日本オペレーションズ・リサーチ学会
2021年 秋季研究発表会

DC問題に対する非厳密 Newton 型近接勾配法

05001310 東京工業大学 *リュウ ティエンシャン LIU Tianxiang
01308490 東京大学/理化学研究所 武田朗子 TAKEDA Akiko

1. はじめに

本発表では [4] の概要を述べる．以下の DC 問題を取り扱う．

$$\min_{\boldsymbol{x}\in\mathbb{R}^n} \quad F(\boldsymbol{x}) = f(\boldsymbol{x}) + h(\boldsymbol{x}) - g(\boldsymbol{x}). \tag{DC}$$

ここで, f は微分可能な関数で, 勾配 ∇f のリプシッツ定数は L である．h は凸関数であり, 有効定義域で連続である．また, g は連続凸関数である．

問題 (DC) を解くアルゴリズムの中で, 主なのは一次法 (例えば, [6, 7]) である．よって, 関数 f のヘッセ行列の情報を使う方法はまだ少ない．一方で, 特に $g = 0$ の場合, 多くの Newton 型近接勾配法 ([1, 5] など) が存在している．それらの文献では, ヘッセ行列の情報を使うと, 数値試験のパフォーマンスはより良くなることが報告されている．

本研究では一般的な場合 ($g \neq 0$) の問題 (DC) を解くため新たな非厳密 Newton 型近接勾配法を提案し，その収束性を解析する．

まず, g が凸より, $\boldsymbol{\xi}^{k+1} \in \partial g(\boldsymbol{x}^k)$ を用いて, 以下の子問題を構築する．

$$\begin{aligned}\min_{\boldsymbol{z}\in\mathbb{R}} G_k(\boldsymbol{z}) :=& \langle \nabla f(\boldsymbol{x}^k) - \boldsymbol{\xi}^{k+1}, \boldsymbol{z} - \boldsymbol{x}^k\rangle \\ &+ \frac{1}{2}(\boldsymbol{z}-\boldsymbol{x}^k)^\top \boldsymbol{B}_k(\boldsymbol{z}-\boldsymbol{x}^k) + h(\boldsymbol{z}).\end{aligned} \tag{Sub}$$

ここで, $\boldsymbol{B}_k$ は対称行列である．子問題の最適解が必要ではなく，近似解で十分である．具体的に, $\epsilon > 0$ が与えると, 以下の条件を満たす近似解 $\boldsymbol{u}^k$ を考える．

$$\operatorname{dist}\left(\boldsymbol{0}, \partial G_k(\boldsymbol{u}^k)\right) \le \epsilon_k \|\boldsymbol{u}^k - \boldsymbol{x}^k\|. \tag{Rule}$$

次に, (Sub) の近似解 $\boldsymbol{u}^k$ を得ると, 目的関数値 F を小さくするために, 直線探索を使う．特に, 以下の三つの直線探索を考える: $\sigma \in (0,1)$, $\beta \in (0,1)$, $M \ge 0$ は整数, $\boldsymbol{d}_k = \boldsymbol{u}^k - \boldsymbol{x}^k$ とし, ステップサイズ α_k を $\{\beta^i : i = 0, 1, \ldots\}$ の中で以下の条件を満たす最大のものとする．

$$F(\boldsymbol{x}^k + \alpha_k \boldsymbol{d}_k) \le \max_{[k-M]_+ \le j \le k} F(\boldsymbol{x}^j) + \sigma\alpha_k \triangle_{k,1}. \tag{LS$_1$}$$

ここで, $\triangle_{k,1} := \langle \nabla f(\boldsymbol{x}^k) - \boldsymbol{\xi}^{k+1}, \boldsymbol{d}_k\rangle + h(\boldsymbol{u}^k) - h(\boldsymbol{x}^k)$ である．

$$F(\boldsymbol{x}^k + \alpha_k \boldsymbol{d}_k) \le \max_{[k-M]_+ \le j \le k} F(\boldsymbol{x}^j) + \sigma \triangle_k(\alpha_k). \tag{LS$_2$}$$

ここで, $\triangle_k(\alpha_k) := \alpha_k \langle \nabla f(\boldsymbol{x}^k) - \boldsymbol{\xi}^{k+1}, \boldsymbol{d}_k\rangle + h(\boldsymbol{x}^k + \alpha_k \boldsymbol{d}_k) - h(\boldsymbol{x}^k)$ である．

$$F(\boldsymbol{x}^k + \alpha_k \boldsymbol{d}_k) \le \max_{[k-M]_+ \le j \le k} F(\boldsymbol{x}^j) - \sigma\alpha_k \|\boldsymbol{d}_k\|^2. \tag{LS$_3$}$$

上記に基づいて, 問題 (DC) に対する非厳密 Newton 型近接勾配法 (iSQA) を示す．

> Step 0. 初期点 $\boldsymbol{x}^0 \in \operatorname{dom} h$, $\beta \in (0,1)$, $\sigma \in (0,1)$, 整数 $M \ge 0$, $\{\epsilon_k\}$ を正の数列とする．また, $k = 0$ とする．
>
> Step 1. $\boldsymbol{B}_k \in \mathcal{S}^n$, $\boldsymbol{\xi}^{k+1} \in \partial g(\boldsymbol{x}^k)$ とする．子問題 (Sub) に対して, $\boldsymbol{x}^k$ を初期点として, (Rule) を満たす近似解 $\boldsymbol{u}^k$ を探す．
>
> Step 2. $\boldsymbol{d}_k = \boldsymbol{u}^k - \boldsymbol{x}^k$ とする．α_k を $\{\beta^i : i = 0, 1, \ldots\}$ の中で (LS$_1$), (LS$_2$) 又は (LS$_3$) を満たす最大のものとする．
>
> Step 3. $\boldsymbol{x}^{k+1} = \boldsymbol{x}^k + \alpha_k \boldsymbol{d}_k$ とし, $k \leftarrow k+1$ とする．そして, Step 1 に戻る．

2. 収束性解析

収束解析を行うため，任意の $k \ge 0$ に対して, $G_k(\cdot)$ が μ_k-強凸関数 ($\mu_k > 0$) と仮定する．また, 以下のパラメータを定義する．

$$\delta_k := \frac{\mu_k + \lambda_{\min}(\boldsymbol{B}_k)}{2} - \epsilon_k.$$

ここで, $\lambda_{\min}(\boldsymbol{B}_k)$ は行列 $\boldsymbol{B}_k$ の最小固有値である．

定理 1 (直線探索の有限回終了)**.** $\delta_k > 0$ のとき, 直線探索 (LS$_1$) と (LS$_2$) は有限回反復で終了する; $\delta_k > \sigma$ のとき, 直線探索 (LS$_3$) は有限回反復で終了する．さらに,

$$\alpha_k \ge \begin{cases} \min\{1,\ 2\beta(1-\sigma)\delta_k/L\}, & \delta_k > 0 : (\mathrm{LS}_1), (\mathrm{LS}_2)\text{ の場合} \\ \min\{1,\ 2\beta(\delta_k - \sigma)/L\}, & \delta_k > \delta : (\mathrm{LS}_3)\text{ の場合}\end{cases}$$

が成り立つ．

定理 2 (部分列の収束)**.** $\{\boldsymbol{B}_k\}$ が有界と仮定する．$\delta := \inf_k \delta_k$ とする．$\{\boldsymbol{x}^k\}$ はアルゴリズム (iSQA) で生成される点列とする．(LS$_1$) 又は (LS$_2$) の場合 $\delta > 0$ であれば, もしくは (LS$_3$) の場合 $\delta > \sigma$ であれば, 以下の結論が成り立つ．

(i) 点列 $\{\boldsymbol{x}^k\}$ は有界;

(ii) $\lim_{k\to\infty}\|\boldsymbol{x}^{k+1}-\boldsymbol{x}^k\|=0$;

(iii) 点列 $\{\boldsymbol{x}^k\}$ の任意の集積点は (DC) の停留点である.

次いで, アルゴリズム (iSQA) の複雑さの分析を行う. そのため, 最初に, $\xi^{k+1}\in\partial g(x^k)$ とし, $\boldsymbol{x}^k$ と最適解の "距離" は次のように定義する.

$$R_k=\left\|\boldsymbol{x}^k-\mathrm{prox}_h\left(\boldsymbol{x}^k-(\nabla f(\boldsymbol{x}^k)-\boldsymbol{\xi}^{k+1})\right)\right\|.$$

ここで, 近接写像は $\mathrm{prox}_h(\boldsymbol{y}):=\arg\min_{\boldsymbol{x}}\{h(\boldsymbol{x})+\frac{1}{2}\|\boldsymbol{x}-\boldsymbol{y}\|^2\}$ として定義される. 複雑さの分析のため, R_k を用いて, 以下の近似停留点を考慮する.

定義 1 (近似停留点)**.** $\varepsilon>0$ が与えると, ある整数 $N\geq 0$ に対して以下の不等式が満たせば, アルゴリズム (iSQA) で「ε-近似停留点が得られる」という.

$$\min_{0\leq k\leq N}R_k^2\leq\varepsilon.$$

定理 3 (定理 3.11 [4])**.** アルゴリズム (iSQA) において $M=0$ とし, 全ての k に対して, 以下の条件を満たすとする.

$$\lambda_{\min}(\boldsymbol{B}_k)-\epsilon_k>\begin{cases}0 & (\mathrm{LS}_1)\text{ 又は }(\mathrm{LS}_2)\text{ の場合;}\\ \sigma & (\mathrm{LS}_3)\text{ の場合.}\end{cases}$$

また, c_k と θ_k を次のように定義する.

$$c_k:=\frac{\sigma(\lambda_{\min}(\boldsymbol{B}_k)-\epsilon_k)\min\{1,\,2\beta(1-\sigma)\,(\lambda_{\min}(\boldsymbol{B}_k)-\epsilon_k)/L\}}{\left(1+1/\lambda_{\min}(\boldsymbol{B}_k)+\sqrt{1-2/\lambda_{\max}(\boldsymbol{B}_k)+1/\lambda_{\min}^2(\boldsymbol{B}_k)}\right)^2},$$

$$\theta_k:=\frac{\sigma\min\{1,\,2\beta(\lambda_{\min}(\boldsymbol{B}_k)-\epsilon_k-\sigma)/L\}}{\left(1+1/\lambda_{\min}(\boldsymbol{B}_k)+\sqrt{1-2/\lambda_{\max}(\boldsymbol{B}_k)+1/\lambda_{\min}^2(\boldsymbol{B}_k)}\right)^2}.$$

ここで, $\lambda_{\max}(\boldsymbol{B}_k)$ は行列 $\boldsymbol{B}_k$ の最大固有値である. $F^*=\min_{\boldsymbol{x}\in\mathbb{R}^n}F(\boldsymbol{x})$ とすると, 全ての $N\geq 0$ に対して, 以下の不等式が成り立つ.

$$\min_{0\leq k\leq N}R_k^2\leq\begin{cases}\left(F(\boldsymbol{x}^0)-F^*\right)/\sum_{k=0}^N c_k & (\mathrm{LS}_1)\text{ 又は }(\mathrm{LS}_2);\\ \left(F(\boldsymbol{x}^0)-F^*\right)/\sum_{k=0}^N\theta_k & (\mathrm{LS}_3)\text{ の場合.}\end{cases}$$

特に, $\inf_k(\lambda_{\min}(\boldsymbol{B}_k)-\epsilon_k)>\sigma$ のとき, アルゴリズム (iSQA) の複雑さは $N=O(\varepsilon^{-1})$ である.

3. 子問題の適切な終了条件

$\boldsymbol{B}_k$ が正定値行列のとき, G_k が強凸函数なので, 子問題は多くのアルゴリズムにより解ける. 例えば,

$$\phi(\boldsymbol{z}):=\langle\nabla f(\boldsymbol{x}^k)-\boldsymbol{\xi}^{k+1},\,\boldsymbol{z}-\boldsymbol{x}^k\rangle+\frac{1}{2}(\boldsymbol{z}-\boldsymbol{x}^k)^\top\boldsymbol{B}_k(\boldsymbol{z}-\boldsymbol{x}^k)$$

を定義し, $\boldsymbol{y}^1=\boldsymbol{z}^0=\boldsymbol{x}^k$ とすると, [2] におけるアルゴリズム V-FISTA の点列は以下の更新式で生成される.

$$\begin{cases}\boldsymbol{z}^\ell=\mathrm{prox}_{\frac{1}{L_\phi}h}\left(\boldsymbol{y}^\ell-\nabla\phi(\boldsymbol{y}^\ell)/L_\phi\right),\\ \boldsymbol{y}^{\ell+1}=\boldsymbol{z}^\ell+\dfrac{\sqrt{\kappa}-1}{\sqrt{\kappa}+1}(\boldsymbol{z}^\ell-\boldsymbol{z}^{\ell-1}).\end{cases}\qquad(\text{V-FISTA})$$

ここで, $L_\phi=\lambda_{\max}(\boldsymbol{B}_k)$ であり, $\kappa=L_\phi/\lambda_{\min}(B_k)$ である. (V-FISTA) の終了条件は次のように設定する.

$$\|\boldsymbol{z}^\ell-\boldsymbol{y}^\ell\|\leq\frac{\epsilon_k}{2L_\phi}\|\boldsymbol{z}^\ell-\boldsymbol{z}^0\|.\qquad(\text{Ter})$$

$\boldsymbol{z}^\ell$ と $\boldsymbol{y}^\ell$ がすでに計算されているので, 終了条件 (Ter) をチェックするための計算量は小さい. さらに, (Ter) により (Rule) を満たすことが保証される.

定理 4. $\boldsymbol{B}_k$ を正定値行列とする. $\boldsymbol{z}^\ell$ が (Ter) を満たせば, (Rule) は $\boldsymbol{u}^k=\boldsymbol{z}^\ell$ で成り立つ.

また, 文献 [3] における Semi-smooth Newton 法を子問題 (Sub) を適用する場合を想定して, (Ter) と同様の計算量が小さい終了条件を提案した.

参考文献

[1] S. Becker, J. Fadili and P. Ochs. On quasi-Newton forward–backward splitting: proximal calculus and convergence. *SIAM Journal on Optimization*, 29, 2445–2482, 2019.

[2] A. Beck. First-order methods in optimization. SIAM, 2017.

[3] X. Li, D. Sun and K. C. Toh. A highly efficient semismooth Newton augmented Lagrangian method for solving Lasso problems. *SIAM Journal on Optimization*, 28, 433–458, 2018.

[4] T. Liu and A. Takeda. An inexact successive quadratic approximation method for a class of difference-of-convex optimization problems. *Optimization online*, 2021.

[5] C. P. Lee and S. J. Wright. Inexact successive quadratic approximation for regularized optimization. *Computational Optimization and Applications*, 72, 641–674, 2019.

[6] B. Wen, X. Chen and T. K. Pong. A proximal difference-of-convex algorithm with extrapolation. *Computational Optimization and Applications*, 69, 297–324, 2018.

[7] S. J. Wright, R. D. Nowak and M. A. T. Figueiredo. Sparse reconstruction by separable approximation. *IEEE Transactions on Signal Processing*, 57, 2479–2493, 2009.