldsymbol{x}')$で表し劣微分(sub-differential)とよぶ

- 劣勾配法:
  - 勾配法において，微分不可能な点では，劣微分のどれかの値を用いる



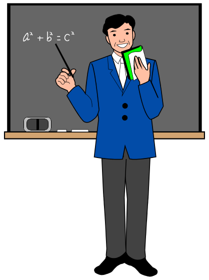
近接勾配法

$$\min_{\boldsymbol{x}\in\mathcal{X}} f(\boldsymbol{x}) + g(\boldsymbol{x})$$

- f : 凸で微分可能, g : 凸で微分不可能
- 適当な初期値から x を次式で更新:

$$\boldsymbol{x}_{k+1} = \operatorname*{argmin}_{\boldsymbol{x}} \underbrace{\nabla f(\boldsymbol{x}_k)^\top (\boldsymbol{x}-\boldsymbol{x}_k)}_{\substack{f\text{の}\boldsymbol{x}_k\text{周りでの}\\ \text{線形近似}}} + \underbrace{\frac{1}{2\varepsilon_k}\|\boldsymbol{x}-\boldsymbol{x}_k\|^2}_{\substack{\text{近接項}\\ \text{（線形近似が成り立つよう}\\ \text{解を}\boldsymbol{x}_k\text{の近傍で探索）}}} + g(\boldsymbol{x})$$

- 近接勾配法（proximal gradient method)とよぶ



講義の流れ

1. 凸最適化問題
2. 最急降下法
3. ニュートン法
4. 準ニュートン法

2階微分

- 勾配法では1階微分しか用いていない
- 2階微分も用いると反復回数を減らせるのでは？

$$\boldsymbol{x}^* = \operatorname*{argmin}_{\boldsymbol{x}\in\mathcal{X}} f(\boldsymbol{x})$$

- 1階微分：勾配ベクトル $\nabla f(\boldsymbol{x})$

$$[\nabla f(\boldsymbol{x})]_j = \frac{\partial f(\boldsymbol{x})}{\partial x^{(j)}} \quad \boldsymbol{x} = (x^{(1)}, \ldots, x^{(d)})^\top$$

- 2階微分：ヘッセ行列 $\nabla^2 f(\boldsymbol{x})$

$$[\nabla^2 f(\boldsymbol{x})]_{j,j'} = \frac{\partial^2 f(\boldsymbol{x})}{\partial x^{(j)} \partial x^{(j')}}$$

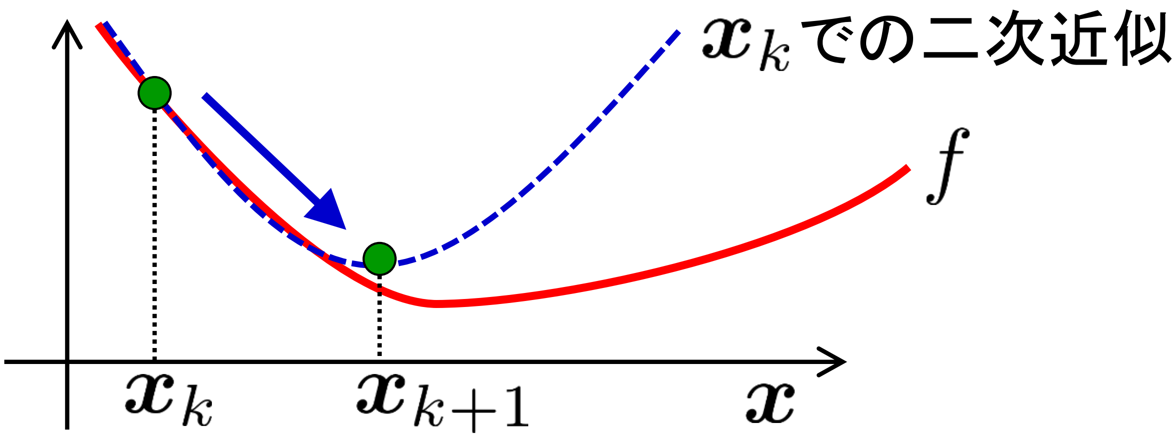
目的関数の二次近似

- 二次のテーラー展開(Taylor expansion)を用いて f を現在の解 $\boldsymbol{x}_k$ の周りで近似する:

$$f(\boldsymbol{x}) \approx f_k(\boldsymbol{x})$$

$$f_k(\boldsymbol{x}) = f(\boldsymbol{x}_k) + \nabla f(\boldsymbol{x}_k)^\top (\boldsymbol{x} - \boldsymbol{x}_k) + \frac{1}{2}(\boldsymbol{x} - \boldsymbol{x}_k)^\top \nabla^2 f(\boldsymbol{x}_k)(\boldsymbol{x} - \boldsymbol{x}_k)$$

- 二次近似を最小にする点に解を更新する



目的関数の二次近似

$$f_k(\boldsymbol{x}) = f(\boldsymbol{x}_k) + \nabla f(\boldsymbol{x}_k)^\top(\boldsymbol{x} - \boldsymbol{x}_k)$$
$$+\frac{1}{2}(\boldsymbol{x} - \boldsymbol{x}_k)^\top \nabla^2 f(\boldsymbol{x}_k)(\boldsymbol{x} - \boldsymbol{x}_k)$$

- f_k の勾配をゼロと置いた方程式

$$\nabla f_k(\boldsymbol{x}) = \nabla f(\boldsymbol{x}_k) + \nabla^2 f(\boldsymbol{x}_k)(\boldsymbol{x} - \boldsymbol{x}_k) = \boldsymbol{0}$$

を解けば, f_k の最小解が得られる:

$$\boldsymbol{x} = \boldsymbol{x}_k - (\nabla^2 f(\boldsymbol{x}_k))^{-1}\nabla f(\boldsymbol{x}_k)$$

- ヘッセ行列が逆を持たないときは, 単位行列の定数倍を加える(正則化, regularization):

$$(\nabla^2 f(\boldsymbol{x}_k) + \mu \boldsymbol{I})^{-1}, \quad \mu > 0$$

ニュートン法

- ニュートン法(Newton method):
  1. 適当に初期値 $\boldsymbol{x}_0$ を定める.
  2. $k = 0, 1, 2, \ldots$ に対して,
     収束するまで以下を繰り返す

$$\boldsymbol{x}_{k+1} = \boldsymbol{x}_k - \varepsilon_k (\nabla^2 f(\boldsymbol{x}_k))^{-1} \nabla f(\boldsymbol{x}_k)$$

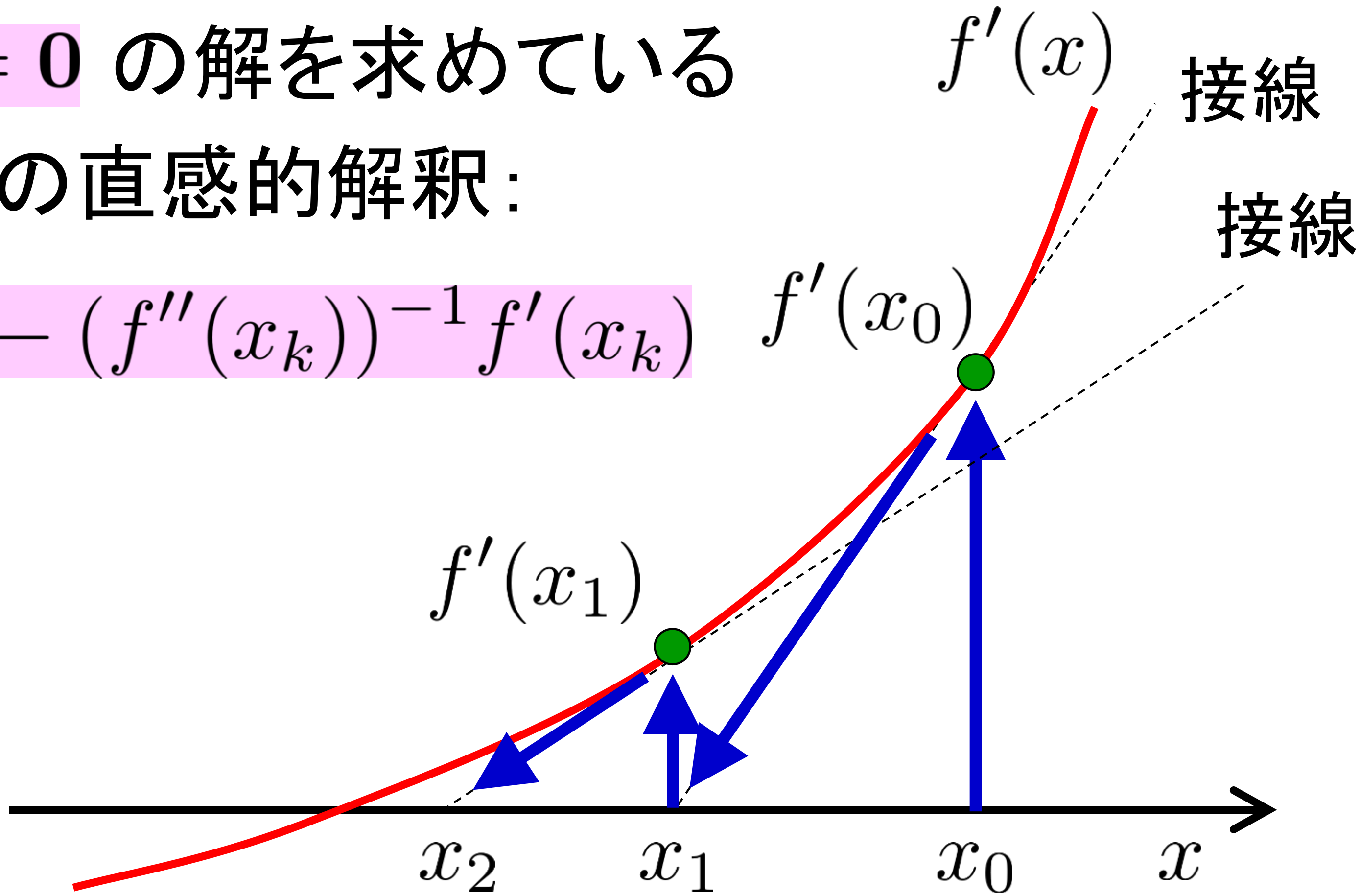
$0 < \varepsilon_k \leq 1$: ステップ幅

- f が二次関数とは限らないので, ステップ幅を導入して更新量を調整する

ニュートン法

- 一般にニュートン法という名称は，方程式の解を求めるアルゴリズムを指す
- 前ページのアルゴリズムでは $\nabla f(\boldsymbol{x}) = \boldsymbol{0}$ の解を求めている
- 1次元での直感的解釈：

$$x_{k+1} = x_k - (f''(x_k))^{-1} f'(x_k)$$



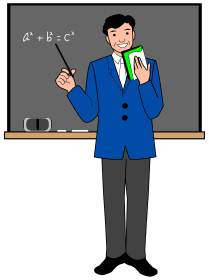
ニュートン法の収束性

- k が十分大きいとき，最適解 x^* の近傍で二次収束(quadratic convergence)する

$$\|\boldsymbol{x}_{k+1} - \boldsymbol{x}^*\| \leq c\|\boldsymbol{x}_k - \boldsymbol{x}^*\|^2$$

$$\boldsymbol{x}^* = \operatorname*{argmin}_{\boldsymbol{x} \in \mathcal{X}} f(\boldsymbol{x}) \qquad 0 < c < 1$$

- しかし，逆行列 $\nabla^2 f(\boldsymbol{x}_k)^{-1}$ を求める必要があるため，各反復の計算に時間がかかる



講義の流れ

1. 凸最適化問題
2. 最急降下法
3. ニュートン法
4. 準ニュートン法

ニュートン法の近似

- ニュートン法は，収束までの反復数は少なくて済むが，各反復でのヘッセ行列の逆

$$\nabla^2 f(\boldsymbol{x}_k)^{-1}$$

の計算に時間がかかる

- 勾配ベクトル $\nabla f(\boldsymbol{x})$ を用いて近似計算することにする

正割条件

- 目的関数 $f(\boldsymbol{x})$ の $\boldsymbol{x}_k$ 周りでの二次近似 $f_k(\boldsymbol{x})$:

$$f_k(\boldsymbol{x}) = f(\boldsymbol{x}_k) + \nabla f(\boldsymbol{x}_k)^\top (\boldsymbol{x} - \boldsymbol{x}_k) + \frac{1}{2}(\boldsymbol{x} - \boldsymbol{x}_k)^\top \nabla^2 f(\boldsymbol{x}_k)(\boldsymbol{x} - \boldsymbol{x}_k)$$

- その勾配は

$$\nabla f_k(\boldsymbol{x}) = \nabla f(\boldsymbol{x}_k) + \nabla^2 f(\boldsymbol{x}_k)(\boldsymbol{x} - \boldsymbol{x}_k)$$

- $\nabla f(\boldsymbol{x}_{k-1}) = \nabla f_k(\boldsymbol{x}_{k-1})$ のとき,

$$\nabla^2 f(\boldsymbol{x}_k)(\boldsymbol{x}_k - \boldsymbol{x}_{k-1}) = \nabla f(\boldsymbol{x}_k) - \nabla f(\boldsymbol{x}_{k-1})$$

を得る. これを正割条件(secant condition)という

準ニュートン法

- $\boldsymbol{H}_k$: ヘッセ行列 $\nabla^2 f(\boldsymbol{x}_k)$の近似
- 準ニュートン法(quasi-Newton method) : 正割条件

$$\boldsymbol{H}_k(\boldsymbol{x}_k - \boldsymbol{x}_{k-1}) = \nabla f(\boldsymbol{x}_k) - \nabla f(\boldsymbol{x}_{k-1})$$

を満たす $\boldsymbol{H}_k$ の中で性質の良いものを選ぶ

- 対称
- 正定値
- 逆行列が直接求められる

BFGSアルゴリズム

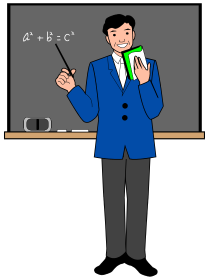
- BFGSアルゴリズム(Broyden-Fletcher-Goldfarb-Shanno algorithm):

$$\boldsymbol{H}_{k+1} = \boldsymbol{H}_k + \frac{\boldsymbol{t}_k \boldsymbol{t}_k^\top}{\boldsymbol{s}_k^\top \boldsymbol{t}_k} - \frac{\boldsymbol{H}_k \boldsymbol{s}_k \boldsymbol{s}_k^\top \boldsymbol{H}_k}{\boldsymbol{s}_k^\top \boldsymbol{H}_k \boldsymbol{s}_k}$$

$$\boldsymbol{s}_k = \boldsymbol{x}_{k+1} - \boldsymbol{x}_k \quad \boldsymbol{t}_k = \nabla f(\boldsymbol{x}_{k+1}) - \nabla f(\boldsymbol{x}_k)$$

  - $\boldsymbol{H}_{k+1}$ は常に対称
  - $\boldsymbol{H}_k$ が正定値かつ $\boldsymbol{s}_k^\top \boldsymbol{t}_k > 0$ ならば $\boldsymbol{H}_{k+1}$も正定値
  - 逆行列が直接求められる:

$$\boldsymbol{H}_{k+1}^{-1} = \boldsymbol{H}_k^{-1} + \frac{(\boldsymbol{s}_k^\top \boldsymbol{t}_k + \boldsymbol{t}_k^\top \boldsymbol{H}_k^{-1} \boldsymbol{t}_k) \boldsymbol{s}_k \boldsymbol{s}_k^\top}{(\boldsymbol{s}_k^\top \boldsymbol{t}_k)^2} - \frac{\boldsymbol{H}_k^{-1} \boldsymbol{t}_k \boldsymbol{s}_k^\top + \boldsymbol{s}_k \boldsymbol{t}_k^\top \boldsymbol{H}_k^{-1}}{\boldsymbol{s}_k^\top \boldsymbol{t}_k}$$



講義の流れ

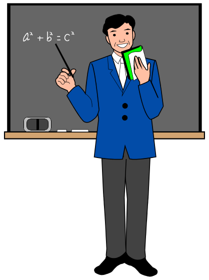
1. 凸最適化問題
2. 最急降下法
3. ニュートン法
4. 準ニュートン法

まとめ

凸最適化問題
- 目的関数が凸関数で定義域が凸集合
- 最適解が一意に定まる

最急降下法
- 勾配を降下するように値を更新
- ステップ幅の選択が重要
- 目的関数が微分不可能な場合は,
  劣勾配法や近接勾配法を用いる

ニュートン法
- 二階微分の情報を利用する
- ヘッセ行列の近似を用いる準ニュートン法が実用的



次回の予告

- 制約付き最適化

宿題1

$$g(\varepsilon_k) \le \alpha \varepsilon_k g'(0)$$

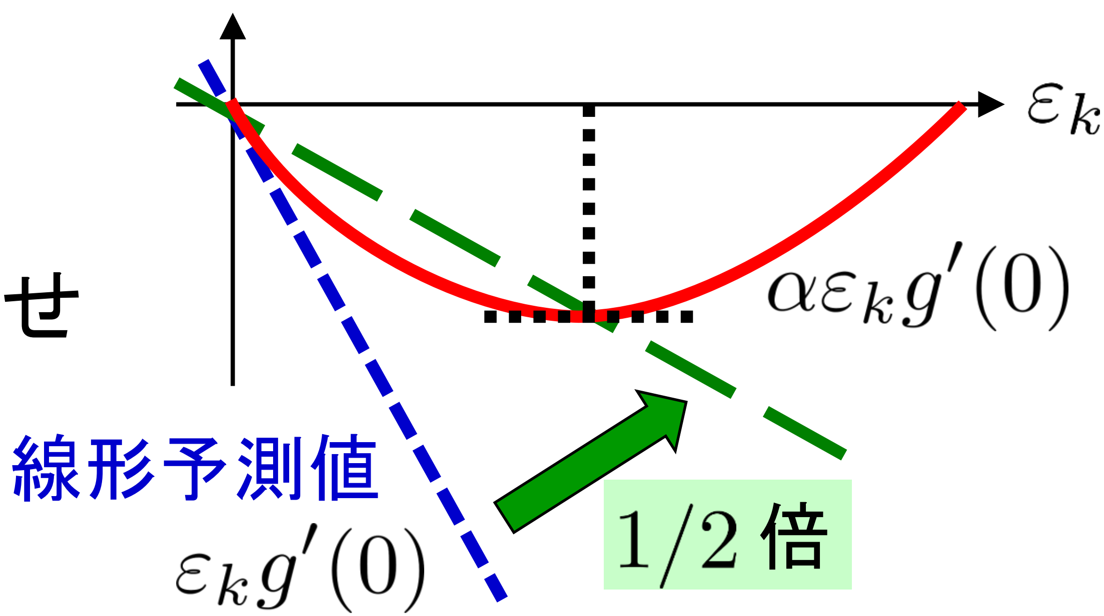
$$g(\varepsilon_k) = f(\boldsymbol{x}_k - \varepsilon_k \nabla f(\boldsymbol{x}_k)) - f(\boldsymbol{x}_k)$$

- $f(\boldsymbol{x}) = \dfrac{1}{2}\boldsymbol{x}^\top \boldsymbol{A}\boldsymbol{x}$ のとき，$\alpha = 1/2$ のアルミホ規準

を満たす最大の ε_k は，最適なステップ幅

$$\varepsilon_k = \frac{\boldsymbol{x}_k^\top \boldsymbol{A}^2 \boldsymbol{x}_k}{\boldsymbol{x}_k^\top \boldsymbol{A}^3 \boldsymbol{x}_k}$$

と一致することを示せ



宿題2

■ 二次関数 $f(\boldsymbol{x}) = \frac{1}{2}\boldsymbol{x}^\top \boldsymbol{A}\boldsymbol{x}$ に対する

厳密直線探索を用いた最急降下法は,

$$\|\boldsymbol{x}_{k+1} - \boldsymbol{x}^*\|_{\boldsymbol{A}} \leq c^2 \|\boldsymbol{x}_k - \boldsymbol{x}^*\|_{\boldsymbol{A}}$$

$$\|\boldsymbol{x}\|_{\boldsymbol{A}} = \boldsymbol{x}^\top \boldsymbol{A}\boldsymbol{x}$$

$$c = \frac{1 - \lambda_{\min}/\lambda_{\max}}{1 + \lambda_{\min}/\lambda_{\max}}$$

$\lambda_{\max} \geq \lambda_{\min} > 0$:
$\boldsymbol{A}$ の最大, 最小固有値

を満たすことを示せ

$$\boldsymbol{x}_{k+1} = \boldsymbol{x}_k - \varepsilon_k \nabla f(\boldsymbol{x}_k)$$

$$\nabla f(\boldsymbol{x}) = \boldsymbol{A}\boldsymbol{x}$$

$$\varepsilon_k = \frac{\|\nabla f(\boldsymbol{x}_k)\|^2}{\nabla f(\boldsymbol{x}_k)^\top \boldsymbol{A} \nabla f(\boldsymbol{x}_k)}$$

宿題2(続き)

- ヒント: カントロビッチの不等式 (Kantorovich's inequality)

$$\forall \boldsymbol{x} \neq \boldsymbol{0}, \quad \frac{4\lambda_{\min}\lambda_{\max}}{(\lambda_{\min}+\lambda_{\max})^2} \leq \frac{\|\boldsymbol{x}\|^4}{\boldsymbol{x}^\top \boldsymbol{A}^{-1}\boldsymbol{x} \cdot \boldsymbol{x}^\top \boldsymbol{A}\boldsymbol{x}}$$

$\lambda_{\max} \geq \lambda_{\min} > 0$:
$\boldsymbol{A}$の最大, 最小固有値

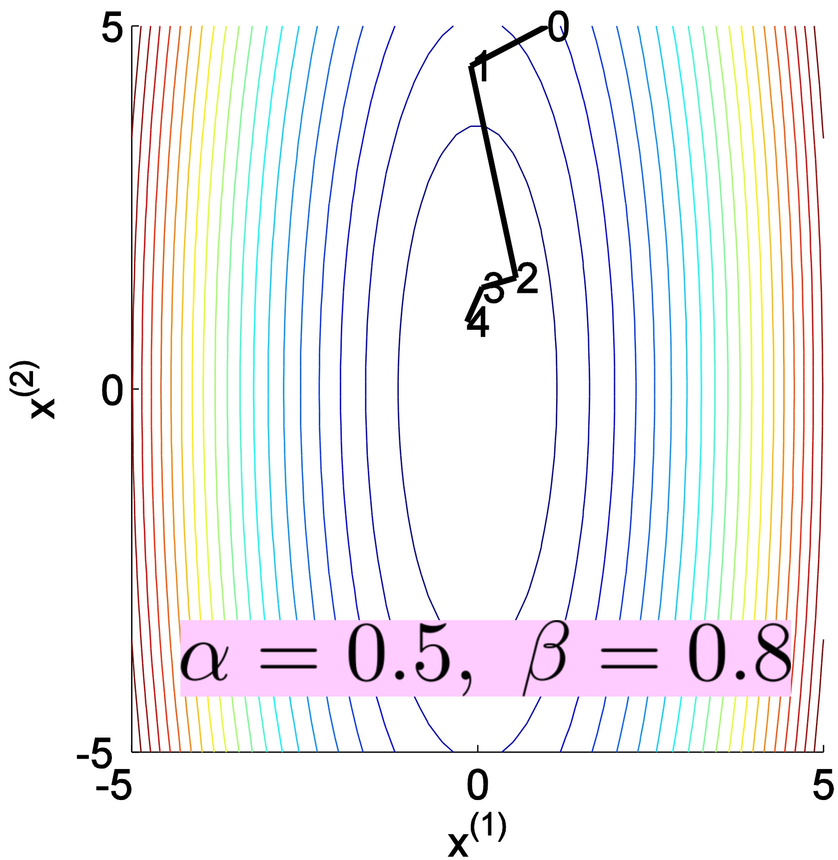
を用いる

宿題3

- バックトラック直線探索を用いた勾配法を実装し，以下の関数を最小化せよ:

$$f(x^{(1)}, x^{(2)}) = 10(x^{(1)})^2 + (x^{(2)})^2$$

  - 実行例:



宿題4

■BFGSアルゴリズムのヘッセ行列の更新式：

$$\boldsymbol{H}_{k+1} = \boldsymbol{H}_k + \frac{\boldsymbol{t}_k \boldsymbol{t}_k^\top}{\boldsymbol{s}_k^\top \boldsymbol{t}_k} - \frac{\boldsymbol{H}_k \boldsymbol{s}_k \boldsymbol{s}_k^\top \boldsymbol{H}_k}{\boldsymbol{s}_k^\top \boldsymbol{H}_k \boldsymbol{s}_k}$$

$$\boldsymbol{s}_k = \boldsymbol{x}_{k+1} - \boldsymbol{x}_k \quad \boldsymbol{t}_k = \nabla f(\boldsymbol{x}_{k+1}) - \nabla f(\boldsymbol{x}_k)$$

A) 逆行列の更新式を求めよ

$$\boldsymbol{H}_{k+1}^{-1} = \boldsymbol{H}_k^{-1} + \frac{(\boldsymbol{s}_k^\top \boldsymbol{t}_k + \boldsymbol{t}_k^\top \boldsymbol{H}_k^{-1} \boldsymbol{t}_k) \boldsymbol{s}_k \boldsymbol{s}_k^\top}{(\boldsymbol{s}_k^\top \boldsymbol{t}_k)^2} - \frac{\boldsymbol{H}_k^{-1} \boldsymbol{t}_k \boldsymbol{s}_k^\top + \boldsymbol{s}_k \boldsymbol{t}_k^\top \boldsymbol{H}_k^{-1}}{\boldsymbol{s}_k^\top \boldsymbol{t}_k}$$

B) $\boldsymbol{H}_k$が正定値かつ$\boldsymbol{s}_k^\top \boldsymbol{t}_k > 0$のとき，
$\boldsymbol{H}_{k+1}$も正定値であることを示せ

宿題4（ヒント）

A) シャーマン・モリソン公式 (Sherman–Morrison formula)を用いる:

$c \neq \boldsymbol{b}^\top \boldsymbol{A}^{-1}\boldsymbol{b}$ のとき $\boldsymbol{B} = \boldsymbol{A} - \frac{1}{c}\boldsymbol{b}^\top \boldsymbol{b}$ に対して

$$\boldsymbol{B}^{-1} = \boldsymbol{A}^{-1} + \frac{\boldsymbol{A}^{-1}\boldsymbol{b}\boldsymbol{b}^\top \boldsymbol{A}^{-1}}{c - \boldsymbol{b}^\top \boldsymbol{A}^{-1}\boldsymbol{b}}$$

B) 更新式の別表現を用いる:

$$\boldsymbol{H}_{k+1}^{-1} = \left(\boldsymbol{I} - \frac{\boldsymbol{s}_k \boldsymbol{t}_k^\top}{\boldsymbol{s}_k^\top \boldsymbol{t}_k}\right) \boldsymbol{H}_k^{-1} \left(\boldsymbol{I} - \frac{\boldsymbol{t}_k \boldsymbol{s}_k^\top}{\boldsymbol{s}_k^\top \boldsymbol{t}_k}\right) + \frac{\boldsymbol{s}_k \boldsymbol{s}_k^\top}{\boldsymbol{s}_k^\top \boldsymbol{t}_k}$$