

大規模無制約最適化問題に対する 数値解法アルゴリズム

～基礎から微分不可能な正則化項を含む問題への拡張まで～

慶應義塾大学 成島康史

2023年8月8日

RIMS共同研究「組合せ最適化セミナー」

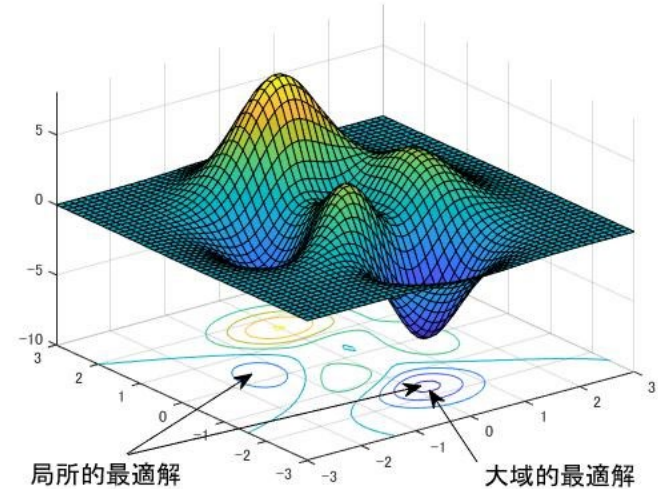
目次

- 第1部 (10:00~11:00)
 - 無制約最適化の基礎
 - 反復法
 - 直線探索
 - 最急勾配法, ニュートン法, 準ニュートン法
- 第2部 (11:15~12:15)
 - 大規模無制約最適化問題に対する反復法
 - 記憶制限準ニュートン法, BB法
 - 非線形共役勾配法
 - メモリーレス準ニュートン法
- 第3部 (13:25~14:25)
 - 微分不可能な正則化項を含む問題への拡張
 - 近接勾配法
 - ニュートン型近接勾配法

無制約最適化問題

無制約最適化問題

minimize $f(x)$ $f: R^n \rightarrow R$



無制約最適化問題のイメージ図

- 任意の $x \in R^n$ に対して, $f(x^*) \leq f(x)$ となる点 $x^* \in S$ を(大域的)最適解という.
- ある近傍 $B(x^*) = \{x \mid \|x^* - x\| \leq \varepsilon\}$ が存在して, 任意の $x \in B(x^*)$ に対して $f(x^*) \leq f(x)$ となる点 x^* を局所的最適解という.

➡ 局所的最適解 (の一つ) を見つけるアルゴリズムを考える
(目的関数の凸性などがあれば大域的最適解の保証あり)

反復法

minimize $f(x)$

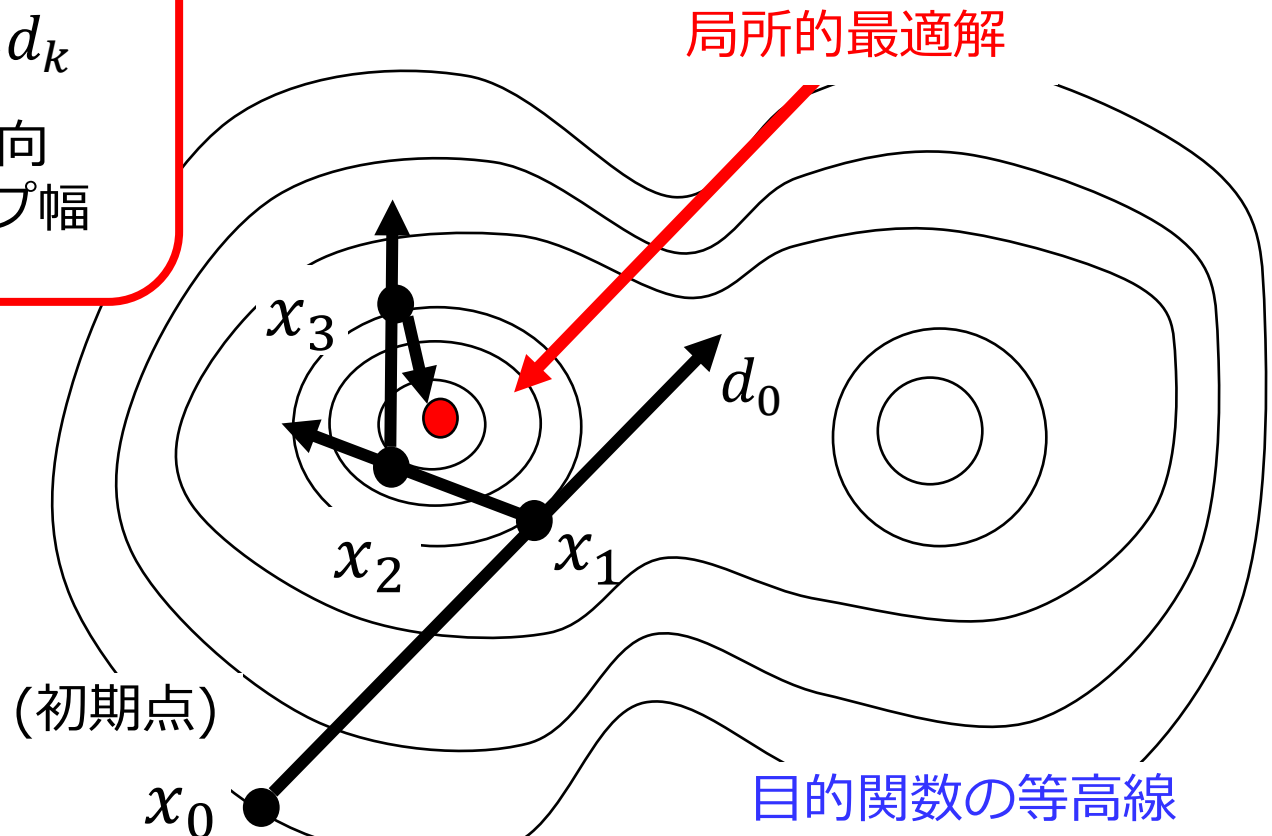
反復法：初期点 $x_0 \in R^n$ からスタートして
逐次的に最適解に近づけていく
数値解法

反復式

$$x_{k+1} = x_k + \alpha_k d_k$$

$d_k \in R^n$: 探索方向

$\alpha_k > 0$: ステップ幅



反復法のアルゴリズム

アルゴリズム (プロトタイプ)

1. 初期点 $x_0 \in R^n$ を与える. $k = 0$ として, ステップ 2. へ進む.
2. 終了判定を行い終了判定条件を満たしていれば終了する.
3. 探索方向 d_k を計算する.
4. 直線探索により ステップ幅 α_k を計算する.
5. $x_{k+1} = x_k + \alpha_k d_k$ により点列を更新.
6. $k \leftarrow k + 1$ としてステップ 2. に戻る.

ステップ 2. の終了判定では, 1次の最適性条件を満たしているかを判定する. 例えば,

- $\|\nabla f(x_k)\|$ が十分小さくなったら終了
- などを使用する

アルゴリズムの大域的収束性

大域的収束性 : 任意の初期点からスタートした時に生成される点列 $\{x_k\}$ が
局所的最適解 x^* に収束する性質
* 「大域的最適解に収束する性質」ではないことに注意

定理 (無制約最適化問題の最適性条件)

1. x^* が無制約最適化問題の局所最適解ならば
 - $\nabla f(x^*) = 0$ (1次の必要条件)
 - $\nabla^2 f(x^*)$ は半正定値 (2次の必要条件)
2. x^* が以下の条件を満たすならば、 x^* は局所最適解である(2次の十分条件)
 - $\nabla f(x^*) = 0$
 - $\nabla^2 f(x^*)$ は正定値

$\nabla f(x) = 0$ を満たす点は解の候補 $\rightarrow \nabla f(x) = 0$ を満たす点は**停留点**と呼ばれる

実際には $\lim_{k \rightarrow \infty} \|x_k - x^*\| = 0$ の代わりに


$$\lim_{k \rightarrow \infty} \|\nabla f(x_k)\| = 0, \quad \liminf_{k \rightarrow \infty} \|\nabla f(x_k)\| = 0$$

をもって大域的収束と呼ぶことが多い

アルゴリズムの局所的収束性

局所的収束性 : 点列 $\{x_k\}$ が解の近くにいるときに, どのように収束するかを表す性質

* 主に収束する速さを考えることが多い

- (q-)1次収束: 正定数 $c < 1$ が存在して以下を満たす

$$\|x_{k+1} - x^*\| \leq c \|x_k - x^*\|$$

- (q-)2次収束: 正定数 c が存在して以下を満たす

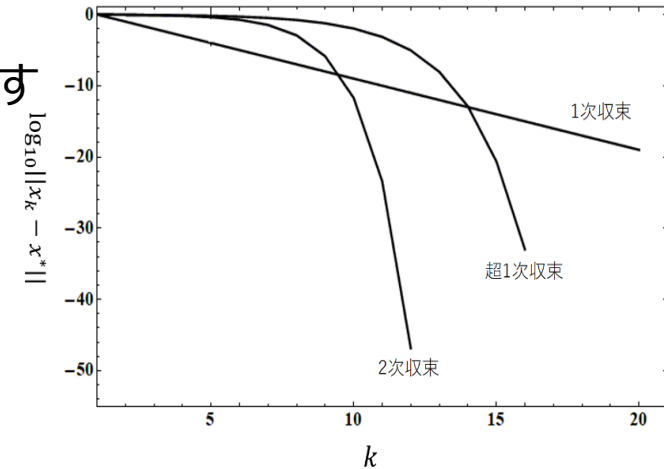
$$\|x_{k+1} - x^*\| \leq c \|x_k - x^*\|^2$$

- (q-)超1次収束: 0に収束する正の数数列 c_k が存在して以下を満たす

$$\|x_{k+1} - x^*\| \leq c_k \|x_k - x^*\|$$

- r-p次収束: 正定数 c とq-p次収束する数列 ξ_k が存在して以下を満たす

$$\|x_k - x^*\| \leq c \xi_k$$



無制約最適化問題に対する反復法

無制約最小化問題

最小化 $f(x)$

アルゴリズム (プロトタイプ)

1. 初期点 $x_0 \in R^n$ を与える. $k = 0$ として, ステップ 2. へ進む.
2. $\|\nabla f(x_k)\|$ が十分小さくなったら終了する.
3. 探索方向 d_k を計算する.
4. 直線探索により ステップ幅 α_k を計算する.
5. $x_{k+1} = x_k + \alpha_k d_k$ により点列を更新.
6. $k \leftarrow k + 1$ としてステップ 2. に戻る.

降下方向

方向微係数が負, つまり,

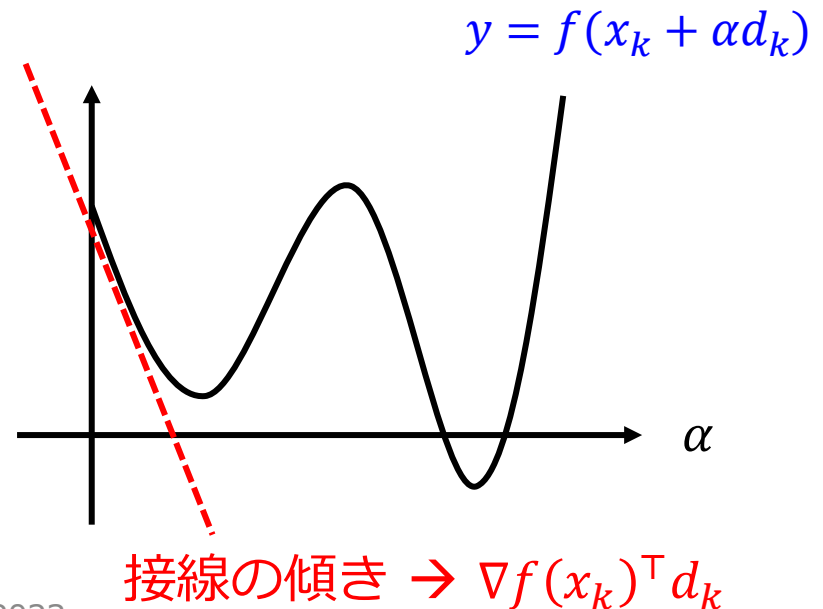
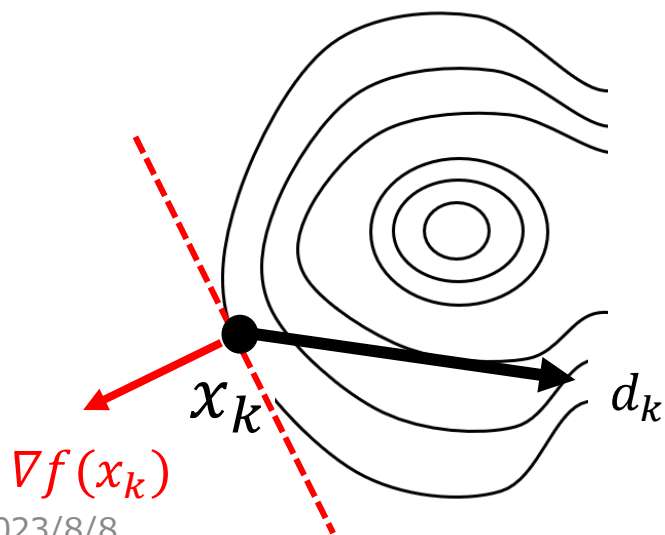
$$\nabla f(x_k)^\top d_k < 0$$

であるとき, d_k は**降下方向**であるという. さらに, ある正の定数 c が存在して

$$\nabla f(x_k)^\top d_k \leq -c \|\nabla f(x_k)\|^2$$

を満たすとき, d_k は**十分な降下方向**であるという.

$$\nabla f(x_k)^\top d_k \rightarrow 0 \text{ の時, } \|\nabla f(x_k)\| \rightarrow 0 \text{ を保証}$$



直線探索(正確な直線探索)

d_k 方向に最小となるステップ幅

$$\alpha_k = \operatorname{argmin}_{\alpha > 0} \{f(x_k + \alpha d_k)\}$$

を選択することを正確な直線探索と呼ぶ.

例) 狭義2次凸関数

$$f(x) = \frac{1}{2} x^\top A x + b^\top x \quad (A \text{は正定値対称行列})$$

の場合, 正確なステップ幅は

$$\alpha_k = -\frac{(Ax_k + b)^\top d_k}{d_k^\top A d_k}$$

で計算可能.

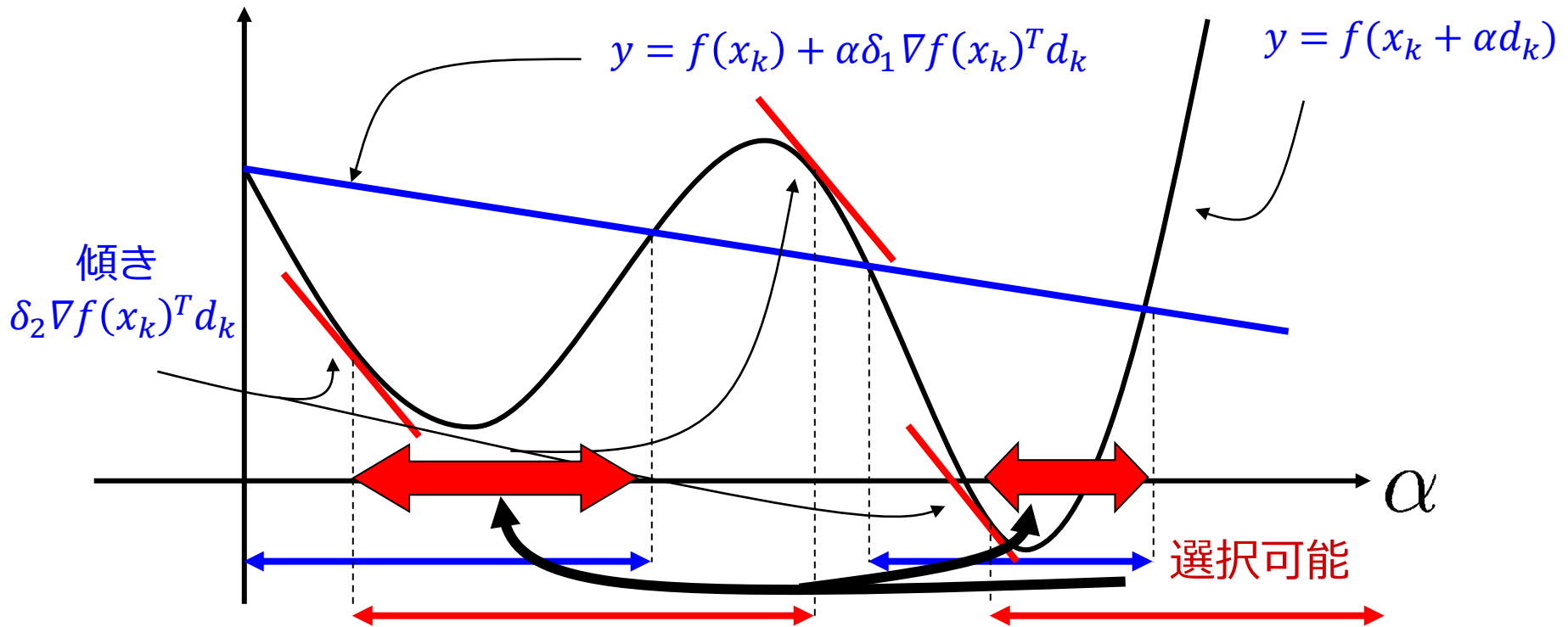
直線探索 (Wolfe条件)

直線探索においては, ステップ幅 α_k に何らかの条件を課すことが多い

Wolfe条件

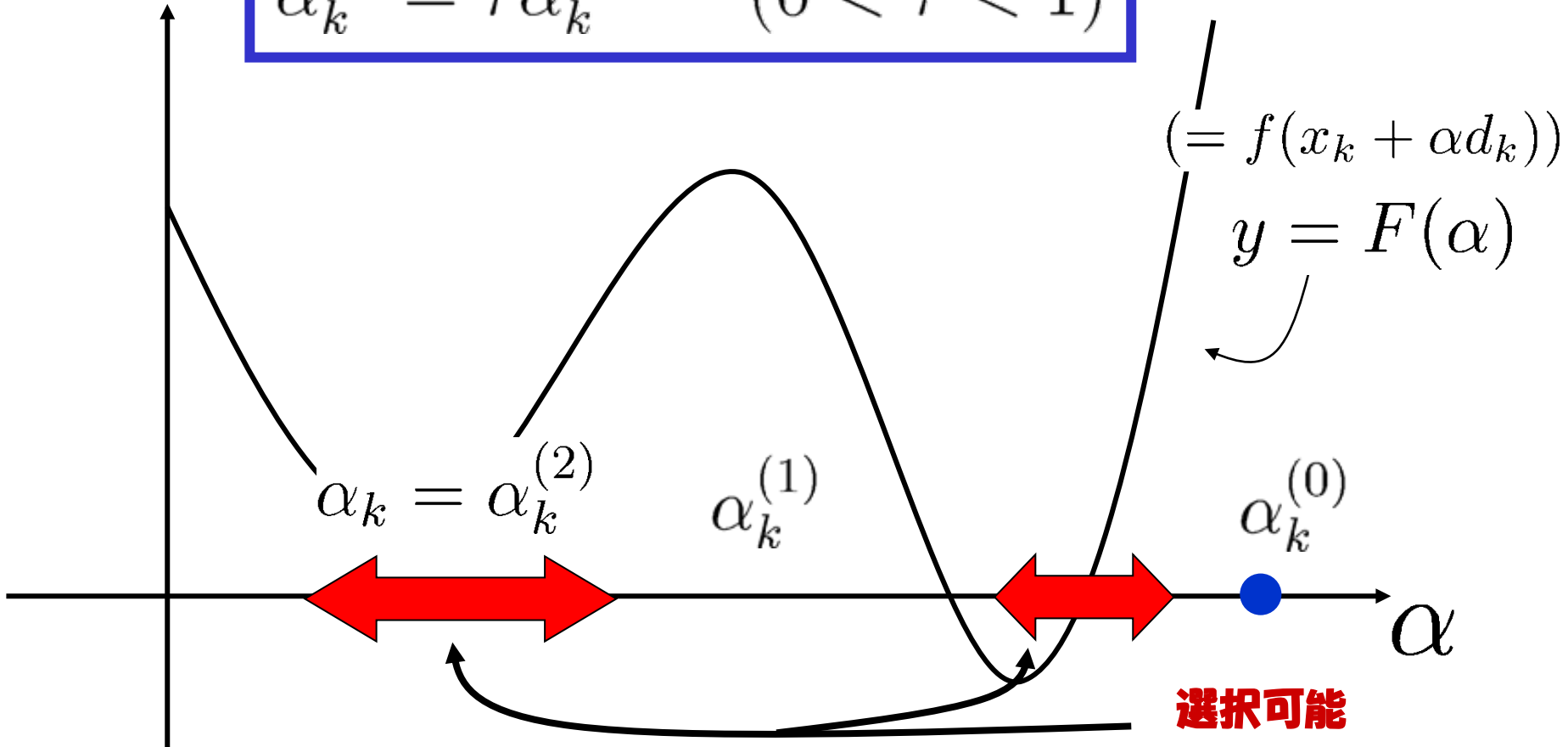
$$f(x_k + \alpha d_k) \leq f(x_k) + \alpha \delta_1 \nabla f(x_k)^\top d_k \quad (\text{単独でArmijo条件})$$

$$\delta_2 \nabla f(x_k)^\top d_k \leq \nabla f(x_k + \alpha d_k)^\top d_k \quad (0 < \delta_1 < \delta_2 < 1)$$



直線探索(分割法)

$$\alpha_k^{(i)} = \tau \alpha_k^{(i+1)} \quad (0 < \tau < 1)$$



反復法の収束性に関連する仮定

仮定

目的関数 f は下に有界, かつ初期点 x_0 における準位集合 $\mathcal{L} = \{x \mid f(x) \leq f(x_0)\}$ の適当な凸近傍 \mathcal{N} において連続微分可能で, 勾配 ∇f はリプシッツ連続, つまり, ある正の定数 L が存在して, 任意の $u, v \in \mathcal{N}$ に対して

$$\|\nabla f(u) - \nabla f(v)\| \leq L\|u - v\|$$

が成立する.

- ∇f の(定数 L を伴った)リプシッツ連続は **L -平滑** とも呼ばれる.
- L -平滑な関数 f に対して

$$f(y) \leq f(x) + \nabla f(x)^\top (y - x) + \frac{L}{2} \|x - y\|^2 \quad \text{for } \forall x, y \in \mathcal{N}$$

が成立する.

ズーテンダイク条件

ズーテンダイク(Zoutendjik)条件

前述の仮定が成り立っているとする。探索方向が降下方向であり、直線探索で Wolfe条件を満たす任意の反復法において、以下が成立する：

$$\left(\sum_{k=0}^{\infty} \|\nabla f(x_k)\|^2 \cos^2 \theta_k = \right) \sum_{k=0}^{\infty} \frac{(\nabla f(x_k)^\top d_k)^2}{\|d_k\|^2} < \infty.$$

- 勾配ベクトルと探索方向のなす角を θ_k とすると $\cos \theta_k = \frac{\nabla f(x_k)^\top d_k}{\|\nabla f(x_k)\| \|d_k\|}$

$$\Rightarrow \sum_{k=0}^{\infty} \|\nabla f(x_k)\|^2 \cos^2 \theta_k < +\infty$$

$$\Rightarrow \exists \bar{c} > 0, \text{ s. t. } \cos \theta_k \geq \bar{c} \text{ for all } k \geq 0 \quad \rightarrow \quad \lim_{k \rightarrow \infty} \|\nabla f(x_k)\| = 0$$

大域的収束性の例

定理

ズーテンダイク条件のときと同様の仮定が成り立っており、

探索方向が十分な降下方向

であるとする。さらに、正の定数 γ が存在して、すべての k に対して、

$$\|d_k\| \leq \gamma \|\nabla f(x_k)\|^b \quad (0 < b \leq 2) \quad (\text{探索方向の有界性})$$

が成立するならば、アルゴリズムによって生成される点列は、ある k 番目の反復において $\nabla f(x_k) = 0$ となるか、もしくは、

$$\liminf_{k \rightarrow \infty} \|\nabla f(x_k)\| = 0$$

の意味で大域的に収束する。

注意：ある k 番目で $\nabla f(x_k) = 0$ となる場合は、そこでアルゴリズムが停止するので、以降では $\nabla f(x_k) \neq 0$ for all $k \geq 0$ として考える。

大域的収束性の例

- 十分な降下条件： $\nabla f(x_k)^\top d_k \leq -c \|\nabla f(x_k)\|^2$
- 探索方向の有界性： $\|d_k\| \leq \gamma \|\nabla f(x_k)\|^b$ ($0 \leq b < 2$)
- ズーテンダイク条件： $\sum_{k=0}^{\infty} \frac{(\nabla f(x_k)^\top d_k)^2}{\|d_k\|^2} < \infty$

証明

背理法の仮定として $\liminf_{k \rightarrow \infty} \|\nabla f(x_k)\| \neq 0$ とすると、 $\varepsilon > 0$ が存在して、

$$\|\nabla f(x_k)\| \geq \varepsilon \quad (\forall k \geq 0)$$

が成立する。したがって、

$$\frac{(\nabla f(x_k)^\top d_k)^2}{\|d_k\|^2} \geq \frac{c^2 \|\nabla f(x_k)\|^4}{\gamma^2 \|\nabla f(x_k)\|^{2b}} = \frac{c^2 \|\nabla f(x_k)\|^{2(2-b)}}{\gamma^2} \geq \frac{c^2 \varepsilon^{2(2-b)}}{\gamma^2}$$

となり、 $\sum_{k=0}^{\infty} \frac{(\nabla f(x_k)^\top d_k)^2}{\|d_k\|^2} = \infty$ が成立するが、これはズーテンダイク条件に反する。

よって証明終了。□

Armijo条件下での大域的収束性

Armijo条件のみの場合の大域的収束性を考えてみる

- 十分な降下条件: $\nabla f(x_k)^\top d_k \leq -c \|\nabla f(x_k)\|^2$
- 探索方向の有界性: $\|d_k\| \leq \gamma \|\nabla f(x_k)\|$
- Armijo条件: $f(x_k + \alpha d_k) \leq f(x_k) + \alpha \delta_1 \nabla f(x_k)^\top d_k$
- ステップ幅の下限が存在: $\exists \bar{\alpha} > 0, \text{ s.t. } \bar{\alpha} \leq \alpha_k \text{ for all } k$

証明 上記の条件より

$$f(x_k) - f(x_{k+1}) \geq -\alpha_k \delta_1 \nabla f(x_k)^\top d_k \geq \bar{\alpha} c \delta_1 \|\nabla f(x_k)\|^2 > 0$$

が成立する. 一方, $\{f(x_k)\}$ は下に有界な単調減少列なので, 極限が存在するので, 上式より

$$\lim_k \|\nabla f(x_k)\| = 0$$

が成立する.

無制約最適化問題に対する反復法

最小化 $f(x)$

反復法

$$x_{k+1} = x_k + \alpha_k d_k$$

代表的な方法 (探索方向の選択法)

- **最急降下法** $d_k = -\nabla f(x_k)$
1次収束する方法だが、(特に悪条件な問題に対して) 効率が悪い
- **ニュートン法** $\nabla^2 f(x_k) d_k = -\nabla f(x_k)$ または $d_k = -\nabla^2 f(x_k)^{-1} \nabla f(x_k)$
2次収束する方法だが、探索方向が**目的関数値の下がる方向(降下方向)**とは限らない
($\nabla f(x_k)^T d_k < 0$)
- **準ニュートン法** $B_k d_k = -\nabla f(x_k)$ または $d_k = -H_k \nabla f(x_k)$
ただし、 $B_k \approx \nabla^2 f(x_k)$, $H_k = B_k^{-1}$
(近似行列をうまく選べば)降下方向を生成し、さらに超1次収束する方法
ただし、 B_k, H_k は密行列なので、大規模な問題には不向き)

最急降下法

最急降下法: 目的関数の1次近似を(局所的に)最小にする方向を選択

$$1\text{次近似: } f(x_k + d) \approx \ell(d) = f(x_k) + \nabla f(x_k)^T d$$

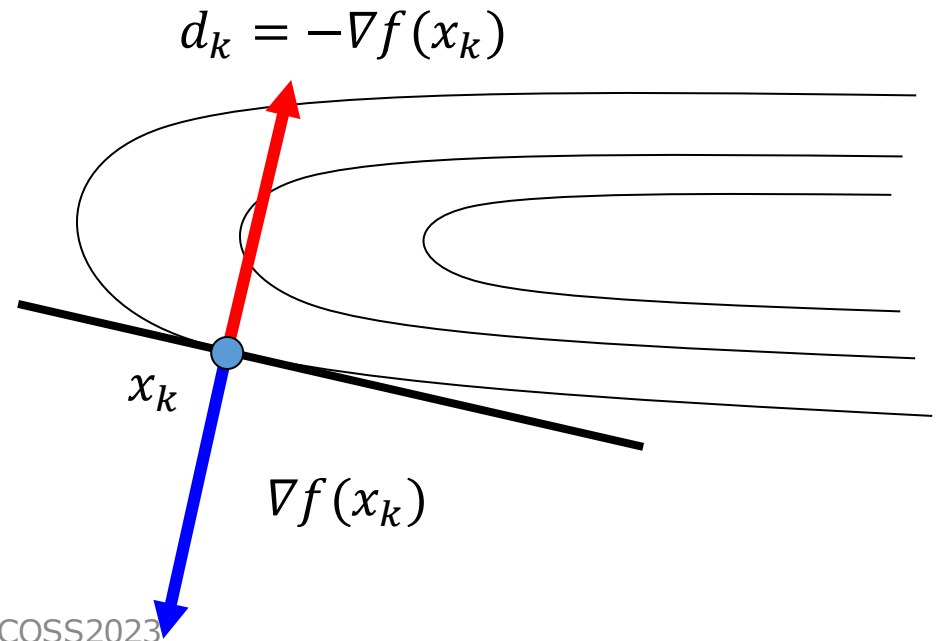
1次近似は d について線形なので, いくらでも小さくできる
→ $\|d\|$ が一定と仮定して $\ell(d)$ を最小化する d を選択

$$\nabla f(x_k)^T d = \|\nabla f(x_k)\| \|d\| \cos \theta \rightarrow \cos \theta = -1 \text{ のときが最小}$$



最急降下方向

$$d_k = -\nabla f(x_k)$$



最急降下法の収束性

大域的収束性

$d_k = -\nabla f(x_k)$ なので,

$$\nabla f(x_k)^\top d_k = -\|\nabla f(x_k)\|^2, \quad \|d_k\| = \|\nabla f(x_k)\|$$

が成立する. よって, ズーテンダイク条件と同様の仮定の下で $\liminf_{k \rightarrow \infty} \|\nabla f(x_k)\| = 0$

が成り立つ.

凸二次関数に対する収束率

目的関数が $f(x) = \frac{1}{2}x^\top Qx + c^\top x$ (Q は正定値対称行列)であるとする. この時,

$$\alpha_k = \operatorname{argmin}_{\alpha > 0} f(x_k + \alpha d_k) \quad \text{正確な直線探索}$$

とした最急降下法によって生成される点列の収束率は以下のとおりである:

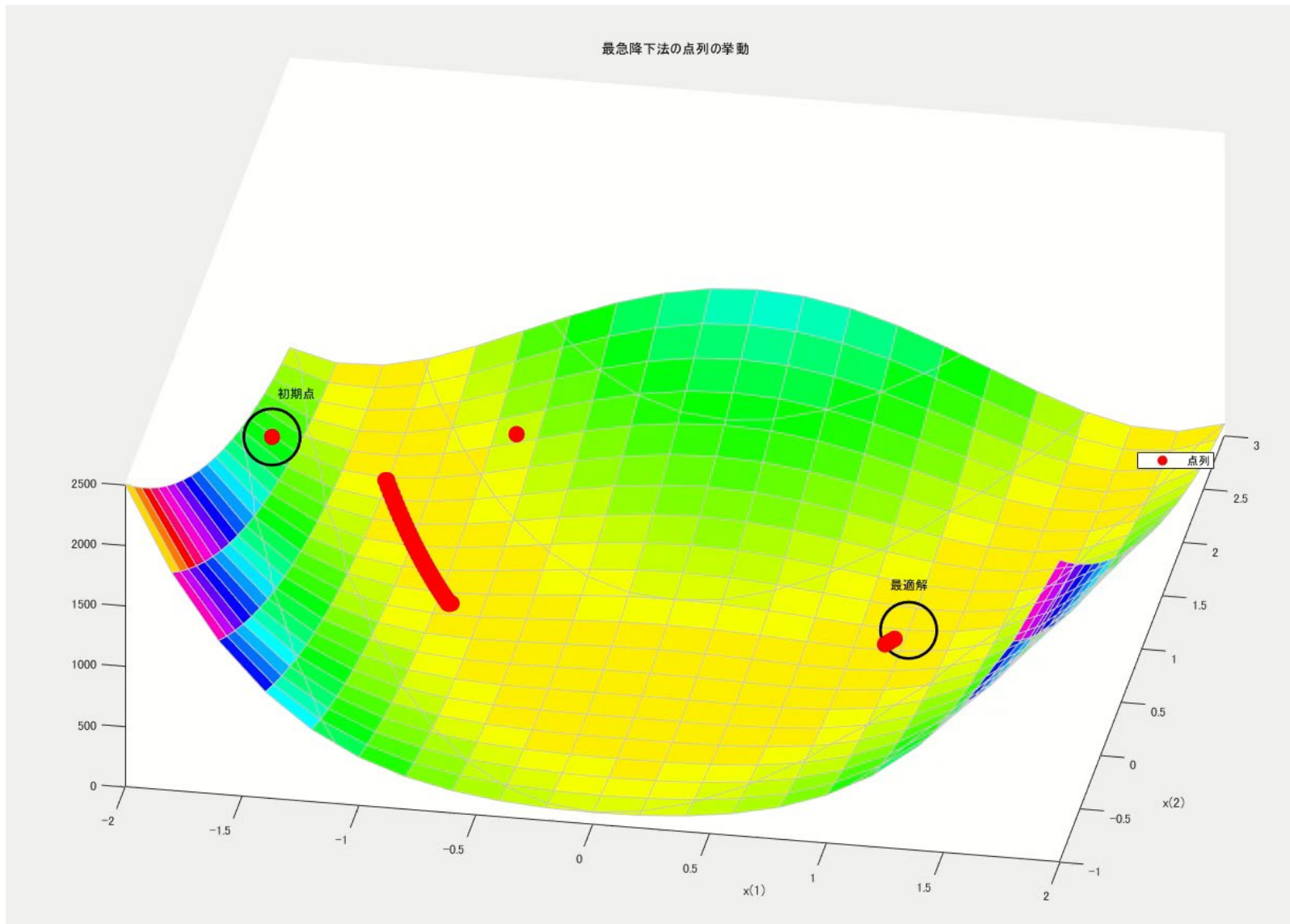
$$\|x_{k+1} - x^*\|_A \leq \left(\frac{\lambda_{\max} - \lambda_{\min}}{\lambda_{\max} + \lambda_{\min}} \right) \|x_k - x^*\|_A$$



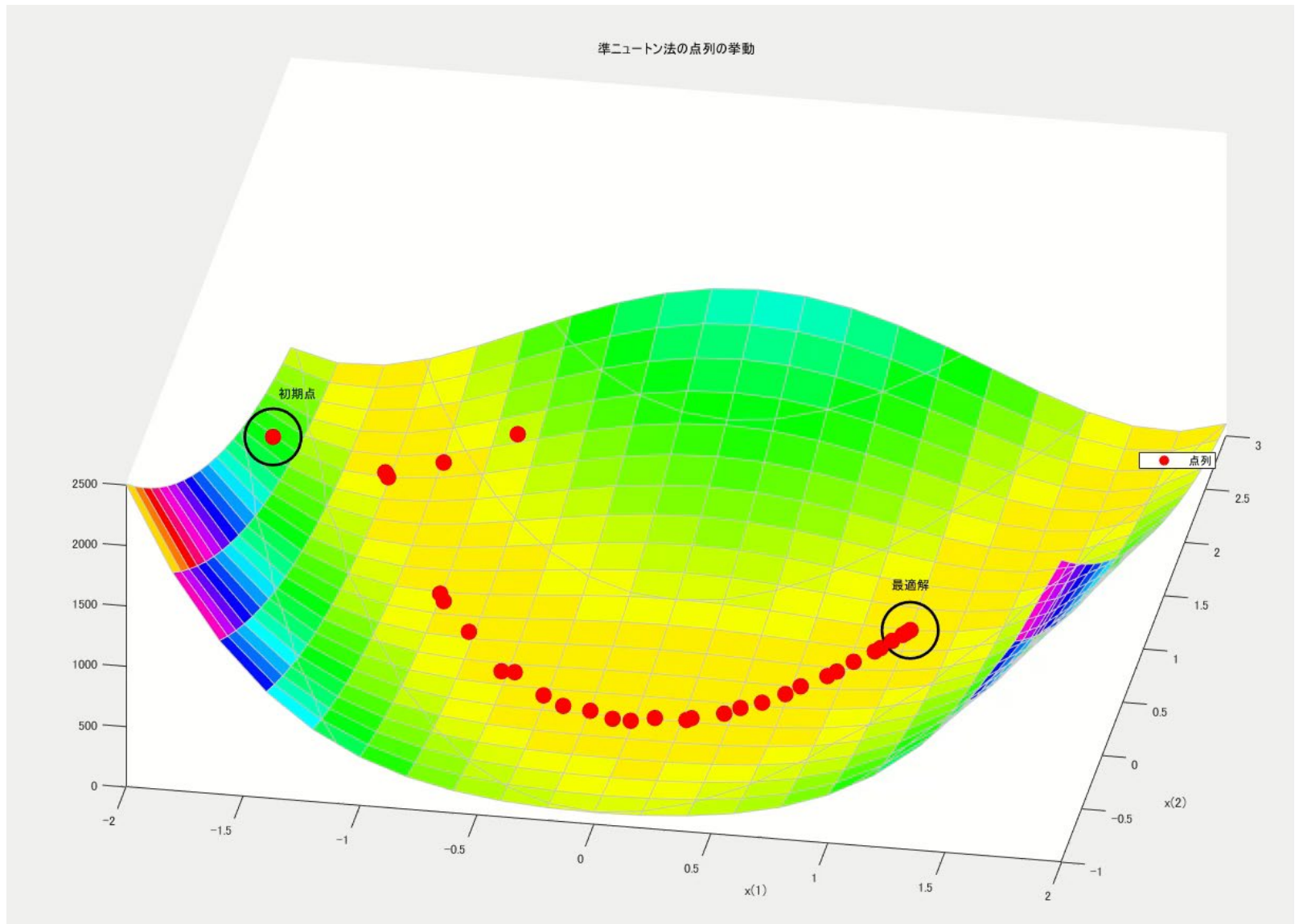
$\lambda_{\max} \gg \lambda_{\min}$ の時に
収束が非常に遅くなる

ただし, λ_{\max} と λ_{\min} は Q の最大固有値と最小固有値であり, $\|x\|_A = \sqrt{x^\top Qx}$ である.

点列の挙動(最急降下法)



点列の挙動(準ニュートン法)

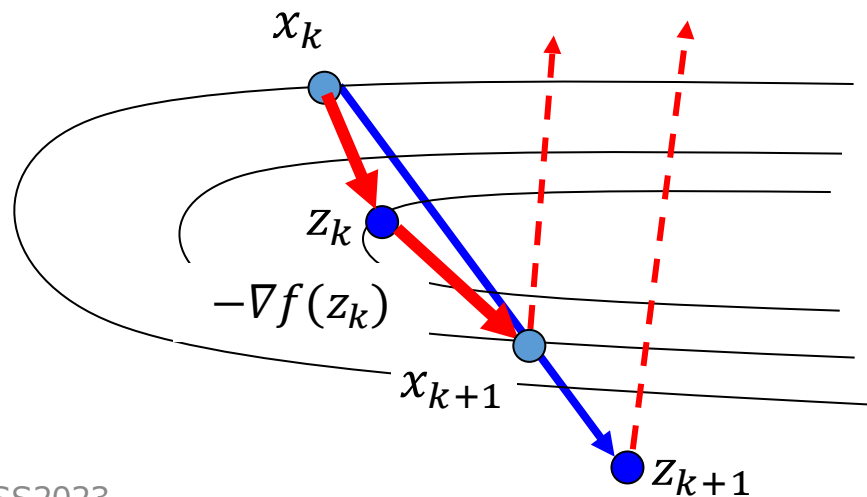


Nesterovの加速法

1. 初期点 $x_0 \in R^n$ を与える. $k = 0, z_0 = x_0, \tau_0 = 0$, として, ステップ2へ.
2. 終了判定を行う.
3. $x_{k+1} = z_k - \frac{1}{L} \nabla f(z_k)$ (定数ステップサイズの最急降下法)
4. $\tau_{k+1} = \frac{1}{2} \left(1 + \sqrt{1 + 4\tau_k^2} \right)$
5. $z_{k+1} = x_{k+1} + \frac{\tau_{k-1}}{\tau_{k+1}} (x_{k+1} - x_k)$
6. $k \leftarrow k + 1$ としてステップ2. に戻る.

- f が L -平滑な凸関数のとき

$$f(x_k) - f(x^*) = O\left(\frac{1}{k^2}\right)$$



ニュートン法

ニュートン法: 目的関数の2次近似を最小にする方向を選択

$$2\text{次近似: } f(x_k + d) \approx q(d) = f(x_k) + \nabla f(x_k)^\top d + \frac{1}{2} d^\top \nabla^2 f(x_k) d$$

ヘッセ行列 $\nabla^2 f(x_k)$ が正定値対称であると仮定すると, $\nabla_d q(d) = 0$ を考えればよいので

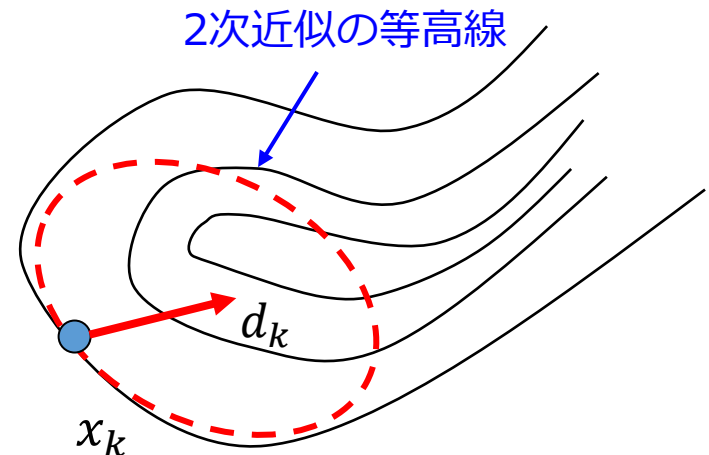
$$\nabla_d q(d) = \nabla f(x_k) + \nabla^2 f(x_k)^\top d = 0$$

ニュートン方向

$$\nabla^2 f(x_k) d_k = -\nabla f(x_k), \text{ または } d_k = -\nabla^2 f(x_k)^{-1} \nabla f(x_k)$$



- 局所的に2次収束する
- 降下方向を生成するとは限らない
- ヘッセ行列 (2回偏微分) を計算する必要がある



準ニュートン法

準ニュートン法: ニュートン法のヘッセ行列を近似行列で置き換えた方法

準ニュートン法

$$B_k d_k = -\nabla f(x_k) \text{ または } d_k = -H_k \nabla f(x_k)$$

$$\text{ただし, } B_k \approx \nabla^2 f(x_k), \quad H_k = B_k^{-1}$$

B_k, H_k は正定値対称であることが望ましい

B_k が正定値対称であるならば, H_k も正定値対称であり,

$$\nabla f(x_k)^\top d_k = -\nabla f(x_k)^\top H_k \nabla f(x_k) < 0$$

となるので, 降下方向を生成する.

準ニュートン法

定理 直線探索でWolfe条件を用いた準ニュートン法を考える。ズーテンダイク条件と同様の仮定が成り立っており、さらに、正の定数 M, m が存在して、すべての k に対して

$$m\|v\|^2 \leq v^\top H_k v \leq M\|v\|^2 \quad \forall v \in R^n$$

が成立するならば、準ニュートン法は $\liminf_{k \rightarrow \infty} \|\nabla f(x_k)\| = 0$ の意味で大域的に収束する。

- 十分な降下条件： $\nabla f(x_k)^\top d_k = -\nabla f(x_k)^\top H_k \nabla f(x_k) \leq -m\|\nabla f(x_k)\|^2$
- 探索方向の有界性： $\|d_k\| \leq \|-H_k \nabla f(x_k)\| \leq \|H_k\| \|\nabla f(x_k)\| \leq M\|\nabla f(x_k)\|$

準ニュートン法の収束率

定理 (Dennis-Moré条件)

目的関数 f は2回連続微分可能であるとする. 点列 $\{x_k\}$ は

$$x_{k+1} = x_k - B_k^{-1} \nabla f(x_k)$$

で生成され, さらに2次の十分条件を満たす局所最適化 x^* に収束するものと仮定する. このとき, B_k が

$$\lim_{k \rightarrow \infty} \frac{\| (B_k - \nabla^2 f(x_k))(x_{k+1} - x_k) \|}{\|x_{k+1} - x_k\|} = 0$$

を満たすならば, 点列 $\{x_k\}$ の収束率は超一次である.

近似行列の選択とセカント条件

近似行列 B_k の選択

$\nabla f(x_k)$ の1次近似を考えると, $\nabla f(x_k) \approx \nabla f(x_{k+1}) - \nabla^2 f(x_{k+1})(x_{k+1} - x_k)$

$$\begin{aligned} \Rightarrow \quad \nabla^2 f(x_{k+1})(x_{k+1} - x_k) &\approx \nabla f(x_{k+1}) - \nabla f(x_k) \\ &= s_k \qquad \qquad \qquad = y_k \end{aligned}$$

セカント条件 (近似行列が満たすべき条件)

$$B_{k+1}s_k = y_k, \quad \text{または} \quad s_k = H_{k+1}y_k$$

セカント条件を満たす更新公式

BFGS(Broyden-Fletcher-Goldfarb-Shanno) 公式

$$B_{k+1} = B_k - \frac{B_k s_k s_k^\top B_k}{s_k^\top B_k s_k} + \frac{y_k y_k^\top}{s_k^\top y_k} \qquad H_{k+1} = \left(I - \frac{s_k y_k^\top}{s_k^\top y_k} \right) H_k \left(I - \frac{y_k s_k^\top}{s_k^\top y_k} \right) + \frac{s_k s_k^\top}{s_k^\top y_k}$$

- もっともよく知られた更新公式で様々なソフトウェアでも使用されている
- Wolfe条件のもとで正定値性が保証される
- 適当な仮定の下, 超一次収束性が保証される

その他の更新公式

DFP(Davidon-Fletcher-Powell) 公式:

$$B_{k+1} = \left(I - \frac{y_k s_k^\top}{s_k^\top y_k} \right) B_k \left(I - \frac{s_k y_k^\top}{s_k^\top y_k} \right) + \frac{y_k y_k^\top}{s_k^\top y_k} \quad H_{k+1} = H_k - \frac{H_k y_k y_k^\top H_k}{y_k^\top H_k y_k} + \frac{s_k s_k^\top}{s_k^\top y_k}$$

- 歴史的に最も古い更新公式の一つ
- 効率的にはBFGS公式のほうが有用

SR1(Symmetric Rank one) 公式:

$$B_{k+1} = B_k + \frac{(y_k - B_k s_k)(y_k - B_k s_k)^\top}{(y_k - B_k s_k)^\top s_k} \quad H_{k+1} = H_k + \frac{(s_k - H_k y_k)(s_k - H_k y_k)^\top}{(s_k - H_k y_k)^\top y_k}$$

- 紹介した中では唯一の1ランクの更新公式
- 必ずしも正定値対称行列になるとは限らない

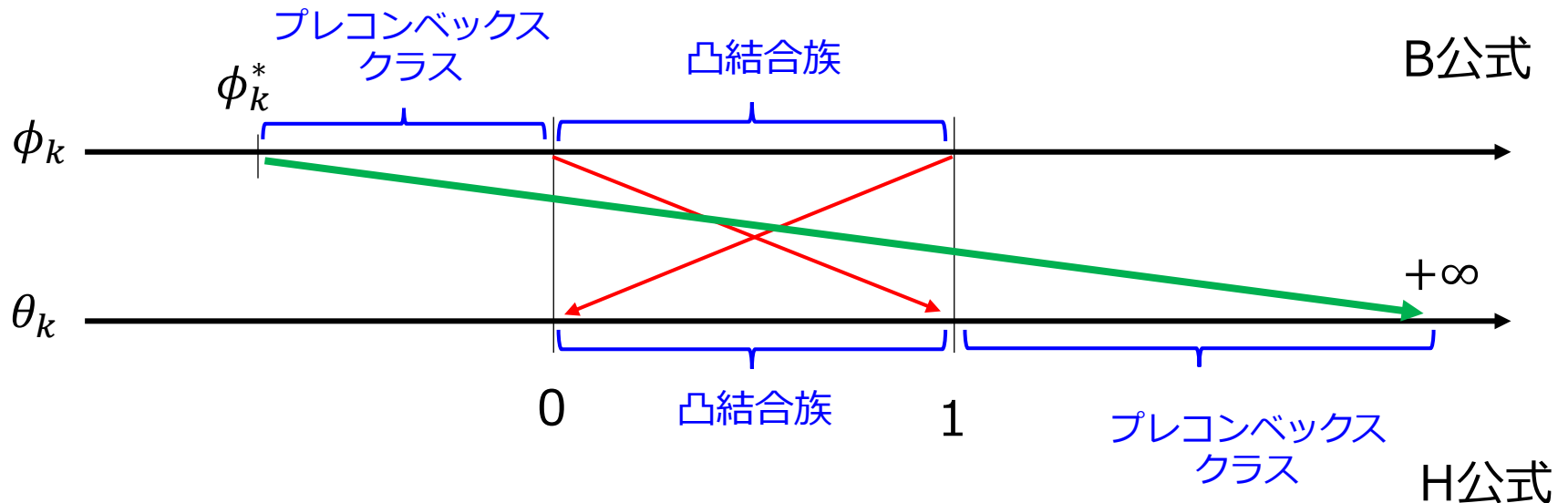
その他の更新公式

Broyden公式族:

$$B_{k+1}^{Broyden} = (1 - \phi_k)B_{k+1}^{BFGS} + \phi_k B_{k+1}^{DFP}$$

$$H_{k+1}^{broyden} = (1 - \theta_k)B_{k+1}^{DFP} + \theta_k B_{k+1}^{BFGS}$$

- BFGS公式とDFP公式の線形結合
- $\phi_k \in [0,1]$ を凸結合族, $\phi_k \in (\phi_k^*, 0)$ をプレコンベックスクラスと呼ぶ.
- $\phi_k = \frac{s_k^\top y_k}{(y_k - B_k s_k)^\top s_k}$ とおくとSR1公式に帰着



数値例

Ex. Rosenblock関数での実験例

	100		10000		100000		1000000	
	反復回数	実行時間	反復回数	実行時間	反復回数	実行時間	反復回数	実行時間
Newton法	305	0.09	305	98.64	memory over		memory over	
BFGS法(H公式)	27	0	35	91.17	memory over		memory over	
CG-HS+法	58	0	26	0.16	67	3.58	36	7.39
加速付き最急降下法	235	0.02	237	0.67	237	2.33	247	14.72
非単調BB法	68	0	68	0	68	0.52	68	3.98

目次

- 第1部 (10:00~11:00)
 - 無制約最適化の基礎
 - 反復法
 - 直線探索
 - 最急勾配法, ニュートン法, 準ニュートン法
- 第2部 (11:15~12:15)
 - 大規模無制約最適化問題に対する反復法
 - 記憶制限準ニュートン法, BB法
 - 非線形共役勾配法
 - メモリーレス準ニュートン法
- 第3部 (13:25~14:25)
 - 微分不可能な正則化項を含む問題への拡張
 - 近接勾配法
 - ニュートン型近接勾配法

大規模無制約最適化問題の解法

いかにして大規模問題に適応させるか？

- ヘッセ行列の疎性を利用する
 - 信頼領域法+ニュートン法
 - スパース準ニュートン法

- 行列を陽には使用しない
 - 記憶制限準ニュートン法
 - Barzilai-Borwein法
 - 非線形共役勾配法
 - メモリーレス準ニュートン法

記憶制限準ニュートン法

逆行列版BFGS公式

$$H_k = V_{k-1}^T H_{k-1} V_{k-1} + \frac{s_{k-1} s_{k-1}^T}{s_{k-1}^T y_{k-1}}, \quad V_{k-1} = I - \frac{y_{k-1} s_{k-1}^T}{s_{k-1}^T y_{k-1}}$$

\hat{H}_0 (通常は単位行列の定数倍) で置き換える

$$\begin{aligned} H_k &= (V_{k-m} V_{k-m+1} \cdots V_{k-2} V_{k-1})^T \hat{H}_{k-m} (V_{k-m} V_{k-m+1} \cdots V_{k-2} V_{k-1}) \\ &+ (V_{k-m+1} V_{k-m+2} \cdots V_{k-2} V_{k-1})^T \frac{s_{k-m} s_{k-m}^T}{s_{k-m}^T y_{k-m}} (V_{k-m+1} V_{k-m+2} \cdots V_{k-2} V_{k-1}) \\ &+ \cdots + (V_{k-2} V_{k-1})^T \frac{s_{k-3} s_{k-3}^T}{s_{k-3}^T y_{k-3}} (V_{k-2} V_{k-1}) + V_{k-1}^T \frac{s_{k-2} s_{k-2}^T}{s_{k-2}^T y_{k-2}} V_{k-1} + \frac{s_{k-1} s_{k-1}^T}{s_{k-1}^T y_{k-1}} \end{aligned}$$

- H_k は $2m$ 本のベクトル $s_{k-1}, \dots, s_{k-m}, y_{k-1}, \dots, y_{k-m}$ のみで計算可能
- $d_k = -H_k g_k$ を求める場合も数 m 回程度の内積計算のみで計算が可能

Barzilai-Borwein(BB)法

注意：以降では, しばらくの間 $g = \nabla f$, $g_k = \nabla f(x_k)$ と表記する

Barzilai-Borwein法

$$x_{k+1} = x_k - \frac{1}{\lambda_k} g_k$$

$$\lambda_k I \approx \nabla^2 f(x_k)$$

$$\lambda_k = \arg \min_{\lambda > 0} \|\lambda I s_{k-1} - y_{k-1}\|$$



$$\lambda_k = \frac{s_{k-1}^T y_{k-1}}{s_{k-1}^T s_{k-1}}$$

準Newton法

更新式

$$x_{k+1} = x_k - B_k^{-1} g_k$$


セカント条件

$$B_k s_{k-1} = y_{k-1}$$

非線形共役勾配法

線形共役勾配法

Solve $Ax = b$ (A : spd)

 Minimize $f(x) = \frac{1}{2}x^T Ax - b^T x$

非線形共役勾配法

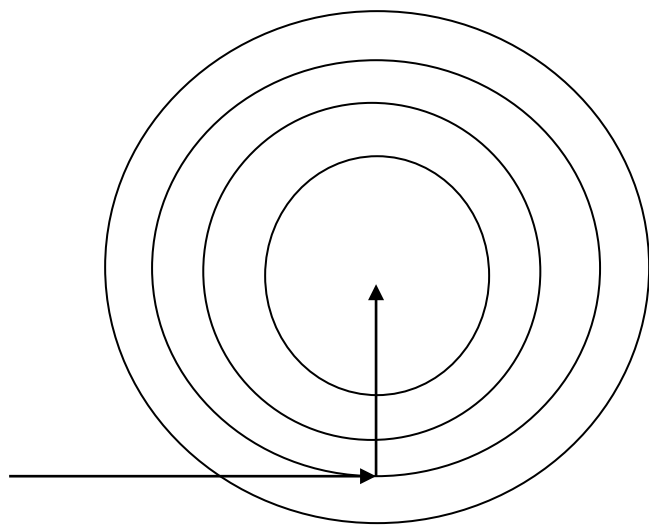
一般の非線形関数最小化問題を解く方法

Minimize $f(x)$

線形共役勾配法 (前準備)

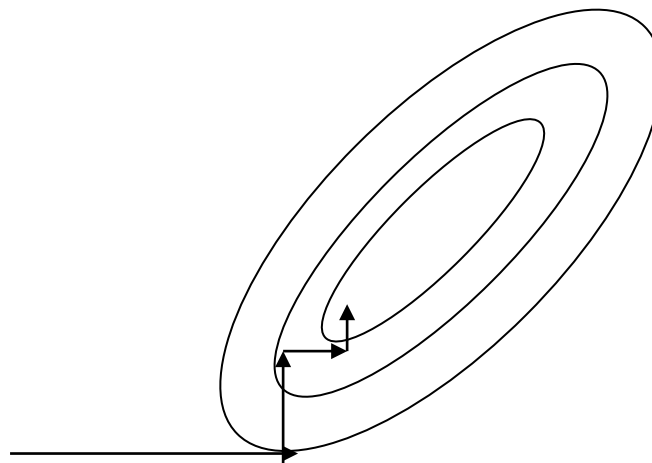
緩和法 (軸探索法)

円の場合 ($A = aI$)



n 回で収束

楕円の場合

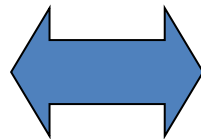


共役方向法（共役性とは？）

$A = P^T P$ と分解して $\hat{x} = Px$ と変数変換を行う

➡ $\hat{x}^T \hat{x} = x^T Ax$ となるので、 \hat{x} の空間では等高線は同心円となる

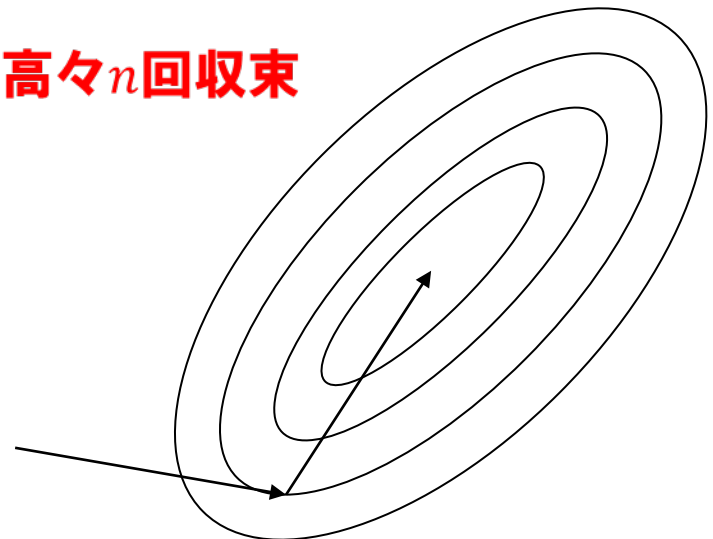
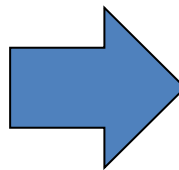
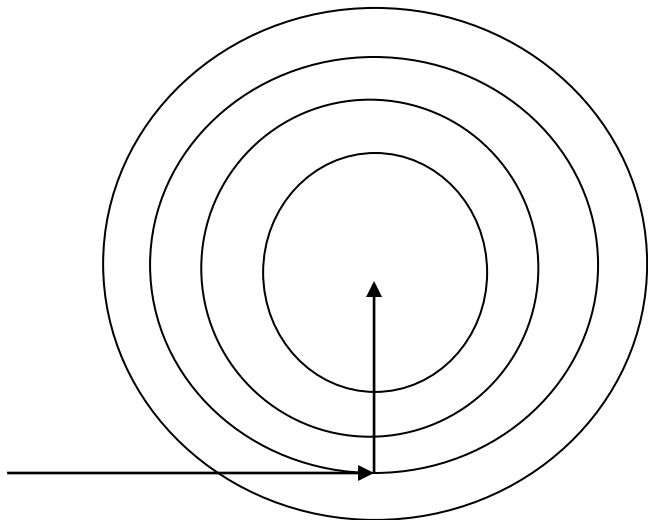
1次変換された空間での直交性
 $(Pu)^T (Pv) = 0$



元の空間
 $u^T Av = 0$

A-直交

高々n回収束



共役方向法の収束性

狭義凸 2 次関数最小化問題に対する反復法を考える。ただし、探索方向 d_0, \dots, d_{n-1} は非零かつ互いに共役であるとし、ステップ幅 α_k は正確な直線探索 (厳密な 1 次元最小化) によって計算されるものとする。このとき、以下が成り立つ。

- (1) k 回目の近似解 x_k はアフィン空間 $x_0 + K_k$ 上で目的関数 f の最小解となる。ただし、 $K_k = \text{Span}\{d_0, d_1, \dots, d_{k-1}\}$ とする。
- (2) 高々 n 回の反復で最小解を得る。

共役勾配法では

$$\nabla f(x_0), \nabla f(x_1), \dots, \nabla f(x_{n-1})$$

を重み付の直交化法によって、逐次、 A -直交する探索方向を生成する。

生成された探索方向は

$$d_i^T A d_j = 0 \quad (i \neq j) \quad (\text{共役性})$$

を満たす

$$* \nabla f(x_k) = A x_k - b = r_k \text{ (残差) に注意}$$

(線形)共役勾配法のアルゴリズム

Step 0 x_0 を与え $k = 0$, $d_0 = -r_0$ として Step 2 へ.

Step 1 $\beta_k = \frac{\|r_k\|^2}{\|r_{k-1}\|^2}$ として $d_k = -r_k + \beta_k d_{k-1}$ とする.

Step 2 $\alpha_k = \frac{r_k^T r_k}{d_k^T A d_k}$ を計算する.

Step 3 $x_{k+1} = x_k + \alpha_k d_k$ とする.

Step 4 収束判定を行う.

Step 5 $k := k + 1$ として Step 1 へ.

$$r_k = Ax_k - b$$

非線形共役勾配法のアルゴリズム

Step 0 x_0 を与え $k = 0$, $d_0 = -g_0$ として Step 2 へ.

Step 1 $\beta_k = \frac{\|g_k\|^2}{\|g_{k-1}\|^2}$ として $d_k = -g_k + \beta_k d_{k-1}$ とする.

Step 2 直線探索によって α_k を求める.

Step 3 $x_{k+1} = x_k + \alpha_k d_k$ とする.

Step 4 収束判定を行う.

Step 5 $k := k + 1$ として Step 1 へ.

$$r_k = Ax_k - b = \nabla f(x_k)$$

非線形共役勾配法(CG法)

探索方向

$$d_k = -g_k + \beta_k d_{k-1}$$

$$y_{k-1} = g_k - g_{k-1}$$

β_k の選び方により様々な共役勾配法が存在する

- Fletcher-Reeves (1964)

$$\beta_k^{FR} = \frac{\|g_k\|^2}{\|g_{k-1}\|^2}$$

- Polak-Ribière (1969)

$$\beta_k^{PR} = \frac{g_k^\top y_{k-1}}{\|g_{k-1}\|^2}$$

- Hestenes-Stiefel (1952)

$$\beta_k^{HS} = \frac{g_k^\top y_{k-1}}{d_{k-1}^\top y_{k-1}}$$

- Dai-Yuan (1999)

$$\beta_k^{DY} = \frac{\|g_k\|^2}{d_{k-1}^\top y_{k-1}}$$

DY法の収束性の概要

$$d_k = -g_k + \beta_k^{DY} d_{k-1}, \quad \beta_k^{DY} = \frac{\|g_k\|^2}{d_{k-1}^\top y_{k-1}}$$

1. $g_k^\top d_k = \beta_k^{DY} g_{k-1}^\top d_{k-1} < 0$ for all $k \geq 1$ を示す

➡ 降下方向を生成することと, $\beta_k^{DY} = \frac{g_k^\top d_k}{g_{k-1}^\top d_{k-1}}$ を得る

2. Zoutendjik条件を用いて $\liminf_{k \rightarrow \infty} \|g_k\| = 0$ を示す.

$\beta_k^{DY} d_{k-1} = g_k + d_k$ より

$$(\beta_k^{DY} \|d_{k-1}\|)^2 = \|g_k\|^2 + 2g_k^\top d_k + \|d_k\|^2$$

➡
$$\frac{\|d_k\|^2}{(g_k^\top d_k)^2} \leq \frac{\|d_{k-1}\|^2}{(g_{k-1}^\top d_{k-1})^2} + \frac{1}{\|g_k\|^2} \leq \frac{\|d_0\|^2}{(g_0^\top d_0)^2} + \sum_{i=1}^k \frac{1}{\|g_i\|^2}$$

DY法の収束性の概要

背理法の仮定として, $\exists \varepsilon > 0, s. t. \|g_k\| \geq \varepsilon$ for all $k \geq 0$ を仮定すると,

$$\frac{\|d_k\|^2}{(g_k^\top d_k)^2} \leq \frac{\|d_0\|^2}{(g_0^\top d_0)^2} + \sum_{i=1}^k \frac{1}{\|g_i\|^2} \leq c_1 + \frac{k}{\varepsilon}$$

となり, Zoutenjdik条件 $\left(\sum_{k=0}^{\infty} \frac{(\nabla f(x_k)^\top d_k)^2}{\|d_k\|^2} < \infty \right)$ に反する \square

HS+法の証明の概要

Property*

正の定数 γ_1, γ_2 が存在して, すべての k に対して $\gamma_1 \leq \|\nabla f(x_k)\| \leq \gamma_2$ が成立していると仮定する. このとき, すべての k に対し,

$$|\beta_k| \leq b, \quad \text{and} \quad \|s_{k-1}\| \leq \eta \rightarrow |\beta_k| \leq \frac{1}{b}$$

を満たす定数 $b > 1$ と $\eta > 0$ が存在するとき, 共役勾配法はProperty*を満たすという.

定理

Zoutendijk条件と同様の仮定が成り立っているとする. このとき, 以下の条件を満たすCG法は $\liminf_{k \rightarrow \infty} \|g_k\| = 0$ の意味で大域的に収束する.

1. $\beta_k \geq 0$ for all $k \geq 1$
2. 十分な降下条件を満たす
3. Property* を満たす

HS+法の証明の概要

$$\beta_k^{HS+} = \max \left\{ 0, \frac{g_k^\top y_{k-1}}{d_{k-1}^\top y_{k-1}} \right\}$$

1. 十分な降下条件は仮定とする（必ずしも満たされるとはかぎらない）。
2. Property* を満たすことを示す。

Property*を満たすことを確認するには、

$$\exists c > 0, \quad \text{s.t. } |\beta_k| \leq c \|s_{k-1}\|$$

を確認すれば十分。

非線形共役勾配法の発展

1. 曲率の情報を取り込むためにセカント条件を利用した方法
 - Dai and Liao (2001)、Yabe and Takano (2003)、Zhou and Zhang (2006)、Ford, Narushima and Yabe (2007) など
2. 直線探索に依存せずに降下方向を生成する方法
 - パラメータ β の修正
 - Hager and Zhang (2005)、Yu, Guan and Chen (2008)、Yu, Guan and Li(2008)、Dai and Kou(2013)
 - スケーリング/3項共役勾配法
 - Zhang, Zhou and Li (2006, 2006, 2007) 、 Cheng (2006) 、
Narushima, Yabe and Ford (2011) 、 Al-Baali, Narushima and Yabe(2015)
3. その他
 - ハイブリッド法
 - Chen and Liu(2010)、Narushima and Yabe(2012)、Sugiki, Narushima and Yabe(2012)
 - 部分空間最小化法との組み合わせ
 - Hager and Zhang(2013)
 - (通常の意味での)無制約最適化問題以外への拡張
 - 多様体上の最適化問題への拡張(Sato and Iwai2015, Sato 2016, Sakai and Iiduka 2020 など)
 - 不動点制約付き問題への拡張(Iiduka2008など)

3項共役勾配法(N, Yabe and Ford 2011)

$$d_k = \begin{cases} -g_k & k = 0 \\ -g_k + \beta_k (g_k^T p_k)^\dagger \{ (g_k^T p_k) d_{k-1} - (g_k^T d_{k-1}) p_k \} & k \geq 1 \end{cases}$$

$$\beta_k \in \mathbf{R} : \text{パラメータ}, \quad a^\dagger = \begin{cases} \frac{1}{a} & a \neq 0, \\ a & a = 0 \end{cases}$$

$p_k \in \mathbf{R}^n : \text{任意ベクトル}$

- $g_k^T p_k = 0$ のとき

$$g_k^T d_k = -\|g\|^2 < 0$$

- $g_k^T p_k \neq 0$ のとき

$$g_k^T d_k = -\|g_k\|^2 + \beta_k g_k^T d_{k-1} - \beta_k \frac{g_k^T d_{k-1}}{g_k^T p_k} g_k^T p_k = -\|g_k\|^2 < 0$$

- 直線探索や β_k に関係なく常に(十分な)降下方向を生成
- 正確な直線探索が使用された場合には通常のCGに帰着

3項共役勾配法

- $p_k = g_k$ の場合

$$d_k = - \left(1 + \beta_k \frac{g_k^T d_{k-1}}{\|g_k\|^2} \right) g_k + \beta_k d_{k-1} \quad \rightarrow \text{スケーリングCG法}$$

- $g_k^T p_k \neq 0$ の場合

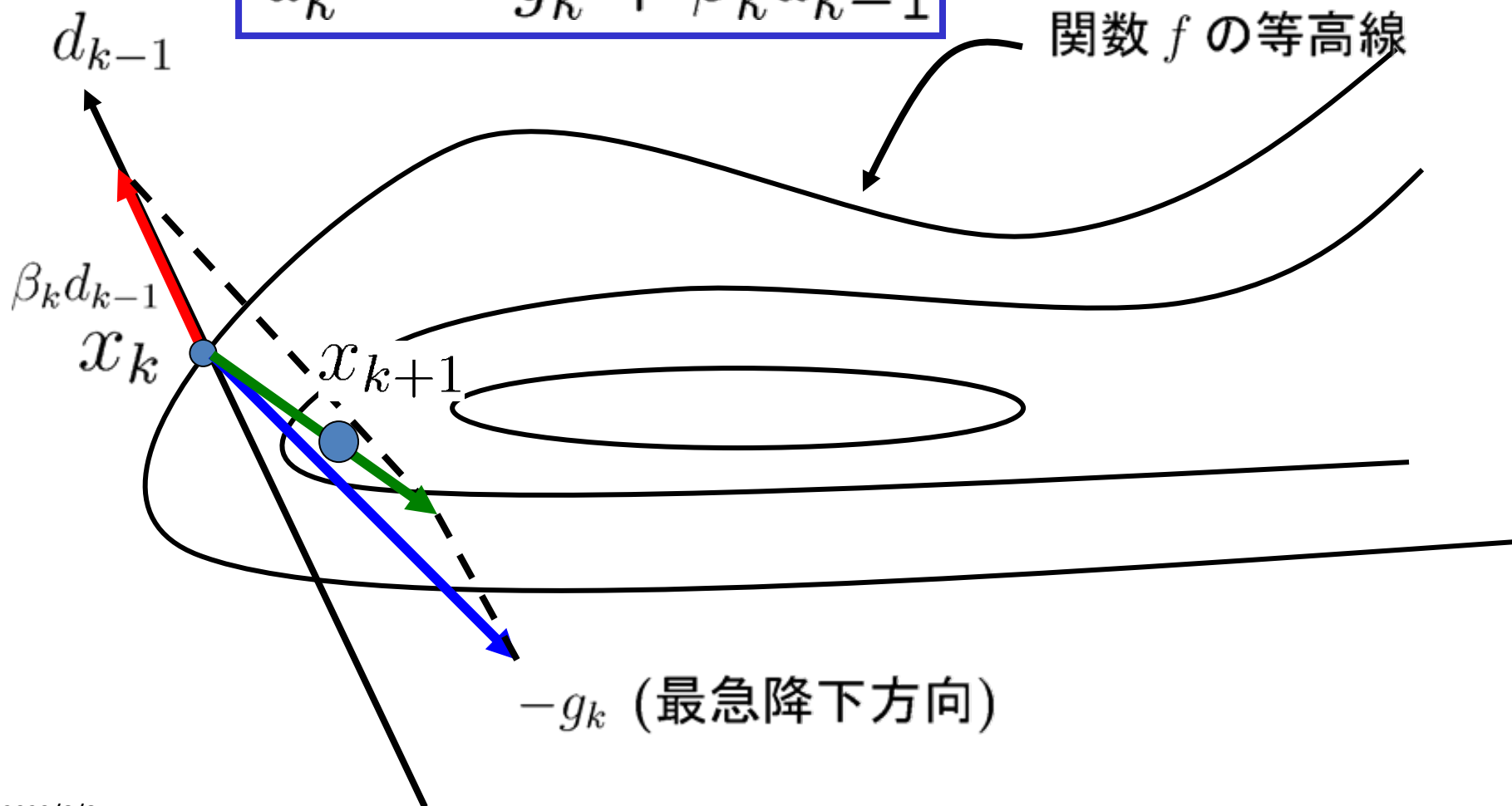
$$\begin{aligned} d_k &= -g_k + \beta_k d_{k-1} - \beta_k \frac{g_k^T d_{k-1}}{g_k^T p_k} p_k \\ &= -g_k + \beta_k \left(I - \frac{p_k g_k^T}{g_k^T p_k} \right) d_{k-1} \end{aligned}$$



Span $\{p_k\}$ に沿った Span $\{g_k\}$ の直交補空間への射影行列

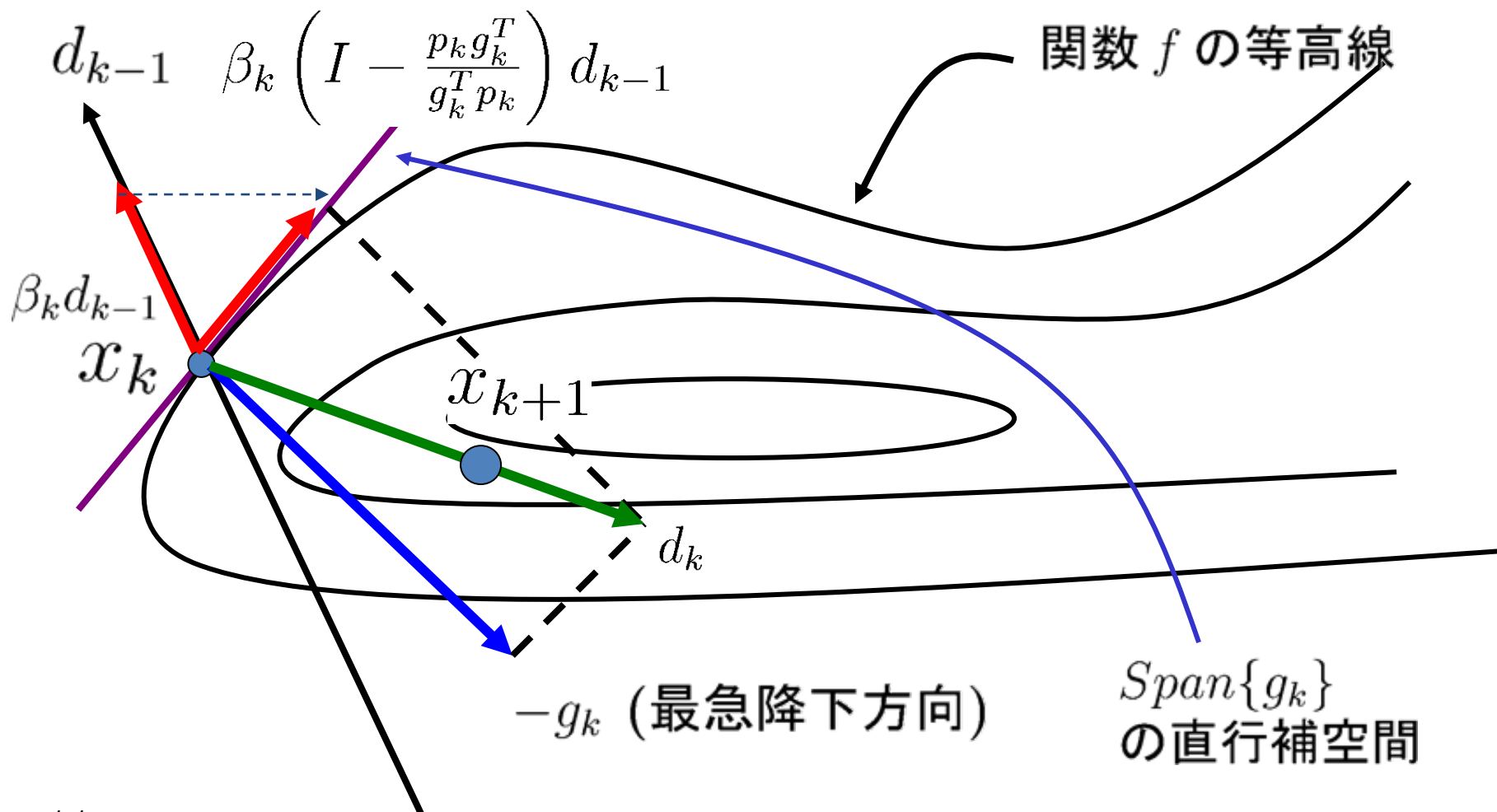
非線形共役勾配法の探索方向

$$d_k = -g_k + \beta_k d_{k-1}$$



非線形共役勾配法の探索方向

- $$d_k = -g_k + \beta_k \left(I - \frac{p_k g_k^T}{g_k^T p_k} \right) d_{k-1}$$



大規模無制約最適化問題の解法

いかにして大規模問題に適応させるか？

- ヘッセ行列の疎性を利用する
 - 信頼領域法+ニュートン法
 - スパース準ニュートン法

- 行列を陽には使用しない
 - 記憶制限準ニュートン法
 - Barzilai-Borwein法
 - 非線形共役勾配法
 - メモリーレス準ニュートン法

記憶制限準ニュートン法 (再掲)

逆行列版BFGS公式

$$H_k = V_{k-1}^T H_{k-1} V_{k-1} + \frac{s_{k-1} s_{k-1}^T}{s_{k-1}^T y_{k-1}}, \quad V_{k-1} = I - \frac{y_{k-1} s_{k-1}^T}{s_{k-1}^T y_{k-1}}$$

\hat{H}_0 (通常は単位行列の定数倍) で置き換える

$$\begin{aligned} H_k &= (V_{k-m} V_{k-m+1} \cdots V_{k-2} V_{k-1})^T \hat{H}_{k-m} (V_{k-m} V_{k-m+1} \cdots V_{k-2} V_{k-1}) \\ &\quad + (V_{k-m+1} V_{k-m+2} \cdots V_{k-2} V_{k-1})^T \frac{s_{k-m} s_{k-m}^T}{s_{k-m}^T y_{k-m}} (V_{k-m+1} V_{k-m+2} \cdots V_{k-2} V_{k-1}) \\ &\quad + \cdots + (V_{k-2} V_{k-1})^T \frac{s_{k-3} s_{k-3}^T}{s_{k-3}^T y_{k-3}} (V_{k-2} V_{k-1}) + V_{k-1}^T \frac{s_{k-2} s_{k-2}^T}{s_{k-2}^T y_{k-2}} V_{k-1} + \frac{s_{k-1} s_{k-1}^T}{s_{k-1}^T y_{k-1}} \end{aligned}$$

- H_k は $2m$ 本のベクトル $s_{k-1}, \dots, s_{k-m}, y_{k-1}, \dots, y_{k-m}$ のみで計算可能
- $d_k = -H_k g_k$ を求める場合も数 m 回程度の内積計算のみで計算が可能

メモリーレス準ニュートン法

準ニュートン法(逆行列版)

$$d_k = -H_k \nabla f(x_k)$$

BFGS公式

$$H_k = V_{k-1}^T H_{k-1} V_{k-1} + \frac{s_{k-1} s_{k-1}^T}{s_{k-1}^T y_{k-1}}, \quad V_{k-1} = I - \frac{y_{k-1} s_{k-1}^T}{s_{k-1}^T y_{k-1}}$$

H_{k-1} を $H_{k-1} = \gamma_{k-1} I$ で置き換える

$$d_k = -H_k g_k = -\gamma_{k-1} V_{k-1}^T V_{k-1} g_k - \frac{s_{k-1} s_{k-1}^T}{s_{k-1}^T s_{k-1}} g_k \quad \leftarrow \text{記憶制限準ニュートン法で}$$

記憶数 $m = 1$ に相当する

$$= -\gamma_{k-1} g_k + \left(\gamma_{k-1} \frac{g_k^T y_{k-1}}{d_{k-1}^T y_{k-1}} - \left(1 + \gamma_{k-1} \frac{\|y_{k-1}\|^2}{d_{k-1}^T y_{k-1}} \right) (g_k^T d_{k-1}) \right) d_{k-1} + \frac{\gamma_{k-1} g_k^T d_{k-1}}{d_{k-1}^T y_{k-1}} y_{k-1}$$

メモリーレス準ニュートン法

$$\begin{aligned}d_k &= -H_k g_k = -\gamma_{k-1} V_{k-1}^\top V_{k-1} g_k - \frac{s_{k-1} s_{k-1}^\top}{s_{k-1}^\top s_{k-1}} g_k \\&= -\gamma_{k-1} g_k + \left(\gamma_{k-1} \frac{g_k^\top y_{k-1}}{d_{k-1}^\top y_{k-1}} - \left(1 + \gamma_{k-1} \frac{\|y_{k-1}\|^2}{d_{k-1}^\top y_{k-1}} \right) (g_k^\top d_{k-1}) \right) d_{k-1} + \frac{\gamma_{k-1} g_k^\top d_{k-1}}{d_{k-1}^\top y_{k-1}} y_{k-1} \\&= -\gamma_{k-1} g_k + \beta_k d_{k-1} + \delta_k y_{k-1} \quad \leftarrow \text{ある種の3項共役勾配法としてとらえられる}\end{aligned}$$

特に、正確な直線探索を用いた場合には $g_k^\top d_{k-1} = 0$ より

$$d_k = \gamma_{k-1} \left(-g_k + \frac{g_k^\top y_{k-1}}{d_{k-1}^\top y_{k-1}} d_{k-1} \right) = \gamma_{k-1} \left(-g_k + \beta_k^{HS} d_{k-1} \right)$$

となるため、非線形共役勾配法(HS公式)に一致する。

メモリーレス準ニュートン法と三項共役勾配法の比較

メモリーレス準ニュートン法(BFGS公式の例)

$$d_k = - \left(\left(I - \frac{s_k y_k^\top}{s_k^\top y_k} \right) \left(I - \frac{y_k s_k^\top}{s_k^\top y_k} \right) + \frac{s_k s_k^\top}{s_k^\top y_k} \right) g_k$$
$$= H_k$$

- $s_k^\top y_k > 0$ ならば正定値対称
- セカント条件を満たしている

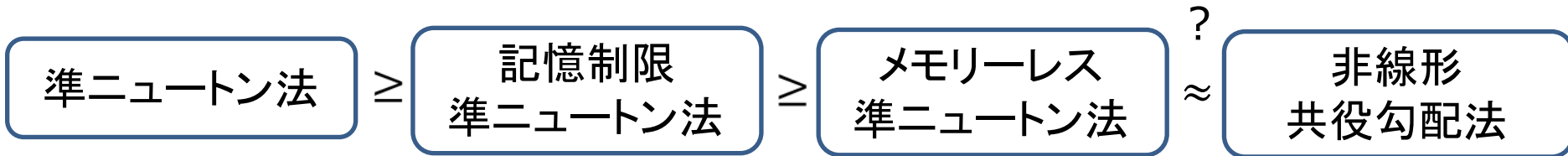
三項共役勾配法(HS公式, $p_k = y_{k-1}$ の場合)

$$d_k = -g_k + \frac{g_k^\top y_{k-1}}{d_{k-1}^\top y_{k-1}} d_{k-1} - \frac{g_k^\top y_{k-1}}{d_{k-1}^\top y_{k-1}} \frac{g_k^\top d_{k-1}}{g_k^\top y_{k-1}} y_{k-1} = - \left(I + \frac{d_{k-1} y_{k-1}^\top}{d_{k-1}^\top y_{k-1}} - \frac{y_{k-1} d_{k-1}^\top}{d_{k-1}^\top y_{k-1}} \right) g_k$$
$$= H_k$$

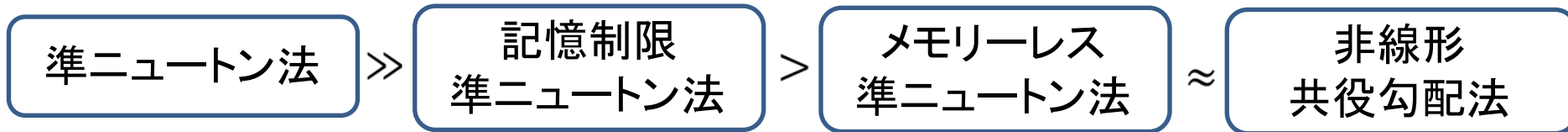
- 対称性, 正定値性, セカント条件などは考慮していない

メモリーレス準ニュートン法の特徴

情報の量



探索方向の計算に必要な計算量



- 記憶制限準ニュートン法と比較すると
 - BFGS公式以外でも簡単に探索方向を考える事ができる
- 非線形共役勾配法や三項共役勾配法と比較すると
 - 対称性を持つので正定値になるように選べば計量として使用可能
 - 無制約以外の問題に拡張しやすい

メモリーレス準ニュートン法の研究

- 最初のメモリーレス準ニュートン法
 - Shanno(1978)
- BFGS公式に基づくメモリーレス準ニュートン法
 - Kou and Dai (2015)
- SR1公式に基づくメモリーレス準ニュートン法
 - Moyi and Leong (2016), Nakayama et al. (2018)
- ブロイデン公式族に基づくメモリーレス準ニュートン法
 - Nakayama N. and Yabe (2019), Nakayama (2018)
- その他
 - 上下制限約付き問題への拡張（有効制約法） Nakayama et al. (2021)
 - 多様体上の最適化問題に対する拡張 N. et al. (2023)
 - 微分不可能な関数を含む最適化問題に対する拡張 Becker et al. (2019), Nakayama et al.(2021), N. and Nakayama (2023) など

Nakayama N. and Yabe (2019)の概要

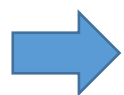
通常の (BFGS公式に基づいた) メモリーレス準ニュートン法

$$d_k = -\gamma_{k-1}g_k + \left(\gamma_{k-1} \frac{g_k^\top y_{k-1}}{d_{k-1}^\top y_{k-1}} - \left(1 + \gamma_{k-1} \frac{\|y_{k-1}\|^2}{d_{k-1}^\top y_{k-1}} \right) (g_k^\top d_{k-1}) \right) d_{k-1} + \frac{\gamma_{k-1} g_k^\top d_{k-1}}{d_{k-1}^\top y_{k-1}} y_{k-1}$$

γ_{k-1} として $\gamma_{k-1}I \approx \nabla f(x_{k-1})^{-1}$ が望ましいため, BB法の時と同様に

$$\gamma_{k-1} = \frac{s_{k-1}^\top y_{k-1}}{y_{k-1}^\top y_{k-1}} \quad \gamma_{k-1} = \frac{s_{k-1}^\top s_{k-1}}{s_{k-1}^\top y_{k-1}}$$

が一般的だが, 大域的収束性を保証するためには $\bar{\gamma} < \gamma_{k-1}$ ($\bar{\gamma} > 0$) が必要



セカント条件自体を修正

Nakayama N. and Yabe (2019)の概要

2次近似

$$\gamma_{k-1}f(x) \approx \gamma_{k-1} \left(f(x_k) + g_k^T(x - x_k) + \frac{1}{2}(x - x_k)^T \nabla^2 f(x_k)(x - x_k) \right)$$

微分して

$$x = x_{k-1}$$



$$\underbrace{\gamma_{k-1} \nabla^2 f(x_k) s_{k-1}}_{\approx B_k} \approx \gamma_{k-1} y_{k-1}$$

スペクトラルスケージング(SS)

セカント条件(Cheng and Li (2010))

$$B_k s_{k-1} = \gamma_{k-1} y_{k-1} \quad \text{または} \quad s_{k-1} = \gamma_{k-1} H_k y_{k-1}$$

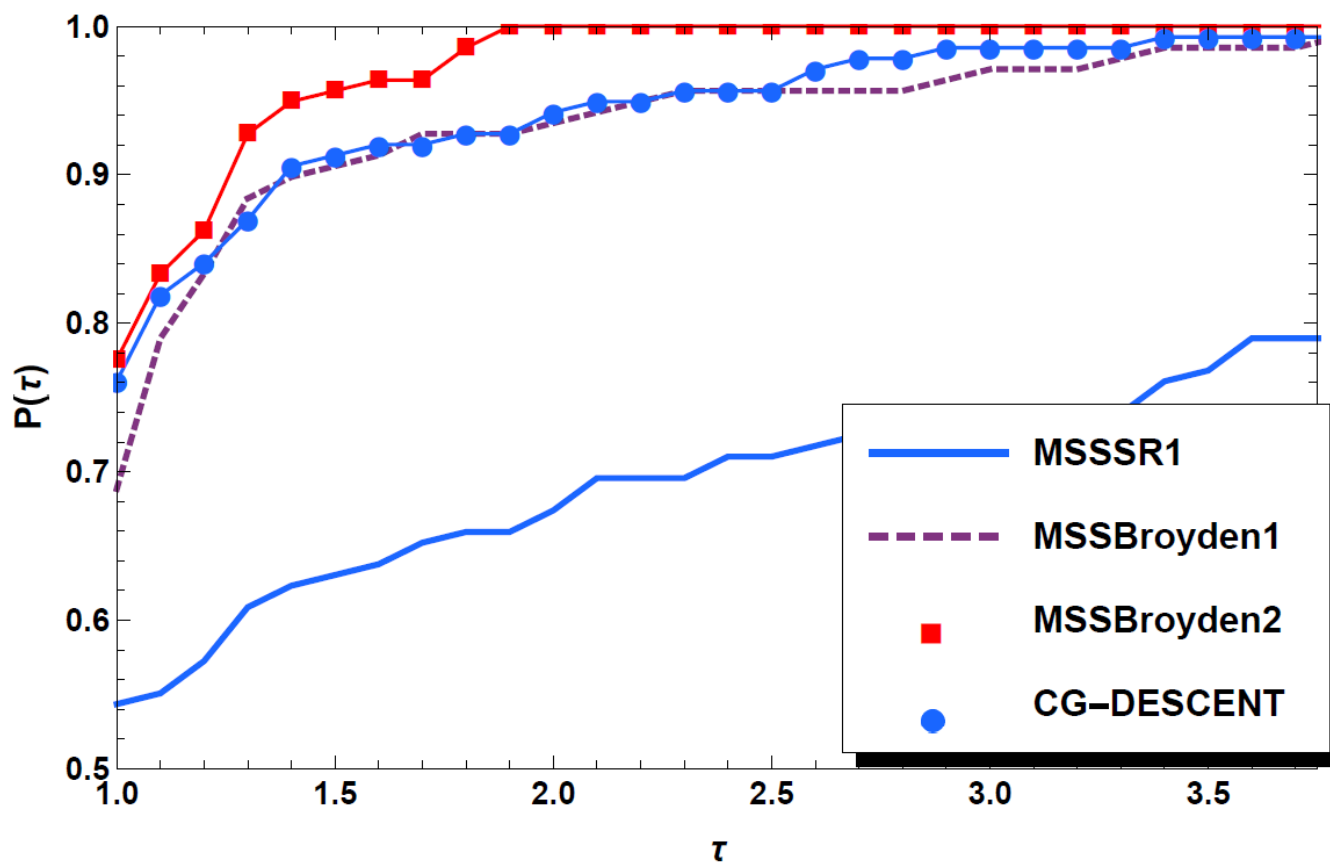
SSセカント条件に基づくメモリーレスBFGS法(実際にはBroyden族を提案)

$$d_k = -g_k + \left(\frac{g_k^T y_{k-1}}{d_{k-1}^T y_{k-1}} - \left(\frac{1}{\gamma_{k-1}} + \frac{\|y_{k-1}\|^2}{d_{k-1}^T y_{k-1}} \right) (g_k^T d_{k-1}) \right) d_{k-1} + \frac{g_k^T d_{k-1}}{d_{k-1}^T y_{k-1}} y_{k-1}$$

メモリーレス準ニュートン法の数値実験例

- CUTER問題集から138問を選択して実験
- 実験の詳細は以下を参照

「無制約最適化問題に対するメモリーレス準ニュートン法について」, 応用数理, 29 (2019), 8-17



修正メモリーレス準ニュートン法

修正メモリーレス準ニュートン法(修正mless-BFGS公式)

$$B_k = I - \frac{S_{k-1}S_{k-1}^\top}{S_{k-1}^\top S_{k-1}} + \frac{Z_{k-1}Z_{k-1}^\top}{S_{k-1}^\top Z_{k-1}}$$

$$H_k = \left(I - \frac{S_{k-1}Z_{k-1}^\top}{S_{k-1}^\top Z_{k-1}} \right) \left(I - \frac{Z_{k-1}S_{k-1}^\top}{S_{k-1}^\top Z_{k-1}} \right) + \frac{S_{k-1}S_{k-1}^\top}{S_{k-1}^\top Z_{k-1}}$$

ただし, Z_{k-1} は

- $\|Z_{k-1}\| \leq c_1 \|S_{k-1}\|$
- $S_{k-1}^\top Z_{k-1} \geq c_2 \|S_{k-1}\|^2$

を満たすベクトル

上記の B_k の固有値は 1 (重複度 $n-2$) と

$$\lambda_k^+ = \frac{1}{2} \left\{ \bar{\lambda}_k + \sqrt{\bar{\lambda}_k^2 - 4 \frac{S_{k-1}^\top Z_{k-1}}{S_{k-1}^\top S_{k-1}}} \right\}, \quad \lambda_k^- = \frac{1}{2} \left\{ \bar{\lambda}_k - \sqrt{\bar{\lambda}_k^2 - 4 \frac{S_{k-1}^\top Z_{k-1}}{S_{k-1}^\top S_{k-1}}} \right\},$$

である. ただし, $\bar{\lambda}_k = 1 + \frac{Z_{k-1}^\top Z_{k-1}}{S_{k-1}^\top Z_{k-1}}$ とする.

修正メモリーレス準ニュートン法

補題 正の定数 m, M が存在して, すべての $k \geq 1$ に対して,

$$m \leq \lambda_k^- \leq \lambda_k^+ \leq M$$

が成立する.

$$m\|v\|^2 \leq v^T B_k v \leq M\|v\|^2 \text{ for } \forall v \in R^n \text{ を意味する}$$

定理 ズーテンダイク条件と同様の仮定が成り立っているとする. このとき, 直線探索でWolfe条件を用いた修正mless-BFGS法は $\liminf_{k \rightarrow \infty} \|\nabla f(x_k)\| = 0$ の意味で大域的に収束する.

目次

- 第1部 (10:00~11:00)
 - 無制約最適化の基礎
 - 反復法
 - 直線探索
 - 最急勾配法, ニュートン法, 準ニュートン法
- 第2部 (11:15~12:15)
 - 大規模無制約最適化問題に対する反復法
 - 記憶制限準ニュートン法, BB法
 - 非線形共役勾配法
 - メモリーレス準ニュートン法
- 第3部 (13:25~14:25)
 - 微分不可能な正則化項を含む問題への拡張
 - 近接勾配法
 - ニュートン型近接勾配法

正則化項付き最適化問題

$$\min_x f(x) = g(x) + h(x)$$

$g: \mathbb{R}^n \rightarrow \mathbb{R}$, L -平滑な関数

$h: \mathbb{R}^n \rightarrow \mathbb{R} \cup \{+\infty\}$, 真凸関数だが微分可能とは限らない

* 前半と g の定義が異なることに注意

LASSO (least absolute shrinkage and selection operator)

$$\min \|Ax - b\|^2 + \lambda \|x\|_1$$

最小2乗問題 + ℓ_1 正則化

凸最適化問題(指示関数)

min $g(x)$
Subject to $x \in S$ (S は凸集合)



min $g(x) + \delta_S(x)$
ただし,

$$\delta_S(x) = \begin{cases} 0 & x \in S \\ +\infty & x \notin S \end{cases}$$

LASSOのイメージ

最小化 $f(\mathbf{x}) + C\|\mathbf{x}\|_1 = f(\mathbf{x}) + C \sum_{k=1}^m |x_k|$

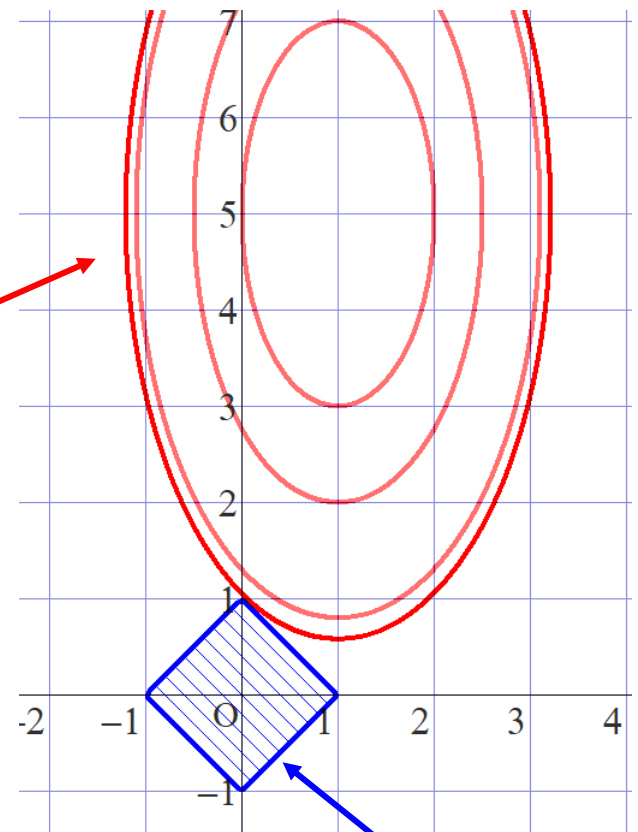


最小化 $f(\mathbf{x})$
制約条件 $\sum_{k=1}^m |x_k| \leq \varepsilon$

目的関数の等高線

左図の例

最小化 $(x-1)^2 + \left(\frac{y-5}{2}\right)^2$
制約条件 $|x| + |y| \leq 1$



実行可能領域

正則化項付き最適化問題

$$\min_x f(x) = g(x) + h(x)$$

$g: R^n \rightarrow R$, L -平滑な関数

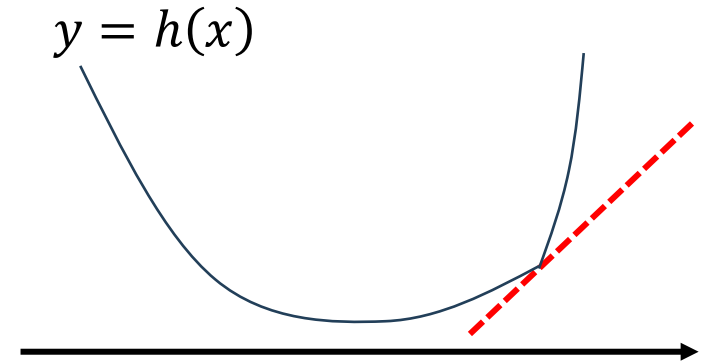
$h: R^n \rightarrow R \cup \{+\infty\}$, 真凸関数だが微分可能とは限らない

1次の最適性条件

$$0 \in \nabla g(x) + \partial h(x)$$

ただし, $\partial h(x)$ は凸関数 h の劣微分とする.

$$\partial h(x) = \{\xi \mid h(y) \geq h(x) + \xi^\top (y - x) \text{ for any } y \in R^n\}$$



1次の最適性条件を満たす点を以下では停留点と呼ぶ.

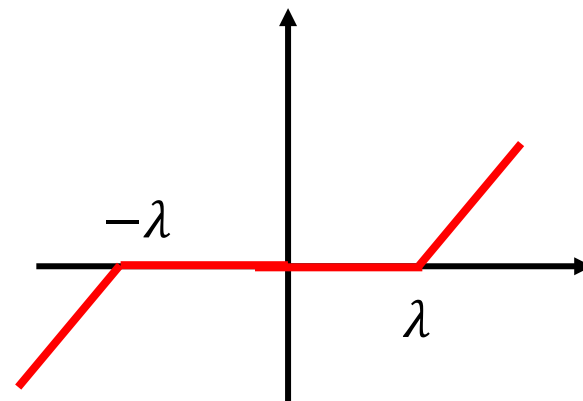
近接写像

近接写像

$$\text{Prox}_h(z) = \operatorname{argmin}_x \left\{ h(x) + \frac{1}{2} \|x - z\|^2 \right\}$$

- ℓ_1 正則化 ($h(x) = \lambda \|x\|_1, \lambda > 0$) の場合

$$(\text{Prox}_h(z))_i = \begin{cases} z_i - \lambda & z_i \geq \lambda \\ 0 & -\lambda \leq z_i \leq \lambda \\ z_i + \lambda & z_i \leq -\lambda \end{cases}$$



- 指示関数 ($h(x) = \delta_S(x)$, S は凸集合) の場合

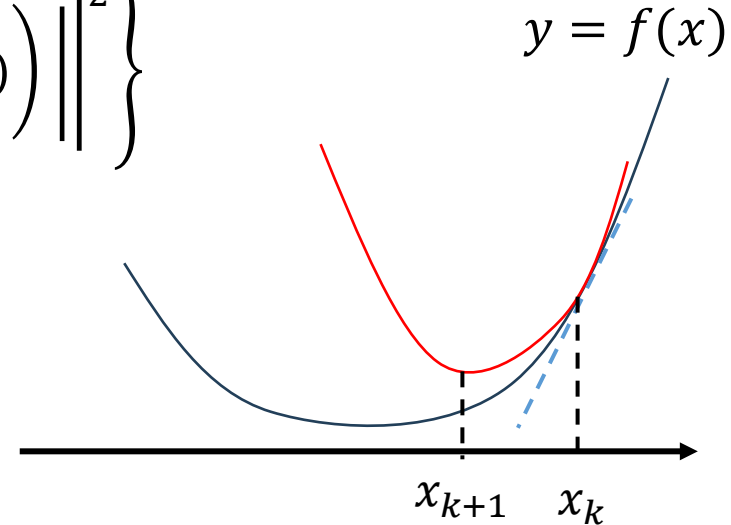
$$\text{Prox}_h(z) = \operatorname{argmin}_{x \in S} \|x - z\| \quad (\text{集合} S \text{への正射影})$$

近接勾配法

$$x_{k+1} = \operatorname{argmin}_x \left\{ \underbrace{g(x_k) + \nabla g(x_k)^\top (x - x_k)}_{g \text{ の1次近似}} + h(x) + \underbrace{\frac{L}{2} \|x - x_k\|^2}_{\text{近接項}} \right\}$$

$$= \operatorname{argmin}_x \left\{ \frac{1}{L} h(x) + \frac{1}{2} \left\| x - \left(x_k - \frac{1}{L} \nabla g(x_k) \right) \right\|^2 \right\}$$

$$= \operatorname{Prox}_{\frac{1}{L}h} \left(\underbrace{x_k - \frac{1}{L} \nabla g(x_k)}_{\text{最急降下法}} \right)$$



- g が強凸の場合の収束率は線形収束, 凸の場合の収束率は $O\left(\frac{1}{k}\right)$
- h が指示関数の場合は射影勾配法に帰着する
- 近接勾配法はISTA (Iterative Shrinkage Thresholding Algorithm) や Forward-backward splitting algorithmなどとも呼ばれる

加速付き近接勾配法(FISTA)

1. 初期点 $x_0 \in R^n$ を与える. $k = 0, z_0 = x_0, \tau_0 = 0$, として, ステップ2へ.
2. 終了判定を行う.
3. $x_{k+1} = \text{Prox}_h \left(z_k - \frac{1}{L} \nabla g(z_k) \right)$
4. $\tau_{k+1} = \frac{1}{2} \left(1 + \sqrt{1 + 4\tau_k^2} \right)$
5. $z_{k+1} = x_{k+1} + \frac{\tau_k - 1}{\tau_{k+1}} (x_{k+1} - x_k)$
6. $k \leftarrow k + 1$ としてステップ2. に戻る.

- g が L -平滑な凸関数のとき

$$f(x_k) - f(x^*) = O\left(\frac{1}{k^2}\right)$$

ニュートン型近接勾配法

$$x_k^+ = \operatorname{argmin}_x \left\{ g(x_k) + \nabla g(x_k)^\top (x - x_k) + h(x) + \frac{1}{2} \underbrace{\|x - x_k\|_{B_k}^2}_{\text{重み付き近接項}} \right\}$$

正定値対称行列に対して

$$\|x\|_A = \sqrt{x^\top A x}$$

$$= \operatorname{argmin}_x \left\{ \underbrace{g(x_k) + \nabla g(x_k)^\top (x - x_k) + \frac{1}{2} (x - x_k)^\top B_k (x - x_k)}_{g \text{ の2次近似}} + h(x) \right\}$$

$$= \operatorname{argmin}_x \left\{ h(x) + \frac{1}{2} \left\| x - \left(x_k - B_k^{-1} \nabla g(x_k) \right) \right\|_{B_k}^2 \right\}$$

$$H_k = B_k^{-1}, \quad \operatorname{Prox}_h^B(z) = \operatorname{argmin}_x \left\{ h(x) + \frac{1}{2} \|x - z\|_B^2 \right\}$$

$$= \operatorname{Prox}_h^B \left(\underbrace{x_k - H_k \nabla g(x_k)}_{\text{(準)ニュートン法}} \right)$$

ニュートン型近接勾配法

- Lee, Sun and Saunders (2014)
 - 最初のニュートン型近接勾配法
- Scheinberg and Tang (2016)
 - 非厳密ニュートン型近接勾配法
- Byrd, Nocedal and Oztoprak (2017)
 - L1正則化付き問題に対する非厳密ニュートン型近接勾配法
- Andersen and Vandenberghe (2017)
 - セルフコンコールド関数に対する非厳密ニュートン型近接勾配法
- Becker, Fadili and Ochs (2019)
 - 特殊な場合での重み付き近接写像の計算法 + それを用いた近接SR1法
- Nakayama, N., and Yabe (2021)
 - メモリーレスプロイデン族に基づく非厳密ニュートン型近接勾配法
- Liu and Takeda(2022)
 - DC関数に対する非厳密ニュートン型近接勾配法
- Nakayama and N. (2023)
 - SR1公式に基づくニュートン型近接勾配法
- Nakayama, N., and Yabe, submitted
 - Nakayama et al. (2021)のDC関数への拡張と部分問題の改良

非厳密ニュートン型近接勾配法(Nakayama et al. 2021)

通常のニュートン型近接勾配法の探索方向と問題点

$$d_k = x_k^+ - x_k$$

$$x_k^+ = \text{Prox}_h^{B_k}(x_k - H_k \nabla g(x_k))$$

- 計量であるため正定値対称
- 密行列の場合は大規模問題では使用できない

$$= \underset{x}{\operatorname{argmin}} \left\{ \nabla g(x_k)(x - x_k) + \frac{1}{2}(x - x_k)^\top B_k(x - x_k) + h(x) \right\}$$

重み付き近接写像の計算に元の問題と同じ次元・構造の2次最適化問題を解く必要がある。

- 重み付き近接写像を近似的に解く
- 近似行列 B_k, H_k をメモリーレス準ニュートン法で構築する

非厳密ニュートン型近接勾配法(Nakayama et al. 2021)

$$\begin{aligned}x_k^+ &= \text{Prox}_h^{B_k}(x_k - H_k \nabla g(x_k)) \\ &= \underset{x}{\text{argmin}} \left\{ \nabla g(x_k)(x - x_k) + \frac{1}{2} (x - x_k)^\top B_k (x - x_k) + h(x) \right\}\end{aligned}$$

部分問題の最適性条件

$$0 \in \nabla g(x) + B_k(x - x_k) + \partial h(x)$$

残差の定義 $r_k \in \nabla g(x_k) + B_k(x_k^+ - x_k) + \partial h(x_k^+)$

非厳密の基準 $\|r_k\|_{H_k} \leq (1 - \theta_k) \|x_k^+ - x_k\|_{B_k} \quad (\theta_k \in [\bar{\theta}, 1], \bar{\theta} \in (0, 1])$

$\theta_k = 1 \rightarrow$ 厳密に解く

$\theta_k = \bar{\theta} \rightarrow$ 緩く解く

近似行列の選択

Nakayama et al. (2019)のメモリーレスブロイデン族を修正

元々のSSセカント条件

$$\gamma_{k-1} \nabla^2 f(x_k) s_{k-1} \approx \gamma_{k-1} y_{k-1} \quad \longrightarrow \quad B_k s_{k-1} = \gamma_{k-1} y_{k-1}$$

修正SSセカント条件

$$\gamma_{k-1} (\nabla^2 f(x_k) + \nu_k I) s_{k-1} \approx \gamma_{k-1} (y_{k-1} + \nu_k s_{k-1})$$

$$\longrightarrow \quad B_k s_{k-1} = \gamma_{k-1} z_{k-1}$$

修正SSセカント条件に基づくブロイデン族

$$B_k = I - \frac{s_{k-1} s_{k-1}^T}{s_{k-1}^T s_{k-1}} + \gamma_k \frac{z_{k-1} z_{k-1}^T}{s_{k-1}^T z_{k-1}} + \phi_k \nu_{k-1} \nu_{k-1}^T, \quad \nu_{k-1} = \sqrt{s_{k-1}^T s_{k-1}} \begin{pmatrix} \frac{z_{k-1}}{s_{k-1}^T z_{k-1}} - \frac{s_{k-1}}{s_{k-1}^T s_{k-1}} \end{pmatrix}$$

パラメータを適切に選択すれば, $0 < m \leq M < \infty$ が存在して,

$$m \|v\|^2 \leq v^T B_k v \leq M \|v\|^2 \quad \text{for } \forall v \in R^n$$

が成立する. (以下はこれを満たす行列を使用していると仮定する)

アルゴリズム

1. 初期設定 : $x_0 \in R^n$ とし, 定数 $\delta, \beta \in (0,1)$ 等のパラメータを与える.
 $k = 0$ として 2. \wedge
2. 近似的に以下の最小化問題を解き, 探索方向を計算する :

$$d_k = x_k^+ - x_k, \quad x_k^+ \approx \operatorname{argmin} \left\{ \nabla g(x_k)(x - x_k) + \frac{1}{2} \|x - x_k\|_{B_k}^2 + h(x) \right\}$$

such that $\|r_k\|_{H_k} \leq (1 - \theta_k) \|x_k^+ - x_k\|$, ($r_k \in \nabla g(x_k) + B_k(x_k^+ - x_k) + \partial h(x_k^+)$).

3. 停止判定を行う.
4. 条件

Armijo条件に相当する条件

$$f(x_k + \beta^i d_k) \leq f(x_k) + \delta \beta^i (\nabla g(x_k)^\top d_k + h(x_k^+) - h(x_k))$$

を満たす最小の非負定数 i を見つけ, $\alpha_k = \beta^i$ とする.

5. $x_{k+1} = x_k + \alpha_k d_k$, $k = k + 1$ として, 2 \wedge 戻る.

直線探索に関する命題

命題 $\alpha \in (0,1]$ に対して, 以下が成立する:

$$f(x_k + \alpha d_k) \leq f(x_k) + \alpha(\nabla g(x_k)^\top d_k + h(x_k^+) - h(x_k)) + \frac{\alpha^2 L}{2} \|d_k\|^2 \quad (\text{A})$$

$$\nabla g(x_k)^\top d_k + h(x_k^+) - h(x_k) \leq -\bar{\theta} \|d_k\|_{B_k}^2 \quad (\text{B})$$

直線探索条件

$$f(x_k + \alpha_k d_k) \leq f(x_k) + \delta \alpha_k \underbrace{(\nabla g(x_k)^\top d_k + h(x_k^+) - h(x_k))}_{\leq -\bar{\theta} \|d_k\|_{B_k}^2}$$

➡ 直線探索は実行可能でさらに $f(x_{k+1}) \leq f(x_k)$ が成立.

➡ $x_k \in \mathcal{L}_0 = \{x \mid f(x) \leq f(x_0)\}$

停留点である条件

命題

ある k 回目の反復において x_k が停留点である必要十分条件は

$$d_k = 0$$

である.

⇒ の証明 省略

⇐ の証明 (概略)

残差の定義 $r_k \in \nabla g(x_k) + B_k(x_k^+ - x_k) + \partial h(x_k^+)$

非厳密の基準 $\|r_k\|_{H_k} \leq (1 - \theta_k) \|x_k^+ - x_k\|_{B_k} \quad (\theta_k \in [\bar{\theta}, 1], \bar{\theta} \in (0, 1])$

$$d_k = x_k^+ - x_k = 0 \text{ より } r_k = 0 \quad \rightarrow \quad 0 \in \nabla g(x_k) + \partial h(x_k) \quad \square$$

大域的収束性

補題 以下が成立する：

$$\beta \min \left\{ 1, \frac{2m}{L} \bar{\theta} (1 - \delta) \right\} \leq \alpha_k \leq 1.$$

正の定数

定理

目的関数が下で有界であるならば

$$\lim_{k \rightarrow \infty} \|d_k\| = 0$$

が成立する。さらに、 $\{x_k\}$ が有界ならば、 $\{x_k\}$ の任意の集積点は停留点である。

大域的収束性の証明 (概略)

直線探索条件より

$$\begin{aligned} f(x_{k+1}) - f(x_k) &\leq \delta \alpha_k (\nabla g(x_k)^\top d_k + h(x_k^+) - h(x_k)) \\ &\leq -\delta \alpha_k \bar{\theta} \|d_k\|_{B_k}^2 \\ &\leq -\delta \bar{\alpha} \bar{\theta} \|d_k\|_{B_k}^2 \\ &\leq -\delta \bar{\alpha} \bar{\theta} m \|d_k\|^2 \leq 0 \end{aligned}$$

$\{f(x_k)\}$ は下に有界な非増加列なので極限が存在する。
よって, $\lim_{k \rightarrow \infty} \|d_k\| = 0$ が成立する.

残差の定義 $r_k \in \nabla g(x_k) + B_k(x_k^+ - x_k) + \partial h(x_k^+)$

非厳密の基準 $\|r_k\|_{H_k} \leq (1 - \theta_k) \|x_k^+ - x_k\|_{B_k} \quad (\theta_k \in [\bar{\theta}, 1], \bar{\theta} \in (0, 1])$

一方, 非厳密の基準 と $\lim_{k \rightarrow \infty} \|d_k\| = 0$ より

$$\lim_{k \rightarrow \infty} \|r_k\| = 0$$

なので, 任意の停留点を \bar{x} とすると, $0 \in \nabla g(\bar{x}) + \partial h(\bar{x})$ が成立する \square

R-1次収束性

定理

関数 g は強凸関数, つまり, 正の定数 μ が存在して

$$g(y) \geq g(x) + \nabla g(x)^\top (y - x) + \frac{\mu}{2} \|y - x\|^2 \quad \text{for any } x, y \in R^n$$

を満たすものとする. このとき, 以下が成立する:

$$\|x_k - x^*\| \leq \rho^k \sqrt{\frac{\mu}{2} (f(x_0) - f(x^*))}$$

ただし,

$$\rho = \sqrt{1 - \frac{1}{2} \frac{\delta \bar{\alpha}}{1 + 4M/(\bar{\theta}\mu) + L/(4m\bar{\theta})}} \in (0, 1)$$

である.

数値実験 その1

テスト問題

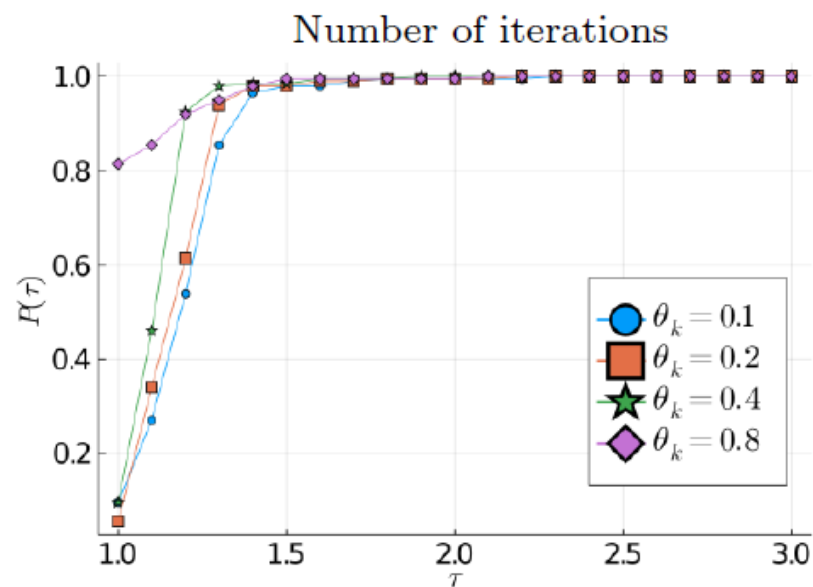
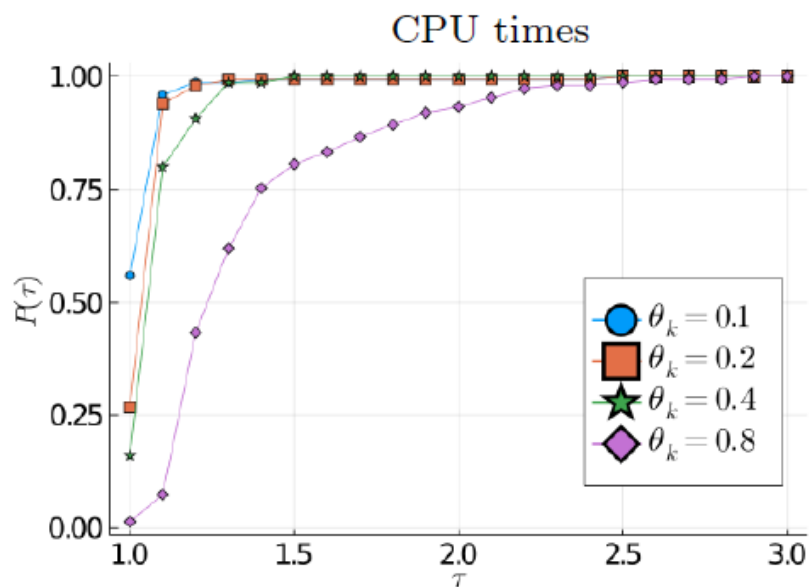
$$\min_x \frac{1}{2} x^\top A x - b^\top x \quad \text{s.t. } x \in \Omega, \quad \Omega = \{x \mid e^\top x = c, x \geq 0\}$$

$$\rightarrow \min \frac{1}{2} x^\top A x - b^\top x + h(x), \quad h(x) = \begin{cases} 0 & x \in \Omega \\ +\infty & \text{otherwise} \end{cases}$$

- 内部反復はV-FISTAで求解
- A, b は乱数で生成
- 次元 $n = 1000, 2500, 5000, 7500$ とし, 各次元で50問 (合計200インスタンス) 作成
- その他詳細は Nakayama, N. and Yabe (2021) を参照
- 比較対象
 - PNOPT : Lee et al. (2014) のニュートン型近接勾配法
 - PG:近接勾配法
 - PGe:Wen et al. (2017)の外挿付き近接勾配法
 - mless-Pre(0.8): 提案法 (プレコンベックスクラス)
 - mless-SR1 : 提案法(SR1公式)

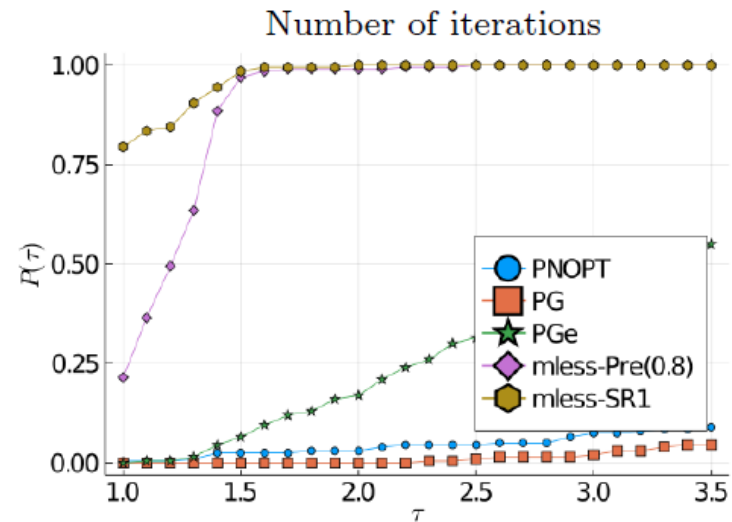
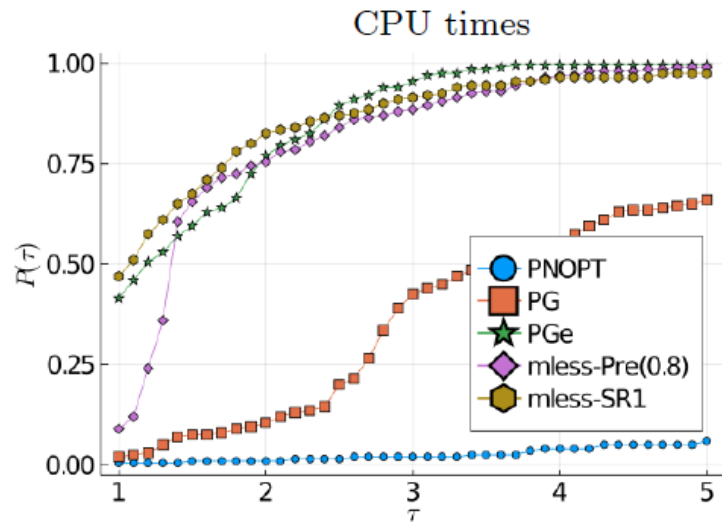
数値実験 その1

BFGS公式で非厳密さのパラメータ θ_k を動かした場合

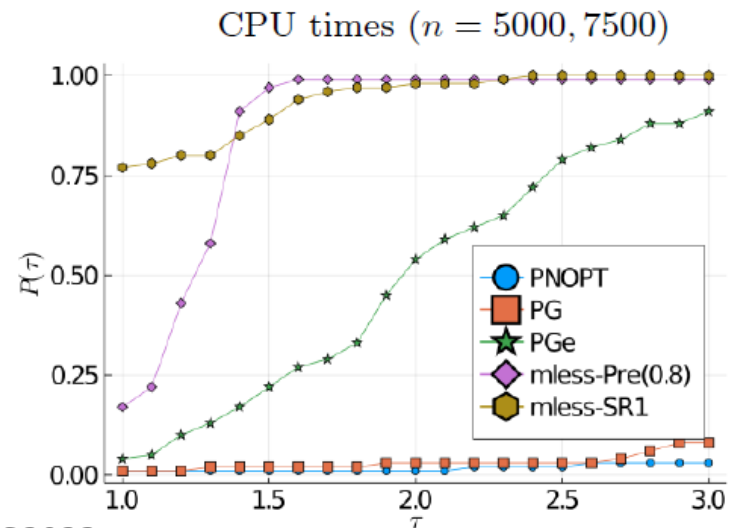
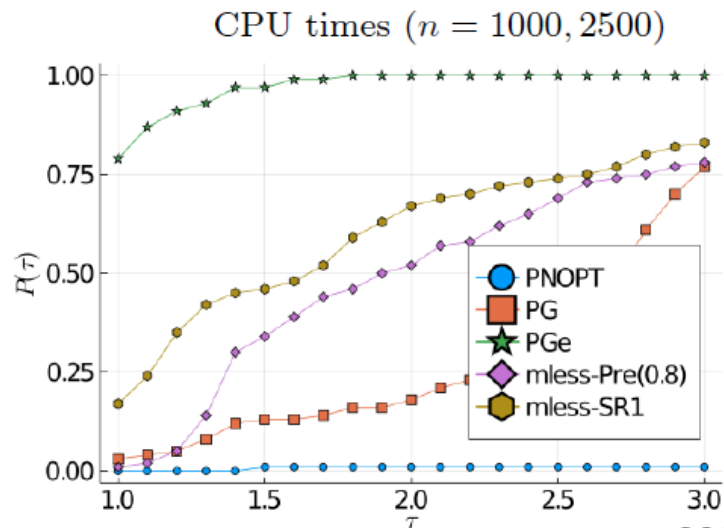


数値実験 その1

全体



次元ごと



数値実験 その2

テスト問題

$$\min_{x \in \mathbb{R}^n} \frac{1}{p} \sum_{i=1}^p \log(1 + \exp(-b_i x^T a_i)) + C \|x\|_1$$

- $a_i \in \mathbb{R}^n, b_i \in \{-1, 1\}, i = 1, \dots, p$ は以下のデータセットを利用

Name	a9a	a1a	w8a	w1a	leukemia	gisette
サンプル数 p	16,281	30,956	292,363	47,272	34	1000
次元 n	123	123	300	300	7129	5000

- 比較方法
 - PNOPT : Lee et al. (2014) のニュートン型近接勾配法
 - TFOCS : FISTAを実装したソフトウェア
 - mless-Pre(0.8): 提案法 (プレコンベックスクラス)
 - mless-SR1 : 提案法(SR1公式)

数値実験 その2

	mless-Pre(0.8)		mless-SR1		PNOPT		TFOCS	
Data	Iter	Time	Iter	Time	Iter	Time	Iter	Time
a9a	134	1.78	148	1.65	31	1.81	1025	7.68
a1a	132	2.65	126	2.13	32	2.34	1074	17.49
w8a	2760	12.07	1198	6.32	29	1.32	331	1.62
w1a	1584	28.26	1321	22.91	28	1.49	286	5.05
leukemia	4179	62.85	777	222.09	512	64.25	1298	5.13
gisette	1088	173.02	1714	269.05	264	119.26	7025	1457.52

ニュートン型近接勾配法の課題(私見)

近接勾配法: $x_{k+1} = \operatorname{argmin}_x \left\{ g(x_k) + \nabla g(x_k)^\top (x - x_k) + h(x) + \frac{L_k}{2} \|x - x_k\|^2 \right\}$

ニュートン型近接勾配法: $d_k = x_k^+ - x_k$, $x_k^+ \approx \operatorname{argmin} \left\{ \nabla g(x_k)(x - x_k) + \frac{1}{2} \|x - x_k\|_{B_k}^2 + h(x) \right\}$

- FISTAなどと比較するとやはり1回あたりの内部反復の手間が大きい
 - Becker et al. (2012)の結果の利用

Proposition (Becker and Fadili (2012))

Let $V = D + uu^\top$, where D is diagonal with positive diagonal elements, and $u \in \mathbb{R}^n$. Then,

$$\operatorname{Prox}_h^V(z) = D^{-1/2} \circ \operatorname{Prox}_{h \circ D^{-1/2}}(D^{1/2}z - v), \quad \dots (A)$$

where $v = \alpha D^{-1/2}u$ and α is the unique root of

$$p(\alpha) = \left\langle u, z - D^{-1/2} \circ \operatorname{Prox}_{h \circ D^{-1/2}} \circ D^{1/2}(z - \alpha D^{-1/2}u) \right\rangle + \alpha.$$

- FISTAのような加速の提案
- 収束率 $O\left(\frac{1}{k^2}\right)$ の保証, KL Property の下での理論保証など
 - 直線探索をProxの計算の外側で実行しているのがネック

まとめ

- 第1部
 - 無制約最適化の基礎
 - 反復法
 - 直線探索
 - 最急勾配法, ニュートン法, 準ニュートン法
- 第2部
 - 大規模無制約最適化問題に対する反復法
 - 記憶制限準ニュートン法, BB法
 - 非線形共役勾配法
 - メモリーレス準ニュートン法
- 第3部
 - 微分不可能な正則化項を含む問題への拡張
 - 近接勾配法
 - ニュートン型近接勾配法