制約付き非凸スパース最適化問題に対する効率のよいアルゴリズム

田中未来　　数理・推論研究系 数理最適化グループ, 統計的機械学習研究センター 助教

1 考える問題

制約領域 S 上での損失関数 $l(x)$ とスパース正則化項 $r(x)$ の和の最小化

$$\text{最小化}\ l(x)+r(x)\ \text{制約条件}\ x\in S \tag{1}$$

仮定 1 $l(x):\mathbb{R}^n\to\mathbb{R}$

- 連続的微分可能 (非凸でも OK)
- ある $L\in\mathbb{R}_+$ が存在して $(L/2)\|x\|_2^2-l(x)$ が凸関数 ($\Longleftarrow \nabla l$ が L-Lipschitz 連続)

仮定 2 $r(x):\mathbb{R}^n\to\mathbb{R}$

- 連続 (非凸でも非平滑でも分離不能でも OK)
- ある凸関数 $\phi(x)$ を用いて $r(x)=\lambda\|x\|_1-\phi(x)$ と書ける

仮定 3 $l(x)+r(x)$: 下に有界かつ強圧的 ($\Longrightarrow$ 収束定理)

仮定 4 $S\subseteq\mathbb{R}^n$: 非空, 凸, “カンタン” ($\Longrightarrow$ 効率よい計算)

多くのスパース正則化関数 $r(x)$ は仮定 2 を満たす: ℓ_1 ノルム罰則 (Tibshirani, 1994), SCAD (Fan and Li, 2001), LSP (Candés et al., 2008), MCP (Zhang, 2010a), capped ℓ_1 罰則 (Zhang, 2010b), ℓ_{1-2} 罰則 (Yin et al., 2015), ℓ_0 ノルム制約の正確な罰則化 (Gotoh et al., 2017)

2 提案手法

次のように凸関数 $g(x), h(x)$ を定める (Gotoh et al. (2017) と同様)

$$g(x)=\frac{L}{2}\|x\|_2^2+\lambda\|x\|_1+\delta(x\mid S),\quad h(x)=\frac{L}{2}\|x\|_2^2-l(x)+\phi(x)$$

問題 (1) は次の DC (difference of convex functions) 最適化問題と等価

$$\text{最小化}\ g(x)-h(x) \tag{2}$$

問題 (2) を DC アルゴリズム (Pham Dinh and Le Thi, 1997) で解く

Algorithm 1 問題 (1) に対する DC アルゴリズム

1: 適当な初期点 $x^{(0)}\in\mathbb{R}^n$ をとる
2: **for** $t=0,1,\ldots$ (収束するまで)
3: 　h の劣勾配 $s^{(t)}\in\partial h(x^{(t)})$ を求める
4: 　$x^{(t+1)}\in\operatorname{argmin}_x\{g(x)-(s^{(t)})^\top x\}$ と更新する

各反復で解く子問題 (“カンタン” := これが効率よく解ける)

$$\text{最小化}\ \frac{1}{2}\|x-v\|_2^2+\gamma\|x\|_1\ \text{制約条件}\ x\in S \tag{3}$$

3 “カンタン” な S に対応する応用例と子問題 (3)

制約領域 S	応用例	子問題 (3) に対するアルゴリズム
箱型制約 $S_\square:=\{x\in\mathbb{R}^n:\underline{x}\le x\le\bar{x}\}$	符号/箱型制約付き線形/ロジスティック回帰	閉じた解 $O(n)$ 時間
ℓ_2 ノルム球制約 $S_\bigcirc:=\{x\in\mathbb{R}^n:\|x\|_2^2\le 1\}$	主成分分析	閉じた解 $O(n)$ 時間
標準単体制約 $S_\triangle:=\{x\in\mathbb{R}^n:\mathbf{1}^\top x=1,x\ge\mathbf{0}\}$	標準単体上の最小 2 乗問題	$S_\triangle$ への射影 $O(n)$ 時間 (Kiwiel, 2008)
箱型 + 単一線形制約 $S_\boxminus:=\{x:a^\top x=b,\underline{x}\le x\le\bar{x}\}$	C-SVM の双対問題, 平均・分散ポートフォリオ最適化問題, 滑らかな非線形連続ナップサック問題	2 分探索 多項式時間 (ϵ 近似)

4 似たような問題に対する既存手法

手法	問題 (1) との差異	問題 (1) への適用可能性
Gong et al. (2013)	r が分離可能,$S=\mathbb{R}^n$	分離不能な場合は不可
Gotoh et al. (2017)	$r(x)=\lambda\|x\|_1-\lambda\|x\|_{1,k}$, 暗に $S=\mathbb{R}^n$ を仮定	$S\neq\mathbb{R}^n$ の場合の効率の良い算法は未解決 (当時)
Tono et al. (2017)	$r(x)=\lambda\|x\|_2^2-\lambda\|x\|_{2,k}^2$, S への射影の計算が容易	$r(x)=\lambda\|x\|_1-\phi(x)$ の場合は不可

5 計算機実験

5.1 汎用ソルバとの比較 (スパース双対 C-SVM)

$$\text{最小化}\ \frac{1}{2}x^\top Kx-\mathbf{1}^\top x+r(x)\ \text{制約条件}\ a^\top x=0,\mathbf{0}\le x\le C\mathbf{1}$$

問題例		提案手法		MATLAB の `fmincon`	
名前	n	目的関数値	時間 (秒)	目的関数値	時間 (秒)
a3a	3185	−0.5114	0.0452	−0.0577	601.1290
a5a	6414	−0.7380	0.1485	−0.2064	22973.7390
a7a	16100	−1.1179	2.9796	—	—
a9a	32561	−1.6711	2.5477	—	—
diabetes	768	−0.1595	0.0027	−0.0595	5.6664
mushrooms	8124	−0.2506	1.8626	−0.0647	4021.3134
phishing	11055	−0.8958	0.0522	—	—

5.2 既存の DC アプローチとの比較 (スパース PCA)

$$\text{最小化}\ -x^\top\boldsymbol{\Sigma}x\ \text{制約条件}\ \|x\|_2^2\le 1,\|x\|_0\le k$$

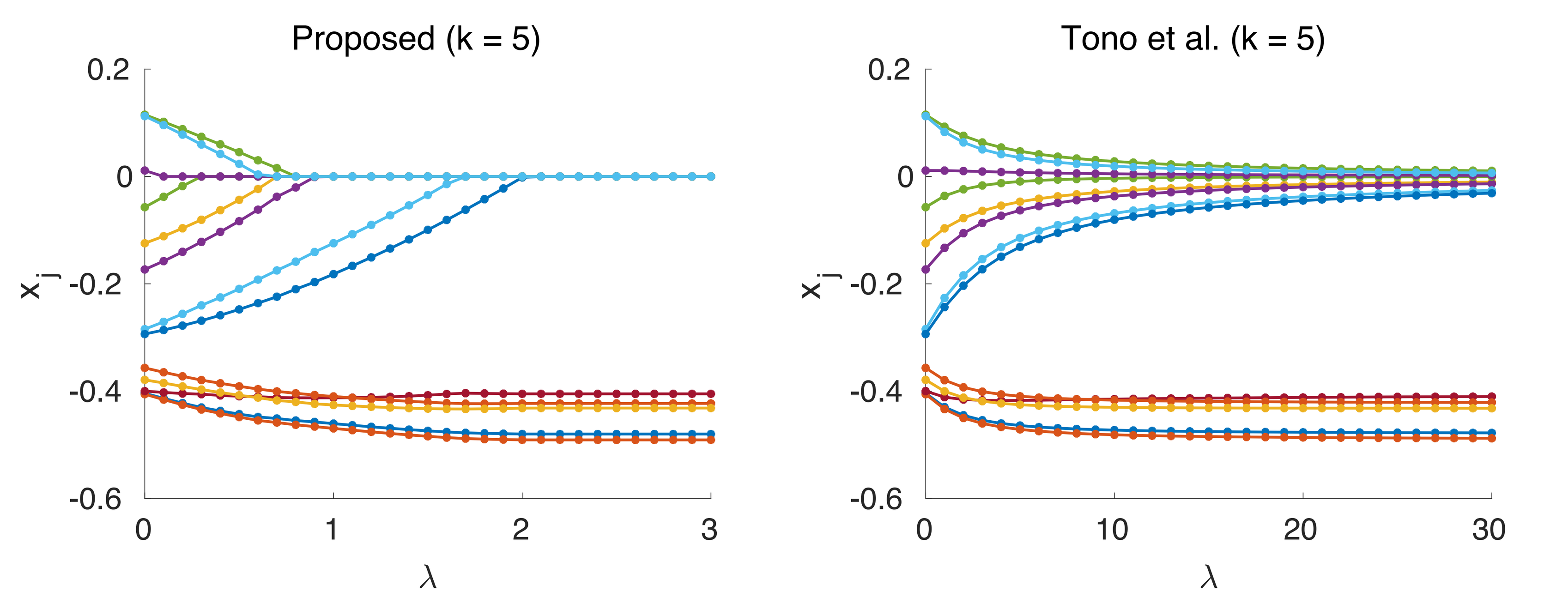

図 1: 解パスの比較 (小規模データ (Jeffers, 1967))

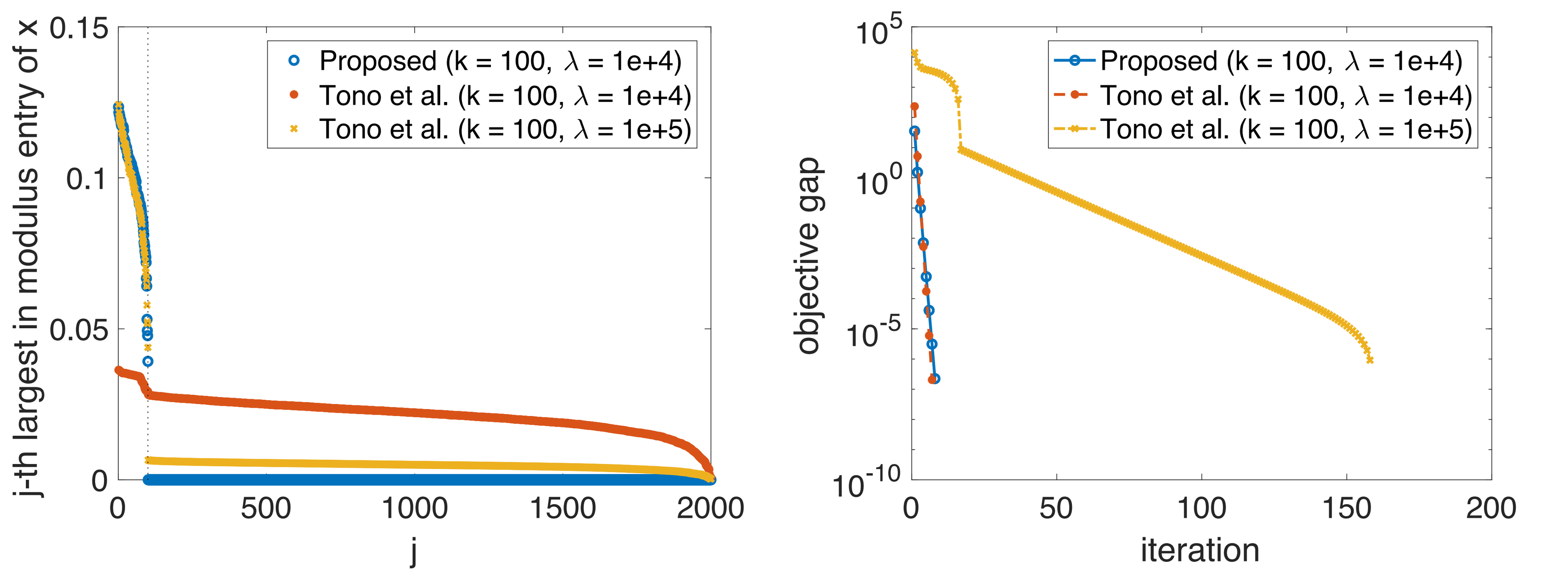

図 2: 収束先と収束の様子の比較 (中規模データ (Alon et al., 1999))

謝辞: 本研究は武田朗子氏 (東大/理研) との共同研究です.