適応型ステップサイズをもつ拡張カルマンフィルタ

京都大学情報学研究科　*森山 裕之　MORIYAMA Hiroyuki
山下 信雄　YAMASHITA Nobuo
福島 雅夫　FUKUSHIMA Masao

1　序論

拡張カルマンフィルタ (Extended Kalman Filter, 以下 EKF と呼ぶ) は, 次式に示される最小二乗問題を解くアルゴリズムである.

$$\begin{array}{ll} \text{minimize} & f(x)=\dfrac{1}{2}\displaystyle\sum_{i=1}^{m}\|g_i(x)\|^2. \quad (1) \\ \text{subject to} & x\in\Re^n \end{array}$$

ただし, $g_i:\Re^n\to\Re^{r_i}$ は連続微分可能とする.

最小二乗問題はカーブフィッティング, 動的システム, 階層型ニューラルネットワークの学習, パターン分類問題等 [2] 非常に広い応用をもつ. とりわけ, ニューラルネットワークの学習においては, 時間とともに生成されるデータに対して, 現在のデータを用いて学習を行うことが望ましい場合がある. そのような場合, $f(x)$ ではなく, 各 $g_i(x)$ の情報のみで反復を行えるアルゴリズムが有効になる. このようなアルゴリズムを一般に逐次型 (on-line, incremental) と呼び, 最も有名なものが, On-line Back Propagation [4] である. ここで考察する EKF もまた逐次型アルゴリズムである.

2　EKF

EKF は, 適当な初期点 x_1^1 から始めて, 点列 $\{x_1^k,\dots,x_m^k\}_{k=1,2,\dots}$ を次の式に従い生成する.

$$\begin{aligned} x_{i+1}^k &:= x_i^k-d_i^k, \quad (i=1,\dots,m) \\ x_1^{k+1} &:= x_{m+1}^k \end{aligned}$$

ここで, 修正方向ベクトル d_i^k は次式で与えられる.

$$d_i^k:=(H_i^k)^{-1}\nabla g_i(x_i^k)g_i(x_i^k). \qquad (2)$$

また, 行列 H_i^k は次式によって更新される.

$$\begin{aligned} H_i^k &:= \lambda^k H_{i-1}^k+\nabla g_i(x_i^k)\nabla g_i(x_i^k)^T && (3) \\ H_0^{k+1} &:= H_m^k && (4) \\ H_0^1 &:= \delta I && (5) \end{aligned}$$

ただし, $0<\delta\ll 1$ であり, 行列 I は 単位行列である.

Pappas [3] は, $\lambda^k\equiv\lambda<1$, $\min_x f(x)=0$ である問題に対して EKF が一次収束することを示し, さらに, $\lambda^k\equiv\lambda<1$, $\min_x f(x)>0$ のときは生成される点列が振動することを示した. また, 多くの文献で $\lambda^k\equiv 1$, $\min_x f(x)=0$ の問題に対する EKF の劣一次収束性が証明されている.

Bertsekas [1] は $\lambda^k\equiv 1$ に対する EKF の劣一次収束性を改善するために,

$$0\le 1-(\lambda^k)^m\le c/k, \quad c>0, \qquad (6)$$

を満たすようにステップサイズ λ^k を選ぶアルゴリズムを提案し, 一般的な $\min_x f(x)\ge 0$ の問題に対して, $\liminf_{k\to\infty}\|(H_i^k)^{-1}\|=0$ であることを用いて大域的収束性を証明した. ここで, (6) 式より $\lambda^k\uparrow 1$ であることに注意する.

従来, 逐次型アルゴリズムは, ステップサイズとして, $O(1/k)$ のように十分遅く 0 に収束するものを用いて大域的収束性を証明していた. これに対し, Tseng [5] は on-line BP に対して, 0 に収束しないステップサイズを用いるアルゴリズムを提案し, その大域的収束性と一次収束性を証明した.

本発表では, 文献 [1] の EKF アルゴリズムに対して, 文献 [5] で提案されたステップサイズルールを応用したアルゴリズムを考え, その大域的収束性の証明を与える.

3　提案法

文献 [1] においては, EKF の大域的収束性を示すために行列 $\nabla g_i(x)\nabla g_i(x)^T$ の正則性を仮定していたが, 階層型ニューラルネットワークの学習では, 一般にこの行列は非正則となる. そのため, 行列 H_i^k を次式によって更新することを考える.

$$H_0^{k+1}:=H_m^k+\delta I, \qquad (7)$$

また, [1] では, $\liminf_{k\to\infty}\|(H_i^k)^{-1}\|=0$ であることを用いて大域的収束性の証明を行っているが,

これは 0 に収束するステップサイズを用いることに相当しており，実用上，収束が遅くなる場合が多いと考えられる．そこで，新たにパラメータ α^k を導入することにより，$\liminf_{k\to\infty} \alpha^k \|(H_i^k)^{-1}\| > 0$ を満たすような適応型ステップサイズをもつアルゴリズムを提案する．これにより，アルゴリズムの収束性が向上することが期待できる．

提案するアルゴリズムは (2)-(5)，(6) および (7) を用いて，次式で与えられる．

$$\begin{aligned} x_{i+1}^k &:= x_i^k - \alpha^k d_i^k, \\ x_1^{k+1} &:= x_{m+1}^k. \end{aligned}$$

また，ステップサイズ α^k は

$$\alpha_*^k = \frac{(1-\epsilon_1)(x_1^{k+1}-x_1^k)^T H_m^k (x_1^{k+1}-x_1^k)}{1.5\ L\|x_{I(k)}^k - x_1^k\|^2} \tag{8}$$

に対して，$\max(1, \epsilon_2 \alpha_*^k) \le \alpha^k \le \max(1, \alpha_*^k)$ を満たすように定めるものとする．ただし，$\epsilon_1, \epsilon_2 \in (0,1)$，$I(k) \equiv \arg\max_{i=1,\ldots,m} \|x_i^k - x_1^k\|$ である．

4 大域的収束性

仮定 1-1 $\nabla g_i(x) g_i(x)$ はリプシッツ連続，つまり $\|\nabla g_i(x) g_i(x) - \nabla g_i(y) g_i(y)\| \le L\|x - y\|;\ \ \forall x, \forall y \in \Re^n$，$i = 1,\ldots,m$ を満たす $L > 0$ が存在する．

仮定 1-2 $\nabla g_i(x) g_i(x)$ は $\Re^n$ 上で有界，すなわち $\|\nabla g_i(x) g_i(x)\| \le M$，$\forall x \in \Re^n$，$i = 1,\ldots,m$，を満たす M が存在する．

アルゴリズムの大域的収束性の証明において最も重要な役割を果たす補題を示す．

補題 1 任意の k に対して行列 H_i^k の固有値が区間 $[c_1 k, c_2 k]$ にあるような 正定数 c_1，c_2 が存在する．

以上の仮定と補題から次の定理が導かれる．

定理 1 (大域的収束性) 仮定 1 のもとで，本稿で提案したアルゴリズムが生成する点列 $\{x_1^k,\ldots,x_m^k\}_{k=1,2,\ldots}$ に対して，$\{\nabla f(x_1^k)\} \to 0$ が成り立つ．つまり点列 $\{x_1^k\}_{k=1,2,\ldots}$ の全ての集積点は *(1)* の停留点である．

5 まとめと今後の課題

本発表では，適応型ステップサイズをもつ拡張カルマンフィルタを提案し，その大域的収束性を示した．その証明の過程から次のことがわかった．

補題1および (8) 式より $\alpha_*^k \ge \sigma k$ を満たす $\sigma > 0$ が存在し，さらに，$\liminf_{k\to\infty} \alpha^k \|(H_i^k)^{-1}\| > 0$ であることがいえるので，従来の EKF よりも速い収束が期待できる．しかし，行列 $\nabla g_i(x) \nabla g_i(x)^T$ が一般に非正則であるような問題（例えば，ニューラルネットワークの学習）では，補題 1 における定数 c_1 の値が δ によって定まる非常に小さな値となるので，ステップサイズが十分大きくなるためには多くの反復回数を要し，実質的には従来の EKF との相違は明らかでない．

逆に，行列 $\nabla g_i(x) \nabla g_i(x)^T$ が常に正則であり，その最小固有値が十分大きい問題に対しては，c_1 の値が比較的大きくとれるため，提案したアルゴリズムの有効性は大きいと期待できる．

今後の課題としては，数値実験による提案法の有効性の検証や，収束速度の理論的な解析があげられる．

参考文献

[1] D. P. Bertsekas, Incremental least squares methods and the extended Kalman filter, *SIAM J. Optim.*, Vol. 3, pp. 807-822, 1996.

[2] D.P. Bertsekas, *Nonlinear Programming*, Athena Scientific, Belmont MA, 1996.

[3] T. N. Pappas, *Solution of Nonlinear Equations by Davidon's Least Squares Method*, M.S. thesis, Dept. of Electrical Engineering and Computer Science, MIT, Cambridge, MA, 1982.

[4] D.E. Rumelhart, G.E. Hinton, and R.J. Williams, Learning representations by back propagating errors, *Nature*, Vol. 323, pp. 533-536, 1986.

[5] P. Tseng, Incremental gradient(-projection) method with momentum term and adaptive stepsize rule, *SIAM J. Optim.*, Vol. 8, pp. 506-531, 1998.