FY2024/2024年度

Continuous Optimization Team
Akiko Takeda
連続最適化チーム　武田 朗子

RIKEN
Center for Advanced Intelligence Project

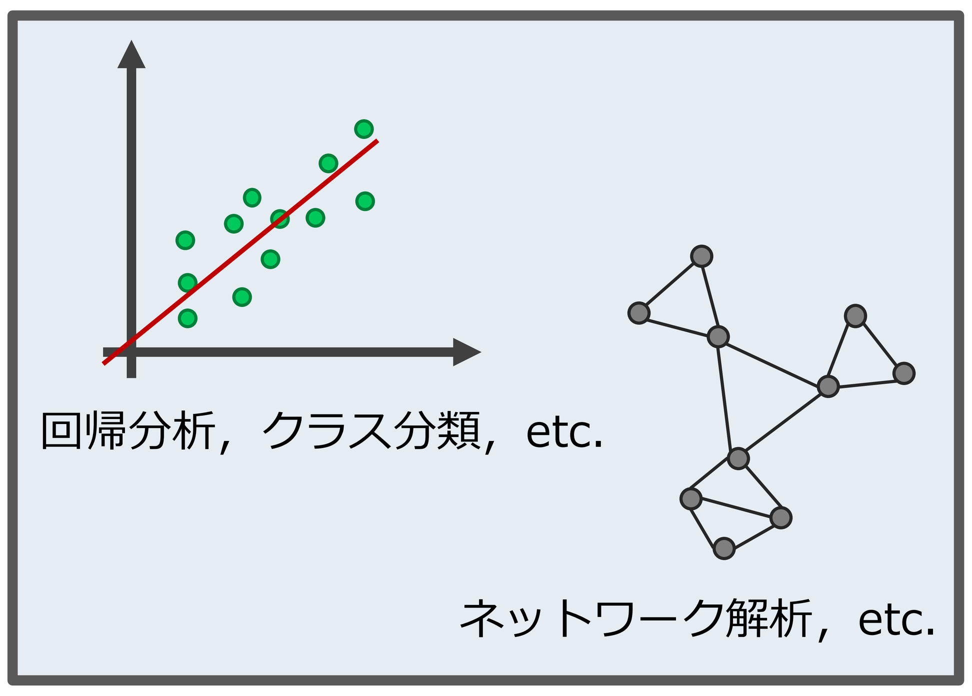
数理最適化（主に連続最適化）に関する研究

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：

- 大規模問題に対するランダム射影を用いた効率的解法
- 非凸最適化問題に対する理論保証つき解法
- 制約付き最適化問題に対する効率的解法

機械学習やデータマイニング



モデル化 →

数理最適化問題

$$\begin{aligned} \min.\ & f(x) \\ \text{s.t.}\ & g_1(x) \le 0, \\ & g_2(x) \le 0, \\ & \dots \end{aligned}$$

解法研究 (例):

- ランダム射影法
- 非平滑・非凸最適化法
- 制約付き最適化法

チームの得意とするところ

Riemann多様体上の2段階最適化
[Han, Mishra, Jawanpuria, Takeda, NeurIPS, 2024]

考える問題　多様体上の2段階最適化問題

$$\min_{x \in \mathcal{M}_x} F(x) := f(x, y^*(x)) \qquad \text{s.t. } y^*(x) = \arg\min_{y \in \mathcal{M}_y} g(x, y)$$

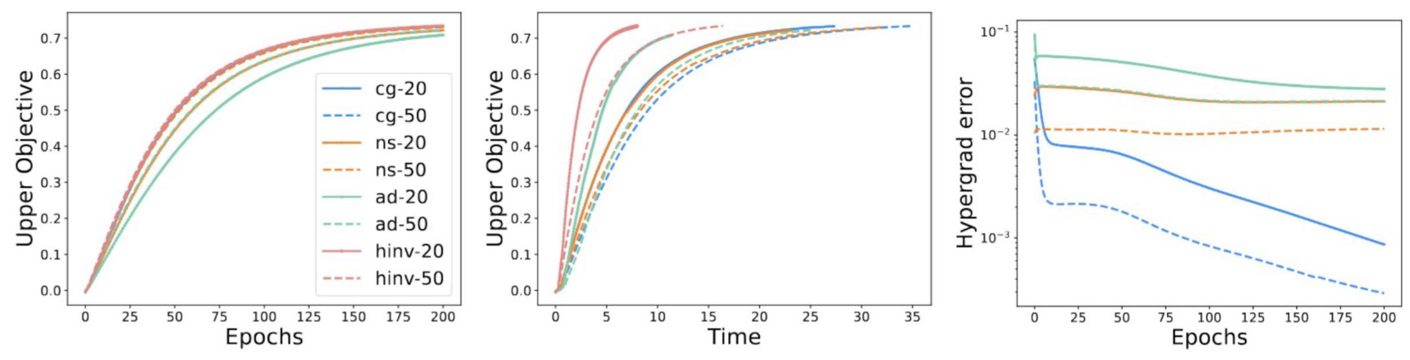
- $\mathcal{M}_x, \mathcal{M}_y$: Riemann多様体
- $f, g: \mathcal{M}_x \times \mathcal{M}_y \to \mathbb{R}$: 平滑
- $g(x, y)$: yに関して測地的強凸

本研究の貢献

✓ Hesse作用素の逆作用素・共軛勾配法・Neumann級数・自動微分を利用した**Riemann超勾配の推定**
✓ **Riemann超勾配の推定誤差**の分析
✓ Riemann超勾配を用いて2段階最適化問題を解く**アルゴリズムの収束性の保証**の分析

数値実験

（1）Riemann超勾配を推定する4手法の比較



（2）教師なしドメイン適用問題

1段階最適輸送問題としての定式化に
データの白色化のアイデアを取り入れて
2段階最適化問題として解く → **分類正解率の向上**

Methods	A→C	A→D	A→W	C→A	C→D	C→W	D→A	D→C	D→W	W→A	W→C	W→D
OT-EMD	66.67	47.77	45.76	67.52	36.31	42.71	62.17	59.71	85.08	55.41	51.16	96.82
OT-SKH	76.83	75.80	**69.83**	84.35	78.34	68.14	80.92	71.57	93.90	74.17	67.02	87.26
Proposed	**78.70**	**80.25**	**69.83**	**88.21**	**80.25**	**68.47**	**82.74**	**75.69**	**97.97**	**83.49**	**73.62**	**98.73**

合成関数最小化のための Levenberg–Marquardt 法
[Marumo, Okuno, Takeda, Math. Program., 2024]

考える問題　凸関数と合成関数の和の最小化

$$\min_{x \in \mathbb{R}^d} \left\{ F(x) := g(x) + h(c(x)) \right\}$$

- $g: \mathbb{R}^d \to \mathbb{R} \cup \{\infty\}$: 凸・シンプル
- $h: \mathbb{R}^n \to \mathbb{R}$: 凸・平滑
- $c: \mathbb{R}^d \to \mathbb{R}^n$: 平滑

応用例：経験損失最小化

$$\min_{x \in \mathbb{R}^d} \frac{1}{N} \sum_{i=1}^{N} \ell(\phi(a_i; x), b_i)$$

- モデル $\phi(\cdot; x)$ が $c(x)$ に
- 損失関数 ℓ が h に

対応

(一般化) Levenberg–Marquardt (LM) 法
[Levenberg, 1944], [Marquardt, 1963]

$$x_{k+1} = \operatorname*{argmin}_{x \in \mathbb{R}^d} \left\{ g(x) + h\Big(\underbrace{c(x_k) + \nabla c(x_k)^\top (x - x_k)}_{c(x)\ \text{の一次近似}} \Big) + \frac{\mu_k}{2} \|x - x_k\|^2 \right\}$$

本研究の貢献

LM法で μ_k を適切に定め 子問題をある近似条件を満たすように解くことで 3つの理論保証「**反復計算量**・**オラクル計算量**・**局所二次収束**」を同時に達成

既存手法との比較

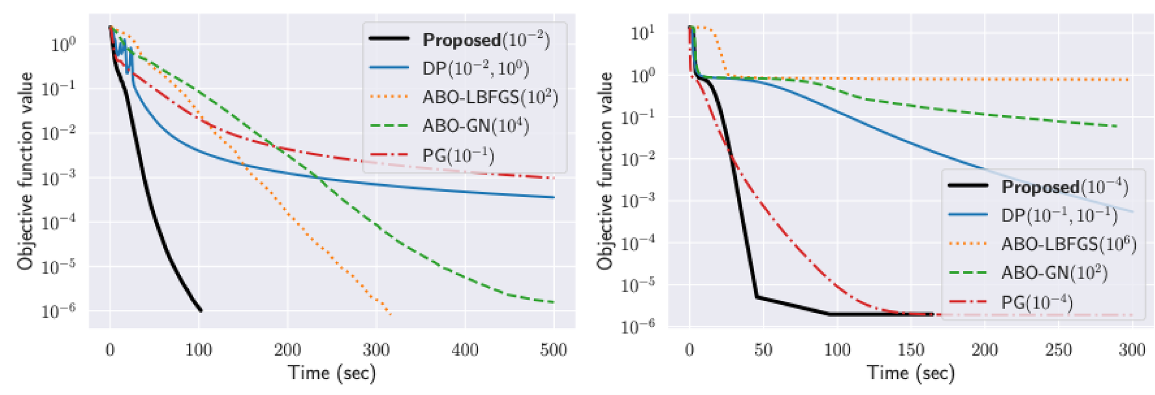
h	文献	反復	オラクル	局所
凸 + 尖	[Burke and Ferris, 1995]			✓
凸 + Lip + 尖	[Drusvyatskiy and Lewis, 2018]	✓		✓
凸 + Lip	[Drusvyatskiy and Lewis, 2018]	✓		
凸 + Lip + 平滑	[Drusvyatskiy and Paquette, 2019]	✓	✓	
凸 + 平滑	**本研究**	✓	✓	✓

数値実験

機械学習やデータ同化の問題例で提案手法（黒線）が既存手法に比べ安定して高速に動く

3層NNによる分類 (MNIST)

NMFによる行列補完 (MovieLens)



2段階最適化を用いたハイパーパラメータ選択に対する統一的な平滑化手法
[Alcantara, Nguyen, Okuno, Takeda, Chen, Math. Program., 2025]

ハイパーパラメータ学習を非平滑非凸2段階最適化問題として定式化

$$\begin{aligned} \min_{\omega_\lambda^*, \lambda} \quad & f(\omega_\lambda^*) \\ \text{s.t} \quad & \omega_\lambda^* \in \operatorname*{argmin}_{\omega \in \Re^n} G(\omega, \bar\lambda) + \lambda_1 R_1(\omega) \\ & (\lambda_1, \bar\lambda) \in \Omega_\epsilon \subset \Re^r, \end{aligned}$$

ただし

$$\begin{aligned} G(\omega, \bar\lambda) &:= g(\omega) + \bar\lambda^T \bar R(w) \\ \bar\lambda &:= (\lambda_2, \dots, \lambda_r)^T \\ \bar R(w) &:= (R_2(w), \dots, R_r(w))^T \\ R_1(\omega) &:= \sum_{i=1}^{n} \psi(|\omega_i|^p) \quad (0 < p \le 1) \\ \Omega_\epsilon &:= \{(\lambda_1, \bar\lambda) \in \Re \times \Re^{r-1} : \lambda_1 \ge \epsilon, \bar\lambda \ge 0\} \end{aligned}$$

非凸
非平滑

提案手法

✓ $R_1(\omega)$ の平滑化近似を利用

$$\varphi_\mu(\omega) := \sum_{j=1}^{n} \psi\left([\phi(\mu, \omega_j)]^p\right)$$

✓ 問題の平滑化緩和

$$\begin{aligned} \min_{\omega, \lambda} \quad & f(\omega) \\ \text{s.t} \quad & \nabla_\omega G(\omega, \bar\lambda) + \lambda_1 \nabla \varphi_\mu(\omega) = 0 \\ & \lambda \in \Omega_\epsilon. \end{aligned}$$

✓ $\mu \downarrow 0$ として半平滑Newton法を適用

Averaged results for $(n, m_2, m_1, p) = (500, 300, 300, 0.5)$using $\psi(t) = \log(1 + at)$

i	val	bkkt	sparsity(%)	time(s)	success(%)
1	**2.06e-03**	5.78e-03	**80.7**	200.4 (4.9)	**95**
2	**2.06e-03**	6.00e-03	73.0	211.8 (4.7)	85
3	2.11e-03	5.90e-03	74.9	**196.3** (5.2)	85
4	**2.06e-03**	**5.68e-03**	79.2	237.3 (6.4)	**95**

本研究の貢献

✓ **2段階KKT条件**の導入
- 最適性の新たな必要条件
- 既存の条件よりも精緻

✓ 収束性に関する統一的な理論分析
- $\phi(\mu, x)$ を生成する密度関数に基づく分析
- **既存研究より弱い制約想定でより強い収束保証**を証明

Averaged results for $(n, m_2, m_1) = (250, 500, 500)$ using $\psi(t) = t$

p	method	val	test	sparsity(%)	time(s)
1	Algorithm 1 w/ ϕ_1	**4.15e-03**	**4.12e-03**	41.7	**39.4**
	Bayesian Optimization	7.10e+00	6.79e+00	**50.4**	314.8
0.5	Algorithm 1 w/ ϕ_1	**6.60e-04**	**6.67e-04**	**72.7**	**39.9**
	Bayesian Optimization	4.47e+01	4.47e+01	19.1	160.8