日本オペレーションズ・リサーチ学会
2023 年春季研究発表会

ヘッセ行列の対角成分を用いた近接勾配法

05001372　中央大学　　*柳下 翔太郎　YAGISHITA Shotaro
05000392　電気通信大学　中山 舜民　NAKAYAMA Shummin

1.　はじめに

本発表では，以下の最適化問題に対して新たな数値解法を提案する：

$$\underset{x\in\mathbb{R}^n}{\text{minimize}}\quad F(x) \coloneqq f(x)+g(x). \tag{1}$$

ただし，$f:\mathbb{R}^n\to\mathbb{R}$ は 2 回連続微分可能，$g:\mathbb{R}^n\to(-\infty,\infty]$ は下半連続で方向微分可能，F は下に有界であるとする．

問題 (1) を解くための代表的な反復法である近接勾配法は，x^t を t 反復目の暫定解として，各反復で適切なステップサイズ $L_t>0$ の選択と

$$x^{t+1}\in\underset{y\in\mathbb{R}^n}{\operatorname{argmin}}\Big\{\nabla f(x^t)^\top(y-x^t) + \frac{L_t}{2}\|y-x^t\|_2^2+g(y)\Big\} \tag{2}$$

によって点列の更新を行う．ここで，$\|\cdot\|_2$ は ℓ_2 ノルムである．一部の g に対しては，閉形式の解を各反復で利用することができる．例えば g が ℓ_1 ノルムなどの関数の場合，子問題 (2) が変数分離の問題になることが効いてくる．一方，各反復で f の 1 次の情報しか用いていないため，収束までの反復回数が多くなってしまうこともある．これに対して，ニュートン型近接勾配法 [2] は f のヘッセ行列 $\nabla^2 f(x^t)$ の正定値な近似行列 H_t を用いて，

$$z^t\in\underset{y\in\mathbb{R}^n}{\operatorname{argmin}}\Big\{\nabla f(x^t)^\top(y-x^t) + \frac{1}{2}\|y-x^t\|_{H_t}^2+g(y)\Big\}$$

とし，適切な $\lambda_t\in(0,1]$ を用いて

$$x^{t+1}=\lambda_t z^t+(1-\lambda_t)x^t$$

とすることで点列の更新を行う．ここで，正定値行列 H に対して，$\|x\|_H=\sqrt{x^\top Hx}$ である．f の 2 次の情報まで用いているため，近接勾配法より反復回数を抑えられる傾向にある．一方，H_t が対角行列でない場合には各反復で閉形式の解を求めることは出来ない．Park ら [4] は，対角正定値行列 U^t を逐次的に更新する対角 Barzilai–Borwein 法を提案し，それを用いて

$$x^{t+1}\in\underset{y\in\mathbb{R}^n}{\operatorname{argmin}}\Big\{\nabla f(x^t)^\top(y-x^t) + \frac{1}{2}\|y-x^t\|_{U_t}^2+g(y)\Big\} \tag{3}$$

によって点列を更新するニュートン型近接勾配法 (の変種) を提案した．一般的なニュートン型近接勾配法と異なり，U_t が対角行列なので 2 次の項が変数分離可能となり，子問題 (2) と同様に一部の g に対しては子問題 (3) の解を陽に求めることができる．Park ら [4] の数値実験では，近接勾配法に対する一定の数値的な優位性は報告されてるが，理論的な優位性は知られていない．本発表では，閉形式の解を各反復で利用することができ，理論的な優位性をもつニュートン型近接勾配法を提案する．

2.　近接対角ニュートン法

以下では f のヘッセ行列の対角成分は正であるとする．我々の提案する**近接対角ニュートン法**は，$x^0\in\operatorname{dom} g$，$\eta>1$，$\beta>0$，$D_t=\operatorname{diag}(\nabla^2 f(x^t))$ を $\nabla^2 f(x^t)$ の対角成分のみを取り出した行列として

$$x^{t+1}\in\underset{y\in\mathbb{R}^n}{\operatorname{argmin}}\Big\{\nabla f(x^t)^\top(y-x^t) + \frac{1}{2}\|y-x^t\|_{\eta^k D_t}^2+g(y)\Big\}$$

によって点列を更新する．η^k の決め方としては

$$f(x^{t+1})\le f(x^t)+\nabla f(x^t)^\top(x^{t+1}-x^t) + \frac{\beta}{2}\|x^{t+1}-x^t\|_{\eta^k D_t}^2$$

を満たす最小の非負整数 k を採用する．$D_t\succeq mI$ であるような $m>0$ が存在し，∇f がリプシッツ連続であれば，上記の条件を満たす k の存在は保証される．ここで，I は単位行列を表す．D_t が対角行列であるため，Park ら [4] の手法と同様に，閉形式の解を各反復で利用できる．

3. 近接対角ニュートン法の収束性

本節では，近接対角ニュートン法の収束性について述べる．まず，近接対角ニュートン法は近接勾配法などと同様の大域的収束性をもつ．

定理 1. $0 < \beta < 1$ とする．$\mathrm{diag}(\nabla^2 f) \succeq mI$ であるような $m > 0$ が存在し，∇f がリプシッツ連続であると仮定する．このとき，$\{x^t\}$ の任意の集積点は (1) の方向停留点である．さらに，$\{x^t\}$ が有界であれば，$\lim_{t\to\infty} \mathrm{dist}(x^t, X^*) = 0$[1]が成立する．$g$ が凸の場合，同様の結果が $1 \leq \beta < 2$ に対しても成り立つ．

f が変数分離であるとき，近接対角ニュートン法は 2 次収束する．

定理 2. $1 < \beta < 2$, $f(x) = \sum_{i=1}^n f_i(x_i)$ とする．f が強凸，∇f と $\nabla^2 f$ がリプシッツ連続，g が凸であると仮定する．このとき，$\{x^t\}$ は (1) の一意な最適解に収束する．さらに，十分大きい t に対しては $k = 0$ が採用され，2 次収束する．

Li ら [3] が微分可能な目的関数 (すなわち, $g = 0$) に対して提案した対角 BFGS 法は，f が変数分離の場合でも高々局所的な超 1 次収束しか得られない．また，Park ら [4] の手法においては，収束率についての議論はされていない．これらの意味で，近接対角ニュートン法は，ヘッセ行列を対角行列で近似する既存の手法に対して理論的な優位性を備えている．現実的には，f が変数分離であるというのは強い仮定である．しかし直観的には，f が変数分離でなくとも，$\nabla^2 f$ が対角行列に近い場合に近接対角ニュートン法が有効であることが期待できる．f が強凸 2 次関数である場合，その有効性が理論的に確認できる．

定理 3. $0 < \beta \leq 1$, A を正定値行列として, $f(x) = \frac{1}{2} x^\top A x + b^\top x$, $D = \mathrm{diag}(A)$ とする．g が凸関数であると仮定し，x^* を (1) の一意な最適解とする．このとき，以下が成立する：

$$\|x^t - x^*\|_D^2 \leq \left(1 - \frac{\sigma}{\overline{\eta}\tau}\right)^t \|x^0 - x^*\|_D^2,$$

$$F(x^{t+1}) - F(x^*) \leq \frac{\overline{\eta}\tau}{2} \left(1 - \frac{\sigma}{\overline{\eta}\tau}\right)^{t+1} \|x^0 - x^*\|_D^2.$$

ここで，$\overline{\eta} = \max\{\frac{1}{\tau}, \frac{\eta}{\beta}\}$, σ は $A \succeq \sigma D$ となる最大の正の数，τ は $\tau D \succeq A$ となる最小の正の数である．

[1] X^* は (1) の方向停留点全体の集合．

一般に，正定値行列 A, B が $A \succeq B$ であることは，$\{x \mid x^\top A x = 1\}$ が $\{x \mid x^\top B x = 1\}$ の内側にあることを意味する．つまり, $\{x \mid x^\top (\sigma D) x = 1\}$ と $\{x \mid x^\top (\tau D) x = 1\}$ はそれぞれ $\{x \mid x^\top A x = 1\}$ に外側と内側から接する楕円であり，$\frac{\sigma}{\tau} \leq 1$ はその大きさの比を表す．A が対角行列であることと $\frac{\sigma}{\tau} = 1$ が同値であるため, $\frac{\sigma}{\tau}$ は A が対角行列に近いかどうかを表す指標であると解釈できる．すなわち，定理 3 は A が対角行列に近いほど速い収束が期待できることを示唆する．我々の知る限り，このような非対角行列に対しての理論的評価は，従来の対角行列を用いるニュートン型近接勾配法に対しては与えられていない．一方，近接勾配法に対しても同様の 1 次収束性を示すことができるが，その際，$\frac{\sigma}{\tau}$ に対応する部分には A の条件数の逆数が現れる [1, Theorem 10.29]．条件数の逆数は，A が単位行列の定数倍に近いかどうかを表す指標と解釈できる．よって，A が対角行列に近いが単位行列の定数倍には近くない場合，近接対角ニュートン法が近接勾配法に対して優位性を持つであろうことが示唆される．

数値実験結果等については，当日報告する．

参考文献

[1] Amir Beck. *First-order methods in optimization.* SIAM, 2017.

[2] Jason D Lee, Yuekai Sun, and Michael A Saunders. Proximal Newton-type methods for minimizing composite functions. *SIAM Journal on Optimization*, 24(3):1420–1443, 2014.

[3] Donghui Li, Xiaozhou Wang, and Jiajian Huang. Diagonal BFGS updates and applications to the limited memory BFGS method. *Computational Optimization and Applications*, 81(3):829–856, 2022.

[4] Youngsuk Park, Sauptik Dhar, Stephen Boyd, and Mohak Shah. Variable metric proximal gradient method with diagonal Barzilai–Borwein stepsize. In *2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 3597–3601. IEEE, 2020.