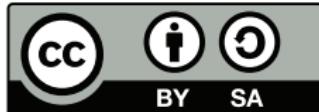


オペレーションズ・リサーチ II (3)

田中 俊二

shunji.tanaka@okayama-u.ac.jp

本文書のライセンスは CC-BY-SA にしたがいます



スケジュール

No.	内容
1	導入 (非線形最適化問題, ゲーム理論, 多目的最適化問題)
2	非線形計画 1 (勾配, ヘッセ行列, 凸性, 最適性条件, ニュートン法)
3	非線形計画 2 (最急降下法, 準ニュートン法, 共役勾配法, 信頼領域法)
4	非線形計画 3 (制約つき問題の最適性条件, KKT 条件, ペナルティ関数法, 2 次計画法, 逐次 2 次計画法)
5	ゲーム理論 1 (種々のゲーム, 標準形, 純粋戦略, 混合戦略, ナッシュ均衡)
6	ゲーム理論 2 (展開形ゲーム, 繰り返しゲーム)
7	多目的最適化 (パレート最適性, 重み付け法, ϵ 制約法, 重み付きメトリック法)

降下法 (descent method)

制約なし非線形計画問題 (P0)

$$\begin{aligned} & \min f(\boldsymbol{x}) \\ \text{s.t. } & \quad \boldsymbol{x} \in \mathbb{R}^n \end{aligned}$$

仮定

$f(\boldsymbol{x})$ は C^1 級

降下法 (descent method)

- $f(\boldsymbol{x}^{(k+1)}) < f(\boldsymbol{x}^{(k)})$ が成立つよう $\boldsymbol{x}^{(k)}$ を更新
- 勾配 $\nabla f(\boldsymbol{x}^{(k)})$ の情報を用いたため、勾配法 (gradient method) とも
- ニュートン法と同様停留点を探す
- 運悪く鞍点が見つかる可能性もあるが、極大点は回避可能
- **大域的収束**
- 降下法: **最急降下法** (勾配降下法), **準ニュートン法**, etc.
- 共役降下法: **共役勾配法**, etc.

最急降下法

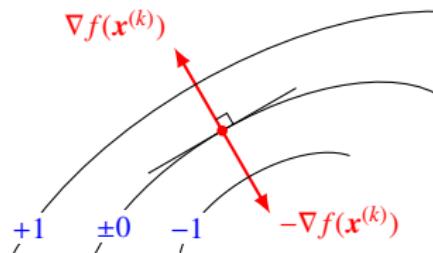
降下法の基本的枠組

$\mathbf{x}^{(k+1)} = \mathbf{x}^{(k)} + \alpha^{(k)} \mathbf{d}^{(k)}$ により解を更新

- $\mathbf{d}^{(k)}$: 探索方向 (search direction)
- $\alpha^{(k)}$: ステップ幅 (step size. $\alpha^{(k)} > 0$)

最急降下法 (steepest descent method)

$\mathbf{d}^{(k)} = -\nabla f(\mathbf{x}^{(k)})$ とする降下法



最急降下とは?

最急降下方向

最急降下方向 (steepest descent direction) \mathbf{d}^*

$\mathbf{x}^{(k)}$ における $f(\mathbf{x})$ の方向微分が最小になる方向

最急降下方向の別の表現

$f(\mathbf{x})$ の 1 次近似を $h_k(\mathbf{x}) = f(\mathbf{x}^{(k)}) + \nabla^\top f(\mathbf{x}^{(k)})(\mathbf{x} - \mathbf{x}^{(k)})$ として, $h_k(\mathbf{x}^{(k)} + \mathbf{d})$ を最小化する \mathbf{d}

最急降下方向の計算

$$h_k(\mathbf{x}^{(k)} + \mathbf{d}) = f(\mathbf{x}^{(k)}) + \nabla^\top f(\mathbf{x}^{(k)})\mathbf{d}$$

$f(\mathbf{x}^{(k)})$ は定数なので, $h_k(\mathbf{x})$ を最小化するには, $\nabla^\top f(\mathbf{x}^{(k)})\mathbf{d}$ を最小化すればよい.

$\nabla^\top f(\mathbf{x}^{(k)})\mathbf{d}$ は $\nabla f(\mathbf{x}^{(k)})$ と \mathbf{d} の内積, \mathbf{d} を単位ベクトルとすると

$$\nabla^\top f(\mathbf{x}^{(k)})\mathbf{d} = \|\nabla f(\mathbf{x}^{(k)})\| \cdot \|\mathbf{d}\| \cos \theta = \|\nabla f(\mathbf{x}^{(k)})\| \cos \theta$$

ただし, θ は $\nabla f(\mathbf{x}^{(k)})$ と \mathbf{d} のなす角. したがって, $\theta = \pi$ のとき最小で, 最急降下方向 \mathbf{d}^* は

$$\mathbf{d}^* = -\frac{\nabla f(\mathbf{x}^{(k)})}{\|\nabla f(\mathbf{x}^{(k)})\|}$$

最急降下法

$\mathbf{x}^{(k+1)} = \mathbf{x}^{(k)} + \alpha^{(k)} \mathbf{d}^{(k)}, \mathbf{d}^{(k)} = -\nabla f(\mathbf{x}^{(k)})$ により解を更新

ステップ幅の決定

降下法の基本的枠組

$\mathbf{x}^{(k+1)} = \mathbf{x}^{(k)} + \alpha^{(k)} \mathbf{d}^{(k)}$ により解を更新

- $\mathbf{d}^{(k)}$: 探索方向 (search direction)
- $\alpha^{(k)}$: ステップ幅 (step size. $\alpha^{(k)} > 0$)

ステップ幅 $\alpha^{(k)}$ の決定

探索方向 $\mathbf{d}^{(k)}$ を決定後, ステップ幅 $\alpha^{(k)}$ を決定 $\Rightarrow f(\mathbf{x}^{(k)} + \alpha^{(k)} \mathbf{d}^{(k)})$ を最小化

直線探索 (line search)

- $f(\mathbf{x}^{(k)} + \alpha \mathbf{d}^{(k)})$ が最小となる α を探索
- $f(\mathbf{x}^{(k)} + \alpha \mathbf{d}^{(k)})$ は α に関する 1 変数関数なので, $f(\mathbf{x})$ の最小化より楽
- とはいっても、最小化する (最適解を求める) には計算時間がかかるので、特定の条件を満たす α を探索
 - アルミホ (Armijo) 条件
 - ウルフ (Wolfe) 条件
 - etc.

直線探索：アルミホ条件

アルミホ条件 (Armijo condition)

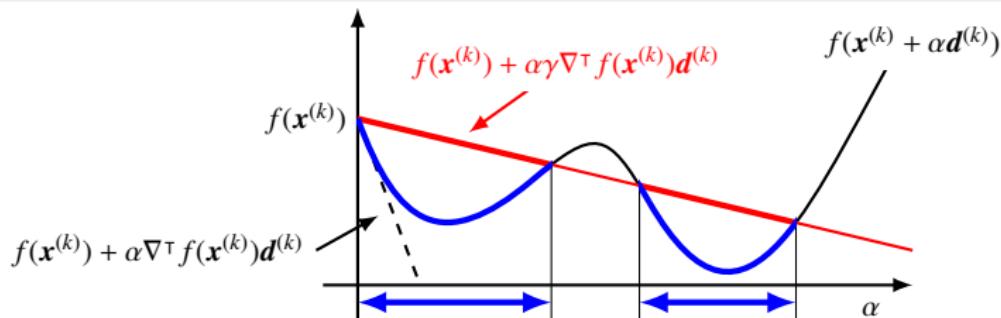
$\gamma (0 < \gamma < 1)$ をパラメータとして,

$$f(\mathbf{x}^{(k)} + \alpha \mathbf{d}^{(k)}) \leq f(\mathbf{x}^{(k)}) + \alpha \gamma \nabla^\top f(\mathbf{x}^{(k)}) \mathbf{d}^{(k)}$$

バックトラック法 (backtracking) によるステップ幅の探索

1. (初期化) $\alpha := 1$ とする.
2. (反復) α がアルミホ条件を満たさないなら, $\alpha := \beta\alpha$ として 2 へ.
3. (終了) $\alpha^{(k)} := \alpha$ として終了.

$\beta (0 < \beta < 1), \gamma (0 < \gamma < 1)$ はパラメータ



直線探索：ウルフ条件

アルミホ条件の問題点

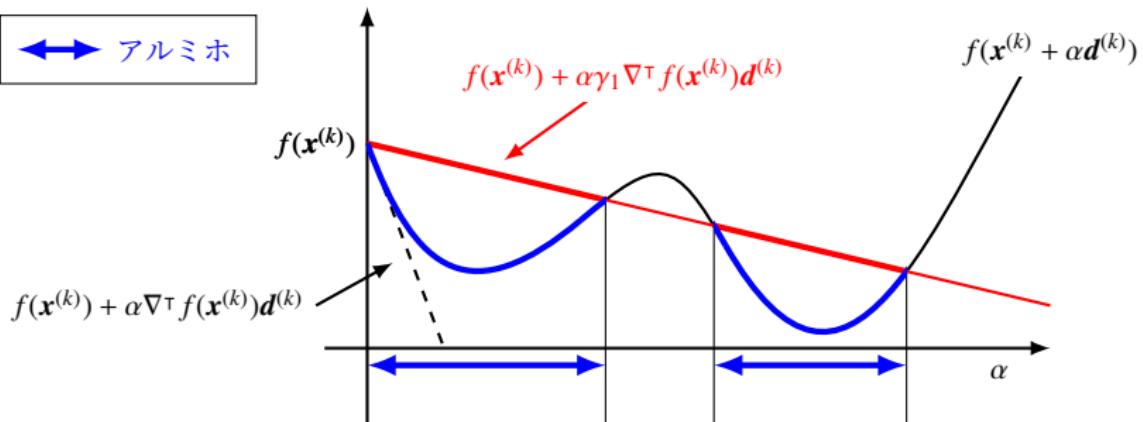
小さな α が選ばれる可能性 $\Rightarrow f(\mathbf{x})$ あまり減少しない

弱 ウルフ条件 (weak Wolfe conditions)

$$f(\mathbf{x}^{(k)} + \alpha \mathbf{d}^{(k)}) \leq f(\mathbf{x}^{(k)}) + \alpha \gamma_1 \nabla^\top f(\mathbf{x}^{(k)}) \mathbf{d}^{(k)} \quad (\text{アルミホ条件})$$

$$\nabla^\top f(\mathbf{x}^{(k)} + \alpha \mathbf{d}^{(k)}) \mathbf{d}^{(k)} \geq \gamma_2 \nabla^\top f(\mathbf{x}^{(k)}) \mathbf{d}^{(k)} \quad (\text{曲率条件})$$

曲率条件 (curvature condition) により、小さすぎる α を回避 ($0 < \gamma_2 < 1$)



直線探索：ウルフ条件

アルミホ条件の問題点

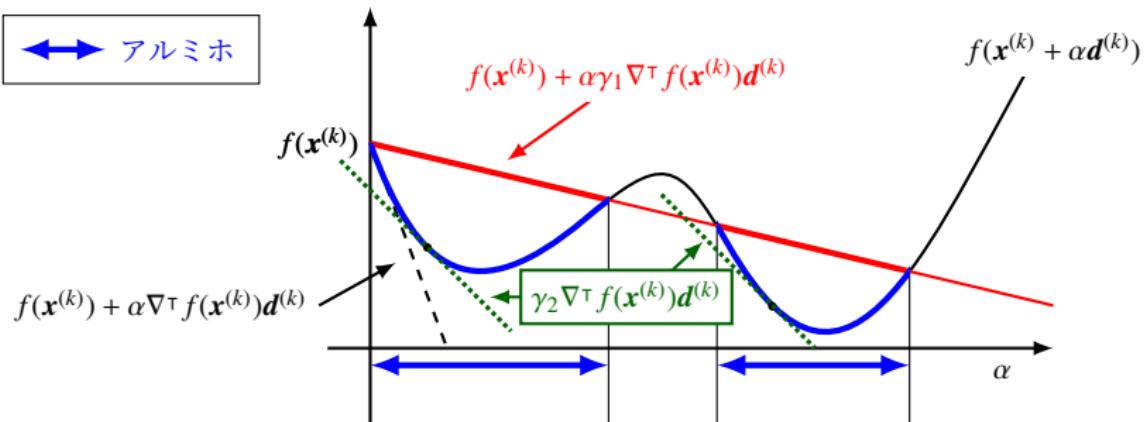
小さな α が選ばれる可能性 $\Rightarrow f(\mathbf{x})$ あまり減少しない

弱 ウルフ条件 (weak Wolfe conditions)

$$f(\mathbf{x}^{(k)} + \alpha \mathbf{d}^{(k)}) \leq f(\mathbf{x}^{(k)}) + \alpha \gamma_1 \nabla^\top f(\mathbf{x}^{(k)}) \mathbf{d}^{(k)} \quad (\text{アルミホ条件})$$

$$\nabla^\top f(\mathbf{x}^{(k)} + \alpha \mathbf{d}^{(k)}) \mathbf{d}^{(k)} \geq \gamma_2 \nabla^\top f(\mathbf{x}^{(k)}) \mathbf{d}^{(k)} \quad (\text{曲率条件})$$

曲率条件 (curvature condition) により、小さすぎる α を回避 ($0 < \gamma_2 < 1$)



直線探索：ウルフ条件

アルミホ条件の問題点

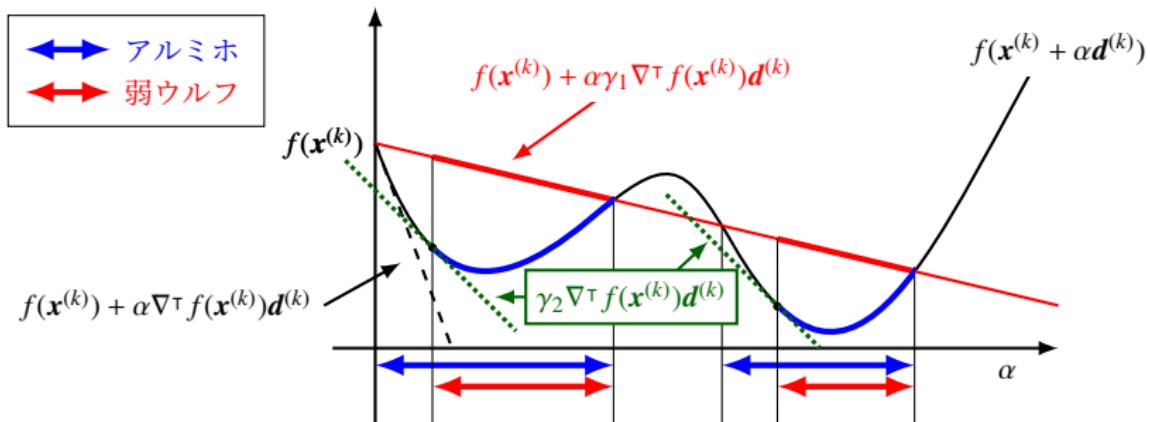
小さな α が選ばれる可能性 $\Rightarrow f(\mathbf{x})$ あまり減少しない

弱ウルフ条件 (weak Wolfe conditions)

$$f(\mathbf{x}^{(k)} + \alpha \mathbf{d}^{(k)}) \leq f(\mathbf{x}^{(k)}) + \alpha \gamma_1 \nabla^\top f(\mathbf{x}^{(k)}) \mathbf{d}^{(k)} \quad (\text{アルミホ条件})$$

$$\nabla^\top f(\mathbf{x}^{(k)} + \alpha \mathbf{d}^{(k)}) \mathbf{d}^{(k)} \geq \gamma_2 \nabla^\top f(\mathbf{x}^{(k)}) \mathbf{d}^{(k)} \quad (\text{曲率条件})$$

曲率条件 (curvature condition) により、小さすぎる α を回避 ($0 < \gamma_2 < 1$)



直線探索：ウルフ条件

アルミホ条件の問題点

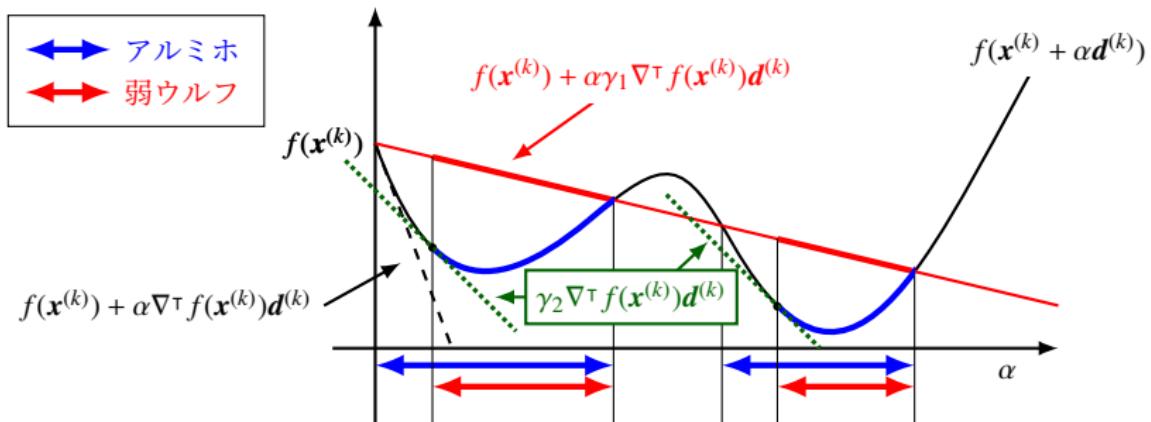
小さな α が選ばれる可能性 $\Rightarrow f(\mathbf{x})$ あまり減少しない

強ウルフ条件 (strong Wolfe conditions)

$$f(\mathbf{x}^{(k)} + \alpha \mathbf{d}^{(k)}) \leq f(\mathbf{x}^{(k)}) + \alpha \gamma_1 \nabla^\top f(\mathbf{x}^{(k)}) \mathbf{d}^{(k)} \quad (\text{アルミホ条件})$$

$$\nabla^\top f(\mathbf{x}^{(k)} + \alpha \mathbf{d}^{(k)}) \mathbf{d}^{(k)} \geq \gamma_2 \nabla^\top f(\mathbf{x}^{(k)}) \mathbf{d}^{(k)} \quad (\text{曲率条件})$$

曲率条件 (curvature condition) により、小さすぎる α を回避 ($0 < \gamma_2 < 1$)



直線探索：ウルフ条件

アルミホ条件の問題点

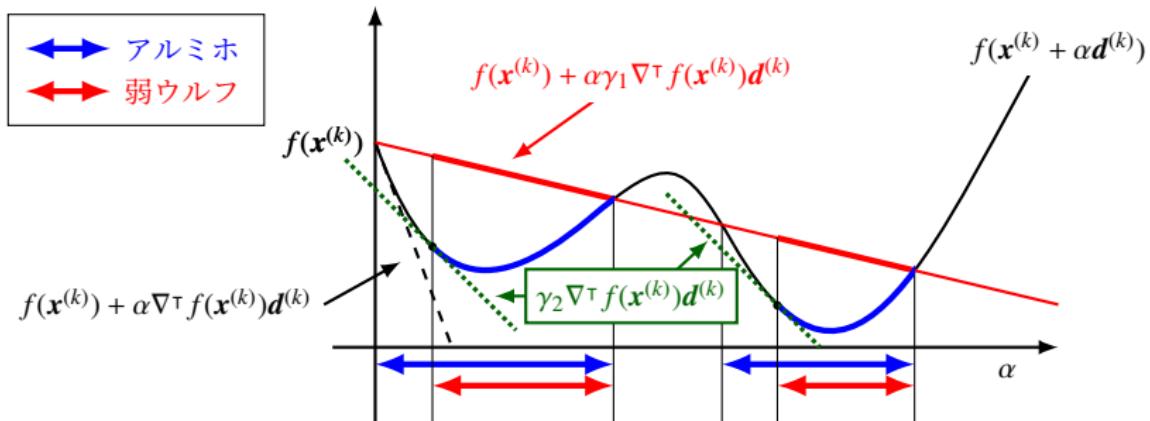
小さな α が選ばれる可能性 $\Rightarrow f(\mathbf{x})$ あまり減少しない

強ウルフ条件 (strong Wolfe conditions)

$$f(\mathbf{x}^{(k)} + \alpha \mathbf{d}^{(k)}) \leq f(\mathbf{x}^{(k)}) + \alpha \gamma_1 \nabla^\top f(\mathbf{x}^{(k)}) \mathbf{d}^{(k)} \quad (\text{アルミホ条件})$$

$$|\nabla^\top f(\mathbf{x}^{(k)} + \alpha \mathbf{d}^{(k)}) \mathbf{d}^{(k)}| \leq \gamma_2 |\nabla^\top f(\mathbf{x}^{(k)}) \mathbf{d}^{(k)}| \quad (\text{曲率条件})$$

正の大きな曲率を禁止 ($0 < \gamma_2 < 1$)



直線探索：ウルフ条件

アルミホ条件の問題点

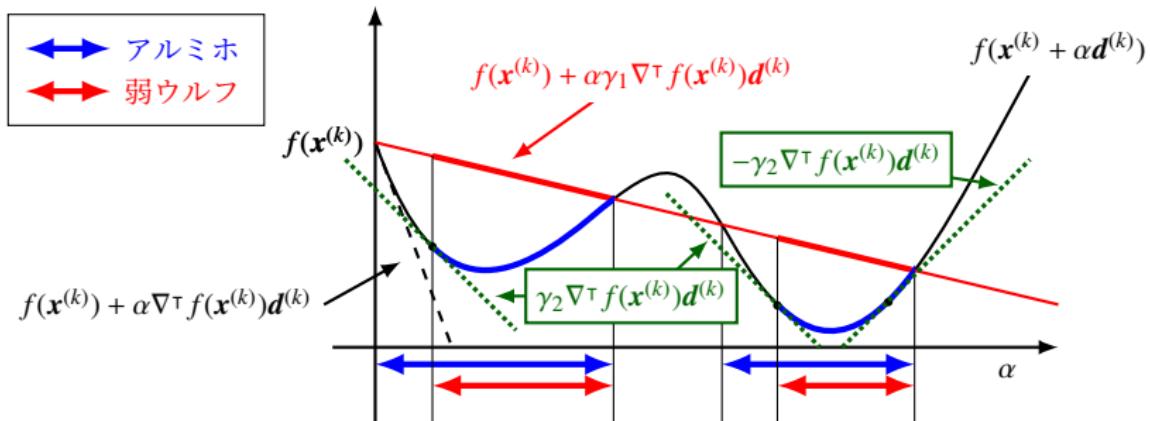
小さな α が選ばれる可能性 $\Rightarrow f(\mathbf{x})$ あまり減少しない

強ウルフ条件 (strong Wolfe conditions)

$$f(\mathbf{x}^{(k)} + \alpha \mathbf{d}^{(k)}) \leq f(\mathbf{x}^{(k)}) + \alpha \gamma_1 \nabla^\top f(\mathbf{x}^{(k)}) \mathbf{d}^{(k)} \quad (\text{アルミホ条件})$$

$$|\nabla^\top f(\mathbf{x}^{(k)} + \alpha \mathbf{d}^{(k)}) \mathbf{d}^{(k)}| \leq \gamma_2 |\nabla^\top f(\mathbf{x}^{(k)}) \mathbf{d}^{(k)}| \quad (\text{曲率条件})$$

正の大きな曲率を禁止 ($0 < \gamma_2 < 1$)



直線探索：ウルフ条件

アルミホ条件の問題点

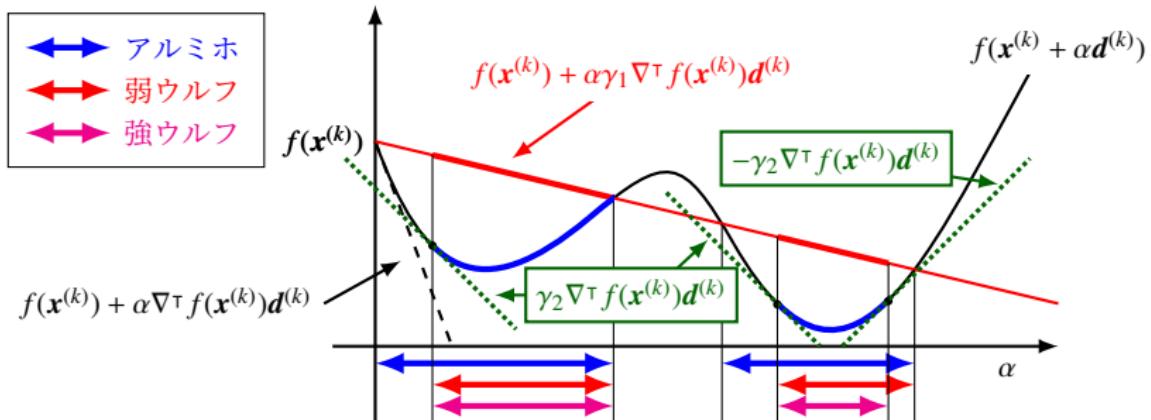
小さな α が選ばれる可能性 $\Rightarrow f(\mathbf{x})$ あまり減少しない

強ウルフ条件 (strong Wolfe conditions)

$$f(\mathbf{x}^{(k)} + \alpha \mathbf{d}^{(k)}) \leq f(\mathbf{x}^{(k)}) + \alpha \gamma_1 \nabla^\top f(\mathbf{x}^{(k)}) \mathbf{d}^{(k)} \quad (\text{アルミホ条件})$$

$$|\nabla^\top f(\mathbf{x}^{(k)} + \alpha \mathbf{d}^{(k)}) \mathbf{d}^{(k)}| \leq \gamma_2 |\nabla^\top f(\mathbf{x}^{(k)}) \mathbf{d}^{(k)}| \quad (\text{曲率条件})$$

正の大きな曲率を禁止 ($0 < \gamma_2 < 1$)



最急降下法の適用例(その1)

例題

$$\begin{aligned} \min \quad & f(x_1, x_2) = 4(x_1 + x_2)^2 + 9(x_1 - x_2)^2 \\ \text{s.t.} \quad & x_1, x_2 \in \mathbb{R} \end{aligned}$$

直線探索にはアルミホ条件(バックトラック法)を用い、ステップ幅の初期値は1,
 $\gamma = 0.1$, $\beta = 0.8$ とする

1. 初期解を $\mathbf{x}^{(0)} = \begin{pmatrix} 1.2000 \\ 1.0000 \end{pmatrix}$ とする

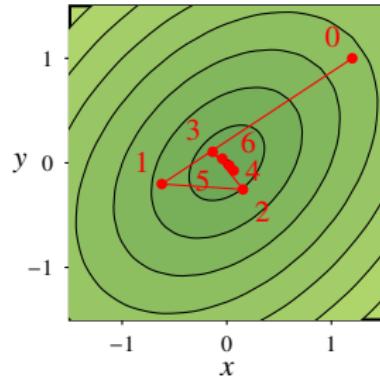
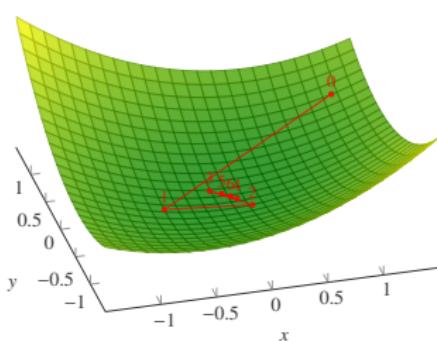
2. 探索方向を $\mathbf{d}^{(0)} = -\nabla f(1.2000, 1.0000) = \begin{pmatrix} -21.2000 \\ -14.0000 \end{pmatrix}$ とする.

3. バックトラック法でアルミホ条件を満たすステップ幅 $\alpha^{(0)}$ を求める.
 $\alpha^{(0)} = \beta^{11} \approx 0.0859$ のときアルミホ条件を満たす.

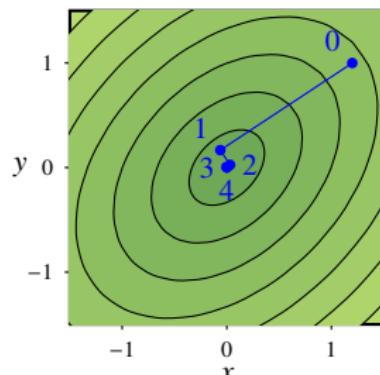
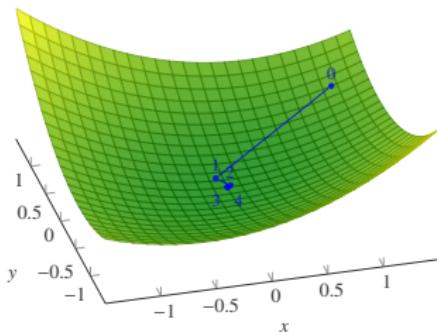
4. $\mathbf{x}^{(1)} := \mathbf{x}^{(0)} + \alpha^{(0)} \mathbf{d}^{(0)} = \begin{pmatrix} -0.6211 \\ -0.2026 \end{pmatrix}$ と更新する.

以下、同様に解を更新すると、
 $\mathbf{x}^{(2)} = \begin{pmatrix} 0.1553 \\ -0.2544 \end{pmatrix}$, $\mathbf{x}^{(3)} = \begin{pmatrix} -0.1342 \\ 0.1048 \end{pmatrix}$, $\mathbf{x}^{(4)} = \begin{pmatrix} 0.0654 \\ -0.0741 \end{pmatrix}$,
 $\mathbf{x}^{(5)} = \begin{pmatrix} -0.0419 \\ 0.0394 \end{pmatrix}$, $\mathbf{x}^{(6)} = \begin{pmatrix} 0.0233 \\ -0.0241 \end{pmatrix}$. (最適解は $\mathbf{x}^* = \mathbf{0}$.)

最急降下法の適用例 (その 2)

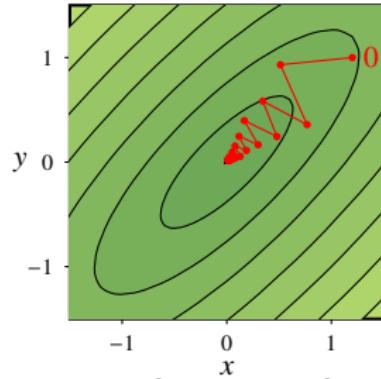
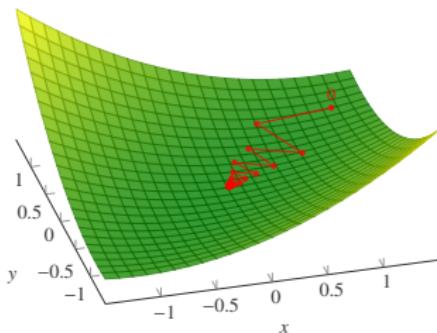


バックトラック法による直線探索

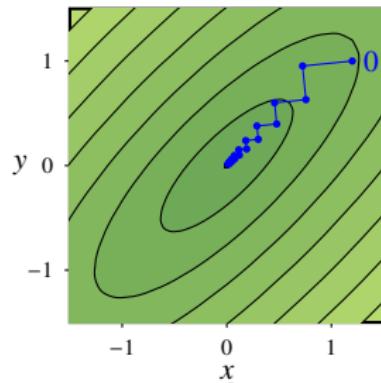
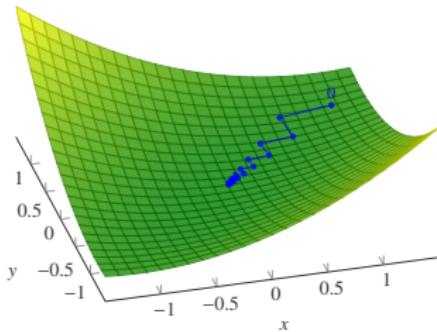


直線探索でステップ幅の最適値を求めた場合

最急降下法の適用例 (その 2)



バックトラック法による直線探索 ($f = (x_1 + x_2)^2 + 9(x_1 - x_2)^2$)



直線探索でステップ幅の最適値を求めた場合 ($f = (x_1 + x_2)^2 + 9(x_1 - x_2)^2$)

最急降下法の長所

- ニュートン法のようにヘッセ行列を用いないため、目的関数が1回微分可能なら適用可能
- 大域的収束
- アルゴリズムが簡単で適用が容易
- 降下方向に向かうため、ニュートン法のように極大点を探すことはない
ただし、運が悪ければ鞍点が見つかることがある

最急降下法の短所

- 収束が遅い。1次収束
 - 探索方向がジグザグ
 - 直線探索で最適なステップ幅を選んでも回避不可能
- 先ほどの2つの例：ニュートン法なら1回で収束(凸2次関数なので)

ニュートン法再び

ニュートン法のアルゴリズム

1. (初期化) 適当に初期解 $\mathbf{x}^{(0)}$ を決める. $k := 0$ とする
2. (終了判定) 終了条件を満たしているなら $\mathbf{x}^{(k)}$ を解として出力し終了
3. (解の更新) ニュートン方向 $\mathbf{d}^{(k)} := -\nabla^2 f(\mathbf{x}^{(k)})^{-1} \nabla f(\mathbf{x}^{(k)})$ を用いて, 解を
 $\mathbf{x}^{(k+1)} := \mathbf{x}^{(k)} + \mathbf{d}^{(k)}$ により更新
4. (次の反復へ) k を 1 増やして 2 へ

- ニュートン法は降下法と類似
 $\mathbf{d}^{(k)}$ 方向に直線探索するニュートン法：直線探索付きニュートン法
- 最急降下法より収束が速い. 2 次収束
- ただし, ニュートン方向が降下方向ではない可能性
- 一般に大域的収束の保証はなし
- ヘッセ行列が正則でない可能性も

ニュートン法再び

ニュートン法のアルゴリズム

1. (初期化) 適当に初期解 $\mathbf{x}^{(0)}$ を決める. $k := 0$ とする
2. (終了判定) 終了条件を満たしているなら $\mathbf{x}^{(k)}$ を解として出力し終了
3. (解の更新) ニュートン方向 $\mathbf{d}^{(k)} := -\nabla^2 f(\mathbf{x}^{(k)})^{-1} \nabla f(\mathbf{x}^{(k)})$ を用いて, 解を $\mathbf{x}^{(k+1)} := \mathbf{x}^{(k)} + \mathbf{d}^{(k)}$ により更新
4. (次の反復へ) k を 1 増やして 2 へ

- ニュートン法は降下法と類似
 $\mathbf{d}^{(k)}$ 方向に直線探索するニュートン法：直線探索付きニュートン法
- 最急降下法より**収束が速い**. 2 次収束
- ただし, ニュートン方向が**降下方向**ではない可能性
- 一般に**大域的収束**の保証はなし
- ヘッセ行列が**正則**でない可能性も

↓

準ニュートン法
(quasi-Newton method)

準ニュートン法 (Quasi-Newton method)

基本方針

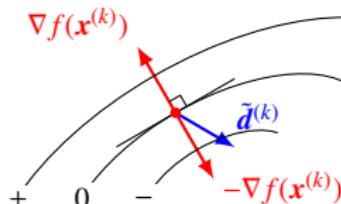
- ヘッセ行列 $\nabla^2 f(\mathbf{x}^{(k)})$ の代わりに正定行列 $B^{(k)}$
- ニュートン方向 $\mathbf{d}^{(k)} = -\nabla^2 f(\mathbf{x}^{(k)})^{-1} \nabla f(\mathbf{x}^{(k)})$ の代わりに
 $\tilde{\mathbf{d}}^{(k)} = -(B^{(k)})^{-1} \nabla f(\mathbf{x}^{(k)})$
- $B^{(k)}$ の選び方
 - $\tilde{\mathbf{d}}^{(k)}$ が降下方向
 - $\nabla^2 f(\mathbf{x}^{(k)})$ を近似

降下方向

最急降下方向 $-\nabla f(\mathbf{x}^{(k)})$ となす角が $\pm 90^\circ$ 以内 (内積が正) なら降下方向

$$(\tilde{\mathbf{d}}^{(k)})^\top (-\nabla f(\mathbf{x}^{(k)})) = \nabla^\top f(\mathbf{x}^{(k)}) (B^{(k)})^{-1} \nabla^\top f(\mathbf{x}^{(k)}) > 0$$

$B^{(k)}$ が正定行列なら $(B^{(k)})^{-1}$ も正定行列であることに注意！



準ニュートン法 (Quasi-Newton method)

基本方針

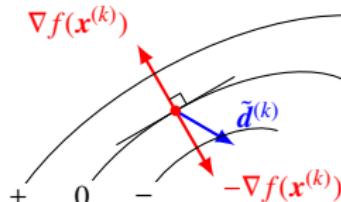
- ヘッセ行列 $\nabla^2 f(\mathbf{x}^{(k)})$ の代わりに正定行列 $B^{(k)}$
- ニュートン方向 $\mathbf{d}^{(k)} = -\nabla^2 f(\mathbf{x}^{(k)})^{-1} \nabla f(\mathbf{x}^{(k)})$ の代わりに
 $\tilde{\mathbf{d}}^{(k)} = -(B^{(k)})^{-1} \nabla f(\mathbf{x}^{(k)})$
- $B^{(k)}$ の選び方
 - $\tilde{\mathbf{d}}^{(k)}$ が降下方向
 - $\nabla^2 f(\mathbf{x}^{(k)})$ を近似

降下方向

最急降下方向 $-\nabla f(\mathbf{x}^{(k)})$ となす角が $\pm 90^\circ$ 以内 (内積が正) なら降下方向

$$(\tilde{\mathbf{d}}^{(k)})^\top (-\nabla f(\mathbf{x}^{(k)})) = \nabla^\top f(\mathbf{x}^{(k)}) (B^{(k)})^{-1} \nabla^\top f(\mathbf{x}^{(k)}) > 0$$

$B^{(k)}$ が正定行列なら $(B^{(k)})^{-1}$ も正定行列であることに注意！



準ニュートン法(続き)

基本方針

- ヘッセ行列 $\nabla^2 f(\mathbf{x}^{(k)})$ の代わりに正定行列 $B^{(k)}$
- ニュートン方向 $\mathbf{d}^{(k)} = -\nabla^2 f(\mathbf{x}^{(k)})^{-1} \nabla f(\mathbf{x}^{(k)})$ の代わりに
 $\tilde{\mathbf{d}}^{(k)} = -(B^{(k)})^{-1} \nabla f(\mathbf{x}^{(k)})$
- $B^{(k)}$ の選び方
 - $\tilde{\mathbf{d}}^{(k)}$ が降下方向
 - $\nabla^2 f(\mathbf{x}^{(k)})$ を近似

ヘッセ行列 $\nabla^2 f(\mathbf{x}^{(k)})$ の近似

- $\nabla f(\mathbf{x})$ の 1 次近似

$$\begin{aligned}\nabla f(\mathbf{x}) &\simeq \nabla f(\mathbf{x}^{(k)}) + \nabla^2 f(\mathbf{x}^{(k)})(\mathbf{x} - \mathbf{x}^{(k)}) \\ \nabla f(\mathbf{x}^{(k+1)}) - \nabla f(\mathbf{x}^{(k)}) &\simeq \nabla^2 f(\mathbf{x}^{(k)})(\mathbf{x}^{(k+1)} - \mathbf{x}^{(k)})\end{aligned}\tag{*}$$

- (*) 式を用いて $B^{(k)}$ を決定 \Rightarrow セカント条件 (secant condition)

$$\nabla f(\mathbf{x}^{(k+1)}) - \nabla f(\mathbf{x}^{(k)}) = B^{(k+1)}(\mathbf{x}^{(k+1)} - \mathbf{x}^{(k)})$$

準ニュートン法のアルゴリズム

準ニュートン法のアルゴリズム

1. (初期化) 適当に初期解 $\mathbf{x}^{(0)}$ および $B^{(0)}$ を決める. $k := 0$ とする
2. (終了判定) 終了条件を満たしているなら $\mathbf{x}^{(k)}$ を解として出力し終了
3. (解の更新) 準ニュートン方向 $\tilde{\mathbf{d}}^{(k)} := -(B^{(k)})^{-1} \nabla f(\mathbf{x}^{(k)})$ を用いて, 解を $\mathbf{x}^{(k+1)} := \mathbf{x}^{(k)} + \alpha^{(k)} \tilde{\mathbf{d}}^{(k)}$ により更新
4. ($B^{(k+1)}$ の計算) $B^{(k+1)}$ を計算
5. (次の反復へ) $k := k + 1$ として 3 へ

- ステップ幅 $\alpha^{(k)}$ は直線探索により計算
- $B^{(k)}$ の計算には, **BFGS 更新**がよく用いられる
- $B^{(0)}$ は正定行列を選ぶ. たとえば $B^{(0)} = I$ など
- 準ニュートン法は大域的収束で**超1次収束**

BFGS 更新 (Broyden-Fletcher-Goldfarb-Shanno update)

$$B^{(k+1)} = B^{(k)} - \frac{(B^{(k)} s^{(k)}) (B^{(k)} s^{(k)})^\top}{(s^{(k)})^\top B^{(k)} s^{(k)}} + \frac{y^{(k)} (y^{(k)})^\top}{(s^{(k)})^\top y^{(k)}}$$

ただし, $y^{(k)} = \nabla f(\mathbf{x}^{(k+1)}) - \nabla f(\mathbf{x}^{(k)})$, $s^{(k)} = \mathbf{x}^{(k+1)} - \mathbf{x}^{(k)}$

準ニュートン法のアルゴリズム(続き)

BFGS 更新 (Broyden-Fletcher-Goldfarb-Shanno update)

$$B^{(k+1)} = B^{(k)} - \frac{(B^{(k)} s^{(k)}) (B^{(k)} s^{(k)})^\top}{(s^{(k)})^\top B^{(k)} s^{(k)}} + \frac{y^{(k)} (y^{(k)})^\top}{(s^{(k)})^\top y^{(k)}}$$

ただし, $y^{(k)} = \nabla f(\mathbf{x}^{(k+1)}) - \nabla f(\mathbf{x}^{(k)})$, $s^{(k)} = \mathbf{x}^{(k+1)} - \mathbf{x}^{(k)}$

- $B^{(k+1)}$ はセカント条件を満たす
- $B^{(k)}$ が正定行列かつ $(s^{(k)})^\top y^{(k)} > 0$ なら, $B^{(k+1)}$ は正定行列
- 弱ウルフ条件の曲率条件を満たすステップ幅を用いると, $(s^{(k)})^\top y^{(k)} > 0$ は成り立つ

セカント条件

$$\nabla f(\mathbf{x}^{(k+1)}) - \nabla f(\mathbf{x}^{(k)}) = B^{(k+1)} (\mathbf{x}^{(k+1)} - \mathbf{x}^{(k)})$$

弱ウルフ条件

$$f(\mathbf{x}^{(k)} + \alpha \tilde{\mathbf{d}}^{(k)}) \leq f(\mathbf{x}^{(k)}) + \alpha \gamma_1 \nabla^\top f(\mathbf{x}^{(k)}) \tilde{\mathbf{d}}^{(k)} \quad (\text{アルミホ条件})$$

$$\nabla^\top f(\mathbf{x}^{(k)} + \alpha \tilde{\mathbf{d}}^{(k)}) \tilde{\mathbf{d}}^{(k)} \geq \gamma_2 \nabla^\top f(\mathbf{x}^{(k)}) \tilde{\mathbf{d}}^{(k)} \quad (\text{曲率条件})$$

$B^{(k+1)}$ の正定性を示すのは少し面倒なので, それ以外の部分を示す

練習問題：BFGS 更新がセカント条件を満たすことの証明

BFGS 更新 (Broyden-Fletcher-Goldfarb-Shanno update)

$$B^{(k+1)} = B^{(k)} - \frac{(B^{(k)} s^{(k)}) (B^{(k)} s^{(k)})^\top}{(s^{(k)})^\top B^{(k)} s^{(k)}} + \frac{y^{(k)} (y^{(k)})^\top}{(s^{(k)})^\top y^{(k)}}$$

セカント条件

$$\nabla f(\mathbf{x}^{(k+1)}) - \nabla f(\mathbf{x}^{(k)}) = B^{(k+1)} (\mathbf{x}^{(k+1)} - \mathbf{x}^{(k)})$$

練習問題：BFGS 更新がセカント条件を満たすことの証明

BFGS 更新 (Broyden-Fletcher-Goldfarb-Shanno update)

$$B^{(k+1)} = B^{(k)} - \frac{(B^{(k)} s^{(k)}) (B^{(k)} s^{(k)})^\top}{(s^{(k)})^\top B^{(k)} s^{(k)}} + \frac{y^{(k)} (y^{(k)})^\top}{(s^{(k)})^\top y^{(k)}}$$

セカント条件

$$\nabla f(\mathbf{x}^{(k+1)}) - \nabla f(\mathbf{x}^{(k)}) = B^{(k+1)} (\mathbf{x}^{(k+1)} - \mathbf{x}^{(k)})$$

$$y^{(k)} = B^{(k+1)} s^{(k)}$$

練習問題：BFGS 更新がセカント条件を満たすことの証明

BFGS 更新 (Broyden-Fletcher-Goldfarb-Shanno update)

$$B^{(k+1)} = B^{(k)} - \frac{(B^{(k)} s^{(k)}) (B^{(k)} s^{(k)})^\top}{(s^{(k)})^\top B^{(k)} s^{(k)}} + \frac{y^{(k)} (y^{(k)})^\top}{(s^{(k)})^\top y^{(k)}}$$

セカント条件

$$\begin{aligned}\nabla f(\mathbf{x}^{(k+1)}) - \nabla f(\mathbf{x}^{(k)}) &= B^{(k+1)}(\mathbf{x}^{(k+1)} - \mathbf{x}^{(k)}) \\ y^{(k)} &= B^{(k+1)} s^{(k)}\end{aligned}$$

$$\begin{aligned}B^{(k+1)} s^{(k)} &= B^{(k)} s^{(k)} - \frac{(B^{(k)} s^{(k)}) (B^{(k)} s^{(k)})^\top}{(s^{(k)})^\top B^{(k)} s^{(k)}} s^{(k)} + \frac{y^{(k)} (y^{(k)})^\top}{(s^{(k)})^\top y^{(k)}} s^{(k)} \\ &= B^{(k)} s^{(k)} - \frac{1}{(s^{(k)})^\top B^{(k)} s^{(k)}} (B^{(k)} s^{(k)}) (B^{(k)} s^{(k)})^\top s^{(k)} + \frac{1}{(s^{(k)})^\top y^{(k)}} y^{(k)} (y^{(k)})^\top s^{(k)} \\ &= B^{(k)} s^{(k)} - \frac{1}{(s^{(k)})^\top B^{(k)} s^{(k)}} (B^{(k)} s^{(k)}) (s^{(k)})^\top B^{(k)} s^{(k)} + \frac{1}{(s^{(k)})^\top y^{(k)}} y^{(k)} (s^{(k)})^\top y^{(k)} \\ &= B^{(k)} s^{(k)} - B^{(k)} s^{(k)} + y^{(k)} \\ &= y^{(k)}\end{aligned}$$

曲率条件を満たすステップ幅に対し条件式が成り立つことの証明

BFGS 更新 (Broyden-Fletcher-Goldfarb-Shanno update)

$$B^{(k+1)} = B^{(k)} - \frac{(B^{(k)} s^{(k)}) (B^{(k)} s^{(k)})^\top}{(s^{(k)})^\top B^{(k)} s^{(k)}} + \frac{y^{(k)} (y^{(k)})^\top}{(s^{(k)})^\top y^{(k)}}$$

条件式

$$(s^{(k)})^\top y^{(k)} > 0$$

曲率条件 ($0 < \gamma_2 < 1$)

$$\nabla^\top f(\mathbf{x}^{(k)} + \alpha^{(k)} \tilde{\mathbf{d}}^{(k)}) \tilde{\mathbf{d}}^{(k)} \geq \gamma_2 \nabla^\top f(\mathbf{x}^{(k)}) \tilde{\mathbf{d}}^{(k)}$$

曲率条件を満たすステップ幅に対し条件式が成り立つことの証明

BFGS 更新 (Broyden-Fletcher-Goldfarb-Shanno update)

$$B^{(k+1)} = B^{(k)} - \frac{(B^{(k)} s^{(k)}) (B^{(k)} s^{(k)})^\top}{(s^{(k)})^\top B^{(k)} s^{(k)}} + \frac{y^{(k)} (y^{(k)})^\top}{(s^{(k)})^\top y^{(k)}}$$

条件式

$$(s^{(k)})^\top y^{(k)} > 0$$

$$\alpha^{(k)} (\nabla^\top f(\mathbf{x}^{(k+1)}) - \nabla^\top f(\mathbf{x}^{(k)})) \tilde{\mathbf{d}}^{(k)} > 0$$

曲率条件 ($0 < \gamma_2 < 1$)

$$\nabla^\top f(\mathbf{x}^{(k)} + \alpha^{(k)} \tilde{\mathbf{d}}^{(k)}) \tilde{\mathbf{d}}^{(k)} \geq \gamma_2 \nabla^\top f(\mathbf{x}^{(k)}) \tilde{\mathbf{d}}^{(k)}$$

$$\nabla^\top f(\mathbf{x}^{(k+1)}) \tilde{\mathbf{d}}^{(k)} \geq \gamma_2 \nabla^\top f(\mathbf{x}^{(k)}) \tilde{\mathbf{d}}^{(k)} \quad (*)$$

曲率条件を満たすステップ幅に対し条件式が成り立つことの証明

BFGS 更新 (Broyden-Fletcher-Goldfarb-Shanno update)

$$B^{(k+1)} = B^{(k)} - \frac{(B^{(k)} s^{(k)}) (B^{(k)} s^{(k)})^\top}{(s^{(k)})^\top B^{(k)} s^{(k)}} + \frac{y^{(k)} (y^{(k)})^\top}{(s^{(k)})^\top y^{(k)}}$$

条件式

$$(s^{(k)})^\top y^{(k)} > 0$$
$$\alpha^{(k)} (\nabla^\top f(\mathbf{x}^{(k+1)}) - \nabla^\top f(\mathbf{x}^{(k)})) \tilde{\mathbf{d}}^{(k)} > 0$$

曲率条件 ($0 < \gamma_2 < 1$)

$$\nabla^\top f(\mathbf{x}^{(k)} + \alpha^{(k)} \tilde{\mathbf{d}}^{(k)}) \tilde{\mathbf{d}}^{(k)} \geq \gamma_2 \nabla^\top f(\mathbf{x}^{(k)}) \tilde{\mathbf{d}}^{(k)}$$
$$\nabla^\top f(\mathbf{x}^{(k+1)}) \tilde{\mathbf{d}}^{(k)} \geq \gamma_2 \nabla^\top f(\mathbf{x}^{(k)}) \tilde{\mathbf{d}}^{(k)} \quad (*)$$

$B^{(k)}$ が正定行列なので $(B^{(k)})^{-1}$ も正定行列であり,

$$\nabla^\top f(\mathbf{x}^{(k)}) \tilde{\mathbf{d}}^{(k)} = -\nabla^\top f(\mathbf{x}^{(k)}) (B^{(k)})^{-1} \nabla f(\mathbf{x}^{(k)}) < 0$$

したがって、条件式は、(*) 式より

$$\alpha^{(k)} (\nabla^\top f(\mathbf{x}^{(k+1)}) - \nabla^\top f(\mathbf{x}^{(k)})) \tilde{\mathbf{d}}^{(k)} \geq \alpha^{(k)} (\gamma_2 - 1) \nabla^\top f(\mathbf{x}^{(k)}) \tilde{\mathbf{d}}^{(k)} > 0$$

共役(きょうやく)勾配法

基本的な考え方

- 最急降下法の問題点：探索方向がジグザグ \Rightarrow 前々回の探索方向が使われる
- 過去に探索した方向とは異なる方向を探索 \Rightarrow **共役方向**

まず凸2次計画問題で考える。すなわち $f(\mathbf{x}) = \frac{1}{2}\mathbf{x}^\top Q\mathbf{x} + \mathbf{q}^\top \mathbf{x}$ (Q は $n \times n$ 正定行列)。

共役方向 (conjugate direction)

$\mathbf{d}_i \in \mathbb{R}^n, \mathbf{d}_j \in \mathbb{R}^n$ が $\mathbf{d}_i^\top Q \mathbf{d}_j = 0$ を満たすとき、 $\mathbf{d}_i, \mathbf{d}_j$ は **Q共役** (Q-conjugate) であるという。

凸2次計画問題に対する共役勾配法 (conjugate gradient method)

探索方向、ステップ幅を以下で決定 ($\mathbf{x}^{(k+1)} = \mathbf{x}^{(k)} + \alpha^{(k)} \bar{\mathbf{d}}^{(k)}$)

$$\alpha^{(k)} = -\frac{\nabla^\top f(\mathbf{x}^{(k)}) \bar{\mathbf{d}}^{(k)}}{(\bar{\mathbf{d}}^{(k)})^\top Q \bar{\mathbf{d}}^{(k)}}, \quad \bar{\mathbf{d}}^{(k)} = \begin{cases} -\nabla f(\mathbf{x}^{(k)}) & (k=0) \\ -\nabla f(\mathbf{x}^{(k)}) + \frac{\nabla^\top f(\mathbf{x}^{(k)}) Q \bar{\mathbf{d}}^{(k-1)}}{(\bar{\mathbf{d}}^{(k-1)})^\top Q \bar{\mathbf{d}}^{(k-1)}} \bar{\mathbf{d}}^{(k-1)} & (k \geq 1) \end{cases}$$

一般の非線形最適化問題に対する共役勾配法

凸 2 次計画問題に対する共役勾配法

探索方向, ステップ幅を以下で決定 ($\mathbf{x}^{(k+1)} = \mathbf{x}^{(k)} + \alpha^{(k)} \bar{\mathbf{d}}^{(k)}$)

$$\alpha^{(k)} = -\frac{\nabla^\top f(\mathbf{x}^{(k)}) \bar{\mathbf{d}}^{(k)}}{(\bar{\mathbf{d}}^{(k)})^\top Q \bar{\mathbf{d}}^{(k)}}, \quad \bar{\mathbf{d}}^{(k)} = \begin{cases} -\nabla f(\mathbf{x}^{(k)}) & (k=0) \\ -\nabla f(\mathbf{x}^{(k)}) + \frac{\nabla^\top f(\mathbf{x}^{(k)}) Q \bar{\mathbf{d}}^{(k-1)}}{(\bar{\mathbf{d}}^{(k-1)})^\top Q \bar{\mathbf{d}}^{(k-1)}} \bar{\mathbf{d}}^{(k-1)} & (k \geq 1) \end{cases}$$

- 共役勾配法の方向 $\bar{\mathbf{d}}_{(k)}$: $\bar{\mathbf{d}}^{(0)}, \dots, \bar{\mathbf{d}}^{(k-1)}$ と Q 共役
- Q が正定行列なら, n 反復以内に (大域的) 最適解に収束
(ニュートン法は 1 反復で収束)

一般の非線形最適化問題に対する共役勾配法

$$\begin{aligned}\mathbf{x}^{(k+1)} &= \mathbf{x}^{(k)} + \alpha^{(k)} \bar{\mathbf{d}}^{(k)} \\ \bar{\mathbf{d}}^{(k)} &= -\nabla f(\mathbf{x}^{(k)}) + \beta^{(k)} \bar{\mathbf{d}}^{(k-1)}\end{aligned}$$

- ステップ幅 $\alpha^{(k)}$ は直線探索で決定
- Q の代わりにヘッセ行列 $\nabla^2 f(\mathbf{x}^{(k)})$. ただし, 直接計算しない
 $\beta^{(k)}$ を Q に依存しない形に変更

一般の非線形最適化問題に対する共役勾配法の探索方向

一般の非線形最適化問題に対する共役勾配法

$$\boldsymbol{x}^{(k+1)} = \boldsymbol{x}^{(k)} + \alpha^{(k)} \bar{\boldsymbol{d}}^{(k)}$$

$$\bar{\boldsymbol{d}}^{(k)} = -\nabla f(\boldsymbol{x}^{(k)}) + \beta^{(k)} \bar{\boldsymbol{d}}^{(k-1)}$$

$\beta^{(k)}$ の決定法 : Fletcher-Reeves 法

$$\beta^{(k)} = \frac{\nabla^\top f(\boldsymbol{x}^{(k)}) \nabla f(\boldsymbol{x}^{(k)})}{\nabla^\top f(\boldsymbol{x}^{(k-1)}) \nabla f(\boldsymbol{x}^{(k-1)})} = \frac{\|\nabla f(\boldsymbol{x}^{(k)})\|^2}{\|\nabla f(\boldsymbol{x}^{(k-1)})\|^2}$$

$\beta^{(k)}$ の決定法 : Polyak-Polak-Ribiére 法

$$\beta^{(k)} = \frac{(\nabla f(\boldsymbol{x}^{(k)}) - \nabla f(\boldsymbol{x}^{(k-1)}))^\top \nabla f(\boldsymbol{x}^{(k)})}{\nabla^\top f(\boldsymbol{x}^{(k-1)}) \nabla f(\boldsymbol{x}^{(k-1)})} = \frac{(\nabla f(\boldsymbol{x}^{(k)}) - \nabla f(\boldsymbol{x}^{(k-1)}))^\top \nabla f(\boldsymbol{x}^{(k)})}{\|\nabla f(\boldsymbol{x}^{(k-1)})\|^2}$$

基本的な考え方

- 探索方向の決定に目的関数の 2 次近似を用いる (ニュートン法と同様)

$$m_k(s) = \tilde{f}_k(\mathbf{x}^{(k)} + s) = f(\mathbf{x}^{(k)}) + \nabla^\top f(\mathbf{x}^{(k)})s + \frac{1}{2}s^\top \nabla^2 f(\mathbf{x}^{(k)})s$$

- $\mathbf{x}^{(k)}$ から半径 $\Delta^{(k)}$ の領域 (**信頼領域**) 内で 2 次近似が信頼できると考える
- 探索方向 $s^{(k)}$ は、信頼領域内で $m_k(s)$ を最小化する点を用いて計算
- 最小化する点は必ず存在。厳密に求める、もしくは降下法などで近似的に求める
- ステップ幅は 1 に固定 (ニュートン法と同様)
- 現在の信頼領域を元の目的関数とその 2 次近似との差 (近似度) で評価し、信頼半径 $\Delta^{(k)}$ を増減させる
- 大域的収束

信頼領域法のアルゴリズム

信頼領域法 (trust region method) のアルゴリズム

1. (初期化) 適当に初期解 $\mathbf{x}^{(0)}$ および初期信頼半径 $\Delta^{(0)}$ を決める。また、パラメータ $\eta_1, \eta_2, \gamma_1, \gamma_2$ を $0 < \eta_1 < \eta_2 < 1, 0 < \gamma_2 < 1 < \gamma_1$ の範囲で設定する。 $k := 0$ とする
2. (終了判定) 終了条件を満たしているなら $\mathbf{x}^{(k)}$ を解として出力し終了
3. (探索方向・近似度の計算) 問題 $(M^{(k)})$ の最適解（もしくは近似解）を求めて $\mathbf{s}^{(k)}$ とし、近似度を $\rho^{(k)} = \frac{f(\mathbf{x}^{(k)} + \mathbf{s}^{(k)}) - f(\mathbf{x}^{(k)})}{m_k(\mathbf{s}^{(k)}) - m_k(\mathbf{0})}$ とする
4. (解の更新) $\rho^{(k)} \geq \eta_1$ なら $\mathbf{x}^{(k+1)} := \mathbf{x}^{(k)} + \mathbf{s}^{(k)}$ 、そうでなければ $\mathbf{x}^{(k+1)} := \mathbf{x}^{(k)}$ とする
5. (信頼半径の更新) $\Delta^{(k+1)}$ を以下で決定
 - $\eta_2 \leq \rho^{(k)} : \Delta^{(k+1)} := \gamma_1 \Delta^{(k)}$
 - $\eta_1 \leq \rho^{(k)} < \eta_2 : \Delta^{(k+1)} := \Delta^{(k)}$
 - $\rho^{(k)} < \eta_1 : \Delta^{(k+1)} := \gamma_2 \Delta^{(k)}$
6. (次の反復へ) k を 1 増やして 2 へ

問題 $(M^{(k)})$

$$\begin{aligned} \min \quad & m_k(\mathbf{s}) - m_k(\mathbf{0}) = \frac{1}{2} \mathbf{s}^\top \nabla^2 f(\mathbf{x}^{(k)}) \mathbf{s} + \nabla f(\mathbf{x}^{(k)}) \mathbf{s} \\ \text{s.t.} \quad & \|\mathbf{s}\| \leq \Delta^{(k)} \end{aligned}$$