

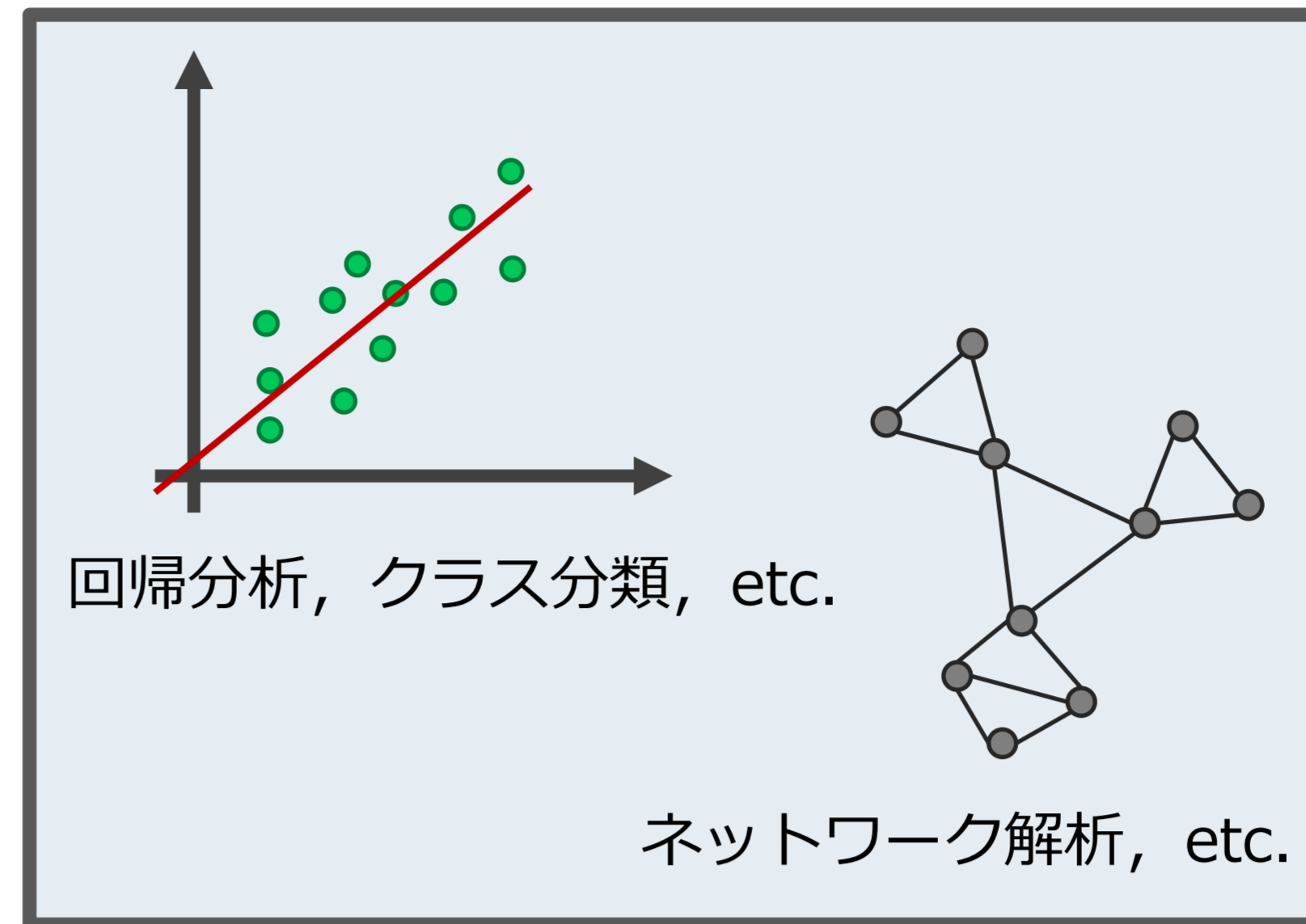
数理最適化（主に連続最適化）に関する研究

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



数理最適化問題

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

- 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

ランダム部分空間信頼領域法

[Higuchi, Poirion, Takeda, ICLR, 2025]

考える問題 大規模変数をもつ非凸最適化問題

研究成果

- ✓ 毎反復ごとに次元縮小した子問題を固有値計算により解く
- ✓ 局所2次収束性と2次停留点保証の両立は初

提案手法のアイデア

横長($s \ll n$)ランダム行列 P_k を各反復で取り直して作った次元縮小最適化問題を2次近似して探索方向を計算

$$\min_{x \in \mathbb{R}^n} f(x) \rightarrow \min_{u \in \mathbb{R}^s} f(x_k + P_k^T u)$$

$$u_{k+1} = \underset{u \in \mathbb{R}^s, \|u\| \leq \Delta_k}{\operatorname{argmin}} f(x_k) + \nabla f(x_k)^T (P_k^T u) + \frac{1}{2} (P_k^T u)^T \nabla^2 f(x_k) (P_k^T u)$$

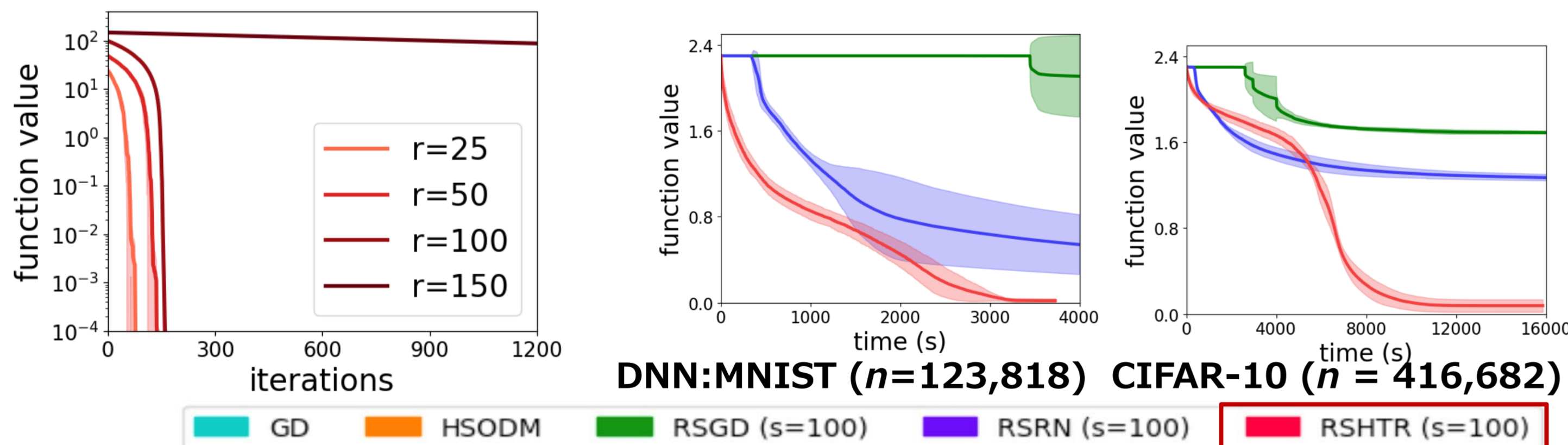
ランダム部分空間信頼領域法の更新式: $x_{k+1} = x_k + P_k^T u_{k+1}$

次元縮小関数を2次近似して得られた解 u_{k+1} を元の空間に戻す

非凸最適化問題に対するランダム部分空間法との比較

	Underlying algo.	Subprob. cost/iter	Global	Local	SOSP
Roberts & Royer (2023)	Direct search	Multi. line-search	$O(\epsilon^{-2})$		
Dzahini & Wild (2024)	Zerotoh order	Finite diff. grad.	$O(\epsilon^{-2})$		
Kozak et al. (2023)	Zerotoh order	Finite diff. grad.	$O(\epsilon^{-2})$	1	
Kozak et al. (2021)	Grad. descent	Gradient	$O(\epsilon^{-2})$	1	
Cartis et al. (2020)	Gauss-Newton	Cond. quad. prog. (QP)	$O(\epsilon^{-2})$		
Shao (2022)	Cubic Newton	Cond. cubic. reg. QP	$O(\epsilon^{-3/2})$		✓
Zhao et al. (2024)	Cubic Newton	Cubic. reg. QP	$O(\epsilon^{-3/2})$		✓
Fuji et al. (2022)	Reg. Newton	Solve eq.	$O(\epsilon^{-2})$	1+†	
Ours	Trust Region	Min eigenvalue	$O(\epsilon^{-3/2})$	2*	✓

- ✓ 次元縮小前と同じオーダーの反復計算量を達成
- ✓ 低ランク関数の局所的収束性解析: 2次収束



非凸・非平滑最適化のための4作用素分割法

[Alcantara, Lee, Takeda, SIAM J. Optim., 2025]

考える問題

$$\min_{x \in \mathbb{R}^n} \Psi(x) := f(x) + g(x) + h(x) + p(x)$$

- f, g : 近接写像が簡単に計算可能
- f, h : 平滑
- p : 非平滑かつ弱凸

提案アルゴリズム

$$x^k \in \operatorname{prox}_{\alpha f}(z^k),$$

$$y^{k+1} \in \operatorname{prox}_{\gamma g} \left(\frac{\gamma}{\alpha} (2x^k - z^k - \alpha \nabla h(x^k)) + \frac{\gamma}{\beta} T_{\beta p}(y^k) \right)$$

$$z^{k+1} = z^k + \tau (y^{k+1} - x^k).$$

追加の非平滑項を考慮可能

統一的枠組み

- 勾配降下法
- 近接勾配法
- 近接DC法
- Douglas-Rachford分割法/ADMM
- Davis-Yin分割法

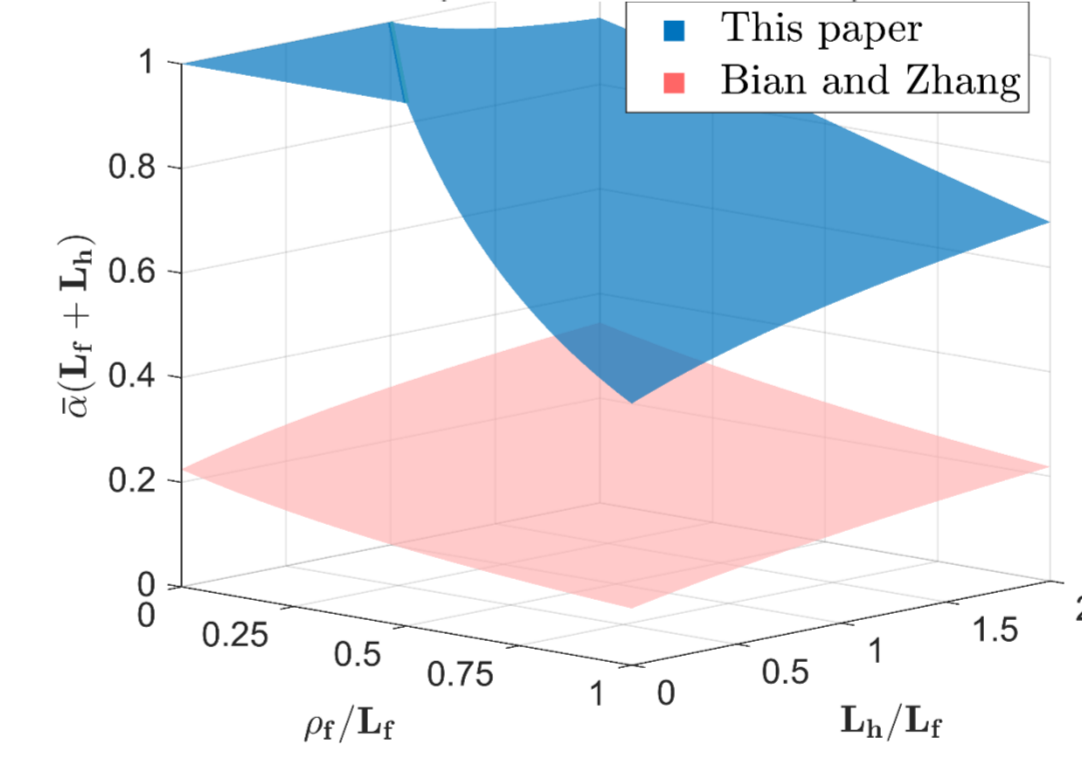
研究成果

- ✓ 非凸な問題に対するDavis-Yin分割法を4項 (うち2つが非平滑項) からなる問題に対して拡張
- ✓ 理論的収束を保証するステップサイズにより精密な上界を導出
実用上もアルゴリズムがより高速に収束

実用面での加速

(非負値低ランク行列補完問題)

Bound	$n = 100, r = 10$		$n = 100, r = 30$	
	iter	time	iter	time
Bian and Zhang (2021)	14371	17.37	27335	32.64
Ours	6516	8.71	12394	16.47



Davis-Yin分割法におけるステップサイズの大幅な改善

基数制約付き最適化問題を用いた数値実験

Method	a9a		cpusmall_scale		ijcnn1	
	m	n	m	n	m	n
PDC	32561	123	8192	12	49900	22
$\tau = 1$	91859	67.21	37132	5.64	2942	1.92
$\tau = 1.1$	91865	6.80	37133	1.34	2948	0.16
$\tau = 1.2$	83514	6.02	33757	1.23	2680	0.14
$\tau = 1.3$	76555	5.74	30944	1.18	2456	0.14
$\tau = 1.4$	70664	5.05	28564	1.03	2267	0.13
$\tau = 1.5$	65618	4.85	26524	1.10	2105	0.13
$\tau = 1.6$	61242	4.64	24756	0.97	1965	0.12
$\tau = 1.7$	51037	3.88	20630	0.74	1637	0.10
$\tau = 1.8$	48350	3.63	19545	0.68	1550	0.09

近接DC法 (PDC)より速い!

局所最適性保証付き修正 K-means 法 [Li, Metel, Takeda, ICML, 2025]

考える問題 K-means クラスタリング

$$\min_{P, C} f(P, C) = \sum_{k=1}^K \sum_{n=1}^N p_{k,n} w_n \|x_n - c_k\|^2$$

s.t. $P \in S, c_k \in \mathbb{R}^d (\forall k \in [K]),$

$$S := \left\{ P \in \mathbb{R}^{K \times N} \mid \begin{aligned} & \sum_{k=1}^K p_{k,n} = 1 (\forall n \in [N]), \\ & p_{k,n} \geq 0 (\forall k \in [K], n \in [N]) \end{aligned} \right\}$$



K-means 法 (Lloyd, 1982)

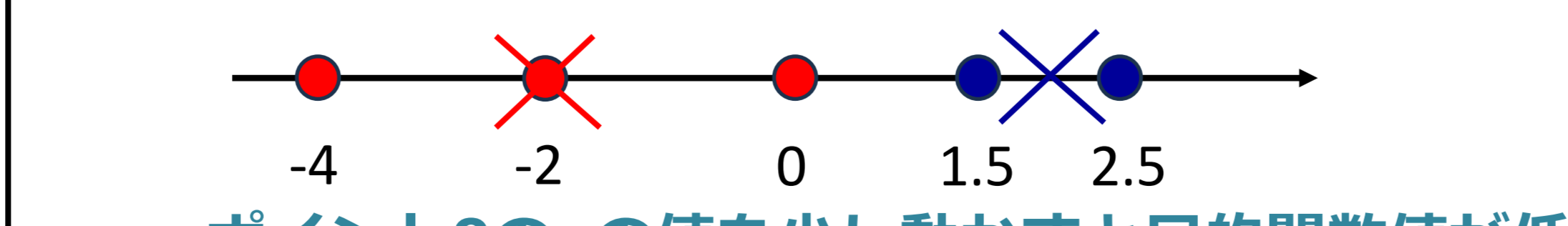
1. クラスタ中心をN点からランダムにK点選択
2. 現クラスタ中心に基づいてN点をK個に分類
3. 各分類内点の重心をクラスタ中心に設定
4. (以下更新が止まるまで2と3の繰り返し)

研究成果

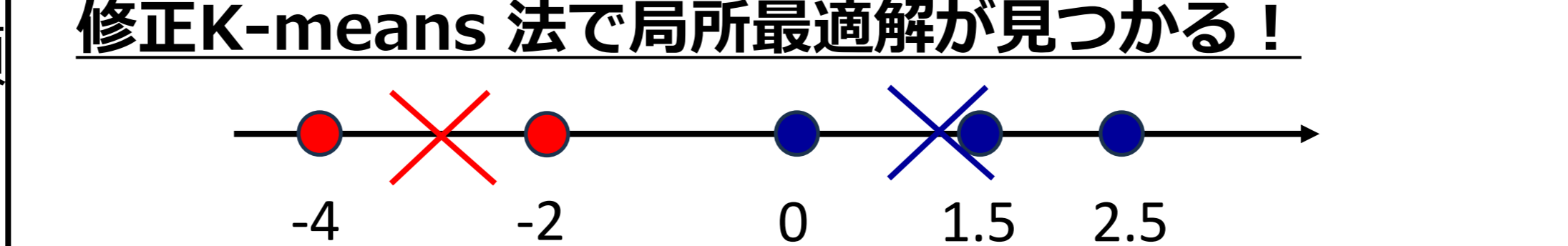
- ✓ K-means 法は局所最適解に収束するとされていた (Selim & Ismail, 1984) が実際には証明に誤りがあり 局所最適解に収束しないことを指摘
- ✓ 2タイプ (離散的・連続的) の局所最適解を定義した上で K-means 法の終了条件を変えることで 局所最適解への収束保証を持つ修正 K-means 法を提案

K-means 法で局所最適解が得られない例

$N = 5, K = 2, x = (-4, -2, 0, 1.5, 2.5)$, 初期中心(0, 2.5)



修正K-means 法で局所最適解が見つかる!



数値実験

修正 K-means 法でより良い解を見つけられることを確認

$K = 5$

Dataset	News20-1 ($N = 2,000, d = 1,089$)				
Algorithm	Mean ± Variance	Minimum	Time(s)	Num Iter	
K-means	984655 ± 55781	864073	0.31	30.8	
D-LO-K-means	808391 ± 0	808391	1.94	167.8	
Accel-D-LO	808032 ± 1566	801207	0.97	81.0	
K-means++	870899 ± 80013	808450	0.20	18.45	
D-LO++	806236 ± 3292	801207	0.82	73.55	
Accel-D-LO++	806595 ± 3111	801207	0.53	47.95	

