

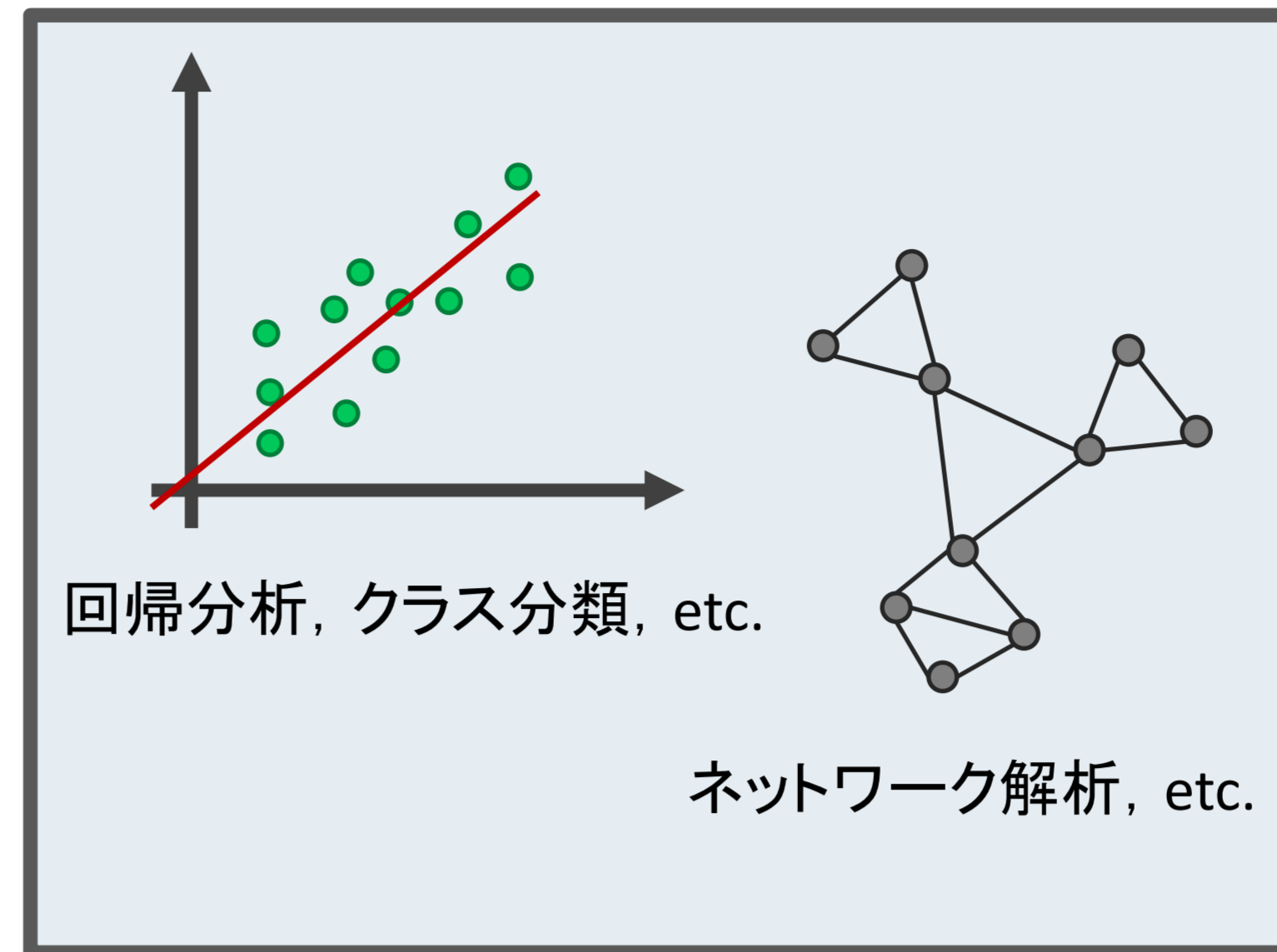
数理最適化(主に**連続最適化**)に関する研究

機械学習やデータマイニングから生じる最適化問題をうまく解き、実社会の問題解決に役立てたい

チームの強み:

- 非平滑・非凸最適化問題に対する効率的解法
- 大規模問題に対するランダム射影を用いた効率的解法
- 多段階最適化問題に対する理論保証つき解法

機械学習やデータマイニング



モデル化

数理最適化問題
 $\min_x f(x)$
 s.t. $g_1(x) \leq 0,$
 $g_2(x) \leq 0,$
 ...

解法研究(例):

- 非平滑・非凸最適化法
- 多段階最適化法
- ランダム射影法

チームの得意とするところ

ランダム射影を用いた非凸2次最適化法

[T. Fuji, P.L. Poirion, A. Takeda, SIOPT, 2021]

問題設定: 非凸2次最適化問題

$$\mathcal{P} \equiv \min_{x \in \mathbb{R}^n} \{x^\top Qx + c^\top x \mid Ax \leq b\},$$

where $Q \in \mathbb{R}^{n \times n}, A \in \mathbb{R}^{m \times n}, c \in \mathbb{R}^n, b \in \mathbb{R}^m$

- 非凸性から、一般に最適解を求めることは困難
- 変数の数 n が大きいときは既存数値解法の適用も難しい

ランダム射影P+凸化で

変数の数が小さく ($n \ll d$), かつ凸な問題に帰着

$$\mathcal{P}' \equiv \min_{u \in \mathbb{R}^d} \{u^\top F^+(PQP^\top)u + (Pc)^\top u \mid AP^\top u \leq b\}$$

where $P \in \mathbb{R}^{d \times n}, F^+$: 半正定値錐への射影 効率的に解ける

定理:

ある条件を満たすように d を適切にとれば高確率で次の不等式が成り立つ:

$$\left(1 - \frac{\varepsilon \|x^*\|_2}{r + \varepsilon \|x^*\|_2}\right) \left(1 + \frac{3\varepsilon_1 + 2\varepsilon_2 + \varepsilon_3}{\cos \theta^*}\right) \text{opt}(\mathcal{P}) \geq \text{opt}(\mathcal{P}') \geq \text{opt}(\mathcal{P})$$

カーネルSVM

表1 \mathcal{P} achieved 66.70% training-accuracy and 53.00% test-accuracy

-データはmnistの

「0」「1」画像

-シン普森カーネル

$$\mathcal{P} \begin{cases} \min_{x \in \mathbb{R}^n} & x^\top YKYx - 2x^\top \mathbf{1} \\ & 0 \leq x \leq C\mathbf{1} \\ & x^\top y = 0 \end{cases}$$

where $K \in \mathbb{R}^{n \times n}$,
不定値カーネル行列
($n = 1000$)

d	Training Accuracy (%)	Test Accuracy (%)
300	55.09 ± 14.40	53.43 ± 12.80
400	64.88 ± 15.64	62.80 ± 14.14
500	89.00 ± 12.61	83.30 ± 11.56
600	95.00 ± 0.85	90.23 ± 1.54
700	92.56 ± 1.60	86.53 ± 2.95
800	84.26 ± 3.09	76.60 ± 4.27
900	72.44 ± 3.20	60.83 ± 4.19
1000	66.96 ± 0.63	53.60 ± 0.78

非平滑2段階最適化によるLpハイパーパラメータ学習

[T. Okuno, A. Takeda, A. Kawana, M. Watanabe, JMLR, 2021]

問題設定: L_p 正則化ハイパーパラメータ選択 ($0 < p \leq 1$)

Lpスパース最適化:

$$\min_{\hat{w}} \|A\hat{w} - b\|^2 + \lambda_1 \sum_{i=1}^n |\hat{w}_i|^p$$

ハイパーパラメータ

- ハイパーパラメータの値によって解 \hat{w} も異なる
- テスト誤差 $f(w)$ を小さくするハイパーパラメータが望ましい

2段階最適化問題への定式化

2段階最適化問題(解くべき問題)

$$\min_{w, \lambda} f(w) \quad \text{連続的微分可能}$$

$$\lambda = (\lambda_1, \bar{\lambda})^\top \geq 0$$

$$w \in \operatorname{argmin}_{\hat{w}} R(\hat{w}, \bar{\lambda}) + \lambda_1 \sum_{i=1}^n |\hat{w}_i|^p$$

$$\bar{\lambda} := (\lambda_2, \dots, \lambda_m)^\top$$

$$w := (w_1, \dots, w_n)^\top$$

- 制約条件の中に別の最適化問題(下位レベル最適化問題)がある
- 本問題は下位レベル最適化問題が微分不可能で非常に取り扱いにくい
- 既存研究は少なく, 理論的な既存結果は皆無

提案アルゴリズム(平滑化法)

下位レベルの問題を下で置き換えた2段階最適化問題を近似的に解くことを繰り返しながら, 同時に $\mu \rightarrow 0+$ とする.

$$w \in \operatorname{argmin}_{\hat{w}} R(\hat{w}, \bar{\lambda}) + \lambda_1 \sum_{i=1}^n (|\hat{w}_i| + \mu)^{\frac{p}{2}}$$

主結果(上の問題の理論的解析は本研究が初)

- 上の最適化問題の最適解が満たすべき理論的性質(SB-KKT条件)を発見
- 提案手法によって生成した点列が適当な仮定の下でSB-KKT条件を満たす点に収束することを数学的に証明
- 実データを用いた数値実験で, ベイズ最適化に比べて提案手法は精度も高く, 高速であることを確認

多レベル最適化問題に対する勾配法 [R. Sato, M. Tanaka, A. Takeda, NeurIPS, 2021]

多レベル最適化問題: 一部の 변수が別の最適化問題の最適解になっているという構造が多重に連なった最適化問題

$$\min_{x_1 \in S_1 \subseteq \mathbb{R}^{d_1}, x_2^*, \dots, x_n^*} f_1(x_1, x_2^*, \dots, x_n^*) \text{ s.t.}$$

$$x_2^* = \operatorname{argmin}_{x_2 \in \mathbb{R}^{d_2}, x_3^*, \dots, x_n^*} f_2(x_1, x_2, x_3^*, \dots, x_n^*) \text{ s.t.}$$

...

$$x_n^* = \operatorname{argmin}_{x_n \in \mathbb{R}^{d_n}} f_n(x_1, x_2, \dots, x_n)$$

主結果

- ✓ 多レベル最適化問題を1レベル最適化問題で近似する方法を提案
 - 下位の問題に反復解法をT反復適用して式変形
 - $T \rightarrow \infty$ で元の問題に収束
- ✓ 近似後の問題の目的関数の勾配を効率よく計算する方法を提案
- ✓ 近似後の問題に射影勾配法を適用した際の収束性を証明(初の理論保証付きアルゴリズム)

反復法で近似 $\min_{x_1 \in S_1} f_1(x_1, x_2^{(T)}, \dots, x_n^{(T)})$

反復解法に対応する写像 $x_i^{(t+1)} = \Phi_i^{(t)}(x_1, x_2^{(T)}, \dots, x_{i-1}^{(t)}, x_i^{(t)})$

式変形 $(i = 2, \dots, n; t_i = 0, \dots, T-1)$

近似的に等価な1レベル最適化問題

$$\min_{x_1 \in S_1} \tilde{F}_1(x_1)$$

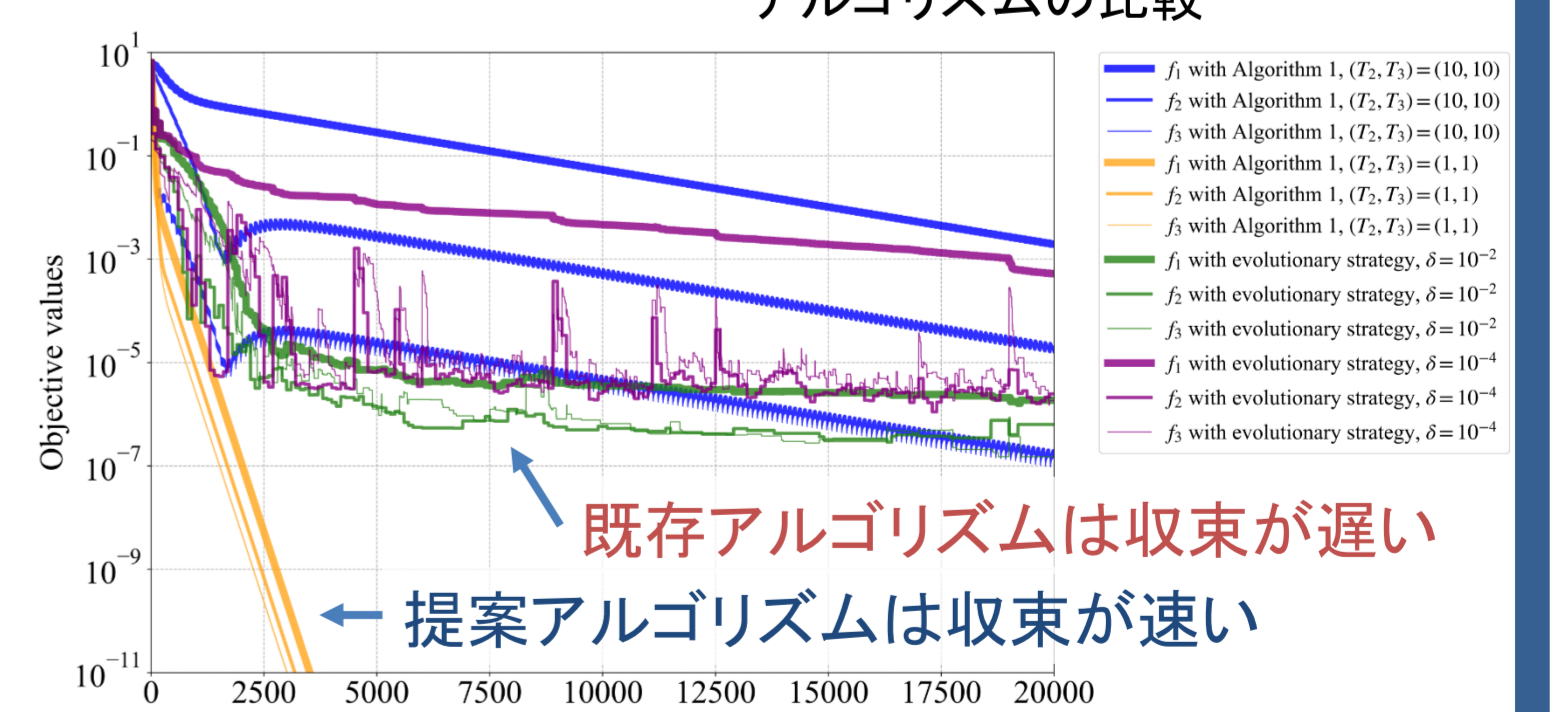
射影勾配法

$$x_1^{(t+1)} = \operatorname{proj}_{S_1}(x_1^{(t)} - \alpha_1^{(t)} \nabla \tilde{F}_1(x_1))$$

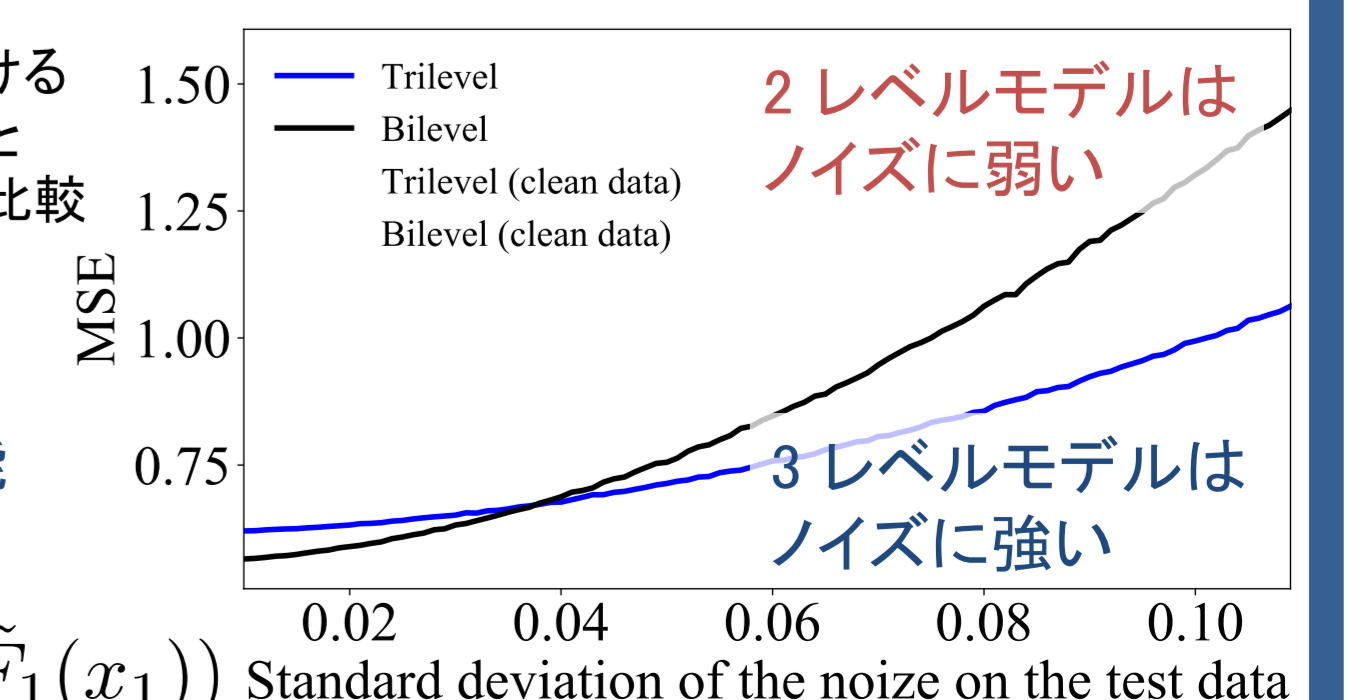
効率よく計算可能

計算機実験結果

最適解が既知の問題を用いたアルゴリズムの比較



ハイパーパラメータ選択における2レベルモデル(非ロバスト)と3レベルモデル(ロバスト)の比較



- ✓ ロバストハイパーパラメータ選択問題や多段階ロバスト最適化問題などを表現可能
- ✓ 一般に解くことはおろか解析も困難(既存アルゴリズムには理論保証なし)