

# Numerical Optimization

Jorge Nocedal, Stephen J. Wright

---

2018/9/8

#夏の理論合宿

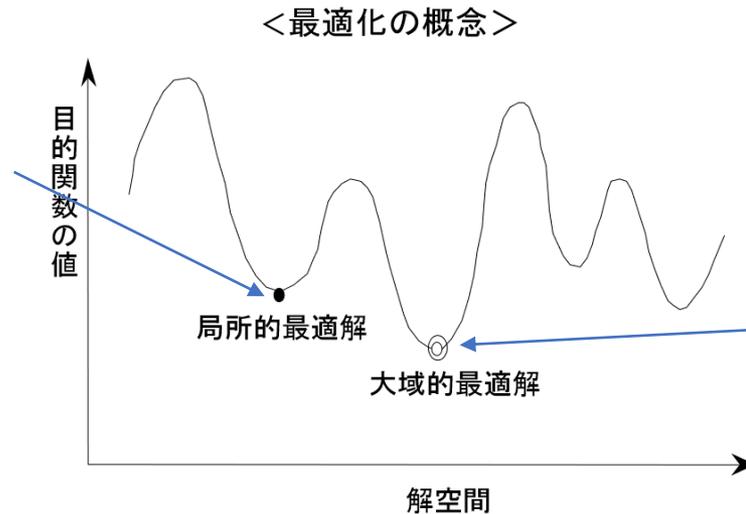
B4 飯塚 卓哉

# 最適化は難しい

非線形計画問題や組み合わせ最適化問題は厳密に解くのはとても難しい

目的関数や制約条件が非線形の問題

一般的な非線形計画問題なら  
ひとまずはここを目指す



凸計画問題なら簡単に  
ここに行ける

関数の形が重要

$\nabla f(x)$ : 点 $x$ における関数 $f$ の勾配     $\nabla^2 f(x)$ : 点 $x$ における関数 $f$ のヘッセ行列

$$\nabla f(x) = \begin{bmatrix} \frac{\partial f(x)}{\partial x_1} \\ \frac{\partial f(x)}{\partial x_2} \\ \vdots \\ \frac{\partial f(x)}{\partial x_n} \end{bmatrix}$$

$$\nabla^2 f(x) = \begin{bmatrix} \frac{\partial^2 f(x)}{\partial x_1^2} & \frac{\partial^2 f(x)}{\partial x_1 x_2} & \cdots & \frac{\partial^2 f(x)}{\partial x_1 x_n} \\ \frac{\partial^2 f(x)}{\partial x_2 \partial x_1} & \frac{\partial^2 f(x)}{\partial x_2^2} & \cdots & \frac{\partial^2 f(x)}{\partial x_2 \partial x_n} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial^2 f(x)}{\partial x_n \partial x_1} & \frac{\partial^2 f(x)}{\partial x_n \partial x_2} & \cdots & \frac{\partial^2 f(x)}{\partial x_n^2} \end{bmatrix}$$

対称行列

例えば...

ヘッセ行列が正定値行列である  
2次関数の等高線は楕円

## • 反復法

基本のアルゴリズム

①初期点を設定

②ある規則に従って次の点を生成

③同じ規則に従ってその次の点を生成

④収束の判定と繰り返し

⑤最終的に収束した点が最適解

(例) 直線探索の場合

$$x_0 \in R^n$$

$$x_1 = x_0 + \alpha_0 d_0$$

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

$$\alpha_k \in R \quad \text{ステップ幅}$$

$$d_k \in R_n \quad \text{探索方向}$$

$\bar{x}$  ← 収束点の表記

# 反復法のアプローチ

連続最適化問題において目的関数の極小値を求めるための基本的な反復アプローチは2つ

$$f: \mathbb{R}^n \rightarrow \mathbb{R} \quad \mathbf{x}^*$$

探索における反復:  $\mathbf{x}_{k+1} = \mathbf{x}_k + \alpha_k \mathbf{p}_k$

ステップ長    探索方向

## ■ 直線探索法(line search method)

最初に目的関数 $f$ が小さくなる降下方向を決め、次にその方向に $\mathbf{x}$ をどれくらい動かすかを表すステップ長を計算する。

探索方向の決め方によって分類され、最急降下法、ニュートン法、準ニュートン法などがある



## ■ 信頼領域法(trust region method)

目的関数があるモデル関数（多くの場合二次関数）で近似することが妥当である、と仮定する部分領域の半径（=信頼領域半径）を決め、降下方向を逐次選択する。

離れたスタート地点から効果的に収束解を得るために、適切に**探索方向**と**ステップ長**を選択する必要がある。

### アルゴリズムの性能指標

**大域的収束性**: どこからスタートしても停留点にたどり着く

**収束率**: 収束するのに時間がかかりすぎてはいけない

<b>3</b>	<b>Line Search Methods</b>	<b>30</b>
3.1	Step Length . . . . .	31
	The Wolfe Conditions . . . . .	33
	The Goldstein Conditions . . . . .	36
	Sufficient Decrease and Backtracking . . . . .	37
3.2	Convergence of Line Search Methods . . . . .	37
3.3	Rate of Convergence . . . . .	41
	Convergence Rate of Steepest Descent . . . . .	42
	Newton's Method . . . . .	44
	Quasi-Newton Methods . . . . .	46
3.4	Newton's Method with Hessian Modification . . . . .	48
	Eigenvalue Modification . . . . .	49
	Adding a Multiple of the Identity . . . . .	51
	Modified Cholesky Factorization . . . . .	52
	Modified Symmetric Indefinite Factorization . . . . .	54
3.5	Step-Length Selection Algorithms . . . . .	56
	Interpolation . . . . .	57
	Initial Step Length . . . . .	59
	A Line Search Algorithm for the Wolfe Conditions . . . . .	60
	Notes and References . . . . .	62

# ステップ長

より効果的に $f$ を減らすステップ長 $\alpha$ を見つけないが、 $\alpha$ を選ぶのに長い時間をかけたくない  
=トレードオフ

最も効果的な $\alpha$  :

関数  $\phi(\alpha) = f(x_k + \alpha p_k) = f(x_{k+1}) \quad \alpha > 0$  の大域的最小解をもたらす $\alpha^*$

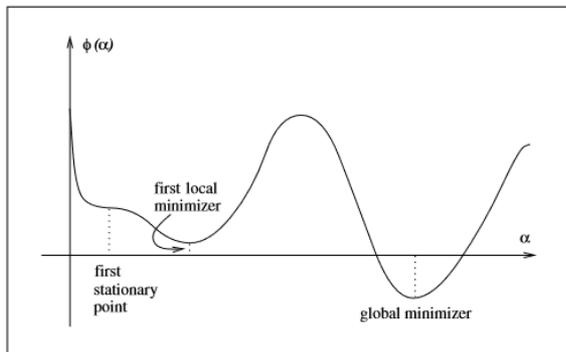


Figure 3.1 The ideal step length is the global minimizer.

↓  
このような $\alpha^*$ を見つけるのは非常に高価

現実的には

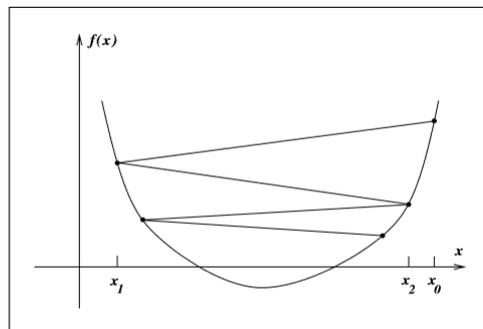
1. ある条件内の**望ましい $\alpha$** を含む範囲を見つけ、
2. その範囲内で良好な $\alpha$ を選択する

という簡易的な方法がとられる。

望ましい $\alpha$ とは? → 必ずしも $\alpha^*$ の近くじゃなくてもいい!

課すべき条件:  $f(x_k + \alpha_k p_k) < f(x_k)$

しかし、減少が不十分な場合がある ↓



いくら $\frac{1}{k}$ 倍していっても  
 $f^* = -1$ にはたどり着かない

Figure 3.2 Insufficient reduction in  $f$ .

十分な減少条件を $\alpha$ に適用する必要がある

# ステップ長：Wolfeの条件

**Armijo条件**：よく知られている不完全な直線探索条件

$$f(x_k + \alpha p_k) \leq f(x_k) + c_1 \alpha \nabla f_k^T p_k, \quad \text{定数 } c_1 \in [0,1]$$

↔  $f$ の減少は、ステップ長 $\alpha_k$ と方向性導関数 $\nabla f_k^T p_k$ の両方に比例する

- 上式の右辺は線形 →  $l(\alpha)$ で表せる
- $l(\cdot)$ は負の傾き $c_1 \nabla f_k^T p_k$ を持つが、 $c_1 \in [0,1]$ なので、小さい $\alpha$ の範囲では $\phi(\alpha)$ の上に来る（実際は $c_1 = 10^{-4}$ くらい）
- つまり $\phi(\alpha) \leq l(\alpha)$ なら十分な減少条件を満たすということにする

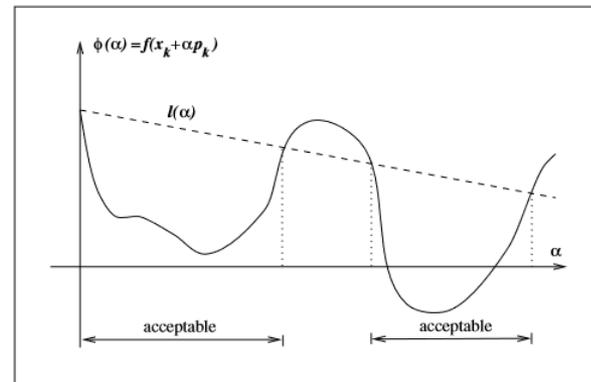


Figure 3.3 Sufficient decrease condition.

一方、 $\alpha$ が小さすぎても意味がない

小さすぎる $\alpha$ を除外するための第2の条件：**曲率条件**

$$\alpha \nabla f(x_k + \alpha_k p_k)^T p_k \geq c_2 \nabla f_k^T p_k, \quad \text{定数 } c_2 \in (c_1, 1) \quad \text{を満たす}$$

- 上式の左辺は導関数 $\phi'(\alpha_k)$
- つまり曲率条件は、 $\alpha_k$ における $\phi$ の傾きが初期傾き $\phi'(0)$ の $c_2$ 倍より大きいという条件
- 傾き $\phi'(\alpha)$ が大きく負なら、その方向に探索することで $f$ を十分減少させることができる
- 一方傾き $\phi'(\alpha)$ がわずかに負、もしくは正ならこの方向では $f$ の減少はそれほど期待できないということだから、探索を終了することは理に適う

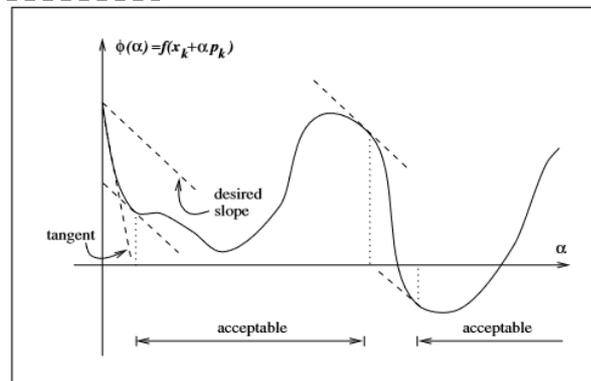


Figure 3.4 The curvature condition.

# ステップ長：Wolfeの条件

**Wolfeの条件**：十分な減少条件と曲率条件

$$\begin{aligned} f(x_k + \alpha_k p_k) &\leq f(x_k) + c_1 \alpha_k \nabla f_k^T p_k, \\ \nabla f(x_k + \alpha_k p_k)^T p_k &\geq c_2 \nabla f_k^T p_k, \end{aligned} \quad 0 < c_1 < c_2 < 1.$$

$\phi(\alpha)$ を最小にする $\alpha$ の近くじゃなくてもWolfeの条件は満たせる

曲率条件を修正して、 $\alpha_k$ を局所最小化解、もしくは停留点の近くに置くことができる

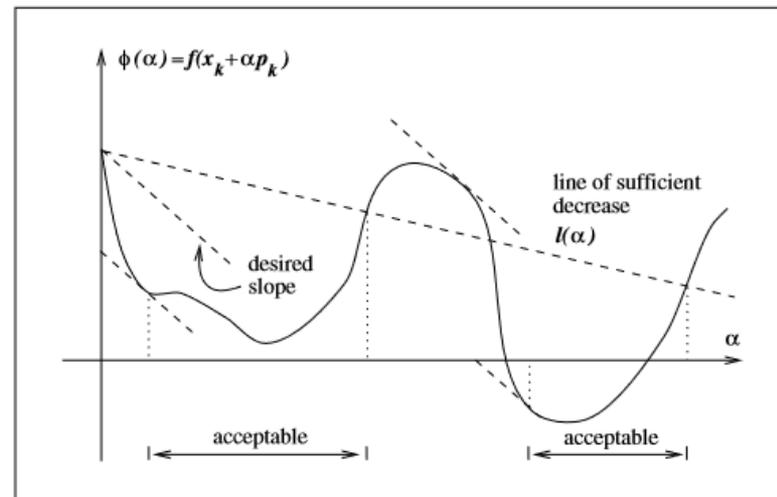


Figure 3.5 Step lengths satisfying the Wolfe conditions.

**強いWolfeの条件**

$$\begin{aligned} f(x_k + \alpha_k p_k) &\leq f(x_k) + c_1 \alpha_k \nabla f_k^T p_k, \\ |\nabla f(x_k + \alpha_k p_k)^T p_k| &\leq c_2 |\nabla f_k^T p_k|, \end{aligned} \quad 0 < c_1 < c_2 < 1.$$

- $\phi'(\alpha_k)$ が大きすぎる正の値にならないので、停留点から遠い点を除外できる
- すべての滑らかで下に有界な関数 $f$ についてWolfeの条件を満たすステップ長 $\alpha$ が存在することが証明されている（証明略）

# ステップ長：Goldsteinの条件

Goldsteinの条件：Wolfeの条件と同じく，十分に減少するが，小さすぎる $\alpha$ を省くための条件

$$f(x_k) + (1 - c)\alpha_k \nabla f_k^T p_k \leq \underline{f(x_k + \alpha_k p_k)} \leq f(x_k) + c\alpha_k \nabla f_k^T p_k, \quad 0 < c < 1/2.$$

Wolfeの十分な減少の条件と同じ

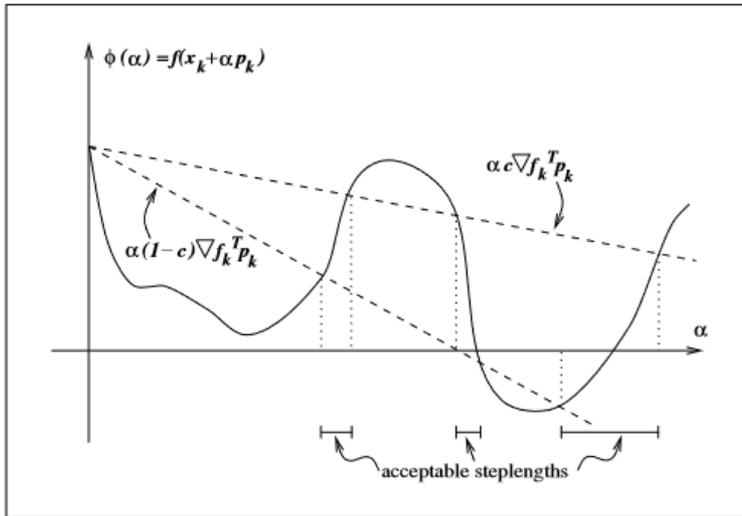


Figure 3.6 The Goldstein conditions.

- Wolfeの条件と比べて， $\phi(\alpha)$ を最小にする $\alpha$ を省いてしまう点で不利
- しかしNewton型の方法でよく使用される。（準ニュートン法を除く）

# 十分な減少とバックトラッキング

バックトラッキング手法を用いて候補ステップ長を適切に選択する場合、曲率条件を用いなくても、直線探索法を行える

**Algorithm 3.1** (Backtracking Line Search).

Choose  $\bar{\alpha} > 0$ ,  $\rho \in (0, 1)$ ,  $c \in (0, 1)$ ; Set  $\alpha \leftarrow \bar{\alpha}$ ;

repeat until  $f(x_k + \alpha p_k) \leq f(x_k) + c\alpha \nabla f_k^T p_k$

$\alpha \leftarrow \rho\alpha$ ;

end (repeat)

Terminate with  $\alpha_k = \alpha$ .

- $\alpha_k$ は最終的には十分に小さくなり、十分な減少条件が成立するため、許容できるステップ長は無限の試行回数後に求められる
- 収縮率 $\rho$ は各反復で変化してもいい
- この方法により、 $\alpha_k$ は、初期値 $\bar{\alpha}$ か、十分な減少条件を満たしてかつ、小さすぎないステップ長になる。
- 後者は、 $\alpha_k$ がひとつ前に十分な減少条件を満たしていないために除外された試行値 $\alpha_k/\rho$ の因数 $\rho$ の範囲内に入っていることにより保証される
- ニュートン法には適しているが、準ニュートン法や、共役勾配法には適していない

# 直線探索法における収束（探索方向について）

大域的収束解を得るには、ステップ長だけでなく、探索方向も適切に選ぶ必要がある

注目するのは探索方向 $p_k$ と、最急降下方向 $-\nabla f_k$ の間の角度 $\theta_k$

$$\cos \theta_k = \frac{-\nabla f_k^T p_k}{\|\nabla f_k\| \|p_k\|}$$

## Zoutendijk条件（大域的収束定理）

$p_k$ が降下方向であり、 $\alpha_k$ がWolfe条件を満たす任意の反復を考える。  $f$ が $\mathbb{R}^n$ 内で下に有界で、初期点 $x_0$ における準位集合 $\{x; f(x) \leq f(x_0)\}$ を含む開集合 $U$ において連続的微分可能であるとする。 また勾配 $\nabla f_k$ は $U$ でLipschitz連続であるとする。

すなわち、ある正定数 $L$ が存在して、

$$\|\nabla f(x) - \nabla f(\tilde{x})\| \leq L\|x - \tilde{x}\|, \quad \text{for all } x, \tilde{x} \in \mathcal{N}. \quad \text{が成り立つとする.}$$

この時 
$$\sum_{k \geq 0} \cos^2 \theta_k \|\nabla f_k\|^2 < \infty. \quad \text{が成り立つ.}$$

- Wolfe条件の代わりに強いWolfe条件や、Goldstein条件を用いても同様の結果が得られる
- 仮定はそれほど限定的ではないのでしばしば満たせる

→この定理から示唆されることは  $\cos^2 \theta_k \|\nabla f_k\|^2 \rightarrow 0.$

この極限は、直線探索アルゴリズムの大域的収束結果を導出するのに使われる。

# 直線探索法における収束（探索方向について）

$$\cos \theta_k = \frac{-\nabla f_k^T p_k}{\|\nabla f_k\| \|p_k\|}$$

探索方向  $p_k$  を選択する方法において、上記で定義した角度  $\theta_k$  が90度から離れるように保証されるなら、

$$\cos \theta_k \geq \delta > 0, \quad \text{for all } k.$$

を満たす定数  $\delta$  が存在する

$$\cos^2 \theta_k \|\nabla f_k\|^2 \rightarrow 0.$$

つまりこれは、上式と合わせて、 $\lim_{k \rightarrow \infty} \|\nabla f_k\| = 0$  を意味する



探索方向が勾配に対して直交方向に近すぎない限り、勾配ノルム  $\|\nabla f_k\|$  は0に収束する

- 特に探索方向  $p_k$  が負の勾配に平行である最急降下法は、Wolfe条件かGoldstein条件を満たす直線探索を行えば、必ず、勾配が0に収束して探索を終了できる
- ニュートン法や準ニュートン法でも、 $p_k = -B_k^{-1} \nabla f_k$  における  $B_k$  が一定の条件を満たせば、 $f$  は大域的に収束する。
- 共役勾配法などの一部のアルゴリズムでもこの性質を示せるがやや弱い結果となる。

※ここでの「大域的に収束」とは、必ずしも  $f$  を最小化するものではなく、停留点にたどり着くことのみを表す

# 収束率

探索方向 $p_k$ が勾配 $\nabla f_k$ に対して直交しない、かつ最急降下ステップが規則的に行われさえすればちゃんと大域的に収束する

→ 毎ステップごとに $\cos \theta_k$ を計算して、それをもとに $p_k$ を最急降下方向に向けられるのでは？ = 角度テスト

But.....これは望ましくない

- 条件の悪い（ヘッセ行列が悪い）問題では、勾配にほぼ直交する探索方法を生成する必要が出てきてしまい、収束速度が悪い
- 高速に収束するアルゴリズムは大域的収束性を破綻させることがあり、逆もありうる  
ex)最急降下法は大域的に収束するが、実際はかなり遅い  
ニュートン法では解の近くから探索を始めれば高速だが、それ以外では降下方向に探索しないことさえある

→ 最急降下法に注目して収束率について考察してみる

## 特徴的な収束率

- **1次収束** ある定数  $q \in (0,1)$  と整数  $k'$  が存在して、

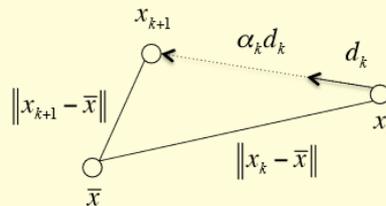
$$\frac{\|x_{k+1} - \bar{x}\|}{\|x_k - \bar{x}\|} \leq q \quad \forall k \geq k' \quad \text{が成り立つこと}$$

- **超1次収束**

$$\lim_{k \rightarrow +\infty} \frac{\|x_{k+1} - \bar{x}\|}{\|x_k - \bar{x}\|} = 0 \quad \text{が成り立つこと}$$

- **p次収束** ある定数  $M > 0$  と整数  $k'$  が存在して、

$$\frac{\|x_{k+1} - \bar{x}\|}{\|x_k - \bar{x}\|^p} \leq M \quad \forall k \geq k' \quad \text{が成り立つこと}$$



# 最急降下法の収束率

目的関数が2次関数の、正確な直線探索を考える

$$f(x) = \frac{1}{2}x^T Qx - b^T x, \quad Q \text{は正定値対称行列}$$

最適解 $x^*$ は、 $f$ の $x$ に関する停留条件 $\nabla f_k = Qx - b = 0$ より $x^* = Q^{-1}b$

最急降下法に基づき、解は $x_{k+1} = x_k - \alpha_k \nabla f_k$ に従って更新.

ステップ幅 $\alpha_k$ を正確な直線探索に従って求めると $\alpha_k$ は

$$f(x_k - \alpha \nabla f_k) = \frac{1}{2}(x_k - \alpha \nabla f_k)^T Q(x_k - \alpha \nabla f_k) - b^T(x_k - \alpha \nabla f_k)$$

の最小解なので,

$$\alpha_k = \frac{\nabla f_k^T \nabla f_k}{\nabla f_k^T Q \nabla f_k}.$$

以上より,

$$x_{k+1} = x_k - \left( \frac{\nabla f_k^T \nabla f_k}{\nabla f_k^T Q \nabla f_k} \right) \nabla f_k.$$

収束率を定量化するために、重み付きノルム  $\|x\|_Q^2 = x^T Qx$ . を導入すると

$x^* = Q^{-1}b$  より,  $\frac{1}{2}\|x - x^*\|_Q^2 = f(x) - f(x^*)$ , 現在の目的関数値と最適化された目的関数値の差

また,  $\nabla f_k = Q(x_k - x^*)$ , となることを用いると

$$\|x_{k+1} - x^*\|_Q^2 = \left\{ 1 - \frac{(\nabla f_k^T \nabla f_k)^2}{(\nabla f_k^T Q \nabla f_k)(\nabla f_k^T Q^{-1} \nabla f_k)} \right\} \|x_k - x^*\|_Q^2$$

$Q$ の固有値を $\lambda_i (0 < \lambda_1 \leq \lambda_2 \leq \dots \leq \lambda_n)$ とおき, Kantorovichの不等式を適用すると理解がしやすくなる

# 最急降下法の収束率

**定理**：目的関数が凸2次関数の、正確な直線探索の下での最急降下法では誤差ノルムが以下を満たす

$$\|x_{k+1} - x^*\|_Q^2 \leq \left( \frac{\lambda_n - \lambda_1}{\lambda_n + \lambda_1} \right)^2 \|x_k - x^*\|_Q^2,$$

意味：関数値 $f_k$ は**線形率**で最小値 $f^*$ に収束する（**1次収束**）

- 特に $Q$ が単位行列の整数倍なら固有値がすべて等しくなって1回で収束する。
- 条件数 $\kappa(Q) = \frac{\lambda_n}{\lambda_1}$ が大きくなるほど、等高線の楕円が長く伸び、収束性が悪化する

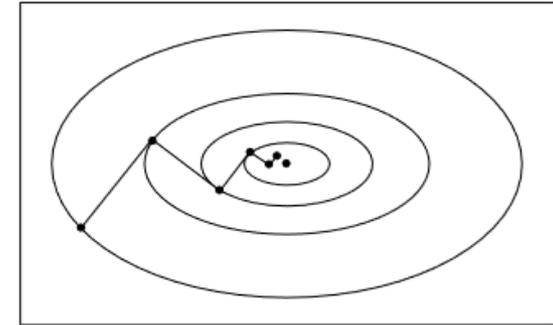


Figure 3.7 Steepest descent steps.

この結果は、一般的な非線形目的関数において本質的に同じ

**定理**： $f: \mathbb{R}^n \rightarrow \mathbb{R}$ が連続的に2回微分可能であり、正確な直線探索による最急降下法による反復で、ヘッセ行列 $\nabla^2 f(x^*)$ が正定値となる点 $x^*$ に収束すると仮定する。

$r$ を  $r \in \left( \frac{\lambda_n - \lambda_1}{\lambda_n + \lambda_1}, 1 \right)$ , を満たす任意のスカラーとすると

十分大きいすべての $k$ について

$$f(x_{k+1}) - f(x^*) \leq r^2 [f(x_k) - f(x^*)], \quad \text{が成り立つ.}$$

この定理から、ヘッセ行列が良い状態でも、最急降下法の収束率は許容できないほど遅いということが分かる

## Newton法

探索方向： $p_k^N = -\nabla^2 f_k^{-1} \nabla f_k$ .

- ヘッセ行列が常に正定値とは限らないため、探索方向 $p_k$ が常に降下方向になるとは限らない  
→ ヘッセ行列の修正や、信頼領域法で解決している
- **2次収束**：誤差（局所最適解との差）の2乗がほぼ一定の比率で収束 = 収束が速い
- **局所的収束性**をもつが、大域的収束性をもたない → 出発点を解の十分近くに選ばないと、解の収束性が保証されない

- 目的関数の**2次近似**を用いて探索方向を決める

$$f(x_k + d) \cong f_k^0(d) = f(x_k) + \langle \nabla f(x_k), d \rangle + \frac{1}{2} \langle \nabla^2 f(x_k) d, d \rangle$$

**2次近似**

1 次の最適性条件  $\nabla f(\bar{x}) = 0$  により、

$$d_k = -(\nabla^2 f(x_k))^{-1} \nabla f(x_k)$$

しかし、このヘッセ行列が  
正定値かどうか保証されていない  
→ 大域的収束性を持たない

# 準Newton法

## 準Newton法

探索方向： $p_k = -B_k^{-1}\nabla f_k$ ,

- 正定値対称行列 $B_k$ は準ニュートン更新式によって各反復ごとに更新される（たとえばBFGS式）
- **超1次収束**：1次収束と2次収束のあいだ。2次収束に近い。
- **ほぼ大域的収束性**→目的関数が凸関数のときは証明されている

それ以外の場合は証明されていないが、実際は工夫することで可能となる

大域的収束性と効率性を満たす良い方法。よく用いられる。

- 目的関数の**修正2次近似**から探索方向を決める

$$f(x_k + d) \cong f_k^Q(d) = f(x_k) + \langle \nabla f(x_k), d \rangle + \frac{1}{2} \langle B_k d, d \rangle$$

$B_k$ ：適当な正定値対称行列 ←これを**BFGS公式**で更新していく

1次の最適性条件より、探索方向を定める

$$d_k = -B_k^{-1}\nabla f(x_k)$$

### BFGS公式

$$B_{k+1} = B_k - \frac{(B_k s_k)(B_k s_k)^T}{\langle B_k s_k, s_k \rangle} + \frac{y_k s_k^T}{\langle y_k, s_k \rangle}$$

ただし

$$s_k = x_{k+1} - x_k$$

$$y_k = \nabla f_{k+1} - \nabla f_k$$

# ヘッセ行列を修正するニュートン法

ニュートン法では、探索開始点が解から遠い場合、ヘッセ行列 $\nabla^2 f(x)$ が正定値ではないことがありうる

この時ニュートン方向 $p_k^N$ は降下方向ではない可能性がありうる→大域的収束性が保証されない

→ヘッセ行列 $\nabla^2 f(x)$ に正の対角行列、またはfull matrixを加えて修正するというアプローチ

## **Algorithm 3.2** (Line Search Newton with Modification).

Given initial point  $x_0$ ;

**for**  $k = 0, 1, 2, \dots$

Factorize the matrix  $B_k = \nabla^2 f(x_k) + E_k$ , where  $E_k = 0$  if  $\nabla^2 f(x_k)$  is sufficiently positive definite; otherwise,  $E_k$  is chosen to ensure that  $B_k$  is sufficiently positive definite;

Solve  $B_k p_k = -\nabla f(x_k)$ ;

Set  $x_{k+1} \leftarrow x_k + \alpha_k p_k$ , where  $\alpha_k$  satisfies the Wolfe, Goldstein, or Armijo backtracking conditions;

**end**

加える行列 $E_k$ がbounded modified factorization propertyを満たしていれば、これは大域的収束性を持つ

# ステップ長選択アルゴリズム

探索方向 $p_k$ を決めた→ステップ長 $\alpha_k$ を決める  $\phi(\alpha) = f(x_k + \alpha p_k), \quad \alpha > 0$

- 1. ある条件内の**望ましい** $\alpha$ を含む範囲を見つけ、←すでに議論した
- 2. その範囲内で良好な $\alpha$ を選択する

➡ 1の範囲を望ましいステップ長の探索中に減少させ、初期段階で収集された関数および導関数の一部を補間して、最小化子 $\alpha^*$ の位置を推測する

以下では、 $\alpha_k, \alpha_{k-1}$ を、それぞれ最適化アルゴリズムの反復 $k$ および $k-1$ で使用されるステップ長とする。  
一方、直線探索中に生成された試行ステップ長を $\alpha_i, \alpha_{i-1}, \alpha_j$ で表す。推測の初期値を $\alpha_0$ とする。

十分な減少条件  $\phi(\alpha_k) \leq \phi(0) + c_1 \alpha_k \phi'(0),$

## 補間のアルゴリズム

初期値 $\alpha_0$ は所与とする。  $\phi(\alpha_0) \leq \phi(0) + c_1 \alpha_0 \phi'(0),$  を満たすなら探索終了。

満たさないとき、 $[0, \alpha_0]$ 内に望ましいステップ長があるので、入手できる3つの情報 $\phi(0), \phi'(0), \phi(\alpha_0)$ を用いて、

$\phi$ の2次近似 $\phi_q(\alpha)$ を以下のように作る

$$\phi_q(\alpha) = \left( \frac{\phi(\alpha_0) - \phi(0) - \alpha_0 \phi'(0)}{\alpha_0^2} \right) \alpha^2 + \phi'(0)\alpha + \phi(0).$$

これは補間条件

$$\phi_q(0) = \phi(0), \phi'_q(0) = \phi'(0), \text{ and } \phi_q(\alpha_0) = \phi(\alpha_0).$$

を満たすように作られている

新しい試行値 $\alpha_1$ はこの関数を最小化する値で

$$\alpha_1 = -\frac{\phi'(0)\alpha_0^2}{2[\phi(\alpha_0) - \phi(0) - \phi'(0)\alpha_0]}.$$

続く

# ステップ長選択アルゴリズム

## 補間のアルゴリズム

$\alpha_1$ が十分な減少条件を満たすなら探索終了. 満たさないとき入手できる4つの情報 $\phi(0), \phi'(0), \phi(\alpha_0), \phi(\alpha_1)$ を用いて3次近似 $\phi_c(\alpha)$ を以下のように作る

$$\phi_c(\alpha) = a\alpha^3 + b\alpha^2 + \alpha\phi'(0) + \phi(0),$$

where

$$\begin{bmatrix} a \\ b \end{bmatrix} = \frac{1}{\alpha_0^2\alpha_1^2(\alpha_1 - \alpha_0)} \begin{bmatrix} \alpha_0^2 & -\alpha_1^2 \\ -\alpha_0^3 & \alpha_1^3 \end{bmatrix} \begin{bmatrix} \phi(\alpha_1) - \phi(0) - \phi'(0)\alpha_1 \\ \phi(\alpha_0) - \phi(0) - \phi'(0)\alpha_0 \end{bmatrix}$$

$\phi_c(\alpha)$ を微分することによりこれを最小化する値 $\alpha_2$ は区間 $[0, \alpha_1]$ にあり,

$$\alpha_2 = \frac{-b + \sqrt{b^2 - 3a\phi'(0)}}{3a}.$$

必要であればこれを繰り返す.

$\alpha_i$ がひとつ前の $\alpha_{i-1}$ に近すぎるかこれより小さすぎるとき,  $\alpha_i = \alpha_{i-1}/2$ でリセットする.

この手順により, 各反復で $\alpha$ が改善し, 最終的な $\alpha$ が小さすぎないことを保証する.

## 3次補間

区間 $[\bar{a}, \bar{b}]$ に望ましいステップ幅があることが分かっている, この区間内に2つの推定値 $\alpha_{i-1}, \alpha_i$ があることが分かっているなら, 補完情報 $\phi(\alpha_{i-1}), \phi'(\alpha_{i-1}), \phi(\alpha_i), \phi'(\alpha_i)$ を用いて

$$\alpha_{i+1} = \alpha_i - (\alpha_i - \alpha_{i-1}) \left[ \frac{\phi'(\alpha_i) + d_2 - d_1}{\phi'(\alpha_i) - \phi'(\alpha_{i-1}) + 2d_2} \right],$$

with

$$d_1 = \phi'(\alpha_{i-1}) + \phi'(\alpha_i) - 3 \frac{\phi(\alpha_{i-1}) - \phi(\alpha_i)}{\alpha_{i-1} - \alpha_i},$$

$$d_2 = \text{sign}(\alpha_i - \alpha_{i-1}) [d_1^2 - \phi'(\alpha_{i-1})\phi'(\alpha_i)]^{1/2}.$$

で次の推定値をつくれればいい

3次補間は2次収束なので高速であり有用である

# 初期値 $\alpha_0$ の決め方

- Newton法, 準Newton法:  $\alpha_0 = 1$ にすれば高速収束性が保持される
- 最急降下法や共役勾配法→良好な探索方向を持たない→初期値を現在の情報から推測するのがいい

ポピュラーな方法: 反復 $x_k$ での関数の一次変化が前のステップで得られたものと同じであるとする

$$\alpha_0 \nabla f_k^T p_k = \alpha_{k-1} \nabla f_{k-1}^T p_{k-1}, \quad \text{つまり} \quad \alpha_0 = \alpha_{k-1} \frac{\nabla f_{k-1}^T p_{k-1}}{\nabla f_k^T p_k}. \quad \text{とする}$$

もう一つの便利な方法:  $f(x_{k-1}), f(x_k), \nabla f_{k-1}^T p_{k-1}$ を二次補間して,  $\alpha_0$ をそれを最小化する値として定義する

$$\alpha_0 = \frac{2(f_k - f_{k-1})}{\phi'(0)}.$$

# Wolfe条件での直線探索アルゴリズム

## 直線探索アルゴリズム（1段階目）

```
Set  $\alpha_0 \leftarrow 0$ , choose  $\alpha_{\max} > 0$  and  $\alpha_1 \in (0, \alpha_{\max})$ ;  
 $i \leftarrow 1$ ;  
repeat  
  Evaluate  $\phi(\alpha_i)$ ;  
  if  $\phi(\alpha_i) > \phi(0) + c_1\alpha_i\phi'(0)$  or  $[\phi(\alpha_i) \geq \phi(\alpha_{i-1})$  and  $i > 1]$   
     $\alpha_* \leftarrow \text{zoom}(\alpha_{i-1}, \alpha_i)$  and stop;  
  Evaluate  $\phi'(\alpha_i)$ ;  
  if  $|\phi'(\alpha_i)| \leq -c_2\phi'(0)$   
    set  $\alpha_* \leftarrow \alpha_i$  and stop;  
  if  $\phi'(\alpha_i) \geq 0$   
    set  $\alpha_* \leftarrow \text{zoom}(\alpha_i, \alpha_{i-1})$  and stop;  
  Choose  $\alpha_{i+1} \in (\alpha_i, \alpha_{\max})$ ;  
   $i \leftarrow i + 1$ ;  
end (repeat)
```

$\alpha_{\max}$  : ユーザ設定の許容最大ステップ長

望ましいステップ長を含むまで  
探索範囲を拡大する段階

## 関数zoomのアルゴリズム（2段階目）

```
repeat  
  Interpolate (using quadratic, cubic, or bisection) to find  
    a trial step length  $\alpha_j$  between  $\alpha_{lo}$  and  $\alpha_{hi}$ ;  
  Evaluate  $\phi(\alpha_j)$ ;  
  if  $\phi(\alpha_j) > \phi(0) + c_1\alpha_j\phi'(0)$  or  $\phi(\alpha_j) \geq \phi(\alpha_{lo})$   
     $\alpha_{hi} \leftarrow \alpha_j$ ;  
  else  
    Evaluate  $\phi'(\alpha_j)$ ;  
    if  $|\phi'(\alpha_j)| \leq -c_2\phi'(0)$   
      Set  $\alpha_* \leftarrow \alpha_j$  and stop;  
    if  $\phi'(\alpha_j)(\alpha_{hi} - \alpha_{lo}) \geq 0$   
       $\alpha_{hi} \leftarrow \alpha_{lo}$ ;  
     $\alpha_{lo} \leftarrow \alpha_j$ ;  
end (repeat)
```

望ましいステップ長を含んだ状態で範囲を縮小して  
いき、望ましいステップ長に収束させる段階

<b>4</b>	<b>Trust-Region Methods</b>	<b>66</b>
	Outline of the Trust-Region Approach . . . . .	68
4.1	Algorithms Based on the Cauchy Point . . . . .	71
	The Cauchy Point . . . . .	71
	Improving on the Cauchy Point . . . . .	73
	The Dogleg Method . . . . .	73
	Two-Dimensional Subspace Minimization . . . . .	76
4.2	Global Convergence . . . . .	77
	Reduction Obtained by the Cauchy Point . . . . .	77
	Convergence to Stationary Points . . . . .	79
4.3	Iterative Solution of the Subproblem . . . . .	83
	The Hard Case . . . . .	87
	Proof of Theorem 4.1 . . . . .	89
	Convergence of Algorithms Based on Nearly Exact Solutions . . . . .	91
4.4	Local Convergence of Trust-Region Newton Methods . . . . .	92
4.5	Other Enhancements . . . . .	95
	Scaling . . . . .	95
	Trust Regions in Other Norms . . . . .	97
	Notes and References . . . . .	98

# 信頼領域法

**信頼領域法**：現在の点の周りで、近似が妥当な範囲内（信頼領域）で目的関数を二次式でモデル化し、その最小点を求めることを繰り返す

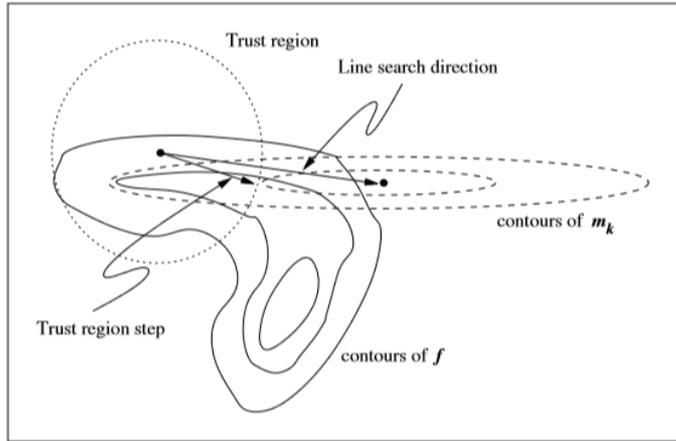


Figure 4.1 Trust-region and line search steps.

利点：直線探索法では $x^*$ に効率的に収束しないような方向に探索してしまうときにより良い探索方向に向かえる

当然ながら、適切な信頼領域の大きさを選ぶことが重要！

目的関数 $f$ の2次モデル $m_k(p)$ を用いた問題の定式化

Taylor展開  $f(x_k + p) = f_k + g_k^T p + \frac{1}{2} p^T \nabla^2 f(x_k + tp) p, \longrightarrow m_k(p) = f_k + g_k^T p + \frac{1}{2} p^T B_k p,$

$m_k(p)$ と $f(x_k + p)$ の誤差 $= O(\|p\|^2)$ だから、 $p$ が小さいほどこの近似は正確

定式化  $\min_{p \in \mathbb{R}^n} m_k(p) = f_k + g_k^T p + \frac{1}{2} p^T B_k p \quad \text{s.t. } \|p\| \leq \Delta_k,$

$\Delta_k$ ：信頼領域半径

信頼領域半径内での $m_k$ を最小化する解 $p^*$

# 信頼領域法の概要

各段階における信頼領域半径 $\Delta_k$ の決定法

→ ひとつ前の段階でのモデル $m_k$ の目的関数 $f$ への当てはまりに基づいて決める

指標:  $\rho_k = \frac{f(x_k) - f(x_k + p_k)}{m_k(0) - m_k(p_k)}$ ; (4.4)

← 実際の目的関数値の減少

← 予測減少 (モデル式の値の減少) : 負値になることはない (増加はしない)

- $\rho_k < 0, \rho_k \approx 0$  :  $f(x_k + p_k) > f(x_k)$ , そのstepは失敗, 次のstepは信頼領域を狭める
- $\rho_k \approx 1$  :  $m_k$ は $f$ の良い近似, 次のstepは信頼領域を広げていい
- $0 < \rho_k \ll 1$  : まずまずの近似, 信頼領域はそのまま次のstepへ

## 信頼領域半径の決定アルゴリズム

**Algorithm 4.1** (Trust Region).

Given  $\hat{\Delta} > 0$ ,  $\Delta_0 \in (0, \hat{\Delta})$ , and  $\eta \in [0, \frac{1}{4}]$ :

for  $k = 0, 1, 2, \dots$

Obtain  $p_k$  by (approximately) solving (4.3);

Evaluate  $\rho_k$  from (4.4);

if  $\rho_k < \frac{1}{4}$

$$\Delta_{k+1} = \frac{1}{4} \Delta_k$$

else

if  $\rho_k > \frac{3}{4}$  and  $\|p_k\| = \Delta_k$

$$\Delta_{k+1} = \min(2\Delta_k, \hat{\Delta})$$

else

$$\Delta_{k+1} = \Delta_k;$$

if  $\rho_k > \eta$

$$x_{k+1} = x_k + p_k$$

else

$$x_{k+1} = x_k;$$

end (for).

$$(4.3) \min_{p \in \mathbb{R}^n} m_k(p) = f_k + g_k^T p + \frac{1}{2} p^T B_k p \quad \text{s.t. } \|p\| \leq \Delta_k,$$

← このアルゴリズムを実際に使うには, 信頼領域内の最適化問題 (4.3)を解かなければならない

ひとまず $k$ を省略して書き直す

$$\min_{p \in \mathbb{R}^n} m(p) \stackrel{\text{def}}{=} f + g^T p + \frac{1}{2} p^T B p \quad \text{s.t. } \|p\| \leq \Delta.$$

この最適化問題の解の特徴を考えていきたい (必ずしも厳密に解けないので)

# 信頼領域法の概要

$$\min_{p \in \mathbb{R}^n} m(p) \stackrel{\text{def}}{=} f + g^T p + \frac{1}{2} p^T B p \quad \text{s.t. } \|p\| \leq \Delta.$$

**定理：** 上の部分問題の大域解ベクトル  $p^*$  が存在するとき、ある定数  $\lambda \geq 0$  が存在して、以下の条件を満たす；

$$(B + \lambda I)p^* = -g, \quad (1)$$

$$\lambda(\Delta - \|p^*\|) = 0, \quad (2)$$

$$(B + \lambda I) \text{ is positive semidefinite.} \quad (3)$$

半正定値

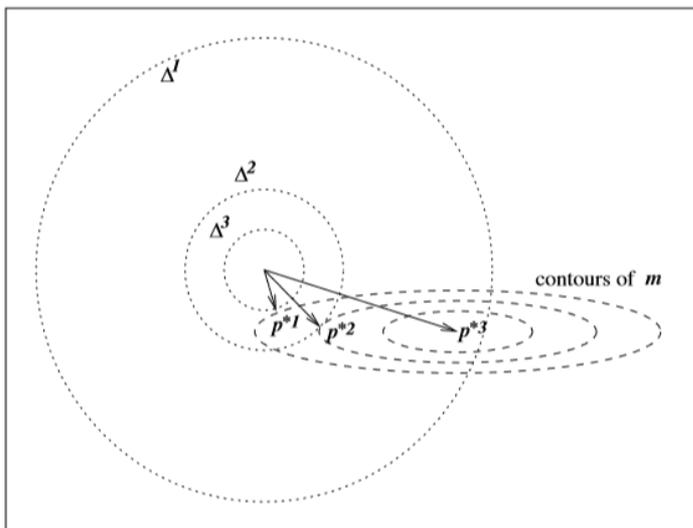


Figure 4.2 Solution of trust-region subproblem for different radii  $\Delta^1, \Delta^2, \Delta^3$ .

## 意味

- (2)は相補性条件： $\lambda$ か $(\Delta - \|p^*\|)$ のどちらかは0
- 解が信頼領域内にある時( $\Delta^1$ ),  $\lambda = 0$ で  $Bp^* = -g$  で  $B$ は半正定値
- 解が信頼領域内にない時( $\Delta^2, \Delta^3$ ),  $\|p^*\| = \Delta$  で  $\lambda > 0$  となり, (1)より

$$\lambda p^* = -Bp^* - g = -\nabla m(p^*).$$

つまり,  $\lambda > 0$  なら解  $p^*$  は  $m$  の負の勾配と同一直線状で, 等高線に垂直

解ベクトル  $p^*$  の特徴が把握できる

# 信頼領域法での主な議論

1. 信頼領域内でのモデル式  $(4.3) \min_{p \in \mathbb{R}^n} m_k(p) = f_k + g_k^T p + \frac{1}{2} p^T B_k p \quad \text{s.t. } \|p\| \leq \Delta_k,$  の最小化の近似解法
2. 信頼領域法の大域的収束性
3. (4.3)の最適化問題の解が満たす条件  $(B + \lambda I)p^* = -g,$  における $\lambda$ を推定する方法 (反復法)  
→ 解ベクトル $p^*$ の特徴が把握できる
4. モデル関数のヘッセ行列 $B_k$ が目的関数のヘッセ行列 $\nabla^2 f(x_k)$ に等しいときに用いられる信頼領域Newton法についての議論  
→ 信頼領域Newton法により, 2階の十分条件を満たす点 $x^*$ に収束するとき, 超一次収束する

1. 信頼領域内でのモデル式

$$(4.3) \min_{p \in \mathbb{R}^n} m_k(p) = f_k + g_k^T p + \frac{1}{2} p^T B_k p \quad \text{s.t. } \|p\| \leq \Delta_k,$$

の最小化の近似解法

直線探索法で**十分な減少**の条件 (Wolfe条件等) は実際はかなり緩かった。

→ 信頼領域法でも領域内の近似解 $p_k$ を見つけて、モデル内で十分な減少をすれば大域的収束性は十分満たせる

## 十分な減少の定義：コーシー点

(4.3)の線形バージョンの解ベクトル $p_k^s$ を見つける

$$p_k^s = \arg \min_{p \in \mathbb{R}^n} f_k + g_k^T p \quad \text{s.t. } \|p\| \leq \Delta_k;$$

信頼領域を満たすことを境界条件として、 $m_k(\tau p_k^s)$ を最小化するスカラー $\tau_k > 0$ を計算

$$\tau_k = \arg \min_{\tau \geq 0} m_k(\tau p_k^s) \quad \text{s.t. } \|\tau p_k^s\| \leq \Delta_k;$$

コーシー点を次のように定義  $p_k^c = \tau_k p_k^s$

これはClosed-formに書き下せる

$$p_k^c = -\tau_k \frac{\Delta_k}{\|g_k\|} g_k,$$

$$\text{ただし, } \tau_k = \begin{cases} 1 & \text{if } g_k^T B_k g_k \leq 0; \\ \min(\|g_k\|^3 / (\Delta_k g_k^T B_k g_k), 1) & \text{otherwise.} \end{cases}$$

→つまりコーシー点の計算は負荷が低い

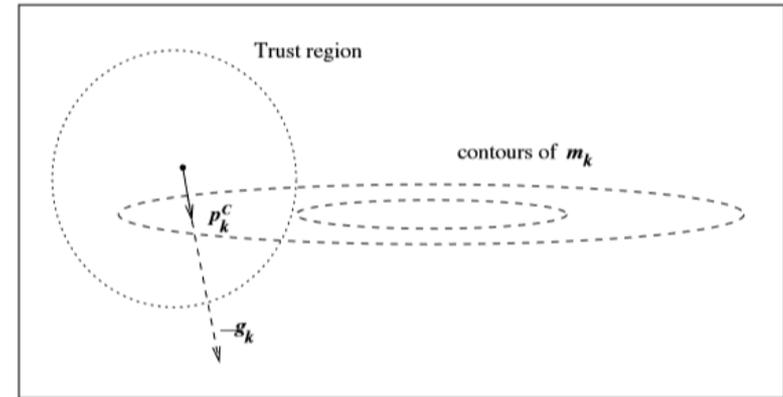


Figure 4.3 The Cauchy point.

信頼領域法は、ステップ $p_k$ が**Cauchy**ステップによって表される減少の少なくともいくつかの正の倍数でモデル $m_k$ を減少させれば、大域的収束性を持つ。

$$(4.3) \min_{p \in \mathbb{R}^n} m_k(p) = f_k + g_k^T p + \frac{1}{2} p^T B_k p \quad \text{s.t. } \|p\| \leq \Delta_k$$

コーシー点がモデル内で大域的収束性を持ち、計算負荷が低いならそれでいいのでは？

→ 信頼領域法で毎回Cauchyステップを取るの、ある一つのステップ長で最急降下法を行うのと等価

つまり遅い……

速くする方法の例

➤  $B_k$ が正定値で  $\|p_k^B\| \leq \Delta_k$  なら  $p_k^B = -B_k^{-1} g_k$  にする

→  $B_k$ が正確なヘシアン  $\nabla^2 f(x_k)$  または準ニュートン近似なら、超一次収束する。

$$(4.3) \min_{p \in \mathbb{R}^n} m_k(p) = f_k + g_k^T p + \frac{1}{2} p^T B_k p \quad \text{s.t. } \|p\| \leq \Delta_k$$

のもっと良い近似解の見つけ方はないか

以降、一回の反復内のみを考えて、添え字kは省略

$$(4.5) \min_{p \in \mathbb{R}^n} m(p) \stackrel{\text{def}}{=} f + g^T p + \frac{1}{2} p^T B p \quad \text{s.t. } \|p\| \leq \Delta$$

• (4.5)の解を  $p^*(\Delta)$  とする ( $\Delta$  に強く関連することを強調するため)

## ■ Dogleg法

■ 二次元部分空間最小化の2つを紹介する

# Dogleg法

$$(4.5) \min_{p \in \mathbb{R}^n} m(p) \stackrel{\text{def}}{=} f + g^T p + \frac{1}{2} p^T B p \quad \text{s.t. } \|p\| \leq \Delta. \quad \text{の近似解の見つけ方}$$

- $B$ が正定値のとき使える手法  $\longrightarrow$   $m$ の無制約最小化子は,  $p^B = -B^{-1}g$

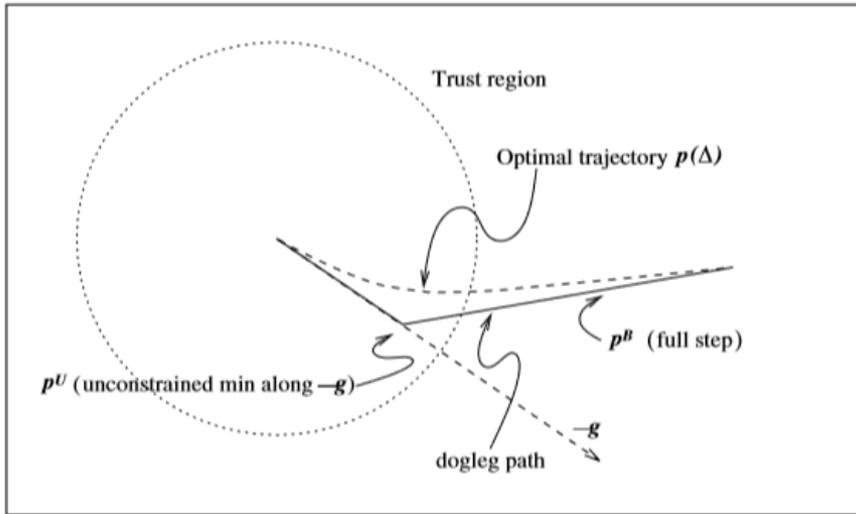


Figure 4.4 Exact trajectory and dogleg approximation.

最適軌跡 $p(\Delta)$ を, 最急降下方向に沿った軌跡 $p^U$ と $p^B$ で近似する

$$p^U = -\frac{g^T g}{g^T B g} g,$$

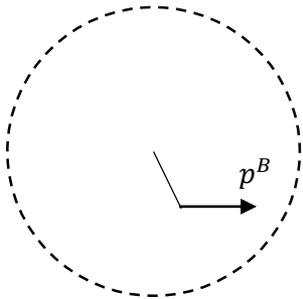
つまり, この近似軌道は

$$\tilde{p}(\tau) = \begin{cases} \tau p^U, & 0 \leq \tau \leq 1, \\ p^U + (\tau - 1)(p^B - p^U), & 1 \leq \tau \leq 2. \end{cases}$$

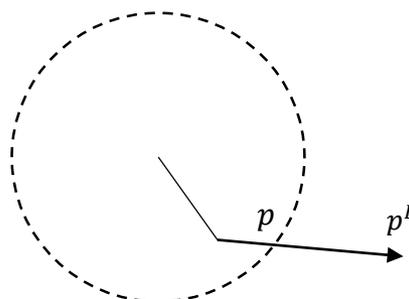
**Dogleg法**: 経路 $\tilde{p}(\tau)$ に沿って信頼領域内でモデル $m$ を最小化する $p$ を見つける

ここで,  $\tau$ について  $\tilde{p}(\tau)$ は増加関数で $m(\tilde{p}(\tau))$ は減少関数であることを示すことができる (証明略) すなわち,

$\|p^B\| \leq \Delta$ のとき  $p = p^B$



$\|p^B\| \geq \Delta$ のとき  $p$ は信頼領域境界との交点



# 二次元部分空間最小化

$B$ が正定値のときdogleg法の探索範囲を $p^U$ と $p^B$ にまたがる二次元部分空間全体に広げれば、より洗練された手法となる

$$\text{つまり } \min_p m(p) = f + g^T p + \frac{1}{2} p^T B p \quad \text{s.t. } \|p\| \leq \Delta, p \in \text{span}[g, B^{-1}g].$$

これに対してコーシー点が計算できるのでこの補問題の最適解は少なくともコーシー点と同じくらいの効率で、大域的に収束する。



この手法を $B$ が正定値でない場合にも使えるように拡張する

二次元部分空間を以下のように修正する

$$\text{span}[g, (B + \alpha I)^{-1}g], \quad \text{for some } \alpha \in (-\lambda_1, -2\lambda_1],$$

ここで、 $\lambda_1$ は $B$ の最も負値な固有値とすると、 $B + \alpha I$ が正定値になることを保証できる

## 2. 信頼領域法の大域的収束性

# コーシー点により達成される減少

コーシー点により達成されるモデル $m$ の減少値の推定値を得て、これを用いて、

**Algorithm 4.1** (Trust Region).

Given  $\hat{\Delta} > 0$ ,  $\Delta_0 \in (0, \hat{\Delta})$ , and  $\eta \in [0, \frac{1}{4}]$ :

for  $k = 0, 1, 2, \dots$

Obtain  $p_k$  by (approximately) solving (4.3);

Evaluate  $\rho_k$  from (4.4);

if  $\rho_k < \frac{1}{4}$

$$\Delta_{k+1} = \frac{1}{4} \Delta_k$$

else

if  $\rho_k > \frac{3}{4}$  and  $\|p_k\| = \Delta_k$

$$\Delta_{k+1} = \min(2\Delta_k, \hat{\Delta})$$

else

$$\Delta_{k+1} = \Delta_k;$$

if  $\rho_k > \eta$

$$x_{k+1} = x_k + p_k$$

else

$$x_{k+1} = x_k;$$

end (for).

左のアルゴリズムで生成される勾配の点列 $\{g_k\}$ が累積的に0に集積する点を有し、 $\eta$ が厳密に正のとき0に収束することを示す。

3.  $(B + \lambda I)p^* = -g$ , における $\lambda$ を推定する方法 (反復法)

# モデル式の最適化の反復法による解き方

$$\min_{p \in \mathbb{R}^n} m(p) \stackrel{\text{def}}{=} f + g^T p + \frac{1}{2} p^T B p \quad \text{s.t. } \|p\| \leq \Delta.$$

**定理：** 上の部分問題の大域解ベクトル  $p^*$  が存在するとき、ある定数  $\lambda \geq 0$  が存在して、以下の条件を満たす；

$$(B + \lambda I)p^* = -g, \quad (1)$$

$$\lambda(\Delta - \|p^*\|) = 0, \quad (2)$$

$$(B + \lambda I) \text{ is positive semidefinite.} \quad (3)$$

半正定値

この  $\lambda$  は  $p^*$  の特徴を把握するのに重要であることを最初に述べた

**Algorithm 4.1** (Trust Region).

Given  $\hat{\Delta} > 0$ ,  $\Delta_0 \in (0, \hat{\Delta})$ , and  $\eta \in [0, \frac{1}{4}]$ :

for  $k = 0, 1, 2, \dots$

    Obtain  $p_k$  by (approximately) solving (4.3);

    Evaluate  $\rho_k$  from (4.4);

    if  $\rho_k < \frac{1}{4}$

$$\Delta_{k+1} = \frac{1}{4} \Delta_k$$

    else

        if  $\rho_k > \frac{3}{4}$  and  $\|p_k\| = \Delta_k$

$$\Delta_{k+1} = \min(2\Delta_k, \hat{\Delta})$$

        else

$$\Delta_{k+1} = \Delta_k;$$

    if  $\rho_k > \eta$

$$x_{k+1} = x_k + p_k$$

    else

$$x_{k+1} = x_k;$$

end (for).

←このアルゴリズムはモデル内で正確な最適解を見つけようとするものではない  
(近似解法)

問題が比較的小さい（反復回数が大きくなりすぎない）なら、モデル式の近似解を  
よりよくすることを目指すのは価値がある



$\lambda$  の値を推定してより  $p^*$  に近い解を求めることを目指す

具体的には上の定理による  $p^*$  の特徴づけと、Newton法を用いる

# モデル式の解 $p$ を見つけるアルゴリズム

$$\min_{p \in \mathbb{R}^n} m(p) \stackrel{\text{def}}{=} f + g^T p + \frac{1}{2} p^T B p \quad \text{s.t. } \|p\| \leq \Delta.$$

**定理**：上の部分問題の大域解ベクトル  $p^*$  が存在するとき、ある定数  $\lambda \geq 0$  が存在して、以下の条件を満たす；

$$(B + \lambda I)p^* = -g, \quad (1)$$

$$\lambda(\Delta - \|p^*\|) = 0, \quad (2)$$

$$(B + \lambda I) \quad \text{is positive semidefinite.} \quad (3)$$

半正定値

**問題**：  $\lambda = 0$  が  $\|p\| \leq \Delta$  で (1), (3) を満たすか、そうでないときは  $B + \lambda I$  が正定値になるような十分大きい  $\lambda$  について (1) より

$$p(\lambda) = -(B + \lambda I)^{-1} g \quad \text{を定義して,}$$

$$\|p(\lambda)\| = \Delta \quad \text{を満たす } \lambda > 0 \text{ を探す. } \quad ((2) \text{ における変数 } \lambda \text{ の一次元の根探索問題})$$

望ましい  $\lambda$  が存在するかを見るために、 **$B$  の固有値組成を見て、 $\|p(\lambda)\|$  の特性を調べる。**

$B$  は対称行列だから、直交行列  $Q$  が存在して、 $B = Q\Lambda Q^T$  となる対角行列  $\Lambda$  が存在する。  $\Lambda = \text{diag}(\lambda_1, \lambda_2, \dots, \lambda_n)$

この時、 $B + \lambda I = Q(\Lambda + \lambda I)Q^T$  で、 $\lambda \neq \lambda_j$  のとき、

$$p(\lambda) = -Q(\Lambda + \lambda I)^{-1} Q^T g = -\sum_{j=1}^n \frac{q_j^T g}{\lambda_j + \lambda} q_j \quad q_j \text{ は行列 } Q \text{ の } j \text{ 列}$$

$q_1, q_2, \dots, q_n$  の直交性から

$$\|p(\lambda)\|^2 = \sum_{j=1}^n \frac{(q_j^T g)^2}{(\lambda_j + \lambda)^2}.$$

$\|p(\lambda)\|$  の特性をいろいろ示唆してくれる

# $\|p(\lambda)\|$ の特性

$$\|p(\lambda)\|^2 = \sum_{j=1}^n \frac{(q_j^T g)^2}{(\lambda_j + \lambda)^2}$$

- $\lambda > -\lambda_1$  のとき,

すべての  $j = 1, 2, \dots, n$  について  $\lambda_j + \lambda > 0$  より,  $\|p(\lambda)\|$  は  $\lambda$  について  $(-\lambda_1, \infty)$  で連続で増加しない関数で,

$$\lim_{\lambda \rightarrow \infty} \|p(\lambda)\| = 0.$$

$q_j^T g \neq 0$  なら,

$$\lim_{\lambda \rightarrow -\lambda_j} \|p(\lambda)\| = \infty$$

特に  $q_1^T g \neq 0$  のとき,  $\|p(\lambda^*)\| = \Delta$  となるような  $\lambda^* \in (-\lambda_1, \infty)$  がある



まず  $q_1^T g \neq 0$  のときに,  $\lambda^*$  を特定する手順を考える

- $B$  が正定値で  $\|B^{-1}g\| \leq \Delta$  なら,  $\lambda^* = 0$  で問題なし
- $B$  が正定値ではないとき, Newton の根探索法を使って

$$\phi_1(\lambda) = \|p(\lambda)\| - \Delta = 0.$$

を解く  $\lambda > -\lambda_1$  を求める.

この手法の問題点:  $\lambda$  が  $-\lambda_1$  より大きいが,  $-\lambda_1$  に非常に近いときに  $\|p(\lambda)\|$  の形を考えると,

$$\phi_1(\lambda) \approx \frac{C_1}{\lambda + \lambda_1} + C_2, \quad \text{で近似できる } (C_1 > 0, C_2 \text{ は定数})$$



非線形  $\rightarrow$  Newton法が信頼できない or 遅い

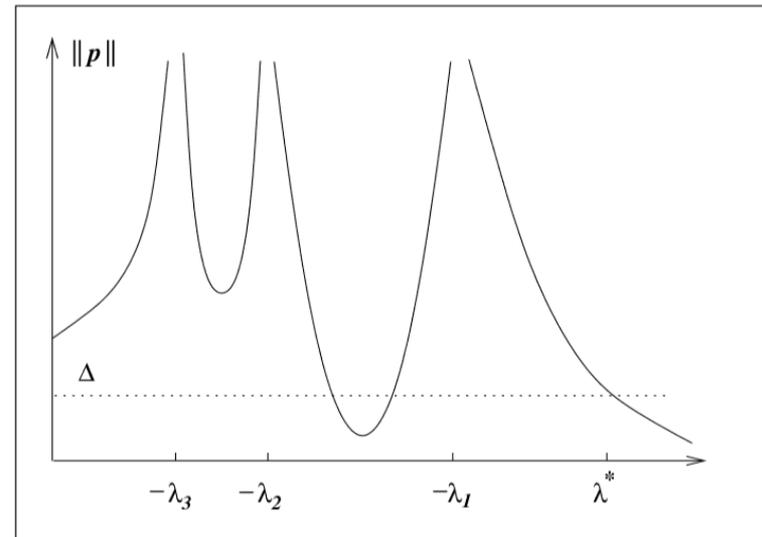


Figure 4.5  $\|p(\lambda)\|$  as a function of  $\lambda$ .

# $\|p(\lambda)\|$ の特性

$$\phi_1(\lambda) = \|p(\lambda)\| - \Delta = 0.$$

これを解 $\lambda$ の周りでほぼ線形とみなして

$$\phi_2(\lambda) = \frac{1}{\Delta} - \frac{1}{\|p(\lambda)\|}, \quad \text{と近似する}$$

この関数で $\lambda$ が $-\lambda_1$ より大きいが、 $-\lambda_1$ に非常に近いときに $\|p(\lambda)\|$ の形を考えると、

$$\phi_2(\lambda) \approx \frac{1}{\Delta} - \frac{\lambda + \lambda_1}{C_3} \quad (C_3 > 0 \text{ は定数})$$

これなら $\phi_2$ は $-\lambda_1$ の周りで線形で、Newton法を使える

$$\text{反復: } \lambda^{(\ell+1)} = \lambda^{(\ell)} - \frac{\phi_2(\lambda^{(\ell)})}{\phi_2'(\lambda^{(\ell)})}.$$

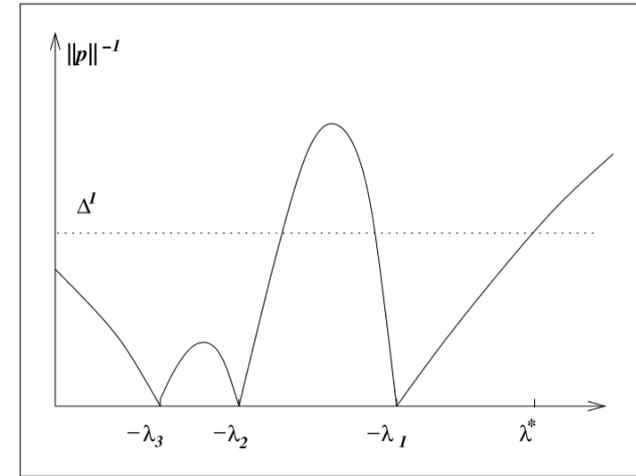


Figure 4.6  $1/\|p(\lambda)\|$  as a function of  $\lambda$ .

## 信頼領域Newton法 (信頼領域法の部分問題の解法アルゴリズム)

### Algorithm 4.3 (Trust Region Subproblem).

Given  $\lambda^{(0)}$ ,  $\Delta > 0$ :

for  $\ell = 0, 1, 2, \dots$

Factor  $B + \lambda^{(\ell)}I = R^T R$ ;

Solve  $R^T R p_\ell = -g$ ,  $R^T q_\ell = p_\ell$ ;

Set

$$\lambda^{(\ell+1)} = \lambda^{(\ell)} + \left( \frac{\|p_\ell\|}{\|q_\ell\|} \right)^2 \left( \frac{\|p_\ell\| - \Delta}{\Delta} \right);$$

end (for).

# 難しいケース

•  $q_1^T g = 0$  のとき

→  $\lambda^* \in (-\lambda_1, \infty)$  の範囲外で  $\|p(\lambda^*)\| = \Delta$  となるような  $\lambda^*$  がある

→ 前述の方法が適用できない

しかし,  $(B + \lambda I)p^* = -g,$  (1)

$\lambda(\Delta - \|p^*\|) = 0,$  (2)

$(B + \lambda I)$  is positive semidefinite. (3)  
半正定値

この定理から,  $\lambda \in [-\lambda_1, -\infty)$  であることを導ける



$\lambda = -\lambda_1$  しかありえない

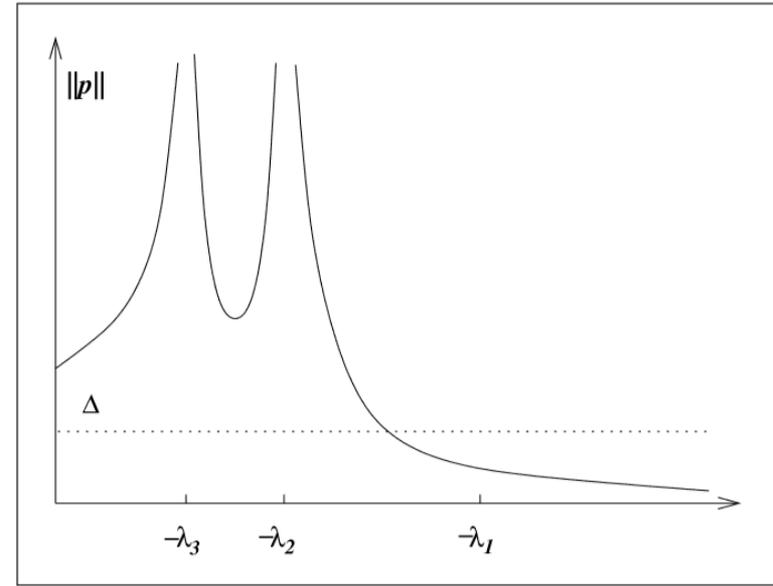


Figure 4.7 The hard case:  $\|p(\lambda)\| < \Delta$  for all  $\lambda \in (-\lambda_1, \infty)$ .

$p(\lambda)$ を見つけるには,  $(B - \lambda_1 I)$ が非正則行列 (行列式が0) であることに着目して  $(B - \lambda_1 I)z = 0$ となる単位ベクトル $z$ があることを考慮すれば,

$$p = \sum_{j:\lambda_j \neq \lambda_1} \frac{q_j^T g}{\lambda_j + \lambda} q_j + \tau z \quad \tau \text{は任意のスカラー}$$

$$\|p\|^2 = \sum_{j:\lambda_j \neq \lambda_1} \frac{(q_j^T g)^2}{(\lambda_j + \lambda)^2} + \tau^2,$$

このように表せば  $\|p(\lambda)\|$  の特性を見ることができる

4. モデル関数のヘッセ行列 $B_k$ が目的関数のヘッセ行列 $\nabla^2 f(x_k)$ に等しいときに用いられる信頼領域Newton法についての議論  
→ 信頼領域Newton法により, 2階の十分条件を満たす点 $x^*$ に収束するとき, 超一次収束する

# 信頼領域法まとめ

- ・流れは信頼領域内のモデル式で解ベクトル $p^*$ を求める→次の信頼領域半径を決めるの繰り返し
- ・ $p^*$ の求め方に、近似的だが負荷が小さいコーシー点を改善していく方法と、問題が比較的小さいときに使えて、正確に $p^*$ を求められる信頼領域Newton法がある

よくわからなかったこと

- ・なぜ大域的収束性や、収束の速さが保証されるかの証明
- ・結局信頼領域法は、直線探索法に比べてどういうメリット・デメリットがあるのか  
(信頼領域法は反復数が少なくなるが、1ステップの計算時間は長い傾向にあり、直線探索法は反復数が多くなるが、1ステップの計算時間は短い傾向にあるらしい)  
(目的関数や近似式のヘッセ行列の形 (正定値か半正定値か非正定値か) によって有利な手法が異なってくる)