

決定変数に依存した分布シフト下における 適応的勾配法

東京大学

東京科学大学

NTT

東京大学 / 理化学研究所

*浜口広樹

引間友也

澤田宏

武田朗子

2026/09/10

(本日の発表資料は浜口広樹の HP でも公表しています)

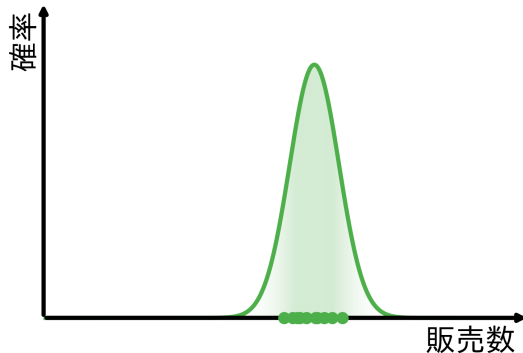
決定変数に依存した分布シフトとは？

確率的なデータ分布を対象とした最適化を扱う。

例: 商品の価格が需要分布に影響を与える価格設定問題

価格:
100 円

需要の期待値:
100 個



需要分布の例:
ガウス分布
ポアソン分布
ガンマ分布
など

$$\text{売上最大化: } \mathbb{E}_{\text{需要}(100 \text{ 円})}[\text{販売数} \cdot 100 \text{ 円}] = 10,000 \text{ 円}$$

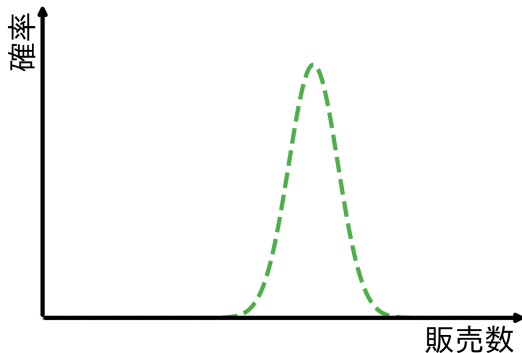
決定変数に依存した分布シフトとは？

確率的なデータ分布を対象とした最適化を扱う。

例: 商品の価格が需要分布に影響を与える価格設定問題

価格:
120 円

需要の期待値:
100 個?



需要分布の例:
ガウス分布
ポアソン分布
ガンマ分布
など

売上最大化: $\mathbb{E}_{\text{需要}(100 \text{ 円})}[\text{販売数} \cdot 120 \text{ 円}] = 12,000 \text{ 円?}$

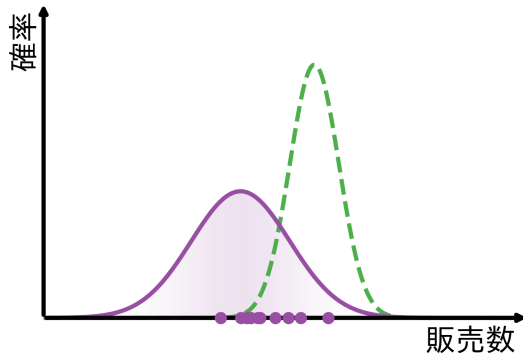
決定変数に依存した分布シフトとは？

確率的なデータ分布を対象とした最適化を扱う。

例: 商品の価格が需要分布に影響を与える価格設定問題

価格:
120 円

需要の期待値:
70 個



需要分布の例:
ガウス分布
ポアソン分布
ガンマ分布
など

$$\text{売上最大化: } \mathbb{E}_{\text{需要}(120 \text{ 円})}[\text{販売数} \cdot 120 \text{ 円}] = 8,400 \text{ 円}$$

Performative Prediction [Hardt and Mendler-Dünner2025, Perdomo et al.2020]

$$\underset{\theta \in \Theta}{\text{minimize}} \quad \mathcal{L}(\theta) := \mathbb{E}_{z \sim \mathcal{D}_{\beta(\theta)}} [\ell(z; \theta)]$$

データ z がサンプルされる分布 $\mathcal{D}$ は、**決定変数 θ の関数 $\beta(\theta)$ が分布パラメータ**。
損失関数 ℓ の期待値 $\mathbb{E}$ を最小化したい。

決定変数に依存して、分布がシフトする例 (いずれも精度最大化が目的):

- 渋滞予測の公表により、運転手が経路を変えて交通量の分布が変化する
- スпам分類器の更新に応じて、攻撃者がメールの特徴量を変化させる
- 推薦システムのパラメータが、ユーザの嗜好分布そのものを変化させる

既存手法の整理

$$\underset{\theta \in \Theta}{\text{minimize}} \quad \mathcal{L}(\theta) := \mathbb{E}_{z \sim \mathcal{D}_{\beta(\theta)}} [\ell(z; \theta)]$$

決定変数に依存してどう分布シフトするか ($\beta(\theta)$) が未知なので、分布からのデータサンプリングが必要であり、その推定が本質的。

既存手法例:

- RRM/RGD [[Perdomo et al.2020](#)]: 分布シフトを無視
- Plug-in [[Lin and Zrnic2024](#)]: β を事前推定
- DFO [[Hikima and Takeda2025](#)]: 全体をブラックボックス化

いずれも、分布パラメータ $\beta(\theta)$ を適応的に推定せず、十全でない。

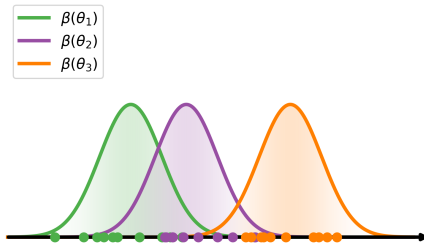
重要な既存手法: PerfGD [Izzo et al. 2021]

PerfGD は、過去のサンプルから分布変化の成分を推定する勾配型手法である。

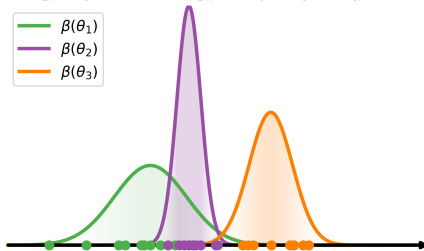
✓ 適応的分布推定・損失関数の活用をする収束保証付き手法としては初。

✗ 理論保証は限定的であり、実際に破綻する例が存在する。

既存手法が扱う分布の例



提案手法が扱う分布の例



理論保証の限定性: 分散固定ガウス分布のみが対象、分布変化 $\beta(\theta)$ が線形、勾配ノルムが小さいと破綻、生成する点列に非退化性が必要、など。

本研究の位置づけ

本研究は、PerfGD の強みを引き継ぎ、適応的な分布パラメータの推定を行う。
その上で、先の限界を全て克服する、より強い収束保証を持つ勾配法を提案する。

手法	分布推定	適応的 サンプル	損失関数 の利用	非ガウス 分布	収束保証	サンプル複雑度
RRM / RGD	無視	✓	✓	✓	–	安定点へ収束
Plug-in	事前推定	–	✓	✓	–	実装依存
DFO	–	✓	–	✓	✓✓	$\mathcal{O}(n^{2.5}\epsilon^{-5})$
PerfGD	履歴ベース	✓	✓	–	✓	$\mathcal{O}(n^{1.5}\epsilon^{-4})$ or fail
提案手法	有限差分	✓	✓	✓	✓✓	$\mathcal{O}(n^{2.5}\epsilon^{-5})$

収束保証の列について、✓ は制限付き、✓✓ はより弱い仮定下での保証。
 n は決定変数の次元、 $\epsilon > 0$ は停留性の精度を表す。

本研究の貢献

リサーチクエスション

分布族や損失関数に対する強い制約を課さずに収束保証を持つ、
performative optimization の勾配型手法を構築できるか？

本研究の位置づけ

分布パラメータ写像 β のヤコビアンを有限差分で適応的に推定する勾配法を提案。
高次元ガウス、ベルヌーイ、ポアソン、対数正規、正則な指数型分布族と、退化損失を含む広い損失関数を扱う。

弱い仮定の下で既存最良に整合するサンプル計算量を示す。

数値実験でも、複数の問題で安定した性能を確認する。

問題設定と勾配の分解 [Izzo et al.2021]

決定変数: $\theta \in \Theta \subseteq \mathbb{R}^n$, 確率的データ: $z \in \mathbb{R}^d$, 確率密度 (質量) 関数: $p(z; \beta)$.
 $\beta: \Theta \rightarrow B \subseteq \mathbb{R}^m$: 決定変数 θ に対して分布パラメータ $\beta(\theta)$ を与える未知の写像

$$\underset{\theta \in \Theta}{\text{minimize}} \quad \mathcal{L}(\theta) = \mathbb{E}_{z \sim \mathcal{D}_{\beta(\theta)}} [\ell(z; \theta)]$$

勾配は通常のリスク勾配と、分布シフトを通じた間接的な勾配に分解できる:

$$\nabla \mathcal{L}(\theta) = \nabla_1 \mathcal{L}(\theta) + \nabla_2 \mathcal{L}(\theta) \quad \text{where} \quad \begin{cases} \nabla_1 \mathcal{L}(\theta) &:= \mathbb{E}_{z \sim \mathcal{D}_{\beta(\theta)}} [\nabla_{\theta} \ell(z; \theta)], \\ \nabla_2 \mathcal{L}(\theta) &:= \frac{\partial \beta}{\partial \theta}(\theta)^{\top} G(\theta; \beta(\theta)). \end{cases}$$

$$G(\theta; \beta_0) := \mathbb{E}_{z \sim \mathcal{D}_{\beta_0}} [\ell(z; \theta) \nabla_{\beta} \ln p(z; \beta_0)], \quad \beta_0 \in B.$$

分布パラメータ $\beta(\theta)$ とその変化率 $\frac{\partial \beta}{\partial \theta}(\theta)$ が未知。これらを推定して勾配計算。

提案アルゴリズム

b : バッチサイズ, δ : 摂動半径, e_i : 第 i 標準基底, α : ステップサイズ, Proj : Θ_δ への射影.

Step 1. 現在の決定変数と摂動した決定変数を配置し、データを集める。

$$\{z_{t,j}\}_{j=1}^b \sim \mathcal{D}_{\beta(\theta_t)}, \quad \underline{\{z_{t,i,j}^\pm\}_{j=1}^b \sim \mathcal{D}_{\beta(\theta_t \pm \delta e_i)}}.$$

Step 2. 各サンプル集合から分布パラメータとヤコビアンを推定する。

$$\hat{\beta}_t \leftarrow \text{Pred}(\{z_{t,j}\}_{j=1}^b), \quad \underline{\hat{\beta}_{t,i}^\pm \leftarrow \text{Pred}(\{z_{t,i,j}^\pm\}_{j=1}^b)},$$

$$\widehat{\frac{\partial \beta}{\partial \theta}}(\theta_t) \leftarrow \left[\frac{\hat{\beta}_{t,1}^+ - \hat{\beta}_{t,1}^-}{2\delta}, \dots, \frac{\hat{\beta}_{t,n}^+ - \hat{\beta}_{t,n}^-}{2\delta} \right].$$

Step 3. 推定勾配 $\widehat{\nabla \mathcal{L}}(\theta_t) = \widehat{\nabla_1 \mathcal{L}}(\theta_t) + \widehat{\nabla_2 \mathcal{L}}(\theta_t)$ を作る。

Step 4. 射影勾配ステップで更新する。

$$\theta_{t+1} \leftarrow \text{Proj}(\theta_t - \alpha \widehat{\nabla \mathcal{L}}(\theta_t)).$$

主仮定とそれを満たす代表例

本研究の主仮定は、分布族と損失関数に対して従来より弱い条件を課す。 仮定 1 の詳細

対象	主仮定を満たし得る代表例
分布族	高次元ガウス、ベルヌーイ、ポアソン、 対数正規、正則な指数型分布族
損失関数	多項式損失、ロジスティック損失、 ソフトマックス交差エントロピー、退化損失

正確には、パラメータ領域のコンパクト性、分布・損失・分布シフトの正則性、一様モーメント条件、分布パラメータの推定誤差条件という、比較的弱い諸条件を追加で満たす場合が対象となる。

既存の PerfGD の理論保証よりも広い分布族と損失関数を扱える。

理論保証の要点

命題 1: 主仮定から導かれる条件

条件の定義

performative loss $\mathcal{L}$ は下に有界で、 $L_{\mathcal{L}}^{\text{sm}}$ -平滑である。

ℓ 、 β 、 G は必要なリプシッツ条件を満たす。

損失の分散とフィッシャー情報量は一様に有界である。

重要な帰結は、ヤコビアン推定の誤差を十分小さく抑えられることである。

中心差分によるヤコビアン推定誤差

$$\mathbb{E} \left[\left\| \frac{\partial \beta}{\partial \theta} - \frac{\widehat{\partial \beta}}{\partial \theta}(\theta_t) \right\|^2 \right] \leq n \left(\frac{(L_{\beta, \theta}^{\text{H}})^2 \delta^4}{12} + \frac{3\sigma_{\max}^2}{2\delta^2 b} \right)$$

第1項は中心差分のバイアス、第2項は分布パラメータの推定における分散に対応。

勾配写像と射影勾配法の一般評価

制約付き最適化の停留性を測る勾配写像 (gradient mapping) を定義する:

$$\mathcal{G}_\alpha(\theta) := \frac{1}{\alpha}(\theta - \text{Proj}(\theta - \alpha \nabla \mathcal{L}(\theta))).$$

$\Theta = \mathbb{R}^n$ なら $\mathcal{G}_\alpha(\theta) = \nabla \mathcal{L}(\theta)$ であり、 $\mathcal{G}_\alpha(\theta) = 0$ は制約付き停留条件を表す。

射影勾配法の一般評価

θ_0 を初期点、 $\mathcal{L}_* := \inf_{\theta \in \Theta} \mathcal{L}(\theta)$ を下限値、 T を反復回数とする。 $0 < \alpha \leq 1/L_{\mathcal{L}}^{\text{sm}}$ とし、 $\theta_{t+1} = \text{Proj}(\theta_t - \alpha \widehat{\nabla \mathcal{L}}(\theta_t))$ とする。このとき、

$$\frac{1}{T} \sum_{t=0}^{T-1} \mathbb{E}[\|\mathcal{G}_\alpha(\theta_t)\|^2] \leq \frac{2(\mathcal{L}(\theta_0) - \mathcal{L}_*)}{\alpha T} + \frac{1}{T} \sum_{t=0}^{T-1} \mathbb{E}\left[\left\|\nabla \mathcal{L}(\theta_t) - \widehat{\nabla \mathcal{L}}(\theta_t)\right\|^2\right].$$

主定理の証明概略

- $\nabla_1 \mathcal{L}$ と $\nabla_2 \mathcal{L}$ を別々に評価し、 $\mathcal{L}$ の $L_{\mathcal{L}}^{\text{sm}}$ -平滑性を示す。 命題 2
- 前頁の一般評価により、停留性を勾配推定誤差へ帰着する。 命題 3
- 分布パラメータと有限差分ヤコビアンの推定誤差を組み合わせる。 命題 4

$$\frac{1}{T} \sum_{t=0}^{T-1} \mathbb{E}[\|\mathcal{G}_{\alpha}(\theta_t)\|^2] \leq \frac{2(\mathcal{L}(\theta_0) - \mathcal{L}_*)}{\alpha T} + \frac{C_1}{b} + \frac{nC_2}{b^{2/3}}.$$

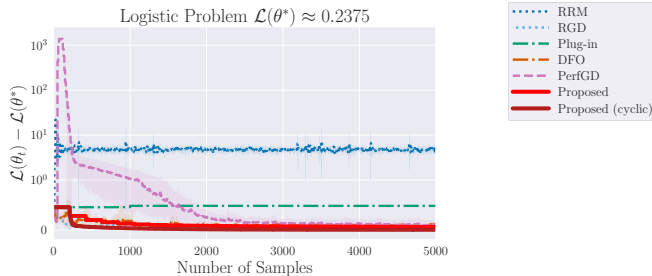
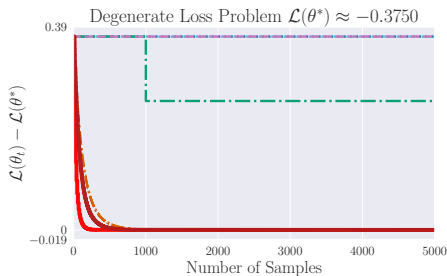
$C_1, C_2 \geq 0$ は問題の正則性から定まる定数。適切に T, b, δ を選べば左辺は ϵ^2 以下。

総サンプル計算量

各反復で $(2n+1)b$ サンプル。総サンプル計算量は $T(2n+1)b = \mathcal{O}(n^{2.5}\epsilon^{-5})$ である。

数値実験

- 指標: 最良点 θ^* に対する目的関数値の差 $\mathcal{L}(\theta) - \mathcal{L}(\theta^*)$ とサンプル数
- 比較手法: RRM / RGD / Plug-in / DFO / PerfGD
- 提案手法: Proposed と Proposed (cyclic) 循環更新



1回あたりのサンプル数を増やしたので、実用上は遅くなるのが自然。しかし、**安定性が増す恩恵がそれを上回る** or **cyclic 更新でサンプル数の増加を抑えられる**。提案手法は他の問題でも同様に安定して高速な性能を示した。

まとめ

本発表のまとめ

- 決定変数に応じてデータ分布が変化する最適化問題を扱った
 - 価格設定: 価格が需要量の分布を変化させる
 - 渋滞予測: 公表する交通情報が運転手の経路選択を変化させる
 - スパム分類: 分類器に応じて攻撃者が正例の特徴量を変化させる
 - 推薦システム: 推薦結果がユーザの嗜好分布を変化させる
- 未知の分布パラメータ写像を適応的に推定する勾配法を提案
- 既存手法 PerfGD よりも、緩い設定で収束保証とサンプル計算量を導出
- 実験では難しい分布シフトの問題設定でも数値的に安定した性能を確認

[Appendix] 実行可能領域: Θ_δ と射影

中心差分では $\theta \pm \delta e_i$ に対応する分布からサンプルを得るため、摂動後も Θ に属する点だけを反復点として用いる。

$$\Theta_\delta := \{\theta \in \Theta \mid \theta \pm \delta e_i \in \Theta, i = 1, \dots, n\}.$$

Θ_δ へのユークリッド射影を次で定義する:

$$\text{Proj}(v) := \underset{x \in \Theta_\delta}{\operatorname{argmin}} \|x - v\|.$$

定義可能性

Θ_δ が空でない閉凸集合なら、任意の $v \in \mathbb{R}^n$ に対して $\text{Proj}(v)$ は一意に定まる。

[Appendix] Proposed (cyclic): ヤコビアン の 循環更新

1. 初回は中心差分によりヤコビアン推定量 $\widehat{\frac{\partial \beta}{\partial \theta}}$ の全 n 列を推定し、メモリに保存。
2. 更新間隔 $K \in \mathbb{N}$ を定め、 K 反復ごとに 1 列だけを循環順に再推定。
3. 再推定しない列は保存した値を再利用。得られた推定ヤコビアンから $\widehat{\nabla_2 \mathcal{L}}(\theta_t)$ を計算。
4. $\widehat{\nabla_1 \mathcal{L}}(\theta_t) + \widehat{\nabla_2 \mathcal{L}}(\theta_t)$ を用いて、Proposed と同じ射影勾配更新を行う。

追加配置数と位置づけ

列を更新しない反復では b サンプル、1 列を更新する反復では $3b$ サンプルを用い、Proposed の $(2n + 1)b$ サンプルから追加配置数を削減する。
なお、本研究ではこれをヒューリスティックな変種として用いている。

[Appendix] 仮定 1: 主仮定の詳細

正則性・モーメント・推定可能性

Θ と Θ_δ は空でない閉凸コンパクト集合であり、 B は凸コンパクト集合で $\beta(\Theta) \subseteq B$ を満たす。

分布パラメータ写像 β は C^3 級である。

$\{p(\cdot; \beta_0)\}_{\beta_0 \in B}$ は共通の台をもち、 $p(z; \cdot)$ は B 上で C^2 級であり、 β_0 に関する積分と微分を交換できる。

スコア関数を $s(z; \beta_0) := \nabla_{\beta} \ln p(z; \beta_0)$ と定義し、 $\|s(z; \beta_0)\|$ と $\left\| \frac{\partial}{\partial \beta} s(z; \beta_0) \right\|$ は $\beta_0 \in B$ について一様に $\|z\|$ の高々多項式増大であると仮定する。

$\ell(z; \theta)$ は C^2 級であり、 $|\ell(z; \theta)|$ 、 $\|\nabla_{\theta} \ell(z; \theta)\|$ 、 $\|\nabla_{\theta}^2 \ell(z; \theta)\|$ は $\theta \in \Theta$ について一様に $\|z\|$ の高々多項式増大である。

分布族は、解析に必要な次数まで一様に有界な有限モーメントをもつ。

ある $\sigma_{\max} \geq 0$ と $b_{\min} \in \mathbb{N}$ が存在し、任意の $b \geq b_{\min}$ について推定器 Pred は $\mathbb{E} \left[\left\| \hat{\beta} - \beta(\theta) \right\|^2 \right] \leq \sigma_{\max}^2 / b$ を満たす。

[Appendix] 命題 1 が導く条件 (1/2): 目的関数と損失

目的関数の下界

performative loss は下に有界である:

$$\mathcal{L}_* := \inf_{\theta \in \Theta} \mathcal{L}(\theta) > -\infty.$$

損失の正則性

損失は期待値の意味で $L_{\ell, \theta}^{\text{Lip}}$ -Lipschitz かつ $L_{\ell, \theta}^{\text{sm}}$ -smooth である:

$$\sup_{\beta_0 \in B} \mathbb{E}_{z \sim \mathcal{D}_{\beta_0}} [\|\nabla_{\theta} \ell(z; \theta)\|^2]^{1/2} \leq L_{\ell, \theta}^{\text{Lip}},$$

$$\mathbb{E}_{z \sim \mathcal{D}_{\beta(\theta)}} [\|\nabla_{\theta} \ell(z; \theta) - \nabla_{\theta} \ell(z; \theta')\|] \leq L_{\ell, \theta}^{\text{sm}} \|\theta - \theta'\|.$$

[Appendix] 命題 1 が導く条件 (2/2): 分布シフト

分布パラメータ写像 β

β は Lipschitz、smooth、Hessian Lipschitz である:

$$\begin{aligned}\|\beta(\theta) - \beta(\theta')\| &\leq L_{\beta,\theta}^{\text{Lip}} \|\theta - \theta'\|, \\ \left\| \frac{\partial \beta}{\partial \theta}(\theta) - \frac{\partial \beta}{\partial \theta}(\theta') \right\| &\leq L_{\beta,\theta}^{\text{sm}} \|\theta - \theta'\|, \quad \left\| \frac{\partial^2 \beta}{\partial \theta^2} - \frac{\partial^2 \beta}{\partial \theta'^2} \right\| \leq L_{\beta,\theta}^{\text{H}} \|\theta - \theta'\|.\end{aligned}$$

関数 G 、損失分散、Fisher 情報量

$I(\beta_0) := \mathbb{E}_{z \sim \mathcal{D}_{\beta_0}} [s(z; \beta_0) s(z; \beta_0)^\top]$ と定義する。

$$\begin{aligned}\|G(\theta; \beta_0) - G(\theta'; \beta'_0)\| &\leq L_{G,\beta}^{\text{Lip}} \|\beta_0 - \beta'_0\| + L_{G,\theta}^{\text{Lip}} \|\theta - \theta'\|, \\ \sup_{\theta \in \Theta, \beta_0 \in B} \text{Var}_{z \sim \mathcal{D}_{\beta_0}} [\ell(z; \theta)] &\leq M_\ell, \quad \sup_{\beta_0 \in B} \text{Tr}(I(\beta_0)) \leq I_{\max}.\end{aligned}$$

[Appendix] 命題 2: 平滑性

命題 2: $\mathcal{L}$ の平滑性

主仮定の下、任意の $\theta, \theta' \in \Theta$ について

$$\|\nabla \mathcal{L}(\theta) - \nabla \mathcal{L}(\theta')\| \leq L_{\mathcal{L}}^{\text{sm}} \|\theta - \theta'\|.$$

ここで、平滑性定数は

$$L_{\mathcal{L}}^{\text{sm}} := L_{\ell, \theta}^{\text{sm}} + L_{\ell, \theta}^{\text{Lip}} L_{\beta, \theta}^{\text{Lip}} \sqrt{I_{\max}} + L_{\beta, \theta}^{\text{sm}} \sqrt{M_{\ell} I_{\max}} + L_{\beta, \theta}^{\text{Lip}} \left(L_{G, \beta}^{\text{Lip}} L_{\beta, \theta}^{\text{Lip}} + L_{G, \theta}^{\text{Lip}} \right).$$

三角不等式により、2つの勾配成分を別々に評価すればよい:

$$\|\nabla \mathcal{L}(\theta) - \nabla \mathcal{L}(\theta')\| \leq \|\nabla_1 \mathcal{L}(\theta) - \nabla_1 \mathcal{L}(\theta')\| + \|\nabla_2 \mathcal{L}(\theta) - \nabla_2 \mathcal{L}(\theta')\|.$$

[Appendix] 命題 2: 証明概略

通常のリスク勾配は、 θ の変化と分布の変化をそれぞれ評価して導かれる:

$$\|\nabla_1 \mathcal{L}(\theta) - \nabla_1 \mathcal{L}(\theta')\| \leq \left(L_{\ell, \theta}^{\text{sm}} + L_{\ell, \theta}^{\text{Lip}} L_{\beta, \theta}^{\text{Lip}} \sqrt{I_{\max}} \right) \|\theta - \theta'\|.$$

間接的な勾配は、ヤコビアンと G をそれぞれ評価して導かれる:

$$\|\nabla_2 \mathcal{L}(\theta) - \nabla_2 \mathcal{L}(\theta')\| \leq \left(L_{\beta, \theta}^{\text{sm}} \sqrt{M_{\ell} I_{\max}} + L_{\beta, \theta}^{\text{Lip}} \left(L_{G, \beta}^{\text{Lip}} L_{\beta, \theta}^{\text{Lip}} + L_{G, \theta}^{\text{Lip}} \right) \right) \|\theta - \theta'\|.$$

2つの係数の和を $L_{\mathcal{L}}^{\text{sm}}$ とおけば命題 2 を得る。

[Appendix] 射影と勾配写像

射影の最適性条件

任意の $v \in \mathbb{R}^n$ と $x \in \Theta_\delta$ に対し、

$$(v - \text{Proj}(v))^\top (x - \text{Proj}(v)) \leq 0.$$

真の勾配と推定勾配に対応する写像をそれぞれ

$$\begin{aligned}\mathcal{G}_\alpha(\theta_t) &:= \frac{1}{\alpha}(\theta_t - \text{Proj}(\theta_t - \alpha \nabla \mathcal{L}(\theta_t))), \\ \widehat{\mathcal{G}}_\alpha(\theta_t) &:= \frac{1}{\alpha} \left(\theta_t - \text{Proj} \left(\theta_t - \alpha \widehat{\nabla \mathcal{L}}(\theta_t) \right) \right) = \frac{1}{\alpha}(\theta_t - \theta_{t+1})\end{aligned}$$

と定義する。

[Appendix] 命題 3: 停留性の評価

命題 3: 停留性の評価

主仮定の下、 $\theta_0 \in \Theta_\delta$ 、 $0 < \alpha \leq 1/L_{\mathcal{L}}^{\text{sm}}$ とする。点列 $\{\theta_t\}_{t=0}^T$ を $\theta_{t+1} = \text{Proj} \left(\theta_t - \alpha \widehat{\nabla \mathcal{L}}(\theta_t) \right)$ により生成し、 $\mathcal{L}_* := \inf_{\theta \in \Theta} \mathcal{L}(\theta)$ と定義する。このとき、

$$\frac{1}{T} \sum_{t=0}^{T-1} \mathbb{E} [\|\mathcal{G}_\alpha(\theta_t)\|^2] \leq \frac{2(\mathcal{L}(\theta_0) - \mathcal{L}_*)}{\alpha T} + \frac{1}{T} \sum_{t=0}^{T-1} \mathbb{E} \left[\left\| \nabla \mathcal{L}(\theta_t) - \widehat{\nabla \mathcal{L}}(\theta_t) \right\|^2 \right].$$

第1項は、反復回数 T とともに減少する。

第2項は、分布パラメータとそのヤコビアン の推定誤差を制御すればよい。

[Appendix] 命題 3: 証明概略

$\mathcal{L}$ の平滑性と $\alpha \leq 1/L_{\mathcal{L}}^{\text{sm}}$ から、

$$\begin{aligned}\mathcal{L}(\theta_{t+1}) &\leq \mathcal{L}(\theta_t) - \alpha \nabla \mathcal{L}(\theta_t)^\top \widehat{\mathcal{G}}_\alpha(\theta_t) + \frac{\alpha}{2} \left\| \widehat{\mathcal{G}}_\alpha(\theta_t) \right\|^2 \\ &\leq \mathcal{L}(\theta_t) - \frac{\alpha}{2} \left\| \mathcal{G}_\alpha(\theta_t) \right\|^2 + \frac{\alpha}{2} \left\| \nabla \mathcal{L}(\theta_t) - \widehat{\nabla \mathcal{L}}(\theta_t) \right\|^2.\end{aligned}$$

最後の不等式は、射影の最適性条件を真の勾配と推定勾配の各射影点に適用した。
一反復の降下評価を $t = 0, \dots, T - 1$ で足し上げると、主張が得られる。

[Appendix] 命題 4: 勾配推定誤差

命題 4: 勾配推定誤差

主仮定の下、 $b \geq b_{\min}$ とし、摂動幅を $\delta = b^{-1/6}$ と定める。このとき、提案手法の任意の反復 t について、

$$\mathbb{E} \left[\left\| \nabla \mathcal{L}(\theta_t) - \widehat{\nabla \mathcal{L}}(\theta_t) \right\|^2 \right] \leq \frac{C_1}{b} + \frac{nC_2}{b^{2/3}},$$

ここで、

$$C_1 := 3 \left(\left(L_{\ell, \theta}^{\text{Lip}} \right)^2 I_{\max} + \left(L_{\beta, \theta}^{\text{Lip}} L_{G, \beta}^{\text{Lip}} \right)^2 \right) \sigma_{\max}^2, \quad C_2 := 3M_{\ell} I_{\max} \left(\frac{(L_{\beta, \theta}^{\text{H}})^2}{12} + \frac{3\sigma_{\max}^2}{2} \right).$$

第1項は現在点における分布パラメータの推定誤差、
第2項は中心差分によるヤコビアンの推定誤差に対応する。

[Appendix] 命題 4: 証明概略

勾配の 2 成分を評価し、 $\|G(\theta; \beta_0)\| \leq \sqrt{M_\ell I_{\max}}$ を用いると、

$$\left\| \nabla \mathcal{L}(\theta_t) - \widehat{\nabla \mathcal{L}}(\theta_t) \right\|^2 \leq 3 \left(\left(L_{\ell, \theta}^{\text{Lip}} \right)^2 I_{\max} + \left(L_{\beta, \theta}^{\text{Lip}} L_{G, \beta}^{\text{Lip}} \right)^2 \right) \left\| \beta(\theta_t) - \hat{\beta}_t \right\|^2 + 3 M_\ell I_{\max} \left\| \frac{\partial \beta}{\partial \theta} - \frac{\partial \hat{\beta}}{\partial \theta}(\theta_t) \right\|^2.$$

一方、テイラー展開と分布パラメータ推定器の誤差評価から、

$$\mathbb{E} \left[\left\| \frac{\partial \beta}{\partial \theta} - \frac{\partial \hat{\beta}}{\partial \theta}(\theta_t) \right\|^2 \right] \leq n \left(\frac{(L_{\beta, \theta}^{\text{H}})^2 \delta^4}{12} + \frac{3 \sigma_{\max}^2}{2 \delta^2 b} \right).$$

$\mathbb{E} \left[\left\| \beta(\theta_t) - \hat{\beta}_t \right\|^2 \right] \leq \sigma_{\max}^2 / b$ と $\delta = b^{-1/6}$ を代入すると、命題 4 の評価を得る。

[Appendix] 主定理とサンプル計算量

命題 3 と命題 4 を組み合わせ、各項を $\epsilon^2/3$ 以下にする。

$$T \geq \frac{6(\mathcal{L}(\theta_0) - \mathcal{L}_*)}{\alpha\epsilon^2}, \quad b \geq \max\left(b_{\min}, \frac{3C_1}{\epsilon^2}, \frac{(3nC_2)^{1.5}}{\epsilon^3}\right), \quad \delta = b^{-1/6}.$$

この選択により、

$$\frac{1}{T} \sum_{t=0}^{T-1} \mathbb{E}[\|\mathcal{G}_\alpha(\theta_t)\|^2] \leq \epsilon^2.$$

総サンプル計算量

1 反復で現在点と $2n$ 個の摂動点から各 b サンプルを得るため、

$$T(2n + 1)b = \mathcal{O}(n^{2.5}\epsilon^{-5}).$$

詳細な仮定・証明・実験設定は [arXiv:2607.26562](https://arxiv.org/abs/2607.26562) をご覧ください。

参考文献 I

Moritz Hardt and Celestine Mendler-Dünnér. Performative Prediction: Past and Future, 2025.

Yuya Hikima and Akiko Takeda. Zeroth-order gradient estimators for stochastic problems with decision-dependent distributions, 2025.

Zachary Izzo, Lexing Ying, and James Zou. How to Learn when Data Reacts to Your Model: Performative Gradient Descent. In *Proceedings of the 38th International Conference on Machine Learning*, pages 4641–4650. PMLR, 2021.

Licong Lin and Tijana Zrnic. Plug-in Performative Optimization. In *Proceedings of the 41st International Conference on Machine Learning*, pages 30546–30565. PMLR, 2024.

Juan Perdomo, Tijana Zrnic, Celestine Mendler-Dünnér, and Moritz Hardt. Performative Prediction. In *Proceedings of the 37th International Conference on Machine Learning*, pages 7599–7609. PMLR, 2020.