

制約なし最小化問題に対する微分を用いない解法

高橋 優輝

森田研究室 B4

2022.12.08

微分不可能な問題に対する最適化

モチベーション

- ・ 微分不可能な問題に対しては，勾配を用いる方法が使えない．
- ・ 理論的には微分可能であっても，その値の評価が数値計算やシミュレーションで行なわれる場合には，目的関数の正確な微分値を求められない．

目次

- ① 8.1 微分不可能な最小化問題に対する直線探索法
- ② 8.2 凸計画問題に対する劣勾配を用いた手法
- ③ 大域的収束性が保証された微分を用いない最適化法
 - 直接探索法
 - simplex gradient を用いた手法

方向微分 (復習)

この節では、 $f: \mathbb{R}^n \rightarrow \mathbb{R}$ とする。

6 章では、降下方向を勾配 $\nabla f(x)$ を用いて定義した。直線探索法において、 f が微分可能でないときには、方向微分を用いて、降下方向を定義する。

定義：方向微分可能

方向 $d \in \mathbb{R}^n$ に対して、極限

$$\lim_{t \rightarrow +0} \frac{f(x + td) - f(x)}{t}$$

が存在するとき、関数 f は点 x において、 d 方向に方向微分可能であるという。

方向微係数

定義：方向微係数

関数 f の点 x における d 方向の方向微係数 $f'(x; d)$ を

$$f'(x; d) := \lim_{t \rightarrow +0} \frac{f(x + td) - f(x)}{t}$$

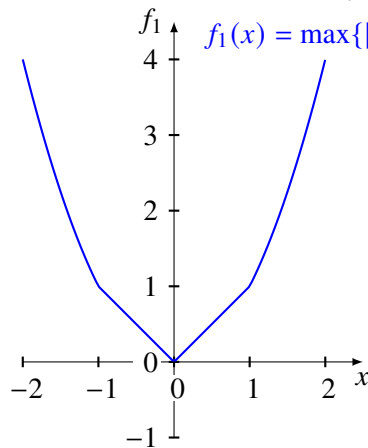
と定義する.

f が微分可能なとき, $f'(x; d) = \nabla f(x)^\top d$ が成り立つ.

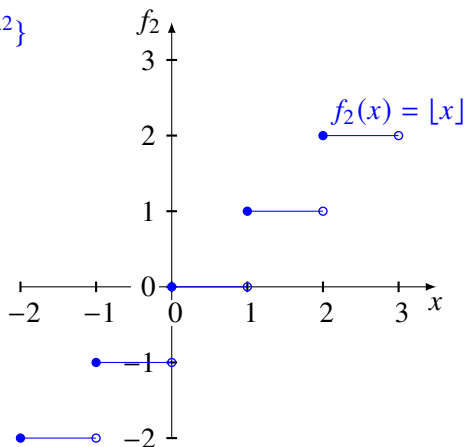
($\because f(x + td) = f(x) + t\nabla f(x)^\top d + o(t)$)

例 (方向微係数)

$$f'(x; d) := \lim_{t \rightarrow +0} \frac{f(x + td) - f(x)}{t}$$



$$\begin{aligned} f'_1(1; 1) &= 2, f'_1(1; -1) = -1 \\ f'_1(0, 1) &= 1, f'_1(0, -1) = -1 \end{aligned}$$



$$\begin{aligned} f'_2(1; 1) &= 0, f'_2(1; -1) = -\infty (\text{不可}) \end{aligned}$$

方向微分を用いた制約なし最小化問題に対する最適性の1次の必要条件

定理 8.1: 方向微分を用いた制約なし最小化問題に対する最適性の1次の必要条件

x^* が, 目的関数 f の制約なし最小化問題の局所的最適解

$$\implies f'(x^*; d) \geq 0, \forall d \in \mathbb{R}^n$$

proof.

x^* が f の局所的最適解であることに注意して, 方向微分の定義から導かれる

$$f(x + td) = f(x) + tf'(x; d) + o(t)$$

について, 両辺 $t(> 0)$ で割った後に, $t \rightarrow +0$ とすれば良い. □

定義: 停留点

$f'(x; d) \geq 0, \forall d \in \mathbb{R}^n$ を満たす点 x を f の停留点と呼ぶ.

→ 3 章の停留点^{*1}の一般化.

^{*1} x^* が目的関数 f の制約なし最小化問題の局所的最適解 $\implies \nabla f(x^*) = 0$.

降下方向

定義:降下方向

点 x において, $f'(x; d) < 0$ が成り立つ方向 $d \in \mathbb{R}^n$ を f の x における降下方向と呼ぶ.

→6章の降下方向^{*2}の一般化.

実際, $d \in \mathbb{R}^n$ は, $f'(x; d) < 0$ を満たすとする, 方向微分の定義より,

$$f(x + td) = f(x) + tf'(x; d) + o(t)$$

となるので, t が十分に小さければ, $f(x + td) < f(x)$ が成り立つので, 降下方向に進むことで, 目的関数値を減らすことができる.

^{*2} $\nabla f(x)^\top d < 0$ が成り立つ方向 d を f の x における降下方向と呼ぶ.

方向微分を用いた直線探索法

以上の議論から，6章の直線探索法を一般化できる．

直線探索法

step 0: 初期点 x^0 を選び， $k := 0$ とする．

step 1: 降下方向 d^k を定める．降下方向が無ければ，停留点なので，終了する．

step 2: アルミホのルール

$$\rho \in (0, 1), f(x^k + t_k d^k) - f(x^k) \leq \rho t_k f'(x; d^k)$$

を満たすステップ幅 t_k を求める．

step 3: $x^{k+1} := x^k + t_k d^k$, $k := k + 1$ として，step 1 へ．

方向微分を用いた直線探索法

任意の k に対して,

$$f'(x; d^k) < -c \|d^k\|^2$$

となるような正の定数 c が存在すれば, 適当な仮定の下で, 直線探索法によって生成された点列 $\{x^k\}$ が停留点に収束することが知られている. 一方で, 一般の微分不可能な関数 f に対して, step 1 での, 降下方向を求めることは容易ではない.

目次

- ① 8.1 微分不可能な最小化問題に対する直線探索法
- ② 8.2 凸計画問題に対する劣勾配を用いた手法
- ③ 大域的収束性が保証された微分を用いない最適化法
 - 直接探索法
 - simplex gradient を用いた手法

8.2 凸計画問題に対する劣勾配を用いた手法

本節では、 $f : \mathbb{R}^n \rightarrow \mathbb{R}$ を微分不可能な凸関数であるとする。
微分不可能な目的関数の (制約なし) 最小化問題を解く手法として、劣勾配法、切除平面法、bundle 法を紹介する。

劣微分と劣勾配 (復習)

定義：劣微分と劣勾配

$$\partial f(x) := \{\eta \in \mathbb{R}^n \mid f(y) - f(x) \geq \eta^\top (y - x), \forall y \in \mathbb{R}^n\}$$

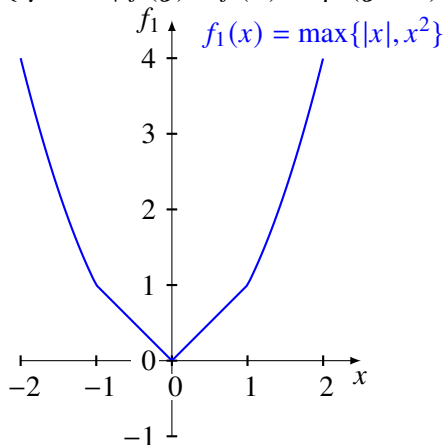
を凸関数 f の x における劣微分と呼ぶ。また, $\eta \in \partial f(x)$ を f の x における劣勾配と呼ぶ。

f が微分可能であれば, $\partial f(x) = \{\nabla f(x)\}$ である。

→ 劣微分は凸関数の微分の一般化。

例 (劣微分)

$$\partial f(x) := \{\eta \in \mathbb{R}^n \mid f(y) - f(x) \geq \eta^\top (y - x), \forall y \in \mathbb{R}^n\}$$



$$\begin{aligned}\partial f(-2) &= \{-4\}, \partial f(-1) = \{k \mid -2 \leq k \leq -1\}, \\ \partial f(0) &= \{k \mid -1 \leq k \leq 1\}, \partial f(0.5) = \{1\}, \partial f(1) = \{k \mid 1 \leq k \leq 2\}\end{aligned}$$

劣勾配法

最急降下法では勾配を用いていた．勾配の代わりに劣勾配を用いるのが，劣勾配法である．

劣勾配法

step 0: $\mathbf{x}^0 \in \mathbb{R}^n$ を選び, $k := 0$ とする．

step 1: 劣勾配 $\boldsymbol{\eta}^k \in \partial f(\mathbf{x}^k)$ を求める．

step 2: \mathbf{x}^k が終了条件を満たしていれば終了する．

step 3: $\mathbf{x}^{k+1} := \mathbf{x}^k - t_k \frac{\boldsymbol{\eta}^k}{\|\boldsymbol{\eta}^k\|}$ とする．

step 4: $k := k + 1$ として, step 1 へ．

step 1 で求めた劣勾配の逆方向 $-\boldsymbol{\eta}^k$ は f の降下方向になるとは限らない．そのため，目的関数を減少させるステップ幅 t_k を定めることができないことがある．そこで，ステップ幅を， $\alpha \geq 0, \beta > 0$ を用いて，

$$t_k = \frac{1}{\alpha + \beta k}$$

と定めることが推奨されている．このとき， f の最小解を \mathbf{x}^* とすると，適切な仮定の下で，関数値の列 $\{f(\mathbf{x}^k)\}$ が最小値 $f(\mathbf{x}^*)$ に収束する．

劣勾配法の収束定理

以下では、点列 $\{x^k\}$ に対応して数列 $\{\epsilon_k\}$ を

$$\epsilon_k := \min_{i=1, \dots, k} \{f(x^i) - f(x^*)\}$$

と定義する。

定理 8.2: 劣勾配法の収束定理

ステップ幅は $\alpha \geq 0, \beta > 0$ を用いて、

$$t_k = \frac{1}{\alpha + \beta k}$$

と決められているとする。さらに、すべての k に対して、不等式

$$\|\eta^k\| \leq L$$

を満たす正の定数 L が存在すると仮定する。このとき、 $\lim_{k \rightarrow \infty} \epsilon_k = 0$ となる。

proof.

任意の i に対して,

$$\begin{aligned}\|x^{i+1} - x^*\| &= \left\| x^i - t_i \frac{\eta^i}{\|\eta^i\|} - x^* \right\|^2 \quad (\because x^i \text{の更新則}) \\ &= \|x^i - x^*\|^2 - 2t_i \frac{(\eta^i)^\top (x^i - x^*)}{\|\eta^i\|} + t_i^2 \\ &\leq \|x^i - x^*\|^2 - 2t_i \frac{f(x^i) - f(x^*)}{\|\eta^i\|} + t_i^2 \quad (\because \eta^i \text{の定義より従う.}^{*3}) \\ &\leq \|x^i - x^*\|^2 - 2t_i \frac{f(x^i) - f(x^*)}{L} + t_i^2 \quad (\because L \text{の仮定})\end{aligned}$$

が成り立つ.

$^{*3} \eta^i \in \partial f(x^i) \implies f(x^*) - f(x^i) \geq (\eta^i)^\top (x^* - x^i) \iff (\eta^i)^\top (x^i - x^*) \geq f(x^i) - f(x^*)$

proof.(続き)

この不等式を $i = 1, \dots, k$ まで足し合わすと,

$$\begin{aligned}\|\mathbf{x}^{k+1} - \mathbf{x}^*\|^2 &\leq \|\mathbf{x}^1 - \mathbf{x}^*\|^2 - \frac{2}{L} \sum_{i=1}^k t_i (f(\mathbf{x}^i) - f(\mathbf{x}^*)) + \sum_{i=1}^k t_i^2 \\ \iff 2 \sum_{i=1}^k t_i (f(\mathbf{x}^i) - f(\mathbf{x}^*)) &\leq L \left(\|\mathbf{x}^1 - \mathbf{x}^*\|^2 - \|\mathbf{x}^{k+1} - \mathbf{x}^*\|^2 + \sum_{i=1}^k t_i^2 \right) \\ &\leq L \left(\|\mathbf{x}^1 - \mathbf{x}^*\|^2 + \sum_{i=1}^k t_i^2 \right)\end{aligned}$$

を得るが, ϵ_k の定義より, 任意の i について, $\epsilon_i \leq f(\mathbf{x}^i) - f(\mathbf{x}^*)$ が成り立ち, かつ $t_i > 0$ なので,

$$2 \sum_{i=1}^k t_i \epsilon_i \leq L \left(\|\mathbf{x}^1 - \mathbf{x}^*\|^2 + \sum_{i=1}^k t_i^2 \right)$$

が成り立つ.

proof.(続き)

ここで、任意の i について、 $\epsilon_i > \epsilon > 0$ となる定数 ϵ が存在すると仮定する。(背理法を用いる。) このとき、

$$2\epsilon \sum_{i=1}^k t_i \leq 2 \sum_{i=1}^k t_i \epsilon_i \leq L \left(\|x^1 - x^*\|^2 + \sum_{i=1}^k t_i^2 \right)$$

が成り立つ。しかし、 $\sum_{i=1}^k t_i = \sum_{i=1}^k \frac{1}{\alpha + \beta i} \rightarrow \infty$ (as $k \rightarrow \infty$) であるが、右辺は有界であることから^{*4}、不等式が矛盾。

よって、任意の i について、 $\epsilon_i > \epsilon > 0$ となる定数 ϵ は存在しない。これと、 ϵ_i の定義より、 ϵ_i は単調減少であることから、 $\lim_{k \rightarrow \infty} \epsilon_k = 0$ となる。 □

^{*4}後に示す。

一般化調和級数

$\alpha \geq 0, \beta > 0, t_k = \frac{1}{\alpha + \beta k}$ とする. このとき,

$$\sum_{k=1}^{\infty} t_k = \sum_{k=1}^{\infty} \frac{1}{\alpha + \beta k} = \infty, \sum_{k=1}^{\infty} t_k^2 = \sum_{k=1}^{\infty} \frac{1}{(\alpha + \beta k)^2} < \infty$$

を (簡単に) 示す.

(i). $\alpha \geq \beta$ のとき,

$\alpha + \beta k \leq (1+k)\alpha$ より, $\frac{1}{\alpha + \beta k} \geq \frac{1}{(1+k)\alpha}$ が成り立つので,

$$\sum_{k=1}^{\infty} \frac{1}{\alpha + \beta k} \geq \frac{1}{\alpha} \sum_{k=1}^{\infty} \frac{1}{1+k} = \infty$$

$\alpha + \beta k \geq (1+k)\beta$ より, $\frac{1}{\alpha + \beta k} \leq \frac{1}{(1+k)\beta}$ が成り立つので,

$$\sum_{k=1}^{\infty} \frac{1}{(\alpha + \beta k)^2} \leq \frac{1}{\beta} \sum_{k=1}^{\infty} \left(\frac{1}{1+k} \right)^2 < \infty$$

(ii). $\alpha \leq \beta$ のときも, 同様の不等式の評価から示される.

劣勾配法

制約条件 $x \in S$ があるような問題に対しては、第 9 章で紹介する射影勾配法のように、点列を $x^{k+1} = P_S \left(x^k - t_k \frac{\eta^k}{\|\eta^k\|} \right)$ として生成すればよい。

ただし、 $P_S(x)$ は x の集合 S への射影^{*5}である。

この場合も上記の定理と同様に目的関数値が最小値に収束することを示せる。

劣勾配法では、過去の (劣) 勾配の情報、ヘッセ行列の情報を利用していないので、高速な収束は期待できない。→ **なぜ？**

そこで、過去の劣勾配の情報を用いて、目的関数 f をモデル化した関数を構成し、それに基づいて、点列を生成することを考える。そのようなアイデアに基づいた手法が切除平面法と bundle 法である。

^{*5}8 章で扱う。

モデル関数

これまでの反復で x^0, \dots, x^k が生成されており，それらの点における目的関数 f の劣勾配 η^0, \dots, η^k が求まっているとする．このとき，目的関数 f の近似モデル関数 \hat{f}_k を以下のように定義する．

$$\hat{f}_k(x) := \max_{i=0, \dots, k} \{f(x^i) + (\eta^i)^\top (x - x^i)\}$$

定理 2.9 より ^{*6}， $\hat{f}_k(x)$ は凸関数である．

さらに，任意の $i = 0, \dots, k$ に対して， η^i の定義より，

$$f(x) \geq f(x^i) + (\eta^i)^\top (x - x^i), \forall x$$

が成り立つので， $f(x) \geq \hat{f}_k(x)$ である．

さらに， \hat{f}_k の定義より，明らかに，

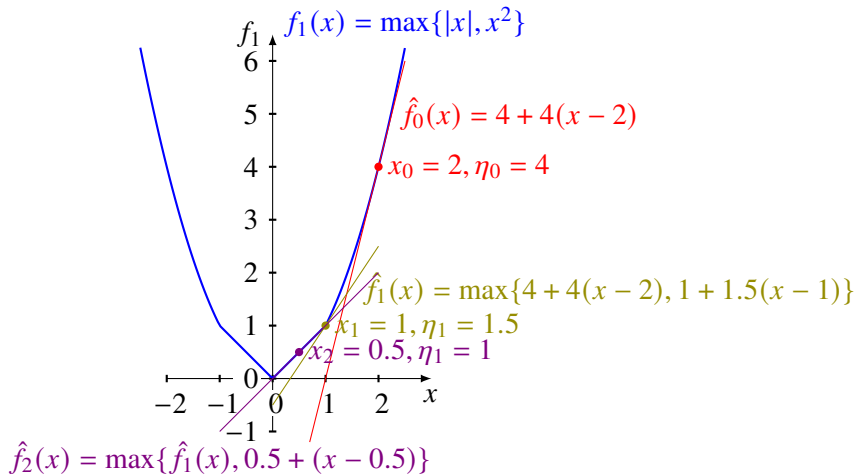
$$\hat{f}_k(x) \geq \hat{f}_{k-1}(x) \geq \dots \geq \hat{f}_1(x) \geq \hat{f}_0(x)$$

が成り立つ．つまり， x^k を適切に選んでいけば， \hat{f}_k は f のよいモデル関数となることが期待できる．

^{*6}関数 $g(x, z)$ は， z を固定したとき， x に関して凸関数であるとする．このとき， $\sup_{z \in Z} g(x, z)$ は凸関数である．

例 (近似モデル関数)

$$\hat{f}_k(x) := \max_{i=0,\dots,k} \{f(x^i) + (\eta^i)^\top (x - x^i)\}$$



切除平面法

モデル関数 $\hat{f}_k(x)$ の最小点を次の反復点とする手法が切除平面法である. :

$$x^{k+1} := \arg \min_{x \in \mathbb{R}^n} \hat{f}_k(x)$$

この問題は, スラック変数 s を導入した次の LP に変換することができる.

$$\begin{array}{ll} \underset{s \in \mathbb{R}}{\text{minimize}} & s \\ \text{subject to} & s \geq f(x^i) + (\eta^i)^\top (x - x^i), \quad i = 1, \dots, k \end{array}$$

先ほどの例が示すように, この LP は有界でない場合がある. そのため, 適当な上下限制約 $C = \{x \in \mathbb{R}^n \mid l_i \leq x_i \leq u_i, i = 1, \dots, n\}$ を用意し, 上記の問題の代わりに, 以下の LP を解いて点列を生成することが多い.

$$\begin{array}{ll} \underset{s \in \mathbb{R}}{\text{minimize}} & s \\ \text{subject to} & s \geq f(x^i) + (\eta^i)^\top (x - x^i), \quad i = 1, \dots, k \\ & x \in C \end{array} \quad (8.4)$$

切除平面法

切除平面法

step 0: $x^0 \in C$ を選び, $\hat{f}_{-1} := -\infty, k := 0$ とする.

step 1: $f(x^k)$ と劣勾配 $\eta^k \in \partial f(x^k)$ を求める.

step 2: 終了条件を満たしていれば終了.

step 3: モデル関数を更新する.

$$\hat{f}_k(x) := \max \{ \hat{f}_{k-1}(x), f(x^k) + (\eta^k)^\top (x - x^k) \}$$

step 4: LP(8.4) の解を求め, その解を x^{k+1} とする.

step 5: $k := k + 1$ として, step 1 へ.

切除平面法では, 反復が進むにつれて, 解くべき LP の制約が増えていく. 単体法で解く場合は, 制約の数に応じて^{*7}, 計算時間も増えるため, 反復の回数が増えるにつれて, 部分問題 (8.4) を解くのが難しくなるという欠点がある.

^{*7} テキストには (梅谷先生本にも), "(制約の数に) 比例して"と書いてあるが, 本当に比例?

bundle 法

そのような欠点を克服するために、問題 (8.4) の目的関数に 2 次関数 $\frac{\mu_k}{2} \|x - x^k\|^2$ を付けた次の問題を考え、その問題の最小解を次の反復点とする手法が bundle 法である. :

$$x^{k+1} := \arg \min_{x \in \mathbb{R}^n} \hat{f}_k(x) + \frac{\mu_k}{2} \|x - x^k\|^2 \quad (8.5)$$

この問題は凸二次計画問題に変換することができる^{*8}. また、2 次関数 $\frac{\mu_k}{2} \|x - x^k\|^2$ の項により、制約条件 $x \in C$ を付けなくても、最小解を持つ (この問題は有界である).

ここで、 μ_k は現在の点との距離を調整するパラメータである.

^{*8} $d := x - x^k$ として、問題 (8.4) のように、スラック変数を導入すればよい.

bundle 法

bundle 法では、大域的収束を保証するために、信頼領域法と同様のアイデアを用いて反復点の更新を行なう。

\mathbf{y}^{k+1} を問題 (8.5) の解として、さらに、このとき、

$$r_k := \frac{f(\mathbf{x}^k) - f(\mathbf{y}^{k+1})}{f(\mathbf{x}^k) - \left(\hat{f}_k(\mathbf{y}^{k+1}) + \frac{\mu_k}{2} \|\mathbf{y}^{k+1} - \mathbf{x}^k\|^2 \right)} \quad (8.6)$$

とする。この r_k を用いて、モデル関数 \hat{f}_k の良し悪しを判定する。

r_k が大きいときは、 $\mathbf{x}^{k+1} := \mathbf{y}^{k+1}$ として、更新する。

そうでなければ、 $\mathbf{x}^{k+1} := \mathbf{x}^k$ として、モデルを改善する。(信頼領域法では、信頼半径を小さくすることでモデル関数を改善したが、bundle 法では \mathbf{y}^{k+1} と劣勾配 $\boldsymbol{\eta}^{k+1} \in \partial f(\mathbf{y}^{k+1})$ を用いてモデルを改善する。)

bundle 法

bundle 法

step 0: $\alpha \in (0, 1)$ とする. $x^0 \in \mathbb{R}^n$ を適当に選び, $y^0 := x^0$ とする.
 $f(x^0)$ と $\eta^0 \in \partial f(x^0)$ を求める.

$\hat{f}_0(x) := f(x^0) + (\eta^0)^\top (x - x^0)$ とする.

step 1: 部分問題 (8.5) を解いて, その解を y^{k+1} とする.

step 2: y^{k+1} が終了条件を満たしていれば終了する.

step 3: $f(y^{k+1})$ と劣勾配 $\eta^{k+1} \in \partial f(y^{k+1})$ を求める.

step 4: 式 (8.6) によって, r_k を計算する.

もし, $r_k \geq \alpha$ であれば, $x^{k+1} := y^{k+1}$, $k := k + 1$ として, step 1 へ.

そうでなければ, $x^{k+1} := x^k$ とする.

step 5: モデル関数を更新する.

$$\hat{f}_{k+1}(x) := \max \{ f_k(x), f(y^{k+1}) + (\eta^{k+1})^\top (x - y^{k+1}) \}$$

$k := k + 1$ として, step 1 へ.

目次

- ① 8.1 微分不可能な最小化問題に対する直線探索法
- ② 8.2 凸計画問題に対する劣勾配を用いた手法
- ③ 大域的収束性が保証された微分を用いない最適化法
 - 直接探索法
 - simplex gradient を用いた手法

本節では、大域的収束性^{*9}が保証された微分を用いない最適化法 (Derivative-Free Optimization; DFO) を紹介する。

DFO はシミュレーション最適化^{*10}など目的関数の評価に時間がかかり、勾配の計算ができない問題に対して用いられる。

DFO には、直接探索法、有限差分近似による勾配を用いた手法、サンプル点から構成されたモデル関数を用いる信頼領域法などがある。

ここでは、直接探索法と有限差分近似による勾配を用いる手法を紹介する。

これらの方法では、現在点 x^k とその近傍の m 点の情報を用いて、最適性の 1 次の必要条件を確認する。

^{*9}反復法において、任意の初期点 x^0 から初めて、非線形計画問題の何らかの解に収束するとき、その反復法は大域的収束するという。ここで「何らかの解」とは、必ずしも大域的解というわけではなく、局所的最適解であったり、最適性の必要条件をみたす実行可能解であったりすることに注意してほしい。(1 章) **大域的最適性ではない！**

^{*10}例) 待ち行列の最適化

目次

- ① 8.1 微分不可能な最小化問題に対する直線探索法
- ② 8.2 凸計画問題に対する劣勾配を用いた手法
- ③ 大域的収束性が保証された微分を用いない最適化法
直接探索法
simplex gradient を用いた手法

(近傍の点の情報を用いた) 方向微分を用いた最適性の 1 次の必要条件

現在の点 x^k とそこからの方向 $d^i, i = 1, \dots, m$ が与えられているとする．このとき， f が方向微分可能であれば， d^i 方向の微分は，十分小さい正数 ϵ を用いて，

$$f'(x^k; d^i) \approx \frac{f(x^k + \epsilon d^i) - f(x^k)}{\epsilon} \quad (8.7)$$

と近似できる．方向微分を用いた最適性の 1 次の必要条件是，全ての方向 $d \in \mathbb{R}^n$ に対して， $f'(x^k, d)$ となることから，すべての $d^i, i = 1, \dots, m$ に対して，

$$f(x^k + \epsilon d^i) - f(x^k) \geq 0$$

となることは，「最適性の 1 次の必要条件」の**必要条件**となる．これが，「最適性の 1 次の必要条件」の**十分条件**となるためには，ベクトル $d^i, i = 1, \dots, m$ に条件を加える必要がある．その条件の 1 つが，ベクトル $d^i, i = 1, \dots, m$ で張られる錐 $\{v \in \mathbb{R}^n \mid v = \sum_{i=1}^m \alpha_i d^i, \alpha_i \geq 0, i = 1, \dots, m\}$ が全空間 \mathbb{R}^n と一致することである．

(近傍の点の情報を用いた) 方向微分を用いた 1 次の必要条件

実際に，そのような条件が成り立つとき，任意の $d \in \mathbb{R}^n$ に対して， $d = \sum_{i=1}^m \alpha_i d^i$ を満たすような $\alpha_i \geq 0, i = 1, \dots, m$ が存在する．さらに， f が x^k で微分可能であるときは，

$$\begin{aligned} f'(x; d) &= \nabla f(x^k)^\top d = \sum_{i=1}^m \alpha_i \nabla f(x^k)^\top d^i = \sum_{i=1}^m \alpha_i f'(x^k; d^i) \\ &\approx \sum_{i=1}^m \alpha_i \frac{f(x^k + \epsilon d^i) - f(x^k)}{\epsilon} \geq 0 \end{aligned} \tag{8.8}$$

となる．この不等式が任意の $d \in \mathbb{R}^n$ に対して成り立つので，最適性の 1 次の必要条件が近似的に成り立つ． f が x^k で微分不可能なときでも，同様の性質を示すことができる．^{*11}

^{*11}他に仮定が必要そう？

直接探索法

直線探索法は反復法の 1 つであり，現在の反復点の近傍のいくつかの候補点の中から，関数値の最も低い点を次の反復点とする手法である．

以下では，候補点を選ぶために必要となる，現在の反復点から探索する方向の集合を $D := \{d^1, \dots, d^m\}$ とする． D の各ベクトルで張られる錐が全空間となるとき，つまり， $\{v \mid v = \sum_{i=1}^m \alpha_i d^i, \alpha_i \geq 0, i = 1, \dots, m\} = \mathbb{R}^n$ のとき， D を positively spanning set という．さらに， D の要素が 1 つでも抜けたら positively spanning set にならないとき， D は正基底であるという．最も単純な正基底として， D に各軸方向をとる手法が考えられる．^{*12}

^{*12} $n = 2$ のとき， $D = \{(1, 0)^\top, (0, 1)^\top, (-1, 0)^\top, (0, -1)^\top\}$

直接探索法

直接探索法

step 0: 正基底 D と $\alpha_0 > 0$, $\mathbf{x}^0 \in \mathbb{R}^n$ を選ぶ. $k := 0$ とする.

step 1: $f(\mathbf{x}^k) > f(\mathbf{x}^k + \alpha_k \mathbf{d})$ となる $\mathbf{d} \in D$ を探す.

そのような \mathbf{d} が見つければ, step 2 へ.

そうでなければ, step 3 へ.

step 2: $\mathbf{x}^{k+1} := \mathbf{x}^k + \alpha_k \mathbf{d}$ とし, 正基底 D を更新する. $k := k + 1$ として, step 1 へ.

step 3: α_k を小さくし, 正基底 D を更新して, step 1 へ.

いま, D が正基底であるから, α_k が十分小さいとき, 式 (8.8) から, よい近似解が求まっていることになる.

直線探索法は, シンプルで大域的収束性も保証されているが, f の勾配やヘッセ行列の情報を使っていないため, 収束が遅い. また, 最悪の場合, 各反復で m 回関数の評価をしなければならない. そのため, 直接探索法は, いくつかの高速化テクニックと組み合わせて使われる. 高速化に使われるテクニックに代理関数と探索ステップがある.

目次

- ① 8.1 微分不可能な最小化問題に対する直線探索法
- ② 8.2 凸計画問題に対する劣勾配を用いた手法
- ③ 大域的収束性が保証された微分を用いない最適化法
 - 直接探索法
 - simplex gradient を用いた手法

simplex gradient

n 個の 1 次独立なベクトル $d^1, \dots, d^n \in \mathbb{R}^n$ が与えられているとする.
点 $x^k, x^k + d^1, \dots, x^k + d^n$ において, 関数 f と同じ関数値をとる 1 次関数を

$$\hat{f}(x) := g^\top x + c$$

とする. ここで, $g \in \mathbb{R}^n$ は 1 次関数 \hat{f} の系列ベクトルであり, $c \in \mathbb{R}$ は定数である.

simplex gradient

関数 \hat{f} は、定義より、

$$\begin{aligned}\hat{f}(\mathbf{x}^k) &= \mathbf{g}^\top \mathbf{x}^k + c = f(\mathbf{x}^k) \\ \hat{f}(\mathbf{x}^k + \mathbf{d}^i) &= \mathbf{g}^\top (\mathbf{x}^k + \mathbf{d}^i) + c = f(\mathbf{x}^k + \mathbf{d}^i), \quad i = 1, \dots, n\end{aligned}$$

を満たすから、

$$\mathbf{g}^\top \mathbf{d}^i = f(\mathbf{x}^k + \mathbf{d}^i) - f(\mathbf{x}^k), \quad i = 1, \dots, n$$

が成り立つ。これは、

$$Y\mathbf{g} = \begin{pmatrix} f(\mathbf{x}^k + \mathbf{d}^1) - f(\mathbf{x}^k) \\ \vdots \\ f(\mathbf{x}^k + \mathbf{d}^n) - f(\mathbf{x}^k) \end{pmatrix}$$

と書ける。ただし、 $Y = [\mathbf{d}^1 \cdots \mathbf{d}^n]^\top$ であり、 $\mathbf{d}^1, \dots, \mathbf{d}^n$ が 1 次独立であることから、 Y は正則である。

simplex gradient

よって、1 次のモデル関数 \hat{f} の勾配 g は、

$$g = Y^{-1} \begin{pmatrix} f(\mathbf{x}^k + \mathbf{d}^1) - f(\mathbf{x}^k) \\ \vdots \\ f(\mathbf{x}^k + \mathbf{d}^n) - f(\mathbf{x}^k) \end{pmatrix} \quad (8.9)$$

で与えられる．この g を simplex gradient あるいは，stencil gradient と呼ぶ．

t を微小量としたとき， $Y = tI$ の simplex gradient は通常の有限差分近似によるものと同じになる．そのため，simplex gradient は有限差分近似の一般化とみなせる．

simplex gradient を用いた手法

直接探索法では目的関数の勾配情報を陽に用いていなかった．ここで紹介する手法は，(近似的な) 勾配の情報を積極的に用いる手法である．
このような手法には，有限差分近似による勾配

$$\hat{\mathbf{g}}^k = \begin{pmatrix} \frac{f(\mathbf{x}^k + h\mathbf{e}^1) - f(\mathbf{x}^k)}{h} \\ \vdots \\ \frac{f(\mathbf{x}^k + h\mathbf{e}^n) - f(\mathbf{x}^k)}{h} \end{pmatrix}$$

を用いた降下法 (例えば，最急降下法や準ニュートン法) がある．ここで， h は十分に小さい正の定数である．しかし，このような有限差分近似では，各反復ごとに， \mathbf{x}^k の周りで n 回の関数の評価が必要である．また，目的関数 f が正確に計算できないときやノイズが入る場合，差分 h を微小の値に固定した近似勾配 $\hat{\mathbf{g}}^k$ は役に立たないことがある．そのような欠点を克服するために，simplex gradient を用いた最急降下法や準ニュートン法が提案されている．

simplex gradient を用いた手法

$d^i = x^{k-i} - x^i$ と選べば, simplex gradient の計算には既に関数値が評価された点の情報を再利用することができるため, 各反復で n 回関数を評価する必要がない. また, アルゴリズムの初期の段階では, 有限差分近似の差分 h に相当する $\|d^i\|$ が大きくなるため, ノイズに強いことが報告されている.

理論的に大域的収束性を保証するためには, 解の周りではノイズが小さくなるなどの厳しい条件が必要となる. ただし, たいていの場合には, 十分実用的な近似解を得ることができる.