# SOME CONTRIBUTIONS TO DATA DRIVEN INDIVIDUALIZED DECISION MAKING PROBLEMS

Zhengling Qi

A dissertation submitted to the faculty of the University of North Carolina at Chapel Hill in partial fulfillment of the requirements for the degree of Doctor of Philosophy in the Department of Statistics and Operations Research.

Chapel Hill
2019

Approved by:

Yufeng Liu

Shu Lu

J.S. Marron

Jong-Shi Pang

Donglin Zeng

# ABSTRACT

ZHENGLING Qi: SOME CONTRIBUTIONS TO DATA DRIVEN INDIVIDUALIZED DECISION MAKING PROBLEMS
(Under the direction of Yufeng Liu)

Recent exploration of the optimal individualized decision rule (IDR) for patients in precision medicine has attracted a lot of attentions due to the potential heterogeneous response of patients to different treatments. In the current literature, an optimal IDR is a decision function based on patients' characteristics for the treatment that maximizes the expected outcome. My dissertation research mainly focuses on how to estimate optimal IDRs under various criteria given experimental data.

In the first part of this dissertation, focusing on maximizing expected outcome, we propose an angle-based direct learning (AD-learning) method to efficiently estimate optimal IDRs with multiple treatments for various types of outcomes. This contributes to the literature, where many existing methods are designed for binary treatment settings with the interest of a continuous outcome. In the second part, motivated by complex individualized decision making procedures, we propose two new robust criteria to estimate optimal IDRs: one is to control the average lower tail of the subjects' outcomes and the other is to control the individualized lower tail of each subject's outcome. In addition to optimizing the individualized expected outcome, our proposed criteria take risks into consideration, and thus the resulting IDRs can prevent adverse events caused by the heavy lower tail of the outcome distribution. In the third part of this dissertation, motivated by the concept of Optimized Certainty Equivalent (OCE), we generalize the second part and propose a decision-rule based optimized covariates dependent equivalent (CDE) for individualized decision making problems. Our proposed IDR-CDE not only broadens the existing expected outcome framework in precision medicine but also enriches the previous concept of the OCE in the risk management. We study the related mathematical problem of estimating an optimal IDRs both theoretically and numerically.

*To my parents,*

*Jianxin Qi and Zhi Zheng,*

*and my beloved wife,*

*Xutong Zhao,*

*who have loved and supported me throughout my life.*

# ACKNOWLEDGEMENTS

I feel so grateful and fortune that I have had the opportunity over the past four years working on exciting and challenging statistical problems, and turning them into my Ph.D. dissertation. It is one of the most enjoyable and fulfilling Ph.D. experiences that anybody could hope for. I could not imagine having better support from my advisor, committee members, colleagues, friends and family.

First of all, I would like to express my deepest gratitude to my advisor Dr. Yufeng Liu for his tremendous help of my Ph.D. study and related research, for his patience, encouraging and motivation. He has taught me how to start doing research, to learn and think with his rich knowledge and extensive experience. His technical and editorial advice was essential to the completion of this dissertation. Besides research, he has many impressive traits that I hope to cultivate. He inspires me for the value of seeking simplest but most useful solutions and persisting on them. He shows me how to be an amazing mentor to students that I hope I could be. Your advice on both research as well as my career has been invaluable.

I would also like to thank my committee members: Dr. Shu Lu, Dr. J.S. Marron, Dr. Jong-Shi Pang and Dr. Donglin Zeng, who provide me insightful suggestions on my dissertation. Those invaluable comments have greatly improved this dissertation. I am also grateful to Dr. Edward Carlstein, who has given me a lot of great comments and suggestions during my job search.

I especially want to thank Dr. Jong-Shi Pang for his continuous guidance on my research. Given his generous support, I had an unforgettable experience as a visiting student at University of Southern California. His domain expertise in mathematical programming has reshaped my understanding of optimization and led me to this fascinating field. Besides that, he has set a role model of a great researcher for me with many memorable traits: hard working, open-minded, modest, etc.

I also want to thank Dr. Ji Zhu for his kind support during my master study at University of Michigan, Ann Arbor. Without his encouragement, I would not pursue a Ph.D. and have the chance to enjoy the beauty of doing research in statistics.

I am also thankful to all my fellow colleagues, especially those lovely basement friends. Special appreciation goes to Wawa, Small little face, ZZQ, Tony Fan, Jonathan P Williams, Iain Carmichael and all the members in Dr. Yufeng Liu's research group.

I am very grateful to all my friends, who always support me through my Ph.D. study. Special appreciation alphabetically goes to Yifan Cui, Ying Cui, Ruituo Fan, Cui Guo, Peng Liao, Xia Lu, Hongsheng Liu, Wenbo Sun, Lu Tang, Nalingna Yuan, Ruofei Zhao and Jinyang Zheng.

Last but not least, nothing would be achieved without the unconditional love and dedication of my family, especially my wife Xutong Zhao, my parents Jianxin Qi and Zhi Zheng, throughout my Ph.D. study and my life.

## TABLE OF CONTENTS

## LIST OF TABLES

## LIST OF FIGURES

# LIST OF ABBREVIATIONS AND SYMBOLS

| | |
|---|---|
| IDR | Individualized Decision Rules |
| AD-learning | Angle-based direct learning |
| OCE | Optimized Certainty Equivalent |
| CDE | Covariates dependent equivalent |
| CVaR | Conditional Value at Risk |
| DC | Difference-of-convex |
| MM | Majorization-minimization |
| D-learning | Direct learning |
| $l_1$-PLS | $l_1$ penalized least square |
| OWL | Outcome weighted learning |
| RWL | Residual weighted learning |
| CPH | Cox Proportional Hazard |
| DL | Decision list |
| ACWL | Adaptive contrast weighted learning |
| VT | Virtual twins |
| DCA | Difference of convex algorithm |
| VaR | Value at Risk |

CHAPTER 1

# Introduction

In this chapter, we outline the contributions in the subsequent development of the dissertation.

## 1.1 Multi-armed Angle-based Direct Learning for Estimating Optimal Individualized Decision Rules with Various Outcomes

Estimating an optimal individualized decision rule (IDR) based on patients' information is an important problem in precision medicine. An optimal IDR is a decision function that optimizes patients' expected clinical outcomes. Many existing methods in the literature are designed for binary treatment settings with the interest of a continuous outcome. Much less work has been done on estimating optimal IDRs in multiple treatment settings with good interpretations. In this paper, we propose an angle-based direct learning (AD-learning) method to efficiently estimate optimal IDRs with multiple treatments. Our proposed method can be applied to high dimensional settings under various types of outcomes, such as continuous, survival or binary outcomes. Moreover, it has an interesting geometric interpretation on the effect of different treatments for each individual patient, which can help doctors and patients make better decisions. Finite sample error bounds have been established to provide a theoretical guarantee for AD-learning. Finally, we demonstrate the superior performance of our method via extensive simulation studies and real data applications.

## 1.2 Estimating Individualized Decision Rules with Tail Controls

Most existing literature has focused on finding optimal IDRs that can maximize the expected outcome for each individual. Motivated by complex individualized decision making procedures and popular conditional value at risk (CVaR) measures, in the second part of this dissertation, we propose two new robust criteria to estimate optimal IDRs: one is to control the average lower tail of the subjects' outcomes and the other is to control the individualized lower tail of each subject's outcome. In addition to optimizing the individualized expected outcome, our proposed criteria take risks into consideration, and thus the resulting IDRs can prevent adverse events caused by the heavy lower tail of the outcome distribution. Interestingly, from the perspective of duality theory, the optimal IDR under our criteria can be interpreted as the decision rule that maximizes the "worst-case" scenario of the individualized outcome within a probability constrained set. The corresponding estimating procedures are implemented using two proposed efficient non-convex optimization algorithms, which are based on the recent developments of difference-of-convex (DC) and majorization-minimization (MM) algorithms. We provide a comprehensive statistical analysis for our estimated optimal IDRs under the proposed criteria such as consistency and finite sample error bounds. Simulation studies and a real data application are used to further demonstrate the robust performance of our methods.

## 1.3 Estimation of Individualized Decision Rules Based on an Optimized Covariate-Dependent Equivalent of Random Outcomes

In third part of this dissertation, motivated by the concept of Optimized Certainty Equivalent (OCE), we propose a decision-rule based optimized covariates dependent equivalent (CDE) for individualized decision making problems. Our proposed IDR-CDE not only broadens the existing expected-mean outcome framework in precision medicine but also enriches the previous concept of the OCE. Under a functional margin description of the decision rule modeled by an indicator function as in the literature of large-margin classifiers, we study the mathematical problem of estimating an optimal IDRs in two cases: in one case, an optimal solution can be obtained "explicitly" that involves the implicit evaluation of an OCE; the other case requires the numerical solution of an empirical minimization problem obtained by sampling the underlying distributions of the random variables involved. A major challenge of the latter optimization problem is that it involves a discontinuous objective function. We show that, under a mild condition at the population level of the model, the epigraphical formulation of this empirical optimization problem is a piecewise affine, thus difference-of-convex (dc), constrained dc, thus nonconvex, program. A simplified dc algorithm is employed to solve the resulting dc program whose convergence to a new kind of stationary solutions is established. Numerical experiments demonstrate that our overall approach outperforms existing methods in estimating optimal IDRs under heavy-tail distributions of the data. In addition to providing a risk-based approach for individualized medical treatments, which is new in the area of precision medicine, the main contributions of this work in general include: the broadening of the concept of the OCE, the epigraphical description of the empirical IDR-CDE minimization problem and its equivalent dc formulation, and the optimization of resulting piecewise affine constrained dc program.

CHAPTER 2

# Multi-armed Angle-based Direct Learning for Estimating Optimal Individualized Decision Rules with Various Outcomes

## 2.1 Introduction

Precision medicine, which recommends different treatments for individual patients, has been a popular research area in the scientific community. Compared with traditional "one-size-fits-all" medical procedures, precision medicine provides an individualized decision for each patient based on their information, such as clinical covariates, genetics, in order to maximize the outcome of each patient. In practice, there exists various types of outcomes such as time to event, health index or the disease indicator.

There are a number of existing statistical methods for estimating optimal IDRs in the literature. These methods can be roughly characterized into two types. The first type includes indirect methods such as Q-learning ((Watkins and Dayan, 1992; Watkins, 1989; Murphy, 2005; Qian and Murphy, 2011) and A-learning ((Murphy, 2003; Robins, 2004)). Q-learning estimates optimal IDRs via modeling the conditional outcome function based on covariates while A-learning models the contrast between rewards of two treatments. The second type of methods directly targets the decision rules. One major approach is to recast the estimating IDRs problem into weighted classification problems and use different machine learning techniques to estimate optimal IDRs ((Zhang et al., 2012; Zhao et al., 2012; Zhou et al., 2017; Zhao et al., 2015a; Tao and Wang, 2017)). In order to enhance interpretability of decision rules, different tree based methods were also proposed ((Zhang et al., 2015; Foster et al., 2011; Laber and Zhao, 2015)). Other direct-search methods include (Tian et al., 2014) and Direct Learning (D-learning) ((Qi and Liu, 2018)), which directly estimate the decision function that leads to optimal IDRs by regression techniques. Recently, a general statistical framework to estimate optimal IDRs was proposed by (Chen et al., 2017).

Censored data are commonly seen in practice because of early drop out or other reasons. Thus, it is also important to develop methods to estimate optimal IDRs for the survival outcome. Various methods have been proposed in the literature to estimate optimal IDRs for survival outcomes, such as (Goldberg and Kosorok, 2012; Zhao et al., 2015b) and (Cui et al., 2017). Recently, (Bai et al., 2016) and (Jiang et al., 2016) proposed several methods to estimate the optimal IDR that can maximize the survival probability of patients.

In the current literature, most of these existing methods are designed for binary treatment settings only. But there are many multi-armed IDR problems in pratice ((Baron et al., 2013)). To the best of our knowledge, not much has been done for estimating the optimal IDR for the multi-armed treatment settings with various outcomes, such binary and survival outcomes. Thus it is essential to develop methods to take multiple treatments into consideration simultaneously and estimate optimal IDRs for various outcomes, which can help to improve the estimating efficiency and the classification accuracy.

Besides the accurate estimation of IDRs, good interpretations are also important for multi-armed treatment settings. For binary treatment settings, indirect methods can report a single value difference function between two treatments to illustrate the relative effectiveness. For classification based methods such as O-learning ((Zhao et al., 2012)), interpretation of the decision rule for binary treatment settings may not be as clear. Meanwhile for $K$-armed treatment settings, at least $\frac{K(K-1)}{2}$ pairwise value difference functions need to be estimated to illustrate the relative performance of treatments for each patient. Although such an extension can be simple to implement, it does not use the data simultaneously and consequently may yield suboptimal decision rules.

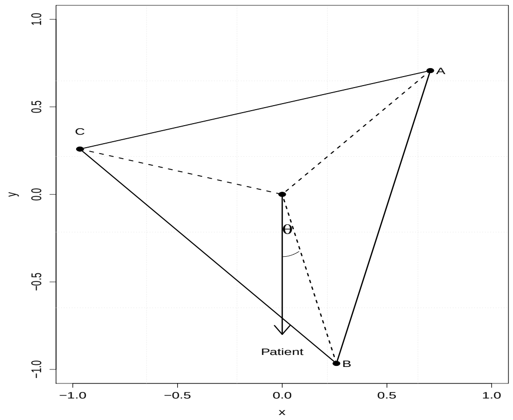


**Figure 2.1:** Graphical illustration of the estimated IDR for a given patient in a three-treatment setting. Vertices $A, B$ and $C$ represent 3 treatments. The estimated IDR of the patient has the least angle with treatment $B$ which is thus more preferable than the other two treatments.

To get accurate estimation of optimal IDRs and obtain a good interpretation jointly under the multi-armed setting, we consider a $K$-vertex simplex structure in an Euclidean space, where each vertex represents one treatment. The simplex lies in a $K$-1 dimensional space with the origin as the center and has equal inner products among vertices. Using the expression of the optimal IDR, we transform the problem of maximizing the value function into maximizing the inner product between the decision function vector and the corresponding vertex in the simplex space. Such a transformation allows us to estimate the optimal IDR using multiple response regression methods. In particular, for each patient, our estimated decision function vector maps the covariates into this $K-1$ dimensional space. The angle between each treatment vertex and the estimation function vector can be interpreted as a measure of preference to this treatment. We recommend a patient to take the corresponding treatment having the least angle with our estimated decision function vector. Figure 2.1 shows an example with our estimated IDR for a given patient. In this case, we recommend treatment $B$ as the best option for this patient. In addition, we can see treatment $C$ is more preferable than treatment $A$ for this patient based on their corresponding angles.

We call our method angle-based direct learning (AD-learning) which can directly estimate optimal IDRs under multi-armed treatment settings using multiple response regression techniques. Furthermore, our proposed AD-learning can be extended to various types of outcome such as binary and survival responses. Compared with existing methods, our proposed AD-learning enjoys several advantages. In particular, our method is robust in the sense that it is not necessary to model the main effect function of the conditional outcome. Due to direct learning scheme, our method does not suffer from the mismatch problem between minimizing prediction errors and maximizing value functions in model based methods such as $l_1$-PLS ((Qian and Murphy, 2011)) and can perform better in high dimensional settings. Moreover, by representing each treatment as a vertex of a standard simplex in the Euclidean space, our proposed method provides an attractive geometric interpretation of the relative effectiveness of all treatments for a given patient. The resulting relative effectiveness of different treatments can be interpreted as the angles between the decision function vector for the patient and each vertex corresponding to the treatment. These angles can be scaled between $[0, \pi]$. In addition, flexible structures such as group and low rank sparsity can be also incorporated to further improve the model interpretation and simplicity, which can be applied in high dimensional settings. Finally, our proposed method is easy to implement with efficient algorithms.

The remainder of Chapter 2 is organized as follows. In Section 2.2, we introduce our AD-learning to estimate optimal IDRs in multiple treatment settings. In Section 2.3, we discuss how to extend our proposed method to binary and survival outcomes. In Section 2.4, we provide a theoretical guarantee for our AD-learning under some mild assumptions. In Section 2.5, we conduct an extensive simulation study to evaluate the finite sample performance of our method with implementation details including algorithms. Furthermore, we illustrate our method using the AIDS data in Section 2.6. We conclude our paper with some discussions and possible future extensions in Section 2.7. Details of proof and additional simulation results are given in the Appendix A.

## 2.2 Angle Based Direct Learning

For notation of this Chapter, we use boldface capital and lowercase symbols to denote matrices and vectors respectively. For a matrix $\mathbf{B}$, we define a mixed $l_1$ and $l_2$ norm as $||\mathbf{B}||_{2,1} = \sum ||\mathbf{B}_j||_2$, where $\mathbf{B}_j$ is the $j$-th row vector of $\mathbf{B}$. We use $\mathrm{Tr}(\mathbf{B})$ to denote the sum of the diagonal value of the matrix $\mathbf{B}$.

We consider a randomized treatment framework for estimating optimal IDRs. For each patient, we observe a triplet random vector $(\mathbf{x}, A, R)$. In particular, $\mathbf{x} = (1, X_1, \cdots, X_p) \in \mathcal{X}$ denotes patients' $p$-dimensional covariates with an intercept. The random variable $A$ represents the randomized treatment that a patient receives. Here we consider the $K$-treatment-armed setting where $A \in \{1, 2, \cdots, K\}$ with a known prior probability distribution $\pi(A, \mathbf{x})$, which is the conditional probability depending on $\mathbf{x}$. In a general setting other than the randomized trial study, $\pi(A, \mathbf{x})$ denotes the propensity score and can be estimated by the generalized linear models such as multinomial logistic regression. The variable $R$ is a patient's outcome after receiving the treatment $A$. Without loss of generality, we assume that the outcome $R$ is bounded and the larger $R$ is, the better the treatment works for this patient.

One of the most important goals of our problem is to estimate the optimal IDR that can maximize the expected clinical outcome of each patient under this IDR. Mathematically speaking, an IDR is a decision function $d(\mathbf{x}) : \mathcal{X} \to \mathcal{A}$, mapping from the covariate space into the treatment space. According to (Qian and Murphy, 2011) and (Zhao et al., 2012), the value function under the IDR $d$ can be expressed as

$$V(d) =: \mathbf{E}[R|d(\mathbf{x}) = A] = \mathbf{E}[\frac{R\mathbb{I}(A = d(\mathbf{x})}{\pi(A, \mathbf{x})}], \tag{2.1}$$

where $\mathbb{I}(\bullet)$ is the indicator function. Then the optimal IDR is defined as

$$d_0(\mathbf{x}) = \text{argmax}_{d \in D} V(d) \tag{2.2}$$

within a pre-specified class of treatment rules $D$. Before introducing our proposed AD-learning, we first discuss the direct learning framework.

### 2.2.1 The Direct Learning Framework

Consider a binary problem with $K = 2$. We encode treatment $A$ to be 1 or $-1$. Then from the value function and optimal IDRs defined in (2.1) and (2.2) respectively, we can further represent the optimal IDR as

$$\begin{aligned} d_0(\mathbf{x}) &= \text{sign}(\mathbf{E}[R|\mathbf{x}, A=1] - E[R|\mathbf{x}, A=-1]) \\ &= \text{sign}(\mathbf{E}[\frac{RA}{\pi(A|\mathbf{x})}|\mathbf{x}]) := \text{sign}(f_0(\mathbf{x})). \end{aligned} \tag{2.3}$$

Using Equation (2.3), similarly in (Tian et al., 2014), the IDR problem is equivalent to estimate the optimal decision function $f_0(\mathbf{x}) = \mathbf{E}[\frac{RA}{\pi(A|\mathbf{x})}|\mathbf{x}]$ via various regression methods such as $l_1$ penalized regression (LASSO). The final decision rule is determined by the sign of this estimator.

Binary D-learning directly estimates the decision rule. It is very different from the outcome weighted learning (OWL) proposed by (Zhao et al., 2012) because binary D-learning uses regression methods to estimate the optimal IDR directly. Note that binary D-learning can be simply extended to the $K$-treatment-arm setting by rewriting the optimal IDR as

$$\begin{aligned} d_0(\mathbf{x}) = & \underset{k\in\{1,\cdots,K\}}{\text{argmax}} \ \mathbf{E}[R|\mathbf{x}, A=k] \\ = & \underset{k\in\{1,\cdots,K\}}{\text{argmax}} \ K\mathbf{E}[R|\mathbf{x}, A=k] - \sum_{i=1}^{K} \mathbf{E}[R|\mathbf{x}, A=i] \\ = & \underset{k\in\{1,\cdots,K\}}{\text{argmax}} \sum_{i\neq k}^{K} \{\mathbf{E}[R|\mathbf{x}, A=k] - \mathbf{E}[R|\mathbf{x}A=i]\} \\ = & \underset{k\in\{1,\cdots,K\}}{\text{argmax}} \sum_{i\neq k}^{K} \mathbf{E}[\frac{RA_{ki}}{\pi_{ki}(A_{ki}, \mathbf{x})}|\mathbf{x}, A = k \text{ or } i] \\ := & \underset{k\in\{1,\cdots,K\}}{\text{argmax}} \sum_{i\neq k}^{K} f_{ki}(\mathbf{x}) := \underset{k\in\{1,\cdots,K\}}{\text{argmax}} f_k(\mathbf{x}), \end{aligned} \tag{2.4}$$

where $A_{ki} \in \{-1, 1\}$ represents treatments $k$ and $i$, and $f_{ki}(\mathbf{x})$ is defined as the optimal decision function between treatment $k$ and $i$. Each pairwise decision function can be estimated by a binary D-learning method. The final treatment decision rule is to compare the cumulative sum of pairwise decison functions $f_k(\mathbf{x})$ for $k = 1, \cdots, K$, and select the largest one. We refer this pairwise method as pairwise D-learning.

Binary D-learning gives us a directed way to estimate optimal IDRs. However, pairwise D-learning, which is based on binary D-learning, focuses only on pairwise comparisons between treatments without considering all treatments simultaneously. Although the proposed effect measure $f_k(\mathbf{x})$ can capture the relative strength of a treatment for a given patient, it may be suboptimal.

To handle multi-armed IDR problems, we propose AD-learning that considers all treatments together to estimate the optimal IDR. Moreover, the AD-learning can provide a more effective measure of treatments for patients with a good interpretation.

### 2.2.2 Angle Based D-learning for Continuous Outcomes

For a $K$-armed IDR problem, one natural approach is to estimate $K$ functions for all treatments. Since only one function is needed for the binary IDR problem, one indeed only needs $K-1$ functions for a $K$-armed problem. Instead of using $K$ functions with a constraint on these functions, we aim to directly estimate $K-1$ functions. To that end, we project the treatment $A$ into $K$ simplex vertices defined on $\mathcal{R}^{K-1}$. Specifically, we encode the $j$-th treatment as a vector $\mathbf{w}_j \in \mathcal{R}^{K-1}$ with

$$\mathbf{w}_j = \begin{cases} (K-1)^{-1/2}\mathbf{1}_{K-1}, & \text{if } A = 1 \\ -(1+\sqrt{K})/(K-1)^{3/2}\mathbf{1}_{K-1} + (\frac{K}{K-1})^{1/2} e_{A-1}, & \text{if } 2 \leq A \leq K, \end{cases} \tag{2.5}$$

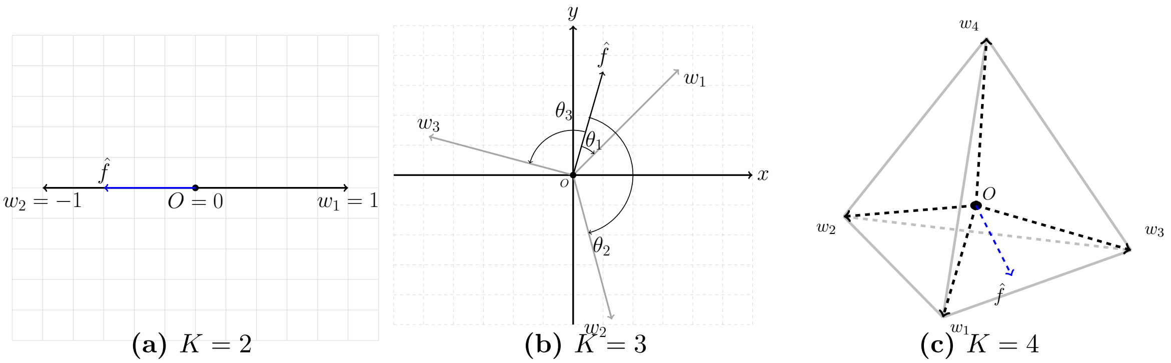
where $e_i$ is a $K-1$ dimensional vector with every element being 0, except the $i$-th location being 1. Define $\mathbf{w}$ as a random vector with $\mathbf{P}[\mathbf{w} = \mathbf{w}_j|\mathbf{x}] = \mathbf{P}[A = j|\mathbf{x}]$. This simplex encoding scheme has several properties. In particular, the center of these vertices is the origin of the space, that is $\sum_{j=1}^{K} \mathbf{w}_j = 0$ with $||\mathbf{w}_j||_2 = 1$ for $j = 1, \cdots, K$. The angle between each pair of vertices is equal, that is $\mathbf{w}_i^T \mathbf{w}_j = C(K) < 1$ for $i \neq j$, where the constant $C$ only depends on

$K$. Interestingly, we can then express the optimal IDR as

$$\begin{aligned}
d_0(\mathbf{x}) &= \underset{k\in\{1,\cdots,K\}}{\text{argmax}}\ \mathbf{E}[R|\mathbf{x}, A=k] \\
&= \underset{k\in\{1,\cdots,K\}}{\text{argmax}}\ (1-c(K))\mathbf{E}[R|\mathbf{x}, A=k] \\
&= \underset{k\in\{1,\cdots,K\}}{\text{argmax}}\ \{(1-c(K))\mathbf{E}[R|\mathbf{x}, A=k] + c(K)\sum_{j=1}^{K}\mathbf{E}[R|\mathbf{x}, A=j]\} \\
&= \underset{k\in\{1,\cdots,K\}}{\text{argmax}}\ \{\mathbf{E}[R|\mathbf{x}, A=k] + c(K)\sum_{j\neq k}^{K}\mathbf{E}[R|\mathbf{x}, A=j]\} \qquad (2.6) \\
&= \underset{k\in\{1,\cdots,K\}}{\text{argmax}}\ \{\mathbf{w}_k^T\mathbf{E}[R\mathbf{w}|\mathbf{x}, A=k] + \mathbf{w}_k^T\sum_{j\neq k}^{K}\mathbf{E}[R\mathbf{w}|\mathbf{x}, A=j]\} \\
&= \underset{k\in\{1,\cdots,K\}}{\text{argmax}}\ \mathbf{w}_k^T\mathbf{E}[\frac{R\mathbf{w}}{\pi(A,\mathbf{x})}|\mathbf{x}] := \underset{k\in\{1,\cdots,K\}}{\text{argmax}}\ \mathbf{w}_k^T\mathbf{f}_0(\mathbf{x}),
\end{aligned}$$

where $\mathbf{f}_0(\mathbf{x})$ is a function vector from $\mathcal{R}^{p+1}$ to $\mathcal{R}^{K-1}$ with some abuse of notation. Then the optimal IDR is given by comparing the inner product between $\mathbf{w}_k$ and $\mathbf{f}_0(\mathbf{x})$ for each treatment $k$. We define the angle between each pair of vertices in $[0, \pi]$. Then $\mathbf{w}_k^T\mathbf{f}_0(\mathbf{x})$ is the largest if and only if the angle between $\mathbf{w}_k$ and $\mathbf{f}_0(\mathbf{x})$ is the least, for $k = 1, \cdots, K$. Thus we call our proposed method as Angle based D-learning (AD-learning). Note that the simplex coding is unique up to the orthogonal rotation.

Our proposed AD-learning has an attractive geometric interpretation. In particular, this least angle decision rule can be understood through newly defined treatment regions for each patient. For example, when $K = 3$, as shown in Figure 2.2 (b), vectors $\mathbf{w}_k$; $k = 1, \cdots, K$ form an equilateral triangle in the $\mathcal{R}^2$ space, where each divided region represents a treatment region. The decision function vector $\mathbf{f}_0(\mathbf{x})$ maps from the covariate space into the treatment region. One can observe that the angles between vertices are the same, and consequently each treatment is treated equally. Such a simplex coding scheme does not require a balanced group size for each treatment since treatment proportions are taken into account by the term $\pi(A, \mathbf{x})$ in Equation (2.6). We name the angle between each $\mathbf{w}_k$ and $\mathbf{f}_0(\mathbf{x})$ as the *treatment score* which lies in a bounded interval $[0, \pi]$. If a patient has the angle of 0 with the $i$-th treatment, then the $i$-th treatment is a perfect fit for this patient compared with other treatments. Figure 2.2 gives a geometric explanation of our AD-learning.



**(a)** $K = 2$ **(b)** $K = 3$ **(c)** $K = 4$

**Figure 2.2:** Geometric interpretation of our least angle decision rule. When $K = 3$ or $K = 4$, the estimate $\hat{f}$ has the smallest angle with treatment 1 so we recommend treatment 1 as the optimal treatment. When $K = 2$, we can see $\hat{f}$ has the smallest angle with vector $w_2$ and the optimal rule for this patient is treatment 2.

To further illustrate our AD-learning, we propose the following alternative interpretation. Suppose the clinical outcome $R$ can be expressed as

$$R = \mu(\mathbf{x}) + \sum_{i=1}^{K} \delta_i(\mathbf{x})\mathbb{I}(A = i) + \epsilon, \tag{2.7}$$

where $\mu(\mathbf{x})$ is main effect function, $\delta_i(\mathbf{x})$ is the interaction effect between covariates and the $i$-th treatment, and $\epsilon$ is mean zero random error. Then we can get

$$\begin{aligned} E[\frac{R\mathbf{w}}{\pi(A,\mathbf{x})}|\mathbf{x}] &= \mu(\mathbf{x})E[\frac{\mathbf{w}}{\pi(A,\mathbf{x})}|\mathbf{x}] + \sum_{i=1}^{K} \delta_i(\mathbf{x})_i\mathbf{E}[\frac{\mathbf{w}\mathbb{I}(A=i)}{\pi(A,\mathbf{x})}|\mathbf{x}] + E[\frac{\mathbf{w}}{\pi(A,\mathbf{x})}|\mathbf{x}]\mathbf{E}[\epsilon|\mathbf{x}] \\ &= \sum_{i=1}^{K} \delta_i(\mathbf{x})\mathbf{w}_i. \end{aligned} \tag{2.8}$$

Furthermore, the optimal IDR is

$$\begin{aligned} d_0(\mathbf{x}) &= \text{argmax}_{k\in\{1,\cdots,K\}}\mathbf{w}_k^T\mathbf{E}[\frac{R\mathbf{w}}{\pi(A|\mathbf{x})}|\mathbf{x}] \\ &= \text{argmax}_{k\in\{1,\cdots,K\}}\mathbf{w}_k^T\sum_{i=1}^{K}\delta_i(\mathbf{x})\mathbf{w}_i \\ &= \text{argmax}_{k\in\{1,\cdots,K\}}C(K)\sum_{i=1}^{K}\delta_i(\mathbf{x}) + (1 - C(K))\delta_k(\mathbf{x}) \\ &= \text{argmax}_{k\in\{1,\cdots,K\}}\delta_k(\mathbf{x}), \end{aligned} \tag{2.9}$$

which is exactly to compare each treatment interaction effect with the covariates.

As a remark, we note that extensions of methods for binary treatment settings to multiple treatment settings using all treatments jointly can be nontrivial since we need to account for multiple treatment effect comparisons without sacrificing too much efficiency. Our proposed AD-learning achieves this by first projecting treatments into a $K$-1 dimensional space. A simplex with $K$ vertices is used to represent the $K$ treatments. Then Equation (2.6) provides an innovative but direct way to efficiently estimate the decision function vector and considers all the data simultaneously. Inherited from the simplex structure, our proposed method has an attractive geometric interpretation to show the relative effectiveness of different treatments for a patient. Thus it provides an informative comparison of all treatments for patients and doctors to make decisions.

Note that the simplex coding scheme has previously been used by (Wu and Lange, 2010) and (Zhang and Liu, 2014) for classification problems. However, our proposed AD-learning is very different because it is not a classification method. Consequently, our method is not an extension of O-learning proposed by (Zhao et al., 2012). Instead, by transforming the problem (2.2) into (2.6), our goal is to estimate the decision function $\mathbf{f}_0(\mathbf{x})$ directly, using multiple response regression introduced in Section 2.2.3.

### 2.2.3 Estimation Procedures of AD-learning

In order to estimate the optimal IDR, it is equivalent to estimating $\mathbf{f}_0(\mathbf{x})$ from Section 2.2.2. The next lemma provides us a way for estimation of $\mathbf{f}_0(\mathbf{x})$.

**Lemma 2.2.1.** *Under the exchange of differential and expectation condition,* $\mathbf{f}_0(\mathbf{x})$ *is an optimal solution to*

$$\underset{\mathbf{f}\in\mathcal{R}^{K-1}}{argmin} \;\; \mathbf{E}[\frac{1}{\pi(A,\mathbf{x})}(KR\mathbf{w}-\mathbf{f}(\mathbf{x}))^T\Sigma(KR\mathbf{w}-\mathbf{f}(\mathbf{x}))], \tag{2.10}$$

*where* $\Sigma$ *can be any positive definite matrix that characterizes the dependency among responses. Without knowing any prior knowledge, one could simply let* $\Sigma = I_{K-1}$.

Assume we observe independent identically distributed data $\{(\mathbf{x}_i, A_i, R_i), i = 1, \cdots, n\}$. Then we can estimate $\mathbf{f}_0(\mathbf{x})$ via empirical average approximation

$$\underset{\mathbf{f}\in\mathcal{F}}{\operatorname{argmin}} \quad \frac{1}{n(K-1)}\sum_{i=1}^{n}\frac{1}{\pi(A_i,\mathbf{x}_i)}(KR_i\mathbf{w}_i-\mathbf{f}(\mathbf{x}_i))^T(KR_i\mathbf{w}_i-\mathbf{f}(\mathbf{x}_i)), \tag{2.11}$$

where $\mathcal{F}$ is a pre-specified class of decision functions. For simplicity, we first consider the class of linear decision rules, that is, $\mathcal{F} := \{\mathbf{f}(\mathbf{x}) = \mathbf{B}^T\mathbf{x}, \mathbf{B} \in \mathbb{R}^{p\times(K-1)}\}$. By observing $KR_i\mathbf{w}_i$ as multivariate responses, one can apply ordinary least square estimates for each of the responses separately. However, since the responses share the same clinical outcome $R_i$ for the $i$-th sample, it is clear that pooling multivariate responses together can efficiently improve the estimation of $\mathbf{f}_0(\mathbf{x})$ ((Breiman and Friedman, 1997)). This motivates us to incorporate shrinkage and selection strategies that explore the correlations among different responses by

$$\underset{\mathbf{B}\in\mathbb{R}^{p\times(K-1)}}{\operatorname{argmin}} \quad \frac{1}{n(K-1)}\sum_{i=1}^{n}\frac{1}{\pi(A_i,\mathbf{x}_i)}(KR_i\mathbf{w}_i - \mathbf{B}^T\mathbf{x}_i)^T(KR_i\mathbf{w}_i - \mathbf{B}^T\mathbf{x}_i) + \lambda J(\mathbf{B}), \tag{2.12}$$

where $\lambda$ is a positive tuning parameter. Then our final least angle decision rule becomes $d_0(\mathbf{x}) = \operatorname{argmax}_{k\in\{1,\cdots,K\}}\mathbf{w}_k^T\mathbf{B}^T\mathbf{x}$. In this decision rule, the corresponding coefficient for the $j$-th variable of $\mathbf{x}$ is $\mathbf{w}_k^T\mathbf{B}_j$, for $j = 1, \cdots, p$, where $\mathbf{B}_j$ is the $j$-th row vector of $\mathbf{B}$. Note that for any orthogonal matrix $\mathbf{\Gamma}$ ,

$$\begin{aligned} ||\mathbf{B}\mathbf{\Gamma}||_{2,1} = \sum_{j=1}^{p}||\mathbf{B}_j^T\mathbf{\Gamma}||_2 \quad &= \sum_{j=1}^{p}\sqrt{\mathbf{B}_j^T\mathbf{\Gamma}\mathbf{\Gamma}^T\mathbf{B}_j} \\ = \sum_{j=1}^{p}||\mathbf{B}_j||_2 \quad &= ||\mathbf{B}||_{2,1}, \end{aligned} \tag{2.13}$$

which implies that $||\mathbf{B}||_{2,1}$ remains to be the same under any orthogonal transformation of $\mathbf{w}$. This is essential since our simplex coding is unique up to the orthogonal rotation. In addition, $\mathbf{B}_j = \mathbf{0}_{K-1}$ implies the $j$-th variable has no effect on our least angle decision rule. These motivate us to use the group sparsity penalty, i.e., the mixed $l_1/l_2$ norm as follows

$$\underset{\mathbf{B}\in\mathbb{R}^{p\times(K-1)}}{\operatorname{argmin}} \quad \frac{1}{n(K-1)}\sum_{i=1}^{n}\frac{1}{\pi(A_i,\mathbf{x}_i)}(KR_i\mathbf{w}_i - \mathbf{B}^T\mathbf{x}_i)^T(KR_i\mathbf{w}_i - \mathbf{B}^T\mathbf{x}_i) + \lambda||\mathbf{B}||_{2,1}. \tag{2.14}$$

Model (2.14) is best suited for the case that all treatments share the common interaction covariates. The group sparsity structure of $\mathbf{B}$ will not change under any orthogonal transformation of $\mathbf{w}$.

It is known that group sparsity of a matrix is a special case of a low rank matrix. If $\mathbf{B} = \mathbf{U}\mathbf{V}^T$ such that $\mathbf{U} \in \mathbb{R}^{p\times r}$ and $\mathbf{V} \in \mathbb{R}^{r\times(K-1)}$ with $r < \min(p, K-1)$. Then $\mathbf{B}^T\mathbf{x} = \mathbf{V}(\mathbf{U}^T\mathbf{x})$

implies potential $r$ orthogonal latent factors in the covariates. Hence we can also use the nuclear norm penalty to control the complexity of coefficient matrix $\mathbf{B}$ if there is a low rank structure or exists latent factors in the covariates by

$$\underset{\mathbf{B}\in\mathbb{R}^{p\times(K-1)}}{\operatorname{argmin}} \quad \frac{1}{n(K-1)}\sum_{i=1}^{n}\frac{1}{\pi(A_i,\mathbf{x}_i)}(KR_i\mathbf{w}_i-\mathbf{B}^T\mathbf{x}_i)^T(KR_i\mathbf{w}_i-\mathbf{B}^T\mathbf{x}_i)+\lambda||\mathbf{B}||_*, \tag{2.15}$$

where $||\mathbf{B}||_*$ denotes the sum of all singular values of coefficient matrix $\mathbf{B}$. The nuclear norm penalty, unlike the rank constraint, provides a soft and stable shrinkage on the singular values. Moreover, it is also invariant to orthogonal rotation of $\mathbf{w}$.

So far, we have only focused on linear decision rules. If $\mathbf{f}_0(\mathbf{x})$ belongs to some classes of nonlinear functions, we can adapt our method to nonlinear learning via kernel learning or basis function expansions. For kernel learning, we can apply kernel ridge regression for each response separately, using Equation (2.11). However, it may lose some efficiency since it does not consider the dependence among the responses. How to perform kernel learning with multiple responses in our setting will be an interesting future research. For basis function expansions, depending on the problem, we can use spline basis functions, interaction functions, wavelet functions, etc. to approximate the nonlinear decision function.

To summarize, Models (2.14) and (2.15) are proposed to control the complexity of coefficient matrix $\mathbf{B}$ and consequently can enhance the estimation and prediction. As our proposed AD-learning directly targets on the decision function $f_0(\mathbf{x})$, it does not suffer the mismatch problem between minimizing prediction errors and maximizing value functions happened for model-based methods such as $l_1$-PLS. Thus our proposed method tends to perform better in high dimensional settings. If there are group signals in the covariates for optimal IDRs, we recommend to use Model (2.14). If there are latent factors in the covariates for optimal IDRs, we recommend to use Model (2.15). One can always use the cross-validation procedure to select Model (2.14) or (2.15) that maximizes the empirical value function on the validation dataset. The computation of these models relies on convex optimization and thus can be solved efficiently.

## 2.3 Extensions to Other Types of Outcomes

In Sections 2, we proposed AD-learning for continuous outcomes. In practice, especially in clinical studies, other types of outcomes such as binary, count responses, or survival time can also be used. In this section, we extend our AD-learning to more general types of outcomes motivated by the following lemma.

**Lemma 2.3.1.** *Under the exchange of differential and expectation condition,* $\mathbf{f}_0(\mathbf{x})$ *is an optimal solution to*

$$\underset{\mathbf{f}\in\mathcal{F}}{argmin} \quad \mathbf{E}[\frac{1}{\pi(A,\mathbf{x})}(\frac{K}{K-1}R - \mathbf{w}^T\mathbf{f}(\mathbf{x}))^2]. \tag{2.16}$$

Based on the optimization problem (2.16), one can write a corresponding working model as

$$\frac{K}{K-1}R = \mathbf{w}^T f(\mathbf{x}) + \epsilon, \tag{2.17}$$

where $\epsilon$ is the random error. Note that when $\mathbf{f} \in \mathcal{F}$, $\mathbf{w}^T\mathbf{f}(\mathbf{x}) = \mathbf{w}^T\mathbf{B}^T\mathbf{x} = \mathrm{Tr}(\mathbf{B}^T(\mathbf{x}\mathbf{w}^T))$. Then $\mathbf{x}\mathbf{w}^T$ can be regarded as modified covariates. Then the multiple response regression model in (2.11) can be extended to a more general model, namely trace regression model ((Rohde et al., 2011)).

Motivated by the optimization problem (2.16) and the corresponding working model, we can extend our proposed AD-learning to more general settings. In particular, instead of the least squared loss for continuous outcome in (2.16), we can use other loss functions for corresponding outcomes.

### 2.3.1 Binary Outcomes

When $R$ is binary, motivated by Lemma 2.3.1 and the connection between (2.16) and working model (2.17), we consider to replace the least squared loss in (2.16) by the deviance loss of logistic regression models. Then we have the following lemma.

**Lemma 2.3.2.** *Under the exchange of differential and expectation condition, an optimal solution to*

$$\underset{\mathbf{f}\in\mathcal{F}}{argmin} \quad \mathbf{E}[-\frac{R\mathbf{w}^T\mathbf{f}}{\pi(A,\mathbf{x})} + \frac{\log(1+\exp(\mathbf{w}^T\mathbf{f}))}{\pi(A,\mathbf{x})}] \tag{2.18}$$

*is the function* $\mathbf{f}_0(\mathbf{x})$ *satisfying*

$$\mathbf{P}[R = 1|\mathbf{x}, A = i] = \frac{\exp(\mathbf{w}_i^T \mathbf{f}_0(\mathbf{x}))}{1 + \exp(\mathbf{w}_i^T \mathbf{f}_0(\mathbf{x}))}. \tag{2.19}$$

Analogous to (2.17), solving (2.18) is equivalent to fitting a logistic regression working model (2.19). Based on Lemma 2.3.2, we can derive the optimal decision rule for the binary outcome as

$$\begin{aligned} d_0(\mathbf{x}) &= \text{argmax}_{k\in\{1,\cdots,K\}} \mathbf{P}[R = 1|\mathbf{x}, A = i] \\ &= \text{argmax}_{k\in\{1,\cdots,K\}} \mathbf{w}_i^T \mathbf{f}_0(\mathbf{x}), \end{aligned} \tag{2.20}$$

which can be also interpreted as the least angle decision rule. Then we can fit a weighted logistic regression with modified covariates $\mathbf{x}^* = \mathbf{x}\mathbf{w}^T$ by modeling

$$\mathbf{P}[R = 1|\mathbf{x}, A] = \frac{\exp(\text{Tr}(\mathbf{B}^T \mathbf{x}^*))}{1 + \exp(\text{Tr}(\mathbf{B}^T \mathbf{x}^*))}, \tag{2.21}$$

and estimate the coefficient matrix $\mathbf{B}$ by maximum likelihood estimation

$$\underset{\mathbf{B}\in\mathbb{R}^{p\times(K-1)}}{\text{argmin}} \quad l(\mathbf{B}) = -\frac{1}{n}\sum_{i=1}^{n} \frac{R_i \text{Tr}(\mathbf{B}^T \mathbf{x}_i^*)}{\pi(A_i, \mathbf{x}_i)} + \frac{1}{n}\sum_{i=1}^{n} \frac{\log(1 + \exp(\text{Tr}(\mathbf{B}^T \mathbf{x}_i^*)))}{\pi(A_i, \mathbf{x}_i)} + \lambda J(\mathbf{B}), \tag{2.22}$$

where $J(\mathbf{B})$ is either the mixed $l_1/l_2$ penalty or the nuclear norm penalty under different model assumptions. We can use the accelerated proximal gradient method to solve this problem ((Beck and Teboulle, 2009)). However, the gradient of the exponential loss function for this model may need relatively large computational time. The efficient group coordinate descent algorithm proposed by (Breheny and Huang, 2015) can be used as an alternative to solve Model (2.22) with the mixed $l_1/l_2$ penalty by vectorizing the modified covariates.

### 2.3.2 Survival Outcomes

When $R$ is the survival outcome, due to the potential censoring of observations, we do not always observe the exact outcomes of patients in clinical studies. Thus $R$ becomes a pair of random variables defined as $R = (Y, \delta) = (\tilde{Y} \wedge C, \delta)$, where $\tilde{Y}$ is the patient's survival time, $C$ is the censoring time, and $\delta$ is an indicator about whether this patient is censored or not. Motivated by Lemma 2.3.1 and a similar derivation as in Section 3.1, we can replace squared error loss in (2.16) for continuous outcomes by the negative log-likelihood of the Cox model for survival outcomes. Then we have the following lemma for survival outcomes.

**Lemma 2.3.3.** *Under the exchange of differential and expectation condition, an optimal solution to*

$$\underset{\mathbf{f}\in\mathcal{F}}{argmin} \ \mathbf{E}[\int_0^\tau \frac{\log \mathbf{E}[e^{\mathbf{f}^T\mathbf{w}}\mathbb{I}(Y \geq u)]}{\pi(A,\mathbf{x})} - \frac{\mathbf{f}^T\mathbf{w}}{\pi(A,\mathbf{x})} dN(u)] \tag{2.23}$$

*is the function $\mathbf{f}^*$ satisfying*

$$\exp(\mathbf{w}_i^T\mathbf{f}^*)\mathbf{E}[\Lambda^*(Y^{(i)})|\mathbf{x}, A=i] = \mathbf{P}[\delta = 1|\mathbf{x}, A=i] \tag{2.24}$$

*for a monotone nondecreasing function $\Lambda^*(u)$, where $N(u) = \mathbb{I}(\tilde{Y} \leq u)\delta$, and $\tau$ is a fixed time point with $\mathbf{P}[\tilde{Y} \geq \tau] > 0$. If the censoring time is non-informative and the censoring rate for each treatment group is the same, then*

$$argmax_{i\in\{1,\cdots,K\}} - \mathbf{w}_i^T f^* = argmax_{i\in\{1,\cdots,K\}}\mathbf{E}[\Lambda(Y)|\mathbf{x}, A=i]. \tag{2.25}$$

Using Lemma 2.3.3, the optimal decision rule for the survival outcome can be written as

$$d_0(\mathbf{x}) = \text{argmax}_{k\in\{1,\cdots,K\}}\mathbf{w}_i^T(-\mathbf{f}^*). \tag{2.26}$$

This is equivalent to fitting a weighted Cox Proportional Hazard (CPH) model with modified covariates $\mathbf{x}^* = \mathbf{x}\mathbf{w}^T$, by defining the hazard function as

$$\lambda(t|\mathbf{x}, A) = \lambda_0(t)e^{\text{Tr}(\mathbf{B}^T\mathbf{x}^*)}, \tag{2.27}$$

where $\lambda_0(t)$ is a baseline hazard function. Then we can estimate the coefficient matrix $\mathbf{B}$ by maximum likelihood estimation such as

$$\underset{\mathbf{B}\in\mathbb{R}^{p\times(K-1)}}{\text{argmin}} \quad l(\mathbf{B}) = \frac{1}{n}\sum_{i:\delta_i=1}\{-\frac{Y_i\text{Tr}(\mathbf{B}^T\mathbf{x}_i^*)}{\pi(A_i,\mathbf{x}_i)} + \frac{1}{\pi(A_i,\mathbf{x}_i)}\log\sum_{j:Y_j\geq Y_i}\exp(\text{Tr}(\mathbf{B}^T\mathbf{x}_i^*))\} + \lambda J(\mathbf{B}), \tag{2.28}$$

where $J(\mathbf{B})$ is either the mixed $l_1/l_2$ penalty or the nuclear norm penalty under different model assumptions. As the gradient of the Cox loss function for this model requires heavy computation, similar to Section 2.3.1, efficient group coordinate descent ((Breheny and Huang, 2015)) can be used to optimize (2.28) with the mixed $l_1/l_2$ penalty through vectorizing the modified covariates.

Note that the modified covariates $\mathbf{x}^*$ in Equation (2.27) contain the treatment information that can be incorporated into the baseline hazard function. Thus baseline hazard functions can be different for different treatments. For Lemma 2.3.3, we assume the censoring rate to be equal for all treatment groups so that our proposed method can be directly extended to the survival outcome. This assumption can possibly be removed by estimating the censoring rate for each group and then adjusting Equation (2.24).

## 2.4 Theoretical Properties of AD-learning

In this section, we show our proposed AD-learning is consistent under some mild conditions and establish finite value reduction bounds for our method. We first state the generalized margin condition used in our theory.

**Assumption 1.** For any $\epsilon > 0$, there exists some constants $C > 0$ and $\alpha > 0$ such that

$$\mathbf{P}[|(\mathbf{w}_i - \mathbf{w}_j)^T \mathbf{f}_0(\mathbf{x})| \leq \epsilon] \leq C\epsilon^{\alpha} \tag{2.29}$$

for every $i, j = 1, \cdots, K$.

Assumption 1 is an extension of margin condition used in binary classification problems to obtain sharper bounds on the excess 0-1 risk ((Audibert et al., 2007)). For our IDR problems, this generalized margin condition characterizes the behavior of the decision function vector $\mathbf{f}_0(\mathbf{x})$ around the boundary among different treatment regions, thus the level of difficulty in finding the optimal IDR. In the literature, (Zhao et al., 2012) used a similar assumption in the binary IDR problem. Using Assumption 1, we have the following theorem for the value reduction bound.

**Theorem 2.4.1.** *For the estimator $\hat{\mathbf{f}}_n$ by our proposed AD-learning and the corresponding IDR $\hat{d}_n$, we have*

$$V(d_0) - V(\hat{d}_n) \leq \frac{2K(K-1)}{1 - C(K)} (E||\mathbf{f}_0 - \hat{\mathbf{f}}_n||_2^2)^{\frac{1}{2}}. \tag{2.30}$$

*Furthermore, if Assumption 1 holds, we can improve the bound by*

$$V(d_0) - V(\hat{d}_n) \leq C_1(K, \alpha)(E||\mathbf{f}_0 - \hat{\mathbf{f}}_n||_2^2)^{\frac{1+\alpha}{2+\alpha}}, \tag{2.31}$$

*where $C_1(K, \alpha)$ is the constant that only depends on $K$ and $\alpha$.*

**Remark 1.** Based on (2.31), we can see that when $\alpha = 0$ and $C = 1$, Assumption (1) always holds for any $\epsilon > 0$. In this case, (2.31) reduces to (2.30). Based on (2.29), if $\alpha$ increases, the outcomes corresponding to various treatments become more different. As a result, the corresponding exponent $\frac{1+\alpha}{2+\alpha}$ becomes larger, and consequently a sharper bound in (2.31) can be obtained.

Theorem 2.4.1 gives an upper bound for the value function reduction in terms of the prediction error. For simplicity, we first consider Model (2.14) with equal $\pi(A_i, \mathbf{x}_i)$ for each treatment. Then we can use the main idea from (Lounici et al., 2009). We first vectorize the multiple responses and the coefficient $\mathbf{B}$ so that the model becomes

$$\underset{\beta \in \mathbb{R}^{p(K-1)}}{\operatorname{argmin}} \quad \frac{1}{n(K-1)} \sum_{k=1}^{K-1} (\mathbf{y}_k - \mathbf{X}\boldsymbol{\beta}_k)^T (\mathbf{y}_k - \mathbf{X}\boldsymbol{\beta}_k) + \lambda ||\boldsymbol{\beta}||_{2,1}, \tag{2.32}$$

where vector $\mathbf{y}_k = KR\mathbf{w}_k \in \mathbb{R}^n$ for $k = 1, \cdots, K-1$ and $\mathbf{X}$ is a design matrix with the $i$-th row being the $i$-th patient covariates $\mathbf{x}_i$. Denote each column of the coefficients $\mathbf{B}$ as $\boldsymbol{\beta}_k$, for $k = 1, \cdots, K-1$. Then $\boldsymbol{\beta} \in \mathbb{R}^{p(K-1)}$ is formed by stacking the coefficient $\boldsymbol{\beta}_k$, for $k = 1, \cdots, K-1$. We further define the $(K-1)n \times p(K-1)$ block diagonal matrix $\mathbf{Z}$ with its $k$-th block formed by the design matrix $X$.

We assume the underlying true $\mathbf{f}_0$ is linear with coefficient $\boldsymbol{\beta}_0$. Define $S(\boldsymbol{\beta}) = \{j : \beta_{kj} \neq 0, k = 1, \cdots, K-1\}$ and the cardinality of $S(\boldsymbol{\beta})$ as $||S(\boldsymbol{\beta})||_0$. We make the following two assumptions as in (Lounici et al., 2009). The first one is the Restricted Eigenvalue (RE) assumption considered by (Bickel et al., 2009) with an extension to the mixed $l_1/l_2$ norm.

**Assumption 2.** [RE(s)] For any nonzero $\boldsymbol{\beta}$ with $||S||_0 \leq s$ and $||\boldsymbol{\beta}_{S^c}||_{2,1} \leq 3||\boldsymbol{\beta}_S||_{2,1}$, there exists a positive real number $\rho(s)$ such that

$$\sqrt{\boldsymbol{\beta}\hat{\boldsymbol{\Sigma}}\boldsymbol{\beta}} \geq \rho(s)||\boldsymbol{\beta}_S||, \tag{2.33}$$

where $S$ denotes the short notation of $S(\boldsymbol{\beta})$ and $\hat{\boldsymbol{\Sigma}} = \frac{1}{n}\mathbf{Z}^T\mathbf{Z}$.

The next assumption is to control the stochastic error term in Model (2.14) with the bounded variance assumption.

**Assumption 3.** (1) Assume that the random error $e_{ki} = (y_{ki} - \mathbf{x}_i^T\boldsymbol{\beta}_k)$; $i = 1, \cdots, n$, $k = 1, \cdots, K-1$, are independent among different $i$ with mean zero and finite variance $\mathbf{E}[e_{ki}^2] \leq \sigma^2$.

(2) There exists a constant $c$ such that $\max_{1\leq i\leq n}\max_{1\leq j\leq p}|x_{ij}| \leq c$.

With the assumptions in place, we have the following theorem.

**Theorem 2.4.2.** *Consider Model* (2.14)*, for $p \geq 3$ and $K, n \geq 1$. Assume $S(\boldsymbol{\beta}_0) \leq s$, Assumptions 2 and 3 and the RE(2s) assumption hold. Let*

$$\lambda = \sigma\sqrt{\frac{(\log p)^{1+\delta}}{n(K-1)}},$$

*for any $\delta > 0$. Then with probability at least $1 - \frac{(2e\log p - e)c^2}{(\log p)^{1+\delta}}$, for the solution $\hat{\mathbf{B}}$ to the Model* (2.14)*, we have*

$$V(d_0) - V(\hat{d}_n) \leq \frac{\sqrt{K-1}K(K-1)}{1-C(K)}\frac{4\sqrt{10}c}{\rho^2(2s)}\sigma\sqrt{\frac{s(\log p)^{1+\delta}}{n}}. \tag{2.34}$$

*Furthermore, if Assumption 1 is satisfied, we can improve the bound by*

$$V(d_0) - V(\hat{d}_n) \leq C(K,\alpha)\frac{32}{\rho^2(s)}\sigma^2 s(\frac{(\log p)^{1+\delta}}{n})^{\frac{1+\alpha}{2+\alpha}}, \tag{2.35}$$

*where $C(K,\alpha)$ only depends on $K$ and the margin condition constant $\alpha$.*

Theorem 2.4.2 gives us the value reduction bound of order nearly $\frac{1}{n}$ as long as $\alpha$ is large enough. This value bound is consistent with $l_1$-PLS proposed by (Qian and Murphy, 2011) if we assume the underlying true function is linear. For a general function approximation, an additional approximation error to $\mathbf{f}_0(\mathbf{x})$ needs to be considered.

For Model (2.15), (Rohde et al., 2011) has obtained the same rate $\mathbf{O}(\frac{1}{n})$ for the prediction error and thus the order of value reduction bound for Model (2.15) is the same as Theorem

2.4.2. For Model (2.22), it can be regarded as usual logistic regression with modified covariates. If we consider the mixed $l_1/l_2$ penalty, error bounds of the same order were developed in (Meier et al., 2008). These results are applicable to our proposed AD-learning. However, to the best of our knowledge, the finite sample properties of other settings such as CPH models with the mixed $l_1/l_2$ penalty or low rank penalty require further developments and we leave it as the future work.

## 2.5 Simulation Study

In this section, we perform an extensive simulation study to investigate the finite sample performance of AD-learning for various types of outcomes. For all simulation settings, we consider four-armed ($K = 4$) randomized trials with equal probabilities of patients being assigned to each treatment group. For the low dimensional simulation setting, we set the sample size $n$ to be 200, 400, and 800. The number of covariates $p$ is set to be 20 and 40. For high dimensional simulation settings, we let the sample size be 400 and $p$ be 1000. Each simulation is repeated for 120 times. Additional simulation results are in the supplementary material, such as settings with $n = 200$, low rank decision function simulation studies, etc.

For the implementation details of AD-learning, two types of algorithms can be applied. The first one is the accelerated proximal gradient method. In particular, Models (2.14) and (2.15) can be represented as

$$\min F(\mathbf{B}) := L(\mathbf{B}) + \lambda J(\mathbf{B}), \tag{2.36}$$

where $L(\mathbf{B})$ is a smooth convex function with its gradient being Lipschitz continuous and $J(\mathbf{B})$ is a non-smooth convex function, of which the proximal operator can be computed efficiently. Then we can use the accelerated proximal gradient method to solve it with low computational complexity. It achieves the optimal converge rate $\mathbf{O}(\frac{1}{m^2})$ for gradient methods, where $m$ is the number of iterations for the algorithm. More details can be found in (Nesterov, 2013) and (Toh and Yun, 2010).

In binary and survival outcome settings, the gradient of function $L(\mathbf{B})$ may need large computational cost to calculate. To address the problem, the stochastic block coordinate decent algorithm can be applied instead when $J(\mathbf{B})$ is the mixed $l_1/l_2$ penalty. By using this algorithm, each gradient decent iteration can be efficiently computed. Thus the stochastic block coordinate decent algorithm may cost less time than the accelerated proximal gradient method.

The tuning parameter $\lambda$ is selected based on the cross-validation procedure. The criterion is to select $\lambda$ that maximizes the average of estimated value functions on the validation data set defined as

$$\hat{V}(d) = \frac{\mathbf{E}_n[R\mathbb{I}(A = d(\mathbf{x}))/\pi(A, \mathbf{x})]}{\mathbf{E}_n[\mathbb{I}(A = d(\mathbf{x}))/\pi(A, \mathbf{x})]}, \tag{2.37}$$

where $\mathbf{E}_n$ denotes the empirical average.

### 2.5.1 Study of Continuous Outcomes

When the clinical outcome $R$ is continuous, we generate our data from Model (2.7). Specifically, for $i = 1, \cdots, n$, let

$$R_i = \mu(\mathbf{x}_i) + \delta(\mathbf{x}_i) + \epsilon_i,$$

where $\delta(\mathbf{x}_i) = \sum_{k=1}^{K}(\mathbf{x}_i^T\boldsymbol{\beta}_k)\mathbb{I}(A = k)$, each covariate is generated by the uniform distribution from $-1$ to 1, and $\epsilon_i$ follows from the standard normal distribution. For each simulation scenario, we consider $\mu(\mathbf{x}) = 1 + X_1 + X_2$ and consider other types of main effect functions in the supplementary material. We design the following three interaction functions similar to those in (Zhou et al., 2017) and (Zhang et al., 2015):

1. $\delta(\mathbf{x}) = (1+X_1+X_2+X_3+X_4)\mathbb{I}(A = 1) + (1+X_1-X_2-X_3+X_4)\mathbb{I}(A = 2) + (1+X_1-X_2+X_3-X_4)\mathbb{I}(A = 3) + (1-X_1-X_2+X_3+X_4)\mathbb{I}(A = 4)$;

2. $\delta(\mathbf{x}) = (3\mathbb{I}(X_1 \leq 0.5)(\mathbb{I}(X_2 > -0.6)-1))\mathbb{I}(A = 1)+((\mathbb{I}(X_3 \leq 1))(2\mathbb{I}(X_4 \leq -0.3)-1)\mathbb{I}(A = 2) + (4\mathbb{I}(X_5 \leq 0) - 2)\mathbb{I}(A = 3) + (4\mathbb{I}(X_6 \leq 0) - 2)\mathbb{I}(A = 4)$;

3. $\delta(\mathbf{x}) = (0.2 + X_1^2 + X_2^2 - X_3^2 - X_4^2)\mathbb{I}(A = 1) + (0.2 + X_2^2 + X_3^2 - X_2^2 - X_4^2)\mathbb{I}(A = 2) + (0.2 + X_1^2 + X_4^2 - X_2^2 - X_3^2)\mathbb{I}(A = 3) + (0.2 + X_2^2 + X_3^2 - X_1^2 - X_4^2)\mathbb{I}(A = 4)$.

The first scenario corresponds to linear interaction effects. For the second scenario, we consider tree-type interaction effects. The last scenario includes polynomial interaction effects and we use degree 2 polynomials as basis functions for all methods. For each simulation scenario, we compare our proposed AD-learning using the group sparsity penalty with the following methods:

(1) $l_1$-PLS proposed by (Qian and Murphy, 2011) with basis $(1, \mathbf{x}, \mathbf{x}A)$;

(2) pairwise D-learning;

(3) the decision list (DL) method proposed by (Zhang et al., 2015);

(4) adaptive contrast weighted learning (ACWL-1 and ACWL-2) methods proposed by (Tao and Wang, 2017);

(5) the method of virtual twins (VT) proposed by (Foster et al., 2011),

where we use degree 2 polynomials as basis functions for all methods in the last scenario. Additional simulation study results on AD-learning using the low rank sparsity penalty are included in the supplementary material. In addition, we also perform the comparison between group $l_1$-PLS and $l_1$-PLS in the supplementary material, which shows little differences between $l_1$-PLS and group $l_1$-PLS in our simulation studies. This confirms our appropriate use of $l_1$-PLS instead of group $l_1$-PLS unless there are some prior information about strong group sparsity structures.

All the tuning parameters are selected via 10-fold cross-validation. We report the value functions and misclassification errors for $p = 40$ on 10000 independently generated test data in Table 2.1. From Table 2.1, we can see that our AD-learning has competitive performance among all methods. When we consider linear interaction effect, it is expected that our proposed AD-learning and $l_1$-PLS perform the best compared with other methods. In particular, our method will potentially be better than $l_1$-PLS because $l_1$-PLS suffers the mismatch problem discussed previously. For the second simulation scenario that corresponds to simple tree type interaction effect, while those tree based methods such as VT, DL and ACWL perform well, our method is still competitive. Similar results for $p = 20$ are included in the supplementary material. An interesting observation for this scenario is that although VT has the largest empirical value function among all methods, its misclassification rate is similar to that of our proposed method when $n = 400$. One potential reason is that VT is focused on model fitting while our method directly targets on decision rules. For the last scenario, since the basis functions we used correctly identify the interaction effect, our proposed AD-learning and $l_1$-PLS enjoy some advantages over other methods.

**Table 2.1:** Results of average means (standard deviations) of empirical value functions and misclassification rates for four continuous-outcome simulation scenarios with 40 covariates. The best value functions and misclassification rates are in bold.

| | $n = 400$ | | $n = 800$ | |
|---|---|---|---|---|
| | Value | Misclassification | Value | Misclassification |
| | | Scenario 1 | | |
| Pair-D | 2.67(0.06) | 0.49(0.02) | 3.01(0.02) | 0.32(0.02) |
| $l_1$-PLS | 3.05(0.04) | 0.24(0.01) | **3.15**(0.01) | 0.16(0.01) |
| DL | 2.6(0.04) | 0.54(0.01) | 2.78(0.02) | 0.47(0.01) |
| ACWL-1 | 2.69(0.05) | 0.46(0.01) | 2.9(0.02) | 0.37(0.01) |
| ACWL-2 | 2.77(0.05) | 0.43(0.01) | 3.02(0.01) | 0.31(0.01) |
| VT | 2.66(0.03) | 0.5(0.01) | 2.81(0.02) | 0.45(0.01) |
| Group-AD | **3.06**(0.05) | **0.22**(0.02) | 3.14(0.03) | **0.15**(0.02) |
| | | Scenario 2 | | |
| Pair-D | 2.84(0.12) | 0.32(0.04) | 2.93(0.1) | 0.3(0.03) |
| $l_1$-PLS | 2.93(0.11) | 0.36(0.04) | 3.01(0.1) | 0.32(0.04) |
| DL | 2.89(0.12) | 0.34(0.04) | 3.04(0.11) | 0.28(0.04) |
| ACWL-1 | 2.76(0.11) | 0.38(0.02) | 2.96(0.11) | 0.32(0.02) |
| ACWL-2 | 2.81(0.11) | 0.38(0.02) | 3.03(0.1) | 0.29(0.03) |
| VT | **3.07**(0.09) | **0.31**(0.02) | **3.12**(0.1) | **0.27**(0.02) |
| Group-AD | 2.97(0.1) | **0.31**(0.03) | 2.97(0.1) | 0.3(0.03) |
| | | Scenario 3 | | |
| Pair-D | 1.2(0.03) | 0.75(0.03) | 1.2(0.03) | 0.75(0.03) |
| $l_1$-PLS | 1.42(0.18) | 0.61(0.13) | 1.58(0.22) | 0.47(0.18) |
| DL | 1.38(0.08) | 0.64(0.06) | 1.5(0.08) | 0.57(0.06) |
| ACWL-1 | 1.29(0.08) | 0.7(0.04) | 1.49(0.07) | 0.56(0.05) |
| ACWL-2 | 1.3(0.07) | 0.69(0.04) | 1.57(0.06) | 0.51(0.05) |
| VT | 1.39(0.05) | 0.64(0.03) | 1.44(0.04) | 0.6(0.03) |
| Group-D | **1.57**(0.14) | **0.5**(0.11) | **1.76**(0.04) | **0.3**(0.05) |

### 2.5.2 Study of Binary and Survival Outcomes

For the binary outcome $R$, the dataset is independently generated by the logistic regression model

$$\text{logit}(\mathbf{P}[R_i = 1]) = \mu(\mathbf{x}_i) + \sum_{k=1}^{K} (\mathbf{x}_i^T \boldsymbol{\beta}_k)\mathbb{I}(A = k),$$

where the link function $\text{logit}(x) = \log \frac{x}{1-x}$. We consider same interaction effects as the first two scenarios of the continuous outcome simulation study.

Since pairwise D-learning and ACWL are not intended for the binary outcome, after modifying the $l_1$-PLS by using $l_1$ penalized logistic regression ($l_1$-PLR), we compare $l_1$-PLR, DL and VT with our AD-learning. Table 2.2 shows the value functions and misclassification rates for $p = 40$ and $n = 400, 800$. We can see that our proposed AD-learning has largest value functions and lowest misclassification rates in both scenarios. Moreover, there are some mismatches in model based methods such as $l_1$-PLS, where the misclassification rates and the value functions are both high. One potential reason is the mismatch between the optimization criterion and the tuning procedure in $l_1$-PLS. The other potential reason is the mismatch between minimizing prediction error and maximizing value function in model based methods.

**Table 2.2:** Results of average means (standard deviations) of empirical value functions and misclassification rates for two binary-outcome simulation scenarios with 40 covariates. The best value functions and misclassification rates are in bold.

| | $n = 400$ | | $n = 800$ | |
|---|---|---|---|---|
| | Value | Misclassification | Value | Misclassification |
| | | Scenario 1 | | |
| $l_1$-PLR | 0.88(0.01) | 0.58(0.02) | 0.91(0) | 0.45(0.02) |
| DL | 0.85(0.01) | 0.67(0.01) | 0.87(0.01) | 0.61(0) |
| VT | 0.84(0.01) | 0.68(0.01) | 0.84(0) | 0.69(0) |
| Binary-AD | **0.9**(0.01) | **0.44**(0.02) | **0.92**(0) | **0.32**(0.02) |
| | | Scenario 2 | | |
| $l_1$-PLR | 0.83(0.01) | 0.66(0.05) | 0.86(0) | 0.61(0.05) |
| DL | 0.81(0.01) | 0.53(0.01) | 0.85(0.01) | 0.44(0.01) |
| VT | 0.83(0.01) | 0.43(0.01) | 0.83(0.01) | 0.51(0) |
| Binary-AD | **0.86**(0.01) | **0.43**(0.04) | **0.87**(0.01) | **0.4**(0.04) |

Next we consider $R$ to be the outcome of time to event. The simulated data are generated by the following model with the exponential distribution

$$R_i = \exp(\lambda_i),$$

where exp denotes the exponential distribution and $\lambda_i = \mu(\mathbf{x}_i) + \sum_{k=1}^{K}(\mathbf{x}_i^T\boldsymbol{\beta}_k)\mathbb{I}(A = k)$ for $i = 1, \cdots, n$. The censoring time $C_i; i = 1, \cdots, n$, are generated from an exponential distribution with mean $\theta$ to induce around 25% censoring rate. We consider the same settings as those in the binary case. For comparisons, we apply the $l_1$ penalized CPH models and compare it with AD-learning, since other methods we use previously are not designed for the survival outcome. From Table 2.3 with $p = 40$, we can see that our proposed AD-learning has clear advantages over $l_1$-CPH. In addition, we also observe the mismatch phenomena of $l_1$-CPH in Scenario 2 of Table 2.3.

**Table 2.3:** Results of average means (standard deviations) of empirical value functions and misclassification rates for two survival-outcome simulation scenarios with 40 covariates. The best value functions and misclassification rates are in bold.

| | $n = 400$ | | $n = 800$ | |
|---|---|---|---|---|
| | Value | Misclassification | Value | Misclassification |
| | | Scenario 1 | | |
| $l_1$-CPH | 41.35(2.2) | 0.33(0.04) | 45.05(1.1) | 0.21(0.02) |
| Surv-AD | **43.91**(1.3) | **0.25**(0.02) | **45.56**(1.06) | **0.18**(0.01) |
| | | Scenario 2 | | |
| $l_1$-CPH | 21.95(0.63) | 0.57(0.04) | **23.21**(0.59) | 0.5(0.04) |
| Surv-AD | **22.1**(0.62) | **0.46**(0.02) | 22.78(0.53) | **0.44**(0.02) |

### 2.5.3 Study of High Dimensional Problems

We evaluate our AD-learning performance for high dimensional settings. We consider the sample size $n = 400$ so that each treatment group has roughly 100 patients and number of covariates $p = 800$. Scenarios 1-2, 3-4, 5-6 correspond to continuous, binary, and survival outcomes respectively. The interaction effects considered here are the same as the first two scenarios in the continuous setting in Section 5.1.

From Table 2.4, we can find that our proposed AD-learning performs better than $l_1$-PLS. One of the possible reasons is that our proposed method tends to select right covariates for the interaction effect function due to the direct learning of the decision rule. An interesting observation is that although pairwise D-learning has the lowest misclassification rate in Scenario 2, its corresponding value function is the lowest. This mismatch comes from the potential suboptimality of pairwise comparisons.

**Table 2.4:** Results of average means (standard deviations) of empirical value functions and misclassification rates for six high dimensional simulation scenarios. The best value functions and misclassification rates are in bold.

| | Method | Value | Misclassification |
|---|---|---|---|
| Scenario 1 | $l_1$-PLS | 5.3(0.02) | 0.17(0.01) |
| | Pair-D | 4.51(0.14) | 0.47(0.03) |
| | Group-AD | **5.31**(0.04) | **0.15**(0.02) |
| Scenario 2 | $l_1$-PLS | 5.64(0.03) | 0.22(0.01) |
| | Pair-D | 5.51(0.02) | **0.2**(0.01) |
| | Group-AD | **5.65(0.04)** | 0.21(0.01) |
| Scenario 3 | $l_1$-PLR | 0.88(0.02) | 0.64(0.04) |
| | Binary-AD | **0.92**(0.02) | **0.46**(0.06) |
| Scenario 4 | $l_1$-PLR | 0.84(0.01) | 0.7(0.02) |
| | Binary-AD | **0.87**(0.01) | **0.45**(0.03) |
| Scenario 5 | $l_1$-CPH | 771.35(126.2) | 0.41(0.09) |
| | Surv-AD | **1004.57**(40.19) | **0.2**(0.02) |
| Scenario 6 | $l_1$-CPH | 150.87(7.71) | 0.63(0.02) |
| | Surv-AD | **158.92**(4.73) | **0.45**(0.02) |

## 2.6 Real Data Applications

In this section, we perform a real data analysis to further evaluate our proposed AD-learning. We consider a clinical trial dataset from "AIDS Clinical Trials Group (ACTG) 175" in (Hammer et al., 1996) to study whether there is a subgroup of patients suitable for different combination treatments of AIDS. In this study, with equal probabilities, a total number of 2139 patients with HIV infection were randomly assigned into four treatment groups: zidovudine (ZDV) monotherapy, ZDV combined with didanosine (ddI), ZDV combined with zalcitabine (ZAL), and ddI monotherapy.

We choose 12 baseline covariates in our model: age (year), weight(kg), CD4+T cells amount at baseline, CD8 amount at baseline, Karnofsky score (scale at 0-100), gender (1 = male, 0 = female), race (1 = non white, 0 = white), homosexual activity (1 = yes, 0 = no), history of intravenous drug use (1 = yes, 0 = no), symptomatic status (1=symptomatic, 0=asymptomatic), antiretroviral history (1=experienced, 0=naive) and hemophilia (1=yes, 0=no). The first five covariates are continuous and have been scaled before estimation. The remaining seven covariates are binary categorical variables.

We consider two outcomes for our analysis. The first outcome is the difference between the early stage (around 25 weeks) CD4+ T (cells/mm$^3$) cell amount and the baseline CD4+ T cells prior to the trial. This was also studied in (Lu et al., 2013) and (Fan et al., 2017). Using this short term outcome, our goal is to use AD-learning to find the short term optimal IDR for each patient with AIDS among four treatment groups. We report the estimator of the coefficient $\mathbf{w}_i^T \mathbf{B}^T$ for each treatment in Table 2.5.

**Table 2.5:** Results of coefficients estimation for comparison functions.

| Variable Name (1-7) | ZDV | ZDV+ddI | ZDV+Zal | ddI |
|---|---|---|---|---|
| Intercept | $-49.86$ | 44.66 | $-3.53$ | 8.73 |
| Age | $-0.47$ | 4.33 | $-3.34$ | $-0.52$ |
| Weight | 0 | 0 | 0 | 0 |
| Karnofsky Score | 0 | 0 | 0 | 0 |
| CD4 baseline | 3.58 | $-14.79$ | $-14.78$ | 9.46 |
| Days pre-anti-retroviral therapy | 0 | 0 | 0 | 0 |
| Hemophilia | 0 | 0 | 0 | 0 |
| Homosexual activity | $-0.28$ | $-3.96$ | 0.65 | 3.60 |
| History of drug use | $-2.50$ | 8.20 | 4.03 | $-9.74$ |
| Race | 0 | 0 | 0 | 0 |
| Gender | 0 | 0 | 0 | 0 |
| Antiretroviral history | 0 | 0 | 0 | 0 |
| Symptomatic indicator | 0 | 0 | 0 | 0 |

In Table 2.5, we can see that four covariates including Age, CD4 baseline, homosexual activity and history of drug use, are identified to play an important role in our estimated optimal IDRs. These variables were also identified in the previous literature such as (Lu et al., 2013) and (Fan et al., 2017). According to the analysis in (Hammer et al., 1996), ZDV alone is inferior to the other treatments, which is also confirmed in our estimated IDR. Based on the CD4 change in the early stage, Zal treatment is generally not recommended in our finding with one possible reason that Zal has the most serious adverse event compared with ZDV and ddI ((Kakuda, 2000)). According to our estimated IDRs, those old patients with small amount of CD4 T cell baseline and having history of drug use but not homosexual activity, are recommended to take ZDV + ddI. The patients with large amount of CD4 T cell baseline and history of homosexual activity but not drug use history, are more advisable to take ddI alone.

To evaluate the performance of our proposed AD-learning, we randomly split the data into five folds and use four folds to train the model. We evaluate our method on the remaining one fold of data based on the empirical value function. We repeat this procedure for 1000 times. From Table 2.6, we can see our AD-learning has the largest value.

**Table 2.6:** Results of empirical value functions on one fold of testing data. The best empirical value function is in bold.

| $l_1$-PLS | Pair-D | DL | ACWL-1 | ACWL-2 | VT | AD low rank | AD group |
|---|---|---|---|---|---|---|---|
| 53.73 (0.33) | 57.17 (0.40) | 53.25 (0.47) | 52.74 (0.45) | 54.04 (0.45) | 54.84 (0.45) | 50.48 (0.38) | **59.69(0.39)** |

The second outcome is patients' time to event. Using this long term outcome, our second goal is to find the long term optimal IDR for patients among four treatment groups. The AIDS data consist of 2139 patient time to event responses with around 75% censor rate during the four-year long trial study. We use our proposed Model (2.23) to estimate the optimal IDR. We report the estimates of the coefficient $\mathbf{w}_i^T\mathbf{B}^T$ for each treatment of 12 covariates in Table 2.7. We can see that all covariates, except the indicator of homosexual activity and symptomatic, play an important role in the estimated optimal IDR. It may not be surprising because it is a long term study and thus more complicated. Since we model via the hazard function, the smaller the coefficient is, the longer the survival time is.

Compared with the previous finding based on the short term CD4 T cells amount, covariates including age, CD4 baseline and history of drug use have the similar effect on the ZDV + ddI and ddI alone treatments. In addition, we also find that ZDV + Zal treatment may not be good to take for the female patients with hemophilia, but may be suitable for the male patients with high Karnosky score and history of drug use. The estimated optimal IDR for other treatments can be interpreted in the similar way. In general, ZDV alone is always the least preferable among other treatments for patients and ZDV+ddI is always preferable for patients. Based on time to event outcome, ZDV + Zal is relatively more preferable than ddI alone. In addition, we evaluate our AD-learning with $l_1$-CPH using the same scheme based on value functions. Our AD-learning has an average value of 911.20, compared with the average value 905.02 for $l_1$-CPH.

**Table 2.7:** Results of coefficient estimation for survival time of failure.

| Variable Name (1-7) | ZDV | ZDV+ddI | ZDV+Zal | ddI |
|---|---|---|---|---|
| Age | 0.04 | −0.11 | 0.04 | 0.03 |
| Weight | 0.11 | 0.02 | 0.02 | −0.14 |
| Karnofsky Score | 0.06 | 0.03 | −0.09 | 0.01 |
| CD4 baseline | −0.04 | 0.04 | −0.00 | 0.00 |
| Days pre-anti-retroviral therapy | 0.09 | −0.07 | -0.04 | 0.02 |
| Hemophilia | 0.05 | −0.06 | 0.16 | −0.15 |
| Homosexual activity | 0.00 | 0.00 | 0.00 | 0.00 |
| History of drug use | 0.04 | −0.11 | −0.12 | 0.18 |
| Race | 0.03 | −0.04 | 0.01 | 0.01 |
| Gender | 0.31 | −0.08 | −0.16 | −0.07 |
| Antiretroviral history | 0.17 | −0.15 | 0.04 | −0.06 |
| Symptomatic Indicator | 0.00 | 0.00 | 0.00 | 0.00 |

## 2.7 Conclusion

In this chapter, we propose an AD-learning method to estimate the optimal IDRs in multiple treatment settings for various types of outcomes. Our proposed method provides a clear geometric interpretation about the relative effectiveness of treatments for patients, which is quantified by angles in the Euclidean space. Our proposed AD-learning is robust to model misspecification. By incorporating group or low rank sparsity, our AD-learning can further improve the estimation of decision rules and interpretation, especially for high dimensional settings. The competitive performance of our method has been demonstrated via the simulation studies and data applications.

Several possible extensions can be explored for future study. Our proposed method for the survival outcome is based on the non-informative censoring and Cox proportional hazard assumption. It will be interesting to develop methods for more complex settings. In order to use nonlinear functions to approximate $\mathbf{f}_0(\mathbf{x})$, we can use different types of basis functions such polynomials or wavelet functions. It will be also interesting to develop kernel methods for our AD-learning, such as multiple kernel learning ((Bach et al., 2004)). Finally, the current AD-learning focuses on a single decision point. It will be worthwhile to develop the corresponding methods for multiple decision points (Zhao et al., 2015a; Liu et al., 2018).

# CHAPTER 3
# Estimating Individualized Decision Rules with Tail Controls

## 3.1 Introduction

Decision making is a long standing research problem in many scientific areas, ranging from engineering, management science to statistics. In the era of big data, the traditional "one fits all" decision rules are no longer ideal in many applications due to data heterogeneity. A decision rule that works for certain subjects may not necessarily work for others. Motivated by this, it is desirable to make individualized decision rules (IDRs) that map from individual characteristics into available decision assignments. Developing effective IDRs has a wide range of applications. For example, a credit card company hopes to send a special offer for each targeted customer tailoring to his/her personal needs. An epidemiologist needs to decide whether to deliver a vaccine plan to a specific region in order to prevent the spread of diseases. From the statistical perspective, the majority of literature is focused on estimating the optimal IDR that can maximize the expected outcome or minimize the expected loss for each subject. However, some effort needs to be made to ensure reasonable outcomes for subjects falling in the tail of the distributions. Risk control needs to be taken into account to prevent adverse consequences. For example, only sending a special offer to a high-risk customer with potentially the largest expected profit may end up generating bad debt expenses for credit card company. The motivation of this paper comes from the IDR problems in precision medicine, also known as personalized medicine. One of the key goals in precision medicine is to develop better preventions and treatment methods that are tailored to each individual patient.

Prior work in precision medicine is focused on estimating the optimal IDR that can maximize expected outcome for each individual. Due to the complicated medical procedure, only targeting on the expected outcome of each patient may not be sufficient. Risk control is necessary to prevent adverse events, i.e., the heavy tail distribution of outcome. We consider a simple and motivating example to illustrate the importance of risk control, in addition to maximize the expected outcome. Figure 3.1 plots the conditional density of a random outcome $R$ under two treatments 1 and -1, given the patient's gender, i.e., male or female. Each curve corresponds to a different Gaussian density curve. If we only consider the optimal IDR that maximizes the expected outcome, treatment 1 is more suitable for female while treatment -1 is better for male. However, the gain is quite little since the mean difference is only 0.1, and thus it is hard to distinguish between these two treatments given the gender information. However, if we consider the effect of variation caused by each treatment, in order to protect each person from risky scenarios, then treatment 1 is more favorable than treatment -1 for male, and similarly, treatment -1 is more preferable than treatment 1 for female. This treatment rule may be more reasonable than the previous one because we do not want to give patients unstable, and potentially high risk treatments. We will revisit this example in our numerical studies for further illustrations.

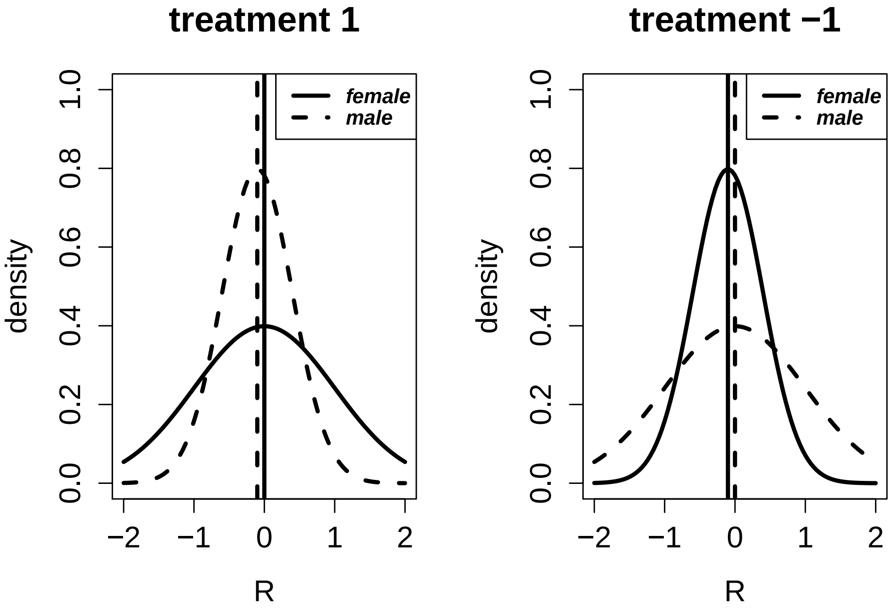


**Figure 3.1:** Plots of a motivating example. The dash and solid lines in the left plot show the probability densities of $\mathcal{N}$(-0.1, 0.5) and $\mathcal{N}(0,1)$ respectively. The dash and solid lines in the right plot correspond to the probability densities of $\mathcal{N}(0,1)$ and $\mathcal{N}$(-0.1, 0.5) respectively. In this example, male is more preferable to treatment 1, while female is more preferable to treatment -1.

Motivated by the conditional value at risk (CVaR) used extensively in finance and risk management, we propose two new criteria that consider the expected outcome and CVaR of outcome as a weighted combination to evaluate IDRs. The resulting IDR under our proposed criteria can optimize the outcome of each individual and control the risk jointly.

The main contributions of this chapter can be summarized as follows:

(a) We develop two innovative approaches to directly estimate the optimal IDR that can maximize the expected outcome while simultaneously control the average or individualized lower tail of the outcome;

(b) Two novel non-convex optimization algorithms are proposed to efficiently compute the solutions with convergence guarantee of the sharpest stationary points, based on some recent developments in optimization;

(c) We develop several important theoretical properties of our proposed methods related to statistical learning theory over two functional spaces such as two different reproducing kernel Hilbert spaces.

The remainder of this Chapter is organized as follows. In Section 3.2, supplementing the previous expected-value function framework, we introduce two new criteria to estimate the optimal IDR by using the concept of CVaR in risk management. We present several properties of our proposed criteria. In Section 3.3, we discuss our statistical estimation procedures to compute optimal IDRs under our proposed criteria. Two novel and efficient non-convex optimization algorithms are presented by using some recent developments in majorization-minimization (MM) and difference of convex algorithms (DCA). In Section 3.4, we establish several important theoretical properties of our methods by making use of statistical learning theory. We demonstrate our methods via extensive simulation studies and a data application in Sections 3.5 and 3.6, respectively. We conclude the paper and discuss some potential future work in Section 3.7. Some technical results are provided in the Appendix B.

### 3.2 Robust Criteria to Estimate Optimal IDRs

Consider a randomized clinical study in a binary-armed treatment setting. We observe each patient's covariate information: $\mathbf{x} = (\mathbf{x}_1, \cdots, \mathbf{x}_p)^T \in \mathcal{X}$, where $\mathcal{X} \subseteq \mathbb{R}^p$. Then each patient will receive a treatment $A \in \mathcal{A} = \{1, -1\}$ randomly. The outcome $R \in \mathbb{R}$ for each patient is measured after treatment. For theoretical simplicity, we assume $R \in \mathcal{R}$ is bounded. Without loss of generality, we assume that the larger $R$ indicates the better condition a patient is in. Define $\pi(a|x) = \mathbf{P}[A = a|\mathbf{x} = x]$ to be the probability of a patient being assigned treatment $a$ given the covariates of this patient. This probability is assumed to be known under a randomized clinical study or needs to be estimated in an observational study by various methods, such as logistic regression. We further assume $\pi(a|\mathbf{x}) > 0$ for $\mathbf{x} \in \mathcal{X}$ almost surely and every $a \in \mathcal{A}$. Furthermore, let $P$ be the probability distribution of a random triplet $(\mathbf{x}, A, R)$, under which the likelihood of $(\mathbf{x}, A, R)$ is defined as $f_0(x)\pi(a|x)f_1(r|x, a)$. In particular, $f_0(x)$ is the probability density function of $\mathbf{x}$ and $f_1(r|x, a)$ is the conditional probability density function of $R$ given $(A, \mathbf{x})$.

An IDR $d$ is defined as a measurable function mapping from the covariate space $\mathcal{X}$ into the treatment space $\mathcal{A}$. For any IDR $d$, define $P^d$ to be the probability measure where the action $A$ follows $d$. Then the probability density function under $P^d$ is defined as $f_0(x)\mathbb{I}(a = d(x))f_1(r|x, a)$, where $\mathbb{I}(\bullet)$ is the indicator function. For notational purpose, we let $L^r(\mathcal{T}, \mathcal{F}_1, P^d)$ be the space of all measurable functions such that $\int_{T\in\mathcal{T}} |f(T)|^r \mathrm{d}P^d < \infty$, where $\mathcal{F}_1$ is the corresponding $\sigma$-field generated by $\mathcal{T} := \mathcal{X} \times \mathcal{A} \times \mathcal{R}$.

#### 3.2.1 Expected Value Function Framework

Before introducing our new criterion and methods, we first present the existing expected-value function framework used by most existing methods, such as (Qian and Murphy, 2011) and (Zhao et al., 2012). The value function was defined in (Qian and Murphy, 2011) as

$$V(d) := \mathbf{E}^d[R] = \mathbf{E}\left[\frac{R\mathbb{I}(A = d(\mathbf{x}))}{\pi(A|\mathbf{x})}\right], \tag{3.1}$$

where the last equality is based on Radon-Nikodym theorem ((Qian and Murphy, 2011)). Note that $\mathbf{E}^d[c(\mathbf{x})] = \mathbf{E}[c(\mathbf{x})]$ for any measurable function $c(\mathbf{x})$. Based on this value function, an optimal IDR $d_0$ is defined as

$$d_0 \in \text{argmax}_d V(d). \tag{3.2}$$

Note that

$$\begin{aligned} V(d) &= \mathbf{E}[\mathbf{E}[R|\mathbf{x}, A = 1]\mathbb{I}(d(\mathbf{x}) = 1) + \mathbf{E}[R|\mathbf{x}, A = -1]\mathbb{I}(d(\mathbf{x}) = -1)] \\ &= \mathbf{E}[(\mathbf{E}[R|\mathbf{x}, A = 1] - \mathbf{E}[R|\mathbf{x}, A = -1])\mathbb{I}(d(\mathbf{x}) = 1)] + \mathbf{E}[\mathbf{E}[R|\mathbf{x}, A = -1]], \end{aligned}$$

and then as a result,

$$d_0(\mathbf{x}) \in \text{argmax}_{a \in \mathcal{A}} \mathbf{E}[R|\mathbf{x}, A = a], \tag{3.3}$$

almost surely. It is observed that under the expected-value function framework, the optimal IDR is to select the treatment with the largest expected outcome among all treatments for each patient.

Despite the progress of developing optimal IDRs in precision medicine, only focusing on obtaining the largest expected outcome for each individual may be too restrictive, especially in precision medicine. For example, doctors may want to know whether a treatment does the best to improve the worst scenario, in particular for a high risk patient. Without such risk consideration, this may lead to potentially severe events, such as exacerbation or hospitalization in practice. Similar concerns may happen in the credit card company, where the “best” policy should not only improve the average profit for the company, but also reduce the chance of incurring heavy loss. This motivates us to control risk exposure caused by decision rules, in addition to maximizing the expected outcome of each individual.

### 3.2.2 Conditional Value at Risk

It is natural to consider some robust metrics such as quantiles of $R$ given $\mathbf{x}$ and $A$ to measure the effect of a treatment. The corresponding optimal IDR $\tilde{d}$ under the quantile can be defined as

$$\tilde{d} \in \text{argmax}_d\, Q_\gamma(P^d), \tag{3.4}$$

where $Q_\gamma(P^d) = \inf\{\alpha : P^d[R < \alpha] \geq 1 - \gamma\}$ and $\gamma \in (0, 1)$, i.e., $\gamma$-quantile of $P^d$. However, the quantile makes the optimization Problem (3.4) hard to solve. Note that $Q_\gamma(P^d)$ is also called $\gamma$-Value at Risk (VaR), an important risk measure in finance ((Jorion, 2001)). In order to address the shortcomings such as the discouragement for diversification, (Artzner et al., 1999) studied an alternative risk measure called Conditional Value at Risk (CVaR), also known as the expected shortfall, average value at risk or expected tail loss. Consider a random outcome $Y$. The $\gamma$-CVaR of $Y$ is given by

$$S(F_Y) := \frac{1}{\gamma}\mathbf{E}[Y\mathbb{I}(Y \leq Q_\gamma(F_Y))], \tag{3.5}$$

where $F_Y$ is the corresponding probability distribution of $Y$. This CVaR can be interpreted as a truncated mean lower than $\gamma$-quantile of $Y$. As a remark, we note that $Y$ is often referred to as a loss in the finance literature. However, here we call $Y$ an outcome to be consistent with the IDR literature. CVaR has several nice properties such as coherence property ((Artzner et al., 1999)) and it is preferable to VaR ((Sarykalin et al., 2008b)). In addition, (Pflug, 2000) showed that $S(F_Y) \leq Q_\gamma(F_Y)$, a lower bound of $\gamma$-VaR. This implies that larger $\gamma$-CVaR of a random outcome indicates larger $\gamma$-VaR. Note that the reverse inequality does not necessarily hold. Interestingly, in addition to several nice properties related to risk measure, CVaR can be viewed as an optimal value of concave maximization by the celebrated work of (Rockafellar and Uryasev, 2000), which is defined as follows:

$$S(F_Y) = \sup_{\alpha \in \mathbb{R}} \left\{ \alpha - \frac{1}{\gamma}\mathbf{E}[(\alpha - Y)_+] \right\}, \tag{3.6}$$

where $[t]_+ = \max(0, t)$. When used in an optimization context, the CVaR-criterion is computationally much easier than the VaR-criterion. The leftmost of the optimal solution set to (3.6) is $Q_\gamma(F_Y)$ (Rockafellar and Uryasev, 2000, Theorem 1). Such a reformulation motivates us to propose a new criterion to study the IDR problem. For related theoretical discussions about CVaR, we refer to (afellar and Uryasev, 2002) and the references therein.

### 3.2.3 Robust Criteria for IDR Problems

In the following two subsections, we combine the existing value function framework with the concept of CVaR in order to incorporate risk control into consideration to estimate an optimal IDR.

#### 3.2.3.1 Average Lower Tail

Motivated by the usage of CVaR, we first propose a robust criterion that combines the value function defined in (4.1) and the lower tail of outcome $R$ by a weighted factor $\tau \in [0, 1]$. Specifically, this combined objective is:

$$M_1(d) := (1-\tau)V(d) + \tau\frac{1}{\gamma}\mathbf{E}^d[R\mathbb{I}(R \le Q_\gamma(P^d))], \tag{3.7}$$

where $0 \le \tau \le 1$. Note that

$$\begin{aligned} P^d(R < \alpha) &= \mathbf{E}^d[\mathbb{I}(R < \alpha)] \\ &= \mathbf{E}\left[\frac{\mathbb{I}(A = d(\mathbf{x}))}{\pi(A|\mathbf{x})}\mathbb{I}(R < \alpha)\right] \\ &= \mathbf{E}[\sum_{a \in \mathcal{A}} \mathbb{I}(d(\mathbf{x}) = a)P(R < \alpha|\mathbf{x}, A = a)] \\ &= \mathbf{E}[P(R < \alpha|\mathbf{x}, A = d(\mathbf{x}))]. \end{aligned} \tag{3.8}$$

Then $Q_\gamma(P^d)$ can be further expressed as

$$Q_\gamma(P^d) = \inf\left\{\alpha \,|\, \mathbf{E}[P(R < \alpha|\mathbf{x}, A = d(\mathbf{x}))] \ge \gamma\right\}, \tag{3.9}$$

which can be interpreted as the average $\gamma$-quantile of $R$ under the decision rule $d$. Correspondingly $\mathbf{E}^d[R\mathbb{I}(R \le Q_\gamma(P^d))]$ can be understood as $\gamma$-average CVaR. Then $M_1(d)$ in (3.7) can be regarded as a convex combination of the value function and the $\gamma$-average CVaR. Similar to (3.6), we can rewrite (3.7) as

$$M_1(d) = (1-\tau)V(d) + \tau\sup_{\alpha \in \mathbb{R}}\left\{\alpha - \frac{1}{\gamma}\mathbf{E}^d[(\alpha - R)_+]\right\}. \tag{3.10}$$

**Proposition 3.2.1.** *The following two statements hold:*

*(a)* $M_1(d) \leq (1-\tau)V(d) + \tau Q_\gamma(P^d)$*;*

*(b)* $M_1(d) \leq V(d)$*.*

*Proof.* Statement (a) is based on the result that CVaR is a lower bound of VaR ((Pflug, 2000)). For statement (b), note that $M_1(d)$ is increasing with respect to $\gamma$. Letting $\gamma = 1$ gives that $Q_\gamma(P^d) = \max_{\omega \in \mathcal{T}} R(\omega)$, where $\mathcal{T}$ is the corresponding event space related to $R$. This further implies that $\mathbf{E}^d[R\mathbb{I}(R \leq Q_1(P^d))] = V(d)$. Therefore, $M_1(d) \leq (1-\tau)V(d) + \tau V(d) = V(d)$. □

According to Proposition 3.2.1, $M_1(d)$ can be regarded as a lower bound of $V(d)$. Maximizing $M_1(d)$ can potentially maximize $V(d)$. Then the optimal IDR under our proposed robust criterion $M_1(d)$ is defined as

$$d_1 \in \text{argmax}_d \; M_1(d). \tag{3.11}$$

The interpretation of the optimal IDR with respect to $M_1(d)$ is to select a treatment with the largest convex combination of the value function and the $\gamma$-average CVaR. The representation of (3.10) gives us a way to compute the optimal IDR $d_1$ and $\alpha$ jointly via optimizing

$$(d_1, \alpha^*) \in \underset{\alpha \in \mathbb{R}, d}{\text{argmax}} \left\{ (1-\tau)V(d) + \tau(\alpha - \frac{1}{\gamma}\mathbf{E}^d[(\alpha - R)_+]) \right\}. \tag{3.12}$$

When $\tau = 0$, the combined objective reduces to original value function $V(d)$. When $\tau = 1$, it becomes $\gamma$-average CVaR of $R$ with respect to $P^d$. The choice of $\tau$ will be discussed later in our numerical studies.

#### 3.2.3.2 Individualized Lower Tails

One natural question is whether we can control the individualized $\gamma$-CVaR instead of average $\gamma$-CVaR of the outcome for subjects. Next, we propose another criterion as an extension of $M_1(d)$ in (3.10) from the average level to individualized level risk control:

$$\begin{aligned} M_2(d) &:= (1-\tau)V(d) + \tau \sup_{\alpha \in L^1(\mathcal{X}, \Xi, P_\mathbf{x})} \left\{ \mathbf{E}[\alpha(\mathbf{x})] - \frac{1}{\gamma}\mathbf{E}^d[(\alpha(\mathbf{x}) - R)_+] \right\} \\ &= (1-\tau)V(d) + \tau \sup_{\alpha \in L^1(\mathcal{X}, \Xi, P_\mathbf{x})} \left\{ \mathbf{E}^d[\alpha(\mathbf{x}) - \frac{1}{\gamma}(\alpha(\mathbf{x}) - R)_+] \right\}, \end{aligned} \tag{3.13}$$

where $\Xi$ is the $\sigma$-field generated by $\mathcal{X}$ and $P_X$ is the corresponding probability measure. In order to understand $M_2(d)$, we first characterize the optimal $\alpha^*$ in (3.13). Define the individualized $\gamma$-VaR as $Q_\gamma(R|\mathbf{x} = x, A = a) := \inf\{\alpha : \ P(R < \alpha(\mathbf{x}, A)|\mathbf{x} = x, A = a) \geq 1 - \gamma\}$ and individualized $\gamma$-CVaR as $\text{CVaR}_\gamma(R|\mathbf{x}, a) := \frac{1}{\gamma}\mathbf{E}[R\mathbb{I}(R \leq Q_\gamma(R|\mathbf{x}, a))|\mathbf{x}, A = a]$ given $\mathbf{x} = x$ and $A = a$. The following theorem gives an explicit expression of the optimal $\alpha^*$ by using the theory of variational analysis ((Rockafellar and Wets, 2009)).

**Theorem 3.2.1.** *Given any decision rule $d$, $\alpha^*$ is optimal to the optimization problem in $M_2(d)$ if and only if*

$$\alpha^*(\mathbf{x}) = Q_\gamma(R|\mathbf{x}, A = d(\mathbf{x})) \quad \textit{almost surely.} \tag{3.14}$$

*Thus*

$$M_2(d) = (1-\tau)V(d) + \tau\mathbf{E}[CVaR_\gamma(R|\mathbf{x}, A = d(\mathbf{x}))]. \tag{3.15}$$

*Proof.* The proof of the claims require the concept of normal integrand and decomposable space from (Rockafellar, 1976) so that we can interchange between supreme operator and expectation in (3.13). Leaving the proof of this interchangeability in the supplementary material, we complete the proof of the theorem as follows.

By definition of $M_2(d)$ in (3.13), we have

$$\begin{aligned}
M_2(d) &= (1-\tau)V(d) + \tau \sup_{\alpha \in L^1(\mathcal{X},\Xi,P_\mathbf{x})} \left\{\mathbf{E}^d[\alpha(\mathbf{x}) - \frac{1}{\gamma}(\alpha(\mathbf{x}) - R)_+]\right\} \\
&= (1-\tau)V(d) + \tau \sup_{\alpha \in L^1(\mathcal{X},\Xi,P_\mathbf{x})} \left\{\mathbf{E}\left[\frac{\mathbb{I}(A = d(\mathbf{x}))}{\pi(A|\mathbf{x})}\left(\alpha(\mathbf{x}) - \frac{1}{\gamma}(\alpha(\mathbf{x}) - R)_+\right)\right]\right\} \\
&= (1-\tau)V(d) + \tau \sup_{\alpha \in L^1(\mathcal{X},\Xi,P_\mathbf{x})} \left\{\mathbf{E}\left[\mathbf{E}\left[\alpha(\mathbf{x}) - \frac{1}{\gamma}(\alpha(\mathbf{x}) - R)_+\middle| \mathbf{x}, A = d(\mathbf{x})\right]\right]\right\} \\
&= (1-\tau)V(d) + \tau \left\{\mathbf{E}\left[\sup_{s\in\mathbb{R}}\left\{s - \frac{1}{\gamma}\mathbf{E}\left[(s - R)_+|\, \mathbf{x}, A = d(\mathbf{x})\right]\right\}\right]\right\} \\
&= (1-\tau)V(d) + \tau\mathbf{E}\left[\text{CVaR}_\gamma(R|\mathbf{x}, A = d(\mathbf{x}))\right],
\end{aligned}$$

with the optimal solution $\alpha^*$ to be $Q_\gamma(R|\mathbf{x}, A = d(\mathbf{x}))$ almost surely by the definition of CVaR. □

According to (3.14), the explicit form of $\alpha^*(\mathbf{x})$ in $M_2(d)$ can be interpreted as the individual $\gamma$-quantile by the decision rule $d$ and $M_2(d)$ in (3.15) takes each individualized CVaR into consideration. We have the following proposition to further illustrate some properties of $M_2(d)$.

**Proposition 3.2.2.** *The following two inequalities hold: $M_1(d) \leq M_2(d) \leq V(d)$.*

*Proof.* The first inequality follows the fact that any constant is an element of $L^1(\mathcal{X}, \Xi, P_{\mathbf{x}})$. The second inequality is similar to $(b)$ in Proposition 3.2.1. □

The first inequality in Proposition 3.2.2 indicates that $M_2(d)$ improves $M_1(d)$ by extending $\alpha$ to incorporate the covariates information $\mathbf{x}$. The second inequality in Proposition 3.2.2 justifies the conservativeness of $M_2(d)$ as a lower bound of $V(d)$. In addition, since $\mathrm{CVaR}_\gamma(R|\mathbf{x}, a) \leq Q_\gamma(R|\mathbf{x}, a)$, one can have $M_2(d) \leq (1-\tau)V(d) + \tau\mathbf{E}[Q_\gamma(R|\mathbf{x}, d(\mathbf{x}))]$. The optimal IDR under $M_2(d)$ is defined as

$$d_2 \in \mathrm{argmax}_d M_2(d). \tag{3.16}$$

**Proposition 3.2.3.** *The optimal IDR under the criterion $M_2(d)$ is given by*

$$d_2(\mathbf{x}) \in \mathit{argmax}_{a \in A} \left\{(1-\tau)\mathbf{E}[R|\mathbf{x}, A = a] + \tau\, \mathit{CVaR}_\gamma(R|\mathbf{x}, a)\right\} \quad \textit{almost surely.}$$

Under $M_2(d)$, the optimal IDR $d_2$ is equivalent to choosing a treatment that has the largest convex combination of the expected outcome and the individualized $\gamma$-CVaR among all treatments. Similar to (3.10), we can compute $\alpha^*(\mathbf{x})$ and $d_2$ jointly via

$$(d_2, \alpha^*) \in \underset{d,\, \alpha \in L^1(\mathcal{X}, \Xi, P_{\mathbf{x}})}{\mathrm{argmax}} \left\{(1-\tau)V(d) + \tau(\mathbf{E}[\alpha(\mathbf{x})] - \frac{1}{\gamma}\mathbf{E}^d[(\alpha(\mathbf{x}) - R)_+])\right\}. \tag{3.17}$$

Although $M_1(d)$ can be viewed as a special case of $M_2(d)$ by letting $\alpha(\mathbf{x})$ to be a constant independent of $\mathbf{x}$, the interpretation is substantially different and each has its own significance as a criterion in choosing an optimal decision rule.

### 3.2.4 Duality Representation

Note that both $M_1(d)$ and $M_2(d)$ involve concave maximization. Thus it would be useful to investigate the dual representation of both $M_1(d)$ and $M_2(d)$ by making use of convex duality theory in (Rockafellar, 1974). To begin with, we first define the following two sets:

$$\mathcal{W}_1^d := \{W \in L^1(\mathcal{T}, \mathcal{F}_1, P^d) \,|\, \mathbf{E}^d[W] = 1,\, \varepsilon_1 \leq W(\omega) \leq \varepsilon_2, \text{ for almost sure } \omega_1 \in \mathcal{T}\}, \quad (3.18)$$

and

$$\mathcal{W}_2^d := \{W \in L^1(\mathcal{T}, \mathcal{F}_1, P^d) \,|\, \epsilon_1 \leq W(\omega_1) \leq \varepsilon_2 \text{ for almost sure } \omega_1 \in \mathcal{T},\, \mathbf{E}[W|\mathbf{x}, A = d(\mathbf{x})] = 1\}, \quad (3.19)$$

where $\varepsilon_1 = 1 - \tau$ and $\varepsilon_2 = 1 - \tau + \frac{\tau}{\gamma}$. It is noted that $0 < \varepsilon_1 < 1$ and $\varepsilon_2 > 1$. We have the following theorem that gives the dual representation of $M_2(d)$.

**Theorem 3.2.2.** *$M_1(d) = \inf_{W \in \mathcal{W}_1^d} \mathbf{E}^d[WR]$ and $M_2(d) = \inf_{W \in \mathcal{W}_2^d} \mathbf{E}^d[WR]$.*

**Duality representation of $M_2(d)$:** According to the proof for duality representation of $M_2(d)$, we can define a conditional probability measure $P_{|\mathbf{x}}^W(B) = \int_B W dP_{|\mathbf{x}}^d$ for any measurable set $B \in \mathcal{T}$, where $W \in \mathcal{W}_2^d$, and then $W = \frac{dP_{|\mathbf{x}}^W}{dP_{|\mathbf{x}}^d}$. Define

$$\zeta(u) = \begin{cases} 0 & \text{if } \varepsilon_1 \leq u \leq \varepsilon_2 \\ +\infty & \text{otherwise} \end{cases},$$

Then we can further rewrite $\mathbf{E}^d[RW]$ in $M_2(d)$ for $W \in \mathcal{W}_1^d$ as

$$\begin{aligned} \mathbf{E}^d[R\mathbf{w}] &= \mathbf{E}^d[R\frac{dP_{|\mathbf{x}}^{\mathbf{w}}}{dP_{|\mathbf{x}}^d}] + \mathbf{E}\left[\zeta(\frac{dP_{|\mathbf{x}}^{\mathbf{w}}}{dP_{|\mathbf{x}}^d})\right] \\ &= \mathbf{E}_{\mathbf{x}}[\mathbf{E}_{P_{|\mathbf{x}}^{\mathbf{w}}}[R]] + \mathbf{E}_{\mathbf{x}}\left[\int \zeta(\frac{dP_{|\mathbf{x}}^{\mathbf{w}}}{dP_{|\mathbf{x}}^d}) dP_{|\mathbf{x}}^d\right] \\ &= \mathbf{E}_{P^{\mathbf{w}}}[R] + \mathbf{E}_{\mathbf{x}}\left[I_\zeta(\frac{dP_{|\mathbf{x}}^{\mathbf{w}}}{dP_{|\mathbf{x}}^d})\right], \end{aligned}$$

where $I_\zeta(\cdot)$ can be interpreted as the $f$-divergence distance between $P_{|\mathbf{x}}^{\mathbf{w}}$ and $P_{|\mathbf{x}}^d$. Then $M_2(d) = \inf_{P_{|\mathbf{x}}^{\mathbf{w}} \ll P_{|\mathbf{x}}^d} \mathbf{E}_{P^{\mathbf{w}}}[R] + \mathbf{E}_{\mathbf{x}}\left[I_\zeta(\frac{dP_{|\mathbf{x}}^{\mathbf{w}}}{dP_{|\mathbf{x}}^d})\right]$, where $u \ll v$ means that the probability measure $u$ is absolutely continuous with respect to the probability measure $v$. Thus the optimal IDR can also be written as

$$d_2 \in \text{argmax}_d \left\{ \inf_{P^{\mathbf{w}}_{|\mathbf{x}} \ll P^d_{|\mathbf{x}}} \mathbf{E}_{P^{\mathbf{w}}}[R] + \mathbf{E}_{\mathbf{x}} \left[ \; I_\zeta(\frac{dP^{\mathbf{w}}_{|\mathbf{x}}}{dP^d_{|\mathbf{x}}}) \right] \right\},$$

which can be interpreted as choosing an optimal decision rule with the highest worst expected outcome within the $f$-divergence distance from the original distribution $P^d$.

According to our problem setting, the density under $P^d_{|\mathbf{x}}$ is $\mathbb{I}(d(x) = a) f_1(r|x, a)$. Since $P^{\mathbf{w}}_{|\mathbf{x}} \ll P^d_{|\mathbf{x}}$, then the density under $P^{\mathbf{w}}_{|\mathbf{x}}$ should be $\mathbb{I}(d(x) = a) w(r|x, a)$ for some conditional probability density $w(r|x, a)$. Then we can have $\frac{dP^{\mathbf{w}}_{|\mathbf{x}}}{dP^d_{|\mathbf{x}}} = \frac{w(r|x,a=d(x))}{f_1(r|x,a=d(x))}$ by the chain rule. Therefore, we can further rewrite $M_2(d)$ as

$$M_2(d) = \inf_{P^{\mathbf{w}}} \left\{ \mathbf{E}_{P^{\mathbf{w}}}[R] \,|\, P^{\mathbf{w}}_{|\mathbf{x}} \ll P^d_{|\mathbf{x}}, \;\; \varepsilon_1 \le \frac{w(r|x, a = d(x))}{f_1(r|x, a = d(x))} \le \varepsilon_2, \text{ almost surely} \right\}. \qquad (3.20)$$

This gives us a natural link to distributionally robust statistical models that can evaluate a decision rule under ambiguity. Maximizing $M_2(d)$ over the decision rule $d$ is equivalent to identifying an optimal IDR that is robust to the contamination of outcome $R$ characterized by a probability constraint set.

**Duality representation of** $M_1(d)$: Similarly, for $V \in \mathcal{W}_1^d$, if we define $P^V(B) = \int_B V \mathrm{d}P^d$ for any measurable set $B \in \mathcal{T}$, then $V = \frac{\mathrm{d}P^V}{\mathrm{d}P^d}$. Thus the optimal IDR can also be written as

$$d_1 \in \text{argmax}_d \left\{ \inf_{P^V \ll P^d} \mathbf{E}_{P^V}[R] + I_\zeta(\frac{dP^V}{dP^d}) \right\},$$

where $I_\zeta(\frac{dP^V}{dP^d}) = \mathbf{E}^d \left[ \zeta(\frac{dP^V}{dP^d}) \right]$. Moreover, the probability density $V$ with respect to $P^d$ can be written as $\frac{v_0(x)v_1(r|x,a=d(x))}{f_0(x)f_1(r|x,a=d(x))}$ by the chain rule, according to our problem setting. Therefore, we can also express $M_1(d)$ as

$$M_1(d) = \inf_{P^V} \left\{ \mathbf{E}_{P^V}[R] \,|\, P^V \ll P^d, \; \varepsilon_1 \le \frac{v_0(x)v_1(r|x, a = d(x))}{f_0(x)f_1(r|x, a = d(x))} \le \varepsilon_2, \text{ almost surely} \right\}. \quad (3.21)$$

Maximizing $M_1(d)$ over the decision rule $d$ is equivalent to identifying an optimal IDR that is robust to the contamination of both outcome $R$ and covariate information $\mathbf{x}$ characterized by a probability constraint set.

**Comparisons between $M_1(d)$ and $M_2(d)$**: From a duality representation perspective, we can see $M_1(d)$ and $M_2(d)$ have substantial differences with regard to their robustness. The "minimax" sense of $M_1(d)$ in (3.21) considers the scenario where both distributions of the covariates $\mathbf{x}$ and outcome $R$ are perturbed from true underlying distributions. For $M_2(d)$, Proposition 3.2.2 shows that $M_2(d) \geq M_1(d)$, which means considering individualized CVaR improves the outcome of a given decision rule $d$. At the same time, however, it also indicates $M_2(d)$ is not as conservative as $M_1(d)$. This can also be justified by the "minimax" representation of (3.20), which considers the contamination of outcome $R$. In the end, both $M_1(d)$ and $M_2(d)$ are more robust than the expected-value framework, i.e., $V(d)$. Therefore $M_1(d)$ and $M_2(d)$ may have the ability to improve generalization.

### 3.3 Statistical Estimation and Optimization

In this section, we discuss the estimation and optimization procedures for Problems (3.10) and (3.17) respectively given observed data. Before that, we first introduce some definitions related to the algorithm convergence of non-convex optimization problems.

Let $\Phi : \mathbb{R}^n \to \mathbb{R}$. The directional derivative of $\Phi$ at a point $x \in \mathbb{R}^n$ along the direction $v \in \mathbb{R}^n$ is given by

$$\Phi'(x, v) = \lim_{\tau \downarrow 0} \frac{\Phi(x + \tau v) - \Phi(x)}{\tau}. \tag{3.22}$$

We say $x_0$ is a directional-stationary (d-stationary) point of $\Phi$ on $\mathbb{R}^n$ if

$$\Phi'(x_0, x - x_0) \geq 0, \ \forall x \in \mathbb{R}^n. \tag{3.23}$$

For a directionally differentiable optimization problem, d-stationary points can be viewed as the first order "sharpest" ones among different kinds of stationary points including Clarke points ((Pang et al., 2016)), and the condition (3.23) is the least relaxed among other types of stationarity conditions. In the following subsections, we develop two algorithms to compute d-stationary points of Problems (3.10) and (3.17) respectively, which is the best we can achieve for non-convex optimization problems in practice.

### 3.3.1 Estimation of Optimal IDRs under $M_1(d)$

The optimization in (3.10) can be further rewritten as

$$\max_{\alpha\in\mathbb{R},d\in\mathcal{D}} \mathbf{E}[\frac{((1-\tau)R+\tau(\alpha-\frac{(\alpha-R)_+}{\gamma}))\mathbb{I}(A=d(\mathbf{x}))}{\pi(A|\mathbf{x})}], \tag{3.24}$$

where $\mathcal{D}$ is some classes of decision rules such as the linear ones.

Consider the binary treatment setting and let $d(\mathbf{x}) = \text{sign}(f(\mathbf{x}))$. Suppose we observe independently and identically distributed data $(\mathbf{x}_i, A_i, R_i); i = 1, \cdots, n$, then we can estimate the optimal IDR via empirical approximation:

$$\max_{\alpha\in\mathbb{R},d\in\mathcal{D}} \frac{1}{n}\sum_{i=1}^{n} \frac{\mathbb{I}(A_i = \text{sign}(f(\mathbf{x}_i)))}{\pi(A_i|\mathbf{x}_i)}((1-\tau)R_i+\tau(\alpha-\frac{(\alpha-R_i)_+}{\gamma})). \tag{3.25}$$

It is well known that optimization over indicator functions is NP hard. Alternatively, we can replace the 0-1 loss function by the following smooth truncated loss,

$$S(u) = \begin{cases} 0 & \text{if } u \le -\delta \\ (1+u/\delta)^2 & \text{if } 0 > u \ge -\delta \\ 2-(1-u/\delta)^2 & \text{if } \delta > u \ge 0 \\ 2 & \text{if } u > \delta, \end{cases}$$

and then use a functional margin representation to express $\mathbb{I}(A_i = \text{sign}(f(\mathbf{x}_i))$ as $\mathbb{I}(A_i f(\mathbf{x}_i) > 0)$ for each $i$. The corresponding function plot of $S(u)$ is shown in Figure 3.2 with $\delta = 1$. From the plot, we can see that the smooth approximation $\frac{S(u)}{2}$ is very close to the 0-1 loss. The parameter $\delta$ can control the closeness of this approximation. In practice, we can simply choose $\delta = 1$.

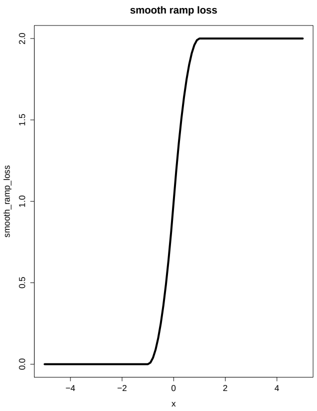


**Figure 3.2:** Plot of smooth surrogate loss function with $\delta = 1$

Let $\mathcal{H}$ be a Reproducing Kernel Hilbert Space (RKHS), then we can estimate the optimal IDR under the mixed value function $M_1(d)$ via computing

$$\min_{\alpha \in \mathbb{R}, f \in \mathcal{H}} \frac{1}{n} \sum_{i=1}^{n} \frac{S(A_i f(\mathbf{x}_i))}{\pi(A_i|\mathbf{x}_i)} (-(1-\tau)R_i - \tau(\alpha - \frac{(\alpha - R_i)_+}{\gamma})) + \frac{\lambda_1}{2}||f||^2_{\mathcal{H}}, \tag{3.26}$$

where $|| \bullet ||$ is the semi-norm in $\mathcal{H}$ and it is used to prevent over-fitting. The estimated IDR is given by $\hat{d}_1(\mathbf{x}) = \text{sign}(\hat{f}(\mathbf{x}))$. Note that Problem (3.26) involves a non-convex and potentially non-smooth optimization problem. Recent development in difference-of-convex (DC) optimization ((Pang et al., 2016)) motivates us to use DC programming to efficiently solve it. Note that $S(u)$ can be expressed as a difference of convex differentiable functions: $S_1(u) - S_2(u)$ where

$$S_1(u) = \begin{cases} 0 & \text{if } u \leq -\delta \\ (1 + u/\delta)^2 & \text{if } -\delta < u \leq 0 \\ 2 + 2u/\delta - 1 & \text{if } 0 < u \end{cases},$$

and

$$S_2(u) = \begin{cases} 0 & \text{if } u \leq 0 \\ (u/\delta)^2 & \text{if } 0 < u \leq \delta \\ 2u/\delta - 1 & \text{if } u > \delta \end{cases}.$$

Define

$$G^{(1)}(f,\alpha) := \frac{1}{n}\sum_{i=1}^{n}\frac{S(A_i f(\mathbf{x}_i))}{\pi(A_i|\mathbf{x}_i)}(-(1-\tau)R_i - \tau(\alpha - \frac{(\alpha - R_i)_+}{\gamma})) + \frac{\lambda_1}{2}||f||^2_{\mathcal{H}}, \tag{3.27}$$

and

$$G_j^{(1)}(f) = \frac{1}{n}\sum_{i=1}^{n}\frac{S(A_i f(\mathbf{x}_i))}{\pi(A_i|\mathbf{x}_i)}(-(1-\tau)R_i - \tau(R_j - \frac{(R_j - R_i)_+}{\gamma})). \tag{3.28}$$

The following proposition gives us a way to express (3.26) as a DC function.

**Proposition 3.3.1.** *The following two optimization problems have the same optimal value, i.e.,*

$$\min_{\alpha\in\mathbb{R}, f\in\mathcal{H}} G^{(1)}(f,\alpha) = \min_{f\in\mathcal{H}}\left\{\tilde{G}^{(1)}(f)\right\}, \tag{3.29}$$

*where $\tilde{G}^{(1)}(f) := \min_{1\le j\le n}\{G_j^{(1)}(f)\} + \frac{\lambda}{2}||f||^2_{\mathcal{H}}$. More importantly, the optimal solution sets of $f$ to both problems are the same.*

*Proof.* Note that for any given $f$, $G^{(1)}(f,\alpha)$ is a convex piecewise affine function with respect to $\alpha$, thus the optimal solution set $\alpha^*$ should contain one of the knots, i.e., $R_1, \cdots, R_n$. Then it follows that

$$\min_{\alpha\in\mathbb{R}} G^{(1)}(f,\alpha) = \min_{j\in\{1,\cdots,n\}} G_j^{(1)}(f) + \frac{\lambda_1}{2}||f||^2_{\mathcal{H}}.$$

Thus

$$\min_{f\in\mathcal{H},\alpha\in\mathbb{R}} G^{(1)}(f,\alpha) = \min_{f\in\mathcal{H}}\left\{\min_{1\le j\le n}\{G_j(f)\} + \frac{\lambda}{2}||f||^2_{\mathcal{H}}\right\},$$

and correspondingly

$$\text{argmin}_{f\in\mathcal{H}} G_1(f,\alpha) = \text{argmin}_{f\in\mathcal{H}}\left\{\min_{1\le j\le n}\{G_j(f)\} + \frac{\lambda}{2}||f||^2_{\mathcal{H}}\right\}.$$

□

Based on Proposition 3.3.1, instead of solving (3.26), we can equivalently solve the optimization problem in the right hand side of (3.29). Let $c_{ij} = \frac{-(1-\tau)R_i - \tau(R_j - \frac{(R_j - R_i)_+}{\gamma})}{\pi(A_i|\mathbf{x}_i)}$ for $i = 1, \cdots, n$ and $j = 1, \cdots, n$, and note that $c_{ij}$ is not necessarily nonnegative. Recall that

$S(u) = S_1(u) - S_2(u)$. Then we can further rewrite $G_j^{(1)}(f)$ as

$$\begin{aligned} G_j^{(1)}(f) = & \frac{1}{n}\sum_{i=1}^{n}(\max(c_{ij},0)S_1(A_i f(\mathbf{x}_i)) + \max(-c_{ij},0)S_2(A_i f(\mathbf{x}_i))) \\ & - \frac{1}{n}\sum_{i=1}^{n}(\max(c_{ij},0)S_2(A_i f(\mathbf{x}_i)) + \max(-c_{ij},0)S_1(A_i f(\mathbf{x}_i))) \\ & := F_j(f) - H_j(f), \end{aligned} \tag{3.30}$$

where both $F_j(f)$ and $H_j(f)$ are convex functions with respect to $f$ for $j = 1, \cdots, n$. Then we can further decompose

$$\begin{aligned} \tilde{G}^{(1)}(f) = & \min_{1\leq j\leq n}\{F_j(f) - H_j(f)\} + \frac{\lambda}{2}||f||_{\mathcal{H}}^2 \\ = & \sum_{i=1}^{n} F_j(f) - \max_{1\leq j\leq n}\{H_j(f) + \sum_{k\neq j}^{n} F_k(f)\} + \frac{\lambda}{2}||f||_{\mathcal{H}}^2 \\ := & F(f) - \max_{1\leq j\leq n} h_j(f) + \frac{\lambda}{2}||f||_{\mathcal{H}}^2, \end{aligned} \tag{3.31}$$

as a DC function, where $h_j(f) := H_j(f) + \sum_{k\neq j}^{n} F_k(f)$. Note that $\tilde{G}^{(1)}(f)$ is a potentially non-smooth function if there exits multiple $k$'s such that $h_k(f) = \max_{1\leq j\leq n} h_j(f)$. As pointed by (Pang et al., 2016), traditional DC programming cannot guarantee the convergence to a d-stationary point of the optimization problem (3.31) and may potentially lead to nonsense points. A failure example by traditional DC programming is given in (Pang et al., 2016). Let $\mathcal{M}_\epsilon(f) := \{j \mid h_j(f) \geq \max_{1\leq k\leq n} h_k(f) - \epsilon\}$, i.e., "$\epsilon$-argmax" index set. Motivated by (Pang et al., 2016), we propose the following enhanced probabilistic DCA summarized in Algorithm 1 below.

**Algorithm 1** Algorithm for (3.26)

```
Given a fixed $\epsilon > 0$, let $f^{(v)}$ be the solution at the $v$ iteration.
Randomly select $j \in \mathcal{M}_\epsilon(f^{(v)})$, and compute
    $$f^{(v+1)} \in \text{argmin}_{f\in\mathcal{H}}\{F(f) - \frac{\partial h_j(f^{(v)})}{\partial f}(f - f^{(v)}) + \frac{\lambda}{2}||f||_{\mathcal{H}}^2\}. \tag{3.32}$$
The algorithm stops when $|\tilde{G}^{(1)}(f^{(v)}) - \tilde{G}^{(1)}(f^{(v+1)})| < \kappa$, for some pre-specified positive constant $\kappa$.
```

The proof of convergence to d-stationary points by the above algorithm can be found in (Pang et al., 2016). For the computation of the subproblem (3.32), efficient algorithms such as quasi-newton methods can be used. If we consider that $f$ belongs to a class of linear functions, we can also compute the solution of (3.26) with the $l_1$ penalty replacing the RKHS norm. Next, we discuss how to estimate optimal IDRs under $M_2(d)$.

### 3.3.2 Estimation of Optimal IDRs under $M_2(d)$

Similar to the previous one in Section 3.1, we can first rewrite the optimization Problem (3.17) as

$$\max_{\substack{\alpha\in L^1(\mathbf{x},\Xi,\mathbf{P_x})\\ d\in\mathcal{D}}} \mathbf{E}\left[\frac{((1-\tau)R+\tau(\alpha(\mathbf{x})-\frac{(\alpha(\mathbf{x})-R)_+}{\gamma}))\mathbb{I}(A=d(\mathbf{x}))}{\pi(A|\mathbf{x})}\right]. \tag{3.33}$$

For illustrative purposes, we consider $\alpha(\mathbf{x})$ to be a class of linear functions and use the $l_1$ penalty to impose sparsity on $\alpha(X)$. Nonlinear functions of $\alpha(X)$ or other types of penalties can also be implemented similarly. Then we can compute the optimal IDR $d$ empirically via minimizing

$$\begin{aligned} G^{(2)}(f,\beta,b) &:= \frac{1}{n}\sum_{i=1}^{n}\frac{S(A_i f(\mathbf{x}_i))}{\pi(A_i|\mathbf{x}_i)}(-(1-\tau)R_i-\tau(\mathbf{x}_i^T\beta+b-\frac{(\mathbf{x}_i^T\beta+b-R_i)_+}{\gamma})) \\ &+\frac{\lambda_1}{2}||f||_{\mathcal{H}}^2+\frac{\lambda_2}{2}||\beta||_1+\frac{\eta}{2}(||\beta||_2^2+b^2) \end{aligned} \tag{3.34}$$

over $f\in\mathcal{H},\beta\in\mathbb{R}^p,b\in\mathbb{R}$ jointly. The regularization term $\frac{\eta}{2}(||\beta||_2^2+b^2)$ is to avoid numerical instability, where $\eta$ is a small positive number, such as $10^{-3}$. In what follows, we derive a convex majorant surrogate function for $G^{(2)}(f,\beta,b)$ and propose to use majorize-minimization (MM) algorithm to solve Problem (3.34). Note that $G^{(2)}(f,\beta,b)$ has two properties:

(i) $G^{(2)}(f,\beta,b)$ is strongly convex with respect to $(\beta,b)$;

(ii) $G^{(2)}(f,\beta,b)$ is a DC function with respect to $f$, i.e., $G^{(2)}(f,\beta,b) := G_1^{(2)}(f,\beta,b) - G_2^{(2)}(f,\beta,b)$ where

$$\begin{aligned} G_1^{(2)}(f,\beta,b) &:= \frac{1}{n}\sum_{i=1}^{n}\frac{S_1(A_i f(\mathbf{x}_i))}{\pi(A_i|\mathbf{x}_i)}(-(1-\tau)R_i-\tau(\mathbf{x}_i^T\beta+b-\frac{(\mathbf{x}_i^T\beta+b-R_i)_+}{\gamma})) \\ &+\frac{\lambda_1}{2}||f||_{\mathcal{H}_1}^2+\frac{\lambda_2}{2}||\beta||_1+\frac{\eta}{2}(||\beta||_2^2+b^2), \end{aligned}$$

and

$$G_2^{(2)}(f,\beta,b) := \frac{1}{n}\sum_{i=1}^{n}\frac{S_2(A_i f(\mathbf{x}_i))}{\pi(A_i|\mathbf{x}_i)}(-(1-\tau)R_i - \tau(\mathbf{x}_i^T\beta + b - \frac{(\mathbf{x}_i^T\beta + b - R_i)_+}{\gamma}));$$

(iii) $\frac{\partial G_2^{(2)}(f,\beta,b)}{\partial f}$ is Lipschitz with respect to $f$ , $\beta$ and $b$.

Given $f^{(v)}$ at the $v$-th iteration, we can compute the unique solution

$$(\beta^{(v)}, b^{(v)}) = \text{argmin}_{\alpha\in\mathcal{H}_2} G^{(2)}(f^{(v)},\beta,b).$$

We define

$$\tilde{G}^{(2)}(f,\beta^{(v)},b^{(v)}) = G_1^{(2)}(f,\beta^{(v)},b^{(v)}) - \frac{\partial G_2^{(2)}(f,\beta^{(v)},b^{(v)})}{\partial f}(f-f^{(v)}) - G_2^{(2)}(f^{(v)},\beta^{(v)},b^{(v)}). \quad (3.35)$$

Then we have the following proposition to justify the use of MM algorithm in order to solve Problem (3.34).

**Proposition 3.3.2.** *$\tilde{G}^{(2)}(f,\beta^{(v)},b^{(v)})$ is a strongly convex majorant of $\min_{\beta\in\mathbb{R}^p,b\in\mathbb{R}} G^{(2)}(f,\beta,b)$ at $f^{(v)}$ and $\frac{\partial \tilde{G}^{(2)}(f,\beta^{(v)},b^{(v)})}{\partial f}|_{f=f^{(v)}} = \frac{\partial \min_{\beta\in\mathbb{R}^p,b\in\mathbb{R}} G^{(2)}(f,\beta,b)}{\partial f}|_{f=f^{(v)}}$.*

*Proof.* One can verify that $\tilde{G}^{(2)}(f,\beta^{(v)},b^{(v)})$ is strongly convex with respect to $f$. By the convexity of $G_2^{(2)}(f,\beta,b)$ with respect to $f$, we have

$$\begin{aligned}\tilde{G}^{(2)}(f,\beta^{(v)},b^{(v)}) &\geq G_1^{(2)}(f,\beta^{(v)},b^{(v)}) - G_2^{(2)}(f,\beta^{(v)},b^{(v)})\\ &= G^{(2)}(f,\beta^{(v)},b^{(v)})\\ &\geq \min_{\beta\in\mathbb{R}^p,b\in\mathbb{R}} G^{(2)}(f,\beta,b).\end{aligned}$$

Moreover,

$$\begin{aligned}\tilde{G}^{(2)}(f^{(v)},\beta^{(v)},b^{(v)}) &= G_1^{(2)}(f^{(v)},\beta^{(v)},b^{(v)}) - G_2^{(2)}(f^{(v)},\beta^{(v)},b^{(v)})\\ &= G^{(2)}(f^{(v)},\beta^{(v)},b^{(v)})\\ &= \min_{\beta\in\mathbb{R}^p,b\in\mathbb{R}} G^{(2)}(f^{(v)},\beta,b).\end{aligned}$$

Given strong convexity of $G^{(2)}(f, \beta, b)$, by the Danskin's theorem, we have equal derivative results. □

Based on Proposition 3.34, we summarize our MM algorithm in Algorithm 2 below.

**Algorithm 2** Algorithm for (3.34)

```
For a given $f^{(v)}$ at the $v$ iteration, compute
    $(\beta^{(v)}, b^{(v)}) = \text{argmin}_{\beta \in \mathbb{R}^p, b \in \mathbb{R}} G^{(2)}(f^{(v)}, \beta, b).$    (3.36)
Given $(\beta^{(v)}, b^{(v)})$, compute
    $f^{(v+1)} = \text{argmin}_{f \in \mathcal{H}_1} \tilde{G}^{(2)}(f, (\beta^{(v)}, b^{(v)})).$    (3.37)
The algorithm stops till some stopping criteria of $f^{(v)}$ are satisfied.
```

Proposition 3.3.2 and three properties of $G^{(2)}(f, \beta)$ are the building blocks for the convergence of our proposed MM-algorithm in Table 2 to a d-stationary point. The related proof can be found in (Mairal, 2015, Example 2.3.4; Proposition 2.5). For recent development in MM-algorithm, see (Cui et al., 2018).

## 3.4 Theoretical Results

In this section, we discuss the statistical theory related to $M_2(d)$, while the method of controlling the average lower tail using $M_1(d)$ can be viewed as a special case. For simplicity, we consider the case that $\tau = 1$ in $M_2(d)$, but the result can be directly generalized for other $\tau$ by combining the existing results under the expected-value function such as (Zhao et al., 2012) or (Zhou et al., 2017). We define two corresponding value functions for $\tau = 1$ as follows:

$$M_0(d, \alpha) = \mathbf{E}[\frac{\mathbb{I}(A = d(\mathbf{x}))}{\pi(A|\mathbf{x})}(\alpha(\mathbf{x}) - \frac{1}{\gamma}(\alpha(\mathbf{x}) - R)_+)], \tag{3.38}$$

and

$$M_0(d) = \sup_{\alpha \in L^1(\mathbf{x}, \Xi, \mathbf{P}_\mathbf{x})} \mathbf{E}[\frac{\mathbb{I}(A = d(\mathbf{x}))}{\pi(A|\mathbf{x})}(\alpha(\mathbf{x}) - \frac{1}{\gamma}(\alpha(\mathbf{x}) - R)_+)]. \tag{3.39}$$

The corresponding optimal solutions of maximizing (3.38) is $d^* = \text{sign}(f^*)$ and $\alpha^*$. Since we use the surrogate loss function $S(u)$, we further define

$$M_T(f,\alpha) = \mathbf{E}[\frac{S(Af(\mathbf{x}))}{\pi(A|\mathbf{x})}(\alpha(\mathbf{x}) - \frac{1}{\gamma}(\alpha(\mathbf{x}) - R)_+)]$$

as the surrogate value function. Our theoretical results are based on statistical learning theory with an extension to two functional classes, since we need to consider both $f$ and $\alpha$ in our problems.

### 3.4.1 Fisher Consistency

We first establish Fisher consistency of estimating optimal ITRs under $M_T(f,\alpha)$ to justify the use of the surrogate loss $S(u)$, compared with $M_0(d,\alpha)$. This is different from the classical Fisher consistency, which only involves one functional class of interest.

**Theorem 3.4.1.** *For any measurable function $f$ and $\alpha$, if $(f_T^*, \alpha_T^*)$ maximizes $M_T(f,\alpha)$, then $(sign(f_T^*), \alpha_T^*)$ maximizes $M_0(d,\alpha)$.*

Based on Theorem 4.2.4, instead of $M_0(d,\alpha)$, we can target on $M_T(d,\alpha)$ alternatively.

### 3.4.2 Excess Value Bound

Based on Theorem 4.2.4, we can further justify the use of the surrogate function $S(u)$ by establishing the following excess value bound for the 0-1 loss in $M_0(d,\alpha)$.

**Theorem 3.4.2.** *For any measurable function $f,\alpha$ and any probability distribution over $(\mathbf{x}, A, R)$,*

$$M_0(d^*,\alpha^*) - M_0(sign(f),\alpha) \leq M_T(f_T^*,\alpha_T^*) - M_T(f,\alpha).$$

Theorem 3.4.2 gives us a way of bounding the difference between the optimal IDR and the estimated IDR under $M_0(d,\alpha)$ by using $M_T(d,\alpha)$ instead.

### 3.4.3 Convergence Rate

In order to obtain the finite sample performance of our estimated optimal IDR under $M_0(d,\alpha)$, it is enough to focus on the difference of $M_T(d,\alpha)$ between the estimated optimal

IDR and the optimal ITR based on Theorem 3.4.2. Define

$$(\hat{f}, \hat{\alpha}) = \text{argmin}_{f \in \mathcal{H}_1, \alpha \in \mathcal{H}_2} O_n(f, \alpha) + \frac{\lambda_{1n}}{2}||f||^2_{\mathcal{H}_1} + \frac{\lambda_{2n}}{2}||\alpha||^2_{\mathcal{H}_2}, \tag{3.40}$$

where $O_n(f, \alpha) := \frac{1}{n}\sum_{i=1}^n \frac{S(A_i f(\mathbf{x}_i))}{\pi(A_i|\mathbf{x}_i)}(\frac{(\alpha(\mathbf{x}_i)-R_i)_+}{\gamma} - \alpha(\mathbf{x}_i))$. We consider different penalty functions. In particular, $|| \bullet ||_{\mathcal{H}_i}$ can be one of the following choices:

(1) $l_1$ norm of coefficients if we consider $\mathcal{H}_i = \{\mathbf{x}^T\beta + b \ |\beta \in \mathbb{R}^p\}$;

(2) $l_2$ norm of coefficients if we consider $\mathcal{H}_i = \{\mathbf{x}^T\beta + b \ |\beta \in \mathbb{R}^p\}$;

(3) RKHS norm if we consider $\mathcal{H}_i$; $i = 1, 2$ to be RKHS with Gaussian radial basis functions.

Before presenting our results, we need following definitions.

**Definition 3.1.** Consider $\mathcal{F}$ to be a class of real value measurable functions $f : \mathcal{Z} \to \mathbb{R}$. The Rademacher complexity of $\mathcal{F}$ is defined as

$$\mathcal{R}_n(\mathcal{F}) := \mathbf{E}[\sup_{f \in \mathcal{F}} \frac{1}{n}\sum_{i=1}^n \sigma_i f(Z_i)], \tag{3.41}$$

where $Z_1, \cdots, Z_n$ are drawn *i.i.d* from some probability distribution $P$ and the Rademacher random variables $\sigma_1, \cdots, \sigma_n$ are drawn *i.i.d* from uniform distribution over $\{1, -1\}$.

If we interpret the Rademacher random variables as noise, the Rademacher complexity is the maximal correlation between functions and the pure noise. Thus it can measure the complexity of classes of functions. The corresponding empirical Rademacher complexity of $\mathcal{F}$ is defined as

$$\hat{\mathcal{R}}_n(\mathcal{F}) := \mathbf{E}[\sup_{f \in \mathcal{F}} \frac{1}{n}\sum_{i=1}^n \sigma_i f(Z_i)|Z_1, \cdots, Z_n], \tag{3.42}$$

where we can see that $\mathbf{E}[\hat{\mathcal{R}}_n(\mathcal{F})] = \mathcal{R}_n(\mathcal{F})$. The following lemma characterizes the Rademacher complexity of the Lipschitz composition operator. It is an extension of Corollary 3.17 in (Ledoux and Talagrand, 2013).

**Lemma 3.4.1.** *If a function $\phi : \mathbb{R}^p \to \mathbb{R}$ is Lipschitz continuous with respect to the $l_1$ norm, i.e*

$$|\phi(t_1, \cdots, t_p) - \phi(s_1, \cdots, s_p)| \le L_\phi \sum_{i=1}^p |t_i - s_i|, \tag{3.43}$$

*for any $t_i, s_i$. Then we have*

$$\mathcal{R}_n(\phi(\mathcal{F}_1, \cdots, \mathcal{F}_p)) \leq L_\phi \sum_{i=1}^{p} \mathcal{R}_n(\mathcal{F}_i), \tag{3.44}$$

*where $\mathcal{F}_i$ is a certain class of functions for $i = 1, \cdots, n$.*

Define $O_T(f, \alpha) = -M_T(f, \alpha)$ and let $(f_{\lambda_{1n}}, \alpha_{\lambda_{2n}}) = \text{argmin}_{f \in \mathcal{H}_1, \alpha \in \mathcal{H}_2} O_T(f, \alpha) + \frac{\lambda_{1n}}{2}||f||^2_{\mathcal{H}_1} + \frac{\lambda_{2n}}{2}||\alpha||^2_{\mathcal{H}_2}$. Then $\mathcal{A}(\lambda_{1n}, \lambda_{2n}) := O_T(f_{\lambda_{1n}}, \alpha_{\lambda_{2n}}) + \frac{\lambda_{1n}}{2}||f_{\lambda_{1n}}||^2_{\mathcal{H}_1} + \frac{\lambda_{2n}}{2}||\alpha_{\lambda_{2n}}||^2_{\mathcal{H}_2} - O_T(f^*_T, \alpha^*_T)$ is considered to be the approximation error. The following theorem gives us a finite sample upper bound of our estimated optimal IDR and the optimal IDR based on $M_0(d, \alpha)$ by the estimation and approximation errors.

**Theorem 3.4.3.** *For any distribution $P$ over $(\mathbf{x}, A, R)$ such that $|R| \leq C_0$ and $\pi(a|\mathbf{x}) \geq a_0$ a.s. for any $a$, then with probability $1 - \epsilon$, one can have*

$$M_0(d^*, \alpha^*) - M_0(sign(\hat{f}), \hat{\alpha}) \leq 4M_5(\mathcal{R}_n(\Pi_1) + \mathcal{R}_n(\Pi_2)) + \sqrt{\frac{8\log(\frac{1}{\epsilon})}{n}} + \mathcal{A}(\lambda_{1n}, \lambda_{2n}),$$

*where $\Pi_1 = \{f|\ f \in \mathcal{H}_1, \frac{\lambda_{1n}}{2}||f||^2_{\mathcal{H}_1} \leq M_3\}$ and $\Pi_2 = \{\alpha|\ \alpha \in \mathcal{H}_2, \frac{\lambda_{2n}}{2}||\alpha||^2_{\mathcal{H}_2} \leq M_3\}$ for some constants $M_3$ and $M_5$.*

In order to obtain the finite sample bound, we need to compute the empirical Rademacher complexity of $\Pi_1$ and $\Pi_2$ and also bound the approximation error. The results will depend on the specific choice of $\mathcal{H}_i$ for $i = 1, 2$. In the following, we give several corollaries in order to establish the finite sample bound for our estimated optimal IDR under $M_0(d, \alpha)$.

**Corollary 3.4.1.** *Consider $\mathcal{H}_i$ to be classes of linear functions with the $l_2$ penalty for $i = 1, 2$ and suppose $\mathbf{E}[||\mathbf{x}||^2_2] \leq C_1^2$. If $f^*_T \in \mathcal{H}_1$ and $\alpha^*_T \in \mathcal{H}_2$, that is $f^*_T = \mathbf{x}^T w^* + b^*_1$ and $\alpha^*_T = \mathbf{x}^T \theta^* + b^*_2$ with $||w^*||^2_2 + (b^*_1)^2 \leq D_1$ and $||\theta^*||^2_2 \leq +(b^*_2)^2 \leq D_2$ for some constants $C_1, D_1, D_2$, then under the assumptions in Theorem 3.4.3, with probability $1 - \epsilon$, one can have*

$$M_0(d^*, \alpha^*) - M_0(sign(\hat{f}), \hat{\alpha}) \leq c_1 n^{-\frac{1}{3}},$$

*for some constant $c_1$.*

**Corollary 3.4.2.** *Consider $\mathcal{H}_i$ to be classes of linear functions with the $l_1$ penalty for $i = 1, 2$ and suppose $||\mathbf{x}||_\infty \leq C_2$. If $f_T^* \in \mathcal{H}_1$ and $\alpha_T^* \in \mathcal{H}_2$, that is $f_T^* = \mathbf{x}^T w^* + b_1^*$ and $\alpha_T^* = \mathbf{x}^T \theta^* + b_2^*$ with $||w^*||_1 + |b_1^*| \leq D_3$ and $||\theta^*||_1 + |b_2^*| \leq D_4$ for some constants $C_2, D_3, D_4$, then under the assumptions in Theorem 3.4.3, with probability $1 - \epsilon$, one can have*

$$M_0(d^*, \alpha^*) - M_0(sign(\hat{f}), \hat{\alpha}) \leq c_2 (\frac{\log(2p)}{n})^{\frac{1}{3}},$$

*for some constant $c_2$.*

**Corollary 3.4.3.** *Consider $\mathcal{H}_i$ to be RKHS with Gaussian radial basis functions for $i = 1, 2$ and suppose assumptions in Theorem 3.4.3 hold. If $\mathcal{A}(\lambda_{1n}, \lambda_{2n}) \leq C_5 \lambda_{1n}^{w_1} + C_6 \lambda_{2n}^{w_2}$, where $w_1, w_2 \in (0, 1]$, then with probability at least $1 - \epsilon$,*

$$M_0(d^*, \alpha^*) - M_0(sign(\hat{f}), \hat{\alpha}) \leq \max(c_3^{(1)}, c_3^{(2)}) \max \left( n^{-\frac{w_1}{2w_1+1}}, n^{-\frac{w_2}{2w_2+1}} \right),$$

*for some constant $c_3^{(1)}$ and $c_3^{(2)}$.*

The above corollary shows that the difference between our estimated IDRs and the optimal IDR under $M_0(d, \alpha)$ converges to 0 in probability under some conditions. The upper bound assumption on the approximation error $\mathcal{A}(\lambda_{1n}, \lambda_{2n})$ is analogous to those in the statistical learning literature such as (Steinwart and Scovel, 2007) to derive the convergence rate.

### 3.5 Simulation Studies

In our numerical analysis, we set $\tau = 0.5$ and $\gamma = 0.5$ to treat expected-value and CVaR value functions equally in most examples, while we show different performances of different $\tau$ and $\gamma$ in our first simulation example below. For all simulation settings, we consider binary-armed randomized trials with equal probabilities of patients being assigned to each treatment group.

We use $l1$-DC-CVaR, $l2$-DC-CVaR and GK-DC-CVaR to represent the methods of estimating optimal IDRs under $M_1(d)$ with three different penalties on $f$ in Problem (3.26) respectively. Here "$l1$" and "$l2$" refer to the $l_1$ and $l_2$ penalties. "GK" represents using Gaussian radial basis functions with bandwidth $\varsigma$ to learn the optimal IDR. Similarly, we use $l1$-MM-CVaR, $l2$-MM-CVaR and GK-MM-CVaR to represent the methods of estimating optimal IDRs under $M_2(d)$ with three different penalties on $f$ in Problem (3.34) respectively.

All tuning parameters are selected based on the 10-fold-cross-validation procedure. We select the tuning parameter that maximizes the empirical average of mixed value functions $M_1(d)$ and $M_2(d)$ on the validation data set defined as

$$\hat{M}_1(d) = \frac{\mathbf{E}_n[\frac{((1-\tau)R+\tau(\hat{\alpha}-\frac{(\hat{\alpha}-R)_+}{\gamma}))\mathbb{I}(A=d(\mathbf{x}))}{\pi(A|\mathbf{x})}]}{\mathbf{E}_n[\frac{\mathbb{I}(A=d(\mathbf{x}))}{\pi(A|\mathbf{x})}]}, \tag{3.45}$$

and

$$\hat{M}_2(d) = \frac{\mathbf{E}_n[\frac{((1-\tau)R+\tau(\hat{\alpha}(\mathbf{x})-\frac{(\hat{\alpha}(\mathbf{x})-R)_+}{\gamma}))\mathbb{I}(A=d(\mathbf{x}))}{\pi(A|\mathbf{x})}]}{\mathbf{E}_n[\frac{\mathbb{I}(A=d(\mathbf{x}))}{\pi(A|\mathbf{x})}]}, \tag{3.46}$$

respectively, where $\mathbf{E}_n$ denotes the empirical average over the validation data.

We compare our methods with the following four methods:

(1) D-learning by (Qi and Liu, 2018);

(2) $l_1$-PLS by (Qian and Murphy, 2011) with basis function $(1, \mathbf{x}, A, \mathbf{x}A)$;

(3) RWL by (Zhou et al., 2017) with linear kernel;

(4) RWL by (Zhou et al., 2017) with Gaussian kernel.

### 3.5.1 A Motivating Example Revisit

Recall the motivating example in Figure 3.1 that shows the importance of risk controls in estimating optimal IDRs. In this subsection, we conduct some numerical analysis to further demonstrate this finding. In particular, the covariate of gender is generated by uniform distribution over $\{1, -1\}$, where 1 and $-1$ denotes male and female respectively. The corresponding outcome $R$ is generated by the following model:

$$R = \mathbb{I}(XA = 1)\epsilon_1 + \mathbb{I}(XA = -1)\epsilon_2,$$

where $\epsilon_1 \sim \mathcal{N}(-0.1, 1)$ and $\epsilon_2 \sim \mathcal{N}(0, 0.5)$. We consider training data with the sample size $n = 200$ and independently generated test data of size 10000. We first set $\tau = \gamma = 0.5$. Based on test data, in Figure 3.3, we plot box plots of three different outcome distributions if treatments follow estimated IDRs by $l_1$-PLS, linear RWL, $l2$-DC-CVaR and $l2$-MM-CVaR correspondingly. Based on these box plots, we can observe that since there is not much difference between these two treatments based on the expected outcome, the empirical mean of value functions resulted from these four methods are indistinguishable. However, besides maximizing the expected outcome for each individual, our methods also control the risk of each individual. Thus the resulting outcome distributions by our methods are more stable, has less variability than those of $l_1$-PLS and linear RWL.

In addition, we also plot medians and standard deviations of value functions of one replication under different combinations of $\tau$ and $\gamma$ by $l2$-DC-CVaR in Figure 3.4. We can see that as $\tau$ gets close to 1 and $\gamma$ gets close to 0, the standard deviations of corresponding value functions are small, which means the resulting optimal IDRs are more stable since we put more weights on lower tails. However, the corresponding medians of value functions are not large. In contrast, if we choose relatively balanced $\tau$ and $\gamma$, the medians of value functions are larger by sacrificing some stability. In practice, users can decide his or her own preferences based on the specific problem.

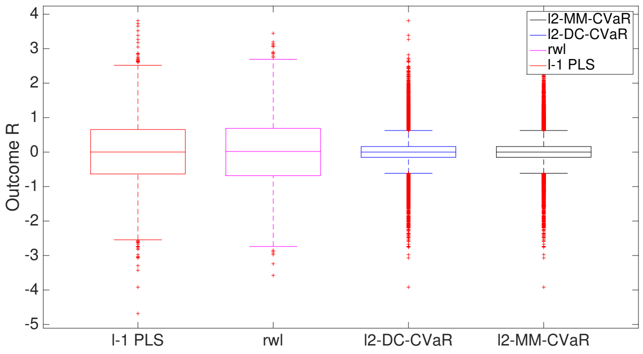


**Figure 3.3:** Box plots of value functions computed by three methods. The left box plot corresponds to the result of $l_1$-PLS under the expected-value function framework. The middle and the right box plots correspond to the result of our proposed methods under $M_1(d)$ and $M_2(d)$ respectively.

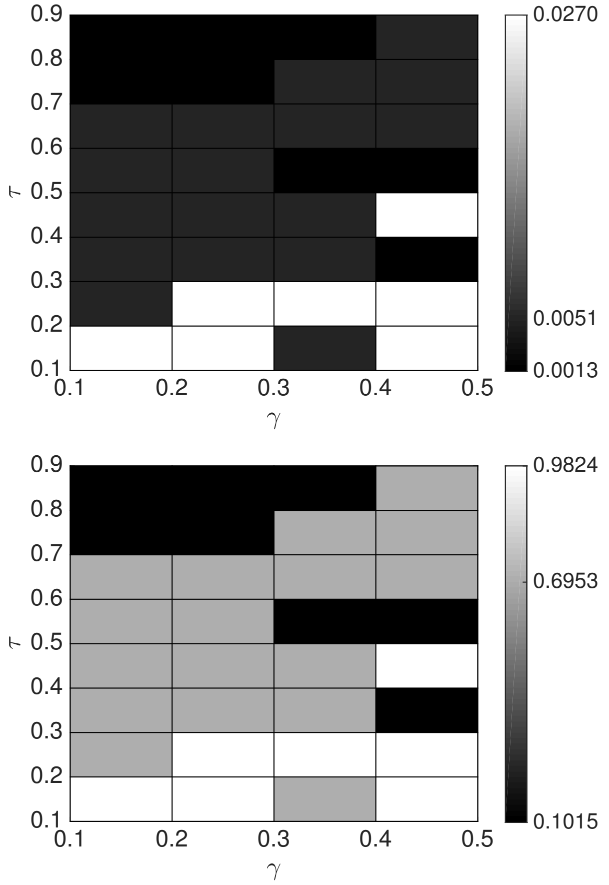


**Figure 3.4:** Medians and standard deviations of value functions under different combinations of $\tau$ and $\gamma$ by $l_2$-DC-CVaR. The left plot corresponds to the medians and the right plot corresponds to the standard deviations of value functions respectively.

### 3.5.2 Distributional Shift Examples

In this section, we demonstrate the superior performance of our methods under distribution shift of covariates $\mathbf{x}$ and outcome $R$ based on the duality representations of $M_1(d)$ and $M_2(d)$ in (3.21) and (3.20) respectively. We consider the sample size $n = 200$ and the dimension $p = 20$. The outcome $R$ is generated by the model: $R = 1 + x_1 + x_2 + A(x_1 - x_2 + x_3) + \epsilon$. We consider the following two distribution shift scenarios:

(1) Each covariate follows a two component Gaussian mixture distribution of $\mathcal{N}(0, 1)$ and $\mathcal{N}(5, 1)$ with probability of mixture to be 0.8 and 0.2 respectively and $\epsilon$ follows standard Gaussian distribution;

(2) Covariates $X$ are generated by the uniform distribution between $-1$ and 1 and $\epsilon$ follows a two component mixture distribution of $\mathcal{N}(0,1)$ and log-normal distribution $lognorm(0,2)$ with probability of mixture to be 0.7 and 0.3 respectively.

The first scenario considers the covariate distribution shift and the second scenario considers the outcome distribution shift. For simplicity, we only report misclassification error rates given by $l_1$-PLS, linear RWL, l2-DC-CVaR and l2-MM-CVaR in Table 3.1. For Scenario (1), since $l_1$-PLS assumes a linear model, it's performance is not affected by covariate distribution shift. In contrast, rwl, which is based on maximizing the value function, depends heavily on correct approximation to value function empirically. Thus the performance of rwl is worse than $l_1$-PLS under this scenario. For the estimated optimal IDR under $M_1(d)$, the performance is superior to rwl because $M_1(d)$ considers the perturbation of the covariate distribution shift, while $M_2(d)$ does not and its corresponding performance is relatively worse. For Scenario (2), since both estimated optimal IDRs under $M_1(d)$ and $M_2(d)$ are minimax estimator under the outcome distribution shift, the performances are much better than two other methods under the value function framework.

**Table 3.1:** Comparisons of misclassification error rates (standard error) for simulated examples with $n = 200$ and $p = 20$.

| | Scenario (1) | Scenario (2) |
|---:|---:|---:|
| $l_1$-PLS | **0.04**(0.004) | 0.38(0.011) |
| rwl | 0.15(0.008) | 0.37(0.01) |
| $l_2$-DC-CVaR | 0.12(0.006) | **0.15**(0.006) |
| $l_2$-MM-CVaR | 0.23(0.009) | 0.31(0.01) |

### 3.5.3 Simulation Scenarios

In this subsection, we further study the performance of our proposed methods via eight simulation examples. We consider the sample size $n = 200$ and the dimension $p = 20$. The covariates $\mathbf{x}$ are generated by the uniform distribution between $-1$ and 1. The outcome $R$ is generated by the model: $R = 1 + x_1 + x_2 + A\delta(\mathbf{x}) + \epsilon$. We consider the following eight different combinations of $\delta(\mathbf{x})$ and $\epsilon$:

(1) $\delta(\mathbf{x}) = x_1 - x_2 + x_3$, and $\epsilon$ follows Gaussian normal $\mathcal{N}(0,1)$;

(2) $\delta(\mathbf{x}) = x_1 - x_2 + x_3$, and $\log(\epsilon)$ follows Gaussian normal $\mathcal{N}(0, 2|1 + x_1 + x_2|)$;

(3) $\delta(\mathbf{x}) = x_1 - x_2 + x_3$, and $\log(\epsilon)$ follows Gaussian normal $\mathcal{N}(0, 2)$;

(4) $\delta(\mathbf{x}) = x_1 - x_2 + x_3$, and $\epsilon$ follows a Weibull distribution with shape parameter 0.3 and scale parameter 0.5;

(5) $\delta(\mathbf{x}) = 3.8(0.8 - x_1^2 - x_2^2)$, and $\epsilon$ follows Gaussian normal $\mathcal{N}(0, 1)$;

(6) $\delta(\mathbf{x}) = 3.8(0.8 - x_1^2 - x_2^2)$, and $\log(\epsilon)$ follows Gaussian normal $\mathcal{N}(0, 2|1 + x_1 + x_2|)$;

(7) $\delta(\mathbf{x}) = 3.8(0.8 - x_1^2 - x_2^2)$, and $\log(\epsilon)$ follows Gaussian normal $\mathcal{N}(0, 2)$;

(8) $\delta(\mathbf{x}) = 3.8(0.8 - x_1^2 - x_2^2)$, and $\epsilon$ follows a Weibull distribution with shape parameter 0.3 and scale parameter 0.5.

We consider different shapes of error distributions to test the robustness of our methods, compared with other methods. The first four scenarios are of linear decision boundaries while the remaining four consider nonlinear decision boundaries. In order to evaluate different methods, we generate test data and use $\text{sign}(\delta(\mathbf{x}))$ as the true optimal decision rule, since treatment $A$ only appears in the interaction term $\delta(\mathbf{x})$. We evaluate different methods based on the misclassification error rates in Table 3.2, mean of expected-value functions in Table 3.3, mean of 50% and 25% quantiles of value functions in Tables 3.4 and 3.5. Overall, our methods show competitive performances among all methods. In particular, for Scenarios (1) and (5), which are standard simulation settings in IDRs literature, our proposed methods performs well in finding optimal IDRs. For Scenarios (2) and (6), the error distributions depend on the covariate information. Although the average of empirical value functions of our proposed methods are smaller than those of RWL, the 50% and 25% quantiles of empirical value functions by our methods are much better. One possible reason is that methods under the expected-value function framework ignore subjects with potentially high risk while only focusing on maximizing the expected-value function. Thus the resulting IDRs by these methods may assign wrong treatments to patients and make them become even worse by delivering the corresponding IDR. Similar observations can be drawn from other simulation scenarios.

**Table 3.2:** Comparisons of misclassification error rates (standard deviation) for simulated examples with $n = 200$ and $p = 20$. From left to right, each column represents Scenarios (1)-(8) respectively. Each row represents one specific method. The last six rows correspond to our proposed methods.

| | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 |
|---:|---:|---:|---:|---:|---:|---:|---:|---:|
| Dlearn | 0.31(0.1) | 0.5(0.01) | 0.49(0.04) | 0.49(0.05) | 0.4(0.04) | 0.53(0.12) | 0.48(0.13) | 0.46(0.12) |
| $l_1$-PLS | **0.28**(0.06) | 0.5(0.06) | 0.44(0.09) | 0.45(0.08) | 0.44(0.05) | 0.52(0.04) | 0.5(0.03) | 0.49(0.04) |
| rwl | 0.34(0.07) | 0.5(0.07) | 0.44(0.07) | 0.44(0.07) | 0.41(0.06) | 0.51(0.05) | 0.47(0.07) | 0.46(0.08) |
| rwl-GK | 0.5(0.01) | 0.5(0.01) | 0.5(0.01) | 0.5(0.01) | **0.38**(0.05) | 0.52(0.12) | 0.46(0.12) | 0.43(0.11) |
| l2-DC-CVaR | 0.32(0.06) | 0.17(0.07) | 0.22(0.07) | 0.13(0.08) | 0.43(0.04) | 0.5(0.02) | 0.48(0.03) | 0.46(0.04) |
| l1-DC-CVaR | 0.32(0.06) | **0.16**(0.06) | **0.19**(0.06) | **0.1**(0.05) | 0.44(0.04) | 0.5(0.01) | 0.49(0.02) | 0.49(0.01) |
| GK-DC-CVaR | 0.49(0.02) | 0.5(0.01) | 0.5(0.01) | 0.5(0.01) | **0.38**(0.03) | 0.51(0.12) | 0.41(0.09) | **0.38**(0.05) |
| l2-MM-CVaR | 0.35(0.07) | 0.44(0.08) | 0.38(0.08) | 0.36(0.07) | 0.43(0.05) | 0.45(0.04) | 0.45(0.05) | 0.44(0.05) |
| l1-MM-CVaR | 0.33(0.08) | 0.47(0.07) | 0.39(0.09) | 0.38(0.1) | 0.41(0.06) | 0.47(0.06) | 0.44(0.07) | 0.43(0.07) |
| GK-MM-CVaR | 0.49(0.02) | 0.5(0.01) | 0.5(0.01) | 0.5(0.02) | **0.38**(0.04) | **0.44**(0.11) | **0.39**(0.07) | 0.4(0.08) |

**Table 3.3:** Comparisons of average value functions (standard deviation) for simulated examples with $n = 200$ and $p = 20$. From left to right, each column represents scenarios (1)-(8) respectively. Each row represents one specific method. The last six rows correspond to our proposed methods.

| | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 |
|---:|---:|---:|---:|---:|---:|---:|---:|---:|
| Dlearn | 1.44(0.23) | 205244.32(2085980.92) | 8.4(0.76) | 5.56(0.33) | 1.42(0.17) | 4759.22(16549.4) | 8.51(0.9) | 5.84(0.57) |
| $l_1$-PLS | **1.51**(0.12) | 205839.37(2090994.07) | 8.55(0.84) | 5.67(0.35) | 1.25(0.22) | 32480.58(287116.51) | 8.42(0.77) | 5.68(0.35) |
| rwl | 1.39(0.17) | 47681.35(412574.85) | 8.58(0.8) | 5.7(0.34) | 1.37(0.23) | 4248.93(13197.34) | 8.46(0.87) | 5.83(0.47) |
| rwl-GK | 1.01(0.07) | **241196.97**(2122609.28) | 8.41(0.71) | 5.55(0.33) | 1.47(0.21) | 30912.36(279385.21) | 8.53(0.94) | 5.96(0.54) |
| l2-DC-CVaR | 1.44(0.13) | 240599.77(2128310.24) | **8.96**(0.68) | 6.27(0.34) | 1.26(0.18) | 32777.67(284170.74) | 8.53(1.05) | 5.78(0.39) |
| l1-DC-CVaR | 1.44(0.13) | 240226.39(2128134.97) | 8.94(0.61) | **6.32**(0.33) | 1.27(0.18) | 32917.95(284157.15) | 8.51(1.05) | 5.68(0.35) |
| GK-DC-CVaR | 1.02(0.07) | 13046.46(72486.56) | 8.4(0.76) | 5.54(0.35) | **1.49**(0.15) | **34718.08**(282511.23) | 8.75(0.86) | **6.11**(0.41) |
| l2-MM-CVaR | 1.37(0.16) | 4121.03(10115.71) | 8.6(0.69) | 5.87(0.33) | 1.26(0.2) | 34297.67(281664.15) | 8.54(0.76) | 5.91(0.4) |
| l1-MM-CVaR | 1.4(0.18) | 4804.45(11856.94) | 8.63(0.7) | 5.82(0.4) | 1.36(0.23) | 30566.94(279921.05) | 8.62(0.82) | 5.92(0.47) |
| GK-MM-CVaR | 1.02(0.08) | 7063.71(24485.57) | 8.41(0.77) | 5.55(0.33) | **1.49**(0.17) | 4728.1(14282.42) | **8.79**(0.83) | 6.05(0.46) |

**Table 3.4:** Comparisons of 50% quantiles (standard deviation) of value functions for simulated examples with $n = 200$ and $p = 20$. From left to right, each column represents scenarios (1)-(8) respectively. Each row represents one specific method. The last six rows correspond to our proposed methods.

| | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 |
|---:|---:|---:|---:|---:|---:|---:|---:|---:|
| Dlearn | 1.45(0.24) | 2.22(0.04) | 2.6(0.1) | 1.84(0.13) | 1.44(0.19) | 2.19(0.67) | 2.89(0.61) | 2.14(0.65) |
| pls | **1.52(0.13)** | 2.21(0.16) | 2.71(0.2) | 1.93(0.19) | 1.28(0.24) | 2.35(0.22) | 2.86(0.18) | 2.03(0.2) |
| rwl | 1.4(0.17) | 2.22(0.19) | 2.71(0.16) | 1.96(0.16) | 1.39(0.25) | 2.36(0.27) | 2.96(0.36) | 2.2(0.44) |
| rwl-GK | 1.01(0.07) | 2.22(0.05) | 2.59(0.05) | 1.81(0.04) | 1.5(0.23) | 2.25(0.65) | 3.01(0.58) | 2.32(0.58) |
| l2-DC-CVaR | 1.45(0.14) | 2.8(0.08) | 3.11(0.09) | 2.46(0.08) | 1.28(0.19) | 2.42(0.14) | 2.97(0.16) | 2.17(0.22) |
| l1-DC-CVaR | 1.45(0.14) | **2.82**(0.06) | **3.13**(0.08) | **2.48**(0.05) | 1.28(0.19) | 2.43(0.06) | 2.91(0.11) | 2.04(0.09) |
| Gaussian-DC-CVaR | 1.01(0.08) | 2.22(0.04) | 2.59(0.05) | 1.81(0.04) | **1.53**(0.16) | 2.33(0.67) | 3.24(0.42) | **2.56**(0.29) |
| l2-MM-CVaR | 1.38(0.17) | 2.37(0.19) | 2.84(0.17) | 2.15(0.14) | 1.28(0.22) | **2.68**(0.22) | 3.08(0.2) | 2.3(0.25) |
| l1-MM-CVaR | 1.41(0.19) | 2.3(0.16) | 2.82(0.19) | 2.09(0.21) | 1.38(0.25) | 2.57(0.35) | 3.1(0.33) | 2.31(0.37) |
| Gaussian-MM-CVaR | 1.02(0.09) | 2.22(0.04) | 2.59(0.04) | 1.81(0.05) | 1.52(0.18) | 2.7(0.61) | **3.31**(0.34) | 2.45(0.46) |

**Table 3.5:** Comparisons of 25% quantiles (standard deviation) of value functions for simulated examples with $n = 200$ and $p = 20$. From left to right, each column represents scenarios (1)-(8) respectively. Each row represents one specific method. The last six rows correspond to our proposed methods.

| | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 |
|---:|---:|---:|---:|---:|---:|---:|---:|---:|
| Dlearn | -1.38(0.25) | 1.02(0.03) | 1.14(0.12) | 0.57(0.16) | -1.54(0.2) | 0.67(0.59) | 1.08(0.67) | 0.53(0.63) |
| $l_1$-PLS | **-1.3**(0.14) | 1.02(0.16) | 1.28(0.26) | 0.69(0.23) | -1.71(0.25) | 0.68(0.18) | 0.89(0.18) | 0.25(0.19) |
| rwl | -1.43(0.19) | 1.04(0.19) | 1.29(0.21) | 0.72(0.2) | -1.59(0.27) | 0.71(0.22) | 1.05(0.41) | 0.51(0.45) |
| rwl-GK | -1.82(0.08) | 1.02(0.04) | 1.12(0.05) | 0.53(0.04) | -1.47(0.23) | 0.72(0.57) | 1.2(0.63) | 0.71(0.57) |
| l2-DC-CVaR | -1.38(0.15) | 1.82(0.12) | 1.86(0.14) | 1.46(0.13) | -1.71(0.2) | 0.75(0.11) | 1.01(0.2) | 0.41(0.25) |
| l1-DC-CVaR | -1.37(0.15) | **1.84**(0.1) | **1.9**(0.12) | **1.5**(0.07) | -1.7(0.2) | 0.75(0.05) | 0.93(0.13) | 0.25(0.09) |
| GK-DC-CVaR | -1.82(0.08) | 1.03(0.03) | 1.12(0.05) | 0.53(0.04) | **-1.45**(0.17) | 0.79(0.6) | 1.46(0.46) | **0.95**(0.28) |
| l2-MM-CVaR | -1.46(0.17) | 1.19(0.21) | 1.45(0.22) | 0.96(0.2) | -1.7(0.22) | 0.99(0.21) | 1.15(0.28) | 0.56(0.31) |
| l1-MM-CVaR | -1.42(0.19) | 1.12(0.17) | 1.43(0.26) | 0.89(0.29) | -1.61(0.26) | 0.91(0.31) | 1.22(0.4) | 0.62(0.4) |
| GK-MM-CVaR | -1.81(0.09) | 1.03(0.04) | 1.12(0.04) | 0.54(0.06) | **-1.45**(0.19) | **1.13**(0.55) | **1.53**(0.37) | 0.85(0.44) |

## 3.6 Real Data Applications

In this section, we perform a real data analysis to further evaluate our proposed robust criteria for estimating optimal IDRs. The clinical trial dataset we used comes from “AIDS Clinical Trials Group (ACTG) 175” in (Hammer et al., 1996) to study whether there exists some subpopulations that are suitable for different combinations of treatments for AIDS. In this study, a total number of 2139 patients with HIV infection were randomly assigned into four treatment groups: zidovudine (ZDV) monotherapy, ZDV combined with didanosine (ddI), ZDV combined with zalcitabine (ZAL), and ddI monotherapy with equal probability. In this data application, we focus on finding optimal IDRs between two treatments: ZDV with ddI and ZDV with ZAL as our interest. The total number of patients receiving these two treatments are 1046.

Similar to the previous studies by (Lu et al., 2013) and (Fan et al., 2017), we select 12 baseline covariates into our model: age (year), weight(kg), CD4+T cells amount at baseline, Karnofsky score (scale at 0-100), CD8 amount at baseline,gender (1 = male, 0 = female), homosexual activity (1 = yes, 0 = no), race (1 = non white, 0 = white), history of intravenous drug use (1 = yes, 0 = no), symptomatic status (1=symptomatic, 0=asymptomatic), antiretroviral history (1=experienced, 0=naive) and hemophilia (1=yes, 0=no). The first five covariates are continuous and have been scaled before estimation. The remaining seven covariates are binary categorical variables. We consider the outcome as the difference between the early stage (around 25 weeks) CD4+ T (cells/mm$^3$) cell amount and the baseline CD4+ T cells before the trial. Using this outcome, we can estimate the optimal IDR under our proposed robust criteria. To evaluate the performance of our proposed methods under robust criteria, we randomly divide the dataset into five folds and use four of them to estimate optimal IDRs by different methods. The remaining one fold of data is used to evaluate the performances of different methods. We repeat this procedure 220 times. For each method, we report the mean, 50% and 25% quantiles of empirical value functions. From Table 3.6, we can see that our proposed methods perform competitively among all methods. In particular, the "GK-DC-CVaR" method performs the best compared with other methods, which indicates the optimal IDR of this problem may be potentially nonlinear. Another observation is that our proposed methods are not consistently better than other methods since robust methods are not necessarily the best for a specific application. However, robustness can be more insensitive to some deviations from model assumptions, which implies that our methods have the potential to be applied for a wide range of problems.

**Table 3.6:** Results of Value function comparison. First column represents the means of empirical value functions. Second and third columns represent means of 50% and 25% quantiles of empirical value functions, respectively.

| | $V_n(d)$ | 50% quantiles | 25% quantiles |
|---:|---:|---:|---:|
| Dlearn | 55.2(11.7) | 45.89(15.5) | −24.74(12.69) |
| $l_1$-pls | 53.02(12.55) | 43.69(16.06) | −26.13(13.6) |
| rwl | 53.74(12.2) | 44.23(15.37) | -26.25(13.46) |
| rwl-GK | 53.29(12.17) | 43.59(12.93) | −25.59(12.93) |
| l2-DC-CVaR | 54.65(13.26) | 43.69(15.57) | -26.24(13.85) |
| l1-DC-CVaR | 50.69(11.82) | 38.22(13.74) | −29.96(12.55) |
| GK-DC-CVaR | **55.33**(11.85) | **45.91**(15.86) | −**23.02**(12.7) |
| l2-MM-CVaR | 52.89(13.18) | 41.68(15.8) | -27.78(13.83) |
| l1-MM-CVaR | 53.83(12.45) | 44.09(15.72) | −26.7(13.32) |
| GK-MM-CVaR | 53.68(12.81) | 44.63(15.77) | −25.92(13.35) |

## 3.7 Conclusion

In this paper, we propose two robust criteria to estimate optimal IDRs by considering individualized risk using the concept of CVaR. The resulting optimal IDRs can not only maximize the individualized expected outcome, but also prevent adverse consequences by controlling the lower tails of the outcome distributions.

Several possible extensions can be explored for future study. In particular, if we observe some additional risk outcome, such as side effect, for each subject, it would be interesting to develop methods to control this risk outcome under a pre-specified level by using our proposed criteria. This model was studied under the expected-value function framework given by (Wang et al., 2018b). Furthermore, (Wang et al., 2018a) recently used quantiles of outcome as criteria to identify optimal IDRs. They proposed a doubly robust estimation method to find the optimal IDR under their proposed criteria. It would be worthwhile to compare the performance of their methods with our proposed methods. Finally, from our numerical analysis, in some scenarios, estimated IDRs under $M_1(d)$ are better than those under $M_2(d)$. One possible reason is the potential model misspecification of $\alpha^*(X)$ given in (3.14), which we specify to be linear in the numerical study. Thus it would be desirable to explore broader structure of $\alpha(\mathbf{x})$ or develop some robust estimation methods to overcome potential model misspecification of $\alpha(\mathbf{x})$.

CHAPTER 4

# Estimation of Individualized Decision Rules Based on An Optimized Covariate-dependent Equivalent of Random Outcomes

## 4.1 Introduction

Most medical treatments are designed for "average patients". Due to the patients' heterogeneity, "one size fits all" medical treatment strategies can be very effective for some patients but not for others. For example, a study of colon cancer (Tan and Du, 2012) found that patients with a surface protein called KRAS are more likely to respond to certain antibody treatments than those without the protein. Thus exploration of precision medicine has recently gained a significant attention in scientific research. Precision medicine is a medical model that provides tailored health care for each specific patient, which has already demonstrated its success in saving lives (Bissonnette and Bergeron, 2012; Kummar et al., 2015). One of the main goals in precision medicine, from the data analytic perspective, is to estimate the optimal individualized decision rules (IDRs) that can improve the outcome of each individual.

### 4.1.1 Estimating optimal IDRs: the expected-outcome approach

An IDR is a decision rule that recommends treatments/actions to patients based on the information of their covariates. Consider the data collected from a single-stage randomized clinical trial involving different treatments. Before the trial, a patient's information $X$, such as blood pressure and past medicine history, is recorded. The enrolled patient will be randomly assigned to take a treatment denoted by $A$. After the patient receiving the treatment/action, the outcome $\mathcal{Z}$ of the patient can be observed. Without loss of generality, we may assume that the larger $\mathcal{Z}$ indicates the better condition a patient is in.

Let $\mathbb{P}$ be the probability distribution of the triplet $Y$ of random variables $(X, A, \mathcal{Z})$ and let $\mathbb{E}$ be the associated expectation operator, where $X$ is a random vector defined on the covariates space $\mathcal{X} \subseteq \mathbb{R}^p$, $A$ is a random variable defined on the finite treatment set $\mathcal{A}$ and $\mathcal{Z}$ is a scalar random variable representing outcome. The likelihood of $(X, A, \mathcal{Z})$ under $\mathbb{P}$ is defined as $f_0(x)\,\pi(a\,|\,x)\,f_1(z\,|\,x, a)$, where $f_0(x)$ is the probability density of $X$, $\pi(a\,|\,x)$ is the probability of patients being assigned treatment $a$ given $X = x$ and $f_1(z\,|\,x, a)$ is the conditional probability density of $\mathcal{Z}$ given covariates $X = x$ and treatment $A = a$. For the clinical trial study, the value of $\pi(a\,|\,x)$ is known; for the observational study, this value can be estimated via various methods such as multinomial logistic regression.

An IDR $d$ is defined as a mapping from the covariate space $\mathcal{X}$ into the action space $\mathcal{A}$. We let $\mathcal{D}$ be the class of all measurable functions mapping from $\mathcal{X}$ into $\mathcal{A}$; that is, $\mathcal{D}$ is the class of all measurable IDRs. For any IDR $d \in \mathcal{D}$, define $\mathbb{P}^{\,d}$ to be the probability distribution under which treatment $A$ is decided by $d$. Then the corresponding likelihood function under $\mathbb{P}^{\,d}$ is $f_0(x)\,\mathbb{I}(a = d(x))\,f_1(z\,|\,x, a)$, where the indicator function $\mathbb{I}(a = d(x))$ equals to 1 if $a = d(x)$ and 0 otherwise. Note that this is a discontinuous step function. The expected-value function (Qian and Murphy, 2011) based on $\mathbb{P}^{\,d}$ is given as $\mathbb{E}^{\,d}[\,\mathcal{Z}\,]$, which can be interpreted as the expected outcome under IDR $d$. It is known that if $\pi(a\,|\,X) \geq a_0 > 0$ almost surely (a.s.) for any $a \in \mathcal{A}$ and some constant $a_0$, then $\mathbb{P}^{\,d}$ is absolutely continuous with respect to $\mathbb{P}$ (Qian and Murphy, 2011). Thus by the Radon-Nikodym theorem,

$$\mathbb{E}^{\,d}[\,\mathcal{Z}\,] \,=\, \mathbb{E}\left[\,\mathcal{Z}\,\frac{\mathrm{d}\mathbb{P}^{\,d}}{\mathrm{d}\mathbb{P}}\,\right] \,=\, \mathbb{E}\left[\,\frac{\mathcal{Z}\,\mathbb{I}(A = d(X))}{\pi(A|X)}\,\right]. \tag{4.1}$$

In particular, $\mathbb{E}^{\,d}[\,c(X)\,] = \mathbb{E}[\,c(X)\,]$ for any integrable function $c$ of the covariate $X$ (Qian and Murphy, 2011). Given the triplet $(X, A, \mathcal{Z})$, an optimal IDR under the expected-value function framework is defined as

$$d_0 \,\in\, \operatorname*{argmax}_{d\in\mathcal{D}}\, \mathbb{E}^{\,d}[\,\mathcal{Z}\,].$$

This is the expected-value function maximization approach to the problem of estimating an optimal IDR to date. However, only maximizing the average of outcome under IDR $d$ may be restrictive in precision medicine. For example, when evaluating several treatments' effects on patients, doctors may want to know which treatment does the best to improve the outcome of a higher-risk patient. More importantly, due to the complex decision-making procedure in precision medicine, an "optimal" IDR that only maximizes the expected outcome of patients may lead to potentially adverse consequences for some patients. Therefore, considering individualized risk exposure is essential in precision medicine. This motivates us to examine the problem of determining optimal IDRs under a broader concept to control the individualized risk of each patient.

### 4.1.2 Optimized certainty equivalent

Estimating optimal IDRs can be regarded as an individualized decision-making problem. Utility functions have played an important role in such problems since they characterize the preference order over random variables, based on which decisions can be made. Guarding against the hazard of adverse decisions, risk measures are needed to balance the sole maximization of such utilities. This bi-objective consideration is well appreciated in portfolio management, leading to many risk measures since the early days of the mean-variance approach in (Markowitz, 1952). We refer the readers to (Rockafellar and Uryasev, 2013) and references therein for a contemporary perspective of diverse risk measures. Among such measures used in investment and economics, one of the most popular is the conditional-value-at-risk (CVaR) that has been extensively discussed in (Rockafellar and Uryasev, 2000; afellar and Uryasev, 2002); see the recent survey in (Sarykalin et al., 2008a). In general, for an essentially bounded random variable $\mathcal{Z}$ with the property that there exists a large enough scalar $B > 0$ such that the set $\{ \omega \in \Omega \mid |\mathcal{Z}(\omega)| > B \}$ has measure zero, where $\Omega$ is the sample space on which the random variable $\mathcal{Z}$ is defined, the $\gamma$-CVaR of $\mathcal{Z}$ is by definition:

$$\mathrm{CVaR}_{\gamma}(\mathcal{Z}) \triangleq \sup_{\eta \in \mathbb{R}} \left[ \eta - \frac{1}{\gamma} \mathbb{E} (\eta - \mathcal{Z})_{+} \right],$$

with $\gamma \in (0,1)$ and $t_{+} \triangleq \max(t,0)$ for a scalar (or vector) $t$. The smallest maximizer of $\mathrm{CVaR}_{\gamma}(\mathcal{Z})$ is the $\gamma$-quantile of $\mathcal{Z}$, which is also known as the value-at-risk (VaR). It turns out that the CVaR is a special case of an **Optimized Certainty Equivalent** (OCE) proposed in (Ben-Tal and Teboulle, 1986, 1987, 2007) that provides a link between utility and risk measures. In fact, the introduction of the OCE predates the popularity of the CVaR in portfolio management.

Let $\mathcal{U}$ denote the family of utility functions $u : \mathbb{R} \to [-\infty, \infty)$ that are upper semicontinuous, concave, and non-decreasing with a nonempty effective domain

$$\operatorname{dom}(u) \triangleq \{ t \in \mathbb{R} \mid u(t) > -\infty \} \neq \emptyset$$

such that $u(0) = 0$ and $1 \in \partial u(0)$, where $\partial u$ denotes the subdifferential map of $u$. Thus in particular,

$$[\, u(t) \geq 0, \; \forall\, t \geq 0 \,] \quad \text{and} \quad [\, u(t) \leq t, \; \forall\, t \in \mathbb{R} \,].$$

The OCE of an essentially bounded random variable $\mathcal{Z}$ is by definition:

$$\mathcal{O}_u(\mathcal{Z}) \triangleq \sup_{\eta \in \mathbb{R}} \left[\, \eta + \mathbb{E}\, u(\mathcal{Z} - \eta) \,\right].$$

According to the above cited references, the scalar $\eta$ is interpreted as the present consumption among the uncertain future income $\mathcal{Z}$. Then the sum $\eta + \mathbb{E}\, u(\mathcal{Z} - \eta)$ is the utility-based present value of $\mathcal{Z}$. Thus the goal of the OCE is to maximize the latter value by choosing an optimal allocation of $\mathcal{Z}$ between present and future consumption. A particular interest of the OCE is the case where $u(t) = \xi_1 \max(0, t) - \xi_2 \max(0, -t)$ for some constants $\xi_1$ and $\xi_2$ satisfying $0 \leq \xi_1 \leq 1 \leq \xi_2$. In this case, a maximizer of $\mathcal{O}_u(\mathcal{Z})$ corresponds to a quantile of the random variable $\mathcal{Z}$. For $\xi_1 = 0$, $\mathcal{O}_u(\mathcal{Z})$ reduces to the CVaR. With a proper truncation, a concave quadratic utility function can also satisfy the non-decreasing property, resulting in a mean-variance combination; see (Ben-Tal and Teboulle, 2007, Example 2.2). One special property of OCE is that $-\mathcal{O}_u(\mathcal{Z})$ gives a convex risk measure (Ben-Tal and Teboulle, 2007, Section 2.2). One of the limitations of the OCE, when applied to our problem of estimating optimal IDRs, is that it does not take into account covariates for the choice of an optimal allocation between present and future consumption when data on the covariates are available.

In this chapter, motivated by applications in the field of precision medicine, we **Individualize** the known concept of the OCE to a **Decision-Rule based Optimized Covariate-Dependent Equivalent** (IDR-CDE) that also incorporates domain covariates. The new equivalent not only broadens the traditional expectation–only based criterion in the estimation of the optimal IDRs in precision medicine, but also enriches the combined concept of utility and risk measures and bring them to individual-based decision making. The proposed IDR-CDE is very flexible so that different utility functions will produce different optimal IDRs for various purposes. It turns out that estimating optimal IDRs under the IDR-CDE is a challenging optimization problem since it involves the discontinuous function $\mathbb{I}(A = d(X))$. A major contribution of our work is that we overcome this technical difficulty by reformulating the estimation problem as a difference-of-convex (dc) constrained dc program under a mild assumption at the population level of the model. This reformulation allows us to employ a dc algorithm for solving the resulting dc program. Numerical results under the settings of binary actions and linear decision rules are presented to demonstrate the performance of our proposed model and algorithm.

### 4.1.3 Contributions and organization

The contributions of this chapter are in two directions: modeling and optimization. In the area of modeling, we extend the expected-value maximization approach in precision medicine to a more general framework by incorporating risk; see Section 4.2. This is accomplished through the extension of the OCE to the IDR-CDE in which we incorporate domain covariates and individualized decision rules. Properties of the IDR-CDE are derived in Subsection 4.2.1. The optimal IDR problem under the IDR-CDE criterion is formally defined in Subsection 4.2.2. Two cases of this problem are considered: the decomposable case (Subsection 4.2.3) and the general case via empirical maximization. Examples of the IDR-CDE given in Subsection 4.2.4 conclude the modeling part of the paper. Beginning in Section 4.3, the solution of the empirical IDR-CDE maximization is the other major topic of our work. The challenge of this problem is the presence of the discontinuous indicator function in the objective function. The cornerstone of our treatment of this problem is its epigraphical formulation which is valid under a mild assumption at the model's population level. We next introduce a piecewise affine description of the epigraphical constraints from which we obtain a difference-of-convex constrained optimization problem to be solved; see Sections 4.3 and 4.4. Although restricted to the empirical IDR-CDE maximization problem, we believe that our novel dc constrained programming treatment of the discontinuous optimization problem on hand can potentially be generalized to the composite optimization of univariate step functions with affine functions. In Section 4.5, we demonstrate the effectiveness of our proposed IDR-CDE optimization over the expected-value maximization via numerical results.

## 4.2 The IDR-based CDE

In this section, we extend the OCE along two directions. The first extension is to take the expectation $\mathbb{E}^d$ with respect to decision-rule based probability distribution $\mathbb{P}^d$ in order to evaluate the outcome under the IDR $d$. The second extension is to allow the deterministic scalar $\eta$ over which the supremum in the OCE is taken to be a family of measurable functions $\mathcal{F}$ defined on the covariate space $\mathcal{X}$. This family $\mathcal{F}$ allows the incorporation of available data representing covariate information for prediction and risk reduction; see the inequality (4.2) below. For notational purpose, we let $\mathcal{L}^r(\mathcal{X}, \Xi, \mathbb{P}_X)$ be the class of all measurable functions $f$ such that $\int |f(X)|^r \, d\mathbb{P}_X < \infty$ with $r \in [1, \infty]$. Here $(\mathcal{X}, \Xi, \mathbb{P}_X)$ is the measure space with $\Xi$ being the $\sigma$-algebra generated by $\mathcal{X}$, and $\mathbb{P}_X$ being the corresponding marginal probability measure of $X$.

### 4.2.1 Definition and properties

For an essentially bounded random variable $\mathcal{Z}$, the *individualized decision-rule based optimized covariate-dependent equivalent* (IDR-CDE) of $\mathcal{Z}$ under decision rule $d$ with respect to a utility function $u \in \mathcal{U}$ and a linear space $\mathcal{F} \subseteq \mathcal{L}^1(\mathcal{X}, \Xi, \mathbb{P}_X)$ is

$$\begin{aligned}
\mathcal{O}^d_{(u,\mathcal{F})}(\mathcal{Z}) &\triangleq \sup_{\alpha \in \mathcal{F}} \left[ \mathbb{E}\,\alpha(X) + \mathbb{E}^d u(\mathcal{Z} - \alpha(X)) \right] \\
&= \sup_{\alpha \in \mathcal{F}} \left[ \mathbb{E}\,\alpha(X) + \mathbb{E}\left( u(\mathcal{Z} - \alpha(X))\, \frac{\mathbb{I}(A = d(X))}{\pi(A|X)} \right) \right] \\
&= \sup_{\alpha \in \mathcal{F}} \mathbb{E}\left[ \left[ \alpha(X) + u(\mathcal{Z} - \alpha(X)) \right] \frac{\mathbb{I}(A = d(X))}{\pi(A|X)} \right],
\end{aligned}$$

where the last equality holds because of $\mathbb{E}[\alpha(X)] = \mathbb{E}^d[\alpha(X)]$ and the change of measure. The space $\mathcal{F}$ is taken to contain all constant functions and such that the expectations in $\mathcal{O}^d_{(u,\mathcal{F})}(\mathcal{Z})$ are taken over integrable functions. One example of such a space is a family of all bounded measurable functions. We will specify $\mathcal{F}$ for different utility functions in later discussion.

The following proposition gives two preliminary properties of the IDR-CDE. In particular, the inequality (4.2) bounds the IDR-CDE $\mathcal{O}^d_{(u,\mathcal{F})}(\mathcal{Z})$ of the random variable $\mathcal{Z}$ in terms of the OCE of $\mathcal{Z}$ in two ways: one is an upper bound in terms of the expected OCE of $\mathcal{Z}$ conditional on $X$ and $A = d(X)$, and the other one is a lower bound in terms of the decision-rule based OCE of $\mathcal{Z}$. A notable mention of both bounds is that they are independent of the family $\mathcal{F}$; see (4.2).

**Proposition 4.2.1.** *The following two statements hold.*
*(a) For any $u \in \mathcal{U}$, one has $\mathcal{O}^d_{(u,\mathcal{F})}(0) = 0$.*
*(b) For any linear space $\mathcal{F}$ containing all constant functions and for which $\mathcal{O}^d_{(u,\mathcal{F})}(\mathcal{Z})$ is finite,*

$$\mathbb{E}\left[\, \mathcal{O}_u(\mathcal{Z}|X, A = d(X))\,\right] \;\geq\; \mathcal{O}^d_{(u,\mathcal{F})}(\mathcal{Z}) \;\geq\; \sup_{\eta\in\mathbb{R}} \mathbb{E}^{\,d}\left[\, \eta + u(\mathcal{Z} - \eta)\,\right]. \tag{4.2}$$

*Proof.* (a) Since $u \in \mathcal{U}$, one has $u(t) \leq t$ and then

$$\mathcal{O}^d_{(u,\mathcal{F})}(0) \;\leq\; \sup_{\alpha\in\mathcal{F}} \left\{ \mathbb{E}\left[\, \alpha(X)\,\right] + \mathbb{E}^{\,d}\left[\, 0 - \alpha(X)\,\right] \right\} \;=\; 0,$$

where the last equality holds since $\mathbb{E}^{\,d}\left(\alpha(X)\right) = \mathbb{E}\left[\alpha(X)\right]$. Meanwhile, $u(0) = 0$ leads to

$$\mathcal{O}^d_{(u,\mathcal{F})}(0) \;\geq\; \mathbb{E}\left[\, 0\,\right] + \mathbb{E}^{\,d}\left[\, 0 - 0\,\right] \;=\; 0,$$

since $0 \in \mathcal{F}$. Combining the two inequalities gives the statement that $\mathcal{O}^d_{(u,\mathcal{F})}(0) = 0$.

(b) We can write

$$\begin{aligned}
\mathcal{O}^d_{(u,\mathcal{F})}(\mathcal{Z}) &= \sup_{\alpha\in\mathcal{F}} \left\{ \mathbb{E}\left[ \sum_{a\in\mathcal{A}} \mathbb{I}(d(X) = a)\, \mathbb{E}\,\left[\, \alpha(X) + u(\mathcal{Z} - \alpha(X)) \mid X, A = a\,\right] \right] \right\} \\
&= \sup_{\alpha\in\mathcal{F}} \left\{\, \mathbb{E}\left[\, \mathbb{E}\left[\, \alpha(X) + u(\mathcal{Z} - \alpha(X)) \mid X, A = d(X)\,\right]\right]\right\} \\
&= \sup_{\alpha\in\mathcal{F}} \left\{\, \mathbb{E}\left[\, \alpha(X) + \mathbb{E}\left[\, u(\mathcal{Z} - \alpha(X)) \mid X, A = d(X)\,\right]\right]\right\} \\
&\leq \mathbb{E}\left[ \sup_{s\in\mathbb{R}} \left\{\, s + \mathbb{E}\left[\, u(\mathcal{Z} - s) \mid X, A = d(X)\,\right]\right\} \right] \\
&= \mathbb{E}\left[\, \mathcal{O}_u(\mathcal{Z} \mid X, A = d(X))\,\right],
\end{aligned}$$

where the inequality holds because for any $\alpha(X)$, we have $\alpha(X) + \mathbb{E}\left[\, u(\mathcal{Z} - \alpha(X)) \mid X, A = d(X) \,\right] \leq \sup_{s\in\mathbb{R}} \left\{ s + \mathbb{E}\left[\, u(\mathcal{Z} - s) \mid X, A = d(X) \,\right] \right\}$. The right-hand inequality in (4.2) holds because $\mathcal{F}$ contains all constant functions. $\square$

Our proposed IDR-CDE measures the outcome $\mathcal{Z}$ via the decision-rule based optimal allocation between the covariate-dependent present value $\alpha(X)$ and the future gain $\mathcal{Z}-\alpha(X)$ under the utility function $u$. Unlike the original OCE, the allocation $\alpha(X)$ depends on the available covariate information $X$ such as environmental factors that can help to decide the optimal allocation. Take linear regression as an example; if the response $\mathcal{Z}$ can be predicted by the linear combination of covariates $X$, then covariates $X$ can explain some variability behind $\mathcal{Z}$; this could result in the reduction in the variance of $\mathcal{Z}$ given the information of $X$. Thus considering the broader covariate-based allocation $\alpha(X)$ could improve the allocation and further reduce the risk. This is also demonstrated via Proposition 4.2.1, by recalling that the negative of the standard OCE is a risk measure; indeed inequality (4.2) confirms that incorporating covariate information may lead to a reduced risk measure. Proposition 4.2.3 provides sufficient conditions for equality to hold between the IDR-CDE and the conditional OCE.

Note that $\mathcal{O}_u(\mathcal{Z} \mid X, A = d(X))$ is a random variable; it is the original OCE corresponding to the random variable with distribution being the conditional distribution of the random variable $\mathcal{Z}$ given $X$ and $A = d(X)$. Thus we may think of it as a conditional OCE. The IDR-CDE preserves many properties of the standard OCE which can be found in (Ben-Tal and Teboulle, 2007). The following are several of these properties.

**Proposition 4.2.2.** *Given the two triplets $(X, A, \mathcal{Z})$ and $(d, u, \mathcal{F})$, the following properties hold:*

*(a)* ***Shift Additivity****: for any essentially bounded random variable $\mathcal{Z}$ and any measurable function $c \in \mathcal{F}$ such that $c(X)$ is essentially bounded, $\mathcal{O}^d_{(u,\mathcal{F})}(\mathcal{Z} + c\,(X)) = \mathcal{O}^d_u(\mathcal{Z}) + \mathbb{E}\left[\, c\,(X) \,\right]$; in particular, $\mathcal{O}^d_{(u,\mathcal{F})}(c\,(X)) = \mathbb{E}\left[\, c\,(X) \,\right]$;*

*(b)* ***Consistency****: for any measurable function $\widehat{c}$ defined over $\mathcal{X} \times \mathcal{A}$ such that $\widehat{c}\,(X, A)$ is essentially bounded, $\mathcal{O}^d_{(u,\mathcal{F})}(\widehat{c}\,(X, A)) = \mathbb{E}\left[\, \widehat{c}\,(X, d(X)) \,\right]$;*

*(c). **Monotonicity**: for any two essentially bounded random variables $\mathcal{Z}_1$ and $\mathcal{Z}_2$ such that $\mathcal{Z}_1(\omega) \leq \mathcal{Z}_2(\omega)$ for almost all $\omega \in \Omega$, $\mathcal{O}^d_{(u,\mathcal{F})}(\mathcal{Z}_1) \leq \mathcal{O}^d_{(u,\mathcal{F})}(\mathcal{Z}_2)$;*

*(d). **Concavity**: for any two essentially bounded random variables $\mathcal{Z}_1$ and $\mathcal{Z}_2$ and any $\lambda \in (0,1)$,*

$$\mathcal{O}^d_{(u,\mathcal{F})}\left(\lambda\,\mathcal{Z}_1 + (1-\lambda)\,\mathcal{Z}_2\right) \,\geq\, \lambda\,\mathcal{O}^d_{(u,\mathcal{F})}(\mathcal{Z}_1) + (1-\lambda)\,\mathcal{O}^d_u(\mathcal{Z}_2).$$

*Proof.* (a) We have

$$\begin{aligned}
&\mathcal{O}^d_{(u,\mathcal{F})}(\mathcal{Z} + c\,(X)) \\
&= \sup_{\alpha \in \mathcal{F}} \Big\{ \mathbb{E}\,[\,\alpha(X)\,] + \mathbb{E}^{\,d}\,[\,u(\mathcal{Z} + c\,(X) - \alpha(X))\,] \Big\} \\
&= \mathbb{E}\,[\,c\,(X)\,] + \sup_{\alpha \in \mathcal{F}} \Big\{ \mathbb{E}[\,\alpha(X) - c\,(X)\,] + \mathbb{E}^{\,d}\,[\,u(\mathcal{Z} + c\,(X) - \alpha(X))\,] \Big\} \\
&= \mathbb{E}\,[\,c\,(X)\,] + \sup_{(\alpha - c) \in \mathcal{F}} \Big\{ \mathbb{E}\,[\,(\alpha - c)(X)\,] + \mathbb{E}^{\,d}\,[\,u(\mathcal{Z} - (\alpha - c)(X))\,] \Big\} \\
&= \mathbb{E}\,[\,c\,(X)\,] + \mathcal{O}^d_{(u,\mathcal{F})}(\mathcal{Z}),
\end{aligned}$$

where the third equality holds since $\mathcal{F}$ is a linear space.

(b) Since $u(t) \leq t$, we have

$$\mathcal{O}^d_{(u,\mathcal{F})}(\mathcal{Z}) \,\leq\, \sup_{\alpha \in \mathcal{F}} \Big\{ \mathbb{E}\,[\,\alpha(X)\,] + \mathbb{E}^{\,d}\,[\,\mathcal{Z} - \alpha(X)\,] \Big\} \,=\, \mathbb{E}^{\,d}\,[\,\mathcal{Z}\,],$$

where the equality holds because $\mathbb{E}^{\,d}\,[\,\alpha(X)\,] = \mathbb{E}\,[\,\alpha(X)\,]$ by the definition of $\mathbb{P}^{\,d}$. Therefore, if $\mathcal{Z} = \hat{c}(X, A)$ is essentially bounded, then

$$\begin{aligned}
\mathcal{O}^d_{(u,\mathcal{F})}(\mathcal{Z}) &\leq \mathbb{E}^{\,d}\,[\,\widehat{c}\,(X, A)\,] \\
&= \mathbb{E}\left[\,\frac{\widehat{c}\,(X, A)\,\mathbb{I}(d(X) = A)}{\pi(A|X)}\,\right] \\
&= \mathbb{E}\left[\,\frac{\widehat{c}\,(X, d(X))\,\mathbb{I}(d(X) = A)}{\pi(A|X)}\,\right] \,=\, \mathbb{E}\,\left[\,\widehat{c}\,(X, d(X))\,\right].
\end{aligned}$$

Since $u(0) = 0$, by the definition of the supreme in $\mathcal{O}^{d}_{(u,\mathcal{F})}$, we derive

$$\begin{aligned}
\mathcal{O}^{d}_{(u,\mathcal{F})}(\widehat{c}\,(X,A)) &\geq \mathbb{E}\,[\,\widehat{c}\,(X,d(X))\,] + \mathbb{E}^{\,d}\,[\,u(\widehat{c}(X,A) - \widehat{c}\,(X,d(X))\,] \\
&= \mathbb{E}\,[\,\widehat{c}\,(X,d(X))\,] + \mathbb{E}^{\,d}\,[\,u(\widehat{c}\,(X,d(X)) - \widehat{c}\,(X,d(X)))\,] \\
&= \mathbb{E}\,[\,\widehat{c}\,(X,d(X))\,].
\end{aligned}$$

Thus, $\mathcal{O}^{d}_{(u,\mathcal{F})}(\widehat{c}\,(X,A)) = \mathbb{E}\,[\,\widehat{c}\,(X,d(X))\,]$.

(c) If $\mathcal{Z}_1 \leq \mathcal{Z}_2$, then $\mathcal{Z}_1 - \alpha(X) \leq \mathcal{Z}_2 - \alpha(X)$ for $\alpha \in \mathcal{F}$. Since $u \in U_0$ is a non-decreasing utility function, it follows that

$$\begin{aligned}
\mathcal{O}^{d}_{(u,\mathcal{F})}(\mathcal{Z}_1) &= \sup_{\alpha\in\mathcal{F}} \Big\{ \mathbb{E}\,[\,\alpha(X)\,] + \mathbb{E}^{\,d}\,[\,u(\mathcal{Z}_1 - \alpha(X))\,] \Big\} \\
&\leq \sup_{\alpha\in\mathcal{F}} \Big\{ \mathbb{E}\,[\,\alpha(X)\,] + \mathbb{E}^{\,d}\,[\,u(\mathcal{Z}_2 - \alpha(X))\,] \Big\} = \mathcal{O}^{d}_{(u,\mathcal{F})}(\mathcal{Z}_2).
\end{aligned}$$

(d) For any $\lambda \in (0,1)$, denote a random variable $\mathcal{Z}_\lambda \triangleq \lambda\,\mathcal{Z}_1 + (1-\lambda)\,\mathcal{Z}_2$ and a measurable function $\alpha_\lambda(X) \triangleq \lambda\,\alpha_1(X) + (1-\lambda)\,\alpha_2(X)$. Clearly $\mathcal{Z}_\lambda$ is essentially bounded and $\alpha_\lambda(X) \in \mathcal{F}$. Then by the concavity of $u$, we have

$$\begin{aligned}
\mathbb{E}\,[\,\alpha_\lambda(X)\,] + \mathbb{E}^{\,d}\,[\,u(\mathcal{Z}_\lambda - \alpha_\lambda(X))\,] \;\geq\; &\lambda \big( \mathbb{E}\,[\,\alpha_1(X)\,] + \mathbb{E}^{\,d}\,[\,u(\mathcal{Z}_1 - \alpha_1(X))\,] \big) + \\
&(1-\lambda) \big( \mathbb{E}\,[\,\alpha_2(X)\,] + \mathbb{E}^{\,d}\,[\,u(\mathcal{Z}_2 - \alpha_2(X))\,] \big).
\end{aligned}$$

Taking supremum over $\alpha_1$ and $\alpha_2$ on both sides, we may derive the stated result. □

Properties (a) and (b) extend corresponding results of the original OCE (Ben-Tal and Teboulle, 2007, Theorem 2.1) from a constant $\eta$ to a measurable function that depends on $X$ and $A$; properties (c) and (d) are essentially the same as those in (Ben-Tal and Teboulle, 2007, Theorem 2.1). These properties justify the use of the IDR-CDE in decision making. Shift Additivity means if the outcome is shifted by some function over covariates, the IDR-CDE measure is shifted by the average of this function. Thus the IDR $d$ is invariant under such a shift. Consistency means that to evaluate the IDR-CDR of a measurable function over

$\mathcal{X} \times \mathcal{A}$ is equivalent to evaluating the expectation of this random function when the action follows the decision rule $d$. Monotonicity and concavity have the same respective meanings as the OCE: the former guarantees a larger CDE for a (stochastically) larger outcome; the latter ensures that the IDR-CDE of a convex combination of two outcomes given a decision rule $d$ is always better than only considering each single outcome separately; this property encourages the simultaneous combination of multiple outcomes for better results.

### 4.2.2 The IDR optimization problem

We employ the IDR-CDE to evaluate the decision rule $d$ of the outcome $\mathcal{Z}$ via its optimized covariate equivalent, with the goal of estimating an optimal IDR that maximizes the IDR-CDE given the pair $(u, \mathcal{F})$ in the following sense.

**Definition 4.1.** Given the triplet $(X, A, \mathcal{Z})$, the pair $(u, \mathcal{F})$, and the family $\mathcal{D}$ of decision rules, an optimal IDR is a rule $d^*$ such that

$$d^*(X) \in \operatorname*{argmax}_{d \in \mathcal{D}} \mathcal{O}^{d}_{(u,\mathcal{F})}(\mathcal{Z}),$$

if such a maximizer exists. □

Thus we can compute $d^*(X)$ and the optimal allocation $\alpha^*(X)$ jointly by solving

$$\sup_{d \in \mathcal{D}, \alpha \in \mathcal{F}} \mathbb{E}\,[\,\alpha(X)\,] + \mathbb{E}^{\,d}\,[\,u(\mathcal{Z} - \alpha(X))\,]. \tag{4.3}$$

The rest of the paper is devoted to the solution of this optimization problem. The discussion is divided into two cases depending on whether we can exchange the supremum over $\alpha$ and the expectation $\mathbb{E}^{\,d}$ in $\mathcal{O}^{d}_{(u,\mathcal{F})}(\mathcal{Z})$. The exchangeable case requires the theory of decomposable space from variational analysis; this leads to an "explicit" determination of the optimal IDR via the evaluation of the conditional OCE given the covariate $X$ and the finite actions $a \in \mathcal{A}$; see Proposition 4.2.4. The general case requires the numerical solution of an empirical optimization problem obtained from sampling of the covariates among available data.

### 4.2.3 Decomposable space and normal integrand

In order to exchange the supreme over $\alpha(X)$ and expectation with respect to $\mathbb{E}^{d}$, we need to first introduce the concept of a decomposable space and the normal integrand.

**Definition 4.2.** (Rockafellar and Wets, 2009, Definitions 14.59 and 14.27). A space $\mathcal{M}$ of $\mathcal{B}_0$-measurable functions is *decomposable* relative to an underlying measure space $(\Omega_0, \mathcal{B}_0, \mu)$ if for every function $x_0 \in \mathcal{M}$, every set $G \in \mathcal{B}_0$ with $\mu(G) < \infty$ and any bounded, measurable function $x_1$, the function $x_2(t) = x_0(t)\mathbb{I}(t \notin G) + x_1(t)\mathbb{I}(t \in G)$ belongs to $\mathcal{M}$. An extended-value function $f : \Omega_0 \times \mathbb{R} \to (-\infty, \infty]$ is a *normal integrand* if its epigraphical mapping $\omega \to \text{epi } f(\omega, \cdot)$ is closed-valued and measurable. □

The space $\mathcal{L}^r(\mathcal{X}, \Xi, \mathbb{P}_{\mathcal{X}})$ is decomposable for $r \in [1, \infty]$ but the family of constant functions is not decomposable. These facts will be used in the examples to be discussed in the next subsection.

We will employ the following simplified version of (Rockafellar and Wets, 2009, Theorem 14.60) that provides the required conditions for the exchange of the supremum and expectation in our context.

**Theorem 4.2.1.** *Let ($\Omega_0$, $\mathcal{B}_0$, $\mu$) be a probability measure space, and $\mathcal{M}$ be a decomposable space of $\mathcal{B}_0$-measurable functions. Let $f : \Omega_0 \times \mathbb{R} \to (-\infty, \infty]$ be a normal integrand; let the integral functional $I_f(x) = \int_{\Omega_0} f(x(\omega), \omega) d\mu(\omega)$ be defined on $\mathcal{M}$. The following two statements hold:*

*(a)* $\inf_{x \in \mathcal{M}} \int_{\Omega_0} f(x(\omega), \omega) d\mu(\omega) = \int_{\Omega_0} \inf_{s \in \mathbb{R}} f(s, \omega) d\mu(\omega)$ *as long as $I_f(x)$ is finite; and*

*(b)* $x_0 \in \operatorname{argmin}_{x \in \mathcal{M}} I_f(x) \Longleftrightarrow x_0(\omega) \in \operatorname{argmin}_{s \in \mathbb{R}} f(s, \omega)$ *almost surely.* □

The following proposition shows that if $\mathcal{F}$ is decomposable, then equality holds between the IDR-CDE and the conditional OCE.

**Proposition 4.2.3.** *If $\mathcal{F}$ is a decomposable space relative to $(\mathcal{X}, \Xi, \mathbb{P}_X)$, then*

$$\mathcal{O}^{d}_{(u,\mathcal{F})}(\mathcal{Z}) = \mathbb{E}\,[\,\mathcal{O}_u(\mathcal{Z} \mid X, A = d(X))\,].$$

*Proof.* Note that $\mathbb{E}\,[\,\alpha(X) + u(\mathcal{Z} - \alpha(X)) \mid X, A = d(X)\,]$ is measurable with respect to $X$ and upper semi-continuous with respect to $\alpha(X)$ for any $X$, thus is a normal integrand (Rockafellar and Wets, 2009, Example 14.31). Hence we have

$$\begin{aligned}\mathcal{O}^d_{(u,\mathcal{F})}(\mathcal{Z}) &= \sup_{\alpha\in\mathcal{F}} \{\, \mathbb{E}\,[\, \mathbb{E}\,[\, \alpha(X) + u(\mathcal{Z} - \alpha(X)) \mid X, A = d(X)\,]\,]\,\} \\ &= \mathbb{E}\left[ \sup_{s\in\mathbb{R}} \{\, s + \mathbb{E}\,[\, u(\mathcal{Z} - s) \mid X, A = d(X)\,]\,\} \right] \\ &= \mathbb{E}\,[\, \mathcal{O}_u(\mathcal{Z} \,|\, X, A = d(X))\,],\end{aligned}$$

where the second equality is by Theorem 4.2.1 because $\mathcal{F}$ is decomposable and $\mathcal{Z}$ is bounded. □

**Remark 2.** Since the conditional OCE is independent of the space $\mathcal{F}$, it follows that so is $\mathcal{O}^d_{(u,\mathcal{F})}(\mathcal{Z})$ provided that $\mathcal{F}$ is decomposable relative to $(\mathcal{X}, \Xi, \mathbb{P}_X)$. Thus, in the following, if we specify $\mathcal{F}$ to be decomposable, then we omit $\mathcal{F}$ and write the IDR-CDE of the random variable $\mathcal{Z}$ as $\mathcal{O}^d_u(\mathcal{Z})$. □

As a result of Proposition 4.2.3, we can characterize the optimal IDR explicitly if $\mathcal{F}$ is a decomposable space. We recall that $\mathcal{A}$ is a finite set.

**Proposition 4.2.4.** *For a given decomposable space $\mathcal{F}$ and utility function $u \in \mathcal{U}$, an optimal IDR is given by*

$$d^*(X) \in \underset{a\in\mathcal{A}}{\operatorname{argmax}}\ \mathcal{O}_u(\mathcal{Z} \,|\, X, A = a). \tag{4.4}$$

*Proof.* By the definition of $\mathcal{O}^d_u(\mathcal{Z})$, we have for any $d \in \mathcal{D}$,

$$\begin{aligned}\mathbb{E}\,[\, \mathcal{O}_u(\mathcal{Z} \,|\, X, A = d(X))\,] &= \mathbb{E}\left[ \sum_{a\in\mathcal{A}} \mathbb{I}(d(X) = a)\, \mathcal{O}_u(\mathcal{Z} \,|\, X, A = a) \right] \\ &\leq \mathbb{E}\left[ \sum_{a\in\mathcal{A}} \mathbb{I}(d(X) = a) \max_{a'\in\mathcal{A}} \mathcal{O}_u(\mathcal{Z} \,|\, X, A = a') \right] \\ &= \mathbb{E}\left[ \left( \max_{a'\in\mathcal{A}} \mathcal{O}_u(\mathcal{Z} \,|\, X, A = a') \right) \sum_{a\in\mathcal{A}} \mathbb{I}(d(X) = a) \right] \\ &= \mathbb{E}\left[ \max_{a'\in\mathcal{A}} \mathcal{O}_u(\mathcal{Z} \,|\, X, A = a') \right].\end{aligned}$$

Therefore if (4.4) holds, then $d^*$ is maximizing. Such a $d^*$ is a measurable function because being an optimal IDR, $d^*(X) = a$ if and only if $\mathcal{O}_u(\mathcal{Z} \,|\, X, A = a) \geq \max_{a'\neq a} \mathcal{O}_u(\mathcal{Z} \,|\, X, A = a')$ and $\mathcal{O}_u(\mathcal{Z} \,|\, X, A = a) \geq \max_{a'\neq a} \mathcal{O}_u(\mathcal{Z} \,|\, X, A = a')$ is a measurable set with respect to $X$. □

**Remark 3.** The explicit expression of an optimal IDR is valid only when the space $\mathcal{F}$ is decomposable. If the conditional distribution of $\mathcal{Z}$ given $X$ and $A = a$ is known, then it is possible to compute the individualized OCE $\mathcal{O}_u(\mathcal{Z} \mid X, A = a)$ directly. For example, if we make certain parametric assumptions on this conditional distribution, we may be able to estimate these parameters based on the collected data and obtain optimal IDRs based on Proposition 4.2.4. This is similar to the model-based methods in the literature of the expected-value function maximization approach. However, the empirical performance could be affected by the possible model misspecification. Therefore the individualized OCE $\mathcal{O}_u(\mathcal{Z} \mid X, A = a)$ is primarily a conceptual notion and the expression (4.4) is mainly for interpretation. □

According to Proposition 4.2.4, an optimal IDR under our proposed CDE can be obtained by choosing the decision rule with the largest individualized OCE. In the next subsection, we will characterize the IDR-OCE via several illustrative examples for both decomposable and non-decomposable families of covariate functions.

### 4.2.4 Illustrative examples

We present several common utility functions to further explain the IDR-CDE for individualized decision making. We will focus on two families: $\mathcal{L}^r(\mathcal{X}, \Xi, \mathbb{P}_X)$ for some $r \in [1, \infty]$ and a family of constant function which we denote $\mathcal{F}_c$. The former family is a decomposable linear space and the latter family is not decomposable.

**Example 4.1.** (Identity utility function) Let $u(t) = t$, then by the definition, we can obtain $\mathcal{O}_u^d(\mathcal{Z}) = \mathbb{E}^d[\mathcal{Z}]$ for both families $L^1(\mathcal{X}, \Xi, \mathbb{P}_X)$ and $\mathcal{F}_c$. This recovers the expected-value maximization framework in the existing literature of precision medicine. By Proposition 4.2.4, for the family $\mathcal{L}^r(\mathcal{X}, \Xi, \mathbb{P}_X)$, an optimal IDR under the identity utility function is given by:

$$d^*(X) \in \underset{a \in \mathcal{A}}{\operatorname{argmax}} \, \mathbb{E}[\mathcal{Z} \mid X, A = a],$$

which is equivalent to the action with the largest expected outcome $\mathcal{Z}$ among all the actions given covariates $X$. ◊

**Example 4.2.** (Piecewise Linear Utility Function) Let

$$u(t) = \xi_1 \max(0, t) - \xi_2 \max(0, -t), \quad \text{where } 0 \le \xi_1 < 1 < \xi_2.$$

It can be verified that $u \in U_0$.

**(a) Decomposable space:** $\mathcal{F} = L^1(\mathcal{X}, \Xi, \mathbb{P}_X)$. The corresponding IDR-CDE is

$$\mathcal{O}_u^d(\mathcal{Z}) = \sup_{\alpha \in \mathcal{F}} \left\{ \begin{array}{l} \mathbb{E}\,[\,\alpha(X)\,] + \\ \\ \mathbb{E}^d\,[\,\xi_1 \max(\,0\,,\, \mathcal{Z} - \alpha(X)\,) - \xi_2 \max(\,0\,,\, \alpha(X) - \mathcal{Z}\,)\,] \end{array} \right\}. \tag{4.5}$$

Based on Proposition 4.2.3 we can write it as: with $\gamma \triangleq \dfrac{1-\xi_1}{\xi_2 - \xi_1}$. Then the $\mathcal{O}_u^d(\mathcal{Z})$ is equal to

$$\begin{aligned} & \mathbb{E}\,\left[\sup_{s \in \mathbb{R}} \left\{\, s + \mathbb{E}\,[\,\xi_1\, \max(\,0, \mathcal{Z} - s\,) - \xi_2\, \max(\,0\,,\, s - \mathcal{Z}\,) \mid X, A = d(X)\,]\,\right\}\right] \\ = \; & \xi_1 \mathbb{E}^d\,[\,\mathcal{Z}\,] + (1 - \xi_1)\, \mathbb{E}\,\left[\sup_{s \in \mathbb{R}} \left\{ s - \tfrac{1}{\gamma}\, \mathbb{E}\,[\max(\,0\,,\, s - \mathcal{Z}\,) \mid X, A = d(X)\,] \right\}\right] \\ = \; & \xi_1\, \mathbb{E}^d\,[\,\mathcal{Z}\,] + (1 - \xi_1)\, \mathbb{E}\,[\,\mathrm{CVaR}_\gamma\,(\mathcal{Z} \mid X, A = d(X))\,], \end{aligned}$$

where given $X$, the corresponding supremum is attained at the $\gamma$-quantile of conditional distribution of $\mathcal{Z}$ on $X$ and $A = d(X)$ almost surely. Therefore, under the piecewise affine utility function, $\mathcal{O}_u^d(\mathcal{Z})$ can be interpreted as a convex combination of the expected value of $\mathcal{Z}$ and its expected CVaR given IDR $d$. Thus this $\mathcal{O}_u^d(\mathcal{Z})$ considers both $\mathbb{E}^d\,[\,\mathcal{Z}\,]$ and CVaR of the outcome simultaneously. In particular, when $\xi_1 = \xi_2 = 1$, this recovers Example 4.1.

By Proposition 4.2.4, a corresponding optimal IDR is

$$d^*(X) \in \operatorname*{argmax}_{a \in \mathcal{A}} \left\{\, \xi_1\, \mathbb{E}\,[\,\mathcal{Z} \,|\, X, A = a\,] + (1 - \xi_1)\, \mathrm{CVaR}_\gamma(\mathcal{Z} \,|\, X, A = a)\,\right\}. \tag{4.6}$$

Therefore, under this piecewise affine utility function, an optimal IDE is to choose the action with the largest convex combination of expected outcome and CVaR of outcome $\mathcal{Z}$ among all the actions given covariates $X$. □

**(b) Family of constant functions:** $\mathcal{F} = \mathcal{F}_c$. The IDE-CDE reduces to (Ben-Tal and Teboulle, 2007, Example 2.3) with IDR $d$ involved:

$$\begin{aligned}\mathcal{O}^{d}_{(u,\mathcal{F}_c)}(\mathcal{Z}) &= \sup_{c\in\mathbb{R}}\Big\{\, c + \mathbb{E}^{\,d}\,[\,\xi_1\,\max\,(\,0\,,\,c-\mathcal{Z}\,) - \xi_2\,\max\,(\,0\,,\,c-\mathcal{Z}\,)\,]\,\Big\} \\ &= \xi_1\,\mathbb{E}^{\,d}\,[\,\mathcal{Z}\,] + (1-\xi_1)\,\sup_{c\in\mathbb{R}}\left\{\, c - \frac{\xi_2-\xi_1}{1-\xi_1}\,\mathbb{E}^{\,d}\,[\,\max\,(\,0\,,\,c-\mathcal{Z}\,)\,]\,\right\}.\end{aligned}$$

The supremum in the right-hand side is any $c^*$ satisfying $\mathbb{P}^{\,d}\,(\mathcal{Z} \le c^*) \ge \gamma$ and $\mathbb{P}^{\,d}\,(\mathcal{Z} \ge c^*) \le 1-\gamma$, which is the $\gamma$-quantile of $\mathcal{Z}$ under the probability distribution $\mathbb{P}^{\,d}$, denoted by $Q^d_\gamma(\mathcal{Z})$. The corresponding maximum value is $\xi_1\,\mathbb{E}^{\,d}\,[\,\mathcal{Z}\,] + (1-\xi_1)\,\mathrm{CVaR}^d_\gamma(\mathcal{Z})$. By definition, an optimal IDR under $\mathcal{F}_c$ is given by

$$d^* \in \underset{d}{\operatorname{argmax}}\;\Big\{\xi_1\mathbb{E}^{\,d}\,[\,\mathcal{Z}\,] + (1-\xi_1)\,\mathrm{CVaR}^d_\gamma(\mathcal{Z})\,\Big\}.$$

While this expression is insightful, the above optimal IDR $d^*$ does not have an explicit form as (4.6) since Proposition 4.2.4 no longer holds by the fact that $\mathcal{F}_c$ is not a decomposable space.

◊

**Example 4.3.** Quadratic Utility Let

$$u(t) \;=\; \left\{\begin{array}{ll} t - \dfrac{1}{2\tau}t^2 & \text{if } t \,\le\, \tau \\ \tau/2 & \text{otherwise,}\end{array}\right\}, \;\text{ where } \tau \,=\, \sup_{\omega\in\Omega}\mathcal{Z}(\omega) - \inf_{\omega\in\Omega}\mathcal{Z}(\omega),$$

be a quadratic function truncated to be an admissible utility function in the family $\mathcal{U}$ and to adopt to the range of the random outcome $\mathcal{Z}$. Note that $u$ is continuously differentiable with derivative $u'(t) = \left(1 - \dfrac{t}{\tau}\right)\,\mathbb{I}(t \le \tau)$.

**(a) Decomposable space:** $\mathcal{F} = \mathcal{L}^2(\mathcal{X}, \Xi, \mathbb{P}_X)$. By Proposition 4.2.3, we have,

$$\begin{aligned}\mathcal{O}^d_u(\mathcal{Z}) \;&=\; \mathbb{E}\left[\sup_{s\in\mathbb{R}}\,\{\,s + \mathbb{E}\,[\,u(\mathcal{Z}-s)\;|\;X, A = d(X)\,]\,\}\right] \\ &= \mathbb{E}^{\,d}\,[\,\mathcal{Z}\,] - \mathbb{E}\left[\frac{1}{2\tau}\,\mathbb{E}\,\Big[\,(\,\mathcal{Z} - \mathbb{E}\,[\,\mathcal{Z}\,|\,X, A = d(X)\,]\,)^2\;|\,X, A = d(X)\Big]\right] \\ &= \mathbb{E}^{\,d}\,[\,\mathcal{Z}\,] - \frac{1}{2\tau}\,\mathbb{E}\,[\,\mathrm{var}(\mathcal{Z}\,|\,X, A = d(X))\,]\,,\end{aligned}$$

where the supreme $\alpha^*(X) = \mathbb{E}[\mathcal{Z} \mid X, A = d(X)]$ almost surely and var$(\bullet)$ is the variance of a random variable. The second equality is based on (Ben-Tal and Teboulle, 2007, Remark 2.1) by noting that $1 + \mathbb{E}[u'(\mathcal{Z} - \alpha^*(X))] = 0$. The interchange between expectation and derivative is justified by the dominated convergence theorem under the restriction that $s \in \left[\inf_{\omega\in\Omega} \mathcal{Z}(\omega), \sup_{\omega\in\Omega} \mathcal{Z}(\omega)\right]$. Thus $\mathcal{O}_u^d(\mathcal{Z})$ can be interpreted as the (individualized) mean-variance risk measure under the decision rule $d$, generalizing the mean-variance criterion in the absence of $A$ and $X$, which is frequently used in portfolio selection. An optimal IDR is given by

$$d^*(X) \in \operatorname*{argmax}_{a\in\mathcal{A}} \left\{ \mathbb{E}[\mathcal{Z} \mid X, A = a] - \frac{1}{2\tau} \operatorname{var}[\mathcal{Z} \mid X, A = a] \right\},$$

which suggests the optimal action to maximize the expected outcome balanced with the variance given covariates $X$.

**(b) Family of constant functions:** $\mathcal{F} = \mathcal{F}_c$. Similar to part (a) above, direct computation yields $\mathcal{O}_u^d(\mathcal{Z}) = \mathbb{E}^d[\mathcal{Z}] - \frac{1}{2\tau} \operatorname{var}^d(Z)$ with $c^* = \mathbb{E}^d[\mathcal{Z}]$, where $\operatorname{var}^d(Z)$ denotes the variance of a random variable $\mathcal{Z}$ under $\mathbb{P}^d$. An optimal IDR under $\mathcal{F}_c$ is

$$\operatorname*{argmax}_{d} \left\{ \mathbb{E}^d[\mathcal{Z}] - \frac{1}{2\tau} \operatorname{var}^d(Z) \right\},$$

which requires further evaluation by a numerical procedure. $\Diamond$

From Example 4.2 and Example 4.3, we see that one of the differences between a covariate-dependent $\alpha(X)$ and a constant $\alpha(X) \in \mathcal{F}_c$ lies in that for the former, the IDR-CDE considers expected individualized OCE given the decision rule $d$, but for a constant $\alpha$, in contrast, the IDR-CDE considers only the OCE of the random variable $\mathcal{Z}$ under $\mathbb{P}^d$. To further understand this difference, consider a toy example with $\mathcal{Z} = X_1 A + \varepsilon$, where both $X_1$ and $\varepsilon$ independently follow the standard normal distribution. Suppose we use the utility function in Example 4.2(b) with $\xi_1 = 0$ and $\xi_2 = 2$ to evaluate an IDR $d(X_1) = 1$. By calculation, $c^* = 0$ and thus we are focused on the median of $\mathcal{Z}$ under the probability distribution $\mathbb{P}^d$. The corresponding $\mathcal{O}^d_{(u,\mathcal{F}_c)}(\mathcal{Z}) = \mathbb{E}[\mathbb{E}[\mathcal{Z}\mathbb{I}(\mathcal{Z} \leq 0)|X, A = 1]]$. If we have one patient with covariate $X_1 = -2$, then $\mathbb{P}(\mathcal{Z} \leq 0|X_1 = -2, A = 1) \approx 84\%$. For this patient, $\mathcal{O}^d_{(u,\mathcal{F}_c)}(\mathcal{Z})$ evaluates the outcome lower than about 84%-quantile, which is not satisfactory. As a result, we may conclude that the optimal IDR cannot be quantified by comparing each action separately of each other when considering $\alpha(X)$ being constant functions only. Consequently such an IDR cannot control the individualized OCE.

So far we only consider single-stage individualized decision making problems. It is also meaningful to extend our proposed IDR-CDE to multi-stage decision-making scenarios in order to deliver time-varying optimal IDRs with risk exposure control. Since it will require advanced modeling and treatment, we leave such an extension for future research.

### 4.3 The Empirical IDR Optimization Problem

In this section, we discuss how to numerically solve the optimization problem (4.3) at the empirical level without assuming any data generating mechanisms. In the following, we focus on estimating the optimal IDR with $\mathcal{A} = \{-1, 1\}$, i.e., a binary action space. Further, for computational purposes, we restrict the decision rule to be given by: $d(X) = \operatorname{sign}(f(X;\theta))$ for a parametric linear estimation function: $f(X;\theta) = \beta^T X + \beta_0 = \theta^T \widehat{X}$, where $\theta \triangleq \begin{pmatrix} \beta \\ \beta_0 \end{pmatrix} \in \mathbb{R}^{p+1}$ contains the unknown coefficients to be estimated and $\widehat{X} \triangleq \begin{pmatrix} X \\ 1 \end{pmatrix}$. Extensions to multi-action space and nonlinear decision rules are possible but will necessitate advanced modeling and treatment. This will be left for future research. Using functional margin representation in standard classification, we then have $\mathbb{I}\,(A = d(X)) = \mathbb{I}\,(A\, f(X;\theta) > 0)$ for any nonzero $f(X;\theta)$. Therefore, the IDR-CDE optimiation problem can be equivalently written as:

$$\sup_{\theta \triangleq (\beta,\beta_0) \in \mathbb{R}^{p+1},\, \alpha \in \mathcal{F}} \left\{ \begin{array}{l} \mathbb{E}\left[ \mathcal{Z}\, \dfrac{\mathbb{I}(A\, f(X;\theta) > 0)}{\pi(A|X)} \right] + \\ \mathbb{E}\left[ [\, \alpha(X) - \mathcal{Z} + u(\mathcal{Z} - \alpha(X)) \,] \, \dfrac{\mathbb{I}(A\, f(X;\theta) > 0)}{\pi(A|X)} \right] \end{array} \right\}. \tag{4.7}$$

Before proceeding, we describe two characteristics of this problem that are important in the algorithmic development and provide our proposal to address them.

**(a) The discontinuity of the indicator function.** The function $\mathbb{I}(A\, f(X;\theta) > 0)$ is a lower semicontinuous, albeit discontinuous function. This seems to prohibit us from employing continuous optimization algorithms to solve problem (4.7). A natural way to resolve this issue is to approximate the indicator function by a continuous function, such as the piecewise truncated hinge loss as in (Wu and Liu, 2007):

$$T_\delta(x) \triangleq \frac{1}{2\,\delta} \underbrace{\left[\max\left(\,x+\delta, 0\,\right) - \max\left(\,x-\delta, 0\,\right)\right]}_{\text{nonnegative}} \quad \text{for some } \delta > 0,$$

so that

$$\begin{aligned} \mathrm{I\!I}\,(A\,f(X;\theta)\,>\,0) \;\;&\approx\;\; T_\delta(A\,f(X;\theta)) \\ &=\;\; \underbrace{\frac{1}{2\,\delta}\,\max\left(\,A\,f(X;\theta)+\delta, 0\,\right)}_{\text{denoted } T_\delta^+(\theta;X,A)} - \underbrace{\frac{1}{2\,\delta}\,\max\left(\,A\,f(X;\theta)-\delta, 0\,\right)}_{\text{denoted } T_\delta^-(\theta;X,A)}, \end{aligned}$$

where both functions $T_\delta^\pm(\bullet;X,A)$ are nonnegative, convex, and piecewise affine; thus the approximating function is non-convex and non-differentiable, making the resulting optimization problem:

$$\sup_{\substack{\theta\triangleq(\beta,\beta_0)\in\mathbb{R}^{p+1},\\ \alpha\in\mathcal{F}}} \left\{ \begin{array}{l} \mathrm{I\!E}\left[\,\mathcal{Z}\,\dfrac{T_\delta^+(\theta;X,A)-T_\delta^-(\theta;X,A)}{\pi(A|X)}\,\right]+ \\[2ex] \mathrm{I\!E}\left[\,[\,\alpha(X)-\mathcal{Z}+u(\mathcal{Z}-\alpha(X))\,]\;\dfrac{T_\delta^+(\theta;X,A)-T_\delta^-(\theta;X,A)}{\pi(A|X)}\,\right] \end{array} \right\} \tag{4.8}$$

difficult to solve. Since we are interested in designing an algorithm that is provably convergent to a properly defined stationary solution, care is needed to handle the combined features of non-convexity and non-differentiability in the approximated problem (4.8) and the discontinuity in (4.7). These features are particularly relevant when we consider the convergence of the former to the latter as $\delta \downarrow 0$. To illustrate the difficulty with some algorithms for solving (4.8), we mention that a majorization-minimization type algorithm (Lange, 2016) may be too complex to implement as a majorizing function may be quite complicated; block coordinate descent type methods may not converge to a stationary point of this problem because the needed regularity assumptions (Tseng, 2001) cannot be expected to be satisfied. Therefore, an alternative way to tackle the discontinuity of the indicator function is needed, which is the focus of Subsection **??**.

**(b) The positive scale-invariance of the indicator function.** The function $\mathbb{I}(A\, f(X;\theta) > 0)$ is positively scale-invariant as any positive scaling of $f(X;\theta)$ will not change the objective value of the problem (4.7). This could cause computational instability, and more seriously, incorrect definition of the indicator function due to round-off errors; these numerical issues become more pronounced when $f(X;\theta)$ is close to 0 in practical implementation of an algorithm. One way to guard against such undesirable characteristics of the indicator function is to solve two optimization problems with the bias term $\beta_0$ set equal to $\pm 1$, respectively, and accept as the solution the one with a smaller objective value. This approach works if the true $\beta_0$ is not equal to 0. In the development below, this safe guard is adopted as can be seen in the formulation (4.10).

In this subsection, we propose a method to transform the discontinuous optimization problem (4.7) that involves the indicator function to a continuous optimization problem by means of a mild assumption. Our approach is to reformulate the discontinuous problem (4.7) via its epigraphical representation. Since $\mathbb{I}(\,\bullet\, > 0)$ is a lower semicontinuous function, its epigraph

$$\operatorname{epi}\mathbb{I}(\,\bullet\, > 0) \triangleq \{\, (t,s) \in \mathbb{R} \times \mathbb{R} \mid t \geq \mathbb{I}(s > 0)\,\}$$

is a closed set (Rockafellar, 1970, Theorem 7.1). However, the random variable $\mathcal{Z}$ may attain positive values, which makes it also essential to consider the hypograph of $\mathbb{I}(\bullet > 0)$, i.e., the set

$$\operatorname{hypo}\mathbb{I}(\,\bullet\, > 0) \triangleq \{\, (t,s) \in \mathbb{R} \times \mathbb{R} \mid t \leq \mathbb{I}(s > 0)\,\}\,.$$

Since the indicator function is not upper semicontinuous, the above set is not closed. We thus consider an approximation of $\mathbb{I}(\bullet > 0)$ by an upper semicontinuous function $\mathbb{I}(\bullet \geq 0)$ that has a closed hypograph

$$\operatorname{hypo}\mathbb{I}(\,\bullet\, \geq 0) \triangleq \{\, (t,s) \in \mathbb{R} \times \mathbb{R} \mid t \leq \mathbb{I}(s \geq 0)\,\}\,.$$

Interestingly, the sets $\operatorname{epi}\mathbb{I}(\,\bullet\, > 0)$ and $\operatorname{hypo}\mathbb{I}(\,\bullet\, \geq 0)$ are each a finite union of polyhedra that admits an extremely simple dc representation given in the next lemma. See also Figures 4.1 and 4.2 for illustration. No proof is required for the lemma.

**Lemma 4.3.1.** *For any $t, s \in \mathbb{R}$, the following two statements hold:*

*(i)* $(t,s) \in \operatorname{epi} \mathbb{I}(\bullet > 0)$ *if and only if* $\max(-t,s) - \max(t+s-1,0) \leq 0$ *;*

*(ii)* $(t,s) \in \operatorname{hypo} \mathbb{I}(\bullet \geq 0)$ *if and only if* $\max(t+s-1,0) - \max(-t,s) \leq 0$. □

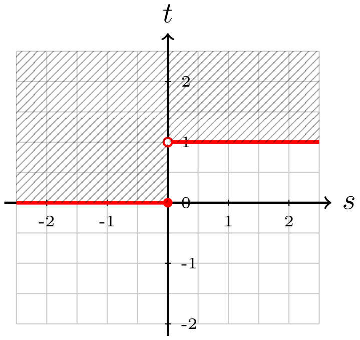


**Figure 4.1:** the region (shaded) for $\operatorname{epi} \mathbb{I}(\bullet > 0)$

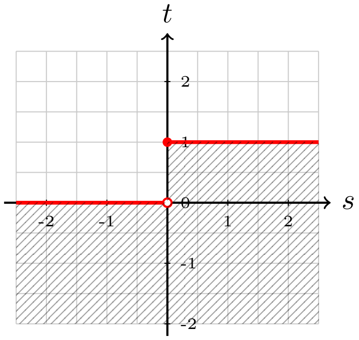


**Figure 4.2:** the region (shaded) for $\operatorname{hypo} \mathbb{I}(\bullet \geq 0)$

Denoting $\mathcal{Z}^- \triangleq \max(-\mathcal{Z}, 0)$ and $\mathcal{Z}^+ \triangleq \max(\mathcal{Z}, 0)$, we assume that

$$\mathbb{E}\left[\mathcal{Z}^+ \frac{\mathbb{I}(A\, f(X;\theta) = 0)}{\pi(A|X)}\right] = 0.$$

Under this assumption, problem (4.7) is equivalent to

$$\underset{\beta \in \mathbb{R}^p, \alpha \in \mathcal{F}}{\text{minimize}} \left\{ \begin{array}{l} \mathbb{E}\left[\mathcal{Z}^- \dfrac{\mathbb{I}(A\, f(X;\theta) > 0)}{\pi(A|X)}\right] - \mathbb{E}\left[\mathcal{Z}^+ \dfrac{\mathbb{I}(A\, f(X;\theta) \geq 0)}{\pi(A|X)}\right] \\ + \mathbb{E}\left[\left(\underbrace{\mathcal{Z} - \alpha(X) - u(\mathcal{Z} - \alpha(X))}_{\text{nonnegative}}\right) \dfrac{\mathbb{I}(A\, f(X;\theta) > 0)}{\pi(A|X)}\right] \end{array} \right\}. \tag{4.9}$$

For further consideration, we take $\alpha(X)$ to be a parameterized family of affine functions $\{b^T X + \beta_0 = w^T \widehat{X}\}$ where $w \triangleq \begin{pmatrix} b \\ b_0 \end{pmatrix}$ is the parameter is be estimated. The use of affine functions to approximate $\alpha^*(X)$ is based on both modeling and computational perspectives. The affine functions are easy for interpretation, but may suffer from model misspecification. The linear assumption can be relaxed by using kernel trick in machine learning. The corresponding computation will be more involved. We approximate the expectation in (4.9) by the sample average that is based on the available data $\{(X^i, A_i, \mathcal{Z}_i)\}_{i=1}^N$. In order to compute a sparse solution that can avoid model overfitting, we add sparsity surrogate functions (Ahn et al., 2017) $P_b$ and $P_\theta$ on the parameters $w$ and $\beta$ in the covariate function $\alpha(X)$ and the function $f(X;\theta)$, respectively, each weighted by the positive scalars $\lambda_b^N$ and $\lambda_\beta^N$. The empirical problem is then given by

$$\underset{\substack{\beta\in\mathbb{R}^p\\ w\triangleq(b,b_0)\in S}}{\text{minimize}}\left\{\begin{array}{l} \lambda_a^N\,P_b(b)+\lambda_\beta^N\,P_\beta(\beta)+\dfrac{1}{N}\displaystyle\sum_{i=1}^{N}\mathcal{Z}_i^-\,\dfrac{\mathrm{I\!I}(A_i\,(\beta^TX^i\pm 1)>0)}{\pi(A_i\,|\,X^i)}- \\ \dfrac{1}{|\mathcal{N}_+|}\displaystyle\sum_{i\in\mathcal{N}_+}\mathcal{Z}_i^+\,\dfrac{\mathrm{I\!I}(A_i\,(\beta^TX^i\pm 1)\geq 0)}{\pi(A_i\,|\,X^i)}+ \\ \dfrac{1}{N}\displaystyle\sum_{i=1}^{N}\left[\,\mathcal{Z}_i-w^T\widehat{X}^i-u(\mathcal{Z}_i-w^T\widehat{X}^i)\right]\,\dfrac{\mathrm{I\!I}(A_i\,(\beta^TX^i\pm 1)>0)}{\pi(A_i\,|\,X^i)} \end{array}\right\}, \tag{4.10}$$

where $\mathcal{N}_+\triangleq\{\,1\leq j\leq N\mid\mathcal{Z}_j>0\,\}$ and $S$ is a closed convex set. [In principle, we may add constraints to the parameter $\beta$ also but refrain from doing this as it does not add value to the methodology.] Based on Lemma 4.3.1, the above problem can be further written as

minimize over $z\triangleq(w,\beta,\sigma^\pm)$; $\beta\in\mathbb{R}^p$, and $w\triangleq(b,b_0)\in S$

$$\varphi(z)\triangleq\left\{\begin{array}{l} \lambda_a^N\,P_b(b)+\lambda_\beta^N\,P_\beta(\beta)+\dfrac{1}{N}\displaystyle\sum_{i=1}^{N}\dfrac{\mathcal{Z}_i^-\sigma_i^-}{\pi(A_i\,|\,X^i)}-\dfrac{1}{|\mathcal{N}_+|}\displaystyle\sum_{j\in\mathcal{N}_+}\dfrac{\mathcal{Z}_j^+\sigma_j^+}{\pi(A_j\,|\,X^j)} \\ \dfrac{1}{N}\displaystyle\sum_{i=1}^{N}\underbrace{\left[\,\mathcal{Z}_i-w^T\widehat{X}^i-u(\mathcal{Z}_i-w^T\widehat{X}^i)\right]\,\dfrac{\sigma_i^-}{\pi(A_i\,|\,X^i)}}_{\text{nonconvex}} \end{array}\right\} \tag{4.11}$$

subject to

$$\max(-\sigma_i^-,\,A_i\,(\beta^TX^i\pm 1))-\max(\sigma_i^-+\,A_i(\beta^TX^i\pm 1)-1,\,0)\,\leq\,0,\;1\leq i\leq N$$

$$\max(\sigma_j^++A_j\,(\beta^TX^j\pm 1)-1)-\max(-\sigma_j^+,\,A_j\,(\beta^TX^j\pm 1))\leq 0,\quad j\in\mathcal{N}_+,$$

where the constraints are of the difference-of-convex, piecewise affine type. Denote $t_i\triangleq\mathcal{Z}_i-w^T\widehat{X}^i$ for any $i=1,\cdots,N$. The last term in the objective function $\varphi$ can be further written as

$$\begin{aligned} &[\,t_i-u(t_i)\,]\,\frac{\sigma_i^-}{\pi(A_i\,|\,X^i)} \\ =\;&\frac{1}{2\,\pi(A_i\,|\,X^i)}\left\{\,\left[\,t_i-u(t_i)+\sigma_i^-\,\right]^2-(\sigma_i^-)^2-[\,t_i-u(t_i)\,]^2\,\right\}. \end{aligned}$$

Since $t_i-u(t_i)\geq 0$ and $\sigma_i^-\geq 0$, the terms $\left[\,t_i-u(t_i)+\sigma_i^-\,\right]^2$ and $[\,t_i-u(t_i)\,]^2$ are convex. Hence each product $[\,t_i-u(t_i)\,]\,\dfrac{\sigma_i^-}{\pi(A_i\,|\,X^i)}$ is the difference of convex functions.

Suppose that the utility function and sparsity surrogate functions are as follows:

$$\begin{aligned} u(t) &= \xi_1 \max(0,t) - \xi_2 \max(0,-t), \quad \text{where } 0 \le \xi_1 < 1 < \xi_2\,; \\ P_b(b) &= \sum_{i=1}^{p} \Big[ \phi_i^b |\, b_i\,| - \rho_i^b(b_i) \Big], \quad \phi_i^b > 0,\ i = 1,\cdots,p\,; \\ P_\beta(\beta) &= \sum_{i=1}^{p} \Big[ \phi_i^\beta |\, \beta_i\,| - \rho_i^\beta(\beta_i) \Big], \quad \phi_i^\beta > 0,\ i = 1,\cdots,p, \end{aligned} \tag{4.12}$$

where $\phi_i^b$ and $\phi_i^\beta$ are given constants and $\rho_i^b$ and $\rho_i^\beta$ are convex differentiable functions (Ahn et al., 2017). We then have

$$\begin{aligned} [\,t_i - u(t_i)\,]\,\sigma_i^- = \frac{1}{2}\Bigg\{ & \underbrace{(1-\xi_1)\big[\max(0,t_i)+\sigma_i^-\,\big]^2 + (1+\xi_2)\big[\max(0,-t_i)+\sigma_i^-\,\big]^2}_{\text{convex}} \\ & - \underbrace{\Big[(2-\xi_1+\xi_2)(\sigma_i^-)^2 - (1-\xi_1)\,[\max(0,t_i)\,]^2 - (1+\xi_2)\,[\max(0,-t_i)\,]^2\Big]}_{\text{convex and continuously differentiable}} \Bigg\}. \end{aligned}$$

Therefore, under the above setting, the objective function $\varphi$ is the difference of two convex functions, $\varphi_1 - \varphi_2$, with $\varphi_2$ being continuously differentiable. In the next section, we present a dc algorithm for solving such a problem.

## 4.4 Solving a Piecewise Affine Constrained DC Program

We consider problem (4.11) cast in the following general form:

$$\begin{aligned} & \underset{x\in X}{\text{minimize}} \quad f(x) - g(x) \\ & \text{subject to} \\ & \quad \max_{1\le j\le J_{1i}} ((a^{ij})^T x + \alpha_{ij}) - \max_{1\le j\le J_{2i}} ((b^{ij})^T x + \beta_{ij}) \le 0, \quad i = 1,\ldots,m, \end{aligned} \tag{4.13}$$

where $f : \mathbb{R}^n \to \mathbb{R}$ is a convex function, $g : \mathbb{R}^n \to \mathbb{R}$ is a continuously differentiable convex function with Lipschitz continuous gradient, each $a^{ij}$ and $b^{ij}$ are $n$-dimensional vectors, each $\alpha_{ij}$ and $\beta_{ij}$ are scalars, each $J_{1i}$ and $J_{2i}$ are positive integers, and $X$ is a polyhedral set. Notice that for any $i = 1, \ldots, m$, it holds that

$$\begin{aligned} & \max_{1 \leq j \leq J_{1i}} ((a^{ij})^T x + \alpha_{ij}) - \max_{1 \leq j \leq J_{2i}} ((b^{ij})^T x + \beta_{ij}) \leq 0 \\ \Longleftrightarrow \quad & (a^{ij_1})^T x + \alpha_{ij_1} - \max_{1 \leq j \leq J_{2i}} ((b^{ij})^T x + \beta_{ij}) \leq 0, \quad \forall\, 1 \leq j_1 \leq J_{1i} \\ \Longleftrightarrow \quad & \max_{1 \leq j_2 \leq J_{2i}} \big( (b^{ij_2} - a^{ij_1})^T x + (\beta_{ij_2} - \alpha_{ij_1}) \big) \geq 0, \quad \forall\, 1 \leq j_1 \leq J_{1i}. \end{aligned}$$

The above equivalences indicate that by properly redefining $(b^{ij}, \beta_{ij})$ and the value of $m$, one can write any piecewise linear constrained dc program (4.13) as the following reverse convex constrained (Hillestad and Jacobsen, 1980) dc program:

$$\begin{aligned} & \underset{x \in X}{\text{minimize}} \quad h(x) \triangleq f(x) - g(x) \\ & \text{subject to} \quad \max_{1 \leq j \leq J_i} (\, (b^{ij})^T x + \beta_{ij}\,) \geq 0, \quad i = 1, \ldots, m. \end{aligned} \tag{4.14}$$

Denote the feasible set of the problem (4.14) as

$$F \triangleq \left\{ x \in X \mid \max_{1 \leq j \leq J_i} (\, (b^{ij})^T x + \beta_{ij}\,) \geq 0, \quad i = 1, \ldots, m \right\}.$$

For any $x \in \mathbb{R}^n$, we also denote

$$\mathcal{I}(x) \triangleq \left\{ 1 \leq i \leq m \mid \max_{1 \leq j \leq J_i} (\, (b^{ij})^T x + \beta_{ij}\,) = 0 \right\}$$

and

$$\mathcal{A}_i(x) \triangleq \underset{1 \leq j \leq J_i}{\text{argmax}} \left\{ (b^{ij})^T x + \beta_{ij} \right\}, \quad i = 1, \ldots, m.$$

We say that $\bar{x} \in X$ is a B(ouligand)-stationary point (Pang, 2007) of the problem (4.14) if

$$h'(\bar{x}; d) \triangleq \lim_{\tau \downarrow 0} \frac{h(\bar{x} + \tau d) - h(\bar{x})}{\tau} = f'(\bar{x}; d) - g'(\bar{x}; d) \geq 0, \quad \forall\, d \in \mathcal{T}_B(\bar{x}; F),$$

where $\mathcal{T}_B(\bar{x}; F)$ is the Bouligand tangent cone of $F$ at $\bar{x} \in F$, i.e., (see, e.g., (Pang et al., 2016, Proposition 3)),

$$\begin{aligned}
\mathcal{T}_B(\bar{x}; F) &\triangleq \left\{ d \in \mathbb{R}^n \mid d = \lim_{\nu \to \infty} \frac{(x^\nu - \bar{x})}{\tau_\nu}, \text{ where } F \ni x^\nu \to \bar{x} \text{ and } \tau_\nu \downarrow 0 \right\} \\
&= \left\{ d \in \mathcal{T}_B(\bar{x}; X) \mid \max_{j \in \mathcal{A}_i(\bar{x})} (b^{ij})^T d \geq 0, \ \forall\, i \in \mathcal{I}(\bar{x}) \right\} \\
&= \bigcap_{i \in \mathcal{I}(\bar{x})} \bigcup_{j \in \mathcal{A}_i(\bar{x})} \left\{ d \in \mathcal{T}_B(\bar{x}; X) \mid (b^{ij})^T d \geq 0 \right\}.
\end{aligned}$$

[Since $X$ is assumed to be polyhedral, $\mathcal{T}_B(\bar{x}; X)$ is a polyhedral cone.] A weaker concept than B-stationarity is that of weak B-stationarity, which pertains to a feasible solution $\bar{x} \in F$ such that $h'(\bar{x}; d) \geq 0$ for any $d \in \mathbb{R}^n$ satisfying

$$\begin{aligned}
d \in \mathcal{T}_B^{\text{weak}}(\bar{x}; F) &\triangleq \left\{ d \in \mathcal{T}_B(\bar{x}; X) \mid \min_{j \in \mathcal{A}_i(\bar{x})} (b^{ij})^T d \geq 0, \ \forall\, i \in \mathcal{I}(\bar{x}) \right\} \\
&= \bigcap_{i \in \mathcal{I}(\bar{x})} \bigcap_{j \in \mathcal{A}_i(\bar{x})} \left\{ d \in \mathcal{T}_B(\bar{x}; X) \mid (b^{ij})^T d \geq 0 \right\}.
\end{aligned}$$

Unlike $\mathcal{T}_B(\bar{x}; F)$, which is not necessarily convex, $\mathcal{T}_B^{\text{weak}}(\bar{x}; F)$ is a polyhedral cone. It is known from (Clarke, 1998, Chapter 2, Proposition 1.1(c) & Exercise 9.10) that

$$\mathcal{T}_C(\bar{x}; F) \subseteq \mathcal{T}_B^{\text{weak}}(\bar{x}; F) \subseteq \mathcal{T}_B(\bar{x}; F),$$

where $\mathcal{T}_C(\bar{x}; F)$ denotes the Clarke tangent cone of $F \subseteq \mathbb{R}^n$ at $\bar{x}$, i.e., $d \in \mathcal{T}_C(\bar{x}; F)$ if for every sequence $\{x^i\} \subseteq S$ converging to $\bar{x}$ and positive scalar sequence $\{t_i\}$ decreasing to 0, there exists a sequence $\{d^i\} \subseteq \mathbb{R}^n$ converging to $d$ such that $x^i + t_i d^i \in F$ for all $i$ (Clarke, 1998, Chapter 2, Proposition 5.2).

In order to better understand the above two stationarity concepts in the context of the piecewise polyhedral structure of the feasible set $F$ and to motivate the algorithm to be presented afterward for solving the problem (4.14), we first introduce a further stationarity concept, which we call A-stationarity (A for Algorithm). Specifically, we note that $F$ is the union of finitely many polyhedra:

$$F = \bigcup_{(j_1,\cdots,j_m)} \left\{ x \in X \mid (b^{ij_i})^T x + \beta_{ij_i} \geq 0, \quad i = 1,\ldots,m \right\},$$

where the union ranges over all tuples $\{j_i\}_{i=1}^m$ with each $j_i \in \{1,\cdots,J_i\}$ for all $i$. Given a vector $\bar{x} \in F$, let $\mathcal{J}(\bar{x})$ be the family of such tuples such that $j_i \in \mathcal{A}_i(\bar{x})$ for all $i = 1,\cdots,m$. We say that $\bar{x} \in F$ is *A-stationary* if there exists a tuple $\bar{j}(\bar{x}) = \{\bar{j}_i\}_{i=1}^m \in \mathcal{J}(\bar{x})$ such that

$$h'(\bar{x};d) \geq 0, \ \forall\, d \in \mathcal{T}_A^{\bar{j}(\bar{x})}(\bar{x};F) \triangleq \left\{ d \in \mathcal{T}_B(\bar{x};X) \mid (b^{i\bar{j}_i})^T d \geq 0, \ \forall\, i \in \mathcal{I}(\bar{x}) \right\}.$$

**Lemma 4.4.1.** *Let $\bar{x} \in F$ be given. Consider the following statements all pertaining to the problem* (4.14)*:*

*(a) $\bar{x}$ is B-stationary;*

*(b) $\bar{x}$ is A-stationary;*

*(c) there exists a tuple $\bar{j}(\bar{x}) = \{\bar{j}_i\}_{i=1}^m \in \mathcal{J}(\bar{x})$ such that*

$$\bar{x} \in \underset{x\in X}{\mathit{argmin}} \left\{ f(x) - [\, g(\bar{x}) + \nabla g(\bar{x})^T (x - \bar{x}) \,] \mid (b^{i\bar{j}_i})^T x + \beta_{i\bar{j}_i} \geq 0, \ i \in \mathcal{I}(\bar{x}) \right\}; \qquad (4.15)$$

*(d) there exists a tuple $\bar{j}(\bar{x}) = \{\bar{j}_i\}_{i=1}^m \in \mathcal{J}(\bar{x})$ such that*

$$\bar{x} \in \underset{x\in X}{\mathit{argmin}} \left\{ f(x) - [\, g(\bar{x}) + \nabla g(\bar{x})^T (x - \bar{x}) \,] \mid (b^{i\bar{j}_i})^T x + \beta_{i\bar{j}_i} \geq 0, \ i = 1,\cdots,m \right\};$$

*(e) $\bar{x}$ is weak B-stationary.*

*It holds that (a) ⇒ (b) ⇔ (c) ⇔ (d) ⇒ (e).*

*Proof.* (a) ⇒ (b). This is because $\mathcal{T}_A^{\bar{j}(\bar{x})}(\bar{x};F) \subseteq \mathcal{T}_B(\bar{x};F)$.

(b) ⇒ (e). This is because $\mathcal{T}_B^{\text{weak}}(\bar{x};F) \subseteq \mathcal{T}_A^{\bar{j}(\bar{x})}(\bar{x};F)$.

(b) ⇔ (c). This is clear because the condition $h'(\bar{x};d) \geq 0$ for all $d \in \mathcal{T}_A^{\bar{j}(\bar{x})}(\bar{x};F)$ is exactly the first-order optimality condition of the convex program in (4.15).

(c) $\Rightarrow$ (d). This is clear because there are more constraints in the feasible region of the optimization problem in (d) than those in (c).

(d) $\Rightarrow$ (c). Let $x \in X$ satisfy $(b^{i\bar{j}_i})^T x + \beta_{i\bar{j}_i} \geq 0$ for all $i \in \mathcal{I}(\bar{x})$. Since $(b^{i\bar{j}_i})^T x + \beta_{i\bar{j}_i} > 0$ for all $i \notin \mathcal{I}(\bar{x})$, it follows that for all $\tau > 0$ sufficiently small, the vector $x^\tau \triangleq x + \tau(\bar{x} - x)$ satisfies $(b^{i\bar{j}_i})^T x^\tau + \beta_{i\bar{j}_i} \geq 0$ for all $i = 1, \cdots, m$. Hence,

$$\begin{aligned} f(\bar{x}) - g(\bar{x}) &\leq f(x^\tau) - \left[ g(\bar{x}) + \nabla g(\bar{x})^T (x^\tau - \bar{x}) \right] \quad \text{by (d)} \\ &\leq \tau \left[ f(\bar{x}) - g(\bar{x}) \right] + (1 - \tau) \left[ f(x) - \left[ g(\bar{x}) + \nabla g(\bar{x})^T (x - \bar{x}) \right] \right], \end{aligned}$$

which yields

$$f(\bar{x}) - g(\bar{x}) \leq f(x) - \left[ g(\bar{x}) + \nabla g(\bar{x})^T (x - \bar{x}) \right],$$

establishing (c). □

In the following, we propose a dc algorithm to compute an A-stationary point of (4.14). The algorithm takes advantage of the reverse convex constraints of the problem in that once initiated at a feasible vector $x^0 \in F$, the algorithm generates a feasible sequence $\{x^\nu\} \subset F$; see Step 1 below.

---

A dc algorithm for solving the reverse convex constrained dc program (4.14).

---

**Initialization.** Given are a scalar $c > 0$, an initial point $x^0 \in F$.

**Step 1.** For each $i = 1, \cdots, m$, choose an index $j_i^\nu \in \mathcal{A}_i(x^\nu)$. Let $x^{\nu+1}$ be the unique optimal solution of the convex program:

$$\begin{aligned} \underset{x \in X}{\text{minimize}} \quad & \widehat{h}_c(x; x^\nu) \triangleq f(x) - \left[ g(x^\nu) + (\nabla g(x^\nu))^T (x - x^\nu) \right] \\ & \qquad + \quad \underbrace{\frac{c}{2} \| x - x^\nu \|^2}_{\text{proximal regularization}} \qquad (4.16) \\ \text{subject to} \quad & (b^{ij_i^\nu})^T x + \beta_{ij_i^\nu} \geq 0, \quad i = 1, \ldots, m. \end{aligned}$$

**Step 2.** If $x^{\nu+1}$ satisfies a prescribed stopping rule, terminate; otherwise, return to Step 1 with $\nu$ replaced by $\nu + 1$. □

An enhanced version of the above algorithm that requires solving multiple subproblems for all indices $j_i^{\nu}$ in a so-called "$\varepsilon$-argmax set" has been suggested in (Pang et al., 2016). For this enhanced algorithm, it can be shown that every accumulation point, if exists, of the generated sequence is a B-stationary point. Although there are theoretical benefits of such an algorithm, it may not be efficient when applied to the empirical CDE problem (4.11), because the number of reverse convex inequalities in the constraint set is proportional to the number of samples, making the "$\varepsilon$-argmax set" potentially very large, thus potentially many subprograms need to be solved at every iteration. There is also a probabilistic variant of the enhanced algorithm that also solves only one convex subprogram of the same type as (4.16). The only difference from the presented deterministic algorithm is that the tuple $\{\bar{j}_i^{\nu}\}_{i=1}^{m}$ is chosen from the $\varepsilon$-argmax sets randomly with positive probabilities. Almost sure convergence of the probabilistic algorithm to a B-stationary point can be established. Since the above (deterministic) algorithm has not been formally introduced in the literature, we provide below a (subsequential) convergence result to an A-stationary solution of the problem (4.14).

It is worth mentioning that each $x^{\nu+1}$ is feasible to the subprogram (4.16) at iteration $\nu+1$ because

$$(b^{i j_i^{\nu+1}})^T x^{\nu+1} + \beta_{i j_i^{\nu+1}} = \max_{1 \leq j \leq J_i} \left( (b^{ij})^T x^{\nu+1} + \beta_{ij} \right) \geq (b^{i j_i^{\nu}})^T x^{\nu+1} + \beta_{i j_i^{\nu}} \geq 0.$$

This inequality also shows that $x^{\nu+1} \in F$ for all $\nu$. The following theorem asserts the subsequential convergence of the sequence generated by the above dc algorithm to an A-stationary point of problem (4.14).

**Theorem 4.4.1.** *Suppose that $h$ is bounded below on the polyhedral set $X$. Then any accumulation point $x^{\infty}$ of the sequence $\{x^{\nu}\}$ generated by the dc algorithm, if it exists, is an A-stationary point of* (4.14).

*Proof.* The sequence of function values $\{h(x^{\nu})\}$ decreases since

$$\begin{aligned} & h(x^{\nu+1}) + \frac{c}{2}\|x^{\nu+1} - x^{\nu}\|^2 \\ \leq \; & \widehat{h}_c(x^{\nu+1}; x^{\nu}) \quad \text{(by the convexity of } g) \\ \leq \; & h(x^{\nu}) \text{ (by the optimality of } x^{\nu+1} \text{ and the feasibility of } x^{\nu} \text{ to (4.16)).} \end{aligned}$$

Since $h$ is bounded below on $X$, we may derive that $\lim\limits_{\nu\to\infty} \|x^{\nu+1} - x^{\nu}\| = 0$. By the definition of the point $x^{\nu+1}$, we obtain that for all $x \in X$ satisfying $(b^{i j_i^{\nu}})^T x + \beta_{i j_i^{\nu}} \geq 0$, $i = 1, \ldots, m$,

$$\begin{aligned} & f(x^{\nu+1}) - \left[\, g(x^{\nu}) + \nabla g(x^{\nu})^T (x^{\nu+1} - x^{\nu}) \,\right] + \frac{c}{2}\, \|x^{\nu+1} - x^{\nu}\|^2 \\ & \leq \; f(x) - \left[\, g(x^{\nu}) + \nabla g(x^{\nu})^T (x - x^{\nu}) \,\right] + \frac{c}{2}\, \|x - x^{\nu}\|^2. \end{aligned} \tag{4.17}$$

Let $\{x^{\nu+1}\}_{\nu\in\kappa}$ be a subsequence of $\{x^{\nu}\}$ that converges to $x^{\infty}$. Then $x^{\infty} \in F$. Since each $\mathcal{A}_i(x^{\nu})$ is finite, we may assume without loss of generality that the selected $j_i^{\nu} \in \mathcal{A}_i(x^{\nu})$ are independent of $\nu$ for any $i = 1, \ldots, m$ on this subsequence, i.e., there exists $\bar{j}_i$ such that $\bar{j}_i = j_i^{\nu}$ for all $i = 1, \ldots, m$ and all $\nu \in \kappa$. For all $x \in X$ satisfying $(b^{i\bar{j}_i})^T x + \beta_{i\bar{j}_i} \geq 0$, the inequality (4.17) holds. Taking limit of $\nu (\in \kappa) \to +\infty$, we obtain that $\bar{j}_i \in \mathcal{A}_i(x^{\infty})$ for $i = 1, \ldots, m$, and for all $x \in X$ satisfying $(b^{i\bar{j}_i})^T x + \beta_{i\bar{j}_i} \geq 0$,

$$f(x^{\infty}) - g(x^{\infty}) \;\leq\; f(x) - [\, g(x^{\infty}) + \nabla g(x^{\infty})^T (x - x^{\infty})\,],$$

which, by Lemma 4.4.1, yields that $x^{\infty}$ is an A-stationary point of the problem (4.14). □

Given $\bar{z} \triangleq (\bar{w}, \bar{\beta}, \bar{\sigma}^{\pm})$ and a positive constant $c > 0$, the strongly convex objective of the subproblem of the dc algorithm in Step 1 for solving the problem (4.11) with $u$, $P_a$ and $P_b$ given in (4.12) can be essentially written as

$$\begin{aligned}
&\lambda_a^N \sum_{i=1}^{p} \left[ \phi_i^a |\, a_{\,i}\,| - \frac{d\rho_i^a(\bar{a}_i)}{da_i}\,(\,a_i - \bar{a}_i\,) \right] + \lambda_\beta^N \sum_{i=1}^{p} \left[ \phi_i^\beta |\, \beta_{\,i}\,| - \frac{d\rho_i^\beta(\bar{\beta}_i)}{d\beta_i}\,(\,\beta_i - \bar{\beta}_i\,) \right] + \\
&\frac{1}{N}\sum_{i=1}^{N} \frac{\mathcal{Z}_i^- \sigma_i^-}{\pi(A_{\,i}\,|\,X^{\,i})} - \frac{1}{|\mathcal{N}_+|}\sum_{i\in\mathcal{N}_+} \frac{\mathcal{Z}_i^+ \sigma_i^+}{\pi(A_{\,i}\,|\,X^{\,i})} + \frac{c}{2}\,||z-\bar{z}||^2 + \\
&\frac{1}{2\,\pi(A_{\,i}\,|\,X^{\,i})}\,\Big\{ (1-\xi_1)\left[\,\max(0,t_i) + \sigma_i^-\,\right]^2 + (1+\xi_2)\left[\,\max(0,-t_i) + \sigma_i^-\,\right]^2 - \\
&2(2-\xi_1+\xi_2)\bar{\sigma}_i^-(\sigma_i^- - \bar{\sigma}_i^-) - 2\,[\,(1-\xi_1)\max(0,\bar{t}_i) - (1+\xi_2)\max(0,-\bar{t}_i)\,]\,(t_i - \bar{t}_i) \Big\},
\end{aligned}$$

where $z \triangleq (w, \beta, \sigma^{\pm})$ with $\beta \in \mathbb{R}^p$, $w \triangleq (a, b) \in S$, $\sigma^- \in \mathbb{R}^N$ and $\sigma^+ \in \mathbb{R}^{|\mathcal{N}_+|}$. The above objective function involves the convex, non-differentiable terms $|a_i|$, $|\beta_i|$, $\left[\,\max(0,t_i) + \sigma_i^-\,\right]^2$, and $\left[\,\max(0,-t_i) + \sigma_i^-\,\right]^2$; the latter two squared terms also make the objective non-separable in the $w$ and $\sigma^-$ variables. All these features make the linear inequality constrained subproblem seemingly complicated. One way to solve this subproblem is via the dual semismooth Newton approach, as discussed in a recent paper (Cui et al., 2018). In fact, by introducing auxiliary variables

$$\begin{cases}
t_i^+ \;=\; \max(t_i, 0), & t_i^- \;=\; \max(-t_i, 0), \\[1ex]
a_i^+ \;=\; \max(a_i, 0), & a_i^- \;=\; \max(-a_i, 0), \\[1ex]
b_i^+ \;=\; \max(b_i, 0), & b_i^- \;=\; \max(-b_i, 0),
\end{cases}$$

we may write

$$\mathcal{Z}_i - w^T \widehat{X}^{\,i} \;=\; t_i = t_i^+ - t_i^-, \quad |a_i| = a_i^+ + a_i^-, \quad |\beta_i| = \beta_i^+ + \beta_i^-.$$

Therefore, an alternative approach for solving (4.11) is to transform it into a standard quadratic programming problem with the additional variables $(t_i^+, t_i^-, a_i^+, a_i^-, b_i^+, b_i^-)$ such that it can be solved by many efficient quadratic programming solvers.

In terms of statistical consistency, as long as the tuning parameters $\lambda_a^N$ and $\lambda_\beta^N$ go to 0 when $N$ goes to infinite, the minimizer of the empirical objective function (4.10) might converge to the minimizer of the corresponding population problem under some regularity conditions ((Van der Vaart, 2000)). If we allow rates of tunning parameters going to 0 faster than $\frac{1}{\sqrt{n}}$, then the convergence rate of empirical minimizers may be $\frac{1}{\sqrt{n}}$ under some regularity conditions.

Similar ideas could be borrowed from (Knight et al., 2000), although their considered settings are different from ours. The convergence results in our settings are more complicated than those standard cases since the empirical loss function here is non-convex and non-smooth.

### 4.5 Numerical Experiments

In this section, we demonstrate the effectiveness of the proposed IDR-CDE in finding optimal IDRs via three synthetic examples. The subproblem of the dc algorithm, being equivalent to a quadratic programming problem, is solved by the commercial solver Gurobi with an academic license. All the numerical results are run in Matlab on Mac OS X with 2.5 GHz Intel Core i7 and 16 GB RAM. We use piecewise linear affine function given by (4.5) with $\xi_1 = 0, \xi_2 = 0.5$ in all the experiments, which is equivalent to estimating the optimal IDR that maximizes $\text{CVaR}_{0.5}(\mathcal{Z})$. In practice, users can decide their own utility functions and values $\xi_1$, $\xi_2$ based on the specific problem settings. If one believes there may have high risks for inappropriate decisions and wants to control the risk of higher-risk individuals, it would be better to use robust utility functions such as the piecewise affine utility function. We consider a binary-action space in a randomized study with $\pi(A_i = \pm 1 \,|\, X_i) = 0.5$. All the tuning parameters such as $\lambda_b^N$ and $\lambda_\beta^N$ are selected via 10-fold-cross-validation that maximizes the following average of the empirical $\mathcal{O}^d_{(u,\mathcal{F})}(\mathcal{Z})$, which is defined as

$$\widehat{\mathcal{O}}^{\widehat{d}}_{(u,\mathcal{F})}(\mathcal{Z}) \triangleq \frac{\displaystyle\sum_{i \in \mathcal{N}} \left[\, \widehat{\alpha}(X_i) + u(\mathcal{Z}_i - \widehat{\alpha}(X_i)) \,\right] \frac{\mathbb{I}(A_i = \widehat{d}(X_i))}{\pi(A_i|X_i)}}{\displaystyle\sum_{i \in \mathcal{N}} \frac{\mathbb{I}(A_i = \widehat{d}(X_i))}{\pi(A_i|X_i)}}.$$

Specifically, we divide the training data into 10 groups. For each fold, we estimate the optimal IDR $\widehat{d}(X)$ using 9 groups of the data (the training set) for a pre-specified series of tuning parameters $\lambda_b^N$ and $\lambda_\beta^N$ and then compute $\widehat{\mathcal{O}}^d_{(u,\mathcal{F})}(\mathcal{Z})$ on the remaining group of data (the test set). The best tuning parameters are the ones that lead to the largest values of $\widehat{\mathcal{O}}^{\widehat{d}}_{(u,\mathcal{F})}(\mathcal{Z})$. The so-obtained parameters are then employed to re-compute the optimal IDR using the entire set of data.

We compare our approach with three existing methods under the expected-value function framework $\mathbb{E}^d[\mathcal{Z}]$. The first one is a model-based method called $l_1$-PLS (Qian and Murphy, 2011) that first fits a penalized least-square regression with covariate function $(1, X, A, X \circ A)$ on $\mathcal{Z}$ to estimate $\mathbb{E}[\mathcal{Z}|X, A = a]$, and then select the action with the largest $\mathbb{E}\,[\,\mathcal{Z}\,|\,X, A = a\,]$, where $X \circ A$ denotes the element-wise product. The second one is a classification-based method called residual weighted learning (RWL) (Zhou et al., 2017) that consists of two steps: (1) fitting a least-square regression on $\mathcal{Z}_i$ with covariates $\widehat{X}_i$ to compute the residual $r_i$ for each data point in order to remove the main effect; (2) applying the support vector machine with truncated loss to compute the optimal IDR with each data point weighted by $r_i$. The third one is the direct learning (DLearn) method (Qi and Liu, 2018) that lies between the model-based and the classification-based method, where the optimal IDR is directly found by weighted penalized least square regression on $\mathcal{Z}A$ with covariates $\widehat{X}$, based on the fact that

$$\mathbb{E}\,[\,\mathcal{Z}\,|\,X, A = 1\,] - \mathbb{E}\,[\,\mathcal{Z}\,|\,X, A = -1\,] = \mathbb{E}\left[\frac{\mathcal{Z}\,A}{\pi(A|X)}\,|\,X\right].$$

The simulation data are generated by the model

$$\mathcal{Z} = m(X) + h(X)A + \varepsilon,$$

where $m(X)$ is the main effect, $h(X)$ is the interaction effect with treatment $A$, and $\varepsilon$ is the random error. We consider the same main effect and interaction effect functions: $m(X) = 1 + X_1 + X_2$ and $h(X) = 0.5 + X_1 - X_2 + X_3$ respectively, but various types of asymmetric error distributions under three simulation scenarios:

(1) $\log(\varepsilon)$ follows a normal distribution with mean 0 and standard deviation 2;

(2) the random error $\varepsilon$ follows a Weibull distribution with scale parameter 0.5 and shape parameter 0.3;

(3) $\log(\varepsilon)$ follows a normal distribution with mean 0 and standard deviation $2|1 + X_1 + X_2|$.

The above scenarios address heavy right tail distributions to test the robustness of different methods. In particular, the log-normal distribution is frequently used in the finance area, the Weibull distribution is commonly considered in survival analysis of clinical trials, and the third scenario considers a heterogeneous error distribution depending on covariates. In all our simulation studies, the error distributions are asymmetric.

The training sample size is set to be 100 and 200, and the number of covariates $p$ is fixed to be 10. Each covariate is generated by uniform distribution on $[-1, 1]$. In Table 4.1, we list the average computational time and the iteration numbers of the dc algorithm for solving the problem (4.11) with $\lambda_a^N = 0.1$ and $\lambda_\beta^N = 0.1$ over 100 simulations. One can see that the proposed algorithm is very efficient and robust for solving the empirical IDR problem.

| | $n = 100$ | | $n = 200$ | |
|---|---|---|---|---|
| | time | iteration numbers | time | iteration numbers |
| Scenario 1 | 0.70 | 18 | 2.10 | 20 |
| Scenario 2 | 0.79 | 18 | 2.08 | 20 |
| Scenario 3 | 0.68 | 16 | 1.88 | 18 |

**Table 4.1:** The average computational times (in seconds) and dc iteration numbers for $p = 10$.

The comparisons of the four methods for finding optimal IDRs over 100 replications are based on the following four criteria:

(1) the misclassification error rate on the test data (this is possible since the optimal IDR under our simulation settings is known, which is $\text{sign}(0.5 + X_1 - X_2 + X_3)$);

(2) the empirical average of outcome under the decision rule over test data, which is defined as

$$\widehat{\mathbb{E}}^d\,[\,\mathcal{Z}\,] = \frac{\displaystyle\sum_{i \in \mathcal{N}_1} \frac{\mathcal{Z}_i\, \mathbb{I}(A_i = \widehat{d}(X_i))}{\pi(A_i|X_i)}}{\displaystyle\sum_{i \in \mathcal{N}_1} \frac{\mathbb{I}(A_i = \widehat{d}(X_i))}{\pi(A_i|X_i)}},$$

where $\mathcal{N}_1$ is the index of test data set. This value evaluates the expected outcome of $\mathcal{Z}$ if the action assignment follows the estimated decision rules $\widehat{d}(X)$;

(3) the empirical 50% quantile of $\mathcal{Z}_i\mathbb{I}(A_i = \widehat{d}(X_i))$ on the test data;

(4) the empirical 25% quantiles of $\mathcal{Z}_i\mathbb{I}(A_i = \widehat{d}(X_i))$ on the test data.

The test data in each scenario are independently generated with size 10,000.

Several observations can be drawn from these simulation examples in Tables 4.2 and 4.3. First of all, our method under the IDR-CDE has the smallest classification error in choosing

|  | $n = 100$ |  | $n = 200$ |  |
|---|---|---|---|---|
|  | Misclass. | Value | Misclass. | Value |
|  |  | Scenario 1 |  |  |
| DLearn | 0.48(0.02) | 8.36(0.09) | 0.47(0.02) | 8.5(0.07) |
| $l_1$-PLS | 0.45(0.01) | 8.46(0.06) | 0.45(0.01) | 8.58(0.09) |
| RWL | 0.42(0.01) | 8.53(0.07) | 0.42(0.01) | 8.59(0.07) |
| IDR-CDE | **0.25(0.01)** | **8.98(0.07)** | **0.17(0.01)** | **9.15(0.08)** |
|  |  | Scenario 2 |  |  |
| DLearn | 0.44(0.02) | 5.82(0.06) | 0.44(0.02) | 5.74(0.06) |
| $l_1$-PLS | 0.42(0.01) | 5.89(0.05) | 0.4(0.01) | 5.86(0.05) |
| RWL | 0.39(0.01) | 5.95(0.04) | 0.37(0.01) | 5.96(0.04) |
| IDR-CDE | **0.21(0.01)** | **6.36(0.04)** | **0.15(0.01)** | **6.41(0.04)** |
|  |  | Scenario 3 |  |  |
| DLearn | 0.5(0.02) | 3948.04(659.88) | 0.51(0.02) | **26588.55(13692.58)** |
| $l_1$-PLS | 0.48(0.01) | **4758.49(801.06)** | 0.5(0.01) | 26209.19(13702.62) |
| RWL | 0.48(0.01) | 4256.27(774.97) | 0.47(0.01) | 24463.43(13592.7) |
| IDR-CDE | **0.24(0.01)** | 4113.85(934.74) | **0.2(0.01)** | 25712.22(13473.72) |

**Table 4.2:** Average misclassification rates (standard errors) and average means (standard errors) of empirical value functions for three simulation scenarios over 100 runs. The best expected value functions and the minimum misclassification rates are in bold.

correct decisions compared with those under the criterion of expected outcome. Under the piecewise utility function, we emphasize more on improving subjects with relative low outcome, in contrast to focusing on average, which may ignore the subjects with higher-risk. As a result, in addition to misclassification rate, the 50% and 25% quantiles of expected-value functions are also the largest among all the methods. Secondly, the advantages of our method become more obvious if comparing the 25% quantiles of the empirical value functions on the test data with 50% quantiles. For example, in the second scenario, the 25% quantiles of empirical value functions of our method are almost twice as large as those by DLearn. Another interesting finding is that in the last scenario, although the average empirical value functions of $l_1$-PLS and RWL are larger than those of our method, our method is indeed much better based on the misclassification error and the quantiles. One possible reason is that these methods under the expected value function framework only correctly identify the decisions for subjects in lower risk while ignoring subjects with potentially higher risk. The estimated optimal IDRs by those methods may lead to serious problems, especially in precision medicine when assigning treatments to patients. Although, on average, patients may gain benefits of following those decision rules, some patients may come across high risk, causing adverse events such as exacerbation in practice by using the recommended treatment using the standard criterion of expected outcome.

|  | $n = 100$ |  | $n = 200$ |  |
|---|---|---|---|---|
|  | 50% quantile | 25% quantile | 50% quantile | 25% quantile |
| Scenario 1 |  |  |  |  |
| DLearn | 2.64(0.04) | 1.17(0.04) | 2.67(0.04) | 1.21(0.05) |
| $l_1$-PLS | 2.73(0.03) | 1.26(0.03) | 2.74(0.03) | 1.25(0.03) |
| RWL | 2.81(0.03) | 1.35(0.03) | 2.83(0.03) | 1.35(0.04) |
| IDR-CDE | **3.17(0.01)** | **1.81(0.02)** | **3.26(0.01)** | **1.99(0.01)** |
| Scenario 2 |  |  |  |  |
| DLearn | 1.96(0.04) | 0.69(0.04) | 1.97(0.04) | 0.7(0.05) |
| $l_1$-PLS | 2.01(0.03) | 0.77(0.03) | 2.08(0.03) | 0.82(0.03) |
| RWL | 2.1(0.03) | 0.85(0.03) | 2.16(0.03) | 0.92(0.03) |
| IDR-CDE | **2.47(0.01)** | **1.36(0.02)** | **2.53(0.01)** | **1.47(0.01)** |
| Scenario 3 |  |  |  |  |
| DLearn | 2.22(0.05) | 1.02(0.05) | 2.2(0.05) | 1.01(0.05) |
| $l_1$-PLS | 2.3(0.03) | 1.04(0.03) | 2.24(0.03) | 0.99(0.03) |
| RWL | 2.29(0.03) | 1.07(0.03) | 2.31(0.03) | 1.09(0.03) |
| IDR-CDE | **2.81(0.01)** | **1.73(0.01)** | **2.86(0.01)** | **1.8(0.02)** |

**Table 4.3:** Results of average 25% (standard errors) and 50% (standard errors) quantiles of empirical value functions for three simulation scenarios over 100 runs. The largest 25% and 50% quantiles are in bold.

In terms of real data applications, there are several possibilities. For example, we can use the piecewise linear utility function to control the lower tails of outcomes for individual patient in AIDS or cancer studies. Another potential application is to use the quadratic utility function to take variance of each decision rules into consideration. The performance of the results by our method depends on the choice of the covariate-dependent $\alpha(X)$ and the utility function $u$. We leave these as the future work.

## APPENDIX A

## SUPPLEMENTARY MATERIAL TO CHAPTER 2

**Proof of Lemma 2.2.1** Let $g(f) = \mathbf{E}[\frac{1}{\pi(A,\mathbf{x})}(KR\mathbf{w} - \mathbf{f}(\mathbf{x}))^T\Sigma(KR\mathbf{w} - \mathbf{f}(\mathbf{x}))]$. Taking the derivative over $f$ and setting it to zero, we get

$$\begin{aligned}\frac{\partial g(f)}{\partial f} &= 2\Sigma\mathbf{E}_{\mathbf{x}}\{\mathbf{E}[(\frac{KRW}{\pi(A,\mathbf{x})} - \frac{f(\mathbf{x})}{\pi(A,\mathbf{x})})|\mathbf{x}]\} \\ &= 2\Sigma\mathbf{E}_{\mathbf{x}}\{K\mathbf{E}[\frac{RW}{\pi(A,\mathbf{x})}|\mathbf{x}] - f(\mathbf{x})|\mathbf{x}]\} = 0.\end{aligned}$$

**Proof of Lemma 2.3.1**

Let $g(f) = \mathbf{E}[\frac{1}{\pi(A,\mathbf{x})}(\frac{K}{K-1}R - \mathbf{w}^T\mathbf{f}(\mathbf{x}))^T(\frac{K}{K-1}R - \mathbf{w}^T\mathbf{f}(\mathbf{x}))]$. Taking the derivative over $f$ and setting it to zero, we get

$$\begin{aligned}\frac{\partial g(f)}{\partial f} &= \mathbf{E}_{\mathbf{x}}\{\mathbf{E}[W(\frac{KR}{(K-1)\pi(A,\mathbf{x})} - \frac{W^T f(\mathbf{x})}{\pi(A,\mathbf{x})})|\mathbf{x}]\} \\ &= \mathbf{E}_{\mathbf{x}}\{\frac{K}{K-1}\mathbf{E}[\frac{RW}{\pi(A,\mathbf{x})}|\mathbf{x}] - \frac{K}{K-1}f(\mathbf{x})|\mathbf{x}]\} = 0,\end{aligned}$$

where the second equality holds because $\mathbf{E}[\frac{WW^T}{\pi(A,\mathbf{x})}|\mathbf{x}] = \frac{K}{K-1}I_{K-1}$ by definition. Thus $\mathbf{f}_0(\mathbf{x})$ is an optimal solution.

**Proof of Lemma 2.3.2**

Let $g(f) = \mathbf{E}[-\frac{R\mathbf{w}^T f}{\pi(A,\mathbf{x})} + \frac{\log(1+\exp(\mathbf{w}^T f))}{\pi(A,\mathbf{x})}]$. Taking the derivative over $f$ and setting it to zero, we get

$$\begin{aligned}\frac{\partial g(f)}{\partial f} &= 2\mathbf{E}_{\mathbf{x}}\{\mathbf{E}[(\frac{RW}{\pi(A,\mathbf{x})} - \frac{W\exp(\mathbf{w}^T f)}{1+\exp(\mathbf{w}^T f))\pi(A,\mathbf{x})})|\mathbf{x}]\} \\ &= 2\mathbf{E}_{\mathbf{x}}\{\sum_{i=1}^{K}\mathbf{w}_i\mathbf{P}[R=1|\mathbf{x},A=i] - \sum_{i=1}^{K}\mathbf{w}_i\frac{\exp(\mathbf{w}_i^T f)}{1+\exp(\mathbf{w}_i^T f)}\} \\ &= 0.\end{aligned}$$

If $\mathbf{P}[R=1|\mathbf{x},A=i] = \frac{\exp(\mathbf{w}_i^T f^*)}{1+\exp(\mathbf{w}_i^T f^*)}$, then $f^*$ is an optimal solution to (2.18).

**Proof of Lemma 2.3.3**

Let $g(f) = \mathbf{E}[\int_0^\tau \frac{\log \mathbf{E}[e^{f^T \mathbf{w}} \mathbb{I}(Y \geq u)]}{\pi(A,\mathbf{x})} - \frac{f^T \mathbf{w}}{\pi(A,\mathbf{x})} dN(u)]$. Taking the derivative over $f$ and setting it to zero, we get

$$\begin{aligned}\frac{\partial g(f)}{\partial f} &= \mathbf{E}_{\mathbf{x}}\{\int_0^\tau \sum_{i=1}^K \mathbf{w}_i \mathbf{E}[\mathbb{I}(Y \geq u)\lambda_i(u,\mathbf{x})|\mathbf{x}, A = i] - \frac{\mathbf{w}_i \exp(\mathbf{w}_i^T f)\mathbb{I}(Y^{(i)} \geq u)\mathbf{E}[\mathbb{I}(Y \geq u)\lambda(u,\mathbf{x})|\mathbf{x}]}{\mathbf{E}[\exp(W^T f)\mathbb{I}(Y \geq u)]} du\} \\ &= \mathbf{E}_{\mathbf{x}}\{\int_0^\tau \sum_{i=1}^K \mathbf{w}_i(\mathbf{E}[\mathbb{I}(Y \geq u)\lambda_i(u,\mathbf{x})|\mathbf{x}, A = i] - \exp(\mathbf{w}_i^T f)\Lambda^*(Y^{(i)}))du\} \\ &= 0,\end{aligned}$$

where $\lambda_i(u,\mathbf{x})$ is the hazard function for the $i$-th treatment and $\Lambda^*(Y)$ is the cumulative hazard function. Then we get a sufficient condition that if $\exp(\mathbf{w}_i^T f)\Lambda^*(Y^{(i)}) = \mathbf{P}[\delta = 1|\mathbf{x}, A = i]$, then $f^*$ is an optimal solution. If the censoring time in each treatment group is the same, then we get (2.25).

**Proof of Theorem 2.4.1**

For any ITR $d$, we have

$$\begin{aligned}
V(d) &= \mathbf{E}[\sum_{k=1}^{K} \mathbf{E}[R|\mathbf{x}, A=k]\mathbb{I}(d(\mathbf{x})=k)] \\
&= \mathbf{E}[\frac{1}{1-C(K)}\{\sum_{k=1}^{K}(1-C(K))\mathbf{E}[R|\mathbf{x}, A=k]\mathbb{I}(d(\mathbf{x})=k) \\
&+ \sum_{j=1}^{K} C(K)\mathbf{E}[R|\mathbf{x}, A=j]\} - \frac{C(K)}{1-C(K)}\sum_{j=1}^{K}\mathbf{E}[R|\mathbf{x}, A=j]] \\
&= \mathbf{E}[\frac{1}{1-C(K)}\{\sum_{k=1}^{K} \mathbf{E}[R|\mathbf{x}, A=k]\mathbb{I}(d(\mathbf{x})=k) \\
&+ \sum_{j=1}^{K} C(K)\mathbf{E}[R|\mathbf{x}, A=j]\sum_{i\neq j}^{K}\mathbb{I}(d(\mathbf{x})=i)\}] - \Delta \\
&= \mathbf{E}[\frac{1}{1-C(K)}\{\sum_{k=1}^{K} \mathbf{E}[R|\mathbf{x}, A=k]\mathbb{I}(d(\mathbf{x})=k) \qquad \text{(A.1)} \\
&+ \sum_{i=1}^{K}\sum_{j\neq i}^{K} C(K)\mathbf{E}[R|\mathbf{x}, A=j]\mathbb{I}(d(\mathbf{x})=i)\}] - \Delta \\
&= \mathbf{E}[\frac{1}{1-C(K)}\{\sum_{k=1}^{K}(\mathbf{E}[R|\mathbf{x}, A=k] \\
&+ \sum_{j\neq k}^{K} C(K)\mathbf{E}[R|\mathbf{x}, A=k])\mathbb{I}(d(\mathbf{x})=k)\}] - \Delta \\
&= \mathbf{E}[\frac{1}{1-C(K)}\{\sum_{k=1}^{K} \mathbf{w}_k^T\mathbf{E}[\frac{RW}{\pi(A,\mathbf{x})}|\mathbf{x}]\mathbb{I}(d(\mathbf{x})=k)\}] - \Delta \\
&= \mathbf{E}[\frac{1}{1-C(K)}\{\sum_{k=1}^{K} \mathbf{w}_k^T\mathbf{f}_0(\mathbf{x})\mathbb{I}(d(\mathbf{x})=k)\}] - \Delta,
\end{aligned}$$

where $\Delta = \mathbf{E}[C(K)\sum_{j=1}^{K}\mathbf{E}[R|\mathbf{x}, A=j]]$ that does not depend on the ITR $d$. Then we can obtain the value reduction bound between the optimal ITR $d_0$ and our estimated ITR $\hat{d}$ by using (A.1):

$$\begin{aligned}
&V(d_0) - V(\hat{d}) \\
\leq &\frac{1}{1-C(K)}\mathbf{E}[\{\sum_{k=1}^{K} \mathbf{w}_k^T \mathbf{f}_0(\mathbf{x})(\mathbb{I}(d(\mathbf{x}) = k) - \mathbb{I}(\hat{d}(\mathbf{x}) = k)\}] \\
\leq &\frac{1}{1-C(K)}\mathbf{E}[\{\sum_{i \neq j} |\mathbf{w}_i^T \mathbf{f}_0(\mathbf{x}) - w_j^T \mathbf{f}_0(\mathbf{x})| \mathbb{I}(d(\mathbf{x}) = i, \hat{d}(\mathbf{x}) = j)\}] \\
\leq &\frac{1}{1-C(K)}\mathbf{E}[\{\sum_{i \neq j} |\mathbf{w}_i^T \mathbf{f}_0(\mathbf{x}) - w_j^T \mathbf{f}_0(\mathbf{x})| \mathbb{I}(\mathbf{w}_i^T \mathbf{f}_0(\mathbf{x}) - w_j^T \mathbf{f}_0(\mathbf{x}))(\mathbf{w}_i^T \hat{f}(\mathbf{x}) - w_j^T \hat{f}(\mathbf{x}) < 0)\}] \\
\leq &\frac{1}{1-C(K)}\mathbf{E}[\{\sum_{i \neq j} |\mathbf{w}_i^T (\mathbf{f}_0(\mathbf{x}) - \hat{f}(\mathbf{x})) - w_j^T (\mathbf{f}_0(\mathbf{x}) - \hat{f}(\mathbf{x}))| \\
&\mathbb{I}(\mathbf{w}_i^T (\mathbf{f}_0(\mathbf{x}) - \hat{f}(\mathbf{x})) w_j^T (\mathbf{f}_0(\mathbf{x}) - \hat{f}(\mathbf{x})) < 0)\}] \\
\leq &\frac{1}{1-C(K)} \sum_{i \neq j} (\mathbf{E}||\mathbf{f}_0(\mathbf{x}) - \hat{f}(\mathbf{x})||_2 + \mathbf{E}||\mathbf{f}_0(\mathbf{x}) - \hat{f}(\mathbf{x})||_2) \\
\leq &\frac{2K(K-1)}{1-C(K)} (\mathbf{E}||\mathbf{f}_0(\mathbf{x}) - \hat{f}(\mathbf{x})||_2^2)^{\frac{1}{2}},
\end{aligned} \tag{A.2}$$

where the second to last inequaltiy holds by using the Hölder and Minkowski inquality together with $||\mathbf{w}_i|| = 1$ for $i = 1, \cdots, K$. Furthermore, if we assume Assumption 1 holds, then we can further bound the value reduction by

$$\begin{aligned}
&V(d_0) - V(\hat{d}) \\
\leq &\frac{1}{1-C(K)}\mathbf{E}[\{\sum_{i \neq j} |\mathbf{w}_i^T \mathbf{f}_0(\mathbf{x}) - w_j^T \mathbf{f}_0(\mathbf{x})| \mathbb{I}(\mathbf{w}_i^T \mathbf{f}_0(\mathbf{x}) - w_j^T \mathbf{f}_0(\mathbf{x})(\mathbf{w}_i^T \hat{f}(\mathbf{x}) - w_j^T \hat{f}(\mathbf{x}) < 0)\}] \\
\leq &\frac{1}{1-C(K)}\mathbf{E}[\{\sum_{i \neq j} \epsilon \mathbb{I}(|(\mathbf{w}_i - w_j)^T \mathbf{f}_0(\mathbf{x})| < \epsilon) \mathbb{I}((\mathbf{w}_i - w_j)^T \mathbf{f}_0(\mathbf{x}))((\mathbf{w}_i - w_j)^T \hat{f}(\mathbf{x})) < 0)\}] \\
&+ \frac{1}{1-C(K)\epsilon}\mathbf{E}[\{\sum_{i \neq j} (\mathbf{w}_i^T \mathbf{f}_0(\mathbf{x}) - w_j^T \mathbf{f}_0(\mathbf{x}))^2 \mathbb{I}((\mathbf{w}_i - w_j)^T \mathbf{f}_0(\mathbf{x}))((\mathbf{w}_i - w_j)^T \hat{f}(\mathbf{x})) < 0)\}] \\
\leq &\frac{1}{1-C(K)} \sum_{i \neq j} \epsilon \mathbf{P}[|(\mathbf{w}_i - w_j)^T \mathbf{f}_0(\mathbf{x})| < \epsilon] + \frac{2}{\epsilon}(\mathbf{E}||\mathbf{f}_0(\mathbf{x}) - \hat{f}(\mathbf{x})||_2^2 + \mathbf{E}||\mathbf{f}_0(\mathbf{x}) - \hat{f}(\mathbf{x})||_2^2) \\
\leq &\frac{1}{1-C(K)} \sum_{i \neq j} C\epsilon^{\alpha+1} + \frac{4}{\epsilon}\mathbf{E}||\mathbf{f}_0(\mathbf{x}) - \hat{f}(\mathbf{x})||_2^2),
\end{aligned} \tag{A.3}$$

for any $\epsilon > 0$. We can then minimize right hand side above over $\epsilon$ and get the desired bound

$$V(d_0) - V(\hat{d}_n) \leq C_1(K, \alpha)(E||\mathbf{f}_0 - \hat{\mathbf{f}}_n||_2^2)^{\frac{1+\alpha}{2+\alpha}}.$$

**Proof of Theorem 2.4.2**

Define $\boldsymbol{\beta}^j = (\boldsymbol{\beta}_{kj}, k = 1, \cdots, (K-1))^T$, and let $\lambda = \sigma\sqrt{\frac{(\log p)^{1+\delta}}{n(K-1)}}$. With probability at least $1 - \frac{(2e \log p - e)c}{(\log p)^{1+\delta}}$, we have the following inequality

$$\begin{aligned} \frac{1}{n(K-1)}||\mathbf{Z}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0)||^2 + \lambda||\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}||_{2,1} \leq & \\ \leq \frac{1}{n(K-1)}||\mathbf{Z}(\boldsymbol{\beta} - \boldsymbol{\beta}_0)||^2 + 4\lambda \sum_{j \in S(\boldsymbol{\beta})} ||\hat{\boldsymbol{\beta}}^j - \boldsymbol{\beta}^j||, \end{aligned} \tag{A.4}$$

for any $\boldsymbol{\beta}$. This was previously shown in Theorem 5.2 by (Lounici et al., 2009). Let $\boldsymbol{\beta} = \boldsymbol{\beta}_0$. Then with probability at least $1 - \frac{(2e \log p - e)c}{(\log p)^{1+\delta}}$, we have

$$\begin{aligned} \frac{1}{n(K-1)}||\mathbf{Z}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0)||^2 &\leq 4\lambda \sum_{j \in S(\boldsymbol{\beta})} ||\hat{\boldsymbol{\beta}}^j - \boldsymbol{\beta}^j|| \\ &\leq 4\lambda\sqrt{s}||(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta})_S|| \end{aligned}$$

and

$$||\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}||_{2,1} \leq 4||(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta})_S||,$$

which implies $||\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}||_{S^c} \leq 3||(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta})_S||$. Then by the RE(s) assumption, with probability at least $1 - \frac{(2e \log p - e)c}{(\log p)^{1+\delta}}$, we have

$$\begin{aligned} \frac{1}{n(K-1)}||\mathbf{Z}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0)||^2 &\leq 4\lambda\sqrt{s}||(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta})_S|| \\ &\leq 4\lambda\sqrt{s}\frac{||\mathbf{Z}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0)||}{\rho(s)\sqrt{n}}, \end{aligned}$$

such that we can bound the empirical error by

$$\frac{1}{n}||\mathbf{Z}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0)||^2 \leq \frac{16(K-1)}{\rho(s)}\sigma^2 s\frac{(\log p)^{1+\delta}}{n}.$$

With the RE(2s) assumption, we can further show that with the same probability

$$\frac{1}{\sqrt{K-1}}||\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0|| \leq \frac{4\sqrt{10}}{\rho^2(2s)}\sigma\sqrt{\frac{s(\log p)^{1+\delta}}{n}}.$$

Combining with Theorem 2.4.1, we get the value reduction bound

$$V(d_0) - V(\hat{d}_n) \leq \frac{\sqrt{K-1}K(K-1)}{1-C(K)}\frac{4\sqrt{10}c}{\rho^2(2s)}\sigma\sqrt{\frac{s(\log p)^{1+\delta}}{n}}.$$

Together with our margin condition, we can directly get the corresponding improved bound (2.31). **Additional Simulation Studies** In this section, we include additional simulation results to further demonstrate the performance of our methods. **Continuous Outcome Studies** When the clinical outcome $R$ is continuous, we report the simulation results with $n = 400, 800, p = 20$ and $n = 200, p = 20, 40$ for the three continuous simulation scenarios in the main paper.

**Table A.1:** Results of average means (standard deviations) of empirical value functions and misclassification rates for four continuous-outcome simulations scenarios with 20 covariates. The best value functions and misclassification rates are in bold.

| | $n = 400$ | | $n = 800$ | |
|---|---|---|---|---|
| | Value | Misclassification | Value | Misclassification |
| Scenario 1 | | | | |
| Pair-D | 2.79(0.04) | 0.45(0.02) | 3.04(0.02) | 0.29(0.02) |
| $l_1$-PLS | **3.12**(0.01) | **0.21**(0.01) | **3.16**(0.01) | **0.14**(0.01) |
| DL | 2.67(0.03) | 0.51(0.01) | 2.78(0.02) | 0.47(0.01) |
| ACWL-1 | 2.77(0.03) | 0.43(0.01) | 2.91(0.02) | 0.37(0.01) |
| ACWL-2 | 2.91(0.02) | 0.37(0.01) | 3.04(0.01) | 0.3(0.01) |
| VT | 2.75(0.02) | 0.48(0.01) | 2.85(0.01) | 0.43(0.01) |
| Group-AD | 3.11(0.03) | **0.21**(0.02) | 3.14(0.03) | 0.16(0.02) |
| Scenario 2 | | | | |
| Pair-D | 2.82(0.11) | 0.33(0.03) | 2.92(0.1) | 0.3(0.03) |
| $l_1$-PLS | 2.93(0.11) | 0.36(0.04) | 2.99(0.1) | 0.32(0.03) |
| DL | 2.88(0.11) | 0.34(0.04) | 2.98(0.12) | 0.28(0.04) |
| ACWL-1 | 2.78(0.11) | 0.38(0.02) | 2.96(0.1) | 0.31(0.02) |
| ACWL-2 | 2.86(0.10) | 0.37(0.02) | 3.04(0.1) | **0.28**(0.03) |
| VT | **3.04**(0.09) | **0.30**(0.02) | **3.09**(0.1) | 0.28(0.03) |
| Group-AD | 2.91(0.1) | 0.32(0.03) | 2.9(0.11) | 0.31(0.03) |
| Scenario 3 | | | | |
| Pair-D | 1.2(0.04) | 0.75(0.03) | 1.21(0.04) | 0.75(0.03) |
| $l_1$-PLS | 1.51(0.19) | 0.54(0.15) | 1.64(0.2) | 0.41(0.18) |
| DL | 1.43(0.1) | 0.61(0.06) | 1.52(0.07) | 0.55(0.06) |
| ACWL-1 | 1.43(0.07) | 0.61(0.05) | 1.49(0.07) | 0.56(0.05) |
| ACWL-2 | 1.47(0.07) | 0.58(0.05) | 1.63(0.06) | 0.45(0.05) |
| VT | 1.42(0.05) | 0.62(0.03) | 1.48(0.04) | 0.57(0.03) |
| Group-D | **1.65**(0.11) | **0.43**(0.09) | **1.77**(0.04) | **0.29**(0.05) |

**Table A.2:** Results of average means (std) of empirical value functions and misclassification rates for four continuous-outcome simulation scenarios with $n = 200$. The best value functions and misclassification rates are in bold.

| | $p = 20$ | | $p = 40$ | |
|---|---|---|---|---|
| | Value | Misclassification | Value | Misclassification |
| | | Scenario 1 | | |
| Pair-D | 2.54(0.06) | 0.56(0.02) | 2.36(0.05) | 0.63(0.02) |
| $l_1$-PLS | 3.04(0.02) | 0.29(0.01) | 2.94(0.03) | 0.36(0.02) |
| DL | 2.51(0.04) | 0.58(0.02) | 2.46(0.04) | 0.6(0.01) |
| ACWL-1 | 2.64(0.04) | 0.5(0.02) | 2.28(0.05) | 0.65(0.02) |
| ACWL-2 | 2.7(0.03) | 0.48(0.01) | 2.3(0.05) | 0.64(0.02) |
| VT | 2.64(0.03) | 0.53(0.01) | 2.57(0.03) | 0.55(0.01) |
| Group-AD | **3.05**(0.03) | **0.28**(0.02) | **3.02**(0.03) | **0.31**(0.02) |
| | | Scenario 2 | | |
| Pair-D | 2.71(0.13) | 0.37(0.04) | 2.73(0.14) | 0.36(0.05) |
| $l_1$-PLS | 2.81(0.12) | 0.41(0.04) | 2.83(0.12) | 0.41(0.05) |
| DL | 2.72(0.12) | 0.39(0.04) | 2.71(0.13) | 0.4(0.04) |
| ACWL-1 | 2.57(0.12) | 0.43(0.03) | 2.05(0.13) | 0.57(0.03) |
| ACWL-2 | 2.63(0.12) | 0.43(0.03) | 2.1(0.13) | 0.57(0.03) |
| VT | **2.97**(0.1) | **0.33**(0.02) | **3.01**(0.09) | **0.33**(0.02) |
| Group-AD | 2.86(0.11) | 0.34(0.03) | 2.9(0.11) | 0.34(0.02) |
| | | Scenario 3 | | |
| Pair-D | 1.2(0.03) | 0.75(0.03) | 1.2(0.03) | 0.75(0.03) |
| $l_1$-PLS | 1.37(0.15) | 0.65(0.1) | 1.31(0.11) | 0.69(0.08) |
| DL | 1.34(0.09) | 0.67(0.06) | 1.29(0.09) | 0.7(0.05) |
| ACWL-1 | 1.27(0.08) | 0.71(0.05) | 1.2(0.04) | 0.74(0.02) |
| ACWL-2 | 1.27(0.08) | 0.71(0.05) | 1.2(0.05) | 0.75(0.03) |
| VT | 1.35(0.07) | 0.66(0.04) | **1.33**(0.06) | **0.68**(0.04) |
| Group-D | **1.38**(0.16) | **0.64**(0.11) | 1.31(0.13) | **0.68**(0.09) |

**Binary Outcome Studies** For the binary outcome $R$, we report the simulation results with $n = 400, 800, p = 20$ and $n = 200, p = 20, 40$ in addition to the results in the main paper.

**Table A.3:** Results of average means (standard deviations) of empirical value functions and misclassification rates for two binary-outcome simulation scenarios with 20 covariates. The best value functions and misclassification rates are in bold.

| | $n = 400$ | | $n = 800$ | |
|---|---|---|---|---|
| | Value | Misclassification | Value | Misclassification |
| | | Scenario 1 | | |
| $l_1$-PLR | 0.89(0.01) | 0.53(0.03) | **0.92**(0) | 0.4(0.02) |
| DL | 0.86(0.01) | 0.64(0.02) | 0.88(0.01) | 0.58(0.02) |
| VT | 0.85(0.01) | 0.66(0.02) | 0.85(0) | 0.67(0.02) |
| Binary-AD | **0.91**(0.01) | **0.41**(0.03) | **0.92**(0) | **0.31**(0.03) |
| | | Scenario 2 | | |
| $l_1$-PLR | 0.84(0.01) | 0.63(0.02) | **0.87**(0) | 0.56(0.03) |
| DL | 0.82(0.01) | 0.53(0.04) | 0.85(0.01) | 0.45(0.04) |
| VT | 0.84(0.01) | 0.42(0.05) | 0.83(0.01) | 0.5(0.05) |
| Binary-AD | **0.86**(0.01) | **0.44**(0.03) | **0.87**(0.01) | **0.42**(0.02) |

**Table A.4:** Results of average means (standard deviation) of empirical value functions and misclassification rates for two binary-outcome simulation scenarios with $n = 200$. The best value functions and misclassification rates are in bold.

| | $p = 20$ | | $p = 40$ | |
|---|---|---|---|---|
| | Value | Misclassification | Value | Misclassification |
| | | Scenario 1 | | |
| $l_1$-PLR | 0.86(0.01) | 0.62(0.02) | 0.85(0.01) | 0.66(0.02) |
| DL | 0.85(0.01) | 0.68(0.01) | 0.84(0.01) | 0.7(0.01) |
| VT | 0.84(0.01) | 0.66(0.01) | 0.83(0.01) | 0.68(0.01) |
| Binary-AD | **0.88**(0.01) | **0.54**(0.02) | **0.87**(0.01) | **0.57**(0.02) |
| | | Scenario 2 | | |
| $l_1$-PLR | 0.81(0.01) | 0.68(0.05) | 0.79(0.01) | 0.7(0.05) |
| DL | 0.78(0.01) | 0.59(0.01) | 0.76(0.01) | 0.61(0.01) |
| VT | **0.84**(0.01) | **0.38**(0.01) | **0.83**(0.01) | **0.36**(0.01) |
| Binary-AD | 0.83(0.01) | 0.49(0.04) | 0.82(0.01) | 0.52(0.04) |

**Survival Outcome Studies** For the survival outcome $R$, we report the simulation results with $n = 400, 800, p = 20$ and $n = 200, p = 20, 40$ in addition to the results in the main paper.

**Table A.5:** Results of average means (standard deviations) of empirical value functions and misclassification rates for two survival-outcome simulation scenarios with 20 covariates. The best value functions and misclassification rates are in bold.

| | $n = 400$ | | $n = 800$ | |
|---|---|---|---|---|
| | Value | Misclassification | Value | Misclassification |
| | | Scenario 1 | | |
| $l_1$-CPH | 43.18(2.02) | 0.26(0.04) | 45.70(1.02) | **0.17**(0.01) |
| Surv-AD | **44.32**(1.27) | **0.23**(0.02) | **45.81**(1.06) | **0.17**(0.01) |
| | | Scenario 2 | | |
| $l_1$-CPH | **22.48**(0.69) | 0.53(0.04) | **23.52**(0.57) | 0.47(0.04) |
| Surv-AD | 22.28(0.61) | **0.45**(0.02) | 22.77(0.49) | **0.43**(0.02) |

**Table A.6:** Results of average means (standard deviation) of empirical value functions and misclassification rates for two survival-outcome simulation scenarios with $n = 200$. The best value functions and misclassification rates are in bold.

| | $p = 20$ | | $p = 40$ | |
|---|---|---|---|---|
| | Value | Misclassification | Value | Misclassification |
| | | Scenario 1 | | |
| $l_1$-CPH | 36.27(3.25) | 0.44(0.06) | 32.19(3.47) | 0.52(0.06) |
| Surv-AD | **40.41**(1.85) | **0.35**(0.03) | **39.46**(2.03) | **0.38**(0.03) |
| | | Scenario 2 | | |
| $l_1$-CPH | 20.98(0.92) | 0.6(0.03) | 19.82(1.08) | 0.63(0.03) |
| Surv-AD | **21.14**(0.95) | **0.49**(0.04) | **20.63**(0.95) | **0.51**(0.04) |

**Low Rank Simulation Studies** When the clinical outcome $R$ is continuous, we generate our data from the following model with

$$R_i = \mu(\mathbf{x}_i) + \sum_{k=1}^{K} (\mathbf{x}_i^T \beta_k) \mathbb{I}(A = k) + \epsilon_i,$$

where $i = 1, \cdots, n$, each covariate is generated by the uniform distribution from $-1$ to 1, and $\epsilon_i$ follows from the standard normal distribution. Let the coefficient matrix $\Gamma = (\beta_1, \cdots, \beta_k)$. We consider the following two simulation scenarios:

1. $\mu(\mathbf{x}) = 1 + X_1 + X_2$ and the coefficient matrix $\Gamma = \mathbf{U}\mathbf{V}^T$, where $\mathbf{U} \in \mathbb{R}^{p\times 2}$ and $\mathbf{V} \in \mathbb{R}^{p\times 2}$. Each element of $\mathbf{U}$ and $\mathbf{V}$ is generated by the uniform distribution from $-1$ to 1;

2. $\mu(\mathbf{x}) = 1 + X_1^2 + X_2^2$ and the coefficient matrix $\Gamma$ is the same as Scenario 1.

The difference in these two scenarios lies in the linear and nonlinear main effect functions. For each simulation scenario, we compare the following methods:

(1) $l_1$-PLS proposed by (Qian and Murphy, 2011) with basis $(1, \mathbf{x}, \mathbf{x}A)$;

(2) Pairwise D-learning;

(3) AD-learning with the group sparsity penalty;

(4) AD-learning with the nuclear norm penalty.

All the tuning parameters are selected via 10-fold cross-validation. We report the value functions and misclassification errors for both $p = 20$ and $p = 40$ on 10000 independently generated test data in the following tables. We can see that our AD-learning has some advantages over $l_1$-PLS and pairwise D-learning.

**Table A.7:** Results of average means (std) of empirical value functions and misclassification rates for four continuous-outcome simulation scenarios with 20 covariates. The best value functions and misclassification rates are in bold.

| | $n = 800$ | | $n = 1600$ | |
|---|---|---|---|---|
| | Value | Misclassification | Value | Misclassification |
| | | Scenario 1 | | |
| $l_1$-PLS | 1.64(0.41) | 0.52(0.29) | 1.7(0.38) | 0.47(0.27) |
| Pair-D | 2(0.05) | 0.32(0.06) | 2.06(0.04) | 0.25(0.05) |
| Group-AD | 1.97(0.06) | 0.35(0.07) | 2.05(0.04) | 0.26(0.05) |
| Low rank-AD | **2.05(0.04)** | **0.22(0.05)** | **2.09(0.04)** | **0.17(0.03)** |
| | | Scenario 2 | | |
| $l_1$-PLS | 2.37(0.36) | 0.47(0.24) | 2.52(0.35) | 0.35(0.23) |
| Pair-D | 2.6(0.1) | 0.38(0.1) | 2.69(0.06) | 0.29(0.08) |
| Group-AD | 2.6(0.07) | 0.38(0.07) | 2.68(0.05) | 0.3(0.06) |
| Low rank-AD | **2.7(0.06)** | **0.24(0.06)** | **2.75(0.04)** | **0.19(0.03)** |

**Table A.8:** Results of average means (std) of empirical value functions and misclassification rates for four continuous-outcome simulation scenarios with 40 covariates. The best value functions and misclassification rates are in bold.

| | $n = 800$ | | $n = 1600$ | |
|---|---|---|---|---|
| | Value | Misclassification | Value | Misclassification |
| | | Scenario 1 | | |
| $l_1$-PLS | 1.64(0.41) | 0.52(0.29) | 1.7(0.38) | 0.47(0.27) |
| Pair-D | 2(0.05) | 0.32(0.06) | 2.06(0.04) | 0.25(0.05) |
| Group-AD | 1.97(0.06) | 0.35(0.07) | 2.05(0.04) | 0.26(0.05) |
| Low rank-AD | **2.05(0.04)** | **0.22(0.05)** | **2.09(0.04)** | **0.17(0.03)** |
| | | Scenario 2 | | |
| $l_1$-PLS | 2.37(0.36) | 0.47(0.24) | 2.52(0.35) | 0.35(0.23) |
| Pair-D | 2.6(0.1) | 0.38(0.1) | 2.69(0.06) | 0.29(0.08) |
| Group-AD | 2.6(0.07) | 0.38(0.07) | 2.68(0.05) | 0.3(0.06) |
| Low rank-AD | **2.7(0.06)** | **0.24(0.06)** | **2.75(0.04)** | **0.19(0.03)** |

**Further Comparison with $l_1$-PLS** In this section, we compare our proposed AD-learning with $l_1$-PLS when the main effect functions $\mu(\mathbf{x}) = 1+X_1^2+X_2^2$ in the first scenario of continuous outcome settings with the non-zero coefficients are generated by uniform distribution from $-1$ to 1. Table A.9 demonstrates the advantage of our proposed method over $l_1$-PLS by avoiding modeling main effect functions.

**Table A.9:** Results of average means (std) of empirical value functions and misclassification rates for two simulation scenarios with 20 covariates. The best value functions and misclassification rates are in bold.

| | $n = 400$ | | $n = 800$ | |
|---|---|---|---|---|
| | Value | Misclassification | Value | Misclassification |
| | | Scenario 1 | | |
| $l_1$-PLS | 1.64(0.41) | 0.52(0.29) | 1.7(0.38) | 0.47(0.27) |
| Group-AD | **1.97**(0.06) | **0.35**(0.07) | **2.05**(0.04) | **0.26**(0.05) |

### Comparison between Group $l_1$-PLS and $l_1$-PLS

In this section, we compare the performance of group $l_1$-PLS and $l_1$-PLS using the first simulation scenario of the continuous outcome study. Table A.10 demonstrates the performance of $l_1$-PLS in our simulation scenarios is similar to group $l_1$-PLS.

**Table A.10:** Results of average means (std) of empirical value functions and misclassification rates for one continuous-outcome simulation scenarios with 20 covariates. The best value functions and misclassification rates are in bold.

| | $n = 400$ | | $n = 800$ | |
|---|---|---|---|---|
| | Value | Misclassification | Value | Misclassification |
| | | Scenario 1 | | |
| Group $l_1$-PLS | 3.12(0.07) | 0.21(0.05) | 3.16(0.04) | 0.14(0.03) |
| $l_1$-PLS | **3.12**(0.06) | **0.2**(0.05) | **3.17** (0.04) | **0.14**(0.03) |

# APPENDIX B
# SUPPLEMENTARY MATERIAL TO CHAPTER 3

**Additional Proof** This supplementary material collects the required concepts for the proof of Theorem 3.2.1 and 3.2.2 in Section 3.2, and all the technical proof in Sections 3.3 and 3.4 of Chapter 3.

## Decomposable Space and Normal Integrand Related to Theorem 3.2.1

In order to exchange the supreme operator over $\alpha(X)$ and the expectation with respect to $\mathbf{E}^d$, we need to first introduce the concept of a decomposable space and the normal integrand.

**Definition B.1.** (Rockafellar and Wets, 2009, Definitions 14.59). A space $\mathcal{L}$ of Borel $\mathcal{B}$-measurable functions is *decomposable* relative to an underlying measure space $(\Omega, \mathcal{B}, \mu)$ if for every function $y_0 \in \mathcal{L}$, every set $A \in \mathcal{B}$ with $\mu(A) < \infty$ and any bounded, measurable function $y_1$, the function $y_2(t) = y_0(t)\mathbb{I}(t \notin G) + y_1(t)\mathbb{I}(t \in G)$ belongs to $\mathcal{L}$.

**Definition B.2.** (Rockafellar and Wets, 2009, Definitions 14.27). An extended-value function $f : \Omega_0 \times \mathbb{R} \to (-\infty, \infty]$ is a *normal integrand* if its epigraphical mapping $\omega \to \text{epi } f(\omega, \cdot)$ is closed-valued and measurable. □

We will employ the following simplified version of (Rockafellar and Wets, 2009, Theorem 14.60) that provides the required conditions for the exchange of the supremum and expectation in our context.

**Theorem B.0.1.** *Let $(\Omega,\ \mathcal{B},\ \mu)$ be a probability measure space, and $\mathcal{L}$ be a decomposable space of $\mathcal{B}$-measurable functions. Let $f : \Omega \times \mathbb{R} \to (-\infty, \infty]$ be a normal integrand; let the integral functional $I_f(x) = \int_\Omega f(x(\omega), \omega) d\mu(\omega)$ be defined on $\mathcal{L}$. The following two statements hold:*

*(a)* $\inf\limits_{x \in \mathcal{L}} \displaystyle\int_\Omega f(x(\omega), \omega) d\mu(\omega) = \int_\Omega \inf_{s \in \mathbb{R}} f(s, \omega) d\mu(\omega)$ *as long as $I_f(x)$ is finite; and*

*(b)* $x_0 \in \mathop{argmin}\limits_{x \in \mathcal{M}} I_f(x) \Longleftrightarrow x_0(\omega) \in \mathop{argmin}\limits_{s \in \mathbb{R}} f(s, \omega)$ *almost surely.* □

## Proof of Theorem 3.2.1

Note that $\mathcal{L}_1(\mathcal{X}, \Xi, P_{\mathbf{x}})$ is a decomposable space by checking the definition. It is enough to justify $-\mathbf{E}\left[\alpha(\mathbf{x}) - \frac{1}{\gamma}(\alpha(\mathbf{x}) - R)_+ \middle| \mathbf{x}, A = d(\mathbf{x})\right]$ is a normal integrand. Since this term is measurable with respect to $\mathbf{x}$ and continuous in $\alpha(X)$, it is a normal integrand ((Rockafellar and Wets, 2009, Example 14.29)).

**Proof of Proposition 3.2.3**

We first show the duality representation of $M_1(d)$. Rewrite $M_1(d)$ in the definition of (3.7) as

$$\begin{aligned} M_1(d) &= \mathbf{E}^d[((1-\tau)R + \frac{\tau}{\gamma} R\mathbb{I}(R \leq Q_\gamma(P^d)))] \\ &:= \rho^d(R) \end{aligned}$$

to connect with the bounded random variable $R \in L^\infty(\mathcal{T}, \mathcal{F}_1, P^d)$. Then $\rho^d(R)$ has the following important properties:

(A1) Monotonicity. If $R_1(\omega) \geq R_2(\omega)$ for $\omega \in \mathcal{T}$, where $R_1, R_2 \in L^\infty(\mathcal{T}, \mathcal{F}_1, P^d)$, then $\rho^d(R_1) \geq \rho^d(R_2)$;

(A2) Concavity. For $\lambda \in [0, 1]$ and $R_1, R_2 \in L^\infty(\mathcal{T}, \mathcal{F}_1, P^d)$, one can have

$$\lambda\rho^d(R_1) + (1-\lambda)\rho^d(R_2) \leq \rho^d(\lambda R_1 + (1-\lambda)R_2);$$

(A3) Translation invariance. For any constant $c \in \mathbb{R}$ and $R \in L^\infty(\mathcal{T}, \mathcal{F}_1, P^d)$, one can have $\rho^d(R + c) = \rho^d(R) + c$;

(A4) Positive homogeneity. For any constant $c' > 0$ and $R \in L^\infty(\mathcal{T}, \mathcal{F}_1, P^d)$, one can have $\rho^d(c'R) = c'\rho^d(R)$,

which one can check directly. For any $R \in L^\infty(\mathcal{T}, \mathcal{F}_1, P^d)$, $\rho^d(R)$ is bounded. Combining with properties (A1) and (A2), one can show $\rho^d(\bullet)$ is continuous in the interior domain of the corresponding $L^\infty(\mathcal{T}, \mathcal{F}_1, P^d)$ with the essential norm (Ruszczyński and Shapiro, 2006b, Proposition 3.1). This ensures Fenchel-Moreau Theorem to hold ((Rockafellar, 1974)). Since (A1)-(A4) hold, Theorem 2.2 in (Ruszczyński and Shapiro, 2006b) gives that

$$\rho^d(R) = \inf_{\mu \in \mathcal{P}^d} \mathbf{E}_\mu[R], \tag{B.1}$$

where $\mathcal{P}^d$ is a subset of all probability measures on the measurable space $(\mathcal{T}, \mathcal{F}_1)$ such that any probability measure $\mu \in \mathcal{P}^d$ is absolutely continuous with respect to $P^d$ and $\frac{d\mu}{dP^d} \in L^1(\mathcal{T}, \mathcal{F}_1, P^d)$, and that

$$\mathcal{P}^d = \{\mu \,|\, \mathbf{E}_\mu[R] \leq \rho^d(R), \, \forall R \in L^\infty(\mathcal{T}, \mathcal{F}_1, P^d)\}. \tag{B.2}$$

Next we simplify $\mathcal{P}^d$. For any $F \in \mathcal{F}_1$ such that $P^d(F) = 0$, denote $R = \mathbb{I}_F(\bullet)$, i.e., 0-1 indicator of a set $F$, then $\rho^d(R) = 0$, which implies $\mathbf{E}_\mu[R] = 0$, i.e., $\mu(F) = 0$ for $\mu \in \mathcal{P}^d$. This also indicates that $\mu$ is absolutely continuous with respect with $P^d$ for $\mu \in \mathcal{P}^d$. Consider the Radon-Nikodym derivative $W = \frac{\mathrm{d}\mu}{\mathrm{d}P^d}$ with $W \in L_1(\mathcal{T}, \mathcal{F}_1, P^d)$. Note that $\mathbf{E}_\mu[R] = \mathbf{E}^d[RW]$ and we can have

$$\begin{aligned}
& \mathbf{E}_\mu[R] \leq \rho^d(R) \\
\Rightarrow \quad & \mathbf{E}^d[R(W - (1-\tau))] \leq \frac{\tau}{\gamma} \mathbf{E}^d[R\mathbb{I}(R \leq Q_\gamma(P^d))] \\
\Rightarrow \quad & \mathbf{E}^d[(R - Q_\gamma(P^d))(W - (1-\tau))] \leq \frac{\tau}{\gamma} \mathbf{E}^d[(R - Q_\gamma(P^d))\mathbb{I}(R \leq Q_\gamma(P^d))] \\
\Rightarrow \quad & \int_{\mathcal{T}} (R(\omega) - Q_\gamma(P^d))(W(\omega) - (1-\tau)) P^d(\mathrm{d}\omega) \\
& \leq \frac{\tau}{\gamma} \int_{\mathcal{T}} (R(\omega) - Q_\gamma(P^d)) \mathbb{I}(R(\omega) \leq Q_\gamma(P^d)) P^d(\mathrm{d}\omega),
\end{aligned}$$

for all $R \in L^\infty(\mathcal{T}, \mathcal{F}_1, P^d)$. Split $R(\omega)$ into three sets: $\mathbb{I}(R(\omega) < Q_\gamma(P^d))$, $\mathbb{I}(R(\omega) = Q_\gamma(P^d))$ and $\mathbb{I}(R(\omega) > Q_\gamma(P^d))$. In order to make the last inequality hold for all $R \in L^\infty(\mathcal{T}, \mathcal{F}_1, P^d)$, if $R(\omega) < Q_\gamma(P^d)$, then $W(\omega) - (1-\tau) \leq \frac{\tau}{\gamma}$, i.e., $W(\omega) \geq \varepsilon_2$. Similarly, when $R(\omega) > Q_\gamma(P^d)$, the right hand-side of the last inequality is 0 and thus $W(\omega) - (1-\tau) \geq 0$, i.e., $W(\omega) \leq \varepsilon_1$. For $R(\omega) = Q_\gamma(P^d)$, $W(\omega)$ must lie between $\varepsilon_1$ and $\varepsilon_2$ to make $\mathbf{E}^d[W] = 1$. Thus, define

$$\mathcal{W}_1^d := \{W \in L^1(\mathcal{T}, \mathcal{F}_1, P^d) \,|\, \mathbf{E}^d[W] = 1, \, \varepsilon_1 \leq W(\omega) \leq \varepsilon_2, \text{ for almost sure } \omega_1 \in \mathcal{T}\}, \tag{B.3}$$

and

$$\rho^d(R) = \inf_{W \in W_1^d} \mathbf{E}[WR]. \tag{B.4}$$

Next, we discuss the derivation of the duality representation of $M_2(d)$, which is an extension of $M_1(d)$. Based on Theorem 3.2.1, we first rewrite $M_2(d)$ as

$$\begin{aligned} M_2(d) &= (1-\tau)\mathbf{E}[\mathbf{E}[R|\mathbf{x}, A = d(\mathbf{x})]] + \tau\mathbf{E}[\mathbf{E}[R\mathbb{I}(R \leq Q_\gamma(R|\mathbf{x}, A = d(\mathbf{x})))|\mathbf{x}, A = d(\mathbf{x})]] \\ &= \mathbf{E}[\mathbf{E}[(1-\tau)R + \tau R\mathbb{I}(R \leq Q_\gamma(R|\mathbf{x}, A = d(\mathbf{x})))|\mathbf{x}, A = d(\mathbf{x})]] \\ &= \mathbf{E}[\rho^d(R|\mathbf{x}, A = d(\mathbf{x}))], \end{aligned}$$

where $\rho^d(R|\mathbf{x}, A = d(\mathbf{x})) := \mathbf{E}[(1-\tau)R + \tau R\mathbb{I}(R \leq Q_\gamma(R|\mathbf{x}, A = d(\mathbf{x})))|\mathbf{x}, A = d(\mathbf{x})]$. Consider the space $L^\infty(\mathcal{T}, \mathcal{F}_2, P^d)$, where $\mathcal{F}_2$ is induced by $\mathbf{x}$. Clearly $\mathcal{F}_2 \subseteq \mathcal{F}_1$ and $L^\infty(\mathcal{T}, \mathcal{F}_2, P^d) \subseteq L^\infty(\mathcal{T}, \mathcal{F}_1, P^d)$. The mapping $\rho^d(\bullet|\mathbf{x}, A = d(\mathbf{x}))$ is defined from $L^\infty(\mathcal{T}, \mathcal{F}_1, P^d)$ into $L^\infty(\mathcal{T}, \mathcal{F}_2, P^d)$. Such a mapping $\rho^d(\bullet|\mathbf{x}, A = d(\mathbf{x}))$ has properties (A1), (A2) and A(4). Property (A3) can be further strengthened to

(A3′) Translation invariance. If $R_1 \in L^\infty(\mathcal{T}, \mathcal{F}_1, P^d)$ and $R_2 \in L^\infty(\mathcal{T}, \mathcal{F}_2, P^d)$, then $\rho^d(R_1 + R_2|\mathbf{x}, A = d(\mathbf{x})) = \rho^d(R_1|\mathbf{x}, A = d(\mathbf{x})) + R_2$.

It can be checked that $\rho^d_\omega(\bullet|\mathbf{x}, A = d(\mathbf{x}))$ satisfies Properties (A1), (A2), (A3′) and (A4). For any $\omega \in \mathcal{T}$, define a real valued function $\rho^d(\bullet|\mathbf{x}, A = d(\mathbf{x}))(\omega)$ as $\rho^d_\omega(\bullet|\mathbf{x}, A = d(\mathbf{x}))$, which maps from $L^\infty(\mathcal{T}, \mathcal{F}_1, P^d)$ into $\mathbb{R}$. As a function of $\omega$, $\rho^d_\omega(\bullet|\mathbf{x}, A = d(\mathbf{x}))$ is measurable with respect to $\mathcal{F}_2$. Similar to the proof related to $M_1(d)$, define $\mathcal{P}^d_\omega$ be a family of conditional probability measures with respect to elements in $\mathcal{P}$ such that any conditional probability measure $\mu_\omega \in \mathcal{P}^d_\omega$ is absolutely continuous with respect to $P^d_\omega$, where $\mu_\omega(\bullet) = [P(\bullet|\mathcal{F}_2)](\omega)$, $P^d_\omega$ is conditional probability measure with respect to $P^d$. Denote the conditional density $W_\omega = \frac{\mathrm{d}\mu_\omega}{\mathrm{d}P^d_\omega} \in L^1(\mathcal{T}, \mathcal{F}_1, P^d)$. Proposition 3.1 and Theorem

4.1 in (Ruszczyński and Shapiro, 2006a) give that

$$\rho^d_\omega(R|\mathbf{x}, A = d(\mathbf{x})) = \inf_{\mu^d_\omega \in \mathcal{P}^d_\omega} \mathbf{E}_{\mu^d_\omega}[R]. \tag{B.5}$$

By the similar argument in the proof of $M_1(d)$, this term can be further expressed as,

$$\rho^d_\omega(R|\mathbf{x}, A = d(\mathbf{x})) = \inf_{W \in \mathcal{W}^d_2} \mathbf{E}[RW|\mathbf{x}, A = d(\mathbf{x})](\omega), \tag{B.6}$$

where

$$\begin{gathered}\mathcal{W}^d_2 = \{W \in L^1(\mathcal{T}, \mathcal{F}_1, P^d) \,|\, \varepsilon_1 \leq W(\omega_1) \leq \varepsilon_2 \\ \text{for almost sure } \ \omega_1 \in \mathcal{T},\ \mathbf{E}[W|\mathbf{x}, A = d(\mathbf{x})] = 1\}.\end{gathered} \tag{B.7}$$

See Example 6.2 in (Ruszczyński and Shapiro, 2006a). Therefore we can have

$$\rho^d(R) = \mathbf{E}\left[\inf_{W \in \mathcal{W}^d_2} \mathbf{E}[RW|\mathbf{x}, A = d(\mathbf{x})]\right]. \tag{B.8}$$

Furthermore, Proposition 5.1 in (Ruszczyński and Shapiro, 2006a) shows that interchange between infimum and expectation holds. This gives us that

$$\rho^d(R) = \inf_{W \in \mathcal{W}^d_2} \mathbf{E}^d[WR]. \tag{B.9}$$

### Proof of Theorem 3.4.1

For any measurable function $\alpha$, by noting that $T(u) + T(-u) = 2$, $\forall u$, then we can write $M_T(f, \alpha)$ as

$$\begin{aligned} M_T(f,\alpha) = &\, \mathbf{E}_\mathbf{x}[T(f(\mathbf{x}))\mathbf{E}[\alpha(\mathbf{x}) - \frac{1}{\gamma}(\alpha(\mathbf{x}) - R)_+|\mathbf{x}, A = 1] \\ &+ T(-f(\mathbf{x}))\mathbf{E}[\alpha(\mathbf{x}) - \frac{1}{\gamma}(\alpha(\mathbf{x}) - R)_+|\mathbf{x}, A = -1]] \\ &= \mathbf{E}_\mathbf{x}[T(f(\mathbf{x}))(\mathbf{E}[\alpha(\mathbf{x}) - \frac{1}{\gamma}(\alpha(\mathbf{x}) - R)_+|\mathbf{x}, A = 1] - \mathbf{E}[\alpha(\mathbf{x}) - \frac{1}{\gamma}(\alpha(\mathbf{x}) - R)_+|\mathbf{x}, A = -1])] \\ &+ 2\mathbf{E}_\mathbf{x}[\mathbf{E}[\alpha(\mathbf{x}) - \frac{1}{\gamma}(\alpha(\mathbf{x}) - R)_+|\mathbf{x}, A = -1]]. \end{aligned}$$

Define

$$\Delta = \mathbf{E}[\alpha(\mathbf{x}) - \frac{1}{\gamma}(\alpha(\mathbf{x}) - R)_+|\mathbf{x}, A = 1] - \mathbf{E}[\alpha(\mathbf{x}) - \frac{1}{\gamma}(\alpha(\mathbf{x}) - R)_+|\mathbf{x}, A = -1]$$

If $\Delta > 0$ for a given $\alpha$, in order to maximize $M_T(f, \alpha)$, we need $T(f(\mathbf{x})) = 2$, that is $f(\mathbf{x}) \geq \delta$. Similarly, if $\Delta < 0$, we need $f(\mathbf{x}) \leq -\delta$ to maximize $M_T(f, \alpha)$. Hence, we can get $|f_T^*(\mathbf{x})| \geq \delta$ for any $\mathbf{x} \in \mathcal{X}$, or equivalently,

$$\max_{\substack{\alpha \in \mathcal{L}_1(\mathbf{x}, \Xi, \mathbf{P}_\mathbf{x}), \\ ||f||_\infty \geq \delta}} M_T(f, \alpha) = \max_{f,\, \alpha \in \mathcal{L}_1(\mathbf{x}, \Xi, \mathbf{P}_\mathbf{x})} M_T(f, \alpha).$$

By the concavity of $-\frac{1}{\gamma}(\alpha(\mathbf{x}) - R)_+$, let $\alpha_1, \alpha_2$ satisfy

$$1 \in \mathbf{E}[\partial \frac{1}{\gamma}(\alpha_1(\mathbf{x}) - R)_+|\mathbf{x}, A = 1],$$

and

$$1 \in \mathbf{E}[\partial \frac{1}{\gamma}(\alpha_2(\mathbf{x}) - R)_+|\mathbf{x}, A = -1],$$

respectively. Then we can have

$$\begin{aligned}
\Delta^* &= \Delta_1 - \Delta_2 \\
&:= \mathbf{E}[\alpha_1(\mathbf{x}) - (\alpha_1(\mathbf{x}) - R)_+|\mathbf{x}, A = 1] - \mathbf{E}[\alpha_2(\mathbf{x}) - (\alpha(\mathbf{x})_2 - R)_+|\mathbf{x}, A = -1] \\
&= \mathrm{CVaR}_\gamma(R|X, A = 1) - \mathrm{CVaR}_\gamma(R|X, A = -1).
\end{aligned}$$

Furthermore, we have the following inequality

$$\begin{aligned}
\max_{\alpha \in \mathcal{F}, ||f||_\infty \geq \delta} M_T(f, \alpha) &\leq 2\mathbf{E}_\mathbf{x}[\max_{\alpha \in \mathcal{F}}\{\mathbf{E}[\alpha(\mathbf{x}) - \frac{1}{\gamma}(\alpha(\mathbf{x}) - R)_+|\mathbf{x}, A = 1], \\
&\qquad \mathbf{E}[\alpha(\mathbf{x}) - \frac{1}{\gamma}(\alpha(\mathbf{x}) - R)_+|\mathbf{x}, A = -1]\}] \\
&\leq 2\mathbf{E}_\mathbf{x}[\max\{\mathrm{CVaR}_\gamma(R|X, A = 1), \mathrm{CVaR}_\gamma(R|X, A = -1)\}] \\
&= 2\mathbf{E}_\mathbf{x}[\mathbb{I}(\Delta^* > 0)\Delta_1 + \mathbb{I}(\Delta^* < 0)\Delta_2] \\
&= M_T(\mathrm{sign}(\Delta^*), \mathbb{I}(\Delta^* > 0)\alpha_1(\mathbf{x}) + \mathbb{I}(\Delta^* < 0)\alpha_2(\mathbf{x}))
\end{aligned}$$

Thus, $\text{sign}(f_T^*) = \text{sign}(\Delta^*)$ and $\alpha_T^* = \mathbb{I}(\Delta^* > 0)\alpha_1(\mathbf{x}) + \mathbb{I}(\Delta^* < 0)\alpha_2(\mathbf{x})$. According to Theorem 2.1 and Proposition 2.3 in the main text, we have

$$d^*(\mathbf{x}) = \text{sign}(\Delta^*), \tag{B.10}$$

and

$$\alpha^*(\mathbf{x}) = \mathbb{I}(\Delta^* > 0)\alpha_1(\mathbf{x}) + \mathbb{I}(\Delta^* < 0)\alpha_2(\mathbf{x}). \tag{B.11}$$

This yields the desired result.

**Proof of Theorem 3.4.2**

Given $\mathbf{x} \in \mathcal{X}$, we first define

$$A_1 := \mathbf{E}[\alpha(\mathbf{x}) - \frac{1}{\gamma}(\alpha(\mathbf{x}) - R)_+|\mathbf{x}, A = 1],$$

and

$$A_2 := \mathbf{E}[\alpha(\mathbf{x}) - \frac{1}{\gamma}(\alpha(\mathbf{x}) - R)_+|\mathbf{x}, A = -1].$$

Then for any measurable functions $f, \alpha$, we have

$$\begin{aligned}
&\mathbf{E}[\frac{S(Af_T^*(\mathbf{x}))}{\pi(A|\mathbf{x})}(\alpha_T^*(\mathbf{x}) - \frac{1}{\gamma}(\alpha_T^*(\mathbf{x}) - R)_+)|\mathbf{x}] - \mathbf{E}[\frac{S(Af(\mathbf{x}))}{\pi(A|\mathbf{x})}(\alpha(\mathbf{x}) - \frac{1}{\gamma}(\alpha(\mathbf{x}) - R)_+)|\mathbf{x}] \\
=&S(f_T^*(\mathbf{x}))\text{CVaR}_\gamma(R|X, A = 1) + S(-f_T^*(\mathbf{x}))\text{CVaR}_\gamma(R|X, A = -1) \\
&-(S(f(\mathbf{x}))\mathbf{E}[\alpha(\mathbf{x}) - \frac{1}{\gamma}(\alpha(\mathbf{x}) - R)_+|\mathbf{x}, A = 1] \\
&+S(-f(\mathbf{x}))\mathbf{E}[\alpha(\mathbf{x}) - \frac{1}{\gamma}(\alpha(\mathbf{x}) - R)_+|\mathbf{x}, A = -1]) \\
=&S(f_T^*(\mathbf{x}))\Delta_1 + S(-f_T^*(\mathbf{x}))\Delta_2 - S(f(\mathbf{x}))A_1 - S(-f(\mathbf{x}))A_2.
\end{aligned}$$

Similarly,

$$\begin{aligned}
&\mathbf{E}[\frac{\mathbb{I}(d^*(\mathbf{x}) = A)}{\pi(A|\mathbf{x})}(\alpha^*(\mathbf{x}) - \frac{1}{\gamma}(\alpha^*(\mathbf{x}) - R)_+)|\mathbf{x}] - \mathbf{E}[\frac{\mathbb{I}(\text{sign}(f(\mathbf{x})) = A)}{\pi(A|\mathbf{x})}(\alpha(\mathbf{x}) - \frac{1}{\gamma}(\alpha(\mathbf{x}) - R)_+)|\mathbf{x}] \\
=&\mathbb{I}(d^*(\mathbf{x}) = 1)\text{CVaR}_\gamma(R|X, A = 1) + \mathbb{I}(d^*(\mathbf{x}) = -1)\text{CVaR}_\gamma(R|X, A = -1) \\
&-(\mathbb{I}(\text{sign}(f(\mathbf{x})) = 1)\mathbf{E}[\alpha(\mathbf{x}) - \frac{1}{\gamma}(\alpha(\mathbf{x}) - R)_+|\mathbf{x}, A = 1] \\
&+\mathbb{I}(\text{sign}(f(\mathbf{x})) = -1)\mathbf{E}[\alpha(\mathbf{x}) - \frac{1}{\gamma}(\alpha(\mathbf{x}) - R)_+|\mathbf{x}, A = -1]) \\
=&\mathbb{I}(d^*(\mathbf{x}) = 1)\Delta_1 + \mathbb{I}(d^*(\mathbf{x}) = -1)\Delta_2 - \mathbb{I}(\text{sign}(f(\mathbf{x})) = 1)A_1 - \mathbb{I}(\text{sign}(f(\mathbf{x})) = -1)A_2
\end{aligned}$$

When $\Delta_1 \geq \Delta_2$, we have $f_T^* \geq \delta$ and $d^*(\mathbf{x}) = 1$, then we have

$$\begin{aligned}
&S(f_T^*(\mathbf{x}))\Delta_1 + S(-f_T^*(\mathbf{x}))\Delta_2 - S(f(\mathbf{x}))A_1 - S(-f(\mathbf{x}))A_2 \\
=&2\Delta_1 - S(f(\mathbf{x}))A_1 - S(-f(\mathbf{x}))A_2 \\
=&2(\Delta_1 - A_1) + S(-f(\mathbf{x}))(A_1 - A_2),
\end{aligned}$$

and

$$\begin{aligned}
&\mathbb{I}(d^*(\mathbf{x}) = 1)\Delta_1 + \mathbb{I}(d^*(\mathbf{x}) = -1)\Delta_2 - \mathbb{I}(\text{sign}(f(\mathbf{x})) = 1)A_1 - \mathbb{I}(\text{sign}(f(\mathbf{x})) = -1)A_2 \\
=&\Delta_1 - \mathbb{I}(\text{sign}(f(\mathbf{x})) = 1)A_1 - \mathbb{I}(\text{sign}(f(\mathbf{x})) = -1)A_2 \\
=&(\Delta_1 - A_1) + \mathbb{I}(\text{sign}(f(\mathbf{x})) = -1)(A_1 - A_2).
\end{aligned}$$

Note that, for any measurable function $f$, $0 \leq T(-f(\mathbf{x})) - \mathbb{I}(\text{sign}(f(\mathbf{x})) = -1) \leq 1$. Thus, we have

$$\begin{aligned}
&2(\Delta_1 - A_1) + S(-f(\mathbf{x}))(A_1 - A_2) - \{(\Delta_1 - A_1) + \mathbb{I}(\text{sign}(f(\mathbf{x})) = -1)(A_1 - A_2)\} \\
=&(\Delta_1 - A_1) + (S(-f(\mathbf{x})) - \mathbb{I}(\text{sign}(f(\mathbf{x})) = -1))(A_1 - A_2) \\
\geq&\min\{\Delta_1 - A_1 + 0, \Delta_1 - A_1 + A_1 - A_2\} \geq \min\{\Delta_1 - A_1, \Delta_1 - A_2\} \\
\geq&0,
\end{aligned}$$

where the last inequality holds because $\Delta_1 \geq A_1$ and $\Delta_1 \geq \Delta_2 \geq A_2$.

When $\Delta_1 < \Delta_2$, we have $f_T^* \leq -\delta$ and $d^*(\mathbf{x}) = -1$, then we have

$$\begin{aligned}
&S(f_T^*(\mathbf{x}))\Delta_1 + S(-f_T^*(\mathbf{x}))\Delta_2 - S(f(\mathbf{x}))A_1 - S(-f(\mathbf{x}))A_2 \\
=&2\Delta_2 - S(f(\mathbf{x}))A_1 - S(-f(\mathbf{x}))A_2 \\
=&2(\Delta_2 - A_2) + S(f(\mathbf{x}))(A_2 - A_1),
\end{aligned}$$

and

$$
\begin{aligned}
&\mathbb{I}(d^*(\mathbf{x})=1)\Delta_1+\mathbb{I}(d^*(\mathbf{x})=-1)\Delta_2-\mathbb{I}(\mathrm{sign}(f(\mathbf{x}))=1)A_1-\mathbb{I}(\mathrm{sign}(f(\mathbf{x}))=-1)A_2\\
=&\Delta_2-\mathbb{I}(\mathrm{sign}(f(\mathbf{x}))=1)A_1-\mathbb{I}(\mathrm{sign}(f(\mathbf{x}))=-1)A_2\\
=&(\Delta_2-A_2)+\mathbb{I}(\mathrm{sign}(f(\mathbf{x}))=1)(A_2-A_2).
\end{aligned}
$$

Note that, for any measurable function $f$, $0\le S(f(\mathbf{x}))-\mathbb{I}(\mathrm{sign}(f(\mathbf{x}))=1)\le 1$. Thus, we have

$$
\begin{aligned}
&2(\Delta_2-A_2)+S(f(\mathbf{x}))(A_2-A_1)-(\Delta_2-A_2)+\mathbb{I}(\mathrm{sign}(f(\mathbf{x}))=1)(A_2-A_2)\\
=&(\Delta_2-A_2)+(S(f(\mathbf{x}))-\mathbb{I}(\mathrm{sign}(f(\mathbf{x}))=1))(A_2-A_1)\\
\ge&\min\{\Delta_2-A_2+0,\Delta_2-A_2+A_2-A_1\}\ge\min\{\Delta_2-A_2,\Delta_2-A_1\}\\
\ge&0,
\end{aligned}
$$

where the last inequality holds because $\Delta_2\ge A_2$ and $\Delta_2\ge\Delta_1\ge A_1$.

Therefore, for both cases, we have

$$
\begin{aligned}
&\mathbf{E}[\frac{S(Af_T^*(\mathbf{x}))}{\pi(A|\mathbf{x})}(\alpha_T^*(\mathbf{x})-\frac{1}{\gamma}(\alpha_T^*(\mathbf{x})-R)_+)|\mathbf{x}]-\mathbf{E}[\frac{S(Af(\mathbf{x}))}{\pi(A|\mathbf{x})}(\alpha(\mathbf{x})-\frac{1}{\gamma}(\alpha(\mathbf{x})-R)_+|\mathbf{x}]\\
\le&\mathbf{E}[\frac{\mathbb{I}(d^*(\mathbf{x})=A)}{\pi(A|\mathbf{x})}(\alpha_1(\mathbf{x})-\frac{1}{\gamma}(\alpha^*(\mathbf{x})-R)_+|\mathbf{x}]\\
-&\mathbf{E}[\frac{\mathbb{I}(\mathrm{sign}(f(\mathbf{x}))=A)}{\pi(A|\mathbf{x})}(\alpha(\mathbf{x})-\frac{1}{\gamma}(\alpha(\mathbf{x})-R)_+|\mathbf{x}]
\end{aligned}
$$

The desired result holds by taking expectations over $\mathbf{x}$ on both sides.

### Proof of Lemma 3.4.1

When $p=1$, it holds automatically by Corollary 3.17 in (Ledoux and Talagrand, 2013). Without loss of generality, it is enough to show the case when $p=2$ and $L_\phi=1$. Consider a class of vector value functions

$$
\Psi=\{\psi=(\psi_1,\cdots,\psi_n),\text{where }\psi_i\text{ is }\phi_i\text{ or }\varphi\},\tag{B.12}
$$

where $\varphi$ is defined as $\varphi(t,s) = t + s$. Then by monotonicity of $\mathcal{R}_n$, we have

$$\mathcal{R}_n(\phi(\mathcal{F}_1, \mathcal{F}_2)) \leq \sup_{\psi \in \Psi} \mathcal{R}_n(\psi(\mathcal{F}_1, \mathcal{F}_2)). \tag{B.13}$$

Suppose there exists at least one $\phi_i$ in $\Psi$, with loss of generality, say $\psi_1 = \phi_1$ and define

$$\begin{aligned} \psi &= (\phi_1, \psi_2, \cdots, \psi_n) \\ \psi' &= (\varphi, \psi_2, \cdots, \psi_n). \end{aligned} \tag{B.14}$$

Then we have

$$\begin{aligned}
&\mathcal{R}_n(\psi(\mathcal{F}_1, \mathcal{F}_2)) \\
=&\mathbf{E}[\sup_{f_i \in \mathcal{F}_1, g_i \in \mathcal{F}_2} \frac{1}{n} \sum_{i=1}^n \sigma_i \psi(f_i, g_i)] \\
=&\frac{1}{2n}\mathbf{E}[\sup_{f_i \in \mathcal{F}_1, g_i \in \mathcal{F}_2} (\phi(f_1, g_1) + \sum_{i=2}^n \psi(f_i, g_i)) + \sup_{f_i \in \mathcal{F}_1, g_i \in \mathcal{F}_2} (-\phi(f_1, g_1) + \sum_{i=2}^n \psi(f_i, g_i))] \\
\leq&\frac{1}{2n}\mathbf{E}[\sup_{f_i, f_i' \in \mathcal{F}_1, g_i, g_i' \in \mathcal{F}_2} (\phi(f_1, g_1) - \phi(f_1', g_1') + \sum_{i=2}^n \psi(f_i, g_i) + \sum_{i=2}^n \psi(f_i', g_i'))] \\
\leq&\frac{1}{2n}\mathbf{E}[\sup_{f_i, f_i' \in \mathcal{F}_1, g_i, g_i' \in \mathcal{F}_2} (|f_1 - f_1'| + |g_1 - g_1'| + \sum_{i=2}^n \psi(f_i, g_i) + \sum_{i=2}^n \psi(f_i', g_i'))] \\
=&\frac{1}{2n}\mathbf{E}[\sup_{f_i, f_i' \in \mathcal{F}_1, g_i, g_i' \in \mathcal{F}_2} (f_1 - f_1' + g_1 - g_1' + \sum_{i=2}^n \psi(f_i, g_i) + \sum_{i=2}^n \psi(f_i', g_i'))] \\
=&\frac{1}{2n}\mathbf{E}[\sup_{f_i \in \mathcal{F}_1, g_i \in \mathcal{F}_2} (f_1 + g_1 + \sum_{i=2}^n \psi(f_i, g_i)) + \sup_{f_i \in \mathcal{F}_1, g_i \in \mathcal{F}_2} (-(f_1 + g_1) + \sum_{i=2}^n \psi(f_i, g_i))] \\
=&\mathbf{E}[\sup_{f_i \in \mathcal{F}_1, g_i \in \mathcal{F}_2} \frac{1}{n} \sum_{i=1}^n \sigma_i \psi'(f_i, g_i)] \\
=&\mathcal{R}_n(\psi'(\mathcal{F}_1, \mathcal{F}_2)).
\end{aligned} \tag{B.15}$$

Thus $\sup_{\psi\in\Psi} \mathcal{R}_n(\phi(\mathcal{F}_1, \mathcal{F}_2)) = \mathcal{R}_n(\varphi(\mathcal{F}_1, \mathcal{F}_2))$. A quick observation shows that

$$\begin{aligned}\mathcal{R}_n(\varphi(\mathcal{F}_1, \mathcal{F}_2)) &= \mathbf{E}[\sup_{f_i\in\mathcal{F}_1, g_i\in\mathcal{F}_2} \frac{1}{n}\sum_{i=1}^{n} \sigma_i(f_i + g_i)] \\ &= \mathbf{E}[\sup_{f_i\in\mathcal{F}_1} \frac{1}{n}\sum_{i=1}^{n} \sigma_i f_i + \sup_{g_i\in\mathcal{F}_2} \frac{1}{n}\sum_{i=1}^{n} \sigma_i g_i] \\ &= \mathcal{R}_n(\mathcal{F}_1) + \mathcal{R}_n(\mathcal{F}_2).\end{aligned} \tag{B.16}$$

**Proof of Theorem 3.4.3**

According to Theorem 3.4.2, we can have

$$\begin{aligned} & M_0(d^*, \alpha^*) - M_0(\mathrm{sign}(\hat{f}), \hat{\alpha}) \\ \le & M_T(f_T^*, \alpha_T^*) - M_T(\hat{f}, \hat{\alpha}) \\ = & O_T(\hat{f}, \hat{\alpha}) - O_T(f_T^*, \alpha_T^*) \\ \le & O_T(\hat{f}, \hat{\alpha}) - O_n(\hat{f}, \hat{\alpha}) + O_n(\hat{f}, \hat{\alpha}) + \frac{\lambda_{1n}}{2}||\hat{f}||_{\mathcal{H}_1}^2 + \frac{\lambda_{2n}}{2}||\hat{\alpha}||_{\mathcal{H}_2}^2 \\ & -(O_n(f_{\lambda_{1n}}, \alpha_{\lambda_{2n}}) + \frac{\lambda_{1n}}{2}||f_{\lambda_{1n}}||_{\mathcal{H}_1}^2 + \frac{\lambda_{2n}}{2}||\alpha_{\lambda_{2n}}||_{\mathcal{H}_2}^2) \\ & +O_n(f_{\lambda_{1n}}, \alpha_{\lambda_{2n}}) - O_T(f_{\lambda_{1n}}, \alpha_{\lambda_{2n}}) + \mathcal{A}(\lambda_{1n}, \lambda_{2n}) \\ \le & [O_T(\hat{f}, \hat{\alpha}) - O_n(\hat{f}, \hat{\alpha})] + [O_n(f_{\lambda_{1n}}, \alpha_{\lambda_{2n}}) - O_T(f_{\lambda_{1n}}, \alpha_{\lambda_{2n}})] + \mathcal{A}(\lambda_{1n}, \lambda_{2n}) \\ = & (I) + (II) + \mathcal{A}(\lambda_{1n}, \lambda_{2n}),\end{aligned}$$

where the first two terms $(I)$ and $(II)$ are estimation errors. The last inequality is due to the definition of $(\hat{f}, \hat{\alpha})$ as the minimizer of $O_n(f, \alpha)$.

In order to bound the estimation error, we first obtain the bounds for $||\hat{f}||_{\mathcal{H}_1}, ||f_{\lambda_{1n}}||_{\mathcal{H}_1}$ and $||\hat{\alpha}||_{\mathcal{H}_2}, ||\alpha_{\lambda_{2n}}||_{\mathcal{H}_2}$ correspondingly. By the definition of $(\hat{f}, \hat{\alpha})$, we have

$$O_n(\hat{f}, \hat{\alpha}) + \frac{\lambda_{1n}}{2}||\hat{f}||_{\mathcal{H}_1}^2 + \frac{\lambda_{2n}}{2}||\hat{\alpha}||_{\mathcal{H}_2}^2 \le O_n(0, 0),$$

where $O_n(0,0) = \frac{1}{n}\sum_{i=1}^n(\frac{1}{\gamma}(-R_i)_+) \leq M_1$, because $R$ is bounded by $C_0$ by assumption. Moreover, since $\alpha - R \leq \frac{1}{\gamma}(\alpha - R)_+$, we have

$$M_1 \geq O_n(0,0) \geq O_n(\hat{f}, \hat{\alpha}) \geq \frac{1}{n}\sum_{i=1}^n \frac{S(A_i f(\mathbf{x}_i))}{\pi(A_i|\mathbf{x}_i)}(-R_i) \geq M_2, \qquad \text{(B.17)}$$

where the last inequality holds by the bounded assumption of $R$. Thus $\frac{\lambda_{1n}}{2}||\hat{f}||_{\mathcal{H}_1}^2 + \frac{\lambda_{2n}}{2}||\hat{\alpha}||_{\mathcal{H}_2}^2 \leq M_1 - M_2 := M_3$, where $M_3 \geq 0$, and $|O_n(\hat{f}, \hat{\alpha})| \leq \max\{M_1, |M_2|\} := M_4$ or equivalently $L_S(\hat{f}, \hat{\alpha}) := \frac{S(A\hat{f}(\mathbf{x}))}{\pi(A|\mathbf{x})}(\frac{1}{\gamma}(\hat{\alpha}(\mathbf{x}) - R)_+ - \hat{\alpha}(\mathbf{x}))$ is bounded by $M_4$. By a similar argument, we can also obtain $\frac{\lambda_{1n}}{2}||f_{\lambda_{1n}}||_{\mathcal{H}_1}^2 + \frac{\lambda_{2n}}{2}||\alpha_{\lambda_{2n}}||_{\mathcal{H}_2}^2 \leq M_3$ and $|L_S(f_{\lambda_{1n}}, \alpha_{\lambda_{2n}})| \leq M_4$.

Define the following functional class

$$\Xi := \{L_S(f, \alpha) \mid f \in \mathcal{H}_1, \alpha \in \mathcal{H}_2, \frac{\lambda_{1n}}{2}||f||_{\mathcal{H}_1}^2 + \frac{\lambda_{2n}}{2}||\alpha||_{\mathcal{H}_2}^2 \leq M_3, |L_S(f_{\lambda_{1n}}, \alpha_{\lambda_{2n}})| \leq M_4\}.$$

Let $\{Z_i\}_{i=1}^n = \{\mathbf{x}_i, A_i, R_i\}_{i=1}^n$ and $P_n$ be the corresponding empirical measure on $Z_n$. We first derive the bound for the estimation error $(I)$ and $(II)$. For the term $(I)$, note that $(I) \leq \sup_\Xi PL_S(f, \alpha) - P_n L_S(f, \alpha)$, where $P$ is probability measure of $(\mathbf{x}, A, R)$. When any $(\mathbf{x}_i, A_i, R_i)$ changes, by the definition of $\Xi$, $\sup_\Xi PL_S(f, \alpha) - P_n L_S(f, \alpha)$ is changed no more than $\frac{M_4}{n}$. Then by the McDiarmid's inequality, with probability at least $1 - \frac{\epsilon}{2}$, we can get

$$\sup_\Xi PL_S(f, \alpha) - P_n L_S(f, \alpha) \leq \mathbf{E}[\sup_\Xi PL_S(f, \alpha) - P_n L_S(f, \alpha)] + \sqrt{\frac{2\log(\frac{1}{\epsilon})}{n}}. \qquad \text{(B.18)}$$

Using the idea of symmetrization by introducing a duplicated data $\{Z_i'\}_{i=1}^n$ and Rademacher variables $\{\sigma_i\}_{i=1}^n$, we can obtain

$$\begin{aligned}
\mathbf{E}[\sup_{\Xi} PL_S(f,\alpha) - P_nL_S(f,\alpha)] &\leq \mathbf{E}[\sup_{\Xi} \mathbf{E}[P_n'L_S(f,\alpha) - P_nL_S(f,\alpha)]] \\
&\leq \mathbf{E}[\sup_{\Xi} P_n'L_S(f,\alpha) - P_nL_S(f,\alpha)] \\
&= \mathbf{E}[\sup_{\Xi} P_n\sigma(L_S(f,\alpha) - L_S'(f,\alpha)] \\
&\leq \mathbf{E}[\sup_{\Xi} P_n\sigma L_S(f,\alpha)] + \mathbf{E}[\sup_{\Xi} -P_n\sigma L_S(f,\alpha)] \\
&= 2\mathbf{E}[\sup_{\Xi} P_n\sigma L_S(f,\alpha)] \\
&= 2\mathcal{R}_n(\Xi).
\end{aligned}$$

For the term $(II)$, by the similar argument, we can show, with probability at least $1-\frac{\epsilon}{2}$,

$$\begin{aligned}
(II) &\leq \sup_{\Xi} P_nL_S(f,\alpha) - PL_S(f,\alpha) \\
&\leq \mathbf{E}[\sup_{\Xi} P_nL_S(f,\alpha) - PL_S(f,\alpha)] + \sqrt{\frac{2\log(\frac{1}{\epsilon})}{n}} \qquad \text{(B.19)} \\
&\leq 2\mathcal{R}_n(\Xi) + \sqrt{\frac{2\log(\frac{1}{\epsilon})}{n}}.
\end{aligned}$$

Then combining bounds of $(I)$ and $(II)$ together gives that, with probability at least $1-\epsilon$, we can have

$$(I) + (II) \leq 4\mathcal{R}_n(\Xi) + \sqrt{\frac{8\log(\frac{1}{\epsilon})}{n}}. \qquad \text{(B.20)}$$

Define a class of functions as

$$\Pi := \{(f,\alpha) \mid f \in \mathcal{H}_1, \alpha \in \mathcal{H}_2, \frac{\lambda_{1n}}{2}||f||_{\mathcal{H}_1}^2 \leq M_3, \frac{\lambda_{2n}}{2}||\alpha||_{\mathcal{H}_2}^2 \leq M_3\}.$$

In order to apply Lemma 3.4.1, we need to show $L_S(t,s)$ is Lipschitz continuous. For any constant $t_1, t_2, s_1, s_2$, we have

$$\begin{aligned}|L_S(t_1,s_1) - L_S(t_2,s_2)| &\leq |L_S(t_1,s_1) - L_S(t_2,s_1)| + |L_S(t_2,s_2) - L_S(t_2,s_1)| \\ &\leq \frac{C_0}{a_0}|t_1 - t_2| + \frac{2(1-\gamma)}{a_0\gamma}|s_1 - s_2| \\ &\leq M_5(|t_1 - t_2| + |s_1 - s_2|),\end{aligned} \tag{B.21}$$

where $M_5 = \max\{\frac{2(1-\gamma)}{a_0\gamma}, \frac{C_0}{a_0}\}$. Then by Lemma 4.1, we have

$$\mathcal{R}_n(\Xi) \leq \mathcal{R}_n(L_S(\Pi)) \leq M_5(\mathcal{R}_n(\Pi_1) + \mathcal{R}_n(\Pi_2)), \tag{B.22}$$

where $\Pi_1 = \{f|\ f \in \mathcal{H}_1, \frac{\lambda_{1n}}{2}||f||^2_{\mathcal{H}_1} \leq M_3\}$ and $\Pi_2 = \{\alpha|\ \alpha \in \mathcal{H}_2, \frac{\lambda_{2n}}{2}||\alpha||^2_{\mathcal{H}_2} \leq M_3\}$.

Thus, combining together, with probability $1-\epsilon$,

$$\begin{aligned}&M_0(d^*, \alpha^*) - M_0(\text{sign}(\hat{f}), \hat{\alpha}) \\ \leq &4M_5(\mathcal{R}_n(\Pi_1) + \mathcal{R}_n(\Pi_2)) + \sqrt{\frac{8\log(\frac{1}{\epsilon})}{n}} + \mathcal{A}(\lambda_{1n}, \lambda_{2n}).\end{aligned}$$

### Proof of Corollary 3.4.1

Based on the definition of $\Pi_1$ and $\Pi_2$, by Lemma B.0.1, we have

$$\begin{aligned}\mathcal{R}_n(\Pi_1) &\leq C_1\sqrt{\frac{2M_3}{n\lambda_{1n}}}, \\ \mathcal{R}_n(\Pi_2) &\leq C_1\sqrt{\frac{2M_3}{n\lambda_{2n}}}.\end{aligned}$$

Since $f_T^* = \mathbf{x}^T w^* + b_1^*$ and $\alpha_T^* = \mathbf{x}^T\theta^* + b_2^*$, we can obtain

$$\begin{aligned}\mathcal{A}(\lambda_{1n}, \lambda_{2n}) &\leq O_T(f_T^*, \alpha_T^*) + \frac{\lambda_{1n}}{2}||w^*||_2^2 + \frac{\lambda_{2n}}{2}||\theta^*||_2^2 - O_T(f_T^*, \alpha_T^*) \\ &\leq \frac{D_1\lambda_{1n}}{2} + \frac{D_2\lambda_{2n}}{2}.\end{aligned} \tag{B.23}$$

According to Theorem 3.4.3, with probability $1-\epsilon$,

$$\begin{aligned} & M_0(d^*, \alpha^*) - M_0(\text{sign}(\hat{f}), \hat{\alpha}) \\ \leq & 4M_5(\mathcal{R}_n(\Pi_1) + \mathcal{R}_n(\Pi_2)) + \sqrt{\frac{8\log(\frac{1}{\epsilon})}{n}} + \mathcal{A}(\lambda_{1n}, \lambda_{2n}) \\ \leq & 4M_5(C_1\sqrt{\frac{2M_3}{n\lambda_{1n}}} + C_1\sqrt{\frac{2M_3}{n\lambda_{2n}}}) + \frac{D_1\lambda_{1n}}{2} + \frac{D_2\lambda_{2n}}{2}. \end{aligned}$$

Then optimizing the right hand side with respect to $\lambda_{1n}$ and $\lambda_{2n}$, we can let $\lambda_{in} = \mathbf{O}(n^{-\frac{1}{3}})$ for $i = 1, 2$ and obtain the final result that with probability $1-\epsilon$,

$$M_0(d^*, \alpha^*) - M_0(\text{sign}(\hat{f}), \hat{\alpha}) \leq c_1 n^{-\frac{1}{3}},$$

for some constant $c_1$.

### Proof of Corollary 3.4.2

Based on the definition of $\Pi_1$ and $\Pi_2$, by Lemma B.0.1, we have

$$\begin{aligned} \mathcal{R}_n(\Pi_1) \leq C_2\sqrt{\frac{4M_3\log(2p)}{n\lambda_{1n}}} \\ \mathcal{R}_n(\Pi_2) \leq C_2\sqrt{\frac{4M_3\log(2p)}{n\lambda_{2n}}}. \end{aligned}$$

Since $f_T^* = \mathbf{x}^T w^* + b_1^*$ and $\alpha_T^* = \mathbf{x}^T\theta^* + b_2^*$, we can obtain

$$\begin{aligned} \mathcal{A}(\lambda_{1n}, \lambda_{2n}) &\leq O_T(f_T^*, \alpha_T^*) + \frac{\lambda_{1n}}{2}||w^*||_1 + \frac{\lambda_{2n}}{2}||\theta||^1 - O_T(f_T^*, \alpha_T^*) \\ &\leq \frac{D_3\lambda_{1n}}{2} + \frac{D_4\lambda_{2n}}{2}. \end{aligned} \tag{B.24}$$

According to Theorem 4.3, with probability $1-\epsilon$,

$$\begin{aligned}
&M_0(d^*, \alpha^*) - M_0(\text{sign}(\hat{f}), \hat{\alpha}) \\
\leq & 4M_5(\mathcal{R}_n(\Pi_1) + \mathcal{R}_n(\Pi_2)) + \sqrt{\frac{8\log(\frac{1}{\epsilon})}{n}} + \mathcal{A}(\lambda_{1n}, \lambda_{2n}) \\
\leq & 4M_5(C_2\sqrt{\frac{4M_3\log(2p)}{n\lambda_{1n}}} + C_2\sqrt{\frac{4M_3\log(2p)}{n\lambda_{2n}}}) + \frac{D_3\lambda_{1n}}{2} + \frac{D_4\lambda_{2n}}{2}.
\end{aligned}$$

Then optimizing right hand side with respect to $\lambda_{1n}$ and $\lambda_{2n}$, we can let $\lambda_{in} = \mathbf{O}((\frac{\log(2p)}{n})^{\frac{1}{3}})$ for $i = 1, 2$ and obtain the final result that with probability $1-\epsilon$,

$$M_0(d^*, \alpha^*) - M_0(\text{sign}(\hat{f}), \hat{\alpha}) \leq c_2(\frac{\log(2p)}{n})^{\frac{1}{3}},$$

for some constant $c_1$.

**Proof of Corollary 3.4.3** Based on the assumptions and the definition of $\Pi_1$ and $\Pi_2$, by Lemma 22 in (Bartlett and Mendelson, 2002), we have

$$\begin{aligned}
\mathcal{R}_n(\Pi_1) &\leq \sqrt{\frac{2M_0}{n\lambda_{1n}}} \\
\mathcal{R}_n(\Pi_2) &\leq \sqrt{\frac{2M_0}{n\lambda_{2n}}}
\end{aligned}$$

According to Theorem 4.3 and assumptions on approximation error, with probability $1-\epsilon$,

$$\begin{aligned}
&M_0(d^*, \alpha^*) - M_0(\text{sign}(\hat{f}), \hat{\alpha}) \\
\leq & 4M_5(\mathcal{R}_n(\Pi_1) + \mathcal{R}_n(\Pi_2)) + \sqrt{\frac{8\log(\frac{1}{\epsilon})}{n}} + \mathcal{A}(\lambda_{1n}, \lambda_{2n}) \\
\leq & 4M_5(\sqrt{\frac{2M_0}{n\lambda_{1n}}} + \sqrt{\frac{2M_0}{n\lambda_{2n}}}) + C_5\lambda_{1n}^{w_1} + C_6\lambda_{2n}^{w_2}.
\end{aligned}$$

Then optimizing right hand side with respect to $\lambda_{1n}$ and $\lambda_{2n}$, we can let $\lambda_{in} = \mathbf{O}((n^{-\frac{1}{2w_i+1}})$ for $i = 1, 2$ and obtain the final result that with probability at least $1 - \epsilon$,

$$\begin{aligned} M_0(d^*, \alpha^*) - M_0(\text{sign}(\hat{f}), \hat{\alpha}) &\leq c_3^{(1)} n^{-\frac{w_1}{2w_1+1}} + c_3^{(2)} n^{-\frac{w_2}{2w_2+1}} \\ &\leq \max(c_3^{(1)}, c_3^{(2)}) \max\left(n^{-\frac{w_1}{2w_1+1}}, n^{-\frac{w_2}{2w_2+1}}\right) \end{aligned}$$

for some constant $c_3^{(1)}$ and $c_3^{(2)}$.

The following two lemmas are the standard results of empirical process. Thus we state them without proofs.

**Lemma B.0.1.** *Let $\mathcal{F} = \{\mathbf{x}^T\theta \mid ||\theta||_2 \leq W_1\}$ be the class of linear functions and suppose $\mathbf{E}[||\mathbf{x}||_2^2] \leq C_1^2$, then*

$$\mathcal{R}_n(\mathcal{F}) \leq \frac{W_1 C_1}{\sqrt{n}}$$

**Lemma B.0.2.** *Let $\mathcal{F} = \{\mathbf{x}^T\theta \mid \theta \in \mathcal{R}^p, ||\theta||_1 \leq W_2\}$ be the class of linear functions and suppose $||\mathbf{x}||_\infty \leq C_2, a.s.$, then*

$$\mathcal{R}_n(\mathcal{F}) \leq \frac{W_2 C_2 \sqrt{2\log(2p)}}{\sqrt{n}}$$

# BIBLIOGRAPHY

R. T. afellar and S. Uryasev. Conditional value-at-risk for general loss distributions. *Journal of Banking & Finance*, 26(7):1443–1471, 2002.

M. Ahn, J.-S. Pang, and J. Xin. Difference-of-convex learning: directional stationarity, optimality, and sparsity. *SIAM Journal on Optimization*, 2017.

P. Artzner, F. Delbaen, J.-M. Eber, and D. Heath. Coherent measures of risk. *Mathematical finance*, 9(3):203–228, 1999.

J.-Y. Audibert, A. B. Tsybakov, et al. Fast learning rates for plug-in classifiers. *The Annals of Statistics*, 35(2):608–633, 2007.

F. R. Bach, G. R. Lanckriet, and M. I. Jordan. Multiple kernel learning, conic duality, and the smo algorithm. In *Proceedings of the twenty-first international conference on Machine learning*, page 6. ACM, 2004.

X. Bai, A. A. Tsiatis, W. Lu, and R. Song. Optimal treatment regimes for survival endpoints using a locally-efficient doubly-robust estimator from a classification perspective. *Lifetime Data Analysis*, pages 1–20, 2016.

G. Baron, E. Perrodeau, I. Boutron, and P. Ravaud. Reporting of analyses from randomized controlled trials with multiple arms: a systematic review. *BMC medicine*, 11(1):84, 2013.

P. L. Bartlett and S. Mendelson. Rademacher and gaussian complexities: Risk bounds and structural results. *Journal of Machine Learning Research*, 3(Nov):463–482, 2002.

A. Beck and M. Teboulle. A fast iterative shrinkage-thresholding algorithm for linear inverse problems. *SIAM Journal on Imaging Sciences*, 2(1):183–202, 2009.

A. Ben-Tal and M. Teboulle. Expected utility, penalty functions, and duality in stochastic nonlinear programming. *Management Science*, 32(11):1445–1466, 1986.

A. Ben-Tal and M. Teboulle. Penalty functions and duality in stochastic programming via $\varphi$-divergence functionals. *Mathematics of Operations Research*, 12(2):224–240, 1987.

A. Ben-Tal and M. Teboulle. An old-new concept of convex risk measures: The optimized certainty equivalent. *Mathematical Finance*, 17(3):449–476, 2007.

P. J. Bickel, Y. Ritov, and A. B. Tsybakov. Simultaneous analysis of lasso and dantzig selector. *The Annals of Statistics*, pages 1705–1732, 2009.

L. Bissonnette and M. G. Bergeron. Infectious disease management through point-of-care personalized medicine molecular diagnostic technologies. *Journal of Personalized Medicine*, 2 (2):50–70, 2012.

P. Breheny and J. Huang. Group descent algorithms for nonconvex penalized linear and logistic regression models with grouped predictors. *Statistics and Computing*, 25(2):173–187, 2015.

L. Breiman and J. H. Friedman. Predicting multivariate responses in multiple linear regression. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 59(1):3–54, 1997.

S. Chen, L. Tian, T. Cai, and M. Yu. A general statistical framework for subgroup identification and comparative treatment scoring. *Biometrics*, 73(4):1199–1209, 2017.

F. H. Clarke. *Nonsmooth analysis and control theory*, volume 178. Springer, 1998.

Y. Cui, R. Zhu, and M. Kosorok. Tree based weighted learning for estimating individualized treatment rules with censored data. *Electronic journal of statistics*, 11(2):3927–3953, 2017.

Y. Cui, J.-S. Pang, and B. Sen. Composite difference-max programs for modern statistical estimation problems. *SIAM Journal on Optimization*, 28(4):3344–3374, 2018.

C. Fan, W. Lu, R. Song, and Y. Zhou. Concordance-assisted learning for estimating optimal individualized treatment regimes. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 79(5):1565–1582, 2017.

J. C. Foster, J. M. Taylor, and S. J. Ruberg. Subgroup identification from randomized clinical trial data. *Statistics in Medicine*, 30(24):2867–2880, 2011.

Y. Goldberg and M. R. Kosorok. Q-learning with censored data. *Annals of statistics*, 40(1): 529, 2012.

S. M. Hammer, D. A. Katzenstein, M. D. Hughes, H. Gundacker, R. T. Schooley, R. H. Haubrich, W. K. Henry, M. M. Lederman, J. P. Phair, M. 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.

R. J. Hillestad and S. E. Jacobsen. Reverse convex programming. *Applied Mathematics and Optimization*, 6(1):63–78, 1980.

R. Jiang, W. Lu, R. Song, and M. Davidian. On estimation of optimal treatment regimes for maximizing t-year survival probability. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 2016.

P. Jorion. *Value at risk: the new benchmark for managing financial risk*, volume 2. McGraw-Hill New York, 2001.

T. N. Kakuda. Pharmacology of nucleoside and nucleotide reverse transcriptase inhibitor-induced mitochondrial toxicity. *Clinical therapeutics*, 22(6):685–708, 2000.

K. Knight, W. Fu, et al. Asymptotics for lasso-type estimators. *The Annals of statistics*, 28 (5):1356–1378, 2000.

S. Kummar, P. M. Williams, C.-J. Lih, E. C. Polley, A. P. Chen, L. V. Rubinstein, Y. Zhao, R. M. Simon, B. A. Conley, and J. H. Doroshow. Application of molecular profiling in clinical trials for advanced metastatic cancers. *JNCI: Journal of the National Cancer Institute*, 107 (4), 2015.

E. Laber and Y. Zhao. Tree-based methods for individualized treatment regimes. *Biometrika*, 102(3):501–514, 2015.

K. Lange. *MM Optimization Algorithms.* Society for Industrial and Applied Mathematics, Philadelphia, PA, 2016. doi: 10.1137/1.9781611974409. URL `https://epubs.siam.org/doi/abs/10.1137/1.9781611974409`.

M. Ledoux and M. Talagrand. *Probability in Banach Spaces: isoperimetry and processes.* Springer Science & Business Media, 2013.

Y. Liu, Y. Wang, M. R. Kosorok, Y. Zhao, and D. Zeng. Augmented outcome-weighted learning for estimating optimal dynamic treatment regimens. *Statistics in Medicine*, 37(26):3776–3788, 2018.

K. Lounici, M. Pontil, A. B. Tsybakov, and S. Van De Geer. Taking advantage of sparsity in multi-task learning. *arXiv preprint arXiv:0903.1468*, 2009.

W. Lu, H. H. Zhang, and D. Zeng. Variable selection for optimal treatment decision. *Statistical methods in medical research*, 22(5):493–504, 2013.

J. Mairal. Incremental majorization-minimization optimization with application to large-scale machine learning. *SIAM Journal on Optimization*, 25(2):829–855, 2015.

H. Markowitz. Portfolio selection. *The Journal of Finance*, 7(1):77–91, 1952.

L. Meier, S. Van De Geer, and P. Bühlmann. The group lasso for logistic regression. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 70(1):53–71, 2008.

S. A. Murphy. Optimal dynamic treatment regimes. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 65(2):331–355, 2003.

S. A. Murphy. A generalization error for q-learning. *Journal of Machine Learning Research*, 6 (Jul):1073–1097, 2005.

Y. Nesterov. *Introductory lectures on convex optimization: A basic course*, volume 87. Springer Science & Business Media, 2013.

J.-S. Pang. Partially B-regular optimization and equilibrium problems. *Mathematics of Operations Research*, 32(3):687–699, 2007.

J.-S. Pang, M. Razaviyayn, and A. Alvarado. Computing b-stationary points of nonsmooth dc programs. *Mathematics of Operations Research*, 42(1):95–118, 2016.

G. C. Pflug. Some remarks on the value-at-risk and the conditional value-at-risk. In *Probabilistic Constrained Optimization*, pages 272–281. Springer, 2000.

Z. Qi and Y. Liu. D-learning to estimate optimal individual treatment rules. *Electronic Journal of Statistics*, 12(2):3601–3638, 2018.

M. Qian and S. A. Murphy. Performance guarantees for individualized treatment rules. *The Annals of Statistics*, 39(2):1180, 2011.

J. M. Robins. Optimal structural nested models for optimal sequential decisions. In *Proceedings of the second seattle Symposium in Biostatistics*, pages 189–326. Springer, 2004.

R. Rockafellar. *Conjugate Duality and Optimization.* Society for Industrial and Applied Mathematics, 1974. doi: 10.1137/1.9781611970524. URL `https://epubs.siam.org/doi/abs/10.1137/1.9781611970524`.

R. T. Rockafellar. *Convex analysis.* Princeton university press, 1970.

R. T. Rockafellar. Integral functionals, normal integrands and measurable selections. In *Nonlinear operators and the calculus of variations*, pages 157–207. Springer, 1976.

R. T. Rockafellar and S. Uryasev. Optimization of conditional value-at-risk. *Journal of Risk*, 2:21–42, 2000.

R. T. Rockafellar and S. Uryasev. The fundamental risk quadrangle in risk management, optimization and statistical estimation. *Surveys in Operations Research and Management Science*, 18(1):33–53, 2013.

R. T. Rockafellar and R. J.-B. Wets. *Variational analysis*, volume 317. Springer Science & Business Media, 2009.

A. Rohde, A. B. Tsybakov, et al. Estimation of high-dimensional low-rank matrices. *The Annals of Statistics*, 39(2):887–930, 2011.

A. Ruszczyński and A. Shapiro. Conditional risk mappings. *Mathematics of Operations Research*, 31(3):544–561, 2006a.

A. Ruszczyński and A. Shapiro. Optimization of convex risk functions. *Mathematics of operations research*, 31(3):433–452, 2006b.

S. Sarykalin, G. Serraino, and S. Uryasev. Value-at-risk vs. conditional value-at-risk in risk management and optimization. In *State-of-the-art decision-making tools in the Information-intensive Age*, pages 270–294. Informs, 2008a.

S. Sarykalin, G. Serraino, and S. Uryasev. Value-at-risk vs. conditional value-at-risk in risk management and optimization. *Tutorials in Operations Research*, pages 270–294, 2008b.

I. Steinwart and C. Scovel. Fast rates for support vector machines using gaussian kernels. *The Annals of Statistics*, pages 575–607, 2007.

C. Tan and X. Du. KRAS mutation testing in metastatic colorectal cancer. *World Journal of Gastroenterology: WJG*, 18(37):5171–5180, 2012.

Y. Tao and L. Wang. Adaptive contrast weighted learning for multi-stage multi-treatment decision-making. *Biometrics*, 73(1):145–155, 2017.

L. Tian, A. A. Alizadeh, A. J. Gentles, and R. 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.

K.-C. Toh and S. Yun. An accelerated proximal gradient algorithm for nuclear norm regularized linear least squares problems. *Pacific Journal of Optimization*, 6(615-640):15, 2010.

P. Tseng. Convergence of a block coordinate descent method for nondifferentiable minimization. *Journal of Optimization Theory and Applications*, 109(3):475–494, 2001.

A. W. Van der Vaart. *Asymptotic statistics*, volume 3. Cambridge university press, 2000.

L. Wang, Y. Zhou, R. Song, and B. Sherwood. Quantile-optimal treatment regimes. *Journal of the American Statistical Association*, 113(523):1243–1254, 2018a.

Y. Wang, H. Fu, and D. Zeng. Learning optimal personalized treatment rules in consideration of benefit and risk: with an application to treating type 2 diabetes patients with insulin therapies. *Journal of the American Statistical Association*, 113(521):1–13, 2018b.

C. J. Watkins and P. Dayan. Q-learning. *Machine learning*, 8(3-4):279–292, 1992.

C. J. C. H. Watkins. *Learning from delayed rewards*. PhD thesis, University of Cambridge England, 1989.

T. T. Wu and K. Lange. Multicategory vertex discriminant analysis for high-dimensional data. *The Annals of Applied Statistics*, pages 1698–1721, 2010.

Y. Wu and Y. Liu. Robust truncated hinge loss support vector machines. *Journal of the American Statistical Association*, 102(479):974–983, 2007.

B. Zhang, A. A. Tsiatis, E. B. Laber, and M. Davidian. A robust method for estimating optimal treatment regimes. *Biometrics*, 68(4):1010–1018, 2012.

C. Zhang and Y. Liu. Multicategory angle-based large-margin classification. *Biometrika*, 101 (3):625–640, 2014.

Y. Zhang, E. B. Laber, A. Tsiatis, and M. Davidian. Using decision lists to construct interpretable and parsimonious treatment regimes. *Biometrics*, 71(4):895–904, 2015.

Y. Zhao, D. Zeng, A. J. Rush, and M. R. Kosorok. Estimating individualized treatment rules using outcome weighted learning. *Journal of the American Statistical Association*, 107(499): 1106–1118, 2012.

Y.-Q. Zhao, D. Zeng, E. B. Laber, and M. R. Kosorok. New statistical learning methods for estimating optimal dynamic treatment regimes. *Journal of the American Statistical Association*, 110(510):583–598, 2015a.

Y.-Q. Zhao, D. Zeng, E. B. Laber, R. Song, M. Yuan, and M. R. Kosorok. Doubly robust learning for estimating individualized treatment with censored data. *Biometrika*, 102(1):151, 2015b.

X. Zhou, N. Mayer-Hamblett, U. Khan, and M. R. Kosorok. Residual weighted learning for estimating individualized treatment rules. *Journal of the American Statistical Association*, 112(517):169–187, 2017.