システム工学第 4 回 (制約なし非線形計画法)

長江 剛志

東北大学大学院工学研究科
技術社会システム専攻

(nagae@m.tohoku.ac.jp)

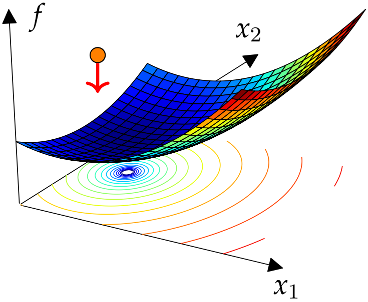
January 16, 2017 (ver1.0)

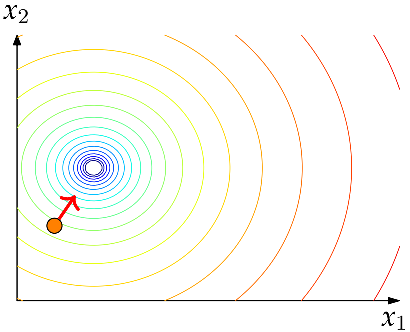
目次

制約無しの最適化問題

$$\min_{\boldsymbol{x}} f(\boldsymbol{x})$$

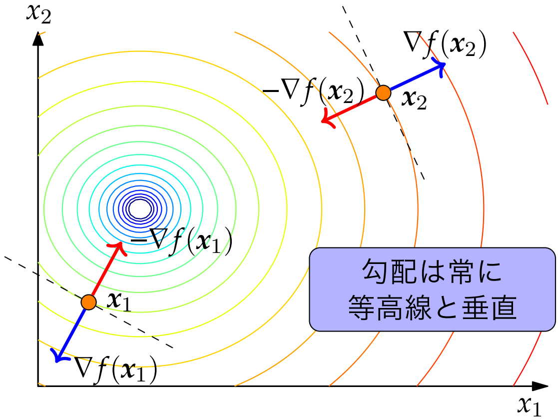
$$\boldsymbol{x} = (x_1, \cdots, x_{\mathrm{N}})^\top$$
$$f : \mathcal{R}^{\mathrm{N}} \to \mathcal{R}$$





関数の勾配

- 勾配 (gradient) : $\nabla f(\boldsymbol{x}) = \left(\frac{\partial f}{\partial x_1}(\boldsymbol{x}), \cdots, \frac{\partial f}{\partial x_N}(\boldsymbol{x})\right)^\top$
- $-\nabla f(\boldsymbol{x})$: 点 $\boldsymbol{x}$ に置いた玉が転がっていく方向 (=目的関数を下げる方向)



関数のヘッセ行列

- ヘッセ行列 (Hessian) :

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

1 変数関数と多変数関数の対応

関数	1 階微分	2 階微分
$f(x)$	$f'(x) = \frac{\mathrm{d}f}{\mathrm{d}x}(x)$	$f''(x) = \frac{\mathrm{d}^2 f}{\mathrm{d}x^2}(x)$
$f(\boldsymbol{x})$	$\nabla f(\boldsymbol{x}) = \begin{bmatrix} \frac{\partial f}{\partial x_1}(\boldsymbol{x}) \\ \vdots \\ \frac{\partial f}{\partial x_N}(\boldsymbol{x}) \end{bmatrix}$	$\nabla^2 f(\boldsymbol{x}) = \begin{bmatrix} \frac{\partial^2 f}{\partial x_1{}^2}(\boldsymbol{x}) & \cdots & \frac{\partial^2 f}{\partial x_1 \partial x_N}(\boldsymbol{x}) \\ \vdots & \ddots & \vdots \\ \frac{\partial^2 f}{\partial x_N \partial x_1}(\boldsymbol{x}) & \cdots & \frac{\partial^2 f}{\partial x_N{}^2}(\boldsymbol{x}) \end{bmatrix}$

勾配とヘッセ行列の例

$$f(\boldsymbol{x}) = (x_1 - 1)^2 + 10(x_1^2 - x_2)^2$$

の勾配および ヘッセ行列は，それぞれ，以下のように求められる.

$$\nabla f(\boldsymbol{x}) = \begin{bmatrix} 40x_1^3 + (2 - 40x_2)x_1 - 2 \\ -20(x_1^2 - x_2) \end{bmatrix}$$

$$\nabla^2 f(\boldsymbol{x}) = \begin{bmatrix} 120x_1^2 - 40x_2 + 2 & -40x_1 \\ -40x_1 & 20 \end{bmatrix}$$

多変数関数の Taylor 展開

1 変数関数の x のまわりでの Taylor 展開

$$f(y) = f(x) + \frac{\mathrm{d}f}{\mathrm{d}x}(x)(y-x) + \frac{1}{2}\frac{\mathrm{d}^2 f}{\mathrm{d}x^2}(x)(y-x)^2 + \cdots$$

多変数関数の x のまわりでの Taylor 展開

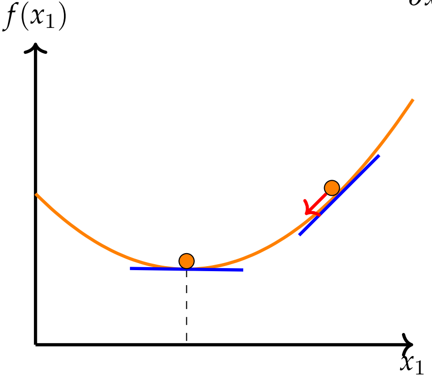
$$\begin{aligned} f(\boldsymbol{y}) &= f(\boldsymbol{x}) + \nabla f(\boldsymbol{x})^\top(\boldsymbol{y}-\boldsymbol{x}) + \frac{1}{2}(\boldsymbol{y}-\boldsymbol{x})^\top \nabla^2 f(\boldsymbol{x})(\boldsymbol{y}-\boldsymbol{x}) + \cdots \\ &= f(\boldsymbol{x}) + \sum_{n=1}^{\mathrm{N}} \frac{\partial f}{\partial x_n}(\boldsymbol{x})(y_n - x_n) \\ &\quad + \frac{1}{2}\sum_{n=1}^{\mathrm{N}}\sum_{m=1}^{\mathrm{N}} \frac{\partial^2 f}{\partial x_n \partial x_m}(\boldsymbol{x})(y_n-x_n)(y_m-x_m) \end{aligned}$$

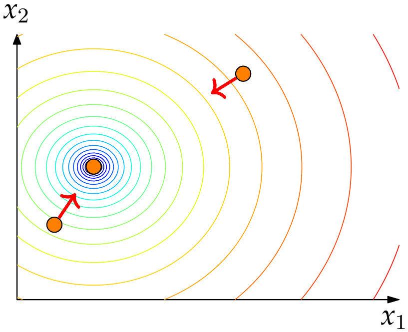
目次

制約なし最適化問題の最適性条件

- $\boldsymbol{x}^*$ が最適解であるための必要条件：
  $\boldsymbol{x}^*$ における勾配のあらゆる要素が 0

$$\nabla f(\boldsymbol{x}^*) = \mathbf{0} \Leftrightarrow \frac{\partial f}{\partial x_n}(\boldsymbol{x}^*) = 0, \quad n = 1, 2, \cdots, \mathrm{N}$$





最適性条件が解析的に解ける場合

2 変数の最適化問題

$$\min_{\boldsymbol{x}} f(\boldsymbol{x}) = (x_1 - 1)^2 + (x_2 - 2)^2 - x_1 x_2$$

目的関数の勾配

$$\nabla f(\boldsymbol{x}) = \begin{bmatrix} 2(x_1 - 1) - x_2 \\ 2(x_2 - 2) - x_1 \end{bmatrix}$$

最適性条件

$$\nabla f(\boldsymbol{x}^*) = \begin{bmatrix} 2x_1^* - x_2^* - 2 \\ -x_1^* + 2x_2^* - 4 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

最適解

$$\begin{bmatrix} x_1^* \\ x_2^* \end{bmatrix} = \begin{bmatrix} \frac{8}{3} \\ \frac{10}{3} \end{bmatrix}$$

目次

最適性条件が解析的に解けない場合

繰り返し計算によって求める

Step 0 (初期化) 適当な初期解 $\boldsymbol{x}^{(0)}$ を与える．$k \leftarrow 0$ とする．

Step 1 (最適性の判定) 最適性条件 $\nabla f(\boldsymbol{x}^{(k)}) = \mathbf{0}$ が満足されていれば終了; $\boldsymbol{x}^{(k)}$ が最適解．

Step 2 (探索方向の決定) 充分に小さい $\alpha > 0$ について $f(\boldsymbol{x}^{(k)}) > f(\boldsymbol{x}^{(k)} + \alpha \boldsymbol{d}^{(k)})$ となるような **探索方向** $\boldsymbol{d}^{(k)}$ を決める．

Step 3 (ステップサイズの決定) 適当な基準に基づいて **ステップサイズ** $\alpha^{(k)} > 0$ を選ぶ．

Step 4 (解の改訂) 次の解を $\boldsymbol{x}^{(k+1)} \leftarrow \boldsymbol{x}^{(k)} + \alpha^{(k)} \boldsymbol{d}^{(k)}$ とする．$k \leftarrow k+1$ として Step 1 に戻る．

Step 2 の具体的な手続き (アルゴリズム) には色々なものが存在する．Step 3 については次のスライドで．

- 最急降下法
- Newton 法・準 Newton 法
- 共役勾配法

ステップサイズの決め方 (一次元探索 / 直線探索) (1)

現在の解 $\boldsymbol{x}$ から探索方向 $\boldsymbol{d}$ に $\alpha > 0$ だけ進んだ解 $\boldsymbol{x} + \alpha\boldsymbol{d}$ における目的関数値を

$$g(\alpha) := f(\boldsymbol{x} + \alpha\boldsymbol{d})$$

とする．$\alpha = 0$ において

$$g(\alpha)\Big|_{\alpha=0} = f(\boldsymbol{x})$$

は自明．最急降下法などで $\boldsymbol{x}$ における勾配 $\nabla f(\boldsymbol{x})$ を求めているなら，一次導関数も

$$\frac{\mathrm{d}g}{\mathrm{d}\alpha}(\alpha)\Big|_{\alpha=0} = \nabla f(\boldsymbol{x})^\top \boldsymbol{d}$$

と簡単に評価できる．

ステップサイズの決め方 (一次元探索 / 直線探索) (2)

Armijo の基準を用いる方法

適当な定数 $0 < \xi < 1$ を予め与えておき，

$$f(\boldsymbol{x} + \alpha \boldsymbol{d}) \leq f(\boldsymbol{x}) + \xi \alpha \nabla f(\boldsymbol{x})^\top \boldsymbol{d}$$

を満足するような α をステップサイズとして採用する.
具体的な手続きとしては，$\alpha^{(0)} = 1$ を初期値として，上記の条件が満足されるまで $\alpha^{(t+1)} = \alpha^{(t)}/2$ とする **2 分割法** など.
この式の右辺は，関数 $g(\alpha)$ の $\alpha = 0$ における接線

$$f(\boldsymbol{x}) + \alpha \nabla f(\boldsymbol{x})^\top \boldsymbol{d}$$

と切片が等しく，傾きがより緩やかな (c.f.$\xi < 1$) 直線

$$f(\boldsymbol{x}) + \xi \alpha \nabla f(\boldsymbol{x})^\top \boldsymbol{d}$$

Armijo の基準:

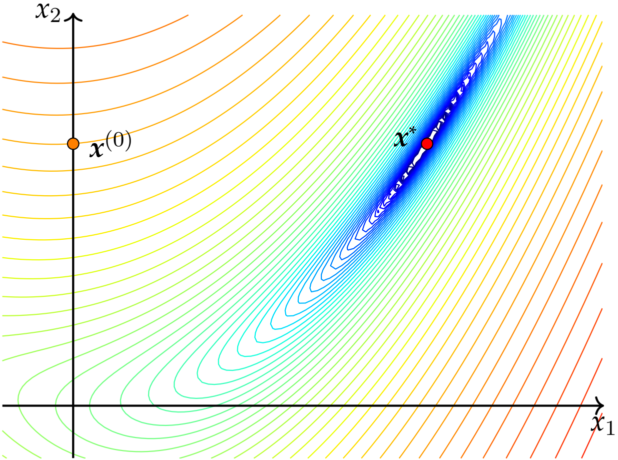
目次

最急降下法

k 回目の探索方向 $\boldsymbol{d}^{(k)}$ として勾配の逆方向 $-\nabla f(\boldsymbol{x}^{(k)})$ を採用.

$$f(\boldsymbol{x}) = (x_1 - 1)^2 + 10(x_1^2 - x_2)^2$$

に最急降下法を適用した場合 (初期値 $\boldsymbol{x}^{(0)} = (0, 1)$).

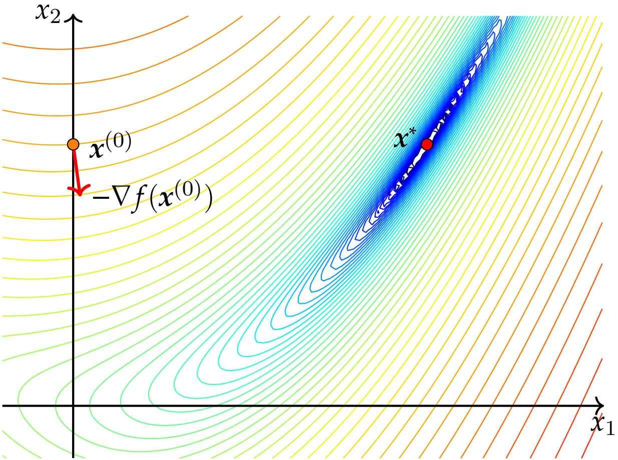


最急降下法

k 回目の探索方向 $\boldsymbol{d}^{(k)}$ として勾配の逆方向 $-\nabla f(\boldsymbol{x}^{(k)})$ を採用.

$$f(\boldsymbol{x}) = (x_1 - 1)^2 + 10(x_1^2 - x_2)^2$$

に最急降下法を適用した場合 (初期値 $\boldsymbol{x}^{(0)} = (0, 1)$).

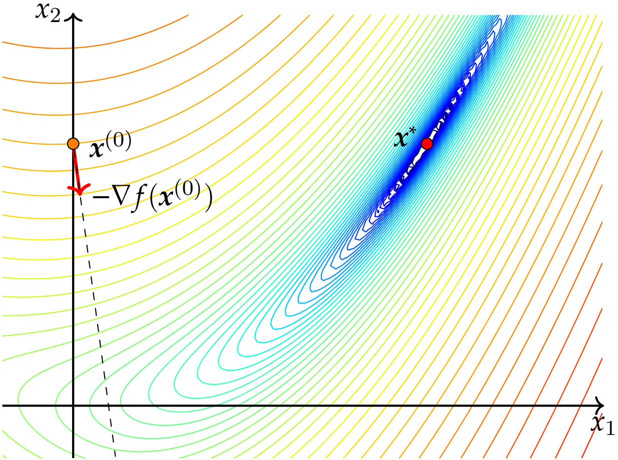


最急降下法

k 回目の探索方向 $\boldsymbol{d}^{(k)}$ として勾配の逆方向 $-\nabla f(\boldsymbol{x}^{(k)})$ を採用.

$$f(\boldsymbol{x}) = (x_1 - 1)^2 + 10(x_1^2 - x_2)^2$$

に最急降下法を適用した場合 (初期値 $\boldsymbol{x}^{(0)} = (0, 1)$).

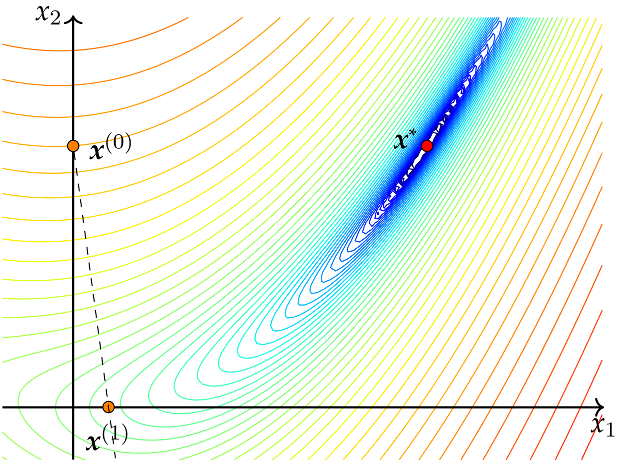


最急降下法

k 回目の探索方向 $\boldsymbol{d}^{(k)}$ として勾配の逆方向 $-\nabla f(\boldsymbol{x}^{(k)})$ を採用.

$$f(\boldsymbol{x}) = (x_1 - 1)^2 + 10(x_1^2 - x_2)^2$$

に最急降下法を適用した場合 (初期値 $\boldsymbol{x}^{(0)} = (0, 1)$).

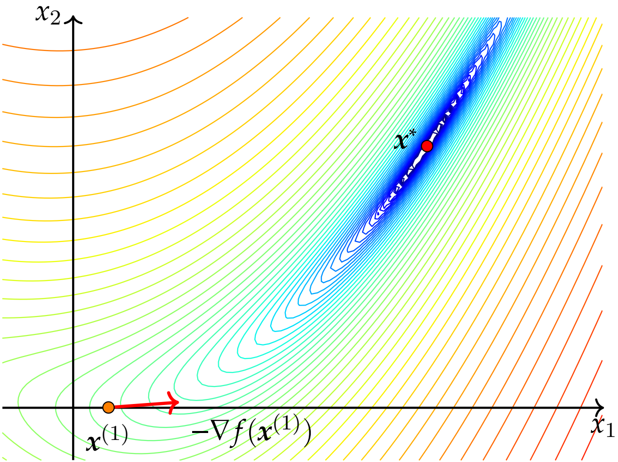


最急降下法

k 回目の探索方向 $\boldsymbol{d}^{(k)}$ として勾配の逆方向 $-\nabla f(\boldsymbol{x}^{(k)})$ を採用.

$$f(\boldsymbol{x}) = (x_1 - 1)^2 + 10(x_1^2 - x_2)^2$$

に最急降下法を適用した場合 (初期値 $\boldsymbol{x}^{(0)} = (0, 1)$).

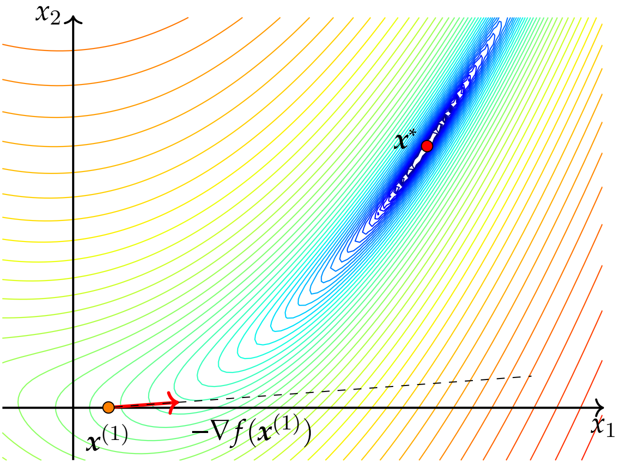


最急降下法

k 回目の探索方向 $\boldsymbol{d}^{(k)}$ として勾配の逆方向 $-\nabla f(\boldsymbol{x}^{(k)})$ を採用.

$$f(\boldsymbol{x}) = (x_1 - 1)^2 + 10(x_1^2 - x_2)^2$$

に最急降下法を適用した場合 (初期値 $\boldsymbol{x}^{(0)} = (0, 1)$).

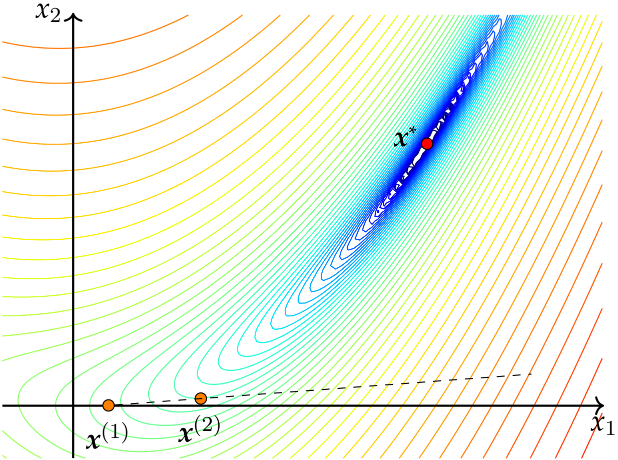


最急降下法

k 回目の探索方向 $\boldsymbol{d}^{(k)}$ として勾配の逆方向 $-\nabla f(\boldsymbol{x}^{(k)})$ を採用.

$$f(\boldsymbol{x}) = (x_1 - 1)^2 + 10(x_1^2 - x_2)^2$$

に最急降下法を適用した場合 (初期値 $\boldsymbol{x}^{(0)} = (0, 1)$).

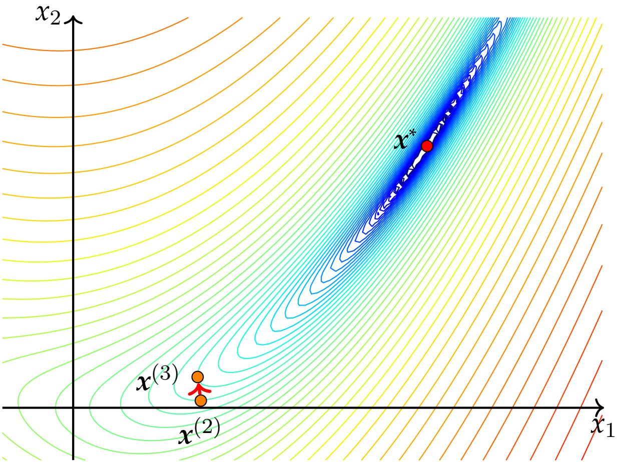


最急降下法

k 回目の探索方向 $\boldsymbol{d}^{(k)}$ として勾配の逆方向 $-\nabla f(\boldsymbol{x}^{(k)})$ を採用.

$$f(\boldsymbol{x}) = (x_1 - 1)^2 + 10(x_1^2 - x_2)^2$$

に最急降下法を適用した場合 (初期値 $\boldsymbol{x}^{(0)} = (0, 1)$).

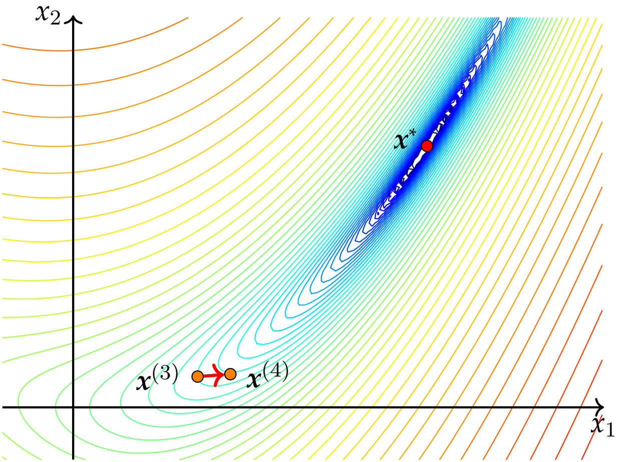


最急降下法

k 回目の探索方向 $\boldsymbol{d}^{(k)}$ として勾配の逆方向 $-\nabla f(\boldsymbol{x}^{(k)})$ を採用.

$$f(\boldsymbol{x}) = (x_1 - 1)^2 + 10(x_1^2 - x_2)^2$$

に最急降下法を適用した場合 (初期値 $\boldsymbol{x}^{(0)} = (0, 1)$).

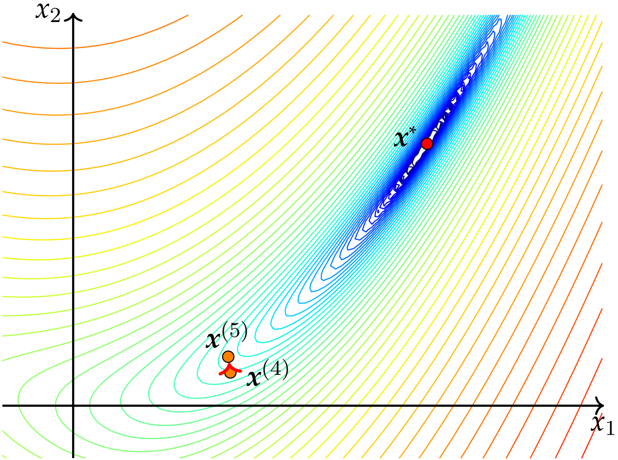


最急降下法

k 回目の探索方向 $\boldsymbol{d}^{(k)}$ として勾配の逆方向 $-\nabla f(\boldsymbol{x}^{(k)})$ を採用.

$$f(\boldsymbol{x}) = (x_1 - 1)^2 + 10(x_1^2 - x_2)^2$$

に最急降下法を適用した場合 (初期値 $\boldsymbol{x}^{(0)} = (0, 1)$).

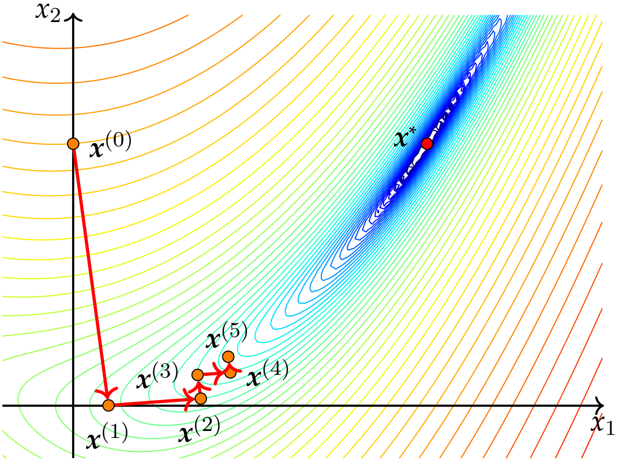


最急降下法

k 回目の探索方向 $\boldsymbol{d}^{(k)}$ として勾配の逆方向 $-\nabla f(\boldsymbol{x}^{(k)})$ を採用.

$$f(\boldsymbol{x}) = (x_1 - 1)^2 + 10(x_1^2 - x_2)^2$$

に最急降下法を適用した場合 (初期値 $\boldsymbol{x}^{(0)} = (0, 1)$).

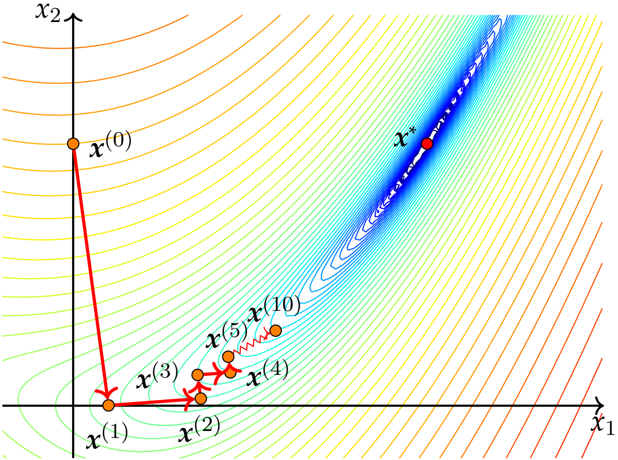


最急降下法

k 回目の探索方向 $\boldsymbol{d}^{(k)}$ として勾配の逆方向 $-\nabla f(\boldsymbol{x}^{(k)})$ を採用.

$$f(\boldsymbol{x}) = (x_1 - 1)^2 + 10(x_1^2 - x_2)^2$$

に最急降下法を適用した場合 (初期値 $\boldsymbol{x}^{(0)} = (0, 1)$).

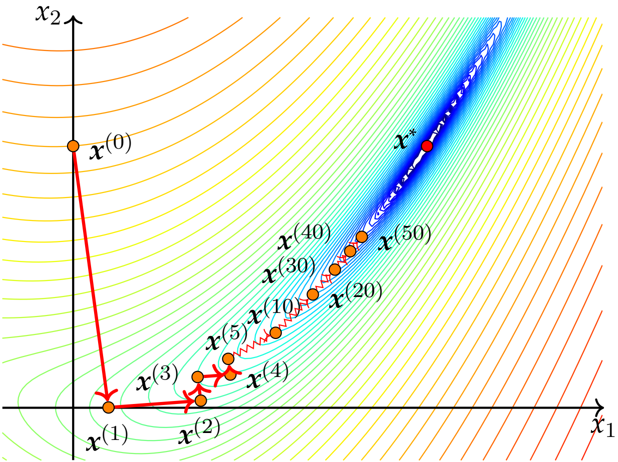


最急降下法

k 回目の探索方向 $\boldsymbol{d}^{(k)}$ として勾配の逆方向 $-\nabla f(\boldsymbol{x}^{(k)})$ を採用.

$$f(\boldsymbol{x}) = (x_1 - 1)^2 + 10(x_1^2 - x_2)^2$$

に最急降下法を適用した場合 (初期値 $\boldsymbol{x}^{(0)} = (0, 1)$).

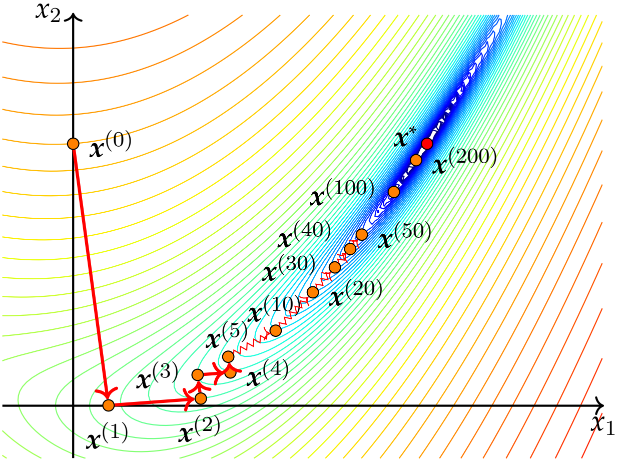


最急降下法

k 回目の探索方向 $\boldsymbol{d}^{(k)}$ として勾配の逆方向 $-\nabla f(\boldsymbol{x}^{(k)})$ を採用.

$$f(\boldsymbol{x}) = (x_1 - 1)^2 + 10(x_1^2 - x_2)^2$$

に最急降下法を適用した場合 (初期値 $\boldsymbol{x}^{(0)} = (0, 1)$).



最急降下法の特徴

長所

- 大域的収束が保証される
- 1 回の繰り返しあたりの計算量が小さい
- 勾配さえ求めれば計算できる

短所

- 探索方向に 2 階微分情報を使わないため，収束が遅い (一次収束). 先の例では $k = 200$ まで繰り返しても最適解に辿りつかない.

(参考) 一次収束

以下を満足するような実数 $0 < \beta < 1$ および自然数 $K > 0$ が存在するとき，点列 $\boldsymbol{x}^{(0)}, \boldsymbol{x}^{(1)}, \boldsymbol{x}^{(2)}, \cdots$ は収束比 β で **一次収束** すると言う.

$$\|\boldsymbol{x}^{(k+1)} - \boldsymbol{x}^{*}\| \leq \beta \|\boldsymbol{x}^{(k)} - \boldsymbol{x}^{*}\|$$

目次

Newton 法 (1)

$f(\boldsymbol{x})$ の 2 階微分情報を使ってより効率的に解を改訂する方法.

関数 f を k 回目の解 $\boldsymbol{x}^{(k)}$ のまわりで Taylor 展開すれば,

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

2 次より大きい項を無視して $\boldsymbol{d} = \boldsymbol{x} - \boldsymbol{x}^{(k)}$ とおけば,

$$f(\boldsymbol{x}^{(k)} + \boldsymbol{d}) \approx f(\boldsymbol{x}^{(k)}) + \nabla f(\boldsymbol{x}^{(k)})^\top \boldsymbol{d} + \frac{1}{2}\boldsymbol{d}^\top \nabla^2 f(\boldsymbol{x}^{(k)})\boldsymbol{d} = g^{(k)}(\boldsymbol{d}) \quad (1)$$

Newton 法は式 (1) の右辺 $g^{(k)}(\boldsymbol{d})$ を最小にするような $\boldsymbol{d}$ (Newton 方向) を用いて解を改訂する方法.

Newton 法 (2)

$\boldsymbol{d}$ についての最適性条件[1]は,

$$\nabla g^{(k)}(\boldsymbol{d}) = \nabla f(\boldsymbol{x}^{(k)}) + \nabla^2 f(\boldsymbol{x}^{(k)})\boldsymbol{d} = \boldsymbol{0}$$

である．ここで，$\nabla f(\boldsymbol{x}^{(k)})$ および $\nabla^2 f(\boldsymbol{x}^{(k)})$ は与件であるから，この最適性条件 (連立方程式) の解を k 回目繰り返しにおける探索方向として採用する:

$$\boldsymbol{d}^{(k)} \leftarrow -\left(\nabla^2 f(\boldsymbol{x}^{(k)})\right)^{-1} \nabla f(\boldsymbol{x}^{(k)})$$

最急勾配法と同様に直線探索を行なってステップサイズ $\alpha^{(k)}$ を求め,

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

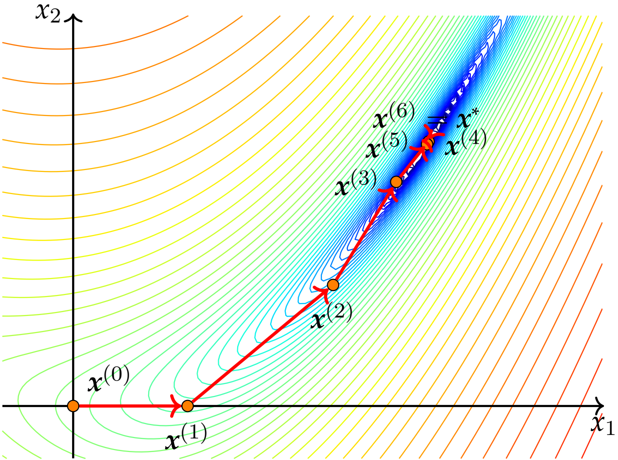
と改訂する．Newton 法では，最適解の近くでは，直線探索を行わず $\alpha^{(k)} = 1$ としても最適解に収束することが知られている．

[1] $\nabla^2 f(\boldsymbol{x}^{(k)})$ が正定値であると仮定する必要がある

Newton 法の実行例

$$f(x_1, x_2) = (x_1 - 1)^2 + 10(x_1^2 - x_2)^2$$

に Newton 法を適用した場合 (初期値 $\boldsymbol{x}^{(0)} = (0, 0)$).



Newton 法の特徴 (1)

長所

- 直線探索を行なわなくても，最適解の近くでは極めて早く収束する (二次収束)
- ヘッセ行列が常に正定値なら Newton 方向は目的関数を下げる方向であるので，大域的収束も保証できる

短所

- 変数が多くなると，繰り返しのたびにヘッセ行列 $\nabla^2 f(\boldsymbol{x}^{(k)})$ を計算するのが大変
- $\nabla^2 f(\boldsymbol{x}^{(k)})$ が計算できても，Newton 方向 $\boldsymbol{d}^{(k)} = -\left(\nabla^2 f(\boldsymbol{x}^{(k)})\right)^{-1} \nabla f(\boldsymbol{x}^{(k)})$ の計算が大変

Newton 法の特徴 (2)

短所 (続き)

- $\nabla^2 f(\boldsymbol{x}^{(k)})$ が正則でない (逆行列を持たない) 場合には, Newton 方向を計算できない
- $\nabla^2 f(\boldsymbol{x}^{(k)})$ が正則であっても，正定値でない場合には, Newton 方向は目的関数を下げる方向とは限らない
  例えば，先の例で初期解を $\boldsymbol{x}^{(0)} = (0, 1)$ とすると, $\nabla^2 f(\boldsymbol{x}^{(0)}) = \begin{bmatrix} -38 & 0 \\ 0 & 20 \end{bmatrix}$ と正定値にならずステップサイズの探索に失敗する.

NEWTON 法の特徴 (3)

(参考) 二次収束

以下を満足するような実数 $C > 0$ および自然数 $\mathrm{K} > 0$ が存在するとき，点列 $\boldsymbol{x}^{(0)}, \boldsymbol{x}^{(1)}, \boldsymbol{x}^{(2)}, \cdots$ は **二次収束** すると言う.

$$\|\boldsymbol{x}^{(k+1)} - \boldsymbol{x}^*\| \leq C\|\boldsymbol{x}^{(k)} - \boldsymbol{x}^*\|^2$$

目次

準 Newton 法 (1)

Newton 法の最適解付近での収束の良さを活かしつつ問題点を克服するため，Newton 法で用いるヘッセ行列を適当な正定値対称行列 $\boldsymbol{B}^{(k)} \approx \nabla^2 f(\boldsymbol{x}^{(k)})$ で近似し，解の改訂と同時に $\boldsymbol{B}^{(k)}$ も更新する.

探索方向 $\boldsymbol{d}^{(k)}$ として，以下の準 Newton 方向を採用する.

$$\boldsymbol{d}^{(k)} \leftarrow -(\boldsymbol{B}^{(k)})^{-1}\nabla f(\boldsymbol{x}^{(k)})$$

Newton 法と異なり，最適解の近くでも直線探索は必要.

正定値行列 $\boldsymbol{B}^{(k)}$ の更新方法としては，以下の BFGS (Broyden-Fletcher-Goldfarb-Shanno) 公式が有名.

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

準 Newton 法 (2)

ただし,

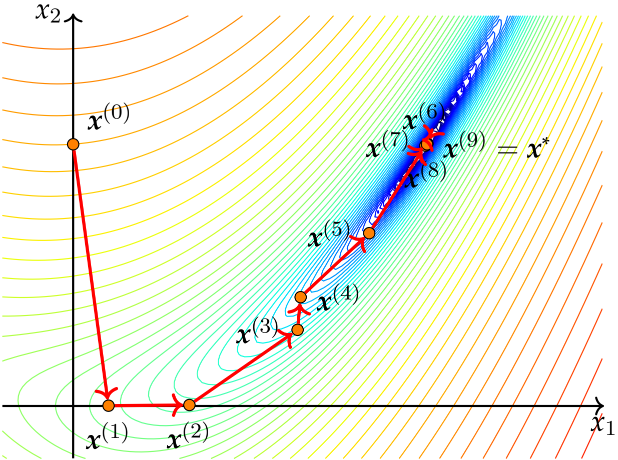
$$\boldsymbol{s}^{(k)} := \boldsymbol{x}^{(k+1)} - \boldsymbol{x}^{(k)} \qquad \boldsymbol{y}^{(k)} := \nabla f(\boldsymbol{x}^{(k+1)}) - \nabla f(\boldsymbol{x}^{(k)})$$

正定値行列の初期値は任意に選んでよいが，特に理由が無ければ $\boldsymbol{B}^{(0)} = \boldsymbol{I}$ と単位行列にすればよい.

準 Newton 法の実行例

$$f(x_1, x_2) = (x_1 - 1)^2 + 10(x_1^2 - x_2)^2$$

に準 Newton 法を適用した場合 (初期値 $\boldsymbol{x}^{(0)} = (0, 1)$)



準 Newton 法の特徴

長所

- ヘッセ行列 $\nabla^2 f(\boldsymbol{x}^{(k)})$ を計算する必要がない
- $\boldsymbol{B}^{(k)}$ を正定値対称と選ぶので Newton 方向が常に目的関数を下げる方向であり，大域的収束が保証される
- 最適解の近くでは非常に早く収束する (超一次収束)
- 実際に探索方向の計算に用いるヘッセ行列の逆行列の近似 $\boldsymbol{H}^{(k)} \approx \left(\nabla^2 f(\boldsymbol{x})\right)^{-1}$ を更新する方法もある

短所

- 変数が大きくなると $\boldsymbol{B}^{(k)}$ の記憶に多くのメモリを必要とする

(参考) 超一次収束

どんな小さな $\beta > 0$ に対しても，それに対して以下を満足する自然数 $\mathrm{K} > 0$ が存在するとき，点列 $\boldsymbol{x}^{(0)}, \boldsymbol{x}^{(1)}, \boldsymbol{x}^{(2)}, \cdots$ は **超一次収束** すると言う.

目次

共役勾配法 (1)

準 Newton 法は正定値行列 $\boldsymbol{B}^{(k)}$ が疎にならないため，大規模問題に適用するには巨大な記憶容量が必要となる．そこで，大規模問題にも適用可能な方法として注目されたのが (非線形) 共役勾配法である．

共役勾配法は，もともと，狭義凸 2 次関数:

$$f(\boldsymbol{x}) = \frac{1}{2}\boldsymbol{x}^\top \boldsymbol{V}\boldsymbol{x} + \boldsymbol{c}^\top \boldsymbol{x} + c_0 \tag{2}$$

の最小化問題を解くための手法．ここで，$\boldsymbol{V} \in \mathcal{R}^{\mathrm{N}\times\mathrm{N}}, \boldsymbol{c} \in \mathcal{R}^{\mathrm{N}}$ で行列 $\boldsymbol{V}$ は正定値対称．

最急降下法の欠点は，探索方向 $\boldsymbol{d}^{(1)}, \boldsymbol{d}^{(2)}, \cdots$ が互いに直交する:

$${\boldsymbol{d}^{(k+1)}}^\top \boldsymbol{d}^{(k)} = \nabla f(\boldsymbol{x}^{(k+1)})^\top \nabla f(\boldsymbol{x}^{(k)}) = 0$$

共役勾配法 (2)

ため，解の挙動がジグザグになることだった．言い換えれば，探索方向の自由度が小さ過ぎることが問題だった．

そこで，k 回目の探索方向 $\boldsymbol{d}^{(k)}$ を，それまでの探索方向 $\boldsymbol{d}^{(0)}, \boldsymbol{d}^{(1)}, \cdots, \boldsymbol{d}^{(k-1)}$ に対して共役 (きょうやく) になる，すなわち，

$$\boldsymbol{d}^{(i)\top} \boldsymbol{V} \boldsymbol{d}^{(k)} = 0, \quad i = 0, 1, \cdots, k-1$$

を満足するように取る方法が **共役方向法** であり，共役な方向の生成に勾配 $\nabla f(\boldsymbol{x})$ を用いるのが **共役勾配法** である．$f(\boldsymbol{x})$ が狭義 2 次凸関数の場合，$\boldsymbol{d}^{(0)} = -\nabla f(\boldsymbol{x}^{(0)})$ から始めて，$k = 1, 2 \cdots$ に対し，

$$\boldsymbol{d}^{(k)} \leftarrow -\nabla f(\boldsymbol{x}^{(k)}) + \frac{\nabla f(\boldsymbol{x}^{(k)})^\top \boldsymbol{V} \boldsymbol{d}^{(k-1)}}{(\boldsymbol{d}^{(k-1)})^\top \boldsymbol{V} \boldsymbol{d}^{(k-1)}} \boldsymbol{d}^{(k-1)}$$

として共役方向を生成する．

共役勾配法 (3)

一般の非線形関数に適用する場合，素朴には，各繰り返しで目的関数を Taylor 2 次展開して $\nabla^2 f(\boldsymbol{x}^{(k)}) \Leftrightarrow \boldsymbol{V}$, $\nabla f(\boldsymbol{x}^{(k)}) \Leftrightarrow \boldsymbol{V}\boldsymbol{x} + \boldsymbol{c}$ と対応づけることで共役勾配法を拡張できる (**非線形共役勾配法**).

しかし，こうしたナイーブな拡張では各反復で $\nabla^2 f(\boldsymbol{x}^{(k)})$ の計算が必要なため，(本来の目的である) 大規模問題への適用の妨げとなる．そこで，ヘッセ行列を使わずに共役方向を生成する方法がいくつか考えられている：

Fletcher-Reeves 法

$$\boldsymbol{d}^{(k+1)} \leftarrow -\nabla f(\boldsymbol{x}^{(k+1)}) + \frac{\|\nabla f(\boldsymbol{x}^{(k+1)})\|^2}{\|\nabla f(\boldsymbol{x}^{(k)})\|^2} \boldsymbol{d}^{(k)}$$

共役勾配法 (4)

Polak-Ribiere 法

$$\boldsymbol{d}^{(k+1)} \leftarrow -\nabla f(\boldsymbol{x}^{(k+1)}) + \frac{\left(\nabla f(\boldsymbol{x}^{(k+1)}) - \nabla f(\boldsymbol{x}^{(k)})\right)^\top \nabla f(\boldsymbol{x}^{(k+1)})}{\|\nabla f(\boldsymbol{x}^{(k)})\|^2} \boldsymbol{d}^{(k)}$$