日本オペレーションズ・リサーチ学会
2022年 春季研究発表会

刈込 ℓ_1 正則化による B-スプライン回帰の節点の選択

05001372　中央大学　*柳下 翔太郎　YAGISHITA Shotaro
01110520　中央大学　後藤 順哉　GOTOH Jun-ya

1. はじめに

本発表では，B-スプライン回帰における節点選択の問題について考える．

自然数 q, l に対し $t_{-q} < t_{-q+1} < \cdots < t_0 < \cdots < t_l < \cdots < t_{l+q}$ なる $l+2q+1$ 個の値を q 次 B-スプライン関数の**節点**と呼ぶ．q 次 B-スプライン回帰とは，$[t_0, t_l] \times \mathbb{R}$ 上のデータ $\{(x_i, y_i)\}_{i=1}^n$ に q 次のスプライン基底 $B_{-q,q}, ..., B_{l-1,q}$ の線形結合で表されるモデル

$$y = \sum_{j=-q}^{l-1} \alpha_{j+q+1} B_{j,q}(x) \tag{1}$$

を適合させるノンパラメトリックな推定手法である．j 番目の q 次の基底は 0 次の基底 $B_{j,0}(x) = \mathbf{1}_{[t_j, t_{j+1})}(x)$ から，$q\,(\geq 1)$ に関する次の再帰式で定義される:

$$B_{j,q}(x) = w_{j,q}(x) B_{j,q-1}(x) + (1 - w_{j+1,q}(x)) B_{j+1,q-1}(x)$$

ここで，$w_{j,q}(x) = \frac{x - t_j}{t_{j+q} - t_j}$ である．

以上の手続きから分かるように，固定された q に対し，スプライン関数 (1) は基底の個数や基底を定義する節点によって規定される．基底の個数や節点を潤沢に用意すればより柔軟なモデル化を許すことができるが，過学習を抑えるためにこれらの自由度を抑える必要がある．これまでも節点の数を抑え適切に配置することを目指したヒューリスティクスが複数提案されている．

本発表ではまず，B-スプライン回帰における節点の選択が，$\boldsymbol{\alpha} := (\alpha_1, ..., \alpha_{l+q})^\top$ のある線型式のスパース性の制御によって達成できることを示し，それを制約として直接取り入れた定式化の等価な非凸最適化アプローチを提案する．また，変数変換を施すことによって問題が扱いやすくなることを示し，近接勾配法に基づくアルゴリズムを提案する．最後に，提案手法の有効性を数値実験により検討する．

2. 定式化

簡単のため，以下では節点が等間隔に配置されていると仮定する．$\mathbb{R}^p$ 上の k 階差分行列を $\boldsymbol{D}^{(k,p)} \in \mathbb{R}^{(p-k)\times p}$ とおき，$\boldsymbol{D}^{(q+1,l+q)}\boldsymbol{\alpha}$ の j 番目の要素を $\Delta^{q+1}\alpha_j$ で表す．このとき，以下が成り立つ．

命題 1. 節点 t_j が区分点となっていないことと，$\Delta^{q+1}\alpha_j = 0$ は同値である．

この事実は [2] によって証明なしに言及されているが，証明を与えることができる．この事実に基づいて，$\boldsymbol{y} := (y_1, ..., y_n)^\top$，$\boldsymbol{B} := (B_{ij})$，$B_{ij} := B_{j-q-1,q}(x_i)$ として，次のような ℓ_0 制約付き最小二乗推定を考える:

$$\begin{aligned} \underset{\boldsymbol{\alpha}}{\text{minimize}} \quad & \frac{1}{2}\|\boldsymbol{y} - \boldsymbol{B}\boldsymbol{\alpha}\|_2^2 \\ \text{subject to} \quad & \|\boldsymbol{D}^{(q+1,l+q)}\boldsymbol{\alpha}\|_0 \leq K. \end{aligned} \tag{2}$$

ただし，$\|\cdot\|_0$ は ℓ_0 擬ノルム，$\|\cdot\|_2$ は ℓ_2 ノルム，K は $K \leq l-1$ なる非負整数である．命題 1 より ℓ_0 制約を直接課すことで，節点を K 個以下に制限しながら，データに適合する滑らかな区分 q 次多項式を見つけることができる．しかし, (2) は大域最適性を保証付きした求解が難しい問題である．また ℓ_0 制約の左辺は $\boldsymbol{\alpha}$ の不連続関数であり, 解析的にも扱い難さが残る．そこで本研究では，$\gamma > 0$ として，次の無制約問題を考える：

$$\underset{\boldsymbol{\alpha}}{\text{minimize}} \quad \frac{1}{2}\|\boldsymbol{y} - \boldsymbol{B}\boldsymbol{\alpha}\|_2^2 + \gamma T_K(\boldsymbol{D}^{(q+1,l+q)}\boldsymbol{\alpha}). \tag{3}$$

ただし，$\boldsymbol{z} \in \mathbb{R}^{l-1}$，$[l-1] := \{1, ..., l-1\}$ として，

$$T_K(\boldsymbol{z}) := \min_{\substack{\Lambda \subset [l-1] \\ |\Lambda| = l-1-K}} \sum_{j\in\Lambda} |z_j|$$

であり，この関数は**刈込 ℓ_1 正則化**と呼ばれる方向微分可能で非凸な連続関数である．$T_K(\boldsymbol{D}^{(q+1,l+q)}\boldsymbol{\alpha}) \geq 0$ であり，$T_K(\boldsymbol{D}^{(q+1,l+q)}\boldsymbol{\alpha}) > 0$ と $\|\boldsymbol{D}^{(q+1,l+q)}\boldsymbol{\alpha}\|_0 > K$ が等価であるので，(3) の第 2 項は ℓ_0 制約に対する罰則関数の役割を果たす．関連して，ある $\bar{\gamma}$ が存在し，$\gamma > \bar{\gamma}$ ならば (3) の任意の方向停留点が (2) の ℓ_0 制約を満たすこと，さらに (3) において，方向停留点と局所最適解が同値であることも示すことができる．また，この問題に交互方向乗数法 (ADMM) を適用した場合，局所最適解への収束性が保証できる [3]．つまり，ℓ_0 制約問題の局所最適解を得ることができる．

しかしながら，ADMM を適用する場合には，実用上の難しさが現れる．これを回避するために，次節で説明する変数変換を施し，それに対して近接勾配法に基づくアルゴリズムを提案する．

3. 変数変換と近接勾配法

変数変換のために，$r = l + q$ として，正方行列 $\hat{\boldsymbol{D}}^k \in \mathbb{R}^{r\times r}$ を再帰式

$$\hat{\boldsymbol{D}}^k = \boldsymbol{D}'_k \hat{\boldsymbol{D}}^{k-1}$$

によって定義する．ただし，

$$\boldsymbol{D}'_k := \begin{pmatrix} \boldsymbol{D}^{(1,r-k+1)} & \boldsymbol{O}_{r-k,k-1} \\ \boldsymbol{O}_{k,r-k+1} & \boldsymbol{O}_{k,k-1} \end{pmatrix} + \begin{pmatrix} \boldsymbol{O}_{r-k,r-k} & \boldsymbol{O}_{r-k,k} \\ \boldsymbol{O}_{k,r-k} & \boldsymbol{I}_k \end{pmatrix}$$

かつ $\hat{\boldsymbol{D}}^1 := \boldsymbol{D}'_1$ で，$\boldsymbol{I}_k$ は k 次元単位行列，$\boldsymbol{O}_{k,l}$ は k 行 l 列の零行列である．$\boldsymbol{D}'_k$ は，1 階差分行列 $\boldsymbol{D}^{(1,r-k+1)}$ に適当な行と列を追加した正方行列である．このとき，$\boldsymbol{D}'_k$ が正則であることから $\hat{\boldsymbol{D}}^k$ も正則であり，$\hat{\boldsymbol{D}}^{q+1}$ の $l-1$ 行目までが $\boldsymbol{D}^{(q+1,r)}$ と一致する．これにより，

$$\boldsymbol{C} := \begin{pmatrix} \boldsymbol{I}_{l-1} & \boldsymbol{O}_{l-1,q+1} \end{pmatrix}$$

とし，さらに $\boldsymbol{\beta} := \hat{\boldsymbol{D}}^{q+1}\boldsymbol{\alpha}$ とおき，$\boldsymbol{A} := (\hat{\boldsymbol{D}}^{q+1})^{-1}$ とすると，問題 (3) は以下に帰着される．

$$\underset{\boldsymbol{\beta}}{\text{minimize}} \quad \frac{1}{2}\|\boldsymbol{y} - \boldsymbol{B}\boldsymbol{A}\boldsymbol{\beta}\|_2^2 + \gamma T_K(\boldsymbol{C}\boldsymbol{\beta}). \tag{4}$$

$\boldsymbol{D}'_k$ の逆行列が陽に表せることから，$\boldsymbol{A}$ は $(\boldsymbol{D}'_1)^{-1} \cdots (\boldsymbol{D}'_{q+1})^{-1}$ として簡単に得られることに注意しておく．この問題と ℓ_0 制約付き問題 (2) との関係として，次の命題が成り立つ．

定理 1. $\boldsymbol{\beta}^*$ を問題 (4) の局所最適解とする．$\gamma > \sigma_{\max}(\boldsymbol{B}\boldsymbol{A}) \cdot \|\boldsymbol{y}\|_2$ ならば，$\boldsymbol{A}\boldsymbol{\beta}^*$ は問題 (2) の ℓ_0 制約を満たす．

ここで，$\sigma_{\max}(\boldsymbol{B}\boldsymbol{A})$ は $\boldsymbol{B}\boldsymbol{A}$ の最大特異値である．この命題より，問題 (4) の（局所）最適解が得られれば，それは問題 (2) の（局所）最適解であることがわかる．

一般の問題 $\min_{\boldsymbol{\beta}} f(\boldsymbol{\beta}) + g(\boldsymbol{\beta})$ に対する近接勾配法は，ステップサイズ $\eta_t > 0$ として，各反復 t で

$$\boldsymbol{\beta}^{t+1} \in \operatorname{prox}_{\frac{g}{\eta_t}}\left(\boldsymbol{\beta}^t - \frac{1}{\eta_t}\nabla f(\boldsymbol{\beta}^t)\right)$$

のように点列を生成する反復法である．ただし，

$$\operatorname{prox}_h(\boldsymbol{x}) := \underset{\boldsymbol{y}}{\operatorname{argmin}} \left\{ h(\boldsymbol{y}) + \frac{1}{2}\|\boldsymbol{y} - \boldsymbol{x}\|_2^2 \right\}$$

であり，これは f の近接写像と呼ばれる．(3) に近接勾配法を適用すると $T_K \circ \boldsymbol{D}^{(q+1,r)}$ の近接写像（の 1 点）を求めることが必要になるが，これを陽に求めることは難しい．一方，$T_K \circ \boldsymbol{C}$ の近接写像（の 1 点）は求めることができる．ことから，変数変換後の問題 (4) に対しては，$f = \frac{1}{2}\|\boldsymbol{y} - \boldsymbol{B}\boldsymbol{A} \cdot\|_2^2$, $g = \gamma T_K(\boldsymbol{C}\cdot)$ として，近接勾配法を適用することができる．また，変数変換前の問題 (3) に ADMM を適用した場合，各反復で線型方程式を解く必要があるが，近接勾配法においてはそれが不要になることも，変数変換の利点である．本発表では，近接勾配法におけるステップサイズの選択に Barzilai-Borwein 法と非単調な直線探索を取り入れた GIST [4] を問題 (4) に対する解法として提案する．既存の GIST に対する収束性の結果 [5, 定理 1] を適用することで以下を得る．

定理 2. $\{\boldsymbol{\beta}^t\}$ を *GIST* によって生成された点列とし，有界であると仮定する．このとき，$\{\boldsymbol{\beta}^t\}$ の任意の集積点は問題 (4) の局所最適解である．

加えて，問題 (4) における GIST によって生成された点列の有界性は，以下の命題の仮定のもとで保証することができる．

定理 3. $\boldsymbol{B}\boldsymbol{A}$ の第 $1, \ldots, l-1$ 列のうちの任意の K 列と第 $l, \ldots, r$ 列が線型独立であると仮定する．このとき，*GIST* によって生成された点列 $\{\boldsymbol{\beta}^t\}$ は有界である．さらに，$\lim_{t\to\infty} \operatorname{dist}(\boldsymbol{\beta}^t, \mathcal{B}_{\mathrm{lopt}}) = 0$ である．

ここで，$\mathcal{B}_{\mathrm{lopt}}$ は問題 (4) の局所最適解の集合であり，$\operatorname{dist}(\boldsymbol{x}, \mathcal{X}) := \inf_{\boldsymbol{y}\in\mathcal{X}} \|\boldsymbol{y} - \boldsymbol{x}\|_2$ である．これらの結果より，ℓ_0 制約付き問題 (2) の局所最適解を得られることがわかる．

数値実験結果等については，当日報告する．

参考文献

[1] C. De Boor. A practical guide to splines, Vol. 27, New York: Springer-Verlag, (1978).

[2] V. Goepp, O. Bouaziz, and G. Nuel. Spline regression with automatic knot selection, arXiv preprint 1808.01770, (2018).

[3] S. Yagishita, J. Gotoh, Pursuit of the Cluster Structure of Network Lasso: Recovery Condition and Non-convex Extension, arXiv preprint 2012.07491, (2020).

[4] P. Gong, C. Zhang, Z. Lu, J.Z. Huang, J. Ye, A General Iterative Shrinkage and Thresholding Algorithm for Non-convex Regularized Optimization Problems, in Proceedings of the 30th International Conference on Machine Learning, (2013), 37–45.

[5] S. Nakayama, and J. Gotoh. On the superiority of PGMs to PDCAs in nonsmooth nonconvex sparse regression, Optimization Letters, (2021), 1–30.