

---

# An effective framework for estimating individualized treatment rules

---

Joowon Lee<sup>1</sup>, Jared D. Huling<sup>2</sup>, Guanhua Chen<sup>1</sup>

<sup>1</sup> University of Wisconsin-Madison, <sup>2</sup> University of Minnesota  
joowon.lee@wisc.edu, huling@umn.edu, gchen25@wisc.edu

## Abstract

Estimating individualized treatment rules (ITRs) is fundamental in causal inference, particularly for precision medicine applications. Traditional ITR estimation methods rely on inverse probability weighting (IPW) to address confounding factors and  $L_1$ -penalization for simplicity and interpretability. However, IPW can introduce statistical bias without precise propensity score modeling, while  $L_1$ -penalization makes the objective non-smooth, leading to computational bias and requiring sub-gradient methods. In this paper, we propose a unified ITR estimation framework formulated as a constrained, weighted, and smooth convex optimization problem. The optimal ITR can be robustly and effectively computed by projected gradient descent. Our comprehensive theoretical analysis reveals that weights that balance the spectrum of a ‘weighted design matrix’ improve both the optimization and likelihood landscapes, yielding improved computational and statistical estimation guarantees. In particular, this is achieved by distributional covariate balancing weights, which are model-free alternatives to IPW. Extensive simulations and applications demonstrate that our framework achieves significant gains in both robustness and effectiveness for ITR learning against existing methods.

## 1 Introduction

Traditional medicine often uses a "one-size-fits-all" approach where the same treatment is applied to all patients regardless of their unique attributes, aiming to find a single optimal treatment that may be most effective for a broad population. However, since not everyone fits the mold, precision medicine has been popular in medical research, emphasizing personalized treatment based on a patient’s unique characteristics [11].

One key element of precision medicine is the estimation of *individualized treatment regime* (ITR), also known as treatment policy or policy. ITRs consider various factors like demographics and socio-psychological aspects to optimize treatment decisions, thereby maximizing individual outcomes. Many frameworks have been proposed for learning ITRs [41, 54, 57, 7, 58, 60, 56] and the related *conditional average treatment effect* (CATE) estimation problem [3, 53, 46, 13, 36]. For model interpretation and other practical concerns, researchers often focus on pre-specified rule classes. For example, the method proposed in [2] focuses on estimating ITRs using shallow-depth decision trees.

The central task in the ITR estimation is controlling confounding factors to isolate the causal effect of treatment from other factors that may influence the outcome. A standard approach is by weighting each sample using their *inverse probability weights* (IPWs), which requires specifying a propensity score model. However, propensity score methods have long been known to be highly sensitive to model misspecification, which yields biased estimates of causal effects [25, 30]. *Distributional covariate balancing weights* (DCBWs) [24, 27] are modern alternatives to IPWs, directly minimizing the distance between empirical covariate distributions induced by the weights.

ITR estimation can be framed as an optimization or *maximum likelihood estimation* (MLE) problem, where the choice of weights directly impacts both the optimization and likelihood landscapes. This, in

turn, affects the computational and statistical errors in the ITR estimation, as well as the performance of the optimization algorithm. The key contribution of this work is to bridge the gap between causal inference and optimization perspectives in ITR estimation. Our main finding is both simple and surprising: *Covariate balancing weights not only address confounding but also yield optimal optimization and likelihood landscapes for the ITR estimation.* We show this claim by formulating the ITR estimation problem as a constrained, weighted, and smooth convex optimization problem. Our formulation is grounded in *angle-based direct learning* (AD-Learning) framework [39], which is a recent framework for multi-category treatments with a linear decision function class assumption. While we focus on multi-category treatments—a relatively underexplored area—we believe that the principles are equally applicable to binary treatments with a linear decision function.

Besides advancing our theoretical understanding of the impact of covariate balancing weights, we propose a unified computational and statistical framework for the ITR estimation. Notably, we introduce a hard  $L_1$ -ball constraint to promote sparsity in regression coefficients, replacing the traditional soft  $L_1$ -penalization. This approach keeps the optimization objective smooth, enabling fast and accurate ITR estimation using *projected gradient descent* (PGD). Furthermore, we integrate techniques such as (1) variable screening, (2) outcome augmentation, and (3) inverse variance weighting to account for heteroscedastic errors, which together show a synergistic effect.

**Related Work.** Most existing ITR approaches use IPW, which requires specifying a propensity score model for confounding control. Recent developments in causal inference have also introduced robust approaches that directly estimate weights based on balance-seeking objectives rather than relying solely on propensity scores. These approaches include entropy balancing weights [20], stable balancing weights [61] as well as DCBWs [24, 27]. However, despite the weights serving a critical role in the estimation process, these advancements have mainly focused on improving *average treatment effect* (ATE) and are largely under-explored in the ITR literature. To the best of our knowledge, this is the first work to apply DCBWs for ITR-Learning under a weighted optimization framework. While recent works have taken a similar approach by utilizing weighting schemes and directly optimizing the weights rather than relying on estimated IPW [28, 26], our method differs in key aspects. Unlike [26], which involves the evaluation of policy effects, our approach directly learns the optimal decision function, bypassing the need for intermediate policy effect evaluations. Although [28] introduced the concept of retargeting the population covariate distribution, which is conceptually related to our analysis, their approach relies on a reference policy, whereas our method does not. We also explore various statistical techniques to improve ITR-Learning. First, we perform *variable screening* using the distance covariance test [48] to retain key effect modifiers and precision variables that directly impact the outcome. Second, [50, 9] introduced *outcome augmentation* to minimize estimator variance in clinical trials and observational studies by IPW involving binary treatments. Building on these foundations, our work extends these approaches to a broader range of weighting schemes, such as DCBW particularly for multi-category treatments. Lastly, with inverse variance weighting, as introduced in *stabilized angle-based direct learning* (SABD-Learning) to address heteroscedasticity [44], this combined strategy enhances estimator precision and robustness.

**Contributions.** We summarize our contributions below.

- **Unified Framework of ITR-Learning:** We introduce a novel framework that addresses the limitations of existing ITR estimation methods by formulating the problem as a constrained, weighted, and smooth convex optimization problem. Under the framework, we propose a PGD algorithm for ITR-Learning under a hard  $L_1$ -constraint.
- **Improved Computational and Statistical Guarantees:** We establish convergence guarantees (Theorem 3.3). Furthermore, under mild assumptions, we demonstrate that the parameter of ITR-Learning can be consistently estimated with high probability (Theorem 3.5) and provide computational and sample complexity. In both cases, we demonstrate that using covariate balancing weights controls confounding factors and leads to better optimization and likelihood landscapes, resulting in improved computational and statistical performance.
- **Statistical Framework of ITR-Learning:** Our main contribution lies in providing a unified framework that combines DCBWs, variable screening, outcome augmentation, and inverse variance weighting in a synergistic and effective way. We demonstrated theoretical justifications for the combined approach, which, to our knowledge, have not been previously established.

## 2 Methods for ITR Estimation

### 2.1 Model Setup

**Individualized Treatment Rule and Optimal Decision Function.** We observe a random vector  $(\mathbf{X}, A, Y)$  where  $\mathbf{X} = (X_1, X_2, \dots, X_p) \in \mathcal{X}$  denotes the  $p$ -dimensional vector of pre-treatment covariates, a received treatment  $A \in \mathcal{A} = \{1, 2, \dots, K\}$ , and the corresponding outcome  $Y$ . Without loss of generality, we assume higher  $Y$  values are more favorable. Let  $Y(a)$  denote the random variable that describes the potential outcome that would have been observed were the individual assigned to treatment  $a \in \mathcal{A}$ . We make the following standard assumptions using the potential outcome framework [43]: (i) Stable Unit Treatment Value Assumption (SUTVA):  $Y = Y(A)$ , (ii) positivity:  $0 < \pi(a, \mathbf{X}) := \mathbb{P}(A = a | \mathbf{X}) < 1$  for all  $a \in \mathcal{A}$  and all  $\mathbf{X} \in \mathcal{X}$ , (iii) no unmeasured confounding:  $Y(a) \perp A | \mathbf{X}$  for any  $a \in \mathcal{A}$ .

An ITR,  $d(\mathbf{x})$ , is a function mapping each covariate  $\mathbf{x}$  to one of the  $K$  treatments. According to [41], the ITR  $d$  can be measured by the value function  $V(d)$ :

$$V(d) := \mathbb{E}[Y(d(\mathbf{X}))] = \mathbb{E}[Y|A = d(\mathbf{X})] = \mathbb{E}\left[\frac{Y \mathbf{1}(A = d(\mathbf{X}))}{\pi(A, \mathbf{X})}\right]$$

Then the optimal ITR  $d^{\text{opt}}$  is defined as a function that maximizes the expected potential outcome  $d^{\text{opt}}(\mathbf{x}) \in \arg \max_{d \in \mathcal{D}} V(d)$  among all functions belonging to the treatment rule class  $(\mathcal{D})$ . Building on this, we adopt the following working model:

$$Y = \mu(\mathbf{X}) + \sum_{k=1}^K \delta_k(\mathbf{X}) \mathbf{I}(A = k) + \epsilon, \quad (1)$$

where  $\mu(\mathbf{X})$  is treatment-free effect,  $\delta_k(\mathbf{X})$  is interaction effect between covariates and  $k$ th treatment, and  $\epsilon$  is a random noise with mean zero and variance  $\sigma^2(A, \mathbf{X})$ . For model identifiability, we assume  $\sum_{k=1}^K \delta_k(\mathbf{x}) = 0$  for all  $\mathbf{x}$ . Under this model,  $\delta_k(\mathbf{X})$  determines the optimal ITR, while  $\mu(\mathbf{x})$  has no impact on the ITR. One approach to identifying the optimal ITR is to use  $K$  separate decision functions for each treatment with the sum-to-zero constraint, but this approach can be computationally inefficient [55].

Instead, one can opt for simplex coding as an alternative, which inherently satisfies the sum-to-zero constraint. In AD-Learning [39], each of the  $K$  treatments is represented as a vertex simplex on  $\mathbb{R}^{K-1}$ , denoted as  $\mathbf{u}_k$ ,  $k = 1, \dots, K$ . In particular,  $\mathbf{u}_k$  is defined as

$$\mathbf{u}_k = \begin{cases} \frac{1}{\sqrt{K-1}} \mathbf{1}_{K-1} & \text{if } k = 1, \\ -\frac{1+\sqrt{K}}{\sqrt{(K-1)^3}} \mathbf{1}_{K-1} + \sqrt{\frac{K}{K-1}} \mathbf{e}_{k-1} & \text{if } k \geq 2, \end{cases}$$

where  $\mathbf{1}_{K-1}$  is a  $K-1$  dimensional vector with entries 1, and  $\mathbf{e}_{k-1}$  is a  $K-1$  dimensional vector with entries 0 except its  $k$ -th entry being 1. This vertex simplex has  $K$  vertices with equal angles between them and an origin at the center of the simplex. All  $\mathbf{u}_k$  have the same Euclidean norm 1. Then the optimal ITR is reformulated as follows:

$$d^{\text{opt}}(\mathbf{x}) = \arg \max_{k \in \{1, \dots, K\}} \mathbb{E}[Y | \mathbf{x}, A = k] = \arg \max_{k \in \{1, \dots, K\}} \underbrace{\mathbf{u}_k^T \mathbb{E}\left[\frac{Y \mathbf{u}}{\pi(A, \mathbf{x})} \middle| \mathbf{x}\right]}_{=: \mathbf{f}_{\text{opt}}(\mathbf{x})}, \quad (2)$$

where  $\mathbf{u}$  is a random treatment vector in  $\mathbb{R}^K$ , corresponding to the random treatment  $A$ . Then estimating the optimal ITR can be converted into estimating the decision function  $\mathbf{f}(\mathbf{x}) = (f_1(\mathbf{x}), \dots, f_{K-1}(\mathbf{x}))^T$ , assigning a  $K-1$  dimensional vector to each covariate  $\mathbf{x}$ . Here, while the optimal decision function  $\mathbf{f}_{\text{opt}}(\cdot)$  can take a generic form, we assume the linear decision function within the linear treatment rule class  $\mathcal{D}$ , induced by the linear decision function class  $\mathcal{F} = \{\mathbf{f}(\mathbf{X}) = \mathbf{B}^T \mathbf{X} : \mathbf{B} \in \mathbb{R}^{p \times (K-1)}\}$ .

**Reduction to Weighted Convex Optimization.** Suppose we have a sample  $(\mathbf{x}_i, a_i, y_i)_{i=1}^n$  of size  $n$  from the joint population distribution for  $(\mathbf{X}, A, Y)$ . Estimating the optimal decision function  $\mathbf{f}_{\text{opt}}$  from the observed finite sample can be reduced to solving a weighted convex optimization problem:

$$\min_{\mathbf{B} \in \mathbb{R}^{p \times (K-1)}} \frac{1}{n} \sum_{i=1}^n w(a_i, \mathbf{x}_i) \ell(y_i, \mathbf{x}_i, a_i; \mathbf{B}) + R(\mathbf{B}). \quad (3)$$

There are three key components in the problem (3): the per-sample objective function  $\ell$ , the per-sample weights  $w$ , and the regularization  $R(\mathbf{B})$ . First,  $\ell$  denotes a model-dependent per-sample convex objective function given by

$$\ell(y_i, \mathbf{x}_i, a_i; \mathbf{B}) = \begin{cases} \frac{1}{2} \left( \frac{K}{K-1} y_i - \mathbf{u}_{a_i}^T \mathbf{B}^T \mathbf{x}_i \right)^2 & \text{for continuous outcome,} \\ \log(1 + \exp(\mathbf{u}_{a_i}^T \mathbf{B}^T \mathbf{x}_i)) - y_i \mathbf{u}_{a_i}^T \mathbf{B}^T \mathbf{x}_i & \text{for binary outcome.} \end{cases} \quad (4)$$

An important characteristic of causal inference problems is that only one of the possible treatments can be observed for each subject. Thus, it is critical to control confounding variables to isolate the causal effect of the treatment from other factors that might influence the outcome. This effectively is done by choosing the appropriate weight  $w_i = w(a_i, \mathbf{x}_i)$  for the  $i$ th subject. A classic approach is to use IPW:  $w(a, \mathbf{x}) = 1/\pi(a, \mathbf{x})$  [39].

## 2.2 Proposed ITR Estimation Framework

**Reducing Finite Sample Bias.** The classic IPW approach requires specifying a propensity score model for confounding control. However, it is well-known that propensity score methods are highly sensitive to model misspecification, which yields biased estimates of causal effects [25, 30]. To address this issue, we investigate incorporating alternative, model-free weighting schemes such as *energy balancing weights* (EBWs) [23, 24], which are a type of DCBWs. Other options for DCBWs include *maximum mean discrepancy* (MMD) balancing weights [19, 27, 8] and Wasserstein distance-based balancing weights [52, 1]. Specifically, under the working model assumption (1), the optimal decision function  $\mathbf{f}_{\text{opt}}$  in (2) can be decomposed as

$$\mathbf{f}_{\text{opt}}(\mathbf{x}) = \underbrace{\mu(\mathbf{x}) \mathbb{E} \left[ \frac{\mathbf{u}}{\pi(A, \mathbf{x})} \middle| \mathbf{x} \right]}_{\text{treatment-free}} + \underbrace{\sum_{k=1}^K \delta_k(\mathbf{x}) \mathbb{E} \left[ \frac{\mathbf{u} I(A=k)}{\pi(A, \mathbf{x})} \middle| \mathbf{x} \right]}_{\text{interaction}} + \mathbb{E} \left[ \frac{\mathbf{u}}{\pi(A, \mathbf{x})} \middle| \mathbf{x} \right] \mathbb{E}[\epsilon | \mathbf{x}]. \quad (5)$$

Since  $\mathbb{E} \left[ \frac{\mathbf{u}}{\pi(A, \mathbf{x})} \middle| \mathbf{x} \right] = 0$  and  $\mathbb{E}[\epsilon | \mathbf{x}] = 0$ , it follows that  $\mathbf{f}_{\text{opt}}(\mathbf{x}) = \sum_{k=1}^K \delta_k(\mathbf{x}) \mathbf{u}_k$ , depending only on the interaction term. However, suppose the true propensity score is not correctly estimated. In that case, the estimated decision function may be biased and include additional factors including the treatment-free effect, leading to sub-optimal treatment decisions. There are two possible sources for this estimation error. First, an incorrectly specified propensity score model might lead to an estimated propensity score  $\hat{\mathbb{P}}_n(A = a | \mathbf{x})$  that deviates from the true propensity score  $\pi(A = a | \mathbf{x})$ . Second, even with correct specification, insufficient sample size may lead to inaccurate finite-sample approximation  $\hat{\mathbb{P}}_n(A = a | \mathbf{x})$  of the true propensity score  $\pi(A = a | \mathbf{x})$  [31]. It results in systematic bias for the estimated coefficient of the treatment-free term:

$$\hat{\mathbb{E}}_n \left[ \frac{\mathbf{u}}{\pi(A, \mathbf{x})} \middle| \mathbf{x} \right] = \sum_{k=1}^K \frac{\mathbf{u}_k}{\pi(k, \mathbf{x})} \hat{\mathbb{P}}_n(A = k | \mathbf{x}) \neq 0.$$

Instead of using IPW, we use DCBWs such as EBWs. Such weights are data-driven and do not rely on specific model assumptions or large sample approximation. Furthermore, EBWs are known to promote independence between treatment and confounders [23], so  $\hat{\mathbb{E}}_n[\mathbf{u} w(A, \mathbf{x}) | \mathbf{x}] \approx \hat{\mathbb{E}}_n[\mathbf{u}] \hat{\mathbb{E}}_n[w(A, \mathbf{x}) | \mathbf{x}] = 0$  since  $\hat{\mathbb{E}}_n[\mathbf{u}] = 0$ . Thus, EBWs may reduce the unfavorable impacts of bias on the finite-sample decision function.

**Improving Optimization Guarantees.** To our best knowledge, the impact of the choice of the regularization  $R(\mathbf{B})$  on the model parameter  $\mathbf{B}$  on the detailed optimization landscape, especially in terms of the convergence rate and statistical estimation error bounds has been under-investigated in the literature. Contrary to the literature, we propose to use the combination of *soft  $L_2$ -regularization and a hard  $L_1$ -ball constraint*:

$$\min_{\mathbf{B} \in \mathbb{R}^{p \times (K-1)}, \|\mathbf{B}\| \leq \lambda_1} \left[ \mathcal{L}(\mathbf{B}) := \frac{1}{n} \sum_{i=1}^n w(a_i, \mathbf{x}_i) \ell(y_i, x_i, a_i; \mathbf{B}) + \frac{\lambda_2}{2} \|\mathbf{B}\|_F^2 \right]. \quad (6)$$

The above formulation has a smooth objective function that is unaffected by the hard  $L_1$ -ball constraint. Also,  $L_2$ -regularization ensures that the objective is at least  $\lambda_2$ -strongly convex. Thus, the

PGD algorithm (with suitable stepsize) enjoys a strong exponential convergence guarantee toward the global optimum. In contrast, standard approaches with  $L_1$ -penalization loss need sub-gradient methods since it is non-smooth. It is known that the convergence rate of the sub-gradient methods is much slower than gradient methods [6, 18].

We use the classical PGD to solve the convex-constrained optimization problem in (6). In each iteration, it involves conducting gradient descent with a suitable stepsize  $\alpha_t$ , followed by a projection  $\Pi_{\mathcal{B}}$  onto the  $L_1$ -ball  $\mathcal{B} := \{\mathbf{B} \in \mathbb{R}^{p \times (K-1)} : \|\mathbf{B}\| \leq \lambda_1\}$ :

$$\mathbf{B}_t \leftarrow \Pi_{\mathcal{B}}(\mathbf{B}_{t-1} - \alpha_t \nabla_{\mathbf{B}} \mathcal{L}(\mathbf{B}_{t-1})). \quad (7)$$

For the projection onto the  $L_1$ -projection, we use the algorithm in [14]. Detailed implementation of this PGD algorithm is discussed in Algorithm 1 in the appendix.

**Additional Improvements by Variance and Dimension Reduction.** To ensure the stability of estimates for both balancing weights and outcome regression models, a variable screening step is recommended. We propose adapting distance covariance test [48] for its computational efficiency and its ability to detect complex dependencies between treatment and outcome. By focusing on screened variables that significantly impact the outcome, we can improve the estimation of outcome regression models. Additionally, estimating balancing weights based on effect modifiers rather than all variables provides more reliable weighting estimates. Hence, the screening approach serves to reduce bias and enhance the effectiveness of ITR-Learning. See Algorithm 2 in the appendix.

Lastly, we propose using both augmented outcomes and inverse variance weighting for variance reduction in learning ITRs. Under the working model assumption in (1), the interaction effect determines optimal treatment assignments, while the treatment-free effect acts as additional noise. Outcome augmentation helps reduce variance by mitigating residual variability due to the treatment-free effect. Nevertheless, misspecification of either the outcome or treatment-free effect models can induce heteroscedastic errors due to residual treatment-free effects [35]. In such scenarios, using inverse variance weighting, introduced in SABD-Learning, enhances robustness against misspecification [44]. Consequently, the combination of outcome augmentation and inverse variance weighting offers a robust and efficient approach to ITR-Learning, as they are complementary methods. Details are discussed in the appendix C.3 and C.4.

### 3 Statement of results

Our theoretical analysis shows that a “weighted design matrix”, specifically a  $p(K-1) \times p(K-1)$  positive semi-definite matrix, plays a central role in the local landscape analysis of the ITR estimation problem:

$$\Psi := \frac{1}{n} \sum_{i=1}^n \omega(a_i, \mathbf{x}_i) (\mathbf{u}_{a_i} \mathbf{u}_{a_i}^T) \otimes (\mathbf{x}_i \mathbf{x}_i^T), \quad (8)$$

where  $\otimes$  denotes the Kronecker product. In order to control the strong convexity and the smoothness parameter of the ITR objective (6), we require that the eigenvalues of  $\Psi$  are uniformly bounded.

Note that the weights  $w_i = \omega(a_i, \mathbf{x}_i)$ ,  $i = 1, \dots, n$  affect the eigenvalues of the weighted design matrix  $\Psi$ . Our analysis reveals that, for the improved convergence rate of the PGD algorithm (Theorem 3.3) as well as a smaller statistical estimation error (Theorem 3.5), we need to choose the weights  $w_i$  such that the minimum eigenvalue of  $\Psi$  is as large as possible. Roughly speaking, this is achieved when  $w_1, \dots, w_n$  are ‘covariate balancing weights’ where the weighted sample covariance matrices conditional for each treatment are approximately the same. In the special case when the covariates are discrete and one-hot encoded, the optimal weights are the ones that exactly balance the covariate distributions given treatments. Thus this provides an explicit spectral characterization of the DCBWs in causal inference literature. See the appendix D for more discussions.

#### 3.1 Computational Guarantees

**Assumption 3.1** (Eigenvalue bounds on the weighted design matrix). There are constants  $0 \leq \lambda^- \leq \lambda^+$  such that the eigenvalues of  $\Psi$  in (8) is between  $\lambda^-$  and  $\lambda^+$ ,

For each subject  $i = 1, \dots, n$ , let  $z_i$  and  $p_i$  denote its ‘activation’ and ‘predictive probability’, where

$$z_i := \mathbf{u}_{\mathbf{a}_i}^T \mathbf{B}^T \mathbf{x}_i, \quad p_i := \exp(z_i)/(1 + \exp(z_i)). \quad (9)$$

We need the activation  $z_i$  to be uniformly bounded for our theoretical analysis of the binary outcome. Since the simplex vectors  $\mathbf{u}_{\mathbf{a}_i}$  have unit length, it is enough to require the following assumption.

**Assumption 3.2** (Bounded activation). There exists a constant  $M > 0$  such that for all model parameter  $\mathbf{B} \in \mathbb{R}^{p \times (K-1)}$  in the constraint set  $\mathcal{B}$  and observed sample  $i = 1, \dots, n$ ,  $\|\mathbf{B}^T \mathbf{x}_i\| \leq M$ . Consequently, there exist constants  $0 < \alpha^- \leq \alpha^+ \leq 1/4$  such that  $\alpha^- \leq p_i(1 - p_i) \leq \alpha^+$  for all  $i$ .

With our hard  $L_1$ -norm constraint on the model parameter  $\mathbf{B}$ , Assumption 3.2 can be easily verified whenever the covariate vectors  $\mathbf{x}_i$  are uniformly bounded, which is a standard in the literature.

We establish the convergence rate of Algorithm 1 for ITR-Learning with stepsizes that are fixed but sufficiently small and diminishing rate. An informal statement is given in Theorem 3.3 below, and a full statement is provided in Theorem E.1 in the appendix. Define

$$\mu := \begin{cases} \lambda^- + \lambda_2 & \text{for continuous outcome} \\ \alpha^- \lambda^- + \lambda_2 & \text{for binary outcome} \end{cases} \quad L := \begin{cases} \lambda^+ + \lambda_2 & \text{for continuous outcome} \\ \alpha^+ \lambda^+ + \lambda_2 & \text{for binary outcome} \end{cases} \quad (10)$$

We say a function  $f : \mathbb{R}^p \rightarrow \mathbb{R}$  is  $\mu$ -strongly convex if  $f(x) - \frac{\mu}{2}\|x\|^2$  is convex and  $L$ -smooth if  $f$  is differentiable and  $\nabla f$  is  $L$ -Lipschitz continuous.

**Theorem 3.3** (Convergence rate of PGD for ITR-Learning). *Let  $(\mathbf{B}_t)_{t \geq 0}$  denote the sequence of parameters obtained by the PGD algorithm (7) for ITR-Learning problem (6) with arbitrary initialization  $\mathbf{B}_0$ . Suppose Assumptions 3.1 and 3.2 hold. Let  $\mathbf{B}^*$  denote the unique global optimum of (6). Then the following hold:*

- (i) (Optimization landscape) *The ITR objective  $\mathcal{L}(\mathbf{B})$  in (6), is  $\mu$ -strongly convex and  $L$ -smooth, where  $\mu, L$  are as in (10).*
- (ii) (Linear convergence with fixed stepsize) *Assume constant stepsize  $\alpha_t \equiv \alpha < 2/L$ . Denote the contraction constant  $\rho(\alpha) := \max\{|1 - \alpha L|, |1 - \alpha \mu|\} \in (0, 1)$ . Then  $\|\mathbf{B}_t - \mathbf{B}^*\|_F^2 \leq \rho(\alpha)^t \|\mathbf{B}_0 - \mathbf{B}^*\|_F^2$  for all  $t \geq 1$ .*
- (iii) (Sublinear convergence with diminishing stepsize) *Assume diminishing stepsize  $\alpha_t = \frac{\beta}{\gamma + t}$ , where  $\beta = \beta(\mu, L) > 0$  is a sufficiently large constant and  $\gamma > 0$  is an arbitrary constant. Then  $\|\mathbf{B}_t - \mathbf{B}^*\|^2 \leq \frac{\nu}{\gamma + t}$  for all  $t \geq 1$  for some constant  $\nu > 0$ .*

Our algorithm 1 with soft  $L_2$ -penalization and hard constraint with  $L_1$ -ball is guaranteed to converge to the true solution  $\mathbf{B}^*$  exponentially fast provided stepsizes  $\alpha_t$  are fixed and sufficiently small.

## 3.2 Statistical Guarantees

Next, we introduce generative models for the ITR estimation problem, for which we will establish statistical estimation guarantees in Theorem 3.5. Fix a joint distribution  $\pi$  for the pair  $(\mathbf{X}, A)$  of covariates  $\mathbf{X} \in \mathbb{R}^p$  and treatment  $A \in \{1, \dots, K\}$ . Fix a true parameter  $\mathbf{B}_* \in \mathbb{R}^{p \times (K-1)}$ . Then we assume  $n$  i.i.d. samples  $(\mathbf{x}_i, a_i, y_i)$  are drawn as  $(\mathbf{x}_i, a_i) \sim \pi$  and

$$y_i | (\mathbf{x}_i, a_i) \sim \begin{cases} N\left(\frac{K-1}{K} z_i, \sigma^2\right) & \text{for continuous outcome} \\ \text{Ber}\left(\frac{\exp(z_i)}{1 + \exp(z_i)}\right) & \text{for binary outcome} \end{cases}, \quad \text{where } z_i := \mathbf{u}_{\mathbf{a}_i}^T \mathbf{B}_*^T \mathbf{x}_i, \quad (11)$$

where  $N(\mu, \sigma^2)$  denotes a normal distribution with mean  $\mu$  and variance  $\sigma^2$ , whereas  $\text{Ber}(p)$  denotes a Bernoulli distribution with mean  $p$ . We assume that  $\pi$  does not depend on  $\mathbf{B}_*$  and the noise variance  $\sigma^2$  for the continuous case is known. We then seek to estimate the true parameter  $\mathbf{B}_*$  from the observed samples. It is easy to check that the negative log-likelihood function (up to additive constants) coincides with the per-sample loss function  $\ell$  in (4). Therefore, the weighted convex optimization problem (3) corresponds to weighted MLE under the generative models in (11) (see Section H in the appendix for details).

We assume that the weighted gradient of the per-sample loss  $\ell$  in (4) has uniformly bounded third moments for our statistical analysis of the generative ITR model above.



**Assumption 3.4** (Bounded moments at  $\mathbf{B}_\star$  and weights). Suppose  $(\mathbf{x}_i, a_i, y_i)$  follows the generative model above. Denote  $U_i := w_i \nabla_{\mathbf{B}} \ell(y_i, \mathbf{x}_i, a_i; \mathbf{B}_\star) \in \mathbb{R}^{p \times (K-1)}$  for  $w_i := w(a_i, \mathbf{x}_i)$  and  $\bar{U}_i := U_i - \mathbb{E}[U_i]$ . Suppose there are constants  $D_1, d_1 \in (0, \infty)$  such that  $\mathbb{E}[\|\bar{U}_i\|_F^3] < D_1$  and  $\min_{k,l} \text{Var}(\bar{U}_i(k, l)) > d_1$ . Also,  $\sum_{i=1}^n w_i^3 / (\sum_{i=1}^n w_i^2)^{3/2} = O(n^{-1/2})$ .

Now we state our main result regarding the statistical estimation guarantee for the true generative model parameter  $\mathbf{B}_\star$ . With high probability, Algorithm 1 can recover  $\mathbf{B}_\star$  up to a statistical error  $O(n^{-1/2})$ .

**Theorem 3.5** (Statistical estimation guarantee). *Let  $(\mathbf{x}_i, a_i, y_i)_{i=1}^n$  be i.i.d. observations from the generative models in (11) with true parameter  $\mathbf{B}_\star \in \mathbb{R}^{p \times (K-1)}$  such that  $\|\mathbf{B}_\star\|_1 \leq \lambda_1$  for some  $\lambda_1 \geq 0$ . Suppose Assumptions 3.1, 3.2, and 3.4 hold. Let  $\hat{\mathbf{B}}_T$  denote the weighted MLE obtained after  $T$  iterations of PGD algorithm (7) for (3). Fix a constant  $\varepsilon > 0$ .*

*Then there exists a constant  $C = C(\varepsilon) > 0$  such that with probability at least  $1 - \varepsilon$  and for  $\mu$  as in (10) but with  $\lambda^\pm = \lambda^\pm(\mathbb{E}[\Psi])$ ,*

$$\|\mathbf{B}_\star - \hat{\mathbf{B}}_T\|_F \leq \frac{C \sqrt{(p/n) \log \varepsilon^{-1}}}{\mu} + \frac{8\lambda_2 \|\mathbf{B}_\star\|_F}{\mu}, \quad (12)$$

*provided  $n, T$  are large enough. More specifically, the above holds when*

- (i) (Sample complexity)  $n \geq 1$  large s.t.  $\frac{(\sum_{i=1}^n w_i^2)^3}{(\sum_{i=1}^n w_i^3)^2} \geq C_1 \varepsilon^{-2}$  for some explicit constant  $C_1 > 0$ ;
- (ii) (Computational complexity)  $T = O(\log n)$  and  $O(n^{-1})$  when constant or diminishing stepsize as in Theorem 3.3 is used, respectively.

*Furthermore, we get  $\sqrt{n}$ -consistent estimation whenever  $\lambda_2 = O(n^{-1/2} \|\mathbf{B}_\star\|_F^{-1})$ .*

Notice that both terms in the error bound (12) are proportional to  $1/\mu$ , which depends both on the minimum eigenvalue  $\lambda^-$  of  $\mathbb{E}[\Psi]$  as well as the  $L_2$ -regularization parameter  $\lambda_2$ . Therefore, choosing the weighting function  $\omega(A, \mathbf{X})$  to maximize  $\lambda^-$  helps minimize the overall statistical estimation error. (Detailed analysis and discussion on this point can be found in the appendix H.) Another easy way to increase  $\mu$  is to use large  $L_2$ -regularization, but as in the second term in (12), using  $L_2$ -regularization gives estimation bias of  $O(\lambda_2 \|\mathbf{B}_\star\|_F / \mu)$ . Hence, we need to choose  $\lambda_2$  small enough so that this bias term is also of  $O(n^{-1/2})$ . Then we obtain an overall  $\sqrt{n}$ -consistent estimator.

The key insight in our proof of Theorem 3.5 is that we can decompose the total estimation error  $\|\mathbf{B}_\star - \hat{\mathbf{B}}_T\|_F$  into computational and statistical parts:

$$\|\mathbf{B}_\star - \hat{\mathbf{B}}_T\|_F \leq \underbrace{\|\hat{\mathbf{B}} - \hat{\mathbf{B}}_T\|_F}_{\text{computational error}} + \underbrace{\|\mathbf{B}_\star - \hat{\mathbf{B}}\|_F}_{\text{statistical error}}, \quad (13)$$

where  $\hat{\mathbf{B}}$  denotes the exact MLE. The computational error vanishes as the number  $T$  of PGD iterations tends to infinity according to Theorem 3.3. For the statistical error, we show

$$\mathbb{P} \left( \|\mathbf{B}_\star - \hat{\mathbf{B}}\|_F \leq \frac{2C}{\sqrt{n}} + \frac{8\lambda_2 \|\mathbf{B}_\star\|_F}{\mu} \right) \geq 1 - O \left( \frac{\sum_{i=1}^n w_i^3}{(\sum_{i=1}^n w_i^2)^{3/2}} \right). \quad (14)$$

Thus, the ‘skewness’ of the weights  $w_i$  measured by the ratio  $\sqrt{n} \sum_{i=1}^n w_i^3 / (\sum_{i=1}^n w_i^2)^{3/2}$  (which arises from Berry-Esseen bound) acts as a scalar multiple to the standard statistical error of  $O(n^{-1/2})$  per the central limit theorem. Hence, more balanced weights lead to fewer statistical estimation errors. If this skewness ratio is uniformly bounded, we can ensure a small statistical error with probability at least  $1 - \varepsilon$ , provided we have at least  $O(\varepsilon^{-2})$  samples. Then, we simply need to ensure the PGD iteration  $T$  is large enough so that the computational error is at most  $C/\sqrt{n}$  to achieve the high-probability total estimation error bound in Theorem 3.5. Another insight from the theorem is that maximizing  $\lambda_{\min}(\mathbb{E}[\Psi])$  promotes heterogeneous balancing weights  $w_i$ , and we get increased sample complexity through the bound  $\frac{(\sum_{i=1}^n w_i^2)^3}{(\sum_{i=1}^n w_i^3)^2} \geq C_1 \varepsilon^{-2}$  (e.g., if  $w_i \equiv 1$ , this is  $n \gtrsim \varepsilon^{-2}$ , but if  $w_1 = n, w_2 = \dots = w_n = 0$ , this may not be satisfied for any  $n$ ).

## 4 Simulation Studies

We investigate four simulation settings, each designed to evaluate various factors influencing the performance of different ITR estimators. For each setting, we mimic a randomized trial (no-confounding) and an observational study (with confounding) as per [44]. We consider 4 treatments with sample sizes  $n = 200, 600, 1000$ , and covariate dimensions  $p = 20, 40, 60$ . The outcome  $Y$  follows the model (1) with  $\epsilon|A, \mathbf{X} \sim N(0, \sigma^2(A, \mathbf{X}))$ . We evaluate method performance using accuracy (i.e., correct identification of the optimal treatment for each observation) and empirical value on a test dataset of 10,000 observations. Each simulation setting is replicated independently 100 times.

In this section, we provide partial experiment results according to the proposed estimation framework in Section 2.2. Specifically, we focus on the results from two scenarios: one mimicking randomized trials with linear ITR as the optimal rule and the other mimicking observational studies with nonlinear ITR as the optimal rule for continuous outcome. Since our proposed method assumes linear treatment rule class, the latter involves a situation where our model’s prespecified treatment rule class is misspecified. Specifically, we use following treatment-free effect function  $\mu(\mathbf{X})$  and interaction effect function  $\delta(\mathbf{X})$  for each scenario:

1. Randomized Trial: Linear ITR as the true optimal

$$\mu(\mathbf{X}) = 1 + 2X_1 + 2X_2,$$

$$\delta(\mathbf{X}) = \begin{cases} 0.75 + 1.5X_1 + 1.5X_2 + 1.5X_3 + 1.5X_4, & A = 1; \\ 0.75 + 1.5X_1 - 1.5X_2 - 1.5X_3 + 1.5X_4, & A = 2; \\ 0.75 + 1.5X_1 - 1.5X_2 + 1.5X_3 - 1.5X_4, & A = 3; \\ 0.75 - 1.5X_1 + 1.5X_2 - 1.5X_3 - 1.5X_4, & A = 4, \end{cases}$$

2. Observational Study: Nonlinear ITR as the true optimal

$$\mu(\mathbf{X}) = 1 + 2X_1 + 2X_2 + 2X_4 - 2X_4^2 + 2X_1X_2 + 2e^{-X_1X_2} + \sin(X_3),$$

$$\delta(\mathbf{X}) = \begin{cases} 0.5 + 1.0X_1 - 2.0X_4 + 0.5X_4^2, & A = 1; \\ 1.0 + 1.0X_1 + 1.0X_4 - 1.0X_4^2, & A = 2; \\ 1.5 + 2.0X_1 - 1.0X_4 - 1.0X_4^2, & A = 3; \\ 1.0 - 1.0X_1 - 1.0X_4 - 1.0X_4^2, & A = 4. \end{cases}$$

in the working model (1). Further details about simulation settings, other scenarios, and corresponding results are available in the appendix I.

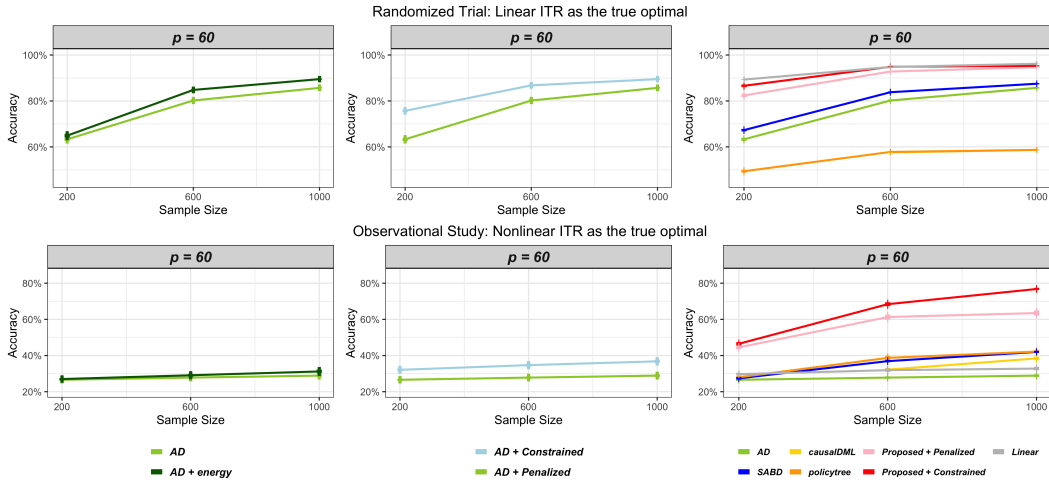


Figure 1: **Accuracy Comparison: proposed methods vs. benchmark methods.** All subplots in the same row share the same simulation setting, focusing on randomized trials with linear ITR as the true optimal rule (top) and observational studies with nonlinear ITR as the true optimal rule (bottom). Each subplot presents (Left) accuracy comparisons based on weights, illustrating the difference between the standard IPW approach of AD-Learning with the proposed approach using EBW, (Middle) accuracy comparisons based on optimization algorithms, illustrating the difference between the standard  $L_1$ -penalized approach against the proposed constrained optimization with PGD, (Right) evaluation of accuracies between existing standard approaches and the proposed method, which integrates EBWs, variance and dimension reduction techniques implemented through PGD. Error bars represent the standard errors of the mean (SEM) of accuracies across multiple simulations.



**Using Distributional Covariate Balancing Weights.** In the left subplots of Figure 1, the incorporation of DCBWs (e.g., EBWs) shows improvements in classification rates for ITR-Learning. While EBWs indeed provide more effective balancing weights in observational studies, as depicted in the lower left subplot in Figure 1, it also results in improvements even in randomized trial settings where the true propensity score is known. This is achieved by reducing finite sample imbalances, as shown in the upper left subplot in Figure 1. Moreover, when the underlying true treatment rule class is nonlinear, the performance of AD-Learning suffers due to a misalignment between the assumed linear treatment rule class and the actual class. In this scenario, using EBWs yields accuracy improvements compared to the conventional IPW approach. Further details can be found in the appendix I.

**Using PGD for Constrained Optimization.** The middle subplots of Figure 1 show accuracy comparisons between standard penalized approach and constrained optimization using PGD. Both optimization algorithms are based on the same proposed statistical approach, integrating EBWs with SABD-Learning, outcome augmentation, and additional variable screening. For solving the  $L_1$ -penalized regression problems, the R package *glmnet* is a standard choice, which is based on cyclic coordinate descent with simulated-annealing-style hyperparameter tuning. However, *glmnet* does not handle hard  $L_1$ -ball constraints required for our method. Therefore, to ensure a fair comparison, we directly implemented standard (projected) (sub)gradient descent to solve optimization problems for both penalized and constrained problems. In our experiment reported in Figure 1, we observe significant improvements using our PGD method, especially in the second scenario. One possible explanation for this improvement is that large penalization may yield a perturbed solution to the ITR problem. If the minimizer of the ITR problem lies on the boundary of the  $L_1$ -ball, the solutions of constrained optimization and penalized optimization coincide. However, if the minimizer is the interior point of  $L_1$ -ball, constrained optimization can find the exact solution, while the penalization may induce a perturbed solution. Furthermore, the trajectory of PGD for the constraint approach appears to be significantly more stable than that of the subgradient descent in the penalization approach (see Figure 2).

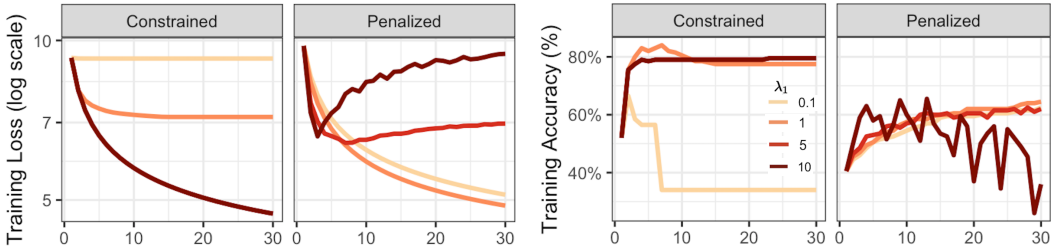


Figure 2: **Comparison of Optimization Methods by  $\lambda_1$ .**  $\lambda_1$  is  $L_1$ -ball size of the constraint set in the constrained optimization and the regularization parameter with additional  $L_1$ -regularization of the model parameter in the penalized regression, respectively.

**Performance Evaluations between the Proposed Method and Benchmark Methods.** In the right subplots of Figure 1, we evaluate accuracies by comparing benchmark methods to our proposed approach. The benchmarks include AD-Learning, SABD-Learning, a treatment-covariate interaction model with  $L_1$ -regularization ‘linear’, and two popular tree-based methods: ‘policytree’ [47, 59] and double-machine learning-based ‘causalDML’ [10, 29]. Our proposed method integrates EBWs with SABD-Learning, outcome augmentation, and additional variable screening for variance and dimension reduction, implemented via PGD. However, causalDML requires a sufficiently large sample size, thereby making its evaluation unfeasible under a sample size of 200. Our numerical analysis provides evidence that our proposed approach outperforms existing methods in both scenarios. ‘linear’ performs well in the upper subplot, which simulates a randomized trial with a true optimal linear rule, as covariates are balanced and the model accurately captures the decision function. However, in the lower subplot, simulating an observational study with a nonlinear rule, the model performs poorly because it lacks balancing weights to address confounding and cannot fully capture nonlinear relationships. Further detailed analysis, including the effectiveness of additional techniques such as variable screening for performance improvement, is provided in the appendix I.

Additionally, our proposed method processes each dataset ( $p = 60$ ,  $n = 1000$ ) in an average of 4.40 seconds, compared to 2.01 seconds for the penalization method using *glmnet* in R. Other methods exhibit scalability challenges, with policytree averaging 12.02 seconds and causalDML taking 46.75 seconds. This demonstrates that computational efficiency is a notable advantage of our approach.

## 5 Applications

We apply the proposed methods to two datasets from AIDS Clinical Trials Group (ACTG) 175 [21] and email marketing [22]. ACTG is a randomized trial for 2,139 patients with HIV infection who were randomly assigned to four different treatments. Twelve covariates including age and gender were used for the analysis. The outcome is the change in CD4 cell count from baseline to 20 weeks, where larger values are preferable. Email marketing dataset has 64,000 customers who were randomly chosen to receive an e-mail campaign among three different marketing methods. Eight covariates with historical customer attributes and their pairwise interactions were used. The outcome is whether customers visited the website in the following two weeks. The goal of the analysis using these two datasets is to find the best treatment/marketing methods based on individuals’ attributes. Similar to [44], we randomly split the data into a training set of {200, 400, 800, 1000, 1200} observations for the ACTG dataset and {1000, 3000, 5000} observations for the email dataset. The remaining observations were used for test data with 10 iterations. We evaluate method performance using the empirical value function on a test dataset. The benchmarking methods include AD-Learning, SABD-Learning, and two tree-based methods (policytree [47] and causalDML [29]). However, causalDML requires a sufficiently large sample size, thereby making its evaluation unfeasible under a sample size of 200. For the binary outcome in Email dataset, SABD-Learning is not used, since the homoscedastic assumption is unnecessary for logistic regression. The results are shown in Table 1.

Table 1: Average empirical value functions across different approaches for ACTG/Email datasets. Mean values with the corresponding standard errors of the mean (SEM) in parentheses are provided. The highest-performing methods are marked in bold.

Dataset	Training size	AD	SABD	policytree	causalDML	Proposed
ACTG	200	37.624 (6.9)	33.645 (6.7)	36.871 (4.6)	-	<b>42.301</b> (5.7)
	400	51.329 (3.2)	53.644 (1.9)	47.191 (2)	49.552 (0.2)	<b>54.868</b> (3.4)
	800	56.021 (1.5)	56.700 (1.6)	52.786 (2.6)	55.506 (2.1)	<b>59.200</b> (2.5)
	1000	56.825 (2.2)	56.963 (2.1)	56.005 (1.7)	56.591 (2.1)	<b>59.552</b> (2.5)
	1200	55.092 (2.6)	55.221 (2.8)	<b>58.819</b> (3.4)	57.104 (2.8)	58.207 (2.9)
Email	1000	0.169 (0.0)	-	0.171 (0.0)	0.172 (0.0)	<b>0.175</b> (0.0)
	3000	0.176 (0.0)	-	0.177 (0.0)	0.177 (0.0)	<b>0.181</b> (0.0)
	5000	0.177 (0.0)	-	0.178 (0.0)	0.178 (0.0)	<b>0.182</b> (0.0)

## 6 Discussion and Limitations

Our method introduces a unified, robust framework for estimating ITR. By formulating the problem as a weighted, constrained optimization problem, we incorporate DCBWs to control confounding and propose the PGD algorithm for sparse ITRs with computational and statistical guarantees. Additionally, we propose variable screening and efficient augmentation, which together show synergistic effects. We demonstrate our method with continuous and binary outcomes, and it can be readily extended to other outcomes, such as censored outcomes. Although we focus on learning linear ITR to facilitate the demonstration of our idea, it can be extended to nonlinear ITR using the basis function [34]. For example, by taking the covariate powers up to  $M$ , the linear function class becomes the class of degree- $M$  polynomial functions. The linear function class can also serve as a useful approximation to nonlinear decision functions.

We acknowledge that although most, if not all, types of ITR-Learning frameworks can benefit from using DCBWs, variable screening, and augmentation, our proposed framework for theoretical analysis relies on the specific assumption that the model can be formulated as a constrained, weighted, and convex optimization problem. Furthermore, we restricted the decision function to the linear function class. Generalizing our theoretical analysis to include nonlinear function classes with nonconvex problems, such as boosting or deep neural networks, is an interesting direction for future investigation. Another limitation of our method is its reliance on standard assumptions to identify optimal ITRs using observational data, including the assumption of no unmeasured confounding. An interesting future direction would be to extend our framework to cases where this assumption does not hold, utilizing instrumental variables [12, 38] and proximal causal learning [40, 45].

Our work significantly impacts society by improving patient health outcomes through personalized care and advancing medical research in disease mechanisms and drug development. Beyond healthcare, these methods can be applied in marketing strategies based on customer characteristics.

## Acknowledgement

Lee J and Chen G’s efforts were partially supported by NSF Grant DMS-2054346.

## References

- [1] Serge Assaad, Shuxi Zeng, Chenyang Tao, Shounak Datta, Nikhil Mehta, Ricardo Henao, Fan Li, and Lawrence Carin. Counterfactual representation learning with balancing weights. In *International Conference on Artificial Intelligence and Statistics*, pages 1972–1980. PMLR, 2021.
- [2] Susan Athey and Stefan Wager. Policy learning with observational data. *Econometrica*, 89(1):133–161, 2021.
- [3] Laura B Balzer, Maya L Petersen, Mark J van der Laan, and Search Collaboration. Targeted estimation and inference for the sample average treatment effect in trials with and without pair-matching. *Statistics in medicine*, 35(21):3717–3732, 2016.
- [4] Léon Bottou. Large-scale machine learning with stochastic gradient descent. In *Proceedings of COMPSTAT’2010*, pages 177–186. Springer, 2010.
- [5] Stephen Boyd, Stephen P Boyd, and Lieven Vandenberghe. *Convex optimization*. Cambridge university press, 2004.
- [6] Stephen Boyd, Lin Xiao, and Almir Mutapcic. Subgradient methods. *lecture notes of EE392o, Stanford University, Autumn Quarter*, 2004(01), 2003.
- [7] Guanhua Chen, Donglin Zeng, and Michael R Kosorok. Personalized dose finding using outcome weighted learning. *J. Amer. Statist. Assoc.*, 111(516):1509–1521, 2016.
- [8] Rui Chen, Jared D Huling, Guanhua Chen, and Menggang Yu. Robust sample weighting to facilitate individualized treatment rule learning for a target population. *Biometrika*, 111(1):309–329, 2024.
- [9] Shuai Chen, Lu Tian, Tianxi Cai, and Menggang Yu. A general statistical framework for subgroup identification and comparative treatment scoring. *Biometrics*, 73(4):1199–1209, 2017.
- [10] Victor Chernozhukov, Denis Chetverikov, Mert Demirer, Esther Duflo, Christian Hansen, and Whitney Newey. Double/debiased/neyman machine learning of treatment effects. *Am. Econ. Rev.*, 107(5):261–65, 2017.
- [11] Francis S Collins and Harold Varmus. A new initiative on precision medicine. *New England journal of medicine*, 372(9):793–795, 2015.
- [12] Yifan Cui and Eric Tchetgen Tchetgen. A semiparametric instrumental variable approach to optimal treatment regimes under endogeneity. *Journal of the American Statistical Association*, 116(533):162–173, 2021.
- [13] Alicia Curth and Mihaela van der Schaar. Nonparametric estimation of heterogeneous treatment effects: From theory to learning algorithms. In *International Conference on Artificial Intelligence and Statistics*, pages 1810–1818. PMLR, 2021.
- [14] John Duchi, Shai Shalev-Shwartz, Yoram Singer, and Tushar Chandra. Efficient projections onto the  $l_1$ -ball for learning in high dimensions. In *Proceedings of the 25th international conference on Machine learning*, pages 272–279, 2008.
- [15] Jianqing Fan and Runze Li. Variable selection via nonconcave penalized likelihood and its oracle properties. *Journal of the American statistical Association*, 96(456):1348–1360, 2001.
- [16] Jianqing Fan and Jinchi Lv. Sure independence screening for ultrahigh dimensional feature space. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 70(5):849–911, 2008.

- [17] William Feller. *An introduction to probability theory and its applications, Volume 2*, volume 81. John Wiley & Sons, 1991.
- [18] Jean-Louis Goffin. On convergence rates of subgradient optimization methods. *Mathematical programming*, 13:329–347, 1977.
- [19] Arthur Gretton, Karsten M Borgwardt, Malte J Rasch, Bernhard Schölkopf, and Alexander Smola. A kernel two-sample test. *The Journal of Machine Learning Research*, 13(1):723–773, 2012.
- [20] Jens Hainmueller. Entropy balancing for causal effects: A multivariate reweighting method to produce balanced samples in observational studies. *Political analysis*, 20(1):25–46, 2012.
- [21] Scott M Hammer, David A Katzenstein, Michael D Hughes, Holly Gundacker, Robert T Schooley, Richard H Haubrich, W Keith Henry, Michael M Lederman, John P Phair, Manette Niu, et al. A trial comparing nucleoside monotherapy with combination therapy in hiv-infected adults with cd4 cell counts from 200 to 500 per cubic millimeter. *New England Journal of Medicine*, 335(15):1081–1090, 1996.
- [22] Kevin Hillstrom. Minethatdata e-mail analytics and data mining challenge. [https://www.uplift-modeling.com/en/v0.3.1/api/datasets/fetch\\_hillstrom.html](https://www.uplift-modeling.com/en/v0.3.1/api/datasets/fetch_hillstrom.html), 2008. Accessed: 2024-04-28.
- [23] Jared D Huling, Noah Greifer, and Guanhua Chen. Independence weights for causal inference with continuous treatments. *Journal of the American Statistical Association*, 119(546):1657–1670, 2024.
- [24] Jared D Huling and Simon Mak. Energy balancing of covariate distributions. *Journal of Causal Inference*, 12(1):20220029, 2024.
- [25] Kosuke Imai and Marc Ratkovic. Covariate balancing propensity score. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 76(1):243–263, 2014.
- [26] Nathan Kallus. Balanced policy evaluation and learning. *Advances in neural information processing systems*, 31, 2018.
- [27] Nathan Kallus. Generalized optimal matching methods for causal inference. *The Journal of Machine Learning Research*, 21(1):2300–2353, 2020.
- [28] Nathan Kallus. More efficient policy learning via optimal retargeting. *Journal of the American Statistical Association*, 116(534):646–658, 2021.
- [29] Michael C Knaus. Double machine learning-based programme evaluation under unconfoundedness. *The Econometrics Journal*, 25(3):602–627, 2022.
- [30] Brian K Lee, Justin Lessler, and Elizabeth A Stuart. Weight trimming and propensity score weighting. *PloS one*, 6(3):e18174, 2011.
- [31] Fan Li, Kari Lock Morgan, and Alan M Zaslavsky. Balancing covariates via propensity score weighting. *Journal of the American Statistical Association*, 113(521):390–400, 2018.
- [32] Runze Li, Wei Zhong, and Liping Zhu. Feature screening via distance correlation learning. *Journal of the American Statistical Association*, 107(499):1129–1139, 2012.
- [33] Haixu Ma, Donglin Zeng, and Yufeng Liu. Learning individualized treatment rules with many treatments: A supervised clustering approach using adaptive fusion. *Advances in Neural Information Processing Systems*, 35:15956–15969, 2022.
- [34] Jacob M Maronge, Jared D Huling, and Guanhua Chen. A reluctant additive model framework for interpretable nonlinear individualized treatment rules. *The Annals of Applied Statistics*, 17(4):3384–3402, 2023.
- [35] Weibin Mo and Yufeng Liu. Efficient learning of optimal individualized treatment rules for heteroscedastic or misspecified treatment-free effect models. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, 84(2):440–472, 2022.

- [36] Xinkun Nie and Stefan Wager. Quasi-oracle estimation of heterogeneous treatment effects. *Biometrika*, 108(2):299–319, 2021.
- [37] Wenliang Pan, Xueqin Wang, Heping Zhang, Hongtu Zhu, and Jin Zhu. Ball covariance: A generic measure of dependence in banach space. *Journal of the American Statistical Association*, 115(529):307–317, 2019.
- [38] Hongming Pu and Bo Zhang. Estimating optimal treatment rules with an instrumental variable: A partial identification learning approach. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, 83(2):318–345, 2021.
- [39] Zhengling Qi, Dacheng Liu, Haoda Fu, and Yufeng Liu. Multi-armed angle-based direct learning for estimating optimal individualized treatment rules with various outcomes. *Journal of the American Statistical Association*, 115(530):678–691, 2020.
- [40] Zhengling Qi, Rui Miao, and Xiaoke Zhang. Proximal learning for individualized treatment regimes under unmeasured confounding. *Journal of the American Statistical Association*, pages 1–14, 2023.
- [41] Min Qian and Susan A Murphy. Performance guarantees for individualized treatment rules. *Annals of statistics*, 39(2):1180, 2011.
- [42] Brian C Ross. Mutual information between discrete and continuous data sets. *PloS one*, 9(2):e87357, 2014.
- [43] Donald B Rubin. Randomization analysis of experimental data: The fisher randomization test comment. *Journal of the American statistical association*, 75(371):591–593, 1980.
- [44] Kushal S Shah, Haoda Fu, and Michael R Kosorok. Stabilized direct learning for efficient estimation of individualized treatment rules. *Biometrics*, 2022. doi: 10.1111/biom.13818.
- [45] Tao Shen and Yifan Cui. Optimal treatment regimes for proximal causal learning. *Advances in Neural Information Processing Systems*, 36, 2024.
- [46] Claudia Shi, David Blei, and Victor Veitch. Adapting neural networks for the estimation of treatment effects. *Advances in neural information processing systems*, 32, 2019.
- [47] Erik Sverdrup, Ayush Kanodia, Zhengyuan Zhou, Susan Athey, and Stefan Wager. policytree: Policy learning via doubly robust empirical welfare maximization over trees. *Journal of Open Source Software*, 5(50):2232, 2020.
- [48] Gábor J Székely, Maria L Rizzo, and Nail K Bakirov. Measuring and testing dependence by correlation of distances. *The Annals of Statistics*, 35(6):2769–2794, 2007.
- [49] Dingke Tang, Dehan Kong, Wenliang Pan, and Linbo Wang. Ultra-high dimensional variable selection for doubly robust causal inference. *Biometrics*, 79(2):903–914, 2023.
- [50] Lu Tian, Ash A Alizadeh, Andrew J Gentles, and Robert Tibshirani. A simple method for estimating interactions between a treatment and a large number of covariates. *Journal of the American Statistical Association*, 109(508):1517–1532, 2014.
- [51] Roman Vershynin. *High-dimensional probability: An introduction with applications in data science*, volume 47. Cambridge university press, 2018.
- [52] Cédric Villani et al. *Optimal transport: old and new*, volume 338. Springer, 2009.
- [53] Stefan Wager and Susan Athey. Estimation and inference of heterogeneous treatment effects using random forests. *Journal of the American Statistical Association*, 113(523):1228–1242, 2018.
- [54] Baqun Zhang, Anastasios A Tsiatis, Eric B Laber, and Marie Davidian. A robust method for estimating optimal treatment regimes. *Biometrics*, 68(4):1010–1018, 2012.
- [55] Chong Zhang and Yufeng Liu. Multicategory angle-based large-margin classification. *Biometrika*, 101(3):625–640, 2014.

- [56] Qingyuan Zhao, Dylan S Small, and Ashkan Ertefaie. Selective inference for effect modification via the lasso. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, 84(2):382–413, 2022.
- [57] Yingqi Zhao, Donglin Zeng, A John Rush, and Michael R Kosorok. Estimating individualized treatment rules using outcome weighted learning. *Journal of the American Statistical Association*, 107(499):1106–1118, 2012.
- [58] Xin Zhou, Nicole Mayer-Hamblett, Umer Khan, and Michael R Kosorok. Residual weighted learning for estimating individualized treatment rules. *Journal of the American Statistical Association*, 112(517):169–187, 2017.
- [59] Zhengyuan Zhou, Susan Athey, and Stefan Wager. Offline multi-action policy learning: Generalization and optimization. *Operations Research*, 71(1):148–183, 2023.
- [60] Ruoqing Zhu, Ying-Qi Zhao, Guanhua Chen, Shuangge Ma, and Hongyu Zhao. Greedy outcome weighted tree learning of optimal personalized treatment rules. *Biometrics*, 73(2):391–400, 2017.
- [61] José R Zubizarreta. Stable weights that balance covariates for estimation with incomplete outcome data. *Journal of the American Statistical Association*, 110(511):910–922, 2015.



## A Preliminaries

In this section, we review several notations and basic facts on linear algebra and matrix calculus.

The Kronecker product, denoted by  $\otimes$ , is a binary operation that combines two matrices to create a new matrix. Given two matrices  $\mathbf{A} = (a_{ij}) \in \mathbb{R}^{m \times n}$  and  $\mathbf{B} \in \mathbb{R}^{p \times q}$ , then  $\mathbf{A} \otimes \mathbf{B}$  is an  $mp \times nq$  matrix defined as

$$\mathbf{A} \otimes \mathbf{B} = \begin{bmatrix} a_{11}\mathbf{B} & \cdots & a_{1n}\mathbf{B} \\ \vdots & \ddots & \vdots \\ a_{m1}\mathbf{B} & \cdots & a_{mn}\mathbf{B} \end{bmatrix} \quad (15)$$

If  $\mathbf{A}' \in \mathbb{R}^{m \times n'}$ , then for the horizontally stacked matrix  $[\mathbf{A}, \mathbf{A}'] \in \mathbb{R}^{m \times (n+n')}$ ,

$$[\mathbf{A}, \mathbf{A}'] \otimes \mathbf{B} = [\mathbf{A} \otimes \mathbf{B}, \mathbf{A}' \otimes \mathbf{B}]. \quad (16)$$

For each  $\mathbf{A} = [\mathbf{a}_1, \dots, \mathbf{a}_n] \in \mathbb{R}^{m \times n}$ , let  $\text{vec}(\mathbf{A}) := [\mathbf{a}_1^T, \dots, \mathbf{a}_n^T]^T \in \mathbb{R}^{mn}$  denote its vectorization. The *commutation matrix*  $\mathbf{C}^{(a,b)}$ , which is a special instance of  $ab \times ab$  permutation matrix. Namely, for each integers  $a, b \geq 1$ , there exists a unique matrix  $\mathbf{C}^{(a,b)} \in \{0, 1\}^{ab \times ab}$  such that for all  $\mathbf{A} \in \mathbb{R}^{a \times b}$ , we have  $\mathbf{C}^{(a,b)} \text{vec}(\mathbf{A}) = \text{vec}(\mathbf{A}^T)$ . Note that  $(\mathbf{C}^{(a,b)})^T = \mathbf{C}^{(b,a)}$ . Furthermore,  $(\mathbf{C}^{(a,b)})^T \mathbf{C}^{(a,b)} = \mathbf{I}_{ab}$  since  $(\mathbf{C}^{(a,b)})^T \mathbf{C}^{(a,b)} \text{vec}(\mathbf{A}) = \mathbf{C}^{(b,a)} \text{vec}(\mathbf{A}^T) = \text{vec}(\mathbf{A})$ . Hence  $\mathbf{C}^{(a,b)}$  is positive semi-definite. Note that  $\mathbf{C}^{(a,1)} = \mathbf{I}_a = \mathbf{C}^{(1,a)}$  since if  $\mathbf{A} \in \mathbb{R}^{a \times 1}$  or  $\mathbf{A} \in \mathbb{R}^{1 \times a}$ ,  $\text{vec}(\mathbf{A}) = \text{vec}(\mathbf{A}^T)$ . Also, for every matrices  $\mathbf{A} \in \mathbb{R}^{m \times n}$  and  $\mathbf{B} \in \mathbb{R}^{r \times q}$ ,

$$\mathbf{C}^{(r,m)}(\mathbf{A} \otimes \mathbf{B}) = (\mathbf{B} \otimes \mathbf{A})\mathbf{C}^{(q,n)}. \quad (17)$$

Also recall the relations for vectorizing product of matrices: for  $\mathbf{A} \in \mathbb{R}^{a \times b}$ ,  $\mathbf{B} \in \mathbb{R}^{b \times c}$ , and  $\mathbf{C} \in \mathbb{R}^{c \times d}$ ,

$$\text{vec}(\mathbf{AB}) = (\mathbf{I}_c \otimes \mathbf{A}) \text{vec}(\mathbf{B}) = (\mathbf{B}^T \otimes \mathbf{I}_a) \text{vec}(\mathbf{A}), \quad (18)$$

$$\text{vec}(\mathbf{ABC}) = (\mathbf{C}^T \otimes \mathbf{A}) \text{vec}(\mathbf{B}) = (\mathbf{I}_d \otimes \mathbf{AB}) \text{vec}(\mathbf{C}) = (\mathbf{C}^T \mathbf{B}^T \otimes \mathbf{I}_a) \text{vec}(\mathbf{A}). \quad (19)$$

For  $\mathbf{A}, \mathbf{B}, \mathbf{C}$ , and  $\mathbf{D}$  are matrices of compatible sizes allowing the matrix products  $\mathbf{AC}$  and  $\mathbf{BD}$ ,

$$(\mathbf{A} \otimes \mathbf{B})(\mathbf{C} \otimes \mathbf{D}) = (\mathbf{AC} \otimes \mathbf{BD}).$$

Next, recall that if  $f : \mathbb{R}^{a \times 1} \rightarrow \mathbb{R}^{b \times 1}$  and  $g : \mathbb{R}^{b \times 1} \rightarrow \mathbb{R}^{c \times 1}$  are differentiable functions, then the Jacobian  $J_f$  can be written by using gradients as  $J_f(x) = (\nabla_x f^T)^T$ . By chain rule  $J_{g \circ f}(x) = J_g(f(x))J_f(x)$ , we have the following chain rule for gradients

$$\nabla_x (g(f(x))^T) = \nabla_x (f(x)^T) \nabla_{f(x)} (g(f(x))^T). \quad (20)$$

We will also frequently use the following fact: For  $\mathbf{A} \in \mathbb{R}^{a \times b}$ ,

$$\nabla_{\text{vec}(\mathbf{A})} \text{vec}(\mathbf{A})^T = \mathbf{I}_{ab}, \quad (21)$$

$$\nabla_{\text{vec}(\mathbf{A})} \text{vec}(\mathbf{A}^T)^T = \nabla_{\text{vec}(\mathbf{A})} \text{vec}(\mathbf{A})^T \mathbf{C}^{(b,a)} = \mathbf{C}^{(b,a)}. \quad (22)$$

For two matrices  $\mathbf{A}, \mathbf{B}$  of the same size, we write  $\mathbf{A} \preceq \mathbf{B}$  if  $\mathbf{B} - \mathbf{A}$  is positive semi-definite. The partial ordering  $\preceq$  is called the *Loewner ordering*.

## B Details of model setup

### B.1 Details of angle-based approach

Consider the working model given by Equation (1) in the main text:

$$Y = \mu(\mathbf{X}) + \sum_{k=1}^K \delta_k(\mathbf{X}) \mathbf{I}(A = k) + \epsilon.$$

To make the optimal ITR identifiable, we impose the constraint  $\sum_{k=1}^K \delta_k(\mathbf{x}) = 0$  for all covariates  $\mathbf{x}$ . Under this model,  $\delta_k(\mathbf{x})$  determines the optimal ITR for an individual with covariate  $\mathbf{x}$ , while  $\mu(\mathbf{x})$  has no impact on the ITR. However, using  $K$  functions with the sum-to-zero constraint can be computationally inefficient [55]. Instead, one can opt for simplex coding as an alternative, which inherently satisfies the sum-to-zero constraint.

In AD-Learning, proposed by [39], each of the  $K$  treatments is represented as a vertex simplex on  $\mathbb{R}^{K-1}$ , denoted as  $\mathbf{u}_k$ ,  $k = 1, \dots, K$ . In particular,  $\mathbf{u}_k$  is defined as

$$\mathbf{u}_k = \begin{cases} \frac{1}{\sqrt{K-1}} \mathbf{1}_{K-1} & \text{if } k = 1, \\ -\frac{1+\sqrt{K}}{\sqrt{(K-1)^3}} \mathbf{1}_{K-1} + \sqrt{\frac{K}{K-1}} \mathbf{e}_{k-1} & \text{if } k \geq 2, \end{cases}$$

where  $\mathbf{1}_{K-1}$  is a  $K-1$  dimensional vector with entries 1, and  $\mathbf{e}_{k-1}$  is a  $K-1$  dimensional vector with entries 0 except its  $k$ -th entry being 1. This vertex simplex has  $K$  vertices with equal angles between them and an origin at the center of the simplex. All  $\mathbf{u}_k$  have the same Euclidean norm 1. To estimate the optimal ITR, a decision function  $\mathbf{f}(\mathbf{x}) = (f_1(\mathbf{x}), \dots, f_{K-1}(\mathbf{x}))^T$  is constructed, assigning a  $K-1$  dimensional vector to each covariate  $\mathbf{x}$ . Here,  $\mathbf{f}(\cdot)$  can take a generic form, while we assume the linear decision function in the main text. The optimal ITR is reformulated by the authors as follows:

$$d^{\text{opt}}(\mathbf{x}) = \arg \max_{k \in \{1, \dots, K\}} \mathbf{u}_k^T \mathbf{f}(\mathbf{x}) = \arg \min_{k \in \{1, \dots, K\}} \angle(\mathbf{u}_k, \mathbf{f}(\mathbf{x})),$$

where  $\angle(\cdot, \cdot)$  represents the angle between two vectors. Note that due to  $\sum_{k=1}^K \mathbf{u}_k = \mathbf{0}$ ,  $\sum_{k=1}^K \mathbf{u}_k^T \mathbf{f}(\mathbf{x}) = 0$  for any given function  $\mathbf{f}$  and covariate value  $\mathbf{x}$ . Then we can use

$$\left( \sum_{k=1}^K \mathbf{u}_k \mathbf{I}(A = k) \right)^T \mathbf{f}(\mathbf{x})$$

instead of  $\sum_{k=1}^K \delta_k(\mathbf{X}) \mathbf{I}(A = k)$  to encode the treatment and covariate interaction. Specifically, we use  $\mathbf{u}_A$  to denote  $\left( \sum_{k=1}^K \mathbf{u}_k \mathbf{I}(A = k) \right)^T$  and assume  $\mathbf{f}(\mathbf{x})$  to be linear in  $\mathbf{x}$  (with the coefficient matrix  $\mathbf{B}$ ). Therefore, terms pertaining to ITRs can be expressed as  $\mathbf{u}_{a_i}^T \mathbf{B}^T \mathbf{x}_i$ .

### B.2 Details of binary outcome

Again, under prespecified treatment rule class  $\mathcal{D}$ , the optimal ITR is given by

$$d^{\text{opt}}(\mathbf{x}) = \arg \max_{k \in \{1, \dots, K\}} \mathbb{E}[Y | \mathbf{x}, A = k] = \arg \max_{k \in \{1, \dots, K\}} \mathbf{u}_k^T \underbrace{\mathbb{E} \left[ \frac{Y \mathbf{u}}{\pi(A, \mathbf{x})} \middle| \mathbf{x} \right]}_{=: \mathbf{f}_{\text{opt}}(\mathbf{x})}.$$

For the binary response setting, we assume the following logistic model

$$\mathbb{P}[Y = 1 | \mathbf{x}, A = k] = \frac{\exp(\mathbf{u}_k^T \mathbf{f}_{\text{opt}}(\mathbf{x}))}{1 + \exp(\mathbf{u}_k^T \mathbf{f}_{\text{opt}}(\mathbf{x}))}.$$

Under this model,  $\mathbf{f}_{\text{opt}}$  is indeed the optimal decision function under the angle-based direct learning framework since

$$\begin{aligned}
d^{\text{opt}}(\mathbf{x}) &= \arg \max_{k \in \{1, \dots, K\}} \mathbb{P}[Y = 1 | \mathbf{x}, A = k] \\
&= \arg \max_{k \in \{1, \dots, K\}} \frac{\exp(\mathbf{u}_k^T \mathbf{f}_{\text{opt}}(\mathbf{x}))}{1 + \exp(\mathbf{u}_k^T \mathbf{f}_{\text{opt}}(\mathbf{x}))} \\
&= \arg \max_{k \in \{1, \dots, K\}} \mathbf{u}_k^T \mathbf{f}_{\text{opt}}(\mathbf{x}).
\end{aligned} \tag{23}$$

Furthermore, it can be checked that the function  $\mathbf{f}_{\text{opt}}$  is an optimal solution to maximizing the IPW-weighted log-likelihood of the logistic regression model:

$$\arg \min_{\mathbf{f}} \mathbb{E} \left[ -\frac{Y \mathbf{u}^T \mathbf{f}}{\pi(A, \mathbf{x})} + \frac{\log(\exp(\mathbf{u}^T \mathbf{f}) + 1)}{\pi(A, \mathbf{x})} \right], \tag{24}$$

under the assumption that the expectation and derivative w.r.t.  $\mathbf{f}$  can be exchanged ([39]).

For the effective and robust ITR estimation, replacing IPW with DCBW leads to

$$\arg \min_{\mathbf{f}} \mathbb{E} [w(\mathbf{x}, A) \{-Y \mathbf{u}^T \mathbf{f} + \log(\exp(\mathbf{u}^T \mathbf{f}) + 1)\}]. \tag{25}$$

Now consider restricting it to the linear decision function class  $\mathcal{F} = \{\mathbf{f}(\mathbf{X}) = \mathbf{B}^T \mathbf{X} : \mathbf{B} \in \mathbb{R}^{p \times (K-1)}\}$ . Then the above (25) becomes

$$\arg \max_{\mathbf{B} \in \mathbb{R}^{p \times (K-1)}} \mathbb{E} [w(\mathbf{x}, A) \{Y \mathbf{u}^T \mathbf{B}^T \mathbf{x} - \log(\exp(\mathbf{u}^T \mathbf{B}^T \mathbf{x}) + 1)\}].$$

Then, we can estimate the optimal parameter  $\mathbf{B}$  by solving the following empirical loss minimization

$$\arg \min_{\mathbf{B} \in \mathbb{R}^{p \times (K-1)}} \frac{1}{n} \sum_{i=1}^n w(\mathbf{x}_i, a_i) (-y_i \mathbf{u}_i^T \mathbf{B}^T \mathbf{x}_i + \log(\exp(\mathbf{u}_i^T \mathbf{B}^T \mathbf{x}_i) + 1)) \quad \text{s.t.} \quad \|\mathbf{B}\|_1 \leq z, \tag{26}$$

where we imposed additionally the  $L_1$ -penalization  $\|\mathbf{B}\|_1 \leq z$ , on  $\mathbf{B}$  to promote a sparse solution. By the proposed algorithm 1, we can efficiently solve the problem (26).

## C Details of proposed approaches

To keep the appendix as self-contained as possible, we provide descriptions of the building blocks of the proposed methods.

### C.1 Use of distributional covariate balancing weights

In this section, we provide the background about the use of DCBWs for constructing effective and robust ITRs. For additional details, please refer to the original papers.

Many existing ITR-Learning approaches such as AD-Learning and SABD-Learning use *inverse probability (of treatment) weighting* (IPW), which requires specifying a propensity score model for confounding control. However, propensity score methods have long been known to be highly sensitive to model misspecification, which yields biased estimates of causal effects [25, 30]. For this reason, methods that directly estimate weights based on balance-seeking objectives instead of using IPW have been proposed. These approaches include entropy balancing weights [20], stable balancing weights [61] as well as DCBWs [24, 27]. However, despite the weights serving a critical role in the estimation process, these advancements have mainly focused on improving *average treatment effect* (ATE) and are largely under-explored in the ITR literature. Significant improvements in the performance of ITRs are often left on the table when the weights are not treated with the same level of emphasis and focus as the other aspects of ITR learning.

To tackle these challenges, we propose the use of DCBWs, such as *energy balancing weights* (EBWs) [24] and *maximum mean discrepancy* (MMD) balancing weights [19, 27, 8], as an alternative to IPW for ITR-Learning. Unlike traditional approaches, these methods do not rely on any pre-specified functional forms for constructing weights. Instead, they focus on minimizing the distance of the weighted empirical distributions of covariates across treatment groups to reduce potential confounding. Even for data from randomized trials, they can help reduce finite sample covariate imbalance. The emphasis on finite sample balance and the model-independent nature of distributional weighting methods effectively mitigate biases due to model misspecification, resulting in more precise and reliable ITR estimation.

In the remainder of this section, we provide additional details about EBWs. The EBWs are derived from the concept of “weighted energy distance” which modifies the energy distance measure to allow for measurement of the distance between a weighted empirical distribution and a target empirical distribution. This measure captures the distance between a weighted *empirical cumulative distribution function* (ECDF) of covariates for a specific treatment group  $a$ , denoted as  $\{\mathbf{X}_i\}_{i:A_i=a}$ , and the combined ECDF of covariates, denoted as  $\{\mathbf{X}_i\}_{i=1}^n$ . Specifically, the weighted energy distance is defined as:

$$\begin{aligned} \mathcal{E}(F_{n,a,w}, F_n) &\equiv \frac{2}{n_a n} \sum_{i=1}^n \sum_{j=1}^n w_i \mathbf{I}(A_i = a) \|\mathbf{X}_i - \mathbf{X}_j\|_2 - \frac{1}{n_a^2} \sum_{i=1}^n \sum_{j=1}^n w_i w_j \mathbf{I}(A_i = A_j = a) \|\mathbf{X}_i - \mathbf{X}_j\|_2 \\ &\quad - \frac{1}{n^2} \sum_{i=1}^n \sum_{j=1}^n \|\mathbf{X}_i - \mathbf{X}_j\|_2. \end{aligned}$$

where  $n_a = \sum_{i=1}^n \mathbf{I}(A_i = a)$ . The weighted energy distance measures the distance between the weighted distribution of covariates among those with  $A_i = a$  and the empirical distribution of covariates in the full sample; it takes value 0 if and only if these two distributions are identical.

The optimal weights are obtained by minimizing the sum of all the weighted energy distances of all treatment groups:

$$\mathbf{w}_n \in \underset{\mathbf{w}=(w_1, \dots, w_n)}{\operatorname{argmin}} \sum_{a \in \mathcal{A}} \mathcal{E}(F_{n,a,w}, F_n) \quad \text{s.t.} \quad \sum_{i=1}^n w_i \mathbf{I}(A_i = a) = n_a \quad (27)$$

for all  $a \in \mathcal{A}$  and  $w_i \geq 0$  for  $i = 1, \dots, n$ .

Thus, the EBWs encourage the distributions of covariates for each treatment group to look like that of the target population, i.e. the full sample. Since (27) is a quadratic objective function along with linear constraints, it can be efficiently solved through interior point methods. Another advantage lies in the fact that it does not require parameter tuning. The estimated weights are effective in balancing

all covariates of the covariate distributions, not just limited to the first or second moments. Lastly, the data-driven characteristic of EBWs that do not rely on specific model assumptions contributes to more robust ITR estimates.

## C.2 Implementation of the PGD algorithm

We provide Algorithm 1 to solve the problem (6) using a *projected gradient descent* (PGD). The algorithm applies to either continuous or binary outcomes. For the projection onto the  $L_1$ -ball, we use the algorithm in [14].

---

### Algorithm 1 Projected Gradient Descent Algorithm to Estimate Decision Function for ITR-Learning

---

**Input:** Data set  $(\mathbf{x}_i, a_i, y_i)_{i=1}^n \in \mathbb{R}^{n \times p} \times \{1, 2, \dots, K\}^n \times \mathbb{R}^n$  ( $\triangleright y_i$  can be continuous or binary); loss function  $L$

**Parameters:**  $\mathbf{B} \in \mathbb{R}^{p \times (K-1)}$  (model coefficient),  $z \geq 0$  ( $L_1$ -ball size);  $\lambda_1 \geq 0$  (additional  $L_2$ -penalization for  $R(\mathbf{B})$ );  $T \in \mathbb{N}$  (number of iterations),  $\alpha_t \geq 0$  (stepsize).

**Constraints:**  $L_1$ -ball  $\mathcal{B} := \{\mathbf{B} \in \mathbb{R}^{p \times (K-1)} \mid \|\mathbf{B}\|_1 \leq z\}$

Compute  $\mathbf{x}_i^* := \mathbf{x}_i \mathbf{u}_{a_i}^T$  for all  $i = 1, \dots, n$  for faster computation.

Choose any point from  $\mathbf{B}_0 \in \mathcal{B}$  as initialization.

**For**  $t = 1, 2, \dots, T$  **do:**

$$\mathbf{B}_t \leftarrow \Pi_{\mathcal{B}} \left( \mathbf{B}_{t-1} - \alpha_t \nabla_{\mathbf{B}} \mathcal{L}(\mathbf{B}_{t-1}) \right).$$

**End for**

**Output:**  $\mathbf{B}_T$

---

To help the convenient implementation of the above PGD algorithm, we provide the gradient of the ITR objective below. For the continuous outcome, we have (see Proposition F.1)

$$\nabla_{\mathbf{B}} \mathcal{L}(\mathbf{B}) = \frac{1}{n} \sum_{i=1}^n \omega(a_i, \mathbf{x}_i) \left( \mathbf{x}_i \mathbf{x}_i^T \mathbf{B} \mathbf{u}_{a_i} \mathbf{u}_{a_i}^T - \frac{K}{K-1} y_i \mathbf{x}_i \mathbf{u}_{a_i}^T \right) + \lambda_2 \mathbf{B}. \quad (28)$$

For the binary outcome, we have (see Proposition F.2)

$$\nabla_{\mathbf{B}} \mathcal{L}(\mathbf{B}) = \frac{1}{n} \left( \sum_{i=1}^n \omega(a_i, \mathbf{x}_i) \left( \frac{\exp(\mathbf{u}_{a_i}^T \mathbf{B}^T \mathbf{x}_i)}{1 + \exp(\mathbf{u}_{a_i}^T \mathbf{B}^T \mathbf{x}_i)} - y_i \right) \mathbf{x}_i \mathbf{u}_{a_i}^T \right) + \lambda_2 \mathbf{B}. \quad (29)$$

## C.3 Additional improvements by dimension reduction: variable screening procedure

High dimensional covariates can significantly affect the performance of methods for estimating weights, outcome regression models, and the final ITR. Specifically, using screened variables is beneficial in estimating outcome regression models by focusing on fewer variables that significantly impact the outcome. Additionally, estimating balancing weights based on effect modifiers rather than all variables provides more reliable weighting estimates. Hence, the screening approach serves to reduce bias and enhance the efficiency of ITR-Learning.

Several techniques have been developed to choose appropriate variables in causal inference. One popular approach in the ITR setting is the group-lasso method introduced in [33] which filters out variables that only affect the treatment-free effect. For the ATE estimation, *sure independence screening* (SIS) method [16, 32] is widely used, which involves an initial screening step based on Pearson correlations between each covariate and the outcome. The covariates are then ranked based on these correlations, and only the top-ranked ones are used in the downstream analysis. While one could also use similar variable screening methods in the ITR setting, the use of Pearson correlation may limit their effectiveness in capturing nonlinear relationships, potentially leading to biased estimates of the ITR.

To address this issue, alternative approaches that employ non-linear dependence tests are explored, including the ball covariance, mutual information test, and distance covariance [37, 42, 48]. Ball covariance, previously applied to variable selection in ATE estimation, has been found to be effective [49]. Distance covariance is another way to capture the nonlinear dependence between two random

vectors/variables [48]. The test compares the observed distances between pairs of observations to the expected distances under the assumption of independence. A zero value of distance covariance indicates independence between two random vectors. If the observed distances are significantly far from the expected distances, it indicates statistical dependence between the random vectors. The  $p$ -value is calculated by comparing the observed empirical distance with the distribution of the same test statistic obtained from random permutations of the data under the assumption of independence. We propose adapting the distance covariance due to its computational efficiency and empirical performance for the settings we explored.

Now, we introduce our proposed variable screening procedure. In a setting with  $K$  treatments, we aim to find which covariates have conditional dependence on the outcome given the treatment. Doing so allows the identification of covariates that either modify the treatment effect or otherwise affect the outcome (i.e. are precision variables). To do so, for each covariate  $X_j$  where  $j \in \{1, \dots, p\}$ , conduct the distance covariance test between covariate and outcome  $K$  times, conditioning on each treatment value. This conditioning on the treatment creates subgroups based on treatment assignments, allowing the assessment of the dependence between the covariate and the outcome within each subgroup. Then we set the  $p$ -value of the covariate  $X_j$  as the minimum  $p$ -value from these  $K$  tests. This approach ensures that we retain as many relevant variables as possible during this preliminary variable screening stage. For the same reason, we avoid the use of multiple comparison adjustments such as Bonferroni correction. Instead, we opt to select variables with a significance level of 5%. We perform the same steps for all covariates in the dataset. These procedures are summarized in Algorithm 2.

---

**Algorithm 2** Variable Screening using Distance Covariance

---

**Input:**  $\mathbf{X} \in \mathbf{R}^{n \times p}$ ;  $Y \in \mathbf{R}^n$ ;  $A \in \{1, \dots, K\}$ .  
Initialize  $\tilde{\mathbf{X}} = \{\}$ .  
**for**  $j = 1, 2, \dots, p$  **do**  
    **for**  $k = 1, 2, \dots, K$  **do**  
        Perform distance covariance test( $X_j, Y|A = k$ ) and obtain  $p$ -value,  $p_{j,k}$ .  
    **end for**  
     $p_j \leftarrow \min(p_{j,1}, p_{j,2}, \dots, p_{j,K})$ .  
    **if**  $p_j < 0.05$  **then**  
         $\tilde{\mathbf{X}} \leftarrow \tilde{\mathbf{X}} \cup \{X_j\}$ .  
    **end if**  
**end for**  
**Output:**  $\tilde{\mathbf{X}}$ , the set of screened variables containing effect modifiers and precision variables.

---

This conditional independence test is crucial to retain effect modifiers and precision variables while filtering out instrumental variables associated with treatment but not the outcome [49]. While acknowledging that the procedure is ad-hoc, the simulation results show that it effectively reduces the number of covariates to estimate covariate balancing weights and outcome regression models, which improves efficiency and helps avoid overfitting.

**C.4 Additional improvements by variance reduction: combining inverse variance weighting with augmented outcome**

Our additional recommendation includes dually using both augmented outcomes and inverse variance weighting for variance reduction. Under our working model (1), the outcome can be explained by the sum of the treatment-free effects of covariates, the interaction effect between covariates and each treatment, and random noise. As the interaction effect alone determines optimal treatment assignments, the treatment-free effect acts as extra noise. This noise can be mitigated using outcome model augmentation. One way to implement outcome augmentation is to use the augmented outcomes  $\tilde{Y} := Y - \hat{\mu}(\mathbf{X})$  in place of the original outcomes for continuous outcomes when applying an ITR estimation method. Flexible techniques such as random forest and neural networks can be utilized to estimate the treatment-free effect. This is the first implementation for multi-category treatments with DCBWs, whereas previous work focused on binary treatments with IPW [44].

Augmented outcomes can refine ITR estimates when correctly specified. Yet, misspecification of either outcome or treatment-free effect models can lead to heteroscedastic errors due to residual treatment-free effects. This is where SABD-Learning’s inverse variance weighting helps, providing



robustness against misspecification. SABD-Learning extends AD-Learning designed for homoscedasticity [39] to account for heteroscedasticity, i.e. when the conditional variance of the (potential) outcome given the covariates is not constant [44]. This method estimates the conditional variance of outcomes and reweights the final objective function by adjusting each sample's contribution based on the inverse of the estimated variance. Therefore, we propose to combine outcome augmentation and inverse variance weighting for robust and efficient ITR learning, as they are complementary methods.

The key idea behind outcome augmentation is to add an asymptotically unbiased quantity to the objective function (negative log-likelihood function) so that the new optimizer for the 'augmented objective function' has a provably smaller variance than the original estimator. To do this, let  $h$  be some unknown (to be determined) augmentation function depending on the covariate. The key property of asymptotic unbiasedness is

$$\frac{1}{n} \sum_{i=1}^n w(a_i, \mathbf{x}_i) \mathbf{u}_{a_i}^T \mathbf{B}^T h(\mathbf{x}_i) \rightarrow 0 \quad \text{as } n \rightarrow \infty \text{ in probability for all } \mathbf{B} \in \mathbb{R}^{p \times (K-1)}. \quad (30)$$

We will justify this property in Corollary C.3 toward the end of this section. Given (30), we have

$$\ell(\mathbf{B}) \approx \ell(\mathbf{B}) + \frac{1}{n} \sum_{i=1}^n w_i \mathbf{u}_{a_i}^T \mathbf{B}^T h(\mathbf{x}_i), \quad (31)$$

where  $\approx$  here means asymptotic equality in probability as  $n \rightarrow \infty$ . Denote the 'augmented objective function' on the right-hand side as  $\ell_{\text{aug}}(\mathbf{B})$ .

In the remainder of this section, we introduce the optimal augmented function  $h_{\text{opt}}(\mathbf{X})$  that minimizes the variance of the estimator, while maintaining asymptotic unbiasedness. We demonstrate that outcome augmentation remains valid when IPW is replaced with DCBWs, such as EBWs and MMD balancing weights. Our approach extends the work of [50, 9] to a more general setting, incorporating other types of covariate balancing weights that have asymptotic consistency (38) in a multi-category treatment setting.

Let  $S(\mathbf{B})$  be the derivative of the original objective function  $\ell(\mathbf{B})$  with respect to  $\mathbf{B}$ . Then setting the optimal  $h(\mathbf{X})$  is equivalent to minimizing the conditional variance of

$$\widehat{\mathbf{B}}_a - \mathbf{B}^* \approx \underbrace{S(\mathbf{B}) - w(A, \mathbf{X}) h(\mathbf{X}) \mathbf{u}_A^T}_{=: \mathbf{Z}} | \mathbf{X} = \mathbf{x}$$

where  $\widehat{\mathbf{B}}_a$  is the minimizer of  $\ell_{\text{aug}}(\mathbf{B})$ , and  $\mathbf{B}^*$  is the minimizer of  $\mathbb{E}[\ell(\mathbf{B})]$ . Using the property of asymptotic unbiasedness Theorem C.2, which is  $\mathbb{E}[\mathbf{Z}] = 0$ , we need to minimize  $\mathbb{E}[\|\mathbf{Z}\|_F^2]$  with respect to  $h$  to minimize the conditional variance of  $\mathbf{Z}$ . Using the property of matrix trace,

$$\nabla_h \mathbb{E}[\|\mathbf{Z}\|_F^2] = \nabla_h \mathbb{E}[\text{tr}(\mathbf{Z}^T \mathbf{Z})] = -2 \mathbb{E}[w(A, \mathbf{X}) \mathbf{Z} \mathbf{u}_A | \mathbf{X} = \mathbf{x}],$$

and setting  $\nabla_h \mathbb{E}[\|\mathbf{Z}\|_F^2] = 0$ . Thus,  $h_{\text{opt}}(\mathbf{X})$  satisfies

$$\mathbb{E}[w(A, \mathbf{X})(S(\mathbf{B}) - w(A, \mathbf{X}) h_{\text{opt}}(\mathbf{X}) \mathbf{u}_A^T) \mathbf{u}_A | \mathbf{X} = \mathbf{x}] = 0.$$

Since  $\mathbf{u}_A^T \mathbf{u}_A = 1$ , we have

$$h_{\text{opt}}(\mathbf{x}) = \frac{\mathbb{E}[w(A, \mathbf{X}) S(\mathbf{B}^*) \mathbf{u}_A | \mathbf{X} = \mathbf{x}]}{\mathbb{E}[w(A, \mathbf{X})^2 | \mathbf{X} = \mathbf{x}]}. \quad (32)$$

Now, we determine the functional form of  $h_{\text{opt}}(\mathbf{X})$  for both continuous and binary outcomes. For simplicity, let us define

$$w := w(A, \mathbf{X}), \quad t_i := \mathbf{u}_{a_i}^T \mathbf{B}^T \mathbf{x}_i \\ R_1(\mathbf{x}) := \mathbb{E}[w^2 \mathbf{u}_A | \mathbf{X} = \mathbf{x}], \quad R_2(\mathbf{x}) := \mathbb{E}[w^2 Y | \mathbf{X} = \mathbf{x}], \quad c_1 = E[w^2 | \mathbf{X} = \mathbf{x}].$$

Then for continuous outcomes, using (28) without regularization term for  $S(\mathbf{B})$  and (32), we have

$$\begin{aligned}
& \ell_{aug}(\mathbf{B}) \\
&= \frac{1}{n} \sum_{i=1}^n \left[ w_i \left( \frac{K}{K-1} y_i - \mathbf{u}_{a_i}^T \mathbf{B}^T \mathbf{x}_i \right)^2 - w_i \mathbf{u}_{a_i}^T \mathbf{B}^T h_{opt}(\mathbf{x}_i) \right] \\
&= \frac{1}{n} \sum_{i=1}^n \left[ w_i \left( \frac{K}{K-1} y_i - \mathbf{u}_{a_i}^T \mathbf{B}^T \mathbf{x}_i \right)^2 - w_i \mathbf{u}_{a_i}^T \mathbf{B}^T \frac{1}{\hat{c}_1} \left( \mathbf{x}_i \mathbf{x}_i^T \mathbf{B} \hat{R}_1(\mathbf{x}_i) - \frac{2K}{K-1} \mathbf{x}_i \hat{R}_2(\mathbf{x}_i) \right) \right] \\
&= \frac{1}{n} \sum_{i=1}^n w_i \left[ \left( \frac{K}{K-1} y_i - t_i \right)^2 - t_i \left( \mathbf{x}_i^T \mathbf{B} \frac{1}{\hat{c}_1} \hat{R}_1(\mathbf{x}_i) \right) + t_i \left( \frac{2K}{K-1} \frac{1}{\hat{c}_1} \hat{R}_2(\mathbf{x}_i) \right) \right] \\
&= \frac{1}{n} \sum_{i=1}^n w_i \left[ \left( t_i - \frac{K}{K-1} \left( y_i - \frac{1}{\hat{c}_1} \hat{R}_2(\mathbf{x}_i) \right) \right)^2 + \text{constant w.r.t } \mathbf{B} \right],
\end{aligned}$$

assuming  $\hat{R}_1(\mathbf{x}_i)/\hat{c}_1 = 0$ . Therefore, we obtain the optimal augmented function with the minimum estimation variance by properly defining  $R_2(\mathbf{X})$ . Our simulation studies demonstrate that setting  $R_2(\mathbf{X}) = \mathbb{E}[Y|\mathbf{X} = \mathbf{x}]$  leads to significant improvement.

Similarly, for binary outcomes, let us define

$$Q_1(\mathbf{x}) := \mathbb{E} \left[ w^2 \frac{\exp(\mathbf{u}_A^T \mathbf{B}^T \mathbf{X})}{1 + \exp(\mathbf{u}_A^T \mathbf{B}^T \mathbf{X})} \middle| \mathbf{X} = \mathbf{x} \right], \quad Q_2(\mathbf{x}) := \mathbb{E}[w^2 Y | \mathbf{X} = \mathbf{x}].$$

Then using (29) without regularization term for  $S(\mathbf{B})$  and (32), we have

$$\begin{aligned}
\ell_{aug}(\mathbf{B}) &= \frac{1}{n} \sum_{i=1}^n w_i \left[ -y_i \mathbf{u}_{a_i}^T \mathbf{B}^T \mathbf{x}_i + \log(1 + \exp(\mathbf{u}_{a_i}^T \mathbf{B}^T \mathbf{x}_i)) + \mathbf{u}_{a_i}^T \mathbf{B}^T h_{opt}(\mathbf{x}_i) \right] \\
&= \frac{1}{n} \sum_{i=1}^n w_i \left[ -y_i t_i + \log(1 + \exp(t_i)) + \frac{1}{\hat{c}_1} t_i (\hat{Q}_1(\mathbf{x}) - \hat{Q}_2(\mathbf{x})) \right] \\
&= \frac{1}{n} \sum_{i=1}^n w_i \left[ -t_i \left( y_i - \frac{1}{\hat{c}_1} (\hat{Q}_1(\mathbf{x}) - \hat{Q}_2(\mathbf{x})) \right) + \log(1 + \exp(t_i)) \right].
\end{aligned}$$

Therefore, by selecting proper  $Q_1(\mathbf{X})$  and  $Q_2(\mathbf{X})$ , the minimizer of  $\ell_{aug}(\mathbf{B})$  has smaller variance than that of the original estimator.

Now we prove asymptotic unbiasedness of  $\frac{1}{n} \sum_{i=1}^n w(a_i, \mathbf{x}_i) \mathbf{u}_{a_i}^T \mathbf{B}^T h(\mathbf{x}_i)$ . First, we consider IPW-based conditions for multi-category treatments.

**Lemma C.1.** *Let  $\mathcal{A} = \{1, \dots, K\}$  be the possible treatment sets. Assume we have i.i.d. observations  $(\mathbf{x}_i, a_i)$  from a population distribution of  $(\mathbf{X}, A)$  such that  $\mathbb{P}(A = a | \mathbf{x})$  is uniformly positive for all  $a, \mathbf{x}$ . Define*

$$T := \frac{1}{n} \sum_{i=1}^n \frac{1}{\mathbb{P}(A = a_i | \mathbf{x}_i)} h(\mathbf{x}_i) g(a_i). \quad (33)$$

*If  $g$  is any function with  $\sum_{j \in \mathcal{A}} g(j) = 0$ , then for any function  $h(\mathbf{X}) : \mathbb{R}^p \rightarrow \mathbb{R}^p$ , then  $T \rightarrow 0$  as  $n \rightarrow \infty$  in probability.*

*Proof.* First note that by iterated expectation,

$$\mathbb{E}[T] = \mathbb{E} \left[ \frac{h(\mathbf{X}) g(A)}{\mathbb{P}(A | \mathbf{X})} \right] = \mathbb{E}_{\mathbf{X}} \left[ \mathbb{E}_{\mathcal{A}} \left[ \frac{h(\mathbf{X}) g(A)}{\mathbb{P}(A | \mathbf{X})} \middle| \mathbf{X} \right] \right] \quad (34)$$

$$= \mathbb{E}_{\mathbf{X}} \left[ \sum_{j \in \mathcal{A}} \frac{h(\mathbf{X}) g(j)}{\mathbb{P}(A = j | \mathbf{X})} \mathbb{P}(A = j | \mathbf{X}) \right] \quad (35)$$

$$= \mathbb{E}_{\mathbf{X}} \left[ h(\mathbf{X}) \sum_{j \in \mathcal{A}} g(j) \right] \quad (36)$$

$$= 0. \quad (37)$$

Then by the standard weak law of large numbers for i.i.d. samples, the empirical mean  $T$  converges in probability to its population mean  $\mathbb{E}[T] = 0$  as  $n \rightarrow \infty$ .  $\square$

Note that in the statement of the Lemma C.1, the choice of such  $g(j)$  can be each treatment vertex under the AD-Learning [39]. Now we replace IPW  $\mathbb{P}(A = a_i | \mathbf{x}_i)$  with covariate balancing weights  $w(a_i, \mathbf{x}_i)$  that has the following asymptotic consistency property:

$$\lim_{n \rightarrow \infty} \frac{1}{n} \sum_{i=1}^n \left( w(a_i, \mathbf{x}_i) - \frac{1}{\mathbb{P}(a_i | \mathbf{x}_i)} \right)^2 = 0. \quad (38)$$

For instance, *maximum mean discrepancy* (MMD) balancing weights, a type of distributional covariate energy balancing weight, satisfy the property (38) [8].

**Theorem C.2** (Asymptotic unbiasedness of augmented function). *Assume we have i.i.d. observations  $(\mathbf{x}_i, a_i)$  from population distribution of  $(\mathbf{X}, A)$ . If  $g : \mathcal{A} \rightarrow \mathbb{R}$  is any bounded function with  $\sum_{j \in \mathcal{A}} g(j) = 0$ , then for any bounded function  $h(\mathbf{x}) : \mathbb{R}^p \rightarrow \mathbb{R}^p$ , define*

$$T_0 := \frac{1}{n} \sum_{i=1}^n w(a_i, \mathbf{x}_i) h(\mathbf{x}_i) g(a_i), \quad (39)$$

where the weights  $w(a_i, \mathbf{x}_i)$  satisfy (38). Then  $T_0 \rightarrow 0$  as  $n \rightarrow \infty$  in probability. That is,  $T_0$  is an asymptotically mean zero, making it suitable to be used as the augmented function.

*Proof.* By the weak law of large numbers for i.i.d. samples, the empirical mean  $T_0$  converges in probability to its population mean  $\mathbb{E}[T_0] = \mathbb{E}[w(A, \mathbf{X})h(\mathbf{X})g(A)]$ . By continuous mapping theorem,  $\|T_0\|^2$  converges to  $\|\mathbb{E}[T_0]\|^2$  as  $n \rightarrow \infty$  in probability. Then we have

$$\begin{aligned} \|\mathbb{E}[T_0]\|^2 &= \|\mathbb{E}[w(A, \mathbf{X})h(\mathbf{X})g(A)]\|^2 \stackrel{(a)}{=} \left\| \mathbb{E}[w(A, \mathbf{X})h(\mathbf{X})g(A)] - \mathbb{E} \left[ \frac{h(\mathbf{X})g(A)}{\mathbb{P}(A|\mathbf{X})} \right] \right\|^2 \\ &= \left\| \mathbb{E} \left[ \left( w(A, \mathbf{X}) - \frac{1}{\mathbb{P}(A|\mathbf{X})} \right) h(\mathbf{X})g(A) \right] \right\|^2 \\ &\stackrel{(b)}{\leq} \mathbb{E} \left[ \left\| \left( w(A, \mathbf{X}) - \frac{1}{\mathbb{P}(A|\mathbf{X})} \right) h(\mathbf{X})g(A) \right\|^2 \right] \\ &\leq C \mathbb{E} \left[ \left( w(A, \mathbf{X}) - \frac{1}{\mathbb{P}(A|\mathbf{X})} \right)^2 \right] \\ &\stackrel{(c)}{=} \lim_{n \rightarrow \infty} C \frac{1}{n} \sum_{i=1}^n \left( w(a_i, \mathbf{x}_i) - \frac{1}{\mathbb{P}(a_i | \mathbf{x}_i)} \right)^2 \rightarrow 0, \end{aligned}$$

where  $C$  is any constant such that  $\|h(\mathbf{X})g(A)\|^2 \leq C$ . For (a), we use Lemma C.1; for (b) and (c), we use Jensen's inequality and the law of large numbers, respectively. The last term converges to 0 due to the property (38). Lastly, since  $\|T_0\|^2$  converges to 0 in probability, we have  $T_0$  converges to 0 in probability by continuous mapping theorem.  $\square$

The following corollary is useful in deriving outcome augmentation under the AD-Learning framework.

**Corollary C.3.** *Assume the AD-learning setting. For any function  $h(\mathbf{X})$  depending only on the covariate  $\mathbf{X}$ , and for each parameter matrix  $\mathbf{B} \in \mathbb{R}^{p \times (K-1)}$ , denoting,*

$$H_0(\mathbf{B}) := w(A, \mathbf{X}) \mathbf{u}_A^T \mathbf{B}^T h(\mathbf{X}),$$

we have  $\mathbb{E}[H_0(\mathbf{B})] = 0$ . Furthermore, assuming the weights  $w(A, \mathbf{X})$  satisfying the consistency property (38), in probability as  $n \rightarrow \infty$ ,

$$\frac{1}{n} \sum_{i=1}^n w(a_i, \mathbf{x}_i) \mathbf{u}_{a_i}^T \mathbf{B}^T h(\mathbf{x}_i) \rightarrow 0. \quad (40)$$

*Proof.* Follows immediately from Theorem C.2 by noting that  $\sum_{a=1}^K \mathbf{u}_a = 0$ .  $\square$

## D Spectral characterization of optimal covariate balancing weight

Recall the definition of the weighted design matrix:

$$\Psi := \frac{1}{n} \sum_{i=1}^n w_i (\mathbf{u}_{a_i} \mathbf{u}_{a_i}^T) \otimes (\mathbf{x}_i \mathbf{x}_i^T). \quad (41)$$

This is the  $p(K-1) \times p(K-1)$  design matrix encoding the sample of patients with covariate  $\mathbf{x}_i$  and observed treatment  $a_i$ , for  $i = 1, \dots, n$ . We are free to choose the weight  $w_i = \omega(a_i, \mathbf{x}_i)$  for the  $i$ th patient as a function as long as they sum to one. What is the best way to choose such weights? Our main results (Theorems 3.3, 3.5) suggest that it is best to choose them in a way that the minimum eigenvalue of  $\Psi$  is as large as possible.

To illustrate this point, suppose for simplicity that we have binary treatments ( $K = 2$  and  $u_1 = 1, u_2 = -1$ ) and two discrete covariates (e.g., ‘Male’ and ‘Female’) encoded as one-hot vectors  $(1, 0)^T$  and  $(0, 1)^T$ . Then we can write  $\Psi$  as

$$n\Psi = w(1, M)n_{1,M} \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix} + w(2, M)n_{2,M} \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix} \quad (42)$$

$$+ w(1, F)n_{1,F} \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix} + w(2, F)n_{2,F} \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix} \quad (43)$$

$$= \begin{bmatrix} w(1, M)n_{1,M} + w(2, M)n_{2,M} & 0 \\ 0 & w(1, F)n_{1,F} + w(2, F)n_{2,F} \end{bmatrix}, \quad (44)$$

where for treatment  $a = 1, 2$  and  $g \in \{M, F\}$ ,  $n(a, g)$  denotes the number of patients of gender  $g$  with treatment  $a$ , and  $w(a, g)$  denotes the weight for patients of gender  $g$  with treatment  $a$ . Notice that the minimum eigenvalue of the last  $2 \times 2$  matrix is the minimum of its two diagonal entries. It is easy to see that the optimal choice of the weights is given by

$$w(1, M) = \frac{n_{1,M} + n_{2,M}}{n_{1,M}}, \quad w(2, M) = \frac{n_{1,M} + n_{2,M}}{n_{2,M}}, \quad (45)$$

$$w(1, F) = \frac{n_{1,F} + n_{2,F}}{n_{1,F}}, \quad w(2, F) = \frac{n_{1,F} + n_{2,F}}{n_{2,F}}. \quad (46)$$

These weights are exactly the ones that make the conditional distributions of the covariates given treatments the same. This observation holds for more general settings.

## E Convergence analysis of PGD for ITR estimation

In this section, we analyze the convergence properties of Algorithm 1 for ITR-Learning problem (6). Below we state the full version of Theorem 3.3 in the main text.

**Theorem E.1** (Convergence rate of PGD for ITR-Learning). *Let  $(\mathbf{B}_t)_{t \geq 0}$  denote the sequence of parameters obtained by the PGD algorithm (7) for ITR-Learning problem (6) with arbitrary initialization  $\mathbf{B}_0$ . Suppose Assumptions 3.1 and 3.2 hold. Let  $\mathbf{B}^*$  denote the unique global optimum of (6). Then the following hold:*

(i) (Optimization landscape) *The ITR objective  $\mathcal{L}(\mathbf{B})$  in (6), is  $\mu$ -strongly convex and  $L$ -smooth, where  $\mu, L$  are as in (10).*

(ii) (Linear convergence with fixed stepsize) *Assume constant stepsize  $\alpha_t \equiv \alpha < 2/L$ . Denote the contraction constant  $\rho(\alpha) := \max\{|1 - \alpha L|, |1 - \alpha \mu|\} \in (0, 1)$ . Then for all  $t \geq 1$ ,*

$$\|\mathbf{B}_t - \mathbf{B}^*\|_F^2 \leq 4\rho(\alpha)^t \lambda_1^2. \quad (47)$$

*In particular, if  $\alpha = \frac{2}{\mu+L}$ , then  $\rho(\alpha) = \frac{L-\mu}{L+\mu} \in [0, 1)$ .*

(iii) (Sublinear convergence with diminishing stepsize) *Assume diminishing stepsize  $\alpha_t = \frac{\beta}{\gamma+t}$ , where  $\rho := \frac{2\mu L}{\mu+L}$  and  $\beta > 1/\rho, \gamma > 0$  are constants. Then for all  $t \geq 1$ ,*

$$\|\mathbf{B}_t - \mathbf{B}^*\|_F^2 \leq \frac{\nu}{\gamma+t}, \quad (48)$$

*where  $\nu := \max\left\{\frac{2\beta^2 \|\nabla f(\mathbf{B}^*)\|_F^2}{\beta\rho-1}, 4(\gamma+1)\lambda_1^2\right\}$ . In particular, the assertion holds for*

$$\beta = 2/\rho = \frac{\mu+L}{\mu L} \text{ and } \nu = \max\left\{\frac{2(\mu+L)^2 \|\nabla f(\mathbf{B}^*)\|_F^2}{\mu^2 L^2}, 4(\gamma+1)\lambda_1^2\right\}.$$

The proof of Theorem E.1 relies on two ingredients: a general result on the convergence of PGD (Lemma E.2) and a local landscape analysis of the ITR objective (Lemma E.3).

We first establish a general result on the convergence rate of PGD to find the minimizer of a strongly convex and smooth objective  $f$  within a convex constraint  $\mathcal{B}$ . Our argument is mostly based on the standard convergence analysis of PGD in the optimization literature [5, 4]. A new element we add is that for analyzing constrained optimization problems, the gradient of the objective  $f$  at the minimizer  $\mathbf{B}^*$  does not need to be zero, which only needs to be in the negative normal cone at the constraint set  $\mathcal{B}$ . In our analysis, we treat the possibly nonzero gradient norm  $\|\nabla f(\mathbf{B}^*)\|_F$  as ‘noise’ (variance) of a stochastic gradient oracle in standard stochastic gradient descent analysis [4].

**Lemma E.2** (Convergence rate of PGD for strongly convex and smooth objectives). *Let  $f : \mathbb{R}^{p \times (K-1)} \rightarrow \mathbb{R}$  be a  $L$ -smooth and  $\mu$ -strongly convex function for some  $\mu, L > 0$ . Suppose  $f^* := \inf_{\mathbf{B} \in \mathcal{B}} f(\mathbf{B}) > -\infty$  and denote  $\mathbf{B}^* := \arg \min_{\mathbf{B} \in \mathcal{B}} f(\mathbf{B})$ , where  $\mathcal{B}$  is a convex constraint set of  $\mathbb{R}^{p \times (K-1)}$ . Consider the following PGD iterates:*

$$\mathbf{B}_{t+1} \leftarrow \Pi_{\mathcal{B}} \{\mathbf{B}_t - \alpha_t \nabla_{\mathbf{B}} f(\mathbf{B}_t)\}. \quad (49)$$

*Then the following hold:*

(i) (Linear convergence with fixed stepsize) *Suppose  $f$  is twice differentiable and  $\alpha_t \equiv \alpha < 2/L$ . Let contraction constant  $\rho(\alpha) := \max\{|1 - \alpha L|, |1 - \alpha \mu|\} \in (0, 1)$ . Then for all  $t \geq 1$ ,*

$$\|\mathbf{B}_t - \mathbf{B}^*\|_F^2 \leq \rho(\alpha)^t \|\mathbf{B}_0 - \mathbf{B}^*\|_F^2. \quad (50)$$

*The bound  $\rho(\alpha)$  on the linear convergence rate is minimized at  $\alpha = \frac{2}{\mu+L}$  with minimum value  $\frac{L-\mu}{L+\mu}$ .*

(ii) (Sublinear convergence with diminishing stepsize) *Denote  $\rho := \frac{2\mu L}{\mu+L}$ . Suppose  $\alpha_t = \frac{\beta}{\gamma+t}$ , where  $\beta > 1/\rho, \gamma > 0$  are constants. Then for all  $t \geq 1$ ,*

$$\|\mathbf{B}_t - \mathbf{B}^*\|_F^2 \leq \frac{\nu}{\gamma+t}, \quad (51)$$

*where  $\nu := \max\left\{\frac{2\beta^2 \|\nabla f(\mathbf{B}^*)\|_F^2}{\beta\rho-1}, (\gamma+1)\|\mathbf{B}_1 - \mathbf{B}^*\|_F^2\right\}$ . In particular, the assertion holds*

$$\text{for } \beta = 2/\rho = \frac{\mu+L}{\mu L} \text{ and } \nu = \max\left\{\frac{2(\mu+L)^2 \|\nabla f(\mathbf{B}^*)\|_F^2}{\mu^2 L^2}, (\gamma+1)\|\mathbf{B}_1 - \mathbf{B}^*\|_F^2\right\}.$$

*Proof.* Let  $\bar{\mathbf{B}}_t := \mathbf{B}_t - \alpha_t \nabla_{\mathbf{B}} f(\mathbf{B}_t)$ . Then

$$\begin{aligned} \|\bar{\mathbf{B}}_t - \mathbf{B}^*\|_F^2 &= \|\bar{\mathbf{B}}_t - \mathbf{B}_t\|_F^2 + \|\mathbf{B}_t - \mathbf{B}^*\|_F^2 + 2\langle \bar{\mathbf{B}}_t - \mathbf{B}_t, \mathbf{B}_t - \mathbf{B}^* \rangle \\ &= \alpha_t^2 \|\nabla_{\mathbf{B}} f(\mathbf{B}_t)\|_F^2 + \|\mathbf{B}_t - \mathbf{B}^*\|_F^2 - 2\alpha_t \langle \mathbf{B}_t - \mathbf{B}^*, \nabla_{\mathbf{B}} f(\mathbf{B}_t) \rangle \\ &\leq \alpha_t^2 \|\nabla_{\mathbf{B}} f(\mathbf{B}_t)\|_F^2 + \|\mathbf{B}_t - \mathbf{B}^*\|_F^2 - 2\alpha_t \langle \mathbf{B}_t - \mathbf{B}^*, \nabla_{\mathbf{B}} f(\mathbf{B}_t) - \nabla_{\mathbf{B}} f(\mathbf{B}^*) \rangle \end{aligned}$$

using that  $\langle \nabla f(\mathbf{B}^*), \mathbf{B}_t - \mathbf{B}^* \rangle \geq 0$  since  $\mathbf{B}^*$  is a stationary point of  $f$  over  $\mathcal{B}$ . Using the co-coercivity of  $\mu$ -strongly convex and  $L$ -smooth functions and, we have

$$\begin{aligned} \|\bar{\mathbf{B}}_t - \mathbf{B}^*\|_F^2 &\leq \alpha_t^2 \|\nabla f(\mathbf{B}_t)\|_F^2 + \|\mathbf{B}_t - \mathbf{B}^*\|_F^2 \\ &\quad - 2\alpha_t \left( \frac{\mu L}{\mu + L} \|\mathbf{B}_t - \mathbf{B}^*\|^2 + \frac{1}{\mu + L} \|\nabla f(\mathbf{B}_t) - \nabla f(\mathbf{B}^*)\|_F^2 \right) \\ &\leq \left( 1 - \frac{2\alpha_t \mu L}{\mu + L} \right) \|\mathbf{B}_t - \mathbf{B}^*\|_F^2 + \alpha_t^2 \|\nabla f(\mathbf{B}_t)\|_F^2 - \frac{2\alpha_t}{\mu + L} \|\nabla f(\mathbf{B}_t) - \nabla f(\mathbf{B}^*)\|_F^2. \end{aligned}$$

By Young's inequality,

$$\|\nabla f(\mathbf{B}_t)\|_F^2 \leq 2\|\nabla f(\mathbf{B}_t) - \nabla f(\mathbf{B}^*)\|_F^2 + 2\|\nabla f(\mathbf{B}^*)\|_F^2.$$

Thus the above yields, given that  $\alpha_t \leq 1/(\mu + L)$ ,

$$\begin{aligned} \|\bar{\mathbf{B}}_t - \mathbf{B}^*\|_F^2 &\leq \left( 1 - \frac{2\alpha_t \mu L}{\mu + L} \right) \|\mathbf{B}_t - \mathbf{B}^*\|_F^2 + 2\alpha_t \left( \alpha_t - \frac{1}{\mu + L} \right) \|\nabla f(\mathbf{B}_t) - \nabla f(\mathbf{B}^*)\|_F^2 \\ &\quad + 2\alpha_t^2 \|\nabla f(\mathbf{B}^*)\|_F^2 \\ &\leq \left( 1 - \frac{2\alpha_t \mu L}{\mu + L} \right) \|\mathbf{B}_t - \mathbf{B}^*\|_F^2 + 2\alpha_t^2 \|\nabla f(\mathbf{B}^*)\|_F^2. \end{aligned}$$

Since the projection onto the convex constraint set  $\mathcal{B}$ ,  $\Pi_{\mathcal{B}}$ , is non-expansive,

$$\|\mathbf{B}_{t+1} - \mathbf{B}^*\|_F^2 \leq \|\bar{\mathbf{B}}_t - \mathbf{B}^*\|_F^2 \leq (1 - \alpha_t \rho) \|\mathbf{B}_t - \mathbf{B}^*\|_F^2 + \alpha_t^2 \sigma^2,$$

where we have denoted  $\rho := \frac{2\mu L}{\mu + L}$  and  $\sigma^2 := 2\|\nabla f(\mathbf{B}^*)\|_F^2$ .

Now, we show **(i)** (50). Suppose  $\alpha_t \equiv \alpha > 0$ . Note that  $\mathbf{B}^*$  is a unique minimizer of  $f$  over  $\mathcal{B}$  because of strong convexity. Since  $-\nabla f(\mathbf{B}^*)$  lies in the normal cone of  $\mathcal{B}$  at  $\mathbf{B}^*$ , we have  $\mathbf{B}^* = \Pi_{\mathcal{B}}(\mathbf{B}^* - \alpha \nabla f(\mathbf{B}^*))$ . Using non-expansiveness of convex projection  $\Pi_{\mathcal{B}}$ ,

$$\begin{aligned} \|\mathbf{B}_{t+1} - \mathbf{B}^*\|_F &= \|\Pi_{\mathcal{B}}(\mathbf{B}_t - \alpha \nabla f(\mathbf{B}_t)) - \Pi_{\mathcal{B}}(\mathbf{B}^* - \alpha \nabla f(\mathbf{B}^*))\|_F \\ &\leq \|(\mathbf{B}_t - \alpha \nabla f(\mathbf{B}_t)) - (\mathbf{B}^* - \alpha \nabla f(\mathbf{B}^*))\|_F \\ &= \left\| \int_0^1 (I - \alpha \nabla^2 f(\mathbf{B}_t + s(\mathbf{B}^* - \mathbf{B}_t))) (\mathbf{B}_t - \mathbf{B}^*) ds \right\|_F \\ &\leq \sup_{\tilde{\mathbf{B}}} \|I - \alpha \nabla^2 f(\tilde{\mathbf{B}})\|_2 \|\mathbf{B}_t - \mathbf{B}^*\|_F \\ &\leq \underbrace{\max\{|1 - \alpha L|, |1 - \alpha \mu|\}}_{=: \rho(\alpha)} \|\mathbf{B}_t - \mathbf{B}^*\|_F. \end{aligned}$$

For the last inequality, since the eigenvalues of  $\nabla^2 f(\tilde{\mathbf{B}})$  are contained in the interval  $[\mu, L]$ , the eigenvalues of  $I - \alpha \nabla^2 f(z)$  are between  $\min(1 - \alpha L, 1 - \alpha \mu)$  and  $\max(1 - \alpha L, 1 - \alpha \mu)$ . Note that  $\rho(\alpha)$  is minimized at  $\alpha = \frac{2}{\mu + L}$  with minimum value  $\frac{L - \mu}{L + \mu}$ . It follows that for any  $t \geq 0$ ,

$$\|\mathbf{B}_{t+1} - \mathbf{B}^*\|_F \leq \rho(\alpha)^{t+1} \|\mathbf{B}_0 - \mathbf{B}^*\|_F.$$

Next, we show **(ii)** (51) by induction. Indeed, it holds for  $t = 1$  by the choice of  $\nu$ . Denoting  $\hat{t} := \gamma + t$ , by the induction hypothesis and by the choices of  $\alpha_t = \beta/\hat{t}$  and  $\nu \geq \frac{\beta^2 \sigma^2}{\beta \rho - 1}$ ,

$$\begin{aligned} \|\mathbf{B}_{t+1} - \mathbf{B}^*\|_F^2 &\leq \left( 1 - \frac{\beta \rho}{\hat{t}} \right) \frac{\nu}{\hat{t}} + \frac{\beta^2 \sigma^2}{\hat{t}^2} \\ &\leq \left( \frac{\hat{t} - 1}{\hat{t}^2} \right) \nu - \underbrace{\left( \frac{\beta \rho - 1}{\hat{t}^2} \right) \nu + \frac{\beta^2 \sigma^2}{\hat{t}^2}}_{\leq 0} \\ &\leq \frac{\hat{t} - 1}{(\hat{t} - 1)(\hat{t} + 1)} \nu = \frac{\nu}{\gamma + t + 1}. \end{aligned}$$



This shows **(ii)**.  $\square$

Next, we establish the following global landscape result of the ITR learning problem (6) by computing the Hessian of the objective  $\mathcal{L}(\mathbf{B})$  in (6).

**Lemma E.3** (Local landscape of the ITR objective). *Let  $\mathcal{L}(\mathbf{B})$  denote the ITR objective in (6).*

**(i)** *Suppose Assumption 3.1 holds. For the continuous outcome,*

$$(\lambda^- + \lambda_2)\mathbf{I}_{p(K-1)} \preceq \nabla_{\text{vec}(\mathbf{B})} \nabla_{\text{vec}(\mathbf{B})}^T \mathcal{L}(\mathbf{B}) \preceq (\lambda^+ + \lambda_2)\mathbf{I}_{p(K-1)}, \quad (52)$$

where the constants  $\lambda^\pm$  are defined in Assumption 3.1.

**(ii)** *Further Assumptions 3.1 and 3.2 hold. For the binary outcome,*

$$(\alpha^- \lambda^- + \lambda_2)\mathbf{I}_{p(K-1)} \preceq \nabla_{\text{vec}(\mathbf{B})} \nabla_{\text{vec}(\mathbf{B})}^T \mathcal{L}(\mathbf{B}) \preceq (\alpha^+ \lambda^+ + \lambda_2)\mathbf{I}_{p(K-1)}. \quad (53)$$

where the constants  $\lambda^\pm$  and  $\alpha^\pm$  are defined in Assumptions 3.1 and 3.2, respectively.

To maintain the flow, we write down the proof of Lemma E.3 to Section F.

Now we can easily deduce Theorem E.1.

**Proof of Theorem E.1.** Lemma E.3 shows that the eigenvalues of the Hessian  $\nabla_{\text{vec}(\mathbf{B})} \nabla_{\text{vec}(\mathbf{B})}^T \mathcal{L}$  of the ITR objective in (6) are uniformly bounded between constants  $\mu$  and  $L$  as defined in (10). Under Assumptions 3.1 and 3.2, we have  $0 < \mu \leq L < \infty$ . This shows that  $\mathcal{L}$  is  $\mu$ -strongly convex and  $L$ -smooth over  $\mathcal{B}$ , as claimed in Theorem E.1 (i).

Next, noting that  $\|\cdot\|_F \leq \|\cdot\|_1$ , triangle inequality and the fact that  $\hat{\mathbf{B}}_0, \mathbf{B}^* \in \mathcal{B}$  yields that, for any  $t \geq 0$ ,

$$\|\mathbf{B}_t - \mathbf{B}^*\|_F^2 \leq (\|\mathbf{B}_t\|_F + \|\mathbf{B}^*\|_F)^2 \quad (54)$$

$$\leq (\|\mathbf{B}_t\|_1 + \|\mathbf{B}^*\|_1)^2 \quad (55)$$

$$\leq 4\lambda_1^2. \quad (56)$$

Hence the assertions (i)-(ii) follow directly from combining Lemmas E.2 and E.3.  $\square$

## F Proof of Lemma E.3

In this section, we prove Lemma E.3. We will first compute the gradient and the Hessian of the ITR objective in (6) separately for the continuous and binary outcome. We will then prove Lemma E.3 at the end of the section.

### F.1 Continuous outcome

**Proposition F.1.** *Let  $\mathcal{L}(\mathbf{B})$  denote the ITR objective in (6) with continuous outcome. Then*

$$\nabla_{\mathbf{B}} \mathcal{L}(\mathbf{B}) = \frac{1}{n} \sum_{i=1}^n \omega(a_i, \mathbf{x}_i) \left( \mathbf{x}_i \mathbf{x}_i^T \mathbf{B} \mathbf{u}_{a_i} \mathbf{u}_{a_i}^T - \frac{K}{K-1} y_i \mathbf{x}_i \mathbf{u}_{a_i}^T \right) + \lambda_2 \mathbf{B}, \quad (57)$$

$$\nabla_{\text{vec}(\mathbf{B})} \nabla_{\text{vec}(\mathbf{B})}^T \mathcal{L}(\mathbf{B}) = \Psi + \lambda_2 \mathbf{I}_{p(K-1)}, \quad (58)$$

where  $\Psi$  is the weighted design matrix defined at (8).

*Proof.* Recall that the ITR objective for continuous outcome is given by

$$\mathcal{L}(\mathbf{B}) = \frac{1}{2n} \sum_{i=1}^n \omega(a_i, \mathbf{x}_i) \left( \frac{K}{K-1} y_i - \mathbf{u}_{a_i}^T \mathbf{B}^T \mathbf{x}_i \right)^2 + \frac{\lambda_2}{2} \|\mathbf{B}\|_F^2.$$

It is trivial to show (57). For the Hessian, using (16),

$$\begin{aligned}
\nabla_{\mathbf{B}}^2 \mathcal{L}(\mathbf{B}) &= \nabla_{\text{vec}(\mathbf{B})} \nabla_{\text{vec}(\mathbf{B})}^T \mathcal{L}(\text{vec}(\mathbf{B})) \\
&= \nabla_{\text{vec}(\mathbf{B})} \text{vec} \left( \frac{1}{n} \sum_{i=1}^n \omega(a_i, \mathbf{x}_i) \left( \mathbf{x}_i \mathbf{x}_i^T \mathbf{B} \mathbf{u}_{a_i} \mathbf{u}_{a_i}^T - \frac{K}{K-1} y_i \mathbf{x}_i \mathbf{u}_{a_i}^T \right) \right)^T + \lambda_2 \nabla_{\mathbf{B}}^2(\mathbf{B}) \\
&= \nabla_{\text{vec}(\mathbf{B})} \text{vec} \left( \frac{1}{n} \sum_{i=1}^n \omega(a_i, \mathbf{x}_i) \left( \mathbf{x}_i \mathbf{x}_i^T \mathbf{B} \mathbf{u}_{a_i} \mathbf{u}_{a_i}^T \right) \right)^T + \lambda_2 \mathbf{I}_{p(K-1)} \\
&= \nabla_{\text{vec}(\mathbf{B})} \left[ \frac{1}{n} \sum_{i=1}^n \omega(a_i, \mathbf{x}_i) \left( \left( (\mathbf{u}_{a_i} \mathbf{u}_{a_i}^T)^T \otimes (\mathbf{x}_i \mathbf{x}_i^T) \right) \text{vec}(\mathbf{B}) \right)^T \right] + \lambda_2 \mathbf{I}_{p(K-1)} \\
&= \nabla_{\text{vec}(\mathbf{B})} \left[ \frac{1}{n} \sum_{i=1}^n \omega(a_i, \mathbf{x}_i) \text{vec}(\mathbf{B})^T \left( (\mathbf{u}_{a_i} \mathbf{u}_{a_i}^T)^T \otimes (\mathbf{x}_i \mathbf{x}_i^T) \right)^T \right] + \lambda_2 \mathbf{I}_{p(K-1)} \\
&= \frac{1}{n} \sum_{i=1}^n \omega(a_i, \mathbf{x}_i) \left( (\mathbf{u}_{a_i} \mathbf{u}_{a_i}^T)^T \otimes (\mathbf{x}_i \mathbf{x}_i^T) \right)^T + \lambda_2 \mathbf{I}_{p(K-1)} \\
&= \frac{1}{n} \sum_{i=1}^n \omega(a_i, \mathbf{x}_i) (\mathbf{u}_{a_i} \mathbf{u}_{a_i}^T) \otimes (\mathbf{x}_i \mathbf{x}_i^T) + \lambda_2 \mathbf{I}_{p(K-1)} \\
&= \Psi + \lambda_2 \mathbf{I}_{p(K-1)}
\end{aligned}$$

This shows the assertion.  $\square$

## F.2 Binary outcome

Given the current model parameter  $\mathbf{B} \in \mathbb{R}^{p \times (K-1)}$ , define the scalars  $z_i, p_i, w_i$  for the  $i$ -th sample by

$$z_i := \mathbf{u}_{a_i}^T \mathbf{B}^T \mathbf{x}_i, \quad p_i := \frac{\exp(z_i)}{1 + \exp(z_i)}, \quad w_i := w(a_i, \mathbf{x}_i). \quad (59)$$

For each binary label  $y \in \{0, 1\}$  and activation  $z \in \mathbb{R}$ , recall the negative log-likelihood of observing  $y$  under the logistic model  $\text{Ber}(p_i)$  is given by

$$\ell(y, z) = \log(1 + \exp(z)) - yz.$$

An easy computation shows

$$\dot{\mathbf{h}}(y, z) := \frac{\partial}{\partial z} \ell(y, z) = \frac{\exp(z)}{1 + \exp(z)} - y, \quad (60)$$

$$\ddot{\mathbf{h}}(y, z) := \frac{\partial^2}{\partial z^2} \ell(y, z) = \left( \frac{\exp(z)}{1 + \exp(z)} \right) \left( 1 - \frac{\exp(z)}{1 + \exp(z)} \right) \leq \frac{1}{4}. \quad (61)$$

For the forthcoming computations, define matrices

$$\mathbf{K} := [\dot{\mathbf{h}}(y_1, z_1), \dots, \dot{\mathbf{h}}(y_n, z_n)] \in \mathbb{R}^{1 \times n}, \quad (62)$$

$$\mathbf{M} := \text{diag} \left( \ddot{\mathbf{h}}(y_1, z_1), \dots, \ddot{\mathbf{h}}(y_n, z_n) \right) \in \mathbb{R}^{n \times n}, \quad (63)$$

$$\Phi := [\mathbf{u}_{a_1} \otimes \mathbf{x}_1, \dots, \mathbf{u}_{a_n} \otimes \mathbf{x}_n] \in \mathbb{R}^{p(K-1) \times n}, \quad (64)$$

$$\mathbf{W} := \text{diag}(w_1, \dots, w_n) \in \mathbb{R}^{n \times n}. \quad (65)$$

**Proposition F.2.** *Let  $\mathcal{L}(\mathbf{B})$  denote the ITR objective in (6) with binary outcome. Suppose Assumption 3.1 holds. Let  $z_i := \mathbf{u}_{a_i}^T \mathbf{B}^T \mathbf{x}_i$  for  $i = 1, \dots, n$ . Then*

$$\nabla_{\mathbf{B}} \mathcal{L}(\mathbf{B}) = \frac{1}{n} \left( \sum_{i=1}^n w_i \dot{\mathbf{h}}(y_i, z_i) \mathbf{x}_i \mathbf{u}_{a_i}^T \right) + \lambda_2 \mathbf{B}, \quad (66)$$

$$\nabla_{\text{vec}(\mathbf{B})} \nabla_{\text{vec}(\mathbf{B})}^T \mathcal{L}(\mathbf{B}) = \frac{1}{n} \Phi (\mathbf{M} \mathbf{W}) \Phi^T + \lambda_2 \mathbf{I}_{p(K-1)} \quad (67)$$

*Proof.* We first claim that

$$\nabla_{\text{vec}(\mathbf{B})} z_i = \mathbf{C}^{(K-1,p)}(\mathbf{x}_i \otimes \mathbf{u}_{a_i}), \quad (68)$$

$$\nabla_{\text{vec}(\mathbf{B})} \ell(y_i, z_i) = \mathbf{C}^{(K-1,p)}(\mathbf{x}_i \otimes \mathbf{u}_{a_i}) \dot{\mathbf{h}}(y_i, z_i), \quad (69)$$

$$\nabla_{\text{vec}(\mathbf{B})} \text{vec}(\mathbf{K})^T = \nabla_{\text{vec}(\mathbf{B})} \mathbf{K} = \Phi \mathbf{M}. \quad (70)$$

Observe that by using (18), we can write

$$z_i = \text{vec}(z_i) = \text{vec}(\mathbf{u}_{a_i}^T \mathbf{B}^T \mathbf{x}_i) = (\mathbf{x}_i^T \otimes \mathbf{u}_{a_i}^T) \text{vec}(\mathbf{B}^T) = \text{vec}(\mathbf{B}^T)^T (\mathbf{x}_i \otimes \mathbf{u}_{a_i}).$$

Noting that  $\text{vec}(\mathbf{B}^T)^T = (\mathbf{C}^{(p,K-1)} \text{vec}(\mathbf{B}))^T = \text{vec}(\mathbf{B})^T \mathbf{C}^{(K-1,p)}$ , Hence we get

$$\begin{aligned} \nabla_{\text{vec}(\mathbf{B})} z_i &= \nabla_{\text{vec}(\mathbf{B})} \text{vec}(\mathbf{B}^T)^T (\mathbf{x}_i \otimes \mathbf{u}_{a_i}) \\ &= \mathbf{C}^{(K-1,p)}(\mathbf{x}_i \otimes \mathbf{u}_{a_i}). \end{aligned}$$

This verifies (68). Then by using the chain rule (20), we get

$$\begin{aligned} \nabla_{\text{vec}(\mathbf{B})} \ell(y_i, z_i) &= \nabla_{\text{vec}(\mathbf{B})} z_i \nabla_{z_i} \ell(y_i, z_i) \\ &= \mathbf{C}^{(K-1,p)}(\mathbf{x}_i \otimes \mathbf{u}_{a_i}) \dot{\mathbf{h}}(y_i, z_i), \end{aligned}$$

which verifies (69).

Next, we compute the gradients of  $\mathbf{K}$  in (66). First, using (68) and the chain rule (20),

$$\begin{aligned} \nabla_{\text{vec}(\mathbf{B})} \dot{\mathbf{h}}(y_i, z_i) &= \nabla_{\text{vec}(\mathbf{B})} z_i \nabla_{z_i} \dot{\mathbf{h}}(y_i, z_i) \\ &= \mathbf{C}^{(K-1,p)}(\mathbf{x}_i \otimes \mathbf{u}_{a_i}) \ddot{\mathbf{h}}(y_i, z_i) \\ &= (\mathbf{u}_{a_i} \otimes \mathbf{x}_i) \ddot{\mathbf{h}}(y_i, z_i), \end{aligned}$$

where the last equality uses commutation matrix relation (17). Then noting that  $\text{vec}(\mathbf{K})^T = \mathbf{K} = [\dot{\mathbf{h}}(y_1, z_1), \dots, \dot{\mathbf{h}}(y_n, z_n)]$ , it follows that

$$\begin{aligned} \nabla_{\text{vec}(\mathbf{B})} \text{vec}(\mathbf{K})^T &\stackrel{(a)}{=} [(\mathbf{u}_{a_1} \otimes \mathbf{x}_1) \ddot{\mathbf{h}}(y_1, z_1), \dots, (\mathbf{u}_{a_n} \otimes \mathbf{x}_n) \ddot{\mathbf{h}}(y_n, z_n)] \\ &\stackrel{(b)}{=} [(\mathbf{u}_{a_1} \otimes \mathbf{x}_1), \dots, (\mathbf{u}_{a_n} \otimes \mathbf{x}_n)] \text{diag}(\ddot{\mathbf{h}}(y_1, z_1), \dots, \ddot{\mathbf{h}}(y_n, z_n)) \\ &\stackrel{(c)}{=} \Phi \mathbf{M}, \end{aligned}$$

where (a) follows from (68) and the chain rule, (b) is an algebra, and (c) follows from the definition. This shows (70).

Now we compute the gradient:

$$\nabla_{\text{vec}(\mathbf{B})} \left[ \frac{1}{n} \sum_{i=1}^n w_i \ell(y_i, z_i) \right] \stackrel{(a)}{=} \frac{1}{n} \sum_{i=1}^n w_i \left[ \mathbf{C}^{(K-1,p)}(\mathbf{x}_i \otimes \mathbf{u}_{a_i}) \dot{\mathbf{h}}(y_i, z_i) \right] \quad (71)$$

$$\stackrel{(b)}{=} \frac{1}{n} \sum_{i=1}^n w_i (\mathbf{u}_{a_i} \otimes \mathbf{x}_i) \dot{\mathbf{h}}(y_i, z_i) \quad (72)$$

$$= \frac{1}{n} [w_1(\mathbf{u}_{a_1} \otimes \mathbf{x}_1), \dots, w_n(\mathbf{u}_{a_n} \otimes \mathbf{x}_n)] \begin{bmatrix} \dot{\mathbf{h}}(y_1, z_1) \\ \vdots \\ \dot{\mathbf{h}}(y_n, z_n) \end{bmatrix} \quad (73)$$

$$\stackrel{(c)}{=} \frac{1}{n} \Phi \mathbf{W} \text{vec}(\mathbf{K}), \quad (74)$$

where (a) follows from (69), (b) uses the commutation matrix relation (17), and (c) follows from the definition. Then by using (72) and (18) with the fact that  $\dot{\mathbf{h}}$  is a scalar, we deduce (66).

Next, we compute the Hessian. From (74),

$$\nabla_{\text{vec}(\mathbf{B})} \mathcal{L}(\mathbf{B}) = \frac{1}{n} \Phi \mathbf{W} \text{vec}(\mathbf{K}) + \lambda_2 \text{vec}(\mathbf{B}). \quad (75)$$

Then using (69) with (75), we get

$$\nabla_{\text{vec}(\mathbf{B})} \nabla_{\text{vec}(\mathbf{B})^T} \mathcal{L}(\mathbf{B}) = \nabla_{\text{vec}(\mathbf{B})} \left[ \frac{1}{n} \text{vec}(\mathbf{K})^T \mathbf{W} \Phi^T + \lambda_2 \mathbf{I}_{p(K-1)} \right] \quad (76)$$

$$= \frac{1}{n} \Phi (\mathbf{M} \mathbf{W}) \Phi^T + \lambda_2 \mathbf{I}_{p(K-1)}. \quad (77)$$

This shows the assertion.  $\square$

### E.3 Proof of Lemma E.3

We are now ready to give proof of Lemma E.3.

*Proof of Lemma E.3.* Recall that  $\Phi := [\mathbf{u}_{a_1} \otimes \mathbf{x}_1, \dots, \mathbf{u}_{a_n} \otimes \mathbf{x}_n]$ . Note that

$$\frac{1}{n} \Phi \mathbf{W} \Phi^T = \frac{1}{n} \sum_{i=1}^n w_i (\mathbf{u}_{a_i} \otimes \mathbf{x}_i) (\mathbf{u}_{a_i} \otimes \mathbf{x}_i)^T \quad (78)$$

$$= \frac{1}{n} \sum_{i=1}^n w_i (\mathbf{u}_{a_i} \mathbf{u}_{a_i}^T \otimes \mathbf{x}_i \mathbf{x}_i^T) \quad (79)$$

$$= \Psi, \quad (80)$$

where  $\Psi$  is the weighted design matrix defined at (8). (In particular, since  $\mathbf{W}$  is a diagonal matrix of nonnegative entries, this shows that  $\Psi$  is positive semi-definite.)

According to Propositions F.1 and F.2, we have

$$\nabla_{\text{vec}(\mathbf{B})} \nabla_{\text{vec}(\mathbf{B})^T} \mathcal{L}(\mathbf{B}) = \begin{cases} n^{-1} \Phi \mathbf{W} \Phi^T + \lambda_2 \mathbf{I}_{p(K-1)} & \text{for continuous outcome} \\ n^{-1} \Phi \mathbf{M} \mathbf{W} \Phi^T + \lambda_2 \mathbf{I}_{p(K-1)} & \text{for binary outcome.} \end{cases} \quad (81)$$

Thus the assertion for the continuous outcome follows immediately. For the case of a binary outcome, by Assumption 3.2 we have

$$\alpha^- \mathbf{I}_n \preceq \mathbf{M} \preceq \alpha^+ \mathbf{I}_n. \quad (82)$$

Then

$$\alpha^- \left( \frac{1}{n} \Phi \mathbf{W} \Phi^T \right) \preceq \frac{1}{n} \Phi (\mathbf{M} \mathbf{W}) \Phi^T \preceq \alpha^+ \left( \frac{1}{n} \Phi \mathbf{W} \Phi^T \right). \quad (83)$$

It follows that

$$\alpha^- \lambda_{\min} \left( \frac{1}{n} \Phi \mathbf{W} \Phi^T \right) \mathbf{I}_{p(K-1)} \preceq \frac{1}{n} \Phi (\mathbf{M} \mathbf{W}) \Phi^T \preceq \alpha^+ \lambda_{\max} \left( \frac{1}{n} \Phi \mathbf{W} \Phi^T \right) \mathbf{I}_{p(K-1)} \quad (84)$$

This shows the assertion for the binary outcome holds.  $\square$

## G A non-asymptotic consistency of weighted, constrained, and regularized MLE

In this section, we provide a general result on the non-asymptotic consistency of MLE in a general setting. Here, observations are assumed to be i.i.d. from a generative model, and the unknown true parameter of the generative model may lie on the boundary of the parameter space.

A generic sample  $(X, Y)$  consists of a random covariate  $X$  and outcome  $Y$ . (In the ITR setting, we will specialize  $X$  to be the pair of the patient's covariate and the corresponding treatment.) We assume the covariate  $X$  follows a probability distribution  $\mu(\cdot)$  and the conditional distribution of  $Y$  given  $X = x$  is parameterized as  $\pi_{\theta_*}(\cdot | x)$  for an unknown parameter  $\theta_*$ . Hence the generic random sample  $(X, Y)$  follows the parameterized distribution  $\pi_{\theta_*}(y | x)\mu(x)$ . Assume we know a smooth parametric family of conditional distributions  $\theta \mapsto \pi_{\theta}(\cdot | x)$ , where the true parameter  $\theta_*$  lies in a convex parameter space  $\Theta \subseteq \mathbb{R}^d$ . Our goal is to estimate  $\theta_*$  from i.i.d. samples  $(x_i, y_i)$ ,  $i = 1, \dots, n$  drawn from  $\pi_{\theta_*}(\cdot | \cdot)\mu(\cdot)$ .

To estimate  $\theta_*$ , a natural approach is to maximize the empirical joint likelihood function. Since the marginal distribution of the covariate  $x$  is independent of the parameter  $\theta$ , we can apply the method of "reweighting" the covariate distribution by introducing an additional random weight  $\omega(x)$  that we can control. Namely, fix a weighting function  $\omega(\cdot) : \mathcal{X} \rightarrow [0, \infty)$ , where  $\mathcal{X}$  is covariate space, that assigns a nonnegative weight to each covariate value  $x$ . Consider then the following weighted expected log-likelihood maximization problem:

$$\arg \min_{\theta \in \Theta} \mathbb{E}_{(X, Y)} [-\omega(X) \log \pi_{\theta}(Y | X)], \quad (85)$$

where we omit the marginal distribution  $\mu(X)$  as it is independent of  $\theta$ .

Recall that we get to design the weighting function  $\omega(\cdot)$ . We first observe that the true parameter  $\theta_*$  is a stationary point of the weighted expected log-likelihood function  $\mathbb{E}[-\omega(X) \log \pi_{\theta}(Y | X)]$  for arbitrary weighting function  $\omega$ . This can be justified under the mild assumption of exchangeability of the expectation and derivatives (which holds under twice continuous differentiability of the log-likelihood function by the dominated convergence theorem). Indeed, note that

$$\begin{aligned} \mathbb{E}_{(X, Y)} [\omega(X) \nabla_{\theta} \log \pi_{\theta}(Y | X)] &= \mathbb{E}_{(X, Y)} \left[ \omega(X) \frac{\nabla_{\theta} \pi_{\theta}(Y | X)}{\pi_{\theta}(Y | X)} \right] \\ &= \mathbb{E}_X \left[ \mathbb{E}_Y \left[ \omega(X) \frac{\nabla_{\theta} \pi_{\theta}(Y | X)}{\pi_{\theta}(Y | X)} \middle| X \right] \right] \\ &= \mathbb{E}_X \left[ \omega(X) \mathbb{E}_{Y \sim \pi_{\theta_*}(\cdot | X)} \left[ \frac{\nabla_{\theta} \pi_{\theta}(Y | X)}{\pi_{\theta}(Y | X)} \middle| X \right] \right] = 0, \end{aligned}$$

where the last equality follows from

$$\begin{aligned} \mathbb{E}_{Y \sim \pi_{\theta_*}(\cdot | X)} \left[ \frac{\nabla_{\theta} \pi_{\theta_*}(Y | X)}{\pi_{\theta_*}(Y | X)} \middle| X \right] &= \int \nabla_{\theta} \pi_{\theta_*}(y | X) dy \\ &= \nabla_{\theta} \int \pi_{\theta_*}(y | X) dy \\ &= \nabla_{\theta} \mathbf{1} = 0. \end{aligned}$$

Therefore,  $\theta_*$  is a critical point of  $\mathbb{E}[\omega(X) \log \pi_{\theta}(Y | X)]$ . In particular, if we assume that the log-likelihood function  $\log \pi_{\theta}(y | x)\mu(x)$  is convex in  $\theta$ , then  $\theta_*$  is a global maximizer of  $\mathbb{E}[\omega(X) \log \pi_{\theta}(Y | X)]$ .

Next, we introduce the finite-sample approximation of (85). Denote  $w_i := \omega(x_i)$  for the weight of the  $i$ th sample  $(x_i, y_i)$ . Then the weighted negative log-likelihood of the observed samples under the model parameter  $\theta$  is

$$\mathcal{L}(\theta) := -\frac{1}{n} \sum_{i=1}^n w_i \underbrace{\log \pi_{\theta}(y_i | \mathbf{x}_i)}_{=: \mathcal{L}_0(\theta)} + R(\theta), \quad (86)$$

where  $R(\theta)$  is a suitable choice of regularizer for parameter  $\theta$ . The regularization term can be considered as prior knowledge about the model parameter  $\theta$ .

Let  $\hat{\theta}_n$  denote a (possibly non-unique) minimizer of the above function over  $\Theta$ . This is a minimizer of the loss function  $\mathcal{L}$  over the constraint set  $\Theta$ , which we call the *weighted, constrained, and regularized MLE* of  $\theta_*$ .

As in the standard MLE analysis, we assume that the negative log-likelihood function  $\theta \mapsto \mathcal{L}_0(\theta)$  is differentiable and strictly convex for all  $x \in \mathbb{R}^p$ . Statistical literature typically assumes that the true parameter lies in the interior of the constrained parameter space,  $\Theta$ , a convex subset of  $\mathbb{R}^d$ . In contrast, we allow the true parameter  $\theta_*$  to lie on the boundary of  $\Theta$ . Regardless of that, we have seen above that the gradient of the weighted expected log-likelihood function vanishes at the true parameter  $\theta_*$ .

In this setting, we aim to provide a high-probability guarantee that there exists a global minimizer of (86) that is close to the true parameter  $\theta_*$  for an arbitrary weighting function  $\omega(\cdot)$ , which we will optimize later for the best statistical guarantee. In Theorem G.1 below, we obtain such a result in the non-asymptotic, weighted, constrained, and regularized setting. For its proof, we combine a classical approach from [15] with concentration inequalities, namely, a classical Berry-Esseen bound for deviations from the standard normal distribution for independent but non-identically distributed random variables and a uniform McDirmid bound (Lemma G.2). The Berry-Esseen bound controls the linear term in the second-order Taylor expansion of the log-likelihood function ( $T_n(\theta)$  in (99)), while the McDirmid bound controls the quadratic term ( $S_n(\theta)$  in (99)). By using an  $\varepsilon$ -net argument, the latter concentration inequality can be extended to a setting where the random variables are parameterized within a compact set.

**Theorem G.1** (Non-asymptotic consistency of weighted, constrained, and regularized MLE). *Consider the constrained, regularized, and weighted MLE problem (86) with unknown parameter  $\theta_*$  from a convex subset  $\Theta \subseteq \mathbb{R}^d$ . Fix an arbitrary weighting function  $\omega(\cdot)$ . Assume the following holds:*

**(a1)** (Smoothness) *For each sample  $(x, y)$ , the per-sample negative log-likelihood function  $\theta \mapsto \log \pi_\theta(y | x)\mu(x)$  is strictly convex, three-times continuously differentiable, and  $R(\theta)$  is differentiable. Furthermore, denote  $U := \omega(X)\nabla_\theta \log \pi_{\theta_*}(Y | X) \in \mathbb{R}^d$  and  $\bar{U} := U - \mathbb{E}[U]$ . Suppose there are constants  $D_1, d_1 \in (0, \infty)$  such that for all  $i$ ,*

$$\mathbb{E}[\|\bar{U}\|^3] < D_1, \quad \min_{1 \leq k \leq d} \text{Var}(U(k)) > d_1. \quad (87)$$

**(a2)** (First-order optimality) *The true parameter  $\theta_*$  is a stationary point of the expected negative weighted log-likelihood function  $\bar{\mathcal{L}}_0(\theta) := \mathbb{E}_{\theta_*}[-\omega(X) \log \pi_\theta(Y | X)]$  over  $\Theta$ :*

$$\langle \nabla_\theta \bar{\mathcal{L}}_0(\theta_*), \theta - \theta_* \rangle \geq 0 \quad \forall \theta \in \Theta. \quad (88)$$

**(a3)** (Approximate second-order optimality) *Let  $\bar{\mathcal{L}}(\theta) := \bar{\mathcal{L}}_0(\theta) + R(\theta)$  denote the expected regularized negative log-likelihood function. Then the regularized Fisher information  $\nabla^2 \bar{\mathcal{L}}(\theta)$  is positive definite at  $\theta = \theta_*$  with minimum eigenvalue  $\rho > 0$ .*

Fix a constant  $C > 0$ . Let  $D = Cn^{-1/2} + \frac{8\|\nabla R(\theta_*)\|}{\rho}$  and  $M = M(D) > 0$  denote the supremum of the absolute values of all third-order partial derivatives of  $\mathcal{L}$  over all  $\theta$  with  $\|\theta - \theta_*\| \leq D$ . Suppose  $\|\nabla R(\theta_*)\|$  is small enough so that

$$D \leq \frac{3\rho}{4M(D)}. \quad (89)$$

Then there are constants  $c_1, c_2, c_3 > 0$  such that

$$\mathbb{P}\left(\left\|\theta_* - \arg \min_{\theta \in \Theta} \mathcal{L}(\theta)\right\| \leq D\right) \geq 1 - c_1 \exp\left(-\frac{C^2 \rho^2}{2 \cdot 32^2 d}\right) - \frac{c_2 \sum_{i=1}^n w_i^3}{\left(\sum_{i=1}^n w_i^2\right)^{3/2}} - O(\exp(-c_3 n)). \quad (90)$$

That is, with high probability explicitly depending on  $C, \rho, d$ , and  $n$ , there exists a global minimizer of  $\theta \mapsto \mathcal{L}(\theta)$  in  $\Theta$  within distance  $D$  from  $\theta_*$ .

It is important to notice that the minimum eigenvalue  $\rho$  of the weighted regularized Fisher information  $\nabla^2 \bar{\mathcal{L}}(\theta)$  depends on the choice of weighting function  $\omega(\cdot)$ . Therefore, Theorem G.1 suggests that

it is best to choose the weighting function to maximize the minimum eigenvalue of the weighted regularized Fisher information in order to minimize the statistical estimation error. We elaborate more on this point in the proof of Theorem 3.5.

We devote the rest of this section to proving Theorem G.1 using Lemma G.2 and Theorem G.3.

**Lemma G.2** (A uniform McDirmid's inequality). *Let  $\mathbf{X}_1, \dots, \mathbf{X}_n$  be independent random vectors in  $\mathbb{R}^p$  from a joint distribution  $\pi$ . Fix a compact parameter space  $\Theta \subseteq \mathbb{R}^d$  and  $f_\theta : \mathbb{R}^d \rightarrow [-J, J]$  is a bounded functional for each  $\theta \in \Theta$  such that*

$$\|f_\theta - f_{\theta'}\|_\infty \leq L\|\theta - \theta'\|, \quad \forall \theta, \theta' \in \Theta \quad (91)$$

for some constant  $L > 0$ . Further assume that  $\mathbb{E}[f_\theta(\mathbf{X}_k)] = 0$  for all  $\theta \in \Theta$  and  $k = 1, \dots, n$ . Then there exist constant  $J > 0$  such that for each  $n \geq 0$ , and  $\eta > 0$ ,

$$\mathbb{P}\left(\sup_{\theta \in \Theta} \left| \frac{1}{n} \sum_{k=1}^n f_\theta(\mathbf{X}_k) \right| \geq \eta\right) \leq \left(\frac{2L \text{diam}(\Theta)}{\eta}\right)^d \exp\left(-\frac{\eta^2 n}{2J^2}\right). \quad (92)$$

*Proof.* Since  $\Theta \subseteq \mathbb{R}^d$  is compact, it can be covered by a finite number of  $L_2$ -balls of any given radius  $\varepsilon > 0$ . Let  $\mathcal{U}_\varepsilon$  be such an open cover using the least number of balls of radius  $\varepsilon > 0$ . Let  $N(\varepsilon) = |\mathcal{U}_\varepsilon|$  denote the least number of such balls to cover  $\Theta$ . Moreover, let  $\text{diam}(\Theta)$  denote the diameter of  $\Theta$ , which is finite since  $\Theta$  is compact. Then  $\Theta$  is contained in a  $d$ -dimensional box of side length  $\text{diam}(\Theta)$ . This box can be covered by  $(\text{diam}(\Theta)/\varepsilon)^d$  cubes of side length  $\varepsilon$ . Covering each such cube of side length  $\varepsilon$  by a ball of radius  $\varepsilon$ , it follows that

$$N(\varepsilon) \leq \left(\frac{\text{diam}(\Theta)}{\varepsilon}\right)^d. \quad (93)$$

Next, fix  $\eta > 0$ ,  $\theta \in \Theta$ , and  $\varepsilon > 0$ . Let  $\theta_1, \dots, \theta_{N(\varepsilon)}$  be the centers of balls in the open cover  $\mathcal{U}_\varepsilon$ . Then there exists  $1 \leq j \leq N(\varepsilon)$  such that  $\|\theta - \theta_j\| < \varepsilon$ . By the hypothesis,  $f_\theta$  depends on  $\theta$  uniformly continuously with respect to the supremum norm. Hence there exists  $\delta = \delta(\varepsilon) > 0$  such that

$$\|f_\theta - f_{\theta_j}\|_\infty \leq L\varepsilon. \quad (94)$$

Denote  $H_n(\theta) := n^{-1} \sum_{k=1}^n f_\theta(\mathbf{X}_k)$ . Then this yields, almost surely,

$$|H_n(\theta) - H_n(\theta_j)| \leq L\varepsilon. \quad (95)$$

Furthermore, by the assumption,  $\|f_\theta\|_\infty$  is uniformly bounded by  $J > 0$ . It follows that for each  $\theta \in \Theta$ ,  $H_n(\theta)$  changes its value at most by  $M$  when one of  $\mathbf{X}_1, \dots, \mathbf{X}_n$  is replaced arbitrarily. Therefore, by the standard McDirmid's inequality (see Theorem 2.9.1. in [51]) and a union bound, with choosing  $\varepsilon = \eta/(2L)$ ,

$$\mathbb{P}(|H_n(\theta)| \geq \eta) \leq \sum_{j=1}^{N(\eta/2L)} \mathbb{P}(|H_n(\theta_j)| \geq \eta/2) \leq K \left(\frac{2L \text{diam}(\Theta)}{\eta}\right)^d \exp\left(-\frac{\eta^2 n}{2J^2}\right). \quad (96)$$

The above holds for all  $n \geq 1$  and  $\eta > 0$ . This shows the assertion.  $\square$

Next, we recall the classical Berry-Esseen theorem for the rate of convergence of normal approximation for the sum of independent random variables due to Feller.

**Theorem G.3** (Berry-Esseen, Theorem 2 in Ch. XVI.5 of Feller '91[17]). *Let  $X_1, X_2, \dots, X_n$  be independent and not necessarily identically distributed random variables with zero means and finite variances. Define  $W = \sum_{i=1}^n X_i$  let  $F$  be the distribution function of  $W$  and  $\Phi$  be the standard normal distribution function. Then*

$$\|F - \Phi\|_\infty \leq \frac{6 \sum_{i=1}^n \mathbb{E}[|X_i|^3]}{(\sum_{i=1}^n \mathbb{E}[X_i^2])^{3/2}}. \quad (97)$$

Now we prove Theorem G.1.



**Proof of Theorem G.1.** Suppose we have  $n$  i.i.d. observed samples  $(x_i, y_i)$  for  $i = 1, \dots, n$  from distribution  $\pi_{\theta_*}(y|x)\mu(x)$ . Also by the assumption,  $\mathcal{L}_0$  is twice continuously differentiable and  $n^{-1} \sum_{i=1}^n w_i = 1$ , so  $\mathbb{E}[\nabla \mathcal{L}_0] = \nabla \mathbb{E}[\mathcal{L}_0]$  and  $\mathbb{E}[\nabla^2 \mathcal{L}_0] = \nabla^2 \mathbb{E}[\mathcal{L}_0]$  by the dominated convergence theorem.

Note that, for any  $r > 0$ ,

$$\left\{ \left\| \theta_* - \arg \min_{\theta \in \Theta} \mathcal{L}(\theta) \right\| \leq r \right\} = \left\{ \inf_{\substack{\theta \in \Theta \\ \|\theta - \theta_*\| = r}} \mathcal{L}(\theta) - \mathcal{L}(\theta_*) > 0 \right\}. \quad (98)$$

The inclusion  $\subseteq$  is clear. Conversely, since  $\mathcal{L}$  is strictly convex, the event on the right implies that its unique global minimizer is within  $r$  from  $\theta_*$ , as desired. Thus in order to show the main result in (90), it suffices to show that the probability of the event on the right with  $r = D = Cn^{-1/2} + \frac{8\|\nabla R(\theta_*)\|}{\rho}$  is at least the right-hand side of (90).

Fix  $\theta \in \Theta$  such that  $\|\theta - \theta_*\| = D$ . We introduce two random variables that we will bound to be small by using some concentration inequalities:

$$T_n(\theta) := \frac{\sqrt{n}}{\|\theta - \theta_*\|} \langle \nabla_{\theta} \mathcal{L}_0(\theta_*) - \mathbb{E}[\omega(X) \nabla_{\theta} \mathcal{L}_0(\theta_*)], \theta - \theta_* \rangle, \quad (99)$$

$$S_n(\theta) := \frac{1}{\|\theta - \theta_*\|^2} (\theta - \theta_*)^T (\nabla_{\theta} \nabla_{\theta^T} \mathcal{L}(\theta_*) - \nabla_{\theta} \nabla_{\theta^T} (\mathbb{E}[\omega(X) \mathcal{L}(\theta_*)])) (\theta - \theta_*). \quad (100)$$

Roughly speaking, these quantities are of order 1 with high probability,  $T_n$  by Central Limit Theorem and  $S_n$  by the law of large numbers. Later in this proof, we will use non-asymptotic versions of these classical limit theorems to obtain high probability bounds for related quantities.

Since  $\theta \mapsto \mathcal{L}(\theta)$  is assumed to be three-time continuously differentiable, the quantity  $M \geq 0$  in the assertion is well-defined and is finite. Then, using the Taylor expansion, we may write

$$\begin{aligned} \mathcal{L}(\theta) - \mathcal{L}(\theta_*) &\geq \langle \nabla_{\theta} \mathcal{L}(\theta_*), \theta - \theta_* \rangle + \frac{1}{2} (\theta - \theta_*)^T \nabla_{\theta} \nabla_{\theta^T} \mathcal{L}(\theta_*) (\theta - \theta_*) \\ &\quad - \frac{M(\|\theta - \theta_*\|)}{6} \|\theta - \theta_*\|^3. \end{aligned} \quad (101)$$

We will lower bound the first two terms on the right-hand side above. Note that

$$\begin{aligned} \langle \nabla_{\theta} \mathcal{L}(\theta_*), \theta - \theta_* \rangle &= \langle \nabla_{\theta} \mathcal{L}_0(\theta_*), \theta - \theta_* \rangle - \mathbb{E}[\omega(X) \langle \nabla_{\theta} \mathcal{L}_0(\theta_*), \theta - \theta_* \rangle] \\ &\quad + \langle \nabla_{\theta} \mathbb{E}[\omega(X) \mathcal{L}_0(\theta_*)], \theta - \theta_* \rangle + \langle \nabla R(\theta_*), \theta - \theta_* \rangle \\ &\stackrel{(a)}{\geq} \langle \nabla_{\theta} \mathcal{L}_0(\theta_*), \theta - \theta_* \rangle - \mathbb{E}[\omega(X) \langle \nabla_{\theta} \mathcal{L}_0(\theta_*), \theta - \theta_* \rangle] \\ &\quad - \|\nabla R(\theta_*)\| \cdot \|\theta - \theta_*\| \\ &\stackrel{(b)}{=} - \frac{\|\theta - \theta_*\|}{\sqrt{n}} T_n(\theta) - \|\nabla R(\theta_*)\| \cdot \|\theta - \theta_*\|, \end{aligned}$$

where for (a) we use the fact that  $\theta_*$  is a stationary point of  $\mathbb{E}[\mathcal{L}_0(\theta)]$  over  $\Theta$  and Cauchy-Schwarz inequality; for (b) we use the definition of  $T_n(\theta)$ .

Next, we turn our attention to the second-order term in the Taylor expansion (101). Recall that from the assumption (a3) in Theorem G.1,

$$\mathbb{E}[\omega(X) \nabla_{\theta} \nabla_{\theta^T} \mathcal{L}(\theta_*)] = \nabla_{\theta} \nabla_{\theta^T} (\mathbb{E}[\omega(X) \mathcal{L}(\theta_*)]) \succeq \rho \mathbf{I}_p, \quad (102)$$

where  $\rho > 0$  is a constant. It follows that

$$(\theta - \theta_*)^T \nabla_{\theta} \nabla_{\theta^T} \mathcal{L}(\theta_*) (\theta - \theta_*) \quad (103)$$

$$\geq (\theta - \theta_*)^T [\nabla_{\theta} \nabla_{\theta^T} \mathcal{L}(\theta_*) - \nabla_{\theta} \nabla_{\theta^T} (\mathbb{E}[\omega(X) \mathcal{L}(\theta_*)])] (\theta - \theta_*) + \rho \|\theta - \theta_*\|^2 \quad (104)$$

$$\geq \|\theta - \theta_*\|^2 (S_n(\theta) + \rho). \quad (105)$$

Combining the above inequalities with noting that  $\|\boldsymbol{\theta} - \boldsymbol{\theta}_*\|^2 = D^2$ , we obtain

$$\begin{aligned}
& \frac{\mathcal{L}(\boldsymbol{\theta}) - \mathcal{L}(\boldsymbol{\theta}_*)}{\|\boldsymbol{\theta} - \boldsymbol{\theta}_*\|^2} \\
& \geq \frac{1}{\|\boldsymbol{\theta} - \boldsymbol{\theta}_*\|^2} \left[ \langle \nabla_{\boldsymbol{\theta}} \mathcal{L}(\boldsymbol{\theta}_*), \boldsymbol{\theta} - \boldsymbol{\theta}_* \rangle + \frac{1}{2} (\boldsymbol{\theta} - \boldsymbol{\theta}_*)^T \nabla_{\boldsymbol{\theta}} \nabla_{\boldsymbol{\theta}^T} \mathcal{L}(\boldsymbol{\theta}_*) (\boldsymbol{\theta} - \boldsymbol{\theta}_*) \right. \\
& \quad \left. - \frac{M(\|\boldsymbol{\theta} - \boldsymbol{\theta}_*\|)}{6} \|\boldsymbol{\theta} - \boldsymbol{\theta}_*\|^3 \right] \\
& \geq -\frac{1}{\|\boldsymbol{\theta} - \boldsymbol{\theta}_*\|} \left[ \frac{T_n(\boldsymbol{\theta})}{\sqrt{n}} + \|\nabla R(\boldsymbol{\theta}_*)\| \right] + \frac{1}{2} (S_n(\boldsymbol{\theta}) + \rho) - \frac{M(\|\boldsymbol{\theta} - \boldsymbol{\theta}_*\|)}{6} \|\boldsymbol{\theta} - \boldsymbol{\theta}_*\| \\
& = \frac{1}{\|\boldsymbol{\theta} - \boldsymbol{\theta}_*\|} \underbrace{\left( -\|\nabla R(\boldsymbol{\theta}_*)\| + \frac{\rho}{4} D - \frac{M(D)}{6} D^2 \right)}_{=: I_1} + \underbrace{\left( \frac{1}{2} (S_n(\boldsymbol{\theta}) + \frac{\rho}{2}) - \frac{1}{\sqrt{n}D} T_n(\boldsymbol{\theta}) \right)}_{=: I_2}.
\end{aligned}$$

Note that  $I_1 \geq 0$  if

$$\frac{\rho}{8} D \geq \|\nabla R(\boldsymbol{\theta}_*)\| \quad \text{and} \quad \frac{\rho}{8} D \geq \frac{M(D)}{6} D^2.$$

The choice of  $D$  holds the former condition, and the latter by the assumption (89). Thus,  $I_1 \geq 0$ .

We now take infimum over all  $\boldsymbol{\theta} \in \Theta$  such that  $\|\boldsymbol{\theta} - \boldsymbol{\theta}_*\| = D$ . It suffices to show that the random variable  $I_2$  defined above is positive with high probability, since  $I_1$  does not depend on  $\boldsymbol{\theta}$ . To this end, write

$$\inf_{\substack{\boldsymbol{\theta} \in \Theta \\ \|\boldsymbol{\theta} - \boldsymbol{\theta}_*\| = D}} I_2 \geq \underbrace{\left( \inf_{\substack{\boldsymbol{\theta} \in \Theta \\ \|\boldsymbol{\theta} - \boldsymbol{\theta}_*\| = D}} -\frac{1}{\sqrt{n}D} T_n(\boldsymbol{\theta}) \right)}_{=: A} + \underbrace{\left( \inf_{\substack{\boldsymbol{\theta} \in \Theta \\ \|\boldsymbol{\theta} - \boldsymbol{\theta}_*\| = D}} \frac{1}{2} (S_n(\boldsymbol{\theta}) + \frac{\rho}{2}) \right)}_{=: B}. \quad (106)$$

Then the last expression in (106) is at least  $\rho/16$  if  $A \geq -\rho/16$  and  $B \geq \rho/8$ . By the hypothesis,  $D = O(1)$  so it is uniformly bounded. Then by the uniform McDirmid's inequality in Lemma G.2, there exist constants  $C', C'' > 0$  such that

$$\mathbb{P} \left( B < \frac{\rho}{8} \right) = \mathbb{P} \left( \inf_{\substack{\boldsymbol{\theta} \in \Theta \\ \|\boldsymbol{\theta} - \boldsymbol{\theta}_*\| = D}} S_n(\boldsymbol{\theta}) < -\frac{\rho}{4} \right) \leq D^d C' \exp(-C'' n). \quad (107)$$

Next, we will show the following inequalities: For  $K = 6D_1/d_1^{3/2}$ ,

$$\begin{aligned}
\mathbb{P} \left( A < -\frac{\rho}{16} \right) &= \mathbb{P} \left( \inf_{\substack{\boldsymbol{\theta} \in \Theta \\ \|\boldsymbol{\theta} - \boldsymbol{\theta}_*\| = D}} T_n(\boldsymbol{\theta}) \geq \frac{\sqrt{n}D\rho}{16} \right) \\
&\stackrel{(c)}{\leq} d \left( \mathbb{P} \left( Z \geq \frac{\sqrt{n}D\rho}{32\sqrt{d}} \right) + \frac{K \sum_{i=1}^n w_i^3}{\sqrt{n} (\sum_{i=1}^n w_i^2)^{3/2}} \right) \\
&\stackrel{(d)}{\leq} d \left( \frac{32\sqrt{d}}{C\rho\sqrt{2\pi}} \exp \left( -\frac{nD^2\rho^2}{2 \cdot 32^2 d} \right) + \frac{K \sum_{i=1}^n w_i^3}{\sqrt{n} (\sum_{i=1}^n w_i^2)^{3/2}} \right) \\
&\stackrel{(e)}{\leq} d \left( \frac{32\sqrt{d}}{C\rho\sqrt{2\pi}} \exp \left( -\frac{C^2\rho^2}{2 \cdot 32^2 d} \right) + \frac{K \sum_{i=1}^n w_i^3}{\sqrt{n} (\sum_{i=1}^n w_i^2)^{3/2}} \right), \quad (108)
\end{aligned}$$

where  $Z \sim N(0, 1)$  is an independent standard normal random variable. (d) above is a simple consequence of the standard Gaussian tail bound  $\mathbb{P}(N(0, 1) > x) \leq \frac{e^{-x^2/2}}{x\sqrt{2\pi}}$  and (d), (e) are from the choice of  $D$  which yields  $nD^2 \geq C^2$ . We will show (c) at the end of this proof. Note that for any two events  $E_1, E_2$  defined on the same probability space,  $\mathbb{P}(E_1 \cap E_2) \geq \mathbb{P}(E_1) + \mathbb{P}(E_2) - 1$ . Then

by combining (106), (107), and (108), we have

$$\begin{aligned}
& \mathbb{P} \left( \inf_{\substack{\boldsymbol{\theta} \in \Theta \\ \|\boldsymbol{\theta} - \boldsymbol{\theta}_*\| = D}} \frac{\mathcal{L}(\boldsymbol{\theta}) - \mathcal{L}(\boldsymbol{\theta}_*)}{\|\boldsymbol{\theta} - \boldsymbol{\theta}_*\|^2} \geq \frac{\rho}{16} \right) \\
& \geq \mathbb{P} \left( A \geq -\frac{\rho}{16} \right) + \mathbb{P} \left( B \geq \frac{\rho}{8} \right) - 1 \\
& \geq 1 - d \left( \frac{32\sqrt{d}}{C\rho\sqrt{2\pi}} \exp \left( -\frac{C^2\rho^2}{2 \cdot 32^2 d} \right) + \frac{K \sum_{i=1}^n w_i^3}{(\sum_{i=1}^n w_i^2)^{3/2}} \right) + (1 - D^d C' \exp(-C''n)) - 1 \\
& = 1 - c_1 \exp \left( -\frac{C^2\rho^2}{2 \cdot 32^2 d} \right) - \frac{c_2 \sum_{i=1}^n w_i^3}{(\sum_{i=1}^n w_i^2)^{3/2}} - O(\exp(-c_3n)),
\end{aligned}$$

with the corresponding constants  $c_1, c_2, c_3 > 0$ . Then the assertion (90) will follow using (98).

It remains to verify (c) in (108). To this end, write

$$\nabla_{\boldsymbol{\theta}} \mathcal{L}_0(\boldsymbol{\theta}_*) - \mathbb{E}[\omega(X) \nabla_{\boldsymbol{\theta}} \mathcal{L}_0(\boldsymbol{\theta}_*)] = -\frac{1}{n} \sum_{i=1}^n w_i \bar{U}_i,$$

where  $U_i = \omega(X_i) \nabla_{\boldsymbol{\theta}} \log \pi_{\boldsymbol{\theta}_*}(Y_i | X_i) \in \mathbb{R}^d$  and  $\bar{U}_i = U_i - \mathbb{E}[U_i]$  as in the statement. Then by using Cauchy-Schwarz inequality and noting that  $\|\boldsymbol{\theta} - \boldsymbol{\theta}_*\| = D$ , we get

$$T_n(\boldsymbol{\theta}) = \left\langle -\frac{1}{\sqrt{n}} \sum_{i=1}^n w_i \bar{U}_i, \frac{\boldsymbol{\theta} - \boldsymbol{\theta}_*}{D} \right\rangle \leq \left\| \frac{1}{\sqrt{n}} \sum_{i=1}^n w_i \bar{U}_i \right\| = \sqrt{\sum_{k=1}^d \left| \frac{1}{\sqrt{n}} \sum_{i=1}^n w_i \bar{U}_i(k) \right|^2}.$$

It is important to note that the distribution of the random variable on the last term above does not depend on  $\boldsymbol{\theta}$ . Note that  $w_i \bar{U}_i$  for  $i = 1, \dots, n$  are independent mean zero (but possibly non-identical distributions due to the weights  $w_i$ ) random vectors in  $\mathbb{R}^d$ , and by the hypothesis, their respective coordinates have uniformly bounded variances. Hence by a union bound,

$$\mathbb{P} \left( \inf_{\substack{\boldsymbol{\theta} \in \Theta \\ \|\boldsymbol{\theta} - \boldsymbol{\theta}_*\| = D}} T_n(\boldsymbol{\theta}) \geq t \right) \leq \sum_{k=1}^d \mathbb{P} \left( \left| \frac{1}{\sqrt{n}} \sum_{i=1}^n w_i \bar{U}_i(k) \right| \geq \frac{t}{\sqrt{d}} \right). \quad (109)$$

Let  $Q_n^k = \frac{1}{\sqrt{n}} \sum_{i=1}^n w_i \bar{U}_i(k)$ . Then by the Berry-Esseen Theorem (Theorem G.3) and the hypothesis, for  $Z \sim N(0, 1)$ ,

$$\begin{aligned}
\sup_{z \in \mathbb{R}} |\mathbb{P}(Q_n^k \leq z) - \mathbb{P}(Z \leq z)| & \leq \frac{6 \sum_{i=1}^n w_i^3 \mathbb{E}[\|\bar{U}_i\|^3]}{(\sum_{i=1}^n w_i^2 \text{Var}(\bar{U}_i(k)))^{3/2}} \\
& = \frac{6 \mathbb{E}[\|\bar{U}_1\|^3] \sum_{i=1}^n w_i^3}{(\text{Var}(\bar{U}_1(k)))^{3/2} (\sum_{i=1}^n w_i^2)^{3/2}} \\
& \leq \frac{6D_1 \sum_{i=1}^n w_i^3}{d_1^{3/2} (\sum_{i=1}^n w_i^2)^{3/2}} \\
& = \frac{K \sum_{i=1}^n w_i^3}{(\sum_{i=1}^n w_i^2)^{3/2}} \quad \text{for } k = 1, \dots, d.
\end{aligned}$$

with  $K = 6D_1/d_1^{3/2}$ . Combining with (109) and using a triangle inequality, we obtain

$$\mathbb{P} \left( \inf_{\substack{\boldsymbol{\theta} \in \Theta \\ \|\boldsymbol{\theta} - \boldsymbol{\theta}_*\| = D}} T_n(\boldsymbol{\theta}) \geq t \right) \leq d \left( \mathbb{P} \left( Z \geq \frac{t}{2\sqrt{d}} \right) + \frac{K \sum_{i=1}^n w_i^3}{(\sum_{i=1}^n w_i^2)^{3/2}} \right). \quad (110)$$

Thus (c) in (108) follows.  $\square$

## H Proof of Theorem 3.5

In this section, we provide proof of the computational and statistical estimation guarantee stated in Theorem (3.5) under the generative ITR model (11). Two key ingredients in our proof are the computational guarantee (Theorem. E.1) and non-asymptotic estimation guarantee of constrained and regularized MLE (Theorem. G.1).

As we have briefly explained in the main text, the connection between the weighted convex optimization (6) for ITR learning and the statistical estimation problem under the generative model (11) is established by writing down the corresponding maximum likelihood estimation problem. To see this, recall that covariate-treatment pair  $(\mathbf{x}_i, a_i)$  follows a joint distribution  $\pi$  that does not depend on the true parameter  $\mathbf{B}_*$ . Further, assume that  $\pi$  admits a density function  $f_\pi$ . Then, for the continuous setting, the joint log-likelihood of observing the triple  $(\mathbf{x}_i, a_i, y_i)$  gives under the generative model with parameter  $\mathbf{B}$

$$-\frac{1}{2\sigma^2} \sum_{i=1}^n \left( y_i - \frac{K-1}{K} \mathbf{u}_{a_i}^T \mathbf{B}^T \mathbf{x}_i \right)^2 + \sum_{i=1}^n \log f_\pi(\mathbf{x}_i, a_i) + \text{constant}.$$

Thus, the standard (unweighted) maximum likelihood estimation of the true parameter  $\mathbf{B}_*$  can be written as

$$\min_{\mathbf{B} \in \mathbb{R}^{p \times (K-1)}, \|\mathbf{B}\| \leq \lambda_1} \sum_{i=1}^n \left( \frac{K}{K-1} y_i - \mathbf{u}_{a_i}^T \mathbf{B}^T \mathbf{x}_i \right)^2,$$

where we imposed additional  $L_1$ -ball constraint for the parameter  $\mathbf{B}$ . Now, if we use covariate-balancing weight  $w(\mathbf{x}_i, a_i)$  for each subject  $i$ , the corresponding weighted MLE problem reads as

$$\min_{\mathbf{B} \in \mathbb{R}^{p \times (K-1)}, \|\mathbf{B}\| \leq \lambda_1} \sum_{i=1}^n w(\mathbf{x}_i, a_i) \left( \frac{K}{K-1} y_i - \mathbf{u}_{a_i}^T \mathbf{B}^T \mathbf{x}_i \right)^2, \quad (111)$$

which is exactly the weighted convex optimization problem (6) for continuous outcome. A similar discussion applies to the binary outcome simply by noting that the joint log-likelihood is given by

$$\sum_{i=1}^n \left( -\log(1 + \exp(\mathbf{u}_{a_i}^T \mathbf{B}^T \mathbf{x}_i)) + y_i \mathbf{u}_{a_i}^T \mathbf{B}^T \mathbf{x}_i \right) + \sum_{i=1}^n \log f_\pi(\mathbf{x}_i, a_i).$$

**Proof of Theorem 3.5.** Consider the generative ITR model (11) with true parameter  $\mathbf{B}_*$ . Fix  $\varepsilon > 0$ . Let  $(\widehat{\mathbf{B}}_t)_{t \geq 0}$  denote the estimated parameters by using the PGD algorithm (7) for the weighted empirical maximum likelihood estimation problem (6). Decompose the total estimation error into computational and statistical parts as

$$\|\mathbf{B}_* - \widehat{\mathbf{B}}_T\|_F \leq \underbrace{\|\widehat{\mathbf{B}} - \widehat{\mathbf{B}}_T\|_F}_{=: I_1} + \underbrace{\|\widehat{\mathbf{B}} - \mathbf{B}_*\|_F}_{=: I_2}.$$

By Theorem E.1, we have

$$I_1 \leq \begin{cases} 4\rho(\alpha)^T \lambda_1^2 & \text{with constant stepsize} \\ \frac{\nu}{\gamma+T} & \text{with diminishing stepsize.} \end{cases}$$

We wish to apply our general result on weighted and constrained MLE in (Theorem G.1) with  $L_2$ -regularization  $R(\mathbf{B}) = \frac{\lambda_2}{2} \|\mathbf{B}\|_F^2$  in order to get a high-probability bound on the term  $I_2$ . For this, we first verify the hypothesis in Theorem G.1 for our generative ITR model (11). Assumption (a1) in Theorem G.1 follows from Assumption 3.4. For Assumption (a3) in Theorem G.1, denoting by  $\rho$  the minimum eigenvalue of the expected regularized Fisher information, by Lemma E.3 (also see (10)),

$$\rho = \mu = \begin{cases} \lambda^- + \lambda_2 & \text{for continuous outcome} \\ \alpha^- \lambda^- + \lambda_2 & \text{for binary outcome.} \end{cases}$$

where  $\lambda_-$  is the minimum eigenvalue of the expected weighted design matrix  $\mathbb{E}[\Psi]$  in (8) (see Assumption 3.1). The above quantity is positive under the hypothesis.

Lastly, for Assumption **(a2)** in Theorem **G.1**, we need to show that the true parameter  $\mathbf{B}_*$  is a global maximizer of the expected log-likelihood function of our generative ITR model **(11)**. This can be shown by a standard argument under the mild assumption that expectations and derivatives are exchangeable. For simplicity, we prove this for the case when  $\mathbf{B}_*$  is in the interior of the constraint set  $\Theta$ . (The general case can be shown similarly by considering the first-order optimality condition for stationary points.) First note that, if we denote the likelihood of observing the triple  $(X, A, Y)$  under the model parameter  $\mathbf{B}$  as  $p_{\mathbf{B}}(X, A, Y) = p_{\mathbf{B}}(Y | X, A) p(X, A)$ , then  $\log p_{\mathbf{B}}(X, A, Y)$  is a concave function in  $\mathbf{B}$ . Thus we only need to show that the true parameter  $\mathbf{B}_*$  is a critical point of the expected weighted log-likelihood function  $\mathbb{E}[\omega(A, X) \log p_{\mathbf{B}}(X, A, Y)]$ . Indeed,

$$\begin{aligned} & \nabla_{\mathbf{B}} \mathbb{E}[\omega(A, X) \log p_{\mathbf{B}}(X, A, Y)] \\ &= \mathbb{E}[\omega(A, X) \nabla_{\mathbf{B}} (\log p_{\mathbf{B}}(Y | X, A) + \log p(X, A))] \\ &= \mathbb{E}[\omega(A, X) \nabla_{\mathbf{B}} \log p_{\mathbf{B}}(Y | X, A)] \\ &= \mathbb{E}\left[\omega(A, X) \frac{\nabla_{\mathbf{B}} p_{\mathbf{B}}(Y | X, A)}{p_{\mathbf{B}}(Y | X, A)}\right] \\ &= \mathbb{E}_X \left[ \mathbb{E}_{(Y, A)} \left[ \omega(A, X) \frac{\nabla_{\mathbf{B}} p_{\mathbf{B}}(Y | X, A)}{p_{\mathbf{B}}(Y | X, A)} \middle| X \right] \right] \\ &= \mathbb{E}_X \left[ \sum_{a=0}^K \omega(a, X) \pi(a, X) \mathbb{E}_{Y \sim p_{\mathbf{B}_*}(\cdot | X, a)} \left[ \frac{\nabla_{\mathbf{B}} p_{\mathbf{B}}(Y | X, A)}{p_{\mathbf{B}}(Y | X, A)} \middle| X, A = a \right] \right], \end{aligned}$$

where we have used iterated expectation to condition first on the covariate  $X$  and then integrate out the treatment-response pair  $(A, Y)$ . Now note that

$$\mathbb{E}_{Y \sim p_{\mathbf{B}_*}(\cdot | X, a)} \left[ \frac{\nabla_{\mathbf{B}} p_{\mathbf{B}_*}(Y | X, A)}{p_{\mathbf{B}_*}(Y | X, A)} \middle| X, A = a \right] = \int \nabla_{\mathbf{B}} p_{\mathbf{B}_*}(Y | X, A) dy \quad (112)$$

$$= \nabla_{\mathbf{B}} \int p_{\mathbf{B}_*}(Y | X, A) dy \quad (113)$$

$$= \nabla_{\mathbf{B}} \mathbf{1} = 0. \quad (114)$$

Thus  $\mathbf{B}_*$  is a critical point of  $\mathbb{E}[\log p_{\mathbf{B}}(X, A, Y)]$ . By the concavity of the log-likelihood function, it follows that  $\mathbf{B}_*$  is a global maximizer of  $\mathbb{E}[\omega(A, X) \log p_{\mathbf{B}}(X, A, Y)]$ . In particular, this verifies the assumption **G.1 (a2)**.

Now we can apply Theorem **G.1** with  $L_2$ -regularization  $R(\mathbf{B}) = \frac{\lambda_2}{2} \|\mathbf{B}\|_F^2$  for our generative ITR model. Denoting

$$D := \frac{C}{\sqrt{n}} + \frac{8\|\nabla R(\mathbf{B}_*)\|_F}{\mu} = \frac{C}{\sqrt{n}} + \frac{8\lambda_2 \|\mathbf{B}_*\|_F}{\mu} \quad (115)$$

for  $C > 0$  any constant, this gives

$$\mathbb{P}(I_2 > D) \leq c_1 \exp\left(-\frac{C^2 \mu^2}{2 \cdot 32^2 d}\right) + \frac{c_2 \sum_{i=1}^n w_i^3}{(\sum_{i=1}^n w_i^2)^{3/2}} + O(\exp(-c_3 n)),$$

where  $c_1, c_2, c_3 > 0$  are constants and  $d = p(K-1)$ . By Assumption **3.4**, the quantities  $\sqrt{n} \sum_{i=1}^n w_i^3 / (\sum_{i=1}^n w_i^2)^{3/2}$  are uniformly bounded in  $n$ . Hence we can first choose  $n$  large enough so that the last two terms combined in the right-hand side in **(116)** are at most  $\varepsilon/2$ , for which  $n \gtrsim \varepsilon^{-2}$  is sufficient. Then we can choose  $C > 0$  large enough so that the first term on the right-hand side of **(116)** is at most  $\varepsilon/2$ . For this, we choose

$$C = \frac{32\sqrt{2d}}{\mu} \sqrt{\log \frac{2c_1}{\varepsilon}} = \Omega\left(\frac{\sqrt{p \log \varepsilon^{-1}}}{\mu}\right)$$

Accordingly, we can choose  $T$  large enough so that  $I_1 \leq C/\sqrt{n}$ . Then the above gives

$$\mathbb{P}\left(\|\mathbf{B}_* - \hat{\mathbf{B}}_T\|_F > \frac{2C}{\sqrt{n}} + \frac{8\lambda_2 \|\mathbf{B}_*\|_F}{\mu}\right) \leq \mathbb{P}(I_2 \geq D) \leq \varepsilon. \quad (116)$$

This shows the assertion.  $\square$

## I Additional simulation details and results

The code supporting this study is available at <https://github.com/ljw9510/effective-ITR>, with plans for release as an R package soon.

### I.1 Details of data-generating mechanism

In this section, we explain how data are generated and treatment assignments are allocated in our simulation studies. The outcome is generated under the working model assumption given by Equation 1 with heteroscedastic random noise  $\epsilon \sim N(0, \sigma^2(A, \mathbf{X}))$ . We have four different settings/cases for the treatment-free effect and interaction effect functions, which are outlined below.

#### Case 1: Linear interaction functions with simple treatment-free effect functions

For Case 1, our randomized study aligns with Scenario 7 and our observational study mirrors Scenario 8 described in [44]. Specifically, we consider the following settings for the treatment-free effects and interaction effects:

Randomized trial:

$$\mu(\mathbf{X}) = 1 + 2X_1 + 2X_2,$$

$$\delta(\mathbf{X}) = \begin{cases} 0.75 + 1.5X_1 + 1.5X_2 + 1.5X_3 + 1.5X_4, & A = 1; \\ 0.75 + 1.5X_1 - 1.5X_2 - 1.5X_3 + 1.5X_4, & A = 2; \\ 0.75 + 1.5X_1 - 1.5X_2 + 1.5X_3 - 1.5X_4, & A = 3; \\ 0.75 - 1.5X_1 + 1.5X_2 - 1.5X_3 - 1.5X_4, & A = 4, \end{cases}$$

Observational study:

$$\mu(\mathbf{X}) = 1 + X_5 + 3X_6 + 2X_1X_2,$$

$$\delta(\mathbf{X}) = \begin{cases} 0.5 + 2X_1 + X_2 + X_3, & A = 1; & 1 + X_1 - X_2 - X_3, & A = 2; \\ 1.5 + 3X_1 - X_2 + X_3, & A = 3; & 1 - X_1 - X_2 + X_3, & A = 4, \end{cases}$$

#### Case 2: Linear interaction functions with complicated treatment-free effect functions

For Case 2, we consider the following treatment-free effect function for model assumptions of  $Y$ :

$$\mu(\mathbf{X}) = 1 + X_5 + X_5^2 + 2e^{-X_1X_2} + \sin(X_3).$$

#### Case 3: Non-linear interaction functions with simple treatment-free effect functions

For Case 3, we consider the following non-linear interaction functions for model assumptions of  $Y$ :

$$\mu(\mathbf{X}) = 1 + 2X_1 + 2X_2 + 2X_4 - 2X_4^2 + 2X_1X_2,$$

$$\delta(\mathbf{X}) = \begin{cases} 0.5 + 1.0X_1 - 2.0X_4 + 0.5X_4^2, & A = 1; & 1.0 + 1.0X_1 + 1.0X_4 - 1.0X_4^2, & A = 2; \\ 1.5 + 2.0X_1 - 1.0X_4 - 1.0X_4^2, & A = 3; & 1.0 - 1.0X_1 - 1.0X_4 - 1.0X_4^2, & A = 4. \end{cases}$$

#### Case 4: Non-linear interaction functions with a complicated treatment-free effect function

For Case 4, we consider the following complicated treatment-free effect function for model assumptions of  $Y$ :

$$\mu(\mathbf{X}) = 1 + 2X_1 + 2X_2 + 2X_4 - 2X_4^2 + 2X_1X_2 + 2e^{-X_1X_2} + \sin(X_3).$$

In all four cases, we adopt the same treatment assignment principles and variance function  $\sigma^2(A, \mathbf{X})$  that induces heteroscedastic error for observational studies as outlined in [44]:

$$\pi(A, \mathbf{X}) = \begin{cases} 0.25 \cdot \mathbf{I}(X_1 < 0) + 0.4 \cdot \mathbf{I}(X_1 > 0), & A = 1, \\ 0.25 \cdot \mathbf{I}(X_1 < 0) + 0.2 \cdot \mathbf{I}(X_1 > 0), & A = 2, 3, 4. \end{cases}$$

$$\sigma^2(A, \mathbf{X}) = 0.25 + 2X_2 \cdot \mathbf{I}(X_2 > 0) + X_3 \cdot \mathbf{I}(X_3 > 0, A = 1) + X_4 \cdot \mathbf{I}(X_4 > 0, A = 2).$$

For randomized trials, each individual has an equal probability of assigning each treatment, therefore, no propensity score estimation was made for AD-learning and SABD-learning. For the heteroscedastic error for randomized trials,  $\sigma^2(A, \mathbf{X}) = 0.25 + 0.2(1.5 - X_2)^2$  was used.

## I.2 Implementation details

All numerical experiments were performed on a 2022 Macbook Air with M1 chip and 16 GB of RAM.

In the computation of the final ITR estimates, we use two different approaches: one with weighted multivariate regression with  $L_1$ -regularization for the penalized approach (named ‘penalized’ in the main text). The other uses the proposed PGD algorithm with  $L_1$ -ball (named ‘constrained’ in the main text). To ensure a fair comparison, both the penalized method and the constrained optimization algorithm were implemented using code developed by the author. We do not consider additional  $L_2$ -regularization in this setting. To determine the regularization constant or  $L_1$ -ball size, we use mean squared error as the criterion. The iterate number  $T$  of the PGD algorithm is 1000.

In our simulation study, we employ the default configuration of the random forest algorithm as implemented in the `randomForest` package in R to estimate the residual variance function in SABD-Learning, the treatment-free effect, and the propensity score. This default configuration includes the construction of an ensemble of 500 decision trees, with each tree’s splitting criterion determined using the Gini index. The number of variables considered for splitting at each node is set to the square root of the total number of covariates. For model performance evaluation, we utilize the *out-of-bag* (OOB) error estimate.

## I.3 Additional simulation results

The goal of these additional simulations in the appendix is to demonstrate, under the pre-specified linear ITR, that DCBW, variable screening, and augmentation can synergistically benefit the final ITR. Methods using energy balancing weights are denoted as “\_e” and those employing the distance covariance test for variable screening are indicated as “\_s” in the table. Note that the augmented “SABD\_e\_s” corresponds to “proposed+penalized” in the main text. In this section, we include the performance of various methods implemented only by the penalized approach to compare the effectiveness of the proposed statistical methods. The simulation was implemented using package `glmnet` in R.

### I.3.1 Treatment Decision Accuracy

Here, we show the averages and standard deviations of treatment decision accuracy from the additional simulation studies. The first four tables present the results of experiments that mimic randomized trials, and the remaining four tables show the results of observational studies.

The simulation results indicate that individual components alone do not significantly improve performance. Instead, the synergistic effect of combining these components is essential for the ITR-Learning. For example, in high-dimensional settings, using DCBW alone balances the empirical distributions of all covariates (up to 60), including a significant portion of irrelevant ones. Additionally, estimated decision functions may include irrelevant variables, reducing the impact of DCBW. Similarly, using variable screening alone remains challenging due to model misspecification or highly nonlinear outcomes. Combining other methods, such as outcome augmentation, helps mitigate model misspecification effects. This combined application ensures efficient and robust optimal ITR estimation across diverse scenarios.

Below is a detailed investigation of each scenario.

First, we examine Case 1, where the treatment rule class is correctly specified (Tables 1, 5). With a relatively simple treatment-free effect, AD-Learning’s performance is comparable to SABD-Learning, particularly following outcome augmentation. When combined with EBWs as AD\_e, SABD\_e, these methods exhibit improved performance over their original versions (AD, SABD). In randomized trials, EBWs effectively reduce finite sample imbalance compared to IPW with true propensity scores. In observational studies, energy balancing yields more effective balancing weights, resulting in higher accuracy.



Next, we delve into Case 2, involving a more complicated treatment-free effect function compared to Case 1 (Tables 2, 6). Notably, since the treatment-free effect does not affect ITR learning, a misspecified treatment-free effect function introduces heterogeneity and increases noise variance. Then we can see the synergistic effectiveness of outcome augmentation and inverse variance weighting in SABD-Learning, effectively handling additional heterogeneity in ITR estimation (See “augmented” panel in Tables 2, 6). However, it’s important to note that all variants of AD-Learning show subpar performance, even with augmented outcomes. AD-Learning results in an intercept-only model frequently as its estimated decision function, which suggests that the mean squared error criterion in LASSO regularization may not be effective for selecting models. In such instances, different penalty types or performance metrics could prove beneficial.

Case 3 and Case 4 handle situations where the underlying treatment rule class is misspecified (Tables 3, 4, 7, and 8). Particularly, in Case 4, estimating ITR becomes even more challenging due to a highly non-linear covariate-outcome relationship, which leads to suboptimal results of the variants of AD-Learning as in Case 2. Our results demonstrate that the more robust EBWs outperform standard IPW, particularly in observational studies. Additionally, variable screening enhances ITR learning in a high-dimensional setting. Employing variable screening (SABD\_s) enhances the performance of SABD, especially with limited sample sizes, as it aids in selecting effect modifiers and precision variables, thereby facilitating decision function estimation. However, even with perfect screening, estimating the decision function remains challenging due to model misspecification or highly nonlinear outcomes. Simultaneous use of EBWs and screening for SABD (SABD\_e\_s) further elevates performance, likely influenced by dimensionality in nonparametric balancing weight methods.

The significant enhancements from integrating DCBWs, variable screening, outcome augmentation, and inverse variance weighting underscore their collective importance. Their combined application ensures efficient and robust optimal ITR estimation across diverse scenarios.

Table 2: Average accuracy and their standard deviations (in parenthesis) in the randomized trial in Case 1: Linear interaction functions with simple treatment-free effect functions.

		$AD$	$AD_e$	$AD_s$	$AD_{e,s}$	$SABD$	$SABD_e$	$SABD_s$	$SABD_{e,s}$	
n	p	Original								
200	20	0.679 (0.06)	0.734 (0.06)	0.703 (0.06)	0.813 (0.06)	0.720 (0.06)	0.745 (0.06)	0.734 (0.06)	0.804 (0.06)	
	40	0.637 (0.07)	0.673 (0.08)	0.658 (0.07)	0.755 (0.08)	0.691 (0.06)	0.700 (0.06)	0.702 (0.07)	0.761 (0.07)	
	60	0.633 (0.07)	0.649 (0.08)	0.644 (0.06)	0.720 (0.08)	0.671 (0.06)	0.686 (0.07)	0.685 (0.05)	0.739 (0.06)	
	20	0.816 (0.05)	0.872 (0.05)	0.828 (0.04)	0.925 (0.04)	0.832 (0.04)	0.867 (0.04)	0.818 (0.04)	0.875 (0.03)	
	40	0.814 (0.04)	0.863 (0.03)	0.808 (0.04)	0.895 (0.07)	0.840 (0.04)	0.867 (0.03)	0.821 (0.04)	0.871 (0.05)	
	60	0.802 (0.04)	0.848 (0.03)	0.783 (0.04)	0.876 (0.06)	0.836 (0.04)	0.859 (0.03)	0.815 (0.04)	0.864 (0.03)	
1000	20	0.865 (0.03)	0.902 (0.05)	0.867 (0.03)	0.953 (0.04)	0.866 (0.03)	0.882 (0.04)	0.846 (0.04)	0.875 (0.03)	
	40	0.854 (0.03)	0.900 (0.03)	0.846 (0.03)	0.927 (0.07)	0.866 (0.04)	0.893 (0.03)	0.839 (0.04)	0.876 (0.04)	
	60	0.857 (0.03)	0.895 (0.02)	0.838 (0.03)	0.922 (0.05)	0.874 (0.03)	0.893 (0.02)	0.846 (0.03)	0.879 (0.03)	
	Augmented									
	20	0.798 (0.04)	0.807 (0.04)	0.806 (0.05)	0.853 (0.05)	0.807 (0.04)	0.807 (0.04)	0.820 (0.05)	0.854 (0.05)	
	40	0.781 (0.04)	0.779 (0.05)	0.776 (0.06)	0.814 (0.07)	0.784 (0.05)	0.776 (0.05)	0.787 (0.06)	0.813 (0.07)	
60	0.769 (0.05)	0.764 (0.05)	0.759 (0.05)	0.789 (0.06)	0.772 (0.05)	0.762 (0.05)	0.776 (0.05)	0.795 (0.06)		
600	20	0.891 (0.02)	0.908 (0.02)	0.898 (0.02)	0.939 (0.02)	0.899 (0.02)	0.912 (0.02)	0.896 (0.02)	0.934 (0.02)	
	40	0.891 (0.02)	0.898 (0.02)	0.892 (0.02)	0.923 (0.03)	0.899 (0.02)	0.902 (0.02)	0.897 (0.02)	0.925 (0.03)	
	60	0.883 (0.03)	0.889 (0.02)	0.873 (0.02)	0.910 (0.03)	0.893 (0.02)	0.893 (0.03)	0.886 (0.02)	0.915 (0.02)	
	20	0.921 (0.02)	0.929 (0.03)	0.924 (0.02)	0.957 (0.02)	0.927 (0.01)	0.931 (0.03)	0.921 (0.02)	0.948 (0.02)	
	40	0.918 (0.02)	0.929 (0.02)	0.916 (0.01)	0.949 (0.02)	0.924 (0.01)	0.933 (0.01)	0.918 (0.02)	0.946 (0.02)	
	60	0.915 (0.02)	0.923 (0.02)	0.908 (0.02)	0.942 (0.03)	0.921 (0.02)	0.926 (0.02)	0.915 (0.02)	0.942 (0.02)	

Table 3: Average accuracy and their standard deviations (in parenthesis) in the randomized trial in Case 2: Linear interaction functions with complicated treatment-free effect functions.

		$AD$	$AD_e$	$AD_s$	$AD_{e,s}$	$SABD$	$SABD_e$	$SABD_s$	$SABD_{e,s}$	
n	p	Original								
200	20	0.348 (0.12)	0.324 (0.14)	0.366 (0.14)	0.373 (0.18)	0.440 (0.13)	0.412 (0.16)	0.488 (0.14)	0.525 (0.21)	
	40	0.316 (0.10)	0.295 (0.10)	0.346 (0.13)	0.329 (0.14)	0.379 (0.12)	0.337 (0.12)	0.454 (0.15)	0.457 (0.19)	
	60	0.309 (0.08)	0.286 (0.08)	0.336 (0.11)	0.310 (0.12)	0.345 (0.10)	0.320 (0.10)	0.454 (0.13)	0.444 (0.18)	
	20	0.440 (0.15)	0.472 (0.19)	0.470 (0.17)	0.523 (0.22)	0.693 (0.08)	0.773 (0.12)	0.695 (0.09)	0.785 (0.12)	
	40	0.391 (0.16)	0.415 (0.20)	0.409 (0.17)	0.448 (0.21)	0.660 (0.11)	0.736 (0.13)	0.692 (0.10)	0.784 (0.12)	
	60	0.367 (0.14)	0.367 (0.16)	0.387 (0.15)	0.408 (0.18)	0.599 (0.13)	0.688 (0.15)	0.669 (0.11)	0.757 (0.12)	
1000	20	0.458 (0.16)	0.473 (0.18)	0.477 (0.17)	0.528 (0.21)	0.769 (0.07)	0.818 (0.10)	0.746 (0.08)	0.816 (0.08)	
	40	0.408 (0.17)	0.410 (0.19)	0.436 (0.18)	0.480 (0.21)	0.741 (0.09)	0.826 (0.09)	0.744 (0.09)	0.827 (0.09)	
	60	0.436 (0.17)	0.467 (0.20)	0.465 (0.18)	0.492 (0.21)	0.746 (0.09)	0.833 (0.10)	0.766 (0.06)	0.816 (0.12)	
	Augmented									
	20	0.417 (0.18)	0.418 (0.18)	0.436 (0.19)	0.474 (0.20)	0.650 (0.10)	0.647 (0.11)	0.716 (0.10)	0.740 (0.10)	
	40	0.367 (0.15)	0.362 (0.15)	0.384 (0.17)	0.413 (0.19)	0.556 (0.14)	0.541 (0.15)	0.679 (0.11)	0.689 (0.12)	
60	0.353 (0.14)	0.348 (0.13)	0.374 (0.15)	0.393 (0.16)	0.517 (0.15)	0.506 (0.14)	0.648 (0.11)	0.665 (0.12)		
600	20	0.492 (0.19)	0.520 (0.20)	0.514 (0.20)	0.583 (0.21)	0.857 (0.04)	0.862 (0.04)	0.873 (0.03)	0.894 (0.06)	
	40	0.437 (0.20)	0.444 (0.20)	0.457 (0.20)	0.481 (0.22)	0.833 (0.05)	0.835 (0.04)	0.864 (0.04)	0.878 (0.04)	
	60	0.396 (0.17)	0.399 (0.18)	0.409 (0.17)	0.444 (0.19)	0.803 (0.06)	0.798 (0.08)	0.843 (0.04)	0.861 (0.04)	
	20	0.485 (0.18)	0.492 (0.19)	0.501 (0.19)	0.541 (0.21)	0.897 (0.03)	0.896 (0.04)	0.905 (0.02)	0.918 (0.04)	
	40	0.433 (0.19)	0.436 (0.20)	0.453 (0.20)	0.504 (0.21)	0.888 (0.03)	0.890 (0.03)	0.903 (0.02)	0.921 (0.03)	
	60	0.469 (0.19)	0.485 (0.20)	0.490 (0.20)	0.513 (0.21)	0.884 (0.03)	0.883 (0.03)	0.899 (0.02)	0.908 (0.03)	

Table 4: Average accuracy and their standard deviations (in parenthesis) in the randomized trial in Case 3: Non-linear interaction functions with simple treatment-free effect functions.

		$AD$	$AD_e$	$AD_s$	$AD_{e,s}$	$SABD$	$SABD_e$	$SABD_s$	$SABD_{e,s}$	
n	p	Original								
200	20	0.385 (0.12)	0.353 (0.13)	0.411 (0.12)	0.404 (0.17)	0.476 (0.12)	0.438 (0.14)	0.521 (0.12)	0.550 (0.16)	
	40	0.339 (0.12)	0.316 (0.11)	0.375 (0.11)	0.358 (0.14)	0.408 (0.12)	0.366 (0.12)	0.473 (0.10)	0.486 (0.14)	
	60	0.319 (0.11)	0.300 (0.10)	0.357 (0.11)	0.342 (0.12)	0.370 (0.12)	0.343 (0.11)	0.454 (0.10)	0.438 (0.13)	
	20	0.534 (0.11)	0.561 (0.11)	0.558 (0.10)	0.630 (0.10)	0.678 (0.07)	0.705 (0.07)	0.659 (0.08)	0.746 (0.07)	
	40	0.490 (0.11)	0.486 (0.13)	0.529 (0.08)	0.576 (0.11)	0.651 (0.07)	0.653 (0.09)	0.651 (0.07)	0.723 (0.08)	
	60	0.472 (0.10)	0.460 (0.12)	0.501 (0.08)	0.536 (0.10)	0.637 (0.07)	0.648 (0.07)	0.639 (0.07)	0.687 (0.07)	
1000	20	0.577 (0.09)	0.598 (0.09)	0.599 (0.08)	0.690 (0.06)	0.740 (0.06)	0.755 (0.05)	0.689 (0.08)	0.768 (0.07)	
	40	0.584 (0.10)	0.583 (0.09)	0.599 (0.09)	0.663 (0.07)	0.724 (0.06)	0.728 (0.08)	0.702 (0.06)	0.779 (0.06)	
	60	0.574 (0.09)	0.579 (0.09)	0.584 (0.06)	0.636 (0.08)	0.723 (0.05)	0.731 (0.05)	0.707 (0.05)	0.760 (0.07)	
	Augmented									
	200	20	0.524 (0.11)	0.524 (0.12)	0.555 (0.10)	0.603 (0.11)	0.608 (0.08)	0.604 (0.09)	0.642 (0.07)	0.687 (0.08)
	40	0.476 (0.12)	0.463 (0.13)	0.513 (0.10)	0.548 (0.11)	0.565 (0.10)	0.545 (0.09)	0.607 (0.07)	0.630 (0.08)	
60	0.443 (0.12)	0.431 (0.13)	0.484 (0.12)	0.508 (0.13)	0.525 (0.11)	0.505 (0.12)	0.601 (0.08)	0.601 (0.09)		
600	20	0.674 (0.06)	0.680 (0.07)	0.685 (0.05)	0.719 (0.06)	0.776 (0.04)	0.784 (0.04)	0.776 (0.04)	0.805 (0.04)	
40	0.654 (0.07)	0.659 (0.06)	0.664 (0.06)	0.697 (0.06)	0.761 (0.04)	0.763 (0.04)	0.768 (0.04)	0.796 (0.04)		
60	0.643 (0.06)	0.645 (0.07)	0.652 (0.05)	0.676 (0.06)	0.750 (0.05)	0.749 (0.05)	0.756 (0.04)	0.780 (0.04)		
1000	20	0.706 (0.05)	0.708 (0.06)	0.715 (0.04)	0.746 (0.04)	0.805 (0.03)	0.805 (0.05)	0.787 (0.04)	0.817 (0.04)	
40	0.704 (0.06)	0.704 (0.05)	0.711 (0.05)	0.738 (0.04)	0.805 (0.03)	0.809 (0.03)	0.807 (0.04)	0.831 (0.03)		
60	0.705 (0.05)	0.707 (0.04)	0.710 (0.04)	0.731 (0.04)	0.800 (0.03)	0.798 (0.03)	0.803 (0.04)	0.820 (0.04)		

Table 5: Average accuracy and their standard deviations (in parenthesis) in the randomized trial in Case 4: Non-linear interaction functions with a complicated treatment-free effect function.

		$AD$	$AD_e$	$AD_s$	$AD_{e,s}$	$SABD$	$SABD_e$	$SABD_s$	$SABD_{e,s}$
n	p	Original							
200	20	0.298 (0.09)	0.300 (0.09)	0.319 (0.10)	0.299 (0.10)	0.351 (0.12)	0.345 (0.12)	0.384 (0.13)	0.379 (0.16)
	40	0.273 (0.06)	0.269 (0.05)	0.283 (0.07)	0.279 (0.07)	0.303 (0.09)	0.293 (0.09)	0.350 (0.11)	0.343 (0.13)
	60	0.279 (0.07)	0.275 (0.06)	0.299 (0.08)	0.289 (0.09)	0.298 (0.09)	0.287 (0.09)	0.359 (0.12)	0.340 (0.13)
	20	0.320 (0.11)	0.331 (0.12)	0.339 (0.12)	0.353 (0.13)	0.570 (0.12)	0.570 (0.13)	0.572 (0.13)	0.622 (0.13)
	40	0.308 (0.10)	0.306 (0.09)	0.329 (0.11)	0.332 (0.13)	0.501 (0.13)	0.511 (0.13)	0.552 (0.10)	0.600 (0.11)
	60	0.297 (0.08)	0.293 (0.08)	0.308 (0.09)	0.307 (0.10)	0.463 (0.13)	0.473 (0.14)	0.529 (0.11)	0.578 (0.14)
600	20	0.325 (0.11)	0.331 (0.12)	0.337 (0.12)	0.356 (0.14)	0.638 (0.09)	0.656 (0.09)	0.614 (0.10)	0.658 (0.10)
	40	0.312 (0.10)	0.316 (0.11)	0.322 (0.11)	0.335 (0.13)	0.631 (0.09)	0.635 (0.12)	0.625 (0.09)	0.674 (0.09)
	60	0.308 (0.09)	0.328 (0.12)	0.323 (0.10)	0.341 (0.13)	0.629 (0.08)	0.637 (0.10)	0.628 (0.08)	0.675 (0.10)
	20	0.347 (0.12)	0.355 (0.13)	0.355 (0.14)	0.380 (0.14)	0.500 (0.13)	0.481 (0.14)	0.558 (0.12)	0.595 (0.13)
	40	0.288 (0.08)	0.287 (0.08)	0.299 (0.10)	0.314 (0.11)	0.376 (0.13)	0.377 (0.12)	0.506 (0.11)	0.499 (0.14)
	60	0.307 (0.10)	0.302 (0.09)	0.319 (0.11)	0.327 (0.12)	0.379 (0.13)	0.362 (0.12)	0.492 (0.13)	0.481 (0.14)
1000	20	0.348 (0.13)	0.369 (0.14)	0.361 (0.14)	0.401 (0.16)	0.725 (0.08)	0.724 (0.08)	0.749 (0.06)	0.764 (0.09)
	40	0.349 (0.13)	0.353 (0.14)	0.360 (0.14)	0.382 (0.15)	0.667 (0.09)	0.653 (0.10)	0.729 (0.06)	0.747 (0.06)
	60	0.318 (0.12)	0.318 (0.11)	0.336 (0.13)	0.343 (0.13)	0.651 (0.10)	0.644 (0.11)	0.720 (0.07)	0.734 (0.07)
	20	0.339 (0.14)	0.352 (0.15)	0.350 (0.14)	0.390 (0.16)	0.775 (0.05)	0.779 (0.06)	0.781 (0.06)	0.800 (0.05)
	40	0.332 (0.12)	0.339 (0.14)	0.346 (0.13)	0.361 (0.14)	0.764 (0.06)	0.756 (0.07)	0.793 (0.05)	0.803 (0.05)
	60	0.331 (0.13)	0.347 (0.13)	0.342 (0.14)	0.358 (0.14)	0.766 (0.05)	0.753 (0.05)	0.788 (0.04)	0.803 (0.04)
Augmented									

Table 6: Average accuracy and their standard deviations (in parenthesis) in the observational study in Case 1: Linear interaction functions with simple treatment-free effect functions.

		$AD$	$AD_e$	$AD_s$	$AD_{e,s}$	$SABD$	$SABD_e$	$SABD_s$	$SABD_{e,s}$	
n	p	Original								
200	20	0.560 (0.11)	0.604 (0.12)	0.580 (0.10)	0.660 (0.11)	0.555 (0.10)	0.634 (0.10)	0.563 (0.10)	0.694 (0.10)	
	40	0.494 (0.12)	0.499 (0.14)	0.532 (0.11)	0.589 (0.13)	0.517 (0.09)	0.550 (0.12)	0.542 (0.10)	0.624 (0.11)	
	60	0.498 (0.11)	0.494 (0.13)	0.520 (0.09)	0.546 (0.13)	0.487 (0.11)	0.533 (0.12)	0.512 (0.10)	0.583 (0.12)	
	20	0.728 (0.07)	0.785 (0.07)	0.738 (0.07)	0.849 (0.06)	0.675 (0.06)	0.824 (0.04)	0.664 (0.05)	0.867 (0.05)	
	40	0.716 (0.08)	0.754 (0.07)	0.721 (0.07)	0.824 (0.07)	0.674 (0.07)	0.802 (0.06)	0.670 (0.07)	0.856 (0.05)	
	60	0.704 (0.08)	0.756 (0.07)	0.693 (0.07)	0.795 (0.08)	0.661 (0.08)	0.778 (0.06)	0.653 (0.07)	0.829 (0.05)	
600	20	0.782 (0.06)	0.845 (0.05)	0.788 (0.06)	0.894 (0.05)	0.708 (0.06)	0.873 (0.03)	0.689 (0.05)	0.905 (0.04)	
	40	0.774 (0.06)	0.831 (0.04)	0.767 (0.06)	0.886 (0.05)	0.700 (0.06)	0.859 (0.04)	0.684 (0.05)	0.901 (0.04)	
	60	0.764 (0.06)	0.825 (0.05)	0.755 (0.06)	0.868 (0.06)	0.702 (0.07)	0.845 (0.04)	0.684 (0.05)	0.885 (0.04)	
	Augmented									
	20	0.692 (0.07)	0.699 (0.07)	0.690 (0.07)	0.727 (0.08)	0.712 (0.07)	0.722 (0.07)	0.702 (0.08)	0.741 (0.08)	
	40	0.660 (0.09)	0.662 (0.08)	0.661 (0.08)	0.687 (0.09)	0.676 (0.07)	0.678 (0.08)	0.677 (0.08)	0.707 (0.08)	
60	0.648 (0.08)	0.632 (0.09)	0.631 (0.08)	0.650 (0.08)	0.660 (0.08)	0.657 (0.08)	0.649 (0.10)	0.672 (0.10)		
1000	20	0.840 (0.04)	0.861 (0.03)	0.841 (0.04)	0.889 (0.04)	0.841 (0.04)	0.881 (0.03)	0.836 (0.04)	0.904 (0.04)	
	40	0.829 (0.05)	0.841 (0.04)	0.828 (0.05)	0.872 (0.04)	0.831 (0.05)	0.870 (0.03)	0.832 (0.04)	0.898 (0.03)	
	60	0.817 (0.05)	0.829 (0.05)	0.811 (0.04)	0.849 (0.05)	0.827 (0.05)	0.856 (0.04)	0.819 (0.04)	0.874 (0.04)	
	20	0.879 (0.03)	0.900 (0.03)	0.879 (0.03)	0.926 (0.02)	0.867 (0.03)	0.919 (0.02)	0.861 (0.03)	0.932 (0.02)	
	40	0.878 (0.04)	0.892 (0.03)	0.874 (0.04)	0.917 (0.03)	0.867 (0.03)	0.909 (0.03)	0.860 (0.03)	0.927 (0.02)	
	60	0.868 (0.03)	0.883 (0.03)	0.862 (0.03)	0.905 (0.04)	0.859 (0.04)	0.904 (0.02)	0.852 (0.04)	0.919 (0.03)	

Table 7: Average accuracy and their standard deviations (in parenthesis) in the observational study in Case 2: Linear interaction functions with complicated treatment-free effect functions.

		$AD$	$AD_e$	$AD_s$	$AD_{e,s}$	$SABD$	$SABD_e$	$SABD_s$	$SABD_{e,s}$
n	p	Original							
	20	0.321 (0.15)	0.327 (0.14)	0.328 (0.15)	0.356 (0.16)	0.292 (0.12)	0.361 (0.17)	0.319 (0.14)	0.394 (0.20)
200	40	0.301 (0.11)	0.300 (0.12)	0.315 (0.12)	0.325 (0.12)	0.284 (0.11)	0.318 (0.10)	0.309 (0.13)	0.345 (0.14)
	60	0.303 (0.11)	0.320 (0.12)	0.315 (0.11)	0.323 (0.12)	0.271 (0.12)	0.341 (0.13)	0.314 (0.12)	0.381 (0.15)
600	20	0.364 (0.15)	0.383 (0.17)	0.373 (0.15)	0.413 (0.18)	0.403 (0.09)	0.659 (0.15)	0.414 (0.09)	0.689 (0.16)
	40	0.355 (0.13)	0.379 (0.15)	0.368 (0.14)	0.388 (0.15)	0.360 (0.10)	0.571 (0.19)	0.404 (0.09)	0.662 (0.16)
1000	60	0.327 (0.11)	0.335 (0.14)	0.329 (0.12)	0.350 (0.14)	0.366 (0.10)	0.515 (0.19)	0.401 (0.09)	0.645 (0.17)
	20	0.348 (0.14)	0.390 (0.16)	0.352 (0.14)	0.393 (0.17)	0.427 (0.06)	0.751 (0.14)	0.417 (0.07)	0.726 (0.14)
200	40	0.364 (0.15)	0.382 (0.16)	0.374 (0.15)	0.420 (0.17)	0.426 (0.07)	0.712 (0.13)	0.430 (0.07)	0.767 (0.11)
	60	0.372 (0.14)	0.415 (0.18)	0.386 (0.15)	0.449 (0.19)	0.413 (0.08)	0.741 (0.13)	0.436 (0.08)	0.771 (0.12)
Augmented									
200	20	0.368 (0.17)	0.393 (0.18)	0.382 (0.18)	0.440 (0.20)	0.576 (0.15)	0.590 (0.15)	0.636 (0.14)	0.662 (0.14)
	40	0.340 (0.15)	0.358 (0.15)	0.353 (0.16)	0.395 (0.18)	0.472 (0.17)	0.493 (0.16)	0.586 (0.14)	0.625 (0.15)
600	60	0.330 (0.12)	0.355 (0.14)	0.339 (0.12)	0.378 (0.14)	0.453 (0.13)	0.447 (0.14)	0.567 (0.14)	0.579 (0.14)
	20	0.403 (0.17)	0.446 (0.18)	0.414 (0.19)	0.497 (0.22)	0.809 (0.06)	0.827 (0.05)	0.828 (0.05)	0.861 (0.05)
1000	40	0.381 (0.16)	0.422 (0.18)	0.397 (0.17)	0.458 (0.18)	0.761 (0.09)	0.766 (0.08)	0.818 (0.05)	0.843 (0.06)
	60	0.341 (0.12)	0.366 (0.14)	0.351 (0.13)	0.395 (0.16)	0.740 (0.10)	0.739 (0.10)	0.794 (0.06)	0.815 (0.06)
200	20	0.374 (0.17)	0.420 (0.18)	0.381 (0.18)	0.449 (0.20)	0.853 (0.07)	0.862 (0.05)	0.864 (0.04)	0.892 (0.03)
	40	0.392 (0.17)	0.429 (0.18)	0.403 (0.18)	0.443 (0.18)	0.836 (0.07)	0.848 (0.04)	0.867 (0.03)	0.885 (0.03)
600	60	0.405 (0.17)	0.446 (0.19)	0.420 (0.18)	0.484 (0.20)	0.839 (0.04)	0.844 (0.05)	0.864 (0.03)	0.885 (0.03)



Table 8: Average accuracy and their standard deviations (in parenthesis) in the observational study in Case 3: Non-linear interaction functions with simple treatment-free effect functions.

		$AD$	$AD_e$	$AD_s$	$AD_{e,s}$	$SABD$	$SABD_e$	$SABD_s$	$SABD_{e,s}$	
n	p	Original								
200	20	0.293 (0.08)	0.301 (0.09)	0.326 (0.10)	0.336 (0.14)	0.382 (0.10)	0.393 (0.12)	0.423 (0.10)	0.474 (0.17)	
	40	0.304 (0.09)	0.314 (0.09)	0.358 (0.10)	0.346 (0.13)	0.368 (0.11)	0.376 (0.12)	0.428 (0.10)	0.465 (0.15)	
	60	0.277 (0.08)	0.284 (0.08)	0.330 (0.10)	0.345 (0.11)	0.337 (0.10)	0.331 (0.10)	0.401 (0.09)	0.416 (0.12)	
	20	0.405 (0.11)	0.491 (0.12)	0.458 (0.11)	0.587 (0.13)	0.547 (0.07)	0.685 (0.07)	0.529 (0.06)	0.747 (0.08)	
	40	0.382 (0.10)	0.462 (0.12)	0.447 (0.10)	0.549 (0.13)	0.523 (0.08)	0.639 (0.08)	0.531 (0.07)	0.729 (0.08)	
	60	0.386 (0.10)	0.438 (0.11)	0.465 (0.08)	0.545 (0.11)	0.519 (0.08)	0.628 (0.09)	0.534 (0.06)	0.696 (0.08)	
1000	20	0.471 (0.09)	0.572 (0.10)	0.518 (0.08)	0.671 (0.08)	0.587 (0.06)	0.739 (0.05)	0.548 (0.06)	0.779 (0.06)	
	40	0.440 (0.08)	0.542 (0.09)	0.502 (0.07)	0.633 (0.08)	0.575 (0.06)	0.725 (0.05)	0.554 (0.05)	0.775 (0.07)	
	60	0.419 (0.09)	0.531 (0.09)	0.495 (0.07)	0.610 (0.10)	0.570 (0.06)	0.719 (0.05)	0.559 (0.05)	0.759 (0.06)	
	Augmented									
	200	20	0.448 (0.12)	0.465 (0.13)	0.492 (0.11)	0.554 (0.12)	0.535 (0.10)	0.568 (0.11)	0.571 (0.08)	0.663 (0.09)
		40	0.456 (0.11)	0.461 (0.12)	0.505 (0.10)	0.544 (0.11)	0.511 (0.11)	0.534 (0.11)	0.562 (0.09)	0.629 (0.09)
60		0.394 (0.11)	0.394 (0.12)	0.445 (0.11)	0.476 (0.11)	0.449 (0.11)	0.458 (0.12)	0.532 (0.09)	0.576 (0.09)	
600	20	0.617 (0.07)	0.661 (0.07)	0.639 (0.07)	0.707 (0.06)	0.700 (0.06)	0.773 (0.04)	0.675 (0.06)	0.798 (0.05)	
	40	0.606 (0.07)	0.643 (0.07)	0.631 (0.06)	0.688 (0.06)	0.691 (0.07)	0.753 (0.05)	0.685 (0.06)	0.792 (0.05)	
	60	0.612 (0.07)	0.639 (0.07)	0.634 (0.07)	0.678 (0.08)	0.692 (0.06)	0.748 (0.04)	0.693 (0.06)	0.776 (0.05)	
1000	20	0.672 (0.05)	0.707 (0.05)	0.686 (0.05)	0.743 (0.05)	0.743 (0.06)	0.811 (0.03)	0.708 (0.05)	0.820 (0.04)	
	40	0.657 (0.05)	0.693 (0.05)	0.674 (0.05)	0.719 (0.07)	0.744 (0.06)	0.797 (0.03)	0.726 (0.05)	0.826 (0.04)	
	60	0.648 (0.05)	0.687 (0.05)	0.662 (0.05)	0.714 (0.05)	0.724 (0.06)	0.794 (0.04)	0.710 (0.05)	0.815 (0.04)	

Table 9: Average accuracy and their standard deviations (in parenthesis) in the observational study in Case 4: Non-linear interaction functions with a complicated treatment-free effect function.

		$AD$	$AD_e$	$AD_s$	$AD_{e,s}$	$SABD$	$SABD_e$	$SABD_s$	$SABD_{e,s}$
n	p	Original							
	20	0.274 (0.07)	0.279 (0.08)	0.281 (0.08)	0.283 (0.09)	0.302 (0.08)	0.322 (0.11)	0.325 (0.09)	0.368 (0.13)
200	40	0.270 (0.07)	0.270 (0.07)	0.285 (0.07)	0.276 (0.07)	0.285 (0.07)	0.305 (0.09)	0.322 (0.09)	0.353 (0.13)
	60	0.266 (0.06)	0.270 (0.06)	0.281 (0.07)	0.272 (0.06)	0.274 (0.07)	0.281 (0.08)	0.313 (0.08)	0.323 (0.10)
600	20	0.294 (0.09)	0.309 (0.11)	0.304 (0.10)	0.338 (0.15)	0.409 (0.06)	0.555 (0.12)	0.422 (0.06)	0.619 (0.12)
	40	0.283 (0.08)	0.286 (0.09)	0.300 (0.09)	0.314 (0.12)	0.369 (0.07)	0.489 (0.13)	0.412 (0.07)	0.580 (0.14)
1000	60	0.278 (0.07)	0.291 (0.08)	0.292 (0.08)	0.308 (0.11)	0.371 (0.07)	0.457 (0.15)	0.408 (0.06)	0.583 (0.12)
	20	0.288 (0.09)	0.314 (0.12)	0.296 (0.09)	0.324 (0.12)	0.441 (0.04)	0.643 (0.10)	0.445 (0.05)	0.663 (0.10)
200	40	0.307 (0.09)	0.314 (0.11)	0.327 (0.11)	0.347 (0.13)	0.431 (0.04)	0.619 (0.09)	0.441 (0.04)	0.693 (0.10)
	60	0.289 (0.07)	0.312 (0.10)	0.306 (0.09)	0.344 (0.13)	0.421 (0.05)	0.603 (0.09)	0.445 (0.04)	0.668 (0.09)
Augmented									
200	20	0.317 (0.11)	0.336 (0.13)	0.335 (0.12)	0.384 (0.15)	0.439 (0.14)	0.456 (0.14)	0.505 (0.12)	0.571 (0.13)
	40	0.291 (0.09)	0.307 (0.10)	0.305 (0.10)	0.329 (0.11)	0.370 (0.13)	0.379 (0.12)	0.481 (0.12)	0.512 (0.14)
600	60	0.287 (0.07)	0.289 (0.07)	0.302 (0.09)	0.328 (0.11)	0.333 (0.10)	0.329 (0.10)	0.446 (0.13)	0.461 (0.13)
	20	0.337 (0.12)	0.361 (0.13)	0.346 (0.13)	0.390 (0.15)	0.660 (0.08)	0.702 (0.07)	0.676 (0.07)	0.761 (0.07)
1000	40	0.329 (0.12)	0.347 (0.13)	0.350 (0.13)	0.373 (0.15)	0.613 (0.10)	0.630 (0.12)	0.684 (0.07)	0.741 (0.06)
	60	0.307 (0.10)	0.322 (0.11)	0.326 (0.11)	0.347 (0.14)	0.612 (0.10)	0.635 (0.11)	0.679 (0.07)	0.726 (0.06)
200	20	0.318 (0.11)	0.342 (0.14)	0.325 (0.12)	0.369 (0.15)	0.734 (0.08)	0.769 (0.08)	0.730 (0.06)	0.801 (0.06)
	40	0.332 (0.12)	0.346 (0.14)	0.348 (0.13)	0.379 (0.15)	0.714 (0.09)	0.741 (0.07)	0.741 (0.06)	0.796 (0.05)
600	60	0.322 (0.11)	0.346 (0.13)	0.334 (0.12)	0.384 (0.15)	0.711 (0.07)	0.748 (0.06)	0.734 (0.06)	0.793 (0.04)

### I.3.2 Empirical values

In this subsection, we show the averages and standard deviations of empirical values from the additional simulation studies. The empirical values of the estimated rule  $\mathbb{E}[Y(\hat{d}(\mathbf{X}))]$  are computed by utilizing the generated potential outcomes on the test dataset.

We include the empirical value of the optimal rule  $\mathbb{E}[Y(d^{\text{opt}}(\mathbf{X}))]$  as well as the values obtained by assigning everyone to each of the four classes  $\mathbb{E}[Y(K)]$  for  $K = 1, \dots, 4$  as benchmarks, assuming the higher value of empirical value is better. Again, the first four tables present the results of experiments that mimic randomized trials, and the remaining four tables show the results of observational studies.

Here, we can observe the benefits of the three components that we found from the treatment decision accuracy results. In addition, the benchmarks of the empirical values reveal some intriguing insights. Firstly, the optimal empirical value provides information on how challenging the current scenario is. For instance, in the randomized trial of Case 1 (Table 9), the optimal empirical value is not significantly different from the mean empirical value evaluated by current methods. Especially with augmented outcomes using a sample size of 1000, it yields a value quite similar to the optimal. On the other hand, in the observational study of Case 4 (Table 16), the obtained empirical values show significant differences from the optimal value. Furthermore, the mean empirical value evaluated from AD-Learning with a sample size of 200 is not markedly different from the empirical value obtained by assigning all individuals to treatment 1. This implies that the estimated ITR provides results akin to a “one-size-fits-all” outcome in small sample sizes, and this finding is not irrelevant to the fact that a majority of AD-Learning results in an intercept-only model as the final decision function under conditions of highly non-linear covariate-outcome relationships. In such instances, we can explore alternative methods to further advance our current results such as other penalty types or performance metrics.

Table 10: Average empirical values and their standard deviations (in parenthesis) in the randomized trial in Case 1: Linear interaction functions with simple treatment-free effect functions. The optimal empirical value is 5.160 and the average empirical values when assigning one treatment are 1.751, 1.752, 1.751, and 1.752, respectively.

		<i>AD</i>	<i>AD<sub>e</sub></i>	<i>AD<sub>s</sub></i>	<i>AD<sub>e,s</sub></i>	<i>SABD</i>	<i>SABD<sub>e</sub></i>	<i>SABD<sub>s</sub></i>	<i>SABD<sub>e,s</sub></i>
n	p	Original							
200	20	4.554 (0.25)	4.725 (0.21)	4.637 (0.23)	4.934 (0.17)	4.701 (0.21)	4.769 (0.19)	4.739 (0.21)	4.925 (0.16)
	40	4.381 (0.31)	4.505 (0.34)	4.465 (0.30)	4.778 (0.28)	4.591 (0.24)	4.616 (0.24)	4.626 (0.27)	4.798 (0.26)
	60	4.364 (0.29)	4.409 (0.36)	4.412 (0.25)	4.670 (0.27)	4.528 (0.21)	4.566 (0.26)	4.578 (0.20)	4.746 (0.21)
600	20	4.963 (0.10)	5.053 (0.10)	4.985 (0.09)	5.117 (0.07)	4.992 (0.09)	5.050 (0.10)	4.956 (0.11)	5.063 (0.06)
	40	4.964 (0.10)	5.049 (0.06)	4.950 (0.09)	5.071 (0.19)	5.013 (0.08)	5.059 (0.06)	4.972 (0.09)	5.056 (0.12)
	60	4.935 (0.11)	5.025 (0.07)	4.891 (0.10)	5.059 (0.11)	5.006 (0.09)	5.045 (0.07)	4.962 (0.09)	5.054 (0.07)
1000	20	5.054 (0.07)	5.092 (0.11)	5.057 (0.07)	5.140 (0.07)	5.053 (0.07)	5.073 (0.08)	5.012 (0.09)	5.060 (0.07)
	40	5.034 (0.07)	5.098 (0.05)	5.020 (0.07)	5.101 (0.16)	5.051 (0.07)	5.090 (0.05)	5.000 (0.09)	5.060 (0.09)
	60	5.046 (0.06)	5.096 (0.04)	5.013 (0.06)	5.112 (0.09)	5.070 (0.06)	5.095 (0.04)	5.021 (0.08)	5.072 (0.07)
Augmented									
200	20	4.928 (0.10)	4.940 (0.10)	4.936 (0.15)	5.022 (0.14)	4.951 (0.09)	4.942 (0.11)	4.964 (0.15)	5.024 (0.14)
	40	4.887 (0.12)	4.873 (0.13)	4.857 (0.22)	4.940 (0.20)	4.888 (0.12)	4.863 (0.13)	4.884 (0.21)	4.936 (0.22)
	60	4.851 (0.12)	4.833 (0.14)	4.817 (0.17)	4.888 (0.16)	4.860 (0.13)	4.826 (0.15)	4.865 (0.16)	4.905 (0.16)
600	20	5.092 (0.04)	5.109 (0.04)	5.100 (0.04)	5.136 (0.04)	5.101 (0.04)	5.114 (0.04)	5.097 (0.04)	5.133 (0.04)
	40	5.095 (0.05)	5.101 (0.04)	5.096 (0.04)	5.124 (0.06)	5.105 (0.04)	5.105 (0.04)	5.102 (0.04)	5.127 (0.05)
	60	5.086 (0.05)	5.091 (0.05)	5.073 (0.05)	5.114 (0.05)	5.097 (0.05)	5.096 (0.05)	5.089 (0.05)	5.120 (0.04)
1000	20	5.125 (0.04)	5.127 (0.05)	5.127 (0.04)	5.148 (0.04)	5.130 (0.04)	5.131 (0.05)	5.125 (0.04)	5.143 (0.04)
	40	5.121 (0.04)	5.129 (0.04)	5.118 (0.04)	5.140 (0.04)	5.126 (0.04)	5.132 (0.04)	5.120 (0.04)	5.139 (0.04)
	60	5.121 (0.04)	5.128 (0.04)	5.115 (0.04)	5.140 (0.04)	5.127 (0.04)	5.131 (0.04)	5.120 (0.04)	5.141 (0.04)

Table 11: Average empirical values and their standard deviations (in parenthesis) in the randomized trial in Case 2: Linear interaction functions with complicated treatment-free effect functions. The optimal empirical value is 10.333 and the average empirical values when assigning one treatment are 6.923, 6.924, 6.923, and 6.924, respectively.

		<i>AD</i>	<i>AD<sub>e</sub></i>	<i>AD<sub>s</sub></i>	<i>AD<sub>e,s</sub></i>	<i>SABD</i>	<i>SABD<sub>e</sub></i>	<i>SABD<sub>s</sub></i>	<i>SABD<sub>e,s</sub></i>
n	p	Original							
200	20	7.704	7.474	7.820	7.764	8.360	8.073	8.677	8.736
		(0.90)	(1.00)	(0.98)	(1.14)	(0.96)	(1.16)	(0.94)	(1.26)
	40	7.430	7.262	7.641	7.482	7.907	7.529	8.415	8.339
		(0.80)	(0.77)	(0.94)	(0.98)	(0.93)	(0.92)	(1.08)	(1.19)
	60	7.413	7.205	7.605	7.359	7.690	7.470	8.475	8.275
		(0.79)	(0.74)	(0.93)	(0.94)	(0.91)	(0.83)	(0.94)	(1.23)
600	20	8.334	8.437	8.488	8.719	9.791	9.966	9.803	10.023
		(1.11)	(1.27)	(1.18)	(1.29)	(0.56)	(0.76)	(0.56)	(0.63)
	40	7.983	8.051	8.079	8.250	9.610	9.821	9.757	9.991
		(1.11)	(1.25)	(1.16)	(1.33)	(0.69)	(0.73)	(0.64)	(0.69)
	60	7.834	7.788	7.956	8.025	9.311	9.648	9.657	9.930
		(1.06)	(1.12)	(1.14)	(1.26)	(0.85)	(0.80)	(0.66)	(0.59)
1000	20	8.395	8.458	8.486	8.706	9.997	10.087	9.929	10.101
		(1.15)	(1.20)	(1.19)	(1.27)	(0.47)	(0.54)	(0.50)	(0.50)
	40	8.033	7.993	8.199	8.400	9.874	10.077	9.893	10.101
		(1.20)	(1.24)	(1.25)	(1.32)	(0.59)	(0.51)	(0.58)	(0.47)
	60	8.275	8.403	8.439	8.506	9.933	10.122	10.021	10.053
		(1.17)	(1.24)	(1.21)	(1.32)	(0.54)	(0.59)	(0.44)	(0.65)
Augmented									
200	20	8.111	8.117	8.201	8.429	9.552	9.526	9.801	9.877
		(1.19)	(1.16)	(1.24)	(1.28)	(0.64)	(0.65)	(0.53)	(0.49)
	40	7.758	7.712	7.859	8.038	9.017	8.912	9.646	9.678
		(1.09)	(1.09)	(1.17)	(1.28)	(0.97)	(1.02)	(0.62)	(0.65)
	60	7.694	7.649	7.834	7.939	8.814	8.764	9.519	9.589
		(1.03)	(1.01)	(1.11)	(1.15)	(0.96)	(0.95)	(0.70)	(0.67)
600	20	8.551	8.713	8.663	9.028	10.236	10.244	10.265	10.278
		(1.25)	(1.23)	(1.28)	(1.21)	(0.36)	(0.35)	(0.35)	(0.39)
	40	8.212	8.237	8.317	8.437	10.173	10.176	10.234	10.249
		(1.25)	(1.27)	(1.27)	(1.34)	(0.37)	(0.37)	(0.36)	(0.36)
	60	7.972	7.998	8.048	8.249	10.100	10.080	10.201	10.230
		(1.16)	(1.16)	(1.20)	(1.27)	(0.38)	(0.44)	(0.34)	(0.33)
1000	20	8.509	8.549	8.587	8.769	10.262	10.255	10.271	10.280
		(1.22)	(1.22)	(1.24)	(1.26)	(0.41)	(0.41)	(0.41)	(0.41)
	40	8.158	8.154	8.265	8.543	10.230	10.232	10.252	10.266
		(1.26)	(1.27)	(1.31)	(1.35)	(0.38)	(0.38)	(0.38)	(0.38)
	60	8.414	8.510	8.532	8.648	10.262	10.260	10.281	10.287
		(1.26)	(1.25)	(1.28)	(1.31)	(0.40)	(0.40)	(0.40)	(0.40)

Table 12: Average empirical values and their standard deviations (in parenthesis) in the randomized trial in Case 3: Non-linear interaction functions with simple treatment-free effect functions. The optimal empirical value is 1.422 and the average empirical values when assigning one treatment are 0.041, -0.919, -0.419, and -0.920, respectively.

n	p	<i>AD</i>	<i>AD<sub>e</sub></i>	<i>AD<sub>s</sub></i>	<i>AD<sub>e,s</sub></i>	<i>SABD</i>	<i>SABD<sub>e</sub></i>	<i>SABD<sub>s</sub></i>	<i>SABD<sub>e,s</sub></i>	
		Original								
200	20	0.210 (0.45)	0.157 (0.48)	0.294 (0.44)	0.382 (0.45)	0.434 (0.47)	0.275 (0.55)	0.546 (0.46)	0.592 (0.59)	
	40	0.023 (0.46)	-0.018 (0.46)	0.131 (0.44)	0.185 (0.45)	0.196 (0.50)	0.044 (0.54)	0.436 (0.38)	0.462 (0.50)	
	60	-0.051 (0.49)	-0.087 (0.45)	0.072 (0.45)	0.100 (0.45)	0.081 (0.51)	-0.008 (0.48)	0.346 (0.43)	0.315 (0.48)	
600	20	0.733 (0.29)	0.826 (0.26)	0.788 (0.27)	0.983 (0.21)	0.963 (0.18)	1.023 (0.17)	0.931 (0.23)	1.095 (0.17)	
	40	0.631 (0.33)	0.646 (0.32)	0.715 (0.22)	0.857 (0.29)	0.910 (0.21)	0.915 (0.24)	0.923 (0.18)	1.025 (0.21)	
	60	0.609 (0.28)	0.585 (0.32)	0.651 (0.22)	0.770 (0.27)	0.905 (0.19)	0.922 (0.18)	0.903 (0.17)	0.993 (0.17)	
1000	20	0.879 (0.22)	0.930 (0.22)	0.920 (0.19)	1.111 (0.12)	1.089 (0.15)	1.104 (0.14)	1.009 (0.18)	1.123 (0.14)	
	40	0.864 (0.23)	0.880 (0.21)	0.887 (0.21)	1.061 (0.14)	1.027 (0.15)	1.020 (0.25)	0.993 (0.16)	1.117 (0.14)	
	60	0.870 (0.20)	0.890 (0.21)	0.878 (0.16)	1.002 (0.24)	1.050 (0.14)	1.064 (0.14)	1.030 (0.14)	1.102 (0.16)	
Augmented										
200	20	0.715 (0.32)	0.697 (0.38)	0.794 (0.24)	0.918 (0.24)	0.883 (0.19)	0.871 (0.26)	0.967 (0.17)	1.056 (0.18)	
	40	0.588 (0.30)	0.545 (0.36)	0.680 (0.27)	0.779 (0.25)	0.791 (0.26)	0.724 (0.30)	0.884 (0.20)	0.926 (0.22)	
	60	0.481 (0.35)	0.444 (0.38)	0.578 (0.32)	0.653 (0.35)	0.661 (0.32)	0.587 (0.42)	0.849 (0.19)	0.840 (0.23)	
600	20	1.082 (0.11)	1.093 (0.12)	1.100 (0.10)	1.151 (0.10)	1.184 (0.10)	1.194 (0.10)	1.216 (0.07)	1.245 (0.08)	
	40	1.051 (0.12)	1.065 (0.11)	1.063 (0.11)	1.118 (0.12)	1.166 (0.11)	1.163 (0.10)	1.189 (0.10)	1.220 (0.09)	
	60	1.039 (0.12)	1.042 (0.13)	1.051 (0.10)	1.092 (0.12)	1.161 (0.10)	1.156 (0.10)	1.173 (0.09)	1.204 (0.09)	
1000	20	1.141 (0.08)	1.134 (0.15)	1.156 (0.08)	1.196 (0.07)	1.219 (0.08)	1.210 (0.12)	1.229 (0.08)	1.260 (0.07)	
	40	1.133 (0.09)	1.137 (0.09)	1.143 (0.09)	1.183 (0.07)	1.207 (0.07)	1.213 (0.08)	1.231 (0.07)	1.258 (0.06)	
	60	1.141 (0.09)	1.146 (0.08)	1.148 (0.08)	1.177 (0.08)	1.212 (0.09)	1.207 (0.08)	1.235 (0.08)	1.250 (0.08)	

Table 13: Average empirical values and their standard deviations (in parenthesis) in the randomized trial in Case 4: Non-linear interaction functions with a complicated treatment-free effect function. The optimal empirical value is 5.623 and the average empirical values when assigning one treatment are 4.243, 3.282, 3.782, and 3.281 respectively.

		<i>AD</i>	<i>AD<sub>e</sub></i>	<i>AD<sub>s</sub></i>	<i>AD<sub>e,s</sub></i>	<i>SABD</i>	<i>SABD<sub>e</sub></i>	<i>SABD<sub>s</sub></i>	<i>SABD<sub>e,s</sub></i>
n	p	Original							
200	20	3.979 (0.55)	4.033 (0.53)	4.058 (0.57)	4.052 (0.58)	4.192 (0.58)	4.132 (0.61)	4.381 (0.58)	4.366 (0.65)
	40	3.890 (0.55)	3.876 (0.54)	3.920 (0.57)	3.913 (0.59)	3.984 (0.65)	3.936 (0.61)	4.184 (0.63)	4.168 (0.65)
	60	3.939 (0.57)	3.931 (0.56)	3.991 (0.58)	3.970 (0.59)	4.018 (0.60)	3.938 (0.59)	4.241 (0.63)	4.137 (0.70)
600	20	4.134 (0.60)	4.198 (0.64)	4.202 (0.64)	4.282 (0.62)	4.995 (0.49)	4.980 (0.52)	5.009 (0.46)	5.127 (0.48)
	40	4.032 (0.64)	4.066 (0.64)	4.111 (0.68)	4.183 (0.66)	4.761 (0.57)	4.774 (0.60)	4.951 (0.48)	5.056 (0.49)
	60	3.989 (0.64)	4.021 (0.62)	4.029 (0.66)	4.042 (0.67)	4.667 (0.61)	4.685 (0.62)	4.885 (0.49)	4.964 (0.60)
1000	20	4.117 (0.66)	4.130 (0.67)	4.152 (0.68)	4.218 (0.71)	5.138 (0.43)	5.168 (0.44)	5.083 (0.46)	5.191 (0.42)
	40	4.047 (0.62)	4.061 (0.64)	4.093 (0.66)	4.188 (0.71)	5.075 (0.49)	5.078 (0.51)	5.070 (0.48)	5.188 (0.43)
	60	4.103 (0.64)	4.190 (0.67)	4.154 (0.65)	4.232 (0.69)	5.144 (0.45)	5.152 (0.49)	5.155 (0.43)	5.237 (0.48)
Augmented									
200	20	4.203 (0.58)	4.247 (0.58)	4.225 (0.60)	4.333 (0.61)	4.733 (0.58)	4.648 (0.66)	4.923 (0.52)	5.016 (0.55)
	40	3.962 (0.60)	3.962 (0.62)	4.010 (0.61)	4.029 (0.65)	4.332 (0.64)	4.288 (0.64)	4.745 (0.54)	4.700 (0.59)
	60	4.011 (0.64)	3.987 (0.61)	4.050 (0.67)	4.095 (0.64)	4.350 (0.61)	4.318 (0.55)	4.707 (0.56)	4.643 (0.63)
600	20	4.245 (0.63)	4.293 (0.68)	4.291 (0.65)	4.425 (0.68)	5.307 (0.38)	5.302 (0.39)	5.364 (0.36)	5.375 (0.39)
	40	4.199 (0.69)	4.262 (0.67)	4.227 (0.71)	4.328 (0.73)	5.202 (0.42)	5.160 (0.39)	5.320 (0.37)	5.334 (0.36)
	60	4.073 (0.71)	4.084 (0.71)	4.140 (0.74)	4.166 (0.74)	5.175 (0.41)	5.161 (0.39)	5.313 (0.34)	5.327 (0.35)
1000	20	4.147 (0.71)	4.199 (0.70)	4.175 (0.72)	4.331 (0.75)	5.356 (0.39)	5.355 (0.41)	5.372 (0.42)	5.393 (0.42)
	40	4.124 (0.69)	4.141 (0.71)	4.168 (0.72)	4.214 (0.77)	5.331 (0.38)	5.310 (0.40)	5.369 (0.38)	5.371 (0.40)
	60	4.183 (0.69)	4.264 (0.69)	4.213 (0.73)	4.298 (0.71)	5.377 (0.40)	5.362 (0.41)	5.418 (0.40)	5.433 (0.40)



Table 14: Average empirical values and their standard deviations (in parenthesis) in the observational study in Case 1: Linear interaction functions with simple treatment-free effect functions. The optimal empirical value is 4.263 and the average empirical values when assigning one treatment are 1.503, 2.003, 2.502, and 2.001, respectively.

		<i>AD</i>	<i>AD<sub>e</sub></i>	<i>AD<sub>s</sub></i>	<i>AD<sub>e,s</sub></i>	<i>SABD</i>	<i>SABD<sub>e</sub></i>	<i>SABD<sub>s</sub></i>	<i>SABD<sub>e,s</sub></i>
n	p	Original							
200	20	3.511 (0.41)	3.621 (0.46)	3.581 (0.36)	3.790 (0.36)	3.553 (0.29)	3.730 (0.33)	3.566 (0.32)	3.879 (0.30)
	40	3.258 (0.51)	3.223 (0.62)	3.415 (0.43)	3.558 (0.51)	3.430 (0.30)	3.450 (0.50)	3.503 (0.29)	3.704 (0.38)
	60	3.289 (0.47)	3.210 (0.58)	3.382 (0.38)	3.412 (0.53)	3.304 (0.44)	3.396 (0.51)	3.391 (0.38)	3.564 (0.47)
600	20	3.992 (0.13)	4.079 (0.11)	4.009 (0.12)	4.165 (0.09)	3.881 (0.15)	4.143 (0.07)	3.851 (0.14)	4.188 (0.08)
	40	3.970 (0.15)	4.027 (0.23)	3.982 (0.14)	4.135 (0.12)	3.888 (0.17)	4.115 (0.12)	3.874 (0.15)	4.182 (0.08)
	60	3.942 (0.16)	4.037 (0.12)	3.918 (0.15)	4.094 (0.13)	3.848 (0.18)	4.079 (0.11)	3.826 (0.16)	4.147 (0.08)
1000	20	4.086 (0.09)	4.166 (0.08)	4.094 (0.09)	4.211 (0.07)	3.942 (0.12)	4.200 (0.05)	3.893 (0.12)	4.224 (0.06)
	40	4.073 (0.11)	4.150 (0.08)	4.061 (0.11)	4.205 (0.07)	3.931 (0.12)	4.183 (0.06)	3.892 (0.12)	4.219 (0.06)
	60	4.069 (0.10)	4.150 (0.07)	4.050 (0.10)	4.190 (0.11)	3.941 (0.14)	4.177 (0.06)	3.900 (0.13)	4.212 (0.07)
Augmented									
200	20	3.913 (0.17)	3.922 (0.18)	3.894 (0.18)	3.970 (0.17)	3.964 (0.16)	3.969 (0.17)	3.923 (0.20)	3.993 (0.18)
	40	3.828 (0.22)	3.827 (0.23)	3.824 (0.21)	3.879 (0.25)	3.877 (0.17)	3.874 (0.20)	3.869 (0.21)	3.927 (0.21)
	60	3.806 (0.22)	3.739 (0.35)	3.737 (0.27)	3.789 (0.24)	3.847 (0.20)	3.817 (0.24)	3.793 (0.31)	3.831 (0.31)
600	20	4.169 (0.06)	4.188 (0.05)	4.168 (0.06)	4.210 (0.06)	4.172 (0.06)	4.207 (0.05)	4.165 (0.06)	4.222 (0.05)
	40	4.158 (0.07)	4.169 (0.06)	4.156 (0.07)	4.200 (0.06)	4.162 (0.06)	4.201 (0.05)	4.162 (0.06)	4.224 (0.06)
	60	4.144 (0.08)	4.153 (0.08)	4.135 (0.07)	4.175 (0.08)	4.154 (0.07)	4.186 (0.06)	4.146 (0.07)	4.201 (0.07)
1000	20	4.208 (0.05)	4.223 (0.05)	4.207 (0.05)	4.239 (0.05)	4.197 (0.05)	4.235 (0.04)	4.191 (0.05)	4.242 (0.04)
	40	4.205 (0.06)	4.215 (0.05)	4.201 (0.06)	4.232 (0.05)	4.196 (0.06)	4.228 (0.05)	4.188 (0.06)	4.238 (0.05)
	60	4.204 (0.05)	4.214 (0.05)	4.198 (0.05)	4.230 (0.05)	4.194 (0.05)	4.232 (0.04)	4.187 (0.06)	4.240 (0.04)

Table 15: Average empirical values and their standard deviations (in parenthesis) in the observational study in Case 2: Linear interaction functions with complicated treatment-free effect functions. The optimal empirical value is 9.433 and the average empirical values when assigning one treatment are 6.672, 7.173, 7.672, and 7.171, respectively.

		<i>AD</i>	<i>AD<sub>e</sub></i>	<i>AD<sub>s</sub></i>	<i>AD<sub>e,s</sub></i>	<i>SABD</i>	<i>SABD<sub>e</sub></i>	<i>SABD<sub>s</sub></i>	<i>SABD<sub>e,s</sub></i>
<i>n</i>	<i>p</i>	Original							
200	20	7.592 (0.73)	7.598 (0.74)	7.608 (0.75)	7.696 (0.78)	7.561 (0.67)	7.736 (0.80)	7.715 (0.72)	7.927 (0.87)
	40	7.462 (0.60)	7.452 (0.66)	7.540 (0.67)	7.554 (0.67)	7.462 (0.64)	7.497 (0.58)	7.647 (0.68)	7.671 (0.69)
	60	7.503 (0.58)	7.531 (0.61)	7.568 (0.58)	7.538 (0.65)	7.380 (0.68)	7.612 (0.66)	7.645 (0.63)	7.815 (0.76)
600	20	7.798 (0.70)	7.824 (0.79)	7.840 (0.72)	7.970 (0.84)	8.147 (0.46)	8.948 (0.58)	8.197 (0.45)	9.024 (0.61)
	40	7.732 (0.70)	7.778 (0.79)	7.787 (0.75)	7.841 (0.77)	7.954 (0.54)	8.610 (0.74)	8.142 (0.47)	8.923 (0.58)
	60	7.615 (0.66)	7.599 (0.71)	7.625 (0.68)	7.666 (0.73)	7.953 (0.59)	8.413 (0.85)	8.125 (0.50)	8.875 (0.70)
1000	20	7.712 (0.77)	7.833 (0.81)	7.731 (0.79)	7.823 (0.87)	8.191 (0.47)	9.122 (0.59)	8.166 (0.48)	9.084 (0.60)
	40	7.748 (0.75)	7.787 (0.81)	7.797 (0.74)	7.958 (0.83)	8.163 (0.42)	9.047 (0.51)	8.206 (0.43)	9.180 (0.43)
	60	7.873 (0.72)	7.967 (0.84)	7.931 (0.74)	8.113 (0.88)	8.151 (0.49)	9.143 (0.55)	8.248 (0.47)	9.211 (0.51)
Augmented									
200	20	7.770 (0.79)	7.887 (0.81)	7.838 (0.82)	8.042 (0.90)	8.667 (0.68)	8.711 (0.67)	8.859 (0.58)	8.936 (0.58)
	40	7.621 (0.78)	7.713 (0.77)	7.680 (0.81)	7.854 (0.88)	8.240 (0.79)	8.308 (0.78)	8.703 (0.65)	8.802 (0.68)
	60	7.585 (0.65)	7.693 (0.69)	7.629 (0.65)	7.812 (0.72)	8.169 (0.68)	8.146 (0.66)	8.647 (0.65)	8.680 (0.65)
600	20	7.920 (0.81)	8.102 (0.84)	7.956 (0.85)	8.263 (0.91)	9.315 (0.35)	9.340 (0.34)	9.343 (0.34)	9.380 (0.33)
	40	7.792 (0.80)	7.962 (0.85)	7.867 (0.80)	8.103 (0.86)	9.194 (0.45)	9.211 (0.41)	9.311 (0.34)	9.342 (0.34)
	60	7.630 (0.68)	7.743 (0.74)	7.676 (0.74)	7.845 (0.84)	9.153 (0.43)	9.159 (0.45)	9.277 (0.34)	9.308 (0.34)
1000	20	7.756 (0.87)	7.949 (0.91)	7.781 (0.89)	8.043 (0.92)	9.318 (0.46)	9.341 (0.41)	9.347 (0.41)	9.371 (0.41)
	40	7.847 (0.87)	7.997 (0.85)	7.885 (0.89)	8.034 (0.87)	9.292 (0.41)	9.318 (0.38)	9.342 (0.39)	9.359 (0.38)
	60	7.948 (0.80)	8.076 (0.85)	8.006 (0.84)	8.246 (0.85)	9.343 (0.41)	9.346 (0.42)	9.373 (0.40)	9.393 (0.41)

Table 16: Average empirical values and their standard deviations (in parenthesis) in the observational study in Case 3: Non-linear interaction functions with simple treatment-free effect functions. The optimal empirical value is 1.422 and the average empirical values when assigning one treatment are 0.042, -0.920, -0.419, and -0.920, respectively.

		<i>AD</i>	<i>AD<sub>e</sub></i>	<i>AD<sub>s</sub></i>	<i>AD<sub>e,s</sub></i>	<i>SABD</i>	<i>SABD<sub>e</sub></i>	<i>SABD<sub>s</sub></i>	<i>SABD<sub>e,s</sub></i>
<i>n</i>	<i>p</i>	Original							
200	20	0.029 (0.36)	-0.014 (0.42)	0.113 (0.41)	0.142 (0.46)	0.312 (0.35)	0.178 (0.49)	0.405 (0.38)	0.411 (0.58)
	40	0.091 (0.33)	0.058 (0.38)	0.244 (0.28)	0.151 (0.42)	0.258 (0.40)	0.149 (0.53)	0.415 (0.36)	0.335 (0.59)
	60	0.014 (0.32)	0.010 (0.32)	0.110 (0.37)	0.129 (0.38)	0.168 (0.37)	0.031 (0.45)	0.349 (0.34)	0.224 (0.56)
600	20	0.482 (0.29)	0.660 (0.36)	0.607 (0.27)	0.910 (0.28)	0.773 (0.16)	1.008 (0.14)	0.721 (0.15)	1.095 (0.16)
	40	0.420 (0.25)	0.592 (0.28)	0.567 (0.23)	0.802 (0.29)	0.717 (0.18)	0.927 (0.18)	0.720 (0.16)	1.076 (0.17)
	60	0.437 (0.27)	0.546 (0.28)	0.612 (0.22)	0.796 (0.29)	0.718 (0.20)	0.907 (0.20)	0.737 (0.15)	1.009 (0.19)
1000	20	0.662 (0.22)	0.883 (0.25)	0.772 (0.19)	1.077 (0.17)	0.832 (0.15)	1.072 (0.13)	0.744 (0.13)	1.132 (0.14)
	40	0.577 (0.21)	0.805 (0.22)	0.721 (0.17)	1.008 (0.17)	0.822 (0.14)	1.070 (0.12)	0.764 (0.12)	1.104 (0.20)
	60	0.533 (0.21)	0.795 (0.21)	0.706 (0.16)	0.964 (0.21)	0.812 (0.13)	1.065 (0.13)	0.781 (0.12)	1.108 (0.15)
Augmented									
200	20	0.521 (0.35)	0.559 (0.36)	0.647 (0.30)	0.783 (0.35)	0.766 (0.26)	0.759 (0.36)	0.862 (0.19)	1.018 (0.19)
	40	0.559 (0.31)	0.544 (0.38)	0.684 (0.25)	0.763 (0.32)	0.715 (0.27)	0.713 (0.30)	0.830 (0.20)	0.922 (0.24)
	60	0.353 (0.37)	0.342 (0.38)	0.475 (0.35)	0.558 (0.35)	0.527 (0.35)	0.474 (0.42)	0.745 (0.24)	0.798 (0.26)
600	20	0.987 (0.15)	1.071 (0.12)	1.025 (0.14)	1.146 (0.10)	1.118 (0.12)	1.188 (0.09)	1.069 (0.12)	1.246 (0.07)
	40	0.960 (0.14)	1.029 (0.14)	1.006 (0.12)	1.107 (0.11)	1.092 (0.13)	1.157 (0.10)	1.086 (0.13)	1.225 (0.09)
	60	0.980 (0.14)	1.030 (0.14)	1.019 (0.13)	1.094 (0.14)	1.106 (0.11)	1.149 (0.09)	1.107 (0.11)	1.198 (0.10)
1000	20	1.090 (0.10)	1.144 (0.11)	1.111 (0.09)	1.191 (0.08)	1.179 (0.10)	1.228 (0.07)	1.124 (0.11)	1.266 (0.07)
	40	1.061 (0.10)	1.121 (0.09)	1.088 (0.09)	1.152 (0.14)	1.176 (0.10)	1.201 (0.09)	1.150 (0.10)	1.260 (0.08)
	60	1.046 (0.11)	1.122 (0.09)	1.072 (0.10)	1.159 (0.08)	1.151 (0.11)	1.212 (0.08)	1.133 (0.10)	1.257 (0.07)

Table 17: Average empirical values and their standard deviations (in parenthesis) in the observational study in Case 4: Non-linear interaction functions with a complicated treatment-free effect function. The optimal empirical value is 5.623 and the average empirical values when assigning one treatment are 4.243, 3.282, 3.782, and 3.281, respectively.

		<i>AD</i>	<i>AD<sub>e</sub></i>	<i>AD<sub>s</sub></i>	<i>AD<sub>e,s</sub></i>	<i>SABD</i>	<i>SABD<sub>e</sub></i>	<i>SABD<sub>s</sub></i>	<i>SABD<sub>e,s</sub></i>
n	p	Original							
200	20	4.034 (0.54)	3.954 (0.59)	4.062 (0.56)	4.006 (0.57)	4.309 (0.55)	4.048 (0.70)	4.381 (0.53)	4.316 (0.71)
	40	3.978 (0.58)	3.883 (0.62)	4.001 (0.59)	3.908 (0.62)	4.219 (0.52)	4.023 (0.64)	4.372 (0.51)	4.184 (0.72)
	60	4.019 (0.50)	3.926 (0.53)	4.038 (0.52)	3.879 (0.53)	4.198 (0.50)	4.007 (0.59)	4.287 (0.54)	4.056 (0.65)
600	20	4.163 (0.53)	4.103 (0.61)	4.191 (0.56)	4.215 (0.69)	4.723 (0.35)	4.992 (0.48)	4.745 (0.37)	5.132 (0.45)
	40	4.096 (0.58)	4.060 (0.60)	4.131 (0.62)	4.132 (0.69)	4.606 (0.41)	4.785 (0.54)	4.703 (0.41)	5.020 (0.48)
	60	4.048 (0.58)	3.980 (0.64)	4.074 (0.60)	4.051 (0.69)	4.619 (0.39)	4.699 (0.62)	4.707 (0.38)	5.042 (0.48)
1000	20	4.047 (0.63)	4.086 (0.68)	4.071 (0.63)	4.089 (0.69)	4.762 (0.41)	5.150 (0.48)	4.764 (0.41)	5.180 (0.47)
	40	4.130 (0.62)	4.103 (0.70)	4.186 (0.66)	4.158 (0.77)	4.727 (0.38)	5.091 (0.44)	4.743 (0.38)	5.236 (0.44)
	60	4.093 (0.62)	4.029 (0.66)	4.128 (0.64)	4.168 (0.71)	4.742 (0.42)	5.117 (0.47)	4.792 (0.40)	5.239 (0.44)
Augmented									
200	20	4.146 (0.59)	4.125 (0.68)	4.193 (0.62)	4.337 (0.72)	4.645 (0.60)	4.576 (0.69)	4.859 (0.50)	4.973 (0.51)
	40	3.975 (0.66)	4.014 (0.68)	4.003 (0.67)	4.099 (0.69)	4.441 (0.56)	4.344 (0.64)	4.795 (0.49)	4.765 (0.56)
	60	3.976 (0.57)	3.957 (0.58)	4.019 (0.60)	4.056 (0.64)	4.276 (0.58)	4.160 (0.59)	4.657 (0.55)	4.577 (0.65)
600	20	4.188 (0.63)	4.267 (0.64)	4.231 (0.64)	4.380 (0.69)	5.266 (0.37)	5.291 (0.36)	5.293 (0.36)	5.396 (0.35)
	40	4.141 (0.70)	4.216 (0.70)	4.212 (0.74)	4.331 (0.76)	5.147 (0.47)	5.106 (0.49)	5.298 (0.37)	5.350 (0.33)
	60	4.061 (0.66)	4.063 (0.71)	4.124 (0.69)	4.186 (0.73)	5.135 (0.47)	5.124 (0.48)	5.297 (0.37)	5.333 (0.37)
1000	20	4.043 (0.69)	4.154 (0.74)	4.072 (0.72)	4.207 (0.76)	5.341 (0.46)	5.346 (0.47)	5.346 (0.41)	5.405 (0.42)
	40	4.149 (0.71)	4.198 (0.72)	4.204 (0.71)	4.302 (0.74)	5.307 (0.40)	5.310 (0.38)	5.352 (0.39)	5.384 (0.37)
	60	4.086 (0.66)	4.131 (0.71)	4.111 (0.68)	4.290 (0.72)	5.349 (0.41)	5.371 (0.41)	5.388 (0.40)	5.433 (0.41)

## NeurIPS Paper Checklist

### 1. Claims

Question: Do the main claims made in the abstract and introduction accurately reflect the paper's contributions and scope?

Answer: [Yes]

Justification: Our main claims made in the abstract and introduction accurately reflect the paper's contributions and scope.

Guidelines:

- The answer NA means that the abstract and introduction do not include the claims made in the paper.
- The abstract and/or introduction should clearly state the claims made, including the contributions made in the paper and important assumptions and limitations. A No or NA answer to this question will not be perceived well by the reviewers.
- The claims made should match theoretical and experimental results, and reflect how much the results can be expected to generalize to other settings.
- It is fine to include aspirational goals as motivation as long as it is clear that these goals are not attained by the paper.

### 2. Limitations

Question: Does the paper discuss the limitations of the work performed by the authors?

Answer: [Yes]

Justification: Our paper discusses the limitations of the work in the section 6.

Guidelines:

- The answer NA means that the paper has no limitation while the answer No means that the paper has limitations, but those are not discussed in the paper.
- The authors are encouraged to create a separate "Limitations" section in their paper.
- The paper should point out any strong assumptions and how robust the results are to violations of these assumptions (e.g., independence assumptions, noiseless settings, model well-specification, asymptotic approximations only holding locally). The authors should reflect on how these assumptions might be violated in practice and what the implications would be.
- The authors should reflect on the scope of the claims made, e.g., if the approach was only tested on a few datasets or with a few runs. In general, empirical results often depend on implicit assumptions, which should be articulated.
- The authors should reflect on the factors that influence the performance of the approach. For example, a facial recognition algorithm may perform poorly when image resolution is low or images are taken in low lighting. Or a speech-to-text system might not be used reliably to provide closed captions for online lectures because it fails to handle technical jargon.
- The authors should discuss the computational efficiency of the proposed algorithms and how they scale with dataset size.
- If applicable, the authors should discuss possible limitations of their approach to address problems of privacy and fairness.
- While the authors might fear that complete honesty about limitations might be used by reviewers as grounds for rejection, a worse outcome might be that reviewers discover limitations that aren't acknowledged in the paper. The authors should use their best judgment and recognize that individual actions in favor of transparency play an important role in developing norms that preserve the integrity of the community. Reviewers will be specifically instructed to not penalize honesty concerning limitations.

### 3. Theory Assumptions and Proofs

Question: For each theoretical result, does the paper provide the full set of assumptions and a complete (and correct) proof?

Answer: [Yes]

Justification: We provide the full set of assumptions in the main text and complete proof in the appendix.

Guidelines:

- The answer NA means that the paper does not include theoretical results.
- All the theorems, formulas, and proofs in the paper should be numbered and cross-referenced.
- All assumptions should be clearly stated or referenced in the statement of any theorems.
- The proofs can either appear in the main paper or the supplemental material, but if they appear in the supplemental material, the authors are encouraged to provide a short proof sketch to provide intuition.
- Inversely, any informal proof provided in the core of the paper should be complemented by formal proofs provided in appendix or supplemental material.
- Theorems and Lemmas that the proof relies upon should be properly referenced.

#### 4. Experimental Result Reproducibility

Question: Does the paper fully disclose all the information needed to reproduce the main experimental results of the paper to the extent that it affects the main claims and/or conclusions of the paper (regardless of whether the code and data are provided or not)?

Answer: [Yes]

Justification: Our paper fully discloses all the information needed to reproduce all simulation results and data applications of the paper.

Guidelines:

- The answer NA means that the paper does not include experiments.
- If the paper includes experiments, a No answer to this question will not be perceived well by the reviewers: Making the paper reproducible is important, regardless of whether the code and data are provided or not.
- If the contribution is a dataset and/or model, the authors should describe the steps taken to make their results reproducible or verifiable.
- Depending on the contribution, reproducibility can be accomplished in various ways. For example, if the contribution is a novel architecture, describing the architecture fully might suffice, or if the contribution is a specific model and empirical evaluation, it may be necessary to either make it possible for others to replicate the model with the same dataset, or provide access to the model. In general, releasing code and data is often one good way to accomplish this, but reproducibility can also be provided via detailed instructions for how to replicate the results, access to a hosted model (e.g., in the case of a large language model), releasing of a model checkpoint, or other means that are appropriate to the research performed.
- While NeurIPS does not require releasing code, the conference does require all submissions to provide some reasonable avenue for reproducibility, which may depend on the nature of the contribution. For example
  - (a) If the contribution is primarily a new algorithm, the paper should make it clear how to reproduce that algorithm.
  - (b) If the contribution is primarily a new model architecture, the paper should describe the architecture clearly and fully.
  - (c) If the contribution is a new model (e.g., a large language model), then there should either be a way to access this model for reproducing the results or a way to reproduce the model (e.g., with an open-source dataset or instructions for how to construct the dataset).
  - (d) We recognize that reproducibility may be tricky in some cases, in which case authors are welcome to describe the particular way they provide for reproducibility. In the case of closed-source models, it may be that access to the model is limited in some way (e.g., to registered users), but it should be possible for other researchers to have some path to reproducing or verifying the results.

#### 5. Open access to data and code

Question: Does the paper provide open access to the data and code, with sufficient instructions to faithfully reproduce the main experimental results, as described in supplemental material?

Answer: [Yes]

Justification: We provide open access to the code, with sufficient instructions to faithfully reproduce the main experimental results.

Guidelines:

- The answer NA means that paper does not include experiments requiring code.
- Please see the NeurIPS code and data submission guidelines (<https://nips.cc/public/guides/CodeSubmissionPolicy>) for more details.
- While we encourage the release of code and data, we understand that this might not be possible, so “No” is an acceptable answer. Papers cannot be rejected simply for not including code, unless this is central to the contribution (e.g., for a new open-source benchmark).
- The instructions should contain the exact command and environment needed to run to reproduce the results. See the NeurIPS code and data submission guidelines (<https://nips.cc/public/guides/CodeSubmissionPolicy>) for more details.
- The authors should provide instructions on data access and preparation, including how to access the raw data, preprocessed data, intermediate data, and generated data, etc.
- The authors should provide scripts to reproduce all experimental results for the new proposed method and baselines. If only a subset of experiments are reproducible, they should state which ones are omitted from the script and why.
- At submission time, to preserve anonymity, the authors should release anonymized versions (if applicable).
- Providing as much information as possible in supplemental material (appended to the paper) is recommended, but including URLs to data and code is permitted.

## 6. Experimental Setting/Details

Question: Does the paper specify all the training and test details (e.g., data splits, hyperparameters, how they were chosen, type of optimizer, etc.) necessary to understand the results?

Answer: [Yes]

Justification: We specify all the training and test details including data splits, hyperparameters, how they were chosen, and type of optimizer for the benchmark.

Guidelines:

- The answer NA means that the paper does not include experiments.
- The experimental setting should be presented in the core of the paper to a level of detail that is necessary to appreciate the results and make sense of them.
- The full details can be provided either with the code, in appendix, or as supplemental material.

## 7. Experiment Statistical Significance

Question: Does the paper report error bars suitably and correctly defined or other appropriate information about the statistical significance of the experiments?

Answer: [Yes]

Justification: We report the means and standard deviations with 100 (10) iterations in our simulation studies (applications), respectively.

Guidelines:

- The answer NA means that the paper does not include experiments.
- The authors should answer "Yes" if the results are accompanied by error bars, confidence intervals, or statistical significance tests, at least for the experiments that support the main claims of the paper.

- The factors of variability that the error bars are capturing should be clearly stated (for example, train/test split, initialization, random drawing of some parameter, or overall run with given experimental conditions).
- The method for calculating the error bars should be explained (closed form formula, call to a library function, bootstrap, etc.)
- The assumptions made should be given (e.g., Normally distributed errors).
- It should be clear whether the error bar is the standard deviation or the standard error of the mean.
- It is OK to report 1-sigma error bars, but one should state it. The authors should preferably report a 2-sigma error bar than state that they have a 96% CI, if the hypothesis of Normality of errors is not verified.
- For asymmetric distributions, the authors should be careful not to show in tables or figures symmetric error bars that would yield results that are out of range (e.g. negative error rates).
- If error bars are reported in tables or plots, The authors should explain in the text how they were calculated and reference the corresponding figures or tables in the text.

#### 8. Experiments Compute Resources

Question: For each experiment, does the paper provide sufficient information on the computer resources (type of compute workers, memory, time of execution) needed to reproduce the experiments?

Answer: [Yes]

Justification: We provide sufficient information on the computer resources in the appendix.

Guidelines:

- The answer NA means that the paper does not include experiments.
- The paper should indicate the type of compute workers CPU or GPU, internal cluster, or cloud provider, including relevant memory and storage.
- The paper should provide the amount of compute required for each of the individual experimental runs as well as estimate the total compute.
- The paper should disclose whether the full research project required more compute than the experiments reported in the paper (e.g., preliminary or failed experiments that didn't make it into the paper).

#### 9. Code Of Ethics

Question: Does the research conducted in the paper conform, in every respect, with the NeurIPS Code of Ethics <https://neurips.cc/public/EthicsGuidelines>?

Answer: [Yes]

Justification: We strictly follow the NeurIPS Code of Ethics.

Guidelines:

- The answer NA means that the authors have not reviewed the NeurIPS Code of Ethics.
- If the authors answer No, they should explain the special circumstances that require a deviation from the Code of Ethics.
- The authors should make sure to preserve anonymity (e.g., if there is a special consideration due to laws or regulations in their jurisdiction).

#### 10. Broader Impacts

Question: Does the paper discuss both potential positive societal impacts and negative societal impacts of the work performed?

Answer: [Yes]

Justification: We discuss both potential positive societal impacts and negative societal impacts of our work in the section 6.

Guidelines:

- The answer NA means that there is no societal impact of the work performed.



- If the authors answer NA or No, they should explain why their work has no societal impact or why the paper does not address societal impact.
- Examples of negative societal impacts include potential malicious or unintended uses (e.g., disinformation, generating fake profiles, surveillance), fairness considerations (e.g., deployment of technologies that could make decisions that unfairly impact specific groups), privacy considerations, and security considerations.
- The conference expects that many papers will be foundational research and not tied to particular applications, let alone deployments. However, if there is a direct path to any negative applications, the authors should point it out. For example, it is legitimate to point out that an improvement in the quality of generative models could be used to generate deepfakes for disinformation. On the other hand, it is not needed to point out that a generic algorithm for optimizing neural networks could enable people to train models that generate Deepfakes faster.
- The authors should consider possible harms that could arise when the technology is being used as intended and functioning correctly, harms that could arise when the technology is being used as intended but gives incorrect results, and harms following from (intentional or unintentional) misuse of the technology.
- If there are negative societal impacts, the authors could also discuss possible mitigation strategies (e.g., gated release of models, providing defenses in addition to attacks, mechanisms for monitoring misuse, mechanisms to monitor how a system learns from feedback over time, improving the efficiency and accessibility of ML).

#### 11. Safeguards

Question: Does the paper describe safeguards that have been put in place for responsible release of data or models that have a high risk for misuse (e.g., pretrained language models, image generators, or scraped datasets)?

Answer: [NA]

Justification: Our paper poses no such risks.

Guidelines:

- The answer NA means that the paper poses no such risks.
- Released models that have a high risk for misuse or dual-use should be released with necessary safeguards to allow for controlled use of the model, for example by requiring that users adhere to usage guidelines or restrictions to access the model or implementing safety filters.
- Datasets that have been scraped from the Internet could pose safety risks. The authors should describe how they avoided releasing unsafe images.
- We recognize that providing effective safeguards is challenging, and many papers do not require this, but we encourage authors to take this into account and make a best faith effort.

#### 12. Licenses for existing assets

Question: Are the creators or original owners of assets (e.g., code, data, models), used in the paper, properly credited and are the license and terms of use explicitly mentioned and properly respected?

Answer: [Yes]

Justification: We cite the original papers and include the URL for the webpage.

Guidelines:

- The answer NA means that the paper does not use existing assets.
- The authors should cite the original paper that produced the code package or dataset.
- The authors should state which version of the asset is used and, if possible, include a URL.
- The name of the license (e.g., CC-BY 4.0) should be included for each asset.
- For scraped data from a particular source (e.g., website), the copyright and terms of service of that source should be provided.

- If assets are released, the license, copyright information, and terms of use in the package should be provided. For popular datasets, [paperswithcode.com/datasets](https://paperswithcode.com/datasets) has curated licenses for some datasets. Their licensing guide can help determine the license of a dataset.
- For existing datasets that are re-packaged, both the original license and the license of the derived asset (if it has changed) should be provided.
- If this information is not available online, the authors are encouraged to reach out to the asset's creators.

### 13. **New Assets**

Question: Are new assets introduced in the paper well documented and is the documentation provided alongside the assets?

Answer: [Yes]

Justification: We introduce our proposed with sufficient details in the paper. The documentation is provided alongside our code.

Guidelines:

- The answer NA means that the paper does not release new assets.
- Researchers should communicate the details of the dataset/code/model as part of their submissions via structured templates. This includes details about training, license, limitations, etc.
- The paper should discuss whether and how consent was obtained from people whose asset is used.
- At submission time, remember to anonymize your assets (if applicable). You can either create an anonymized URL or include an anonymized zip file.

### 14. **Crowdsourcing and Research with Human Subjects**

Question: For crowdsourcing experiments and research with human subjects, does the paper include the full text of instructions given to participants and screenshots, if applicable, as well as details about compensation (if any)?

Answer: [NA]

Justification: We used open-source datasets and our paper does not involve crowdsourcing nor research with human subjects.

Guidelines:

- The answer NA means that the paper does not involve crowdsourcing nor research with human subjects.
- Including this information in the supplemental material is fine, but if the main contribution of the paper involves human subjects, then as much detail as possible should be included in the main paper.
- According to the NeurIPS Code of Ethics, workers involved in data collection, curation, or other labor should be paid at least the minimum wage in the country of the data collector.

### 15. **Institutional Review Board (IRB) Approvals or Equivalent for Research with Human Subjects**

Question: Does the paper describe potential risks incurred by study participants, whether such risks were disclosed to the subjects, and whether Institutional Review Board (IRB) approvals (or an equivalent approval/review based on the requirements of your country or institution) were obtained?

Answer: [NA]

Justification: our paper does not involve crowdsourcing nor research with human subjects.

Guidelines:

- The answer NA means that the paper does not involve crowdsourcing nor research with human subjects.

- Depending on the country in which research is conducted, IRB approval (or equivalent) may be required for any human subjects research. If you obtained IRB approval, you should clearly state this in the paper.
- We recognize that the procedures for this may vary significantly between institutions and locations, and we expect authors to adhere to the NeurIPS Code of Ethics and the guidelines for their institution.
- For initial submissions, do not include any information that would break anonymity (if applicable), such as the institution conducting the review.