ガウス・ニュートン法とレーベンバーグ・マーカート法

ほげ

2021 年 2 月 7 日

1　ニュートン法

ニュートン法は, 2 階微分可能な関数 $f(\boldsymbol{x})$ を, $\boldsymbol{x}$ に関して最小化するための逐次的な手法である [6]. その派生であるガウス・ニュートン法や, レーベンバーグ・マーカート法は, グラフベース SLAM の基礎となるポーズ調整 [2][4][5] や, スキャンマッチング [1][3] などで用いられている.

関数 $f(\boldsymbol{x}) : \mathbb{R}^N \to \mathbb{R}$ の最適化を, 初期値 $\breve{\boldsymbol{x}} \in \mathbb{R}^N$ から始めるとする. 初期値 $\breve{\boldsymbol{x}}$ からの変位を $\Delta\boldsymbol{x}$ として, 元の関数 $f(\boldsymbol{x})$ を変位に対する関数 $f(\breve{\boldsymbol{x}} + \Delta\boldsymbol{x})$ として書き直し, $\Delta\boldsymbol{x}$ に関して最小化する. そのような $\Delta\boldsymbol{x}$ が求まったら, 現在の値を $\breve{\boldsymbol{x}}$ から $\breve{\boldsymbol{x}} + \Delta\boldsymbol{x}$ に更新する. これがニュートン法の大まかな流れとなる.

変位に対する関数 $f(\breve{\boldsymbol{x}} + \Delta\boldsymbol{x})$ を, 2 次の項までテイラー展開すると

$$f(\breve{\boldsymbol{x}} + \Delta\boldsymbol{x}) \simeq f(\breve{\boldsymbol{x}}) + \boldsymbol{g}^\top \Delta\boldsymbol{x} + \frac{1}{2}\Delta\boldsymbol{x}^\top \boldsymbol{H} \Delta\boldsymbol{x} \tag{1}$$

(1) 式において, $\boldsymbol{g} = \nabla f(\breve{\boldsymbol{x}}) \in \mathbb{R}^N$ と $\boldsymbol{H} = \nabla^2 f(\breve{\boldsymbol{x}}) \in \mathbb{R}^{N\times N}$ は, それぞれ $f(\boldsymbol{x})$ の勾配およびヘッセ行列であり, 次のように書ける.

$$\nabla f(\boldsymbol{x}) = \frac{\partial}{\partial \boldsymbol{x}} f(\boldsymbol{x}) = \left[\frac{\partial}{\partial x_1} f(\boldsymbol{x}), \cdots, \frac{\partial}{\partial x_N} f(\boldsymbol{x})\right]^\top \in \mathbb{R}^N \tag{2}$$

$$\nabla^2 f(\boldsymbol{x}) = \begin{bmatrix} \frac{\partial^2}{\partial x_1^2} f(\boldsymbol{x}) & \frac{\partial^2}{\partial x_1 \partial x_2} f(\boldsymbol{x}) & \cdots & \frac{\partial^2}{\partial x_1 \partial x_N} f(\boldsymbol{x}) \\ \frac{\partial^2}{\partial x_2 \partial x_1} f(\boldsymbol{x}) & \frac{\partial^2}{\partial x_2^2} f(\boldsymbol{x}) & \cdots & \frac{\partial^2}{\partial x_2 \partial x_N} f(\boldsymbol{x}) \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial^2}{\partial x_N \partial x_1} f(\boldsymbol{x}) & \frac{\partial^2}{\partial x_N \partial x_2} f(\boldsymbol{x}) & \cdots & \frac{\partial^2}{\partial x_N^2} f(\boldsymbol{x}) \end{bmatrix} \in \mathbb{R}^{N\times N} \tag{3}$$

(1) 式を $\Delta\boldsymbol{x}$ に関して最小化するために, $f(\breve{\boldsymbol{x}} + \Delta\boldsymbol{x})$ を $\Delta\boldsymbol{x}$ で偏微分すると

$$\frac{\partial}{\partial \Delta\boldsymbol{x}} f(\breve{\boldsymbol{x}} + \Delta\boldsymbol{x}) = \frac{\partial}{\partial \Delta\boldsymbol{x}} \left(f(\breve{\boldsymbol{x}}) + \boldsymbol{g}^\top \Delta\boldsymbol{x} + \frac{1}{2}\Delta\boldsymbol{x}^\top \boldsymbol{H} \Delta\boldsymbol{x} \right) = \boldsymbol{g} + \boldsymbol{H}\Delta\boldsymbol{x} \tag{4}$$

ここでベクトルの微分に関する次の関係式を用いた.

$$\frac{\partial}{\partial \boldsymbol{x}} \boldsymbol{a}^\top \boldsymbol{x} = \boldsymbol{a} \tag{5}$$

$$\frac{\partial}{\partial \boldsymbol{x}} \boldsymbol{x}^\top \boldsymbol{A} \boldsymbol{x} = \left(\boldsymbol{A} + \boldsymbol{A}^\top\right) \boldsymbol{x} \tag{6}$$

$f(\boldsymbol{x})$ は 2 階微分可能であるので，ヘッセ行列 $\boldsymbol{H} = \nabla^2 f(\boldsymbol{x})$ を構成する各要素は，微分の順序を入れ替えられる．従って $\partial^2/\partial x_i \partial x_j f(\boldsymbol{x}) = \partial^2/\partial x_j \partial x_i f(\boldsymbol{x})$ であるから，ヘッセ行列 $\boldsymbol{H}$ の (i, j) 要素と (j, i) 要素は等しく，対称行列になる．(6) 式の結果は，行列 $\boldsymbol{A}$ が対称であれば $2\boldsymbol{A}\boldsymbol{x}$ とできる ($\boldsymbol{A} = \boldsymbol{A}^\top$)．(4) 式を 0 とおいて $\Delta\boldsymbol{x}$ について解けば，(1) 式を最小化する $\Delta\boldsymbol{x}^*$ は次のようになる．

$$\Delta\boldsymbol{x}^* = -\boldsymbol{H}^{-1}\boldsymbol{g} \tag{7}$$

以上より，ニュートン法のアルゴリズムは次のようにまとめられる．

アルゴリズム 1 ニュートン法

Input: 初期値 $\breve{\boldsymbol{x}} \in \mathbb{R}^N$，収束判定に用いる閾値 $\varepsilon \ll 1$

1: $\boldsymbol{x}_1 \leftarrow \breve{\boldsymbol{x}}$ として初期化する
2: **for** $i = 1, 2, \cdots$ **do**
3: $\quad$ $\boldsymbol{x} = \boldsymbol{x}_i$ における $f(\boldsymbol{x})$ の勾配 $\boldsymbol{g} = \nabla f(\boldsymbol{x}_i)$ とヘッセ行列 $\boldsymbol{H} = \nabla^2 f(\boldsymbol{x}_i)$ を求める
4: $\quad$ $\boldsymbol{x}_i$ に対する変位を $\Delta\boldsymbol{x}^* \leftarrow -\boldsymbol{H}^{-1}\boldsymbol{g}$ として求める
5: $\quad$ $\boldsymbol{x}_{i+1} \leftarrow \boldsymbol{x}_i + \Delta\boldsymbol{x}^*$ として解を更新する
6: $\quad$ $|f(\boldsymbol{x}_i) - f(\boldsymbol{x}_{i+1})| < \varepsilon$ であれば終了する

2 ガウス・ニュートン法

ガウス・ニュートン (Gauss-Newton) 法は，関数 $f(\boldsymbol{x})$ が次のように，M 個の関数 $e_1(\boldsymbol{x}), \cdots, e_M(\boldsymbol{x})$ の二乗和で表される場合に利用できる．

$$f(\boldsymbol{x}) = \frac{1}{2}\sum_{i=1}^{M} e_i(\boldsymbol{x})^2 \tag{8}$$

例えば，M 個の入力と教師データの組 $\{(\boldsymbol{a}_1, b_1), \cdots, (\boldsymbol{a}_M, b_M)\}$ があるとして，これらのデータに当てはまるように，モデル $b = \phi(\boldsymbol{a}; \boldsymbol{x})$ のパラメータ $\boldsymbol{x} \in \mathbb{R}^N$ を決めたいとする．このとき (8) 式の $e_i(\boldsymbol{x})$ を，i 番目の入力データ $\boldsymbol{a}_i$ に対するモデルの予測値 $\phi(\boldsymbol{a}_i; \boldsymbol{x})$ と，期待される出力値 b_i との残差 $e_i(\boldsymbol{x}) = b_i - \phi(\boldsymbol{a}_i; \boldsymbol{x})$ とする．(8) 式の $f(\boldsymbol{x})$ は残差の二乗和となるが，これをパラメータ $\boldsymbol{x}$ に関して最小化すれば，データに最もよく適合するパラメータ $\boldsymbol{x}$ が得られる．グラフベース SLAM のポーズ調整であれば，$e_i(\boldsymbol{x})$ は i 番目のポーズグラフのエッジについての残差となる．

先程のニュートン法と同様に，関数 $f(\boldsymbol{x})$ を，初期値 $\breve{\boldsymbol{x}}$ からの変位 $\Delta\boldsymbol{x}$ についての関数 $f(\breve{\boldsymbol{x}} + \Delta\boldsymbol{x})$ に置き換えて考える．$f(\breve{\boldsymbol{x}} + \Delta\boldsymbol{x})$ を最小化するような $\Delta\boldsymbol{x}$ は，(7) 式のように，$f(\boldsymbol{x})$ のヘッセ行列の逆行列 $\boldsymbol{H}^{-1}$ と，勾配 $\boldsymbol{g}$ との積で表現される．残差 $e_i(\boldsymbol{x})$ の勾配 $\nabla e_i(\boldsymbol{x}) \in \mathbb{R}^N$ とヘッセ行列 $\nabla^2 e_i(\boldsymbol{x}) \in \mathbb{R}^{N \times N}$ は，j 番目の要素 $(\nabla e_i(\boldsymbol{x}))_j$ と，(j, k) 番目の要素 $\left(\nabla^2 e_i(\boldsymbol{x})\right)_{jk}$ がそれぞれ

$$(\nabla e_i(\boldsymbol{x}))_j = \frac{\partial}{\partial x_j} e_i(\boldsymbol{x}), \quad \left(\nabla^2 e_i(\boldsymbol{x})\right)_{jk} = \frac{\partial^2}{\partial x_j \partial x_k} e_i(\boldsymbol{x}) \tag{9}$$

で与えられるので，$f(\boldsymbol{x})$ の勾配 $\boldsymbol{g}$ の第 k 要素 g_k と，ヘッセ行列 $\boldsymbol{H}$ の (j, k) 要素 H_{jk} は

$$g_k = \frac{\partial}{\partial x_k} f(\boldsymbol{x}) = \sum_{i=1}^{M} e_i(\boldsymbol{x}) \frac{\partial}{\partial x_k} e_i(\boldsymbol{x}) = \sum_{i=1}^{M} e_i(\boldsymbol{x}) \left(\nabla e_i(\boldsymbol{x})\right)_k \tag{10}$$

$$
\begin{aligned}
H_{jk} &= \frac{\partial^2}{\partial x_j \partial x_k} f(\boldsymbol{x}) = \frac{\partial}{\partial x_j}\left(\sum_{i=1}^{M} e_i(\boldsymbol{x}) \frac{\partial}{\partial x_k} e_i(\boldsymbol{x})\right) \\
&= \sum_{i=1}^{M}\left(\frac{\partial}{\partial x_j} e_i(\boldsymbol{x}) \frac{\partial}{\partial x_k} e_i(\boldsymbol{x}) + e_i(\boldsymbol{x}) \frac{\partial^2}{\partial x_j \partial x_k} e_i(\boldsymbol{x})\right) \\
&= \sum_{i=1}^{M} \frac{\partial}{\partial x_j} e_i(\boldsymbol{x}) \frac{\partial}{\partial x_k} e_i(\boldsymbol{x}) + \sum_{i=1}^{M} e_i(\boldsymbol{x}) \frac{\partial^2}{\partial x_j \partial x_k} e_i(\boldsymbol{x}) \\
&= \sum_{i=1}^{M} \left(\nabla e_i(\boldsymbol{x})\right)_j \left(\nabla e_i(\boldsymbol{x})\right)_k + \sum_{i=1}^{M} e_i(\boldsymbol{x}) \left(\nabla^2 e_i(\boldsymbol{x})\right)_{jk}
\end{aligned} \tag{11}
$$

これより $f(\breve{\boldsymbol{x}} + \Delta\boldsymbol{x})$ の勾配 $\boldsymbol{g} = \nabla f(\breve{\boldsymbol{x}})$ とヘッセ行列 $\boldsymbol{H} = \nabla^2 f(\breve{\boldsymbol{x}})$ は

$$
\boldsymbol{g} = [g_1, \cdots, g_N]^\top = \sum_{i=1}^{M} e_i(\boldsymbol{x}) \nabla e_i(\breve{\boldsymbol{x}}) \tag{12}
$$

$$
\boldsymbol{H} = [H_{jk}] = \sum_{i=1}^{M} \nabla e_i(\breve{\boldsymbol{x}}) \nabla e_i(\breve{\boldsymbol{x}})^\top + \sum_{i=1}^{M} e_i(\breve{\boldsymbol{x}}) \nabla^2 e_i(\breve{\boldsymbol{x}}) \tag{13}
$$

のように書ける．ニュートン法では, 最適な $\Delta\boldsymbol{x}$ を $-\boldsymbol{H}^{-1}\boldsymbol{g}$ として求める．

ヘッセ行列 $\boldsymbol{H}$ を得るには, $e_i(\boldsymbol{x})$ に関する $N \times N$ ヘッセ行列 $\nabla^2 e_i(\boldsymbol{x})$ を M 個, 即ち $M \cdot N(N+1)/2$ 種類の 2 階微分を計算する必要がある．N や M が大きければ, ヘッセ行列を求めるのは困難である．そこで, (13) 式の $\nabla^2 e_i(\boldsymbol{x})$ を省略する．このとき $\boldsymbol{H}$ は, $e_i(\boldsymbol{x})$ に関する勾配 $\nabla e_i(\boldsymbol{x})$ のみから計算でき, ヘッセ行列 $\nabla^2 e_i(\boldsymbol{x})$ は不要になる．新たな $\tilde{\boldsymbol{H}}$ は

$$
\tilde{\boldsymbol{H}} = \sum_{i=1}^{M} \nabla e_i(\breve{\boldsymbol{x}}) \nabla e_i(\breve{\boldsymbol{x}})^\top \tag{14}
$$

であるから, 最適な $\Delta\boldsymbol{x}$ は

$$
\Delta\boldsymbol{x} = -\left(\sum_{i=1}^{M} \nabla e_i(\breve{\boldsymbol{x}}) \nabla e_i(\breve{\boldsymbol{x}})^\top\right)^{-1} \sum_{i=1}^{M} e_i(\boldsymbol{x}) \nabla e_i(\breve{\boldsymbol{x}}) \tag{15}
$$

となる．$N \times M$ 行列 $\boldsymbol{J}(\breve{\boldsymbol{x}})$ を

$$
\boldsymbol{J}(\breve{\boldsymbol{x}}) = [\nabla e_1(\breve{\boldsymbol{x}}), \nabla e_2(\breve{\boldsymbol{x}}), \cdots, \nabla e_M(\breve{\boldsymbol{x}})] \in \mathbb{R}^{N \times M} \tag{16}
$$

とおけば $\tilde{\boldsymbol{H}} = \boldsymbol{J}(\breve{\boldsymbol{x}}) \boldsymbol{J}(\breve{\boldsymbol{x}})^\top$ であるから, $\Delta\boldsymbol{x}$ は

$$
\Delta\boldsymbol{x} = -\left(\boldsymbol{J}(\breve{\boldsymbol{x}}) \boldsymbol{J}(\breve{\boldsymbol{x}})^\top\right)^{-1} \nabla f(\breve{\boldsymbol{x}}) \tag{17}
$$

のように書ける．以上より, ガウス・ニュートン法のアルゴリズムが得られる．

アルゴリズム 2 ガウス・ニュートン法

Input: 初期値 $\breve{\boldsymbol{x}} \in \mathbb{R}^N$, 収束判定に用いる閾値 $\varepsilon \ll 1$

1: $\boldsymbol{x}_1 \leftarrow \breve{\boldsymbol{x}}$ として初期化する
2: **for** $i = 1, 2, \cdots$ **do**
3: $\boldsymbol{x} = \boldsymbol{x}_i$ における勾配 $\boldsymbol{g}$ とヘッセ行列 $\tilde{\boldsymbol{H}}$ を, (12) 式と (14) 式から求める
4: $\boldsymbol{x}_i$ に対する変位を $\Delta\boldsymbol{x}^* \leftarrow -\tilde{\boldsymbol{H}}^{-1}\boldsymbol{g}$ として求める
5: $\boldsymbol{x}_{i+1} \leftarrow \boldsymbol{x}_i + \Delta\boldsymbol{x}^*$ として解を更新する
6: $|f(\boldsymbol{x}_i) - f(\boldsymbol{x}_{i+1})| < \varepsilon$ であれば終了する

$f(\boldsymbol{x})$ を構成する各残差 $e_i(\boldsymbol{x}) = e_i(\breve{\boldsymbol{x}} + \Delta\boldsymbol{x})$ を, 1 次の項までテイラー展開して近似すれば

$$e_i(\breve{\boldsymbol{x}} + \Delta\boldsymbol{x}) \simeq e_i(\breve{\boldsymbol{x}}) + \nabla e_i(\breve{\boldsymbol{x}})^\top \Delta\boldsymbol{x} \tag{18}$$

となる．これを元の関数 $f(\boldsymbol{x}) = f(\breve{\boldsymbol{x}} + \Delta\boldsymbol{x})$ に代入すれば

$$f(\breve{\boldsymbol{x}} + \Delta\boldsymbol{x}) \simeq \frac{1}{2}\sum_{i=1}^{M}\left(e_i(\breve{\boldsymbol{x}}) + \nabla e_i(\breve{\boldsymbol{x}})^\top \Delta\boldsymbol{x}\right)^2 \tag{19}$$

$$= \frac{1}{2}\sum_{i=1}^{M}\left(e_i(\breve{\boldsymbol{x}})^2 + 2e_i(\breve{\boldsymbol{x}})\nabla e_i(\breve{\boldsymbol{x}})^\top \Delta\boldsymbol{x} + \Delta\boldsymbol{x}^\top \nabla e_i(\breve{\boldsymbol{x}})\nabla e_i(\breve{\boldsymbol{x}})^\top \Delta\boldsymbol{x}\right) \tag{20}$$

を得る．$f(\breve{\boldsymbol{x}} + \Delta\boldsymbol{x})$ を $\Delta\boldsymbol{x}$ で偏微分すれば

$$\frac{\partial}{\partial \Delta\boldsymbol{x}} f(\breve{\boldsymbol{x}} + \Delta\boldsymbol{x}) = \sum_{i=1}^{M}\left(e_i(\breve{\boldsymbol{x}}) + \nabla e_i(\breve{\boldsymbol{x}})^\top \Delta\boldsymbol{x}\right)\nabla e_i(\breve{\boldsymbol{x}}) \tag{21}$$

$$= \sum_{i=1}^{M}\left(e_i(\breve{\boldsymbol{x}})\nabla e_i(\breve{\boldsymbol{x}}) + \nabla e_i(\breve{\boldsymbol{x}})\nabla e_i(\breve{\boldsymbol{x}})^\top \Delta\boldsymbol{x}\right) \tag{22}$$

となるから, 0 とおいて $\Delta\boldsymbol{x}$ について解けば, $f(\breve{\boldsymbol{x}} + \Delta\boldsymbol{x})$ を最小化する $\Delta\boldsymbol{x}$ として

$$\Delta\boldsymbol{x}^* = -\left(\sum_{i=1}^{M}\nabla e_i(\breve{\boldsymbol{x}})\nabla e_i(\breve{\boldsymbol{x}})^\top\right)^{-1}\sum_{i=1}^{M} e_i(\boldsymbol{x})\nabla e_i(\breve{\boldsymbol{x}}) \tag{23}$$

が得られ, (15) 式と同一になる．従って, ガウス・ニュートン法では, 残差 $e_i(\boldsymbol{x})$ を初期値 $\breve{\boldsymbol{x}}$ の周りで線形近似することで, 目的関数 $f(\boldsymbol{x})$ を近似している．一方, ニュートン法では, 目的関数 $f(\boldsymbol{x})$ を初期値 $\breve{\boldsymbol{x}}$ の周りで 2 次近似している．

3 レーベンバーグ・マーカート法

ガウス・ニュートン法では, 目的関数の近似である (19) 式を, (初期値 $\breve{\boldsymbol{x}}$ に加える) 変位 $\Delta\boldsymbol{x}$ に関して最小化した．(18) 式の残差を, 初期値 $\breve{\boldsymbol{x}}$ のまわりで線形近似しているため, 初期値から離れた点では近似による誤差が大きくなる ($e_i(\boldsymbol{x})$ が三角関数 sin や cos で表現される場合などに相当する)．レーベンバーグ・マーカート (Levenberg-Marquardt) 法では, ガウス・ニュートン法の目的関数 (19) 式に, 変位 $\Delta\boldsymbol{x}$ が大きくなり過ぎないための罰則項 ($\Delta\boldsymbol{x}$ の各要素の二乗和) を加えた新たな目的関数を, $\Delta\boldsymbol{x}$ に関して最小化する．

$$f(\breve{\boldsymbol{x}} + \Delta\boldsymbol{x}) + \frac{\lambda}{2}\|\Delta\boldsymbol{x}\|^2 = \frac{1}{2}\sum_{i=1}^{M}\left(e_i(\breve{\boldsymbol{x}}) + \nabla e_i(\breve{\boldsymbol{x}})^\top \Delta\boldsymbol{x}\right)^2 + \frac{\lambda}{2}\|\Delta\boldsymbol{x}\|^2 \tag{24}$$

(24) 式を $\Delta\boldsymbol{x}$ で偏微分すれば

$$\frac{\partial}{\partial\Delta\boldsymbol{x}}\left(f(\breve{\boldsymbol{x}}+\Delta\boldsymbol{x})+\frac{\lambda}{2}\|\Delta\boldsymbol{x}\|^2\right)=\left(\sum_{i=1}^{M}\left(e_i(\breve{\boldsymbol{x}})\nabla e_i(\breve{\boldsymbol{x}})+\nabla e_i(\breve{\boldsymbol{x}})\nabla e_i(\breve{\boldsymbol{x}})^\top\Delta\boldsymbol{x}\right)\right)+\lambda\Delta\boldsymbol{x} \tag{25}$$

となるから, 0 とおいて $\Delta\boldsymbol{x}$ について解けば, (24) 式を最小化する $\Delta\boldsymbol{x}$ として

$$\begin{aligned}\Delta\boldsymbol{x}^* &= -\left(\sum_{i=1}^{M}\nabla e_i(\breve{\boldsymbol{x}})\nabla e_i(\breve{\boldsymbol{x}})^\top+\lambda\boldsymbol{I}\right)^{-1}\sum_{i=1}^{M}e_i(\boldsymbol{x})\nabla e_i(\breve{\boldsymbol{x}}) &(26)\\ &= -\left(\boldsymbol{J}(\breve{\boldsymbol{x}})\boldsymbol{J}(\breve{\boldsymbol{x}})^\top+\lambda\boldsymbol{I}\right)^{-1}\nabla f(\breve{\boldsymbol{x}}) &(27)\end{aligned}$$

が得られる ($\boldsymbol{I}$ は単位行列). 罰則項に対応する項 $\lambda\boldsymbol{I}$ を, $\tilde{\boldsymbol{H}}=\boldsymbol{J}(\breve{\boldsymbol{x}})\boldsymbol{J}(\breve{\boldsymbol{x}})^\top$ の対角要素を使って, $\lambda\,\mathrm{diag}(\tilde{\boldsymbol{H}})$ とすることもある ($\mathrm{diag}(\boldsymbol{A})$ は, 行列 $\boldsymbol{A}$ の対角要素のみを残し, 他の要素を 0 で置き換えた対角行列).

$$\Delta\boldsymbol{x}^* = -\left(\boldsymbol{J}(\breve{\boldsymbol{x}})\boldsymbol{J}(\breve{\boldsymbol{x}})^\top+\lambda\,\mathrm{diag}(\tilde{\boldsymbol{H}})\right)^{-1}\nabla f(\breve{\boldsymbol{x}}) \tag{28}$$

(15) 式に示すガウス・ニュートン法の更新量 $\Delta\boldsymbol{x}$ を求める際は, ヘッセ行列 $\tilde{\boldsymbol{H}}=\boldsymbol{J}(\breve{\boldsymbol{x}})\boldsymbol{J}(\breve{\boldsymbol{x}})^\top$ の正則性 (逆行列の存在) を仮定している. 数値誤差などによって $\tilde{\boldsymbol{H}}$ の逆行列の計算が不安定になる場合があるが, $\lambda\boldsymbol{I}$ を加えることで安定化する. レーベンバーグ・マーカート法では目的関数 $f(\breve{\boldsymbol{x}}+\Delta\boldsymbol{x})$ が減少するように, λ が自動的に調節される.

$$\lambda\Delta\boldsymbol{x}^* = -\left(\frac{1}{\lambda}\boldsymbol{J}(\breve{\boldsymbol{x}})\boldsymbol{J}(\breve{\boldsymbol{x}})^\top+\boldsymbol{I}\right)^{-1}\nabla f(\breve{\boldsymbol{x}}) \tag{29}$$

上式より $\lambda\to\infty$ とすれば更新量は $\Delta\boldsymbol{x}\to-\nabla f(\breve{\boldsymbol{x}})$ となり, 勾配降下法に近づく. また (26) 式から, $\lambda\to 0$ とすれば $\Delta\boldsymbol{x}$ は (15) 式に示すガウス・ニュートン法となる. 現在の λ の下で, 解を $\breve{\boldsymbol{x}}+\Delta\boldsymbol{x}^*$ により更新し, 目的関数の値 $f(\breve{\boldsymbol{x}}+\Delta\boldsymbol{x}^*)$ が $f(\breve{\boldsymbol{x}})$ よりも増加 (悪化) してしまった場合は, λ を大きくして勾配降下法に近づける (更新量を小さくする). また目的関数の値が改善 (減少) した場合は, λ を小さくしてガウス・ニュートン法に近づける (更新量を大きくする). アルゴリズムは最初, 目的関数の値が改善し続けるので更新量が大きくなり, 最適解の近くに移動するまでは粗い探索を行う. 最適解にある程度近づいたら, 更新量が小さくなり細かな探索を行うようになる (粗く探索すると, 最適解から離れて目的関数の値が悪化してしまう). 以上より, レーベンバーグ・マーカート法のアルゴリズムが得られる.

アルゴリズム 3 レーベンバーグ・マーカート法

Input: 初期値 $\breve{\boldsymbol{x}}\in\mathbb{R}^N$, 収束判定に用いる閾値 $\varepsilon\ll 1$, λ の初期値 (0.001 など), 係数 $\rho>1(2,10$ など$)$

$\boldsymbol{x}_1\leftarrow\breve{\boldsymbol{x}}$ として初期化する
for $i=1,2,\cdots$ **do**
  $\boldsymbol{x}=\boldsymbol{x}_i$ における勾配 $\boldsymbol{g}$ とヘッセ行列 $\tilde{\boldsymbol{H}}$ を, (12) 式と (14) 式から求める
  $\boldsymbol{x}_i$ に対する変位を $\Delta\boldsymbol{x}^*\leftarrow-\left(\tilde{\boldsymbol{H}}+\lambda\boldsymbol{I}\right)^{-1}\boldsymbol{g}$ として求める
  $\boldsymbol{x}_{i+1}\leftarrow\boldsymbol{x}_i+\Delta\boldsymbol{x}^*$ として解を更新する
  $|f(\boldsymbol{x}_i)-f(\boldsymbol{x}_{i+1})|<\varepsilon$ であれば終了する
  if $f(\boldsymbol{x}_i)<f(\boldsymbol{x}_{i+1})$ **then**
    $\lambda\leftarrow\lambda\rho$ に設定
  else
    $\lambda\leftarrow\lambda\rho^{-1}$ に設定

参考文献

[1] Peter Biber and Wolfgang Straßer. The Normal Distributions Transform: A New Approach to Laser Scan Matching. In *Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pages 2743–2748, 2003.

[2] Giorgio Grisetti, Rainer Kuemmerle, Cyrill Stachniss, and Wolfram Burgard. A Tutorial on Graph-Based SLAM. *IEEE Transactions on Intelligent Transportation Systems Magazine*, 2(4):31–43, December 2010.

[3] Stefan Kohlbrecher, Oskar von Stryk, Johannes Meyer, and Uwe Klingauf. A Flexible and Scalable SLAM System with Full 3D Motion Estimation. In *Proceedings of the IEEE International Symposium on Safety, Security, and Rescue Robotics (SSRR)*, pages 155–160, 2011.

[4] Kurt Konolige, Giorgio Grisetti, Rainer Kümmerle, Wolfram Burgard, Benson Limketkai, and Regis Vincent. Efficient Sparse Pose Adjustment for 2D Mapping. In *Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pages 22–29, 2010.

[5] Rainer Kümmerle, Giorgio Grisetti, Hauke Strasdat, Kurt Konolige, and Wolfram Burgard. G2o: A General Framework for Graph Optimization. In *Proceedings of the IEEE International Conference on Robotics and Automation (ICRA)*, pages 3607–3613, 2011.

[6] 金森 敬文, 鈴木 大慈, 竹内 一郎, and 佐藤 一誠. **機械学習プロフェッショナルシリーズ　機械学習のための連続最適化**. 講談社, 2016.