

Similarity between Newton's flow and Backtracking New Q-Newton's method

中部大学創発学術院 渡邊天鵬^{たかゆき}

Chubu University Academy of Emerging Sciences, Takayuki Watanabe

概要

ニュートン法は、最も有名な求根アルゴリズムである。本稿では、ニュートン法を改善するアルゴリズムとして Tuyen Trung Truong によって提案された、Backtracking New Q-Newton's method というアルゴリズムについて考える。Backtracking New Q-Newton's method は離散時間力学系と捉えられるがその吸引領域は、連続時間力学系であるニュートン・フローのそれと類似していることを数値的に発見した。これを含め、Backtracking New Q-Newton's method の局所的または大域的な性質について論じた論文 [FHTW1, FHTW2] について本稿で解説したい。本研究は John Erik Fornæss と Mi Hu, Tuyen Trung Truong との共同研究に基づく。

1 Backtracking New Q-Newton's method の大域収束性

自然数 $m \in \mathbb{N}$ と十分滑らかな関数 $f: \mathbb{R}^m \rightarrow \mathbb{R}$ に対し、 f の (局所) 最小値を求める、という最適化問題を考える。以下では例えば、複素多項式 $F(z) = z(z-i)^2$ に対して $f(x, y) = |F(x+iy)|^2/2$ と定め、 $f: \mathbb{R}^2 \rightarrow \mathbb{R}$ の局所最小値を求めるような状況を想定している。ここで、 i は虚数単位である。

この設定におけるニュートン法とは、以下のアルゴリズムを指すものとする。関数 f が最小値を取る点では f の勾配が消えるので、その点を求めることで最小値の候補を見つける。関数 f は C^2 級とし、点 a での勾配を $\nabla f(a)$ 、ヘッセ行列を $\nabla^2 f(a)$ と表記する。以下の漸化式

$$z_{n+1} = z_n - \nabla^2 f(z_n)^{-1} \nabla f(z_n)$$

で点列を (定められる限りは) 定める。もし $z_{n+1} = z_n$ かつ $\nabla^2 f(z_n)$ が可逆行列なら、 $\nabla f(z_n)$ がゼロベクトルとなり、 z_n は最小値の候補である。また、 z^* が f の最小値かつ $\nabla^2 f(z^*)$ が可逆なら、初期値 z_0 を z^* の十分近くとすれば漸化式を満たす点列は $z_n \rightarrow z^* (n \rightarrow \infty)$ となることが知られている。

ニュートン法は、大域収束性を持たないことが知られている。つまり、最小値の近くではなく別の、ある初期値からアルゴリズムを始めると f の最小値に収束せず、かつそのような初期値は非空開集合 (特にルベグ測度正) ぶん存在する。大域収束性を改善するために、Truong は Backtracking New Q-Newton's method を考案した。

Backtracking New Q-Newton's method (BNQN) とは、ニュートン法におけるヘッセ行列や、

その負の固有方向に手を加えたアルゴリズムである．まずは記号を導入したい．実行列 A を可逆な $m \times m$ 対称行列とする．それに対し，正の固有値に対応する固有ベクトルが生成する部分空間を V_+ ，負の固有値に対応するそれを V_- と書く．ベクトル空間 $\mathbb{R}^m$ から V_+ への直交射影を $pr_{A,+}$ ， V_- への直交射影を $pr_{A,-}$ と書く．また， $\text{minsp}(A)$ で A の固有値 λ 全てにおける絶対値 $|\lambda|$ の最小値を表す．以上の準備のもと，以下で Backtracking New Q-Newton's method を定義する．

Algorithm 1: Backtracking New Q-Newton's method New Variant

Result: Find a minimum of $f: \mathbb{R}^m \rightarrow \mathbb{R}$

Given: $\{\delta_0, \delta_1, \dots, \delta_m\} \subset \mathbb{R}$ $\theta \geq 0$, $0 < \tau$ and $0 < \gamma_0 \leq 1$;

Initialization: $z_0 \in \mathbb{R}^m$;

$\kappa := \frac{1}{2} \min_{i \neq j} |\delta_i - \delta_j|$;

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

$j = 0$

if $\|\nabla f(z_k)\| \neq 0$ **then**

while $\text{minsp}(\nabla^2 f(z_k) + \delta_j \|\nabla f(z_k)\|^\tau Id) < \kappa \|\nabla f(z_k)\|^\tau$ **do**

$j = j + 1$

end

end

$A_k := \nabla^2 f(z_k) + \delta_j \|\nabla f(z_k)\|^\tau Id$

$v_k := A_k^{-1} \nabla f(z_k) = pr_{A_k,+}(v_k) + pr_{A_k,-}(v_k)$

$w_k := pr_{A_k,+}(v_k) - pr_{A_k,-}(v_k)$

$\widehat{w}_k := w_k / \max\{1, \theta \|w_k\|\}$

$\gamma := \gamma_0$

if $\|\nabla f(z_k)\| \neq 0$ **then**

while $f(z_k - \gamma \widehat{w}_k) - f(z_k) > -\gamma \langle \widehat{w}_k, \nabla f(z_k) \rangle / 3$ **do**

$\gamma = \gamma / 3$

end

end

$z_{k+1} := z_k - \gamma \widehat{w}_k$

end

オリジナルのニュートン法との違いを述べたい．漸化式 $z_{k+1} := z_k - \gamma \widehat{w}_k$ を見ると，ベクトル $\widehat{w}_k$ がヘッセ行列から定まっているという点はニュートン法と共通である．しかし，ヘッセ行列 $\nabla^2 f(z_k)$ ではなくその摂動 $A_k := \nabla^2 f(z_k) + \delta_j \|\nabla f(z_k)\|^\tau Id$ ，及び A_k の負の固有方向を -1 倍したものをを用いている．摂動を行うことにより，ヘッセ行列が非可逆になったり，軌道 $\{z_n\}_{n=1}^\infty$ がサドルにトラップされたりすることを防ぐ目的がある．加えて，漸化式 $z_{k+1} := z_k - \gamma \widehat{w}_k$ に現れる γ は， $f(z_n)$ が十分早く減少するように調整している．同様の考えは backtracking line search で用いられており，いわゆる Armijo 条件 [A] の亜種である．以上のようにニュートン法を修正することにより，次の定理で述べるように大域収束性が得られる．

Theorem 1.1 ([T23]). 関数 $f: \mathbb{R}^m \rightarrow \mathbb{R}$ は C^3 級とする. 点列 $\{z_n\}$ を BNQN から定まるものとする. このとき, 任意の n で $f(z_{n+1}) \leq f(z_n)$ が成り立つ.

また, f が Morse, つまり $\nabla f(z) = 0$ となる点 z ではヘッセ行列 $\nabla^2 f(z)$ が可逆であると仮定する. このとき, $\lim_{n \rightarrow \infty} \|z_n\| = \infty$ となるか, または $\{z_n\}$ はある点 z_∞ に収束する. 後者 $\lim_{n \rightarrow \infty} z_n = z_\infty$ の場合, $\nabla f(z_\infty) = 0$ である. さらに, ルベーク測度に関してほとんど全ての初期点 z_0 に対して, $\lim_{n \rightarrow \infty} z_n = z_\infty$ ならば z_∞ は f の局所最小値であり, 収束の速度は二次収束である.

定理を理論的な背景として, BNQN を用いて generic な関数 f の (局所) 最小値を求めることができる期待される. 具体例を観察することにより, 実際に最小値を求められる場合があることを実証したい.

Example 1.2. 関数 $f: \mathbb{R}^2 \rightarrow \mathbb{R}$ を, 複素多項式 $F(z) = z(z-i)(z-3-2i)$ を用いて $f(x, y) = |F(x+iy)|^2/2$ と定義する. このとき, f の最小は $(x, y) = (0, 0), (0, 1), (3, 2)$ の三つで達成され, これらは F の零点に由来する. この f に BNQN を適用したときの大域的構造を図に描画する^{*1}.

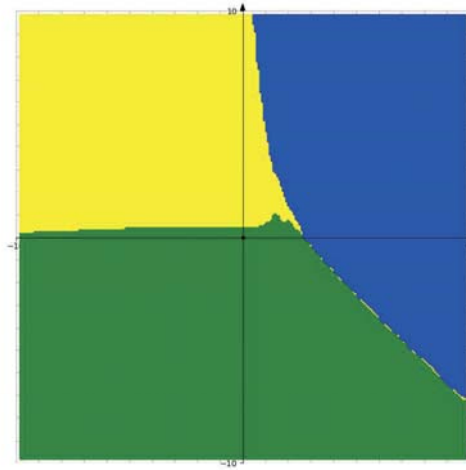


図1 各初期点に対して, BNQN で生成された軌道がどこに収束したかに応じて着色した図. 点 $(0, 0)$ に収束した初期点は緑, $(0, 1)$ に収束した初期点は黄, $(3, 2)$ に収束した初期点は青色である. つまり, それぞれの色で吸引領域を表している. 三つのうちどれにも収束しない初期点は黒色で表示するように実装しているが, そのような初期点はほとんど存在しないことが見て取れる.

^{*1} 本稿は冊子版では白黒で印刷され, web 版ではカラーで表示される. 色に関する説明は, カラー版での表示を前提としている. また, 本稿で用いた図は全て Creative Commons Attribution 4.0 International License に基づいて [FHTW1] より引用している.

上図で見られる通り、緑・黄・青の領域を合わせたほとんど全ての初期点に対して、BNQN が生成した点列は f の最小を与える点に収束している。もう一つ観察される特徴として、吸引領域の境界が比較的滑らかであることがある。オリジナルのニュートン法では、3 次以上の多項式 F に同様の図を描画すると、吸引領域の境界は必ずフラクタル的になる。吸引領域の構造は未解明である。

2 ニュートン・フローとの比較

BNQN が作る吸引領域の構造について、非フラクタル性に着目して、連続時間力学系との比較を行った。その結果、BNQN の吸引領域の構造とニュートン・フローのそれとが類似していることを数値的に発見した。まずはニュートン・フローについて説明したい。

Definition 2.1. 関数 $g: \mathbb{R}^m \rightarrow \mathbb{R}$ を C^1 級関数とする。初期条件 $z(0) = z_0$ の常微分方程式

$$\frac{dz}{dt} = -\frac{g(z)}{g'(z)}$$

およびその解 $z(t)$ を g のニュートン・フローという。

関数が $g = f$ のとき、ニュートン・フローの平衡点は f の零点である。また、初期点 z_0 が f の零点 z^* に十分近いとき、解 $z(t)$ は $t \rightarrow \infty$ で z^* に収束することが知られている。(ただし、収束速度は一次収束であり、二次収束するニュートン法より収束速度は遅い。)

実は、以下の具体例で示すように、 f を標的関数とする BNQN の吸引領域の構造は、 $g = f/f'$ のニュートン・フローのそれと似ている。

Example 2.2. 関数を $f(x, y) = |F(x + iy)|^2/2$, $F(z) = z(z - i)(z - 3 - 2i)$ とする。

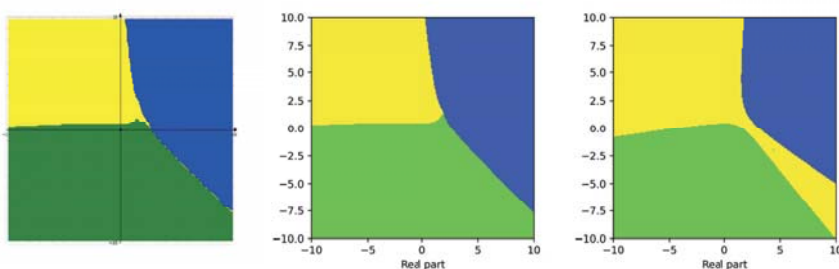


図 2 左から BNQN, f/f' のニュートン・フロー, f のニュートン・フローが生成する吸引領域の図。色は Example 1.2 と同じであり、濃い緑と薄い緑はどちらも同じ最小点に対応している。

Example 2.2 で見て取れるように、BNQN と f/f' のニュートン・フローは非常によく似ている。一方、 f/f' のニュートン・フローと f のニュートン・フローは、やや似ているものの完全に一致しているとは言い難い。ニュートン・フローで $g = f/f'$ としたときの右边は $g/g' = ff'/((f')^2 - ff'')$

であるから、平衡点は f の零点および f の臨界点である。これが f/f' と f とでニュートン・フローに差が生まれる要因である。

Example 2.2 の観察はあくまで数値的である。[FHTW1] では 25 個以上の例について計算をし、同様のことが見て取れることを確認している。また、[FHTW2] では、2 次多項式 F から定まる $f(x, y) = |F(x + iy)|^2/2$ について、BNQN の構造が Shröder 型になっていることを厳密に証明し、その結果として吸引領域の境界が直線（特に滑らか）になっていることを証明した。しかし、 F が 3 次以上の多項式のときは、吸引領域の類似性に関する数学的に厳密な結果はまだない。Example 2.2 での観察を一層深めることで、力学系理論の観点から BNQN を理解し、結果として数値計算や最適化分野に貢献できると筆者は期待している。

参考文献

- [A] Larry Armijo. Minimization of functions having Lipschitz continuous first partial derivatives. Pacific J. Math. 16(1), 1–3 (1966)
- [FHTW1] John Erik Fornæss, Mi Hu, Tuyen Trung Truong, Takayuki Watanabe. Backtracking New Q-Newton’s method, Newton’s flow, Voronoi’s diagram and Stochastic Root Finding. Complex Analysis and Operator Theory 18 (2024), 112.
- [FHTW2] John Erik Fornæss, Mi Hu, Tuyen Trung Truong, Takayuki Watanabe. Backtracking New Q-Newton’s method, Schröder’s theorem, and Linear Conjugacy. arXiv:2312.12166
- [T23] Tuyen Trung Truong. Backtracking New Q-Newton’s Method: a good algorithm for optimization and solving systems of equations. arXiv:2209.05378