Journal of Machine Learning Research 1 (2000) 1-48 Submitted 4/00; Published 10/00

Sparse concordance-assisted learning for optimal treatment decision

Shuhan Liang SLIANG4@NCSU.EDU
Department of Statistics
North Carolina State University
Raleigh, NC 27695, USA

Wenbin Lu LU@STAT.NCSU.EDU
Department of Statistics
North Carolina State University
Raleigh, NC 27695, USA

Rui Song RSONG@NCSU.EDU
Department of Statistics
North Carolina State University
Raleigh, NC 27695, USA

Lan Wang WANGX346@UMN.EDU
School of Statistics
University of Minnesota
Minneapolis, MN 55455, USA

Editor:

Abstract

To find optimal decision rule, Fan et al. (2016) proposed an innovative concordance-assisted learning algorithm which is based on maximum rank correlation estimator. It makes better use of the available information through pairwise comparison. However the objective function is discontinuous and computationally hard to optimize. In this paper, we consider a convex surrogate loss function to solve this problem. In addition, our algorithm ensures sparsity of decision rule and renders easy interpretation. We derive the L_2 error bound of the estimated coefficients under ultra-high dimension. Simulation results of various settings and application to STAR*D both illustrate that the proposed method can still estimate optimal treatment regime successfully when the number of covariates is large.

Keywords: concordance-assisted learning, optimal treatment regime, L_1 norm, support vector machine, variable selection.

1. Introduction

A treatment regime is a decision rule that tailors treatment for each individual. Instead of randomly assigning treatments, we can select a specific treatment among a few options for each patient based on his or her clinical, genetic and other health information. A decision rule is a procedure to decide which treatment should be picked and it is a function of available information for each patient. Optimal treatment regime aims to find the decision rule that would yield the most favorable outcome. Besides treatment type, treatments

©2000 Authoră .

of interest also include different treatment combinations and dosage level variation. In reality, it often occurs that large number of patient level covariates are available. However, many of them have no qualitative interaction with treatment effect. Covariates may also be correlated with each other. Under such circumstances, variable selection for optimal treatment regime is necessary to avoid overfitting and increase model interpretation.

Many learning algorithms have been proposed to estimate optimal treatment regime (Qian et al. (2012)). Watkins and Dayan (1992) modeled the conditional expectation of outcome (Q-function) and obtained the optimal treatment regime through maximizing the Q-function. Qian and Murphy (2011) extended Q-learning using l_1-penalized least square (PLS). The estimator derived from the two-step procedure may not be consistent if the conditional mean is misspecified. Instead of modeling the outcome, Murphy (2003) proposed the advantage learning (A-learning) algorithm, which is based on modeling contrast function. A contrast function is the difference in potential outcome given different treatments. Lu et al. (2011) considered model selection for estimating optimal treatment regime via penalized least square. A-learning is more robust than Q-learning since it does not require a correct specification of the baseline function but a correct model of interaction term is still needed. In literature, it is common to adopt parametric models for Q-function or contract function. As a result, the corresponding decision rule derived from Q-learning or A-learning may be biased.

To further reduce the impact of model misspecification, Zhang et al. (2012) proposed a value function estimator using inverse probability weighting. The optimal decision rule is derived by maximizing the value function estimator. Zhao et al. (2012) proposed the outcome weighted learning (OWL) algorithm. The OWL approximates optimal treatment decision estimation by transforming an objective function in Zhang et al. (2012) to a classification loss. Larger reward observed indicates higher chance that the optimal decision rule would recommend the same treatment as the patient actually received. Song et al. (2015) extended this method to penalized outcome weighted learning (POWL). Penalty functions include lasso (Tibshirani (1996)) and SCAD (Fan and Li (2001)).

Value search methods suffer from slow convergence and computation difficulties. Fan et al. (2016) proposed a novel concordance-assisted learning (CAL) algorithm to estimate optimal decision rule. Concordance function is motivated by maximizing value function using pairwise comparison between patients. Since concordance function can be estimated by a much smoother function, better asymptotic results can be obtained.

In this paper, we show that concordance-assisted learning algorithm can be transformed to a classification problem. We replace 0-1 loss by a continuous surrogate function. In order to improve the accuracy of optimal treatment regime and interpretation of decision rule, we conduct variable selection by adding lasso penalty to the objective function. We derive error bound of the proposed estimator under ultra-high dimension. We illustrate that the proposed estimator has better performance than existing popular methods under different scenarios together with a clinical trial study.

In Section 2, we reviewed and developed the concordance-assisted learning algorithm. We continued to derive the L_2 error bound of coefficient estimation in Section 3. Section 4 demonstrates the performance of sparse concordance-assisted learning at different settings. We present results of the proposed method for the STAR*D clinical trial in Section 5. The proofs of all lemmas and theorems are provided in the Appendix.

2. Method

In this section, we first introduce notations and explain its usage. It is followed by a concordance-assisted learning overview. We then propose the sparse concordance-assisted learning algorithm and provide an algorithm to calculate the proposed estimator using Douglas-Rachford splitting method.

2.1 Notation

Let $\boldsymbol{X}_i = (X_{i1}, X_{i2}, ..., X_{ip})^T$ denote the vector of covariates measured for the i-th patient, A_i the assigned treatment and Y_i the outcome after treatment. Let $\boldsymbol{X} = (\boldsymbol{X}_1, \boldsymbol{X}_2, ..., \boldsymbol{X}_n)^T$ denote the feature matrix. Assume that $(\boldsymbol{X}_i, A_i, Y_i)$ are independent, identically distributed. Y is a continuous variable and larger value of Y indicates better treatment effect. Denote $g(\boldsymbol{X})$ as the individualized treatment regime (ITR), P^g as the joint distribution of $(\boldsymbol{X}, A = g(\boldsymbol{X}), Y)$, $E^g(Y)$ as the expected outcome if all treatments follow $g(\boldsymbol{X})$. From now on we consider the case of a binary treatment, i.e., A takes values in {0,1}. Denote $\mu(a, \boldsymbol{X}) = E(Y|A = a, \boldsymbol{X})$, Zhang et al. (2012) shows that $g^{opt}(\boldsymbol{X}) = I\{\mu(1, \boldsymbol{X}) > \mu(0, \boldsymbol{X})\}$. Here $g^{opt}(\boldsymbol{X})$ represents optimal treatment regime.

We also assume stable unit treatment value assumption (SUTVA) and no-unmeasured-confounders assumption holds. SUTVA(Rubin (1980)), i.e., $Y = I(A = 0)Y^*(0) + I(A = 1)Y^*(1)$, assumes that no interference exists between treatments of different units and no same treatment variation exists for different units. $Y_i^*(a)$ is the potential outcome after receiving treatment a for subject i. The no-unmeasured-confounders condition, i.e., $\{Y_i^*(0), Y_i^*(1)\} \perp A_i|\boldsymbol{X}_i$, implies all variables that affect treatment assignment or treatment-specific outcomes are observed. The second assumption holds in a randomized trial.

2.2 Concordance-assisted learning overview

Concordance-assisted learning estimates optimal treatment regime by comparing the outcome gain of different treatments between individuals. Maximum Rank Correlation (MRC) estimator (Kendall (1938), Han (1987), Cavanagh and Sherman (1998)) is chosen to estimate the concordance function. CAL further relaxes parametric assumptions and allows for more flexibility. Fan et al. (2016) showed that under certain conditions, optimal treatment regime estimated by concordance-assisted learning is the same as optimal treatment regime estimated by maximizing value functions.

The true optimal decision rule may not be linear, however, throughout the paper, we only search the optimal decision rule within the class of linear decision rules, i.e. $g(\boldsymbol{X}) = I(\boldsymbol{\beta}^T\boldsymbol{X} \geq \beta_0)$. This is partly because that linear decision rules are much easier to compute and interpret compared with nonlinear decision rules, and they generally can achieve high accuracy. CAL is a two-step procedure that first estimates the prescriptive index, i.e., a set of decision rules with fixed covariate weights by maximizing the concordance function:

$$C(\boldsymbol{\beta}) = E\Big\{\{[Y_i^*(1) - Y_i^*(0)] - [Y_j^*(1) - Y_j^*(0)]\}I(\boldsymbol{\beta}^T\boldsymbol{X}_i > \boldsymbol{\beta}^T\boldsymbol{X}_j)\Big\}$$

and then threshold estimator is optimized based on the prescriptive index estimator. Let $D(\boldsymbol{X}_i)$ be the expected outcome gain of treatment 1 for the i^{th} subject, i.e. $D(\boldsymbol{X}_i) = E(Y_i|A_i = 1, \boldsymbol{X}_i) - E(Y_i|A_i = 0, \boldsymbol{X}_i)$. Concordance function is motivated by the following

idea: for pairwise subjects i and j, larger $D(\boldsymbol{X}_i) - D(\boldsymbol{X}_j)$, which means subject i would benefit more by taking treatment 1 compared to subject j, requires larger $\boldsymbol{\beta}^T\boldsymbol{X}_i$ compared to $\boldsymbol{\beta}^T\boldsymbol{X}_j$.

Define $w_i = \frac{\{Y_i - \nu(\boldsymbol{X}_i)\}\{A_i - \pi(\boldsymbol{X}_i)\}}{\pi(\boldsymbol{X}_i)\{1-\pi(\boldsymbol{X}_i)\}}$,here $\nu(\boldsymbol{X}_i)$ is any arbitrary function and $\pi(\boldsymbol{X}_i) = P(A_i = 1|\boldsymbol{X}_i)$ is the propensity score. In practice we choose $\nu(\boldsymbol{X}_i)$ to be the mean response of the patients who receive treatment 0. Given $\boldsymbol{X}_i$, w_i is an unbiased estimator of $D(\boldsymbol{X}_i)$. The proof is given in Appendix C. Concordance-assisted learning can be summarized as follows:

1 Estimate the prescriptive index:

$$\hat{\boldsymbol{\beta}} = \underset{||\boldsymbol{\beta}||=1}{argmax} \frac{1}{n(n-1)} \sum_{i \neq j} (w_i - w_j) I(\boldsymbol{\beta}^T\boldsymbol{X}_i > \boldsymbol{\beta}^T\boldsymbol{X}_j).$$

2 Estimate the threshold using the inverse probability weighted estimator (IPW) proposed by Zhang et al. (2012):

$$\hat{\beta}_0 = \underset{\beta_0}{argmax} \frac{1}{n} \sum_{i=1}^{n} \frac{\{Y_i - \nu(\boldsymbol{X}_i)\} I\{A_i = \hat{g}(\boldsymbol{X}_i)\}}{A_i\pi(\boldsymbol{X}_i) + (1 - A_i)\{1 - \pi(\boldsymbol{X}_i)\}},$$
$$\hat{g}(\boldsymbol{X}_i) = I(\hat{\boldsymbol{\beta}}^T\boldsymbol{X}_i > \beta_0).$$

Although in general the concordance-based estimator does not always lead to the actual optimal decision rule, under certain conditions, concordance-based estimator is the maximizer of value function (Fan et al. (2016)). Concordance-based estimator has attractive properties, including faster convergence rates, known asymptotic distribution (normal) and easy optimization (Fan et al. (2016)). It is a very promising approach for optimal treatment regime estimation. In the next section, we will introduce sparse concordance-assisted learning (SCAL). Compared to CAL, it is easier to optimize and can achieve satisfactory accuracy under high dimension.

2.3 Sparse concordance-assisted learning

Notice that solving for $\hat{\boldsymbol{\beta}}$ is equivalent to minimize:

$$\begin{aligned} &\frac{1}{n(n-1)} \sum_{i \neq j} (w_i - w_j) I(\boldsymbol{\beta}^T\boldsymbol{X}_i < \boldsymbol{\beta}^T\boldsymbol{X}_j), \\ &\text{subject to } ||\boldsymbol{\beta}|| = 1. \end{aligned} \tag{1}$$

(1) is equivalent to minimizing (see Appendix B):

$$\begin{aligned} &\sum_{w_i > w_j} (w_i - w_j) I(\boldsymbol{\beta}^T\boldsymbol{X}_i < \boldsymbol{\beta}^T\boldsymbol{X}_j), \\ &\text{subject to } ||\boldsymbol{\beta}|| = 1. \end{aligned}$$

This alternative expression reduces computation cost and ensures the convexity of the objective function. We replace the indicator loss function with the hinge loss. Hinge loss

is a convex upper bound of the 0-1 loss function. It is often used for support vector machine (Cortes and Vapnik (1995)), a popular classification method with good performance (Gordon (2004)). The optimization of hinge loss function can be solved in polynomial time. Due to the high dimension of p, we use lasso penalty to estimate the optimal treatment regime and perform variable selection simultaneously. Lasso penalty also helps reduce the variance of the fitted coefficients (Zhu et al. (2004)). The prescriptive index estimated by the sparse concordance-assisted learning algorithm (SCAL) is:

$$\hat{\boldsymbol{\beta}} = \underset{\boldsymbol{\beta}}{\operatorname{argmin}} \frac{2}{n(n-1)} \sum_{w_i > w_j} (w_i - w_j) \Big[1 - \boldsymbol{\beta}^T (\boldsymbol{X}_i - \boldsymbol{X}_j)\Big]_+ + \lambda \sum_{j=1}^{p} |\beta_j|.$$

We then estimate the threshold parameter β_0 by:

$$\hat{\beta}_0 = \underset{\beta_0}{\operatorname{argmax}} \frac{1}{n} \sum_{i=1}^{n} w_i I(\hat{\boldsymbol{\beta}}^T \boldsymbol{X}_i > \beta_0).$$

The threshold parameter β_0 is estimated through grid search. In practice, the search range is $[\min(\hat{\boldsymbol{\beta}}^T \boldsymbol{X}_i), \max(\hat{\boldsymbol{\beta}}^T \boldsymbol{X}_j)]$, $1 \le i, j \le n$.

We sort the subjects by descending order of w_i. Therefore,

$$\sum_{i=1}^{n} \sum_{j=1}^{n} (w_i - w_j) I(w_i - w_j > 0) = \sum_{i=1}^{n} \sum_{j=i+1}^{n} (w_i - w_j). \tag{2}$$

The objective function of SCAL can be written as:

$$\hat{\boldsymbol{\beta}} = \underset{\boldsymbol{\beta}}{argmin} \frac{2}{n(n-1)} \sum_{i=1}^{\binom{n}{2}} \delta w_i (1 - \boldsymbol{\beta}^T \boldsymbol{D}_i)_+ + \lambda \sum_{j=1}^{p} |\beta_j|.$$

$$\boldsymbol{D} = \begin{pmatrix} \boldsymbol{X}_1^T - \boldsymbol{X}_2^T \\ \boldsymbol{X}_1^T - \boldsymbol{X}_3^T \\ \cdots \\ \boldsymbol{X}_1^T - \boldsymbol{X}_n^T \\ \boldsymbol{X}_2^T - \boldsymbol{X}_3^T \\ \boldsymbol{X}_2^T - \boldsymbol{X}_4^T \\ \cdots \end{pmatrix}, \quad \delta w = \begin{pmatrix} w_1 - w_2 \\ w_1 - w_3 \\ \cdots \\ w_1 - w_n \\ w_2 - w_3 \\ w_2 - w_4 \\ \cdots \end{pmatrix}.$$

The optimization problem in step 1 is a weighted L_1-SVM problem. The objective function is convex and piecewise linear and many algorithms have been proposed to solve this problem. It can be solved by various linear programming and convex packages. Zhu et al. (2004) proposed an algorithm to compute the whole solution path. Iterative algorithm like Spingarn's Method is another good way to solve this problem. We use three methods: CVX, a package for specifying and solving convex programs (Michael and Stephen (2014), Michael and Stephen (2008)), GLPK (GNU Linear Programming Kit, glp (2016)) and the method proposed by Spingarn (1985) to find the minimizer. Spingarn's method of partial

inverses implements Douglas-Rachford splitting for equality constrained convex problem (Douglas and Rachford (1956)). We add ancillary variables $\boldsymbol{\theta}$ and reformulate (1) as:

$$\begin{aligned} &min f_1(\boldsymbol{\beta}) + f_2(\boldsymbol{\theta}) \\ &subject\ to\ \boldsymbol{\theta} = \boldsymbol{D}\boldsymbol{\beta} \end{aligned}$$

where $f_1(\boldsymbol{\beta}) = \lambda||\boldsymbol{\beta}||_1, f_2(\boldsymbol{\theta}) = \sum_{i=1}^{\binom{n}{2}} \delta w_i (1 - \boldsymbol{D}_i\boldsymbol{\beta})_+$.
The iterative algorithm is as follows:

Repeat

1. $\boldsymbol{V}_1^+ = \Big(prox_{tf_1}(\boldsymbol{\beta}), prox_{tf_2}(\boldsymbol{\theta})\Big)$, where
   $[prox_{tf_1}(\boldsymbol{\beta})]_j = S(\beta_j, t\lambda)$,
   S is soft thresholding operator: $S(x, \lambda) = sgn(x)(|x| - \lambda)_+$.
   $[prox_{tf_2}(\boldsymbol{\beta})]_i = \begin{cases} 1 & \theta \in [1 - t\delta w_i, 1], \\ \theta_i & \theta > 1, \\ \theta_i + t\delta w_i & \theta < 1 - t\delta w_i. \end{cases}$
2. $\boldsymbol{V}_2^+ = \begin{pmatrix} \boldsymbol{I} \\ \boldsymbol{D} \end{pmatrix} \boldsymbol{R}^{-T}\boldsymbol{R}^{-1}\Big[P_1(2\boldsymbol{V}_1^+ - \boldsymbol{V}_3) + \boldsymbol{D}^T P_2(2\boldsymbol{V}_1^+ - \boldsymbol{V}_3)\Big]$,
   $P_1(\boldsymbol{\beta}, \boldsymbol{\theta}) = \boldsymbol{\beta}$, $P_2(\boldsymbol{\beta}, \boldsymbol{\theta}) = \boldsymbol{\theta}$, $\boldsymbol{R}\boldsymbol{R}^T = \boldsymbol{I} + \boldsymbol{D}^T\boldsymbol{D}$ is Cholesky decomposition.
3. $\boldsymbol{V}_3^+ = \boldsymbol{V}_3 + \boldsymbol{V}_2^+ - \boldsymbol{V}_1^+$.

until convergence.
Output: $\boldsymbol{\beta}$

We keep step-size parameter t fixed at 1. The convergence is guaranteed (Spingarn (1985)). The iterative algorithm greatly reduces memory and time cost. Under the case of ultra-high dimension, preprocessing requires $O(p^3)$ work to form and compute Cholesky decomposition and $O(p^2)$ work per iteration. In summary, CVX is the least computational efficient way to estimate prescriptive index and Spingarn's Method is the only approach that can handle STAR*D trial in terms of its scale.

3. Error bound for order-2 U statistics

Define $\boldsymbol{\beta}^* = argmin_{\boldsymbol{\beta}} L(\boldsymbol{\beta})$ where

$$L(\boldsymbol{\beta}) = E\Big\{(w_i - w_j)I(w_i - w_j > 0)[1 - (\boldsymbol{X}_i - \boldsymbol{X}_j)^T\boldsymbol{\beta}]_+\Big\}.$$

Then the gradient vector and Hessian matrix of the loss function $L(\boldsymbol{\beta})$ are:

$$\begin{aligned} S(\boldsymbol{\beta}) &= -E\Big\{(w_i - w_j)I(w_i - w_j > 0)I[1 - (\boldsymbol{X}_i^T - \boldsymbol{X}_j^T)\boldsymbol{\beta} \geq 0](\boldsymbol{X}_i - \boldsymbol{X}_j)\Big\}, \\ H(\boldsymbol{\beta}) &= E\Big\{(w_i - w_j)I(w_i - w_j > 0)Dirac\delta[1 - (\boldsymbol{X}_i^T - \boldsymbol{X}_j^T)\boldsymbol{\beta}](\boldsymbol{X}_i - \boldsymbol{X}_j)(\boldsymbol{X}_i^T - \boldsymbol{X}_j^T)\Big\}, \end{aligned}$$

where $Diracδ$ is the Dirac delta function. Denote the index set of active features as $T = \{1 \le j \le p : \beta_j^* \neq 0\}$ and $|T| = q$. $\hat{\boldsymbol{\beta}}(\lambda) = \underset{\boldsymbol{\beta}}{argmin}\ l_n(\boldsymbol{\beta}, \lambda)$ is an estimator of $\boldsymbol{\beta}^*$, where

$$l_n(\boldsymbol{\beta}, \lambda) = \frac{2}{n(n-1)} \sum_{i=1}^{n} \sum_{j=1}^{n} (w_i - w_j) I(w_i - w_j > 0)[1 - (\boldsymbol{X}_i^T - \boldsymbol{X}_j^T)\boldsymbol{\beta}]_+ + \lambda ||\boldsymbol{\beta}||_1.$$

We assume the following regularity conditions:

(A1) The densities of $\boldsymbol{X}_i$, i = 1, 2, ... are continuous and have common support in $\boldsymbol{R}^p$, and there exists a constant $M_0 > 0$ such that $|\boldsymbol{X}_{ij}| \le M_0$, $i \in R^+,\ j \in 1, ..., p$
(A2) Denote $\boldsymbol{Z}_{ij} = \boldsymbol{X}_i^T - \boldsymbol{X}_j^T$ with probability density function $f^*(z)$. There exists $B(0, \delta_0)$, a ball centered at 0 with radius $\delta_0 > 0$ such that $E[(w_i - w_j) I(w_i - w_j > 0) | \boldsymbol{Z}_{ij} = \boldsymbol{z}_{ij}] f^*(\boldsymbol{z}_{ij}) > C_3$ for every $\boldsymbol{z}_{ij} \in B(0, \delta_0)$.
(A3) $\int E[(w_i - w_j) I(w_i - w_j > 0) | \boldsymbol{z}] z_k f^*(\boldsymbol{z}) d\boldsymbol{z} \neq 0$ for some k.
(A4) There exists a constant M_1 s.t. $\underset{\boldsymbol{d} \in \boldsymbol{R}^p : ||\boldsymbol{d}||_0 \le 2q}{max} \frac{\boldsymbol{d}^T \boldsymbol{D}^T \boldsymbol{D} \boldsymbol{d}}{\frac{n(n-1)}{2} ||\boldsymbol{d}||_2^2} \le M_1$ almost surely.
(A5) Denote $\bar{c} = \frac{c-1}{c+1}$ where c is a constant satisfying $\lambda \ge c||S(\hat{\boldsymbol{\beta}}^*)||_\infty$, T is the set of significant coefficients (non-zero coefficients). There exists a constant $M_2 > 0$ such that $\underset{\boldsymbol{d} \in \boldsymbol{R}^p : ||d||_0 \le q, ||\boldsymbol{d}_T||_1 \ge \bar{c}||\boldsymbol{d}_{T^c}||_1}{min} \frac{\boldsymbol{d}^T H(\boldsymbol{\beta}^*) \boldsymbol{d}}{||\boldsymbol{d}||_2^2} \ge M_2$.
(A6) $q = O(n^{c_1})$ for some $0 \le c_1 < \frac{1}{2}$.
(A7) There exists a constant M_3 such that for any w_M, $P(|w_i| > w_M) < \exp(-\frac{w_M}{M_3})$.

Condition (A1) ensures $H(\boldsymbol{\beta})$ is well-defined and continuous in $\boldsymbol{\beta}$. Condition (A2) is similar to condition (A2) in Koo et al. (2008). It guarantees $L(\boldsymbol{\beta}) \to \infty$ as $||\boldsymbol{\beta}|| \to \infty$ and further guarantees the existence of $\boldsymbol{\beta}^*$. Condition (A3) implies that $\boldsymbol{\beta}^* \neq 0$. Condition (A4) gives the upper bound of restricted eigenvalue (RE). It can guarantee the Gram matrix is positive definite over a subset of vectors (Bickel et al. (2009)). Condition (A5) gives the lower bound for restricted eigenvalue of $H(\boldsymbol{\beta}^*)$. Condition (A6) restricts the divergence rate for the number of none-zero variables. Condition (A7) is a popular distribution assumption in literature.

Lemma 1: *Assume condition (A1) and (A7) satisfied. Suppose* $\lambda = c\sqrt{32A(\alpha)(\log p)^3/n}$, *c is some given constant,* α *is a small probability and* $A(\alpha)$ *is a constant such that* $(2 + n)p^{-A(\alpha)^{1/3} M_0^{-2/3} M_3^{-2/3} + 1} \le \alpha$, *we have*

$$P(\lambda \ge c||\hat{S}(\boldsymbol{\beta}^*)||_\infty) \ge 1 - \alpha.$$

Lemma 2: *Assume conditions (A1), (A4), (A6) and (A7) are satisfied, $p > n$. Let*

$$\begin{aligned} B(h) = \frac{2}{n(n-1)}\Big| &\sum_{i=1}^{n}\sum_{j=1}^{n}(w_i - w_j)I(w_i - w_j > 0)[1 - (\boldsymbol{X}_i^T - \boldsymbol{X}_j^T)(\boldsymbol{\beta}^* - h)]_+ \\ &- \sum_{i=1}^{n}\sum_{j=1}^{n}(w_i - w_j)I(w_i - w_j > 0)[1 - (\boldsymbol{X}_i^T - \boldsymbol{X}_j^T)\boldsymbol{\beta}^*]_+ \\ &- \sum_{i=1}^{n}\sum_{j=1}^{n}E\{(w_i - w_j)I(w_i - w_j > 0)[1 - (\boldsymbol{X}_i^T - \boldsymbol{X}_j^T)(\boldsymbol{\beta}^* - h)]_+ \\ &- \sum_{i=1}^{n}\sum_{j=1}^{n}(w_i - w_j)I(w_i - w_j > 0)[1 - (\boldsymbol{X}_i^T - \boldsymbol{X}_j^T)\boldsymbol{\beta}^*]_+\}\Big|. \end{aligned}$$

Then for sufficiently large n,

$$P\Big(\sup_{||\boldsymbol{h}||_0 \le q, ||\boldsymbol{h}||_2 \neq 0} \frac{B(\boldsymbol{h})}{||\boldsymbol{h}||_2} \ge (1 + C_2\sqrt{M_1})q\sqrt{\frac{32\log p}{n}}[M_3 q(C_2^2 - 2)\log p + M_3 \log 2n]\Big) \le 3p^{-q(C_2^2-2)}$$

Lemma 2 guarantees that with high probability,

$$\frac{2}{n(n-1)}\sum_{i=1}^{n}\sum_{j=1}^{n}(w_i - w_j)I(w_i - w_j > 0)\Big\{[1 - (\boldsymbol{X}_i^T - \boldsymbol{X}_j^T)(\boldsymbol{\beta}^* - h)]_+ - [1 - (\boldsymbol{X}_i^T - \boldsymbol{X}_j^T)\boldsymbol{\beta}^*]_+\Big\}.$$

is within a small range of its expectation. From now on, we choose h to be $h = \boldsymbol{\beta}^* - \hat{\boldsymbol{\beta}}(\lambda)$.

Lemma 3: *For $\lambda \ge c||S(\hat{\boldsymbol{\beta}}^*)||_\infty$,*

$$||\boldsymbol{h}_T||_1 \ge \bar{c}||\boldsymbol{h}_{T^c}||_1,$$

where $\bar{c} = \frac{c-1}{c+1}$, T is the set of significant coefficients (non-zero coefficients) and $|T| \le q$.

Theorem 4: *Suppose (A1) - (A7) hold, then $\hat{\boldsymbol{\beta}}$ satisfies*

$$||\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}^*||_2 \le \sqrt{1 + \frac{1}{\bar{c}}}[\frac{2\lambda\sqrt{q}}{M_2} + \frac{2C_4}{M_2}q^2\sqrt{\frac{(\log p)^3}{n}}(\frac{5}{4} + \frac{1}{\bar{c}})]$$

with probability at least $1 - 3p^{-q(C_2^2-2)+1}$, where C_4 is a constant.

When $\lambda = c\sqrt{32A(\alpha)(\log p)^3/n}$, the first term has order $\sqrt{(\log p)^3 q/n}$ and the second term has order $q^2\sqrt{(\log p)^3/n}$. Therefore, with high probability,

$$||\hat{\boldsymbol{\beta}}(\lambda) - \boldsymbol{\beta}^*||_2 = O_p(q^2\sqrt{(\log p)^3/n}).$$

The proofs of the lemmas and theorems are given in the Appendix. We first show the existence of $\boldsymbol{\beta}^*$ and it is non-trivial (formally stated and proved in Appendix A). Lemma 1

indicates a certain λ to bound the infinity norm of gradient vector with large probability. Lemma 2, and Lemma 3 are stepping stones on the path to proving Theorem 4. To be specific, Lemma 2 establishes the relationship between difference in loss functions and its expectation. Lemma 3. Similar results can be found in Zhang et al. (2016) and Peng et al. (2015). Compared to their proofs, our proof uses Hoeffding's inequality for u-statistics of order 2. Another difference is that an unbounded weight exists in our objective function. To handle this challenge, we make assumptions on the tail distribution of weights and adjust the bound correspondingly.

4. Simulation studies

In this section, we demonstrate the numerical performance of the proposed method. We simulate data from a randomized experiment and evaluate the estimated optimal treatment regimes using sparse concordance-assisted learning and penalized outcome weighted learning. Objective function of POWL is:

$$\frac{1}{n}\sum_{i=1}^{n}\Big\{\frac{Y_i}{A_i\pi(\boldsymbol{X_i})+(1-A_i)[1-\pi(\boldsymbol{X_i})]}[1-(2A_i-1)g(\boldsymbol{X}_i)]_+\Big\}+\lambda\sum_{j=1}^{p}|\beta_j|. \tag{3}$$

Here g is a linear function, i.e. $g(\boldsymbol{\beta},\boldsymbol{X}_i)=\boldsymbol{\beta}^T\boldsymbol{X}_i+\beta_0$. Notice that the solution to (2) will remain the same if every Y_i is added to a constant c. In order to guarantee the objective function is convex and the optimization problem is feasible, we add a constant to all Y_i to make sure the smallest response is positive. The constant is chosen so that the smallest shifted response is 0.01. POWL is implemented using convex toolbox in MATLAB. We compute the IPW estimator using Monte Carlo simulations with 1000 replicates and select the tuning parameter λ with the largest $\hat{Y}_{\hat{opt}}$.

To evaluate the estimated decision rule, we report the mean outcome following the estimated optimal treatment regime (Estimated Value) and the percentage of correct decision (PCD) of the estimated optimal treatment regime. The mean of value function following estimated treatment regime is calculated by plugging estimated decisions in the real model using Monte Carlo simulations with 1000 replicates. The mean of value function following the true optimal treatment regime (True Value) is also listed. In addition, we report the mean square error of $\hat{\boldsymbol{\beta}}$. For variable selection, we report correct number of zero coefficients (Corr0) and incorrect number of zero coefficients (Incorr0) compared to the true optimal treatment regime. Results are evaluated and compared under various settings. The associated sample standard deviations are included in the parentheses.

4.1 Low Dimension

We follow the first simulation scenario in Zhao et al. (2012): $X_{i1}, X_{i2}, ..., X_{i50}$ are generated independently from a uniform distribution on $[-1,1]$, $i=1,\cdots,n$. The treatment indicator A is generated from Bernoulli distribution with $p=0.5$. The conditional density of the response Y given $\boldsymbol{X}$ and A is normal, with mean $Q_0(\boldsymbol{X}_i)=1+2X_{i1}+X_{i2}+0.5X_{i3}+0.442(1-X_{i1}-X_{i2})(2A_i-1)$ and variance 1. Here only X_{i1} and X_{i2} have linear interaction

with treatment. We ran 100 simulations with n=30, 100 and 200 respectively to estimate the individualized treatment rule using SCAL and POWL. Table 1 summarizes the results.

From Table 1 we have the following observations. First, sparse concordance-assisted learning leads to more accurate estimates of $\boldsymbol{\beta}$ and better variable selection results. Sparse concordance-assisted learning achieves smaller mean square error (MSE) smaller Incorr0 and smaller Corr0 than that of penalized outcome weighted learning. Although the model size of POWL is smaller and closer to the real model size, its value function estimation is smaller. This further demonstrates that SCAL can select covariates that have strong interaction with treatment and compensate for the influence of model complexity.

The mean of value function following the estimated treatment regime gets closer to the real optimal value as sample size increases. We also notice that SCAL estimator does not vary much from sample to sample: both PCD and value function estimated by SCAL have smaller variance.

In general two methods lead to comparable results. The difference between methods are small, and this is especially true when the sample size is large. When n=30, SCAL leads to much closer value function estimation to true optimal value function estimation than that of POWL. But when n=200, the difference between value functions estimated from SCAL and POWL is only 1.39% of the true value function estimation. It is not surprising since compared to concordance-assisted learning, outcome weighted learning uses information less efficiently, which, can be made up of by increasing available information.

	n	MSE	Incorr0(0)	Corr0(48)	PCD	Estimated Value	True Value
POWL	30	1.60	1.70	42.23	0.615(0.02)	1.09(0.02)	1.44
	100	1.27	1.94	46.64	0.768(0.02)	1.27(0.02)	1.44
	200	1.09	1.99	47.78	0.786(0.02)	1.30(0.03)	1.44
SCAL	30	1.40	0.73	35.79	0.659(0.01)	1.16(0.01)	1.44
	100	0.52	0.11	41.97	0.764(0.01)	1.31(0.01)	1.44
	200	0.19	0.01	46.03	0.749(0.01)	1.32(0.01)	1.44

Table 1: Simulation results of sparse concordance-assisted learning (SCAL) and penalized outcome weighted learning (POWL): low-dimensional case

4.2 High Dimension

We consider the following six models to generate simulation data:

Model I: $\boldsymbol{Y} = \boldsymbol{X}\gamma_1 + \boldsymbol{X}\boldsymbol{\beta}A + \epsilon$, $\gamma_1 = (3, -1, 1, \mathbf{0_{p-2}})^T$, $\boldsymbol{\beta} = (2, 1.8, 0, 0, 0, -1.6, \mathbf{0_{p-6}})^T$.

Model II: $\boldsymbol{Y} = 3 - 0.5(\boldsymbol{X}\gamma_1)^2 + 0.625(\boldsymbol{X}\gamma_2)^2 + \boldsymbol{X}\boldsymbol{\beta}A + \epsilon$, $\gamma_1 = (1, 0.5, \mathbf{0_{p-2}})^T$, $\gamma_2 = (0, 1, \mathbf{0_{p-2}})^T$,$\boldsymbol{\beta} = (2, 1.8, 0, 0, 0, -1.6, \mathbf{0_{p-6}})^T$.

Model III: $\boldsymbol{Y} = 1-\sin(\boldsymbol{X}\gamma_1)+\sin(\boldsymbol{X}\gamma_2)+\boldsymbol{X}\boldsymbol{\beta}A+\epsilon$, $\gamma_1 = (1, \mathbf{0_{p-1}})^T$,$\gamma_2 = (0, 1, \mathbf{0_{p-2}})^T$, $\boldsymbol{\beta} = (2, 1.8, 0, 0, 0, -1.6, \mathbf{0_{p-6}})^T$.

Model IV: $\boldsymbol{Y} = \boldsymbol{X}\gamma_1+(\boldsymbol{X}\boldsymbol{\beta})^3A+\epsilon$, $\gamma_1 = (3, -1, 1, \mathbf{0_{p-2}})^T$, $\boldsymbol{\beta} = (1, 0.9, 0, 0, 0, -0.8, \mathbf{0_{p-6}})^T$.

Model V: $\boldsymbol{Y} = 3 - 0.5(\boldsymbol{X}\gamma_1)^2 + 0.625(\boldsymbol{X}\gamma_2)^2 + (\boldsymbol{X}\boldsymbol{\beta})^3A + \epsilon$, $\gamma_1 = (1, 0.5, \mathbf{0_{p-2}})^T$, $\gamma_2 = (0, 1, \mathbf{0_{p-2}})^T$,$\boldsymbol{\beta} = (1, 0.9, 0, 0, 0, -0.8, \mathbf{0_{p-6}})^T$.

Model VI: $\boldsymbol{Y} = 1-\sin(\boldsymbol{X}\gamma_1)+\sin(\boldsymbol{X}\gamma_2)+(\boldsymbol{X}\boldsymbol{\beta})^3A+\epsilon$, $\gamma_1 = (1, \mathbf{0_{p-1}})^T$,$\gamma_2 = (0, 1, \mathbf{0_{p-2}})^T$, $\boldsymbol{\beta} = (1, 0.9, 0, 0, 0, -0.8, \mathbf{0_{p-6}})^T$.

There are three baseline functions: Models I and III share the same linear baseline function; Models II and IV share the same higher order polynomial baseline function; Models III and VI share the same complex baseline function. In the first three models there is a linear interaction between covariates and treatment; in the last three models treatment and a cubic function of prescriptive index are interacted with each other. All six models have the same important variables $\boldsymbol{X}_{i1}$, $\boldsymbol{X}_{i2}$ and $\boldsymbol{X}_{i6}$. Covariates $\boldsymbol{X}_i = (X_{i1}, X_{i2}, ..., X_{ip})^T$ are generated from a multivariate normal distribution: each entry is standard normal and the correlation between covariates is $Corr(X_{ij}, X_{ik}) = \rho^{|j-k|}$ for $1 \leq j \neq k \leq p$. ρ is chosen to be 0 and 0.2 respectively. The error term ϵ is generated from standard normal distribution. We ran 100 simulations for each scenario with n=100 and p=500, 1000 respectively.

We consider randomized studies where A is generated from Bernoulli distribution with p = 0.5. The performance of variable selection and treatment regime estimation of both methods are summarized in Tables 2 and 3. Conclusions are similar to low-dimensional case. SCAL selects more important variables and fewer unimportant variables than POWL. SCAL also provides more accurate decision rule estimation. Its value function estimate is 0.73 higher than POWL, which is 18.3% of true optimal value function estimate and its PCD is 22.4% higher. The comparison between MSE of $\boldsymbol{\beta}$ estimate further supports the advantage of SCAL.

The performance of SCAL continues to improve as magnitude of interaction between treatment and covariates increases. However we are unable to see this trend from POWL. In general, SCAL can recover important variables better under cubic prescriptive index and treatment interaction than under linear interaction. When $\rho = 0$ and $p = 500$, comparing Model 1 with Model 4, we can see that MSE and Incorr0 dropped a lot. On the contrary, for POWL under the same circumstances, MSE remains almost the same and Incorr0 even increases. Results of Corr0 as well as PCD agree with this statement.

SCAL has demonstrated its performance under high dimension. It also shows the potential to identify important variables when p goes even larger. Performance of variable selection and optimal treatment regime estimation become slightly worse when p increases from 500 to 1000. In reality it is common that hundreds of covariates are available for each patient. Reliable results can still be obtained by SCAL under such circumstances.

The PCD slightly increases when the correlation between covariates increases, suggesting that correlated covariate structure can reduce the impact of falsely selected unimportant variables and missing important variables. When ρ increases from 0 to 0.5, $p = 500$, the PCD of Model 1 increases 4.2% for SCAL and 0.9% for POWL. Due to the fact that correlation exists in most of the real-world data, SCAL proves itself to be a desirable approach.

Next, we consider observational studies where the propensity score is estimated from data. To be specific, the treatment indicator A is generated from Bernoulli($\frac{1}{1+e^{-u}}$), where $u(\boldsymbol{X}) = 0.01 - 0.5 * X_1 + 0.4 * X_{10}$. Here, we consider the same high-dimensional settings with $p = 500$ and $\rho = 0.2$, and the propensity score is estimated using the l_1-penalized logistic regression. The results of SCAL and POWL are summarized in Table 4. Overall, SCAL outperforms POWL in terms of variable selection and estimating optimal treatment regime in all cases as observed for randomization studies.

p	ρ	Model	MSE	Incorr0(0)	Corr0(497/997)	PCD	Estimated Value	True Value
500	0	Model 1	0.61	0.75	482.62	0.744(0.01)	3.80(0.02)	4.21
		Model 2	0.56	0.57	485.34	0.763(0.01)	3.79(0.03)	4.16
		Model 3	0.44	0.49	488.12	0.786(0.01)	1.92(0.02)	2.22
		Model 4	0.35	0.35	486.81	0.801(0.01)	5.67(0.04)	5.93
		Model 5	0.32	0.25	487.00	0.810(0.01)	5.63(0.05)	5.88
		Model 6	0.29	0.20	485.33	0.820(0.01)	3.74(0.04)	3.94
	0.2	Model 1	0.50	0.71	487.37	0.783(0.01)	4.00(0.02)	4.31
		Model 2	0.47	0.68	488.52	0.788(0.01)	3.86(0.02)	4.16
		Model 3	0.38	0.46	490.32	0.816(0.01)	2.08(0.02)	2.32
		Model 4	0.28	0.25	487.45	0.831(0.01)	6.53(0.02)	6.68
		Model 5	0.28	0.23	487.86	0.832(0.06)	6.36(0.05)	6.53
		Model 6	0.25	0.18	485.98	0.845(0.01)	4.56(0.02)	4.68
1000	0	Model 1	0.67	0.89	981.92	0.741(0.01)	3.81(0.02)	4.26
		Model 2	0.56	0.70	985.66	0.755(0.01)	3.84(0.03)	4.20
		Model 3	0.49	0.58	988.22	0.782(0.01)	1.93(0.01)	2.26
		Model 4	0.38	0.34	986.01	0.797(0.01)	5.70(0.03)	5.97
		Model 5	0.34	0.20	985.83	0.800(0.01)	5.68(0.04)	5.92
		Model 6	0.32	0.20	985.25	0.813(0.01)	3.76(0.03)	3.98
	0.2	Model 1	0.56	0.94	989.91	0.762(0.01)	4.01(0.02)	4.35
		Model 2	0.51	0.64	986.95	0.779(0.01)	3.88(0.02)	4.20
		Model 3	0.43	0.54	989.14	0.805(0.01)	2.09(0.02)	2.36
		Model 4	0.31	0.27	987.40	0.819(0.01)	6.55(0.02)	6.72
		Model 5	0.30	0.24	985.37	0.827(0.01)	6.41(0.02)	6.57
		Model 6	0.26	0.18	987.92	0.832(0.01)	4.59(0.02)	4.73

Table 2: Simulation results of sparse concordance-assisted learning (SCAL): high dimensional case

p	ρ	Model	MSE	Incorr0(0)	Corr0(497/997)	PCD	Estimated Value	True Value
500	0	Model 1	1.81	2.37	449.06	0.521(0.01)	3.03(0.02)	4.21
		Model 2	1.76	2.40	456.11	0.520(0.01)	2.98(0.01)	4.16
		Model 3	1.76	2.33	450.64	0.521(0.01)	1.04(0.01)	2.22
		Model 4	1.81	2.53	450.63	0.516(0.01)	3.00(0.05)	5.93
		Model 5	1.77	2.53	453.97	0.517(0.01)	2.95(0.05)	5.88
		Model 6	1.78	2.53	454.53	0.517(0.01)	1.02(0.06)	3.94
	0.2	Model 1	1.76	2.30	456.63	0.528(0.01)	3.08(0.02)	4.31
		Model 2	1.76	2.38	458.01	0.528(0.07)	2.92(0.02)	4.16
		Model 3	1.75	2.24	458.19	0.530(0.01)	1.09(0.02)	2.32
		Model 4	1.80	2.48	453.13	0.520(0.01)	3.13(0.07)	6.68
		Model 5	1.81	2.53	456.69	0.519(0.01)	2.98(0.07)	6.53
		Model 6	1.79	2.52	453.29	0.521(0.01)	1.15(0.08)	4.68
1000	0	Model 1	1.79	2.67	948.08	0.512(0.01)	3.04(0.01)	4.26
		Model 2	1.82	2.68	947.05	0.512(0.01)	2.99(0.01)	4.20
		Model 3	1.84	2.66	949.14	0.514(0.01)	1.06(0.01)	2.26
		Model 4	1.83	2.78	950.22	0.508(0.01)	2.88(0.04)	5.97
		Model 5	1.83	2.78	947.72	0.508(0.01)	2.86(0.04)	5.92
		Model 6	1.82	2.78	947.83	0.507(0.01)	0.91(0.04)	3.98
	0.2	Model 1	1.83	2.68	950.87	0.517(0.01)	3.06(0.02)	4.35
		Model 2	1.82	2.66	946.94	0.516(0.01)	2.91(0.01)	4.20
		Model 3	1.78	2.66	951.12	0.517(0.01)	1.08(0.02)	2.36
		Model 4	1.79	2.77	948.49	0.508(0.01)	2.93(0.05)	6.72
		Model 5	1.81	2.76	947.55	0.507(0.01)	2.76(0.05)	6.57
		Model 6	1.76	2.77	952.16	0.510(0.01)	0.96(0.05)	4.73

Table 3: Simulation results of penalized outcome weighted learning(POWL): high dimensional case

p=500	ρ	Model	MSE	Incorr0(0)	Corr0(497)	PCD	Estimated Value	True Value
SCAL	0.2	Model 1	0.82	0.85	484.96	0.732(0.01)	3.83(0.04)	4.33
		Model 2	0.73	0.77	487.52	0.744(0.01)	3.72(0.04)	4.18
		Model 3	0.67	0.70	489.03	0.768(0.01)	1.93(0.04)	2.34
		Model 4	0.41	0.47	483.77	0.807(0.01)	6.42(0.04)	6.70
		Model 5	0.32	0.31	483.05	0.815(0.01)	6.36(0.04)	6.55
		Model 6	0.36	0.41	486.40	0.818(0.01)	4.46(0.05)	4.71
POWL	0.2	Model 1	1.64	2.49	460.97	0.505(0.01)	2.99(0.01)	4.33
		Model 2	1.56	2.70	467.86	0.509(0.01)	2.87(0.01)	4.18
		Model 3	1.62	2.56	467.20	0.512(0.01)	1.04(0.01)	2.34
		Model 4	1.52	2.66	470.31	0.500(0.01)	2.92(0.04)	6.70
		Model 5	1.51	2.70	472.09	0.501(0.01)	2.79(0.04)	6.55
		Model 6	1.48	2.66	471.58	0.500(0.01)	0.93(0.05)	4.71

Table 4: Simulation results for observational studies: sparse concordance-assisted learning (SCAL) and penalized outcome weighted learning (POWL):

5. Application to STAR*D Study

We apply the proposed method to STAR*D Study, the largest and longest study ever conducted to assess effectiveness of depression treatments. 4041 outpatients who are diagnosed with major depressive disorder (MDD), representing of various ethnic and socioeconomic groups are collected. There are four levels in this clinical trial and at each level different treatments are evaluated and compared. See Fava et al. (2003) for design and measurement details of STAR*D study.

In the data analysis, we focus on patients who received bupropion (BUP) or sertraline (SER) in the second level to illustrate our method. Among the 309 selected subjects, 153 of them received bupropion (BUP) and 166 received sertraline (SER). In order to be consistent with our previous notation, we use 0 to represent SER and 1 to represent BUP. We consider all 305 covariates collected from enrollment, IVR call, ROA interviews, clinic visit and other events (such as suicide, non-serious adverse event and protocol deviation) to recommend individualized treatment for each patient. We choose negative 16-item Quick Inventory of Depressive Symptomatology-Clinician-Rated (QIDS-C16) as our response variable. QIDS-C16 is reverse coded so that it satisfies larger outcome indicates better treatment effect. The negative QIDS-C16 is in the range of -24 to 0.

We apply SCAL using Spingarn's method to this data set. λ_{opt} is tuned using 5-fold cross validation. A pre-defined range (0, 4) is searched and λ is chosen based on the IPW estimator. Propensity score is estimated by proportion of subjects who receive treatment 1 in the training dataset. For comparison, we also evaluate the performance of POWL using 5-fold cross validation. The response is shifted with the smallest value to be 0.01. Note that the adjusted response is only used to optimize objective function; Value function is estimated using original response.

To compare the estimated treatment regimes on STAR*D data, we draw bootstrap samples over 1,000 times and estimate the 95% confident interval of difference between expected outcome following estimated treatment regime from SCAL and the non-dynamic treatment regimes. The expected outcome difference between SCAL and POWL is also calculated. See Table 5 for estimate of value function based on estimated optimal treatment regime and 95% confident interval of the differences.

Treatment Regime	Estimated Value	Diff	95% CI on Diff
Optimal Regime(SCAL)	-6.77		
Optimal Regime(POWL)	-9.46	2.69	(1.18, 4.24)
BUP	-10.52	3.75	(2.38, 5.19)
SER	-10.74	3.97	(2.57, 5.50)

Table 5: Estimated values, difference in estimated values and its 95% CI

The estimated value function by SCAL is significantly larger than either of the non-dynamic treatment regimes. It is also significantly larger than the expected outcome following the optimal treatment regime based on POWL at $\alpha = 0.05$. Compared to POWL, SCAL achieves a reasonable model size. It keeps good balance of including important features and controlling the complexity of the model. Estimate based on POWL is too sparse and many covariates which have interaction with treatment effects are missed. Table 5 and

Table 6 are summaries of treatment actually received versus estimated optimal treatment. We can see that estimated optimal treatment regime based on SCAL tends to be more balance. Based on SCAL, among all 171 subjects who were assigned to SER, 92 of them should stay in the same treatment and 79 of them should be assigned to BUP. While the result of POWL indicates 111 of them were assigned to the optimal treatment and only 56 patients should switch to BUP.

	Estimated treatment: SER	Estimated treatment: BUP
Randomized treatment: SER	92	75
Randomized treatment: BUP	79	73

Table 6: Summary of treatment recommended by SCAL

	Estimated treatment: SER	Estimated treatment: BUP
Randomized treatment: SER	111	56
Randomized treatment: BUP	106	46

Table 7: Summary of treatment recommended by POWL

6. Conclusion

We propose a variable selection method based on concordance-assisted learning for estimating optimal treatment regime. Our method can minimize the weighted misclassification rate and select prescriptive index simultaneously. The proposed method gives much more accurate decision rule and value function estimation than existing popular methods under various simulation settings. Moreover, inputs that are correlated with treatments effects are also successfully identified. Sparse concordance-assisted learning achieves promising result in constructing real-world decision. We also study the error bound of SCAL in ultra-high dimension.

The proposed method does not require model specification except for propensity score. It is based on an estimate of contrast function which can be defined easily under binary treatment circumstance. SCAL can solve problems in other fields as well. One popular example is to determine the best move in a game. SCAL can be implemented to choose the best action. In the future, one interesting direction of our study would be to extend the definition of contrast function when more than two treatment arms are available. We may also replace linear support vector machine by other kernels to see whether a better treatment regime can be found.

Acknowledgments

We would like to acknowledge support for this project from National Institute of Mental Health for providing the STAR*D data. We would also like to thank Dr. Eric Chi for his help in developing Spingarn's method.

Appendix A.

In this appendix we first state and prove two fundamental lemmas.

Lemma 0.1: *If condition (A1) and (A2) are satisfied, $\boldsymbol{\beta}^*$ exists.*

Proof. of Lemma 0.1:

$$
\begin{aligned}
&E\Big\{(w_i - w_j)I(w_i - w_j > 0)[1 - \boldsymbol{\beta}^*(\boldsymbol{X}_i^T - \boldsymbol{X}_j^T)]_+\Big\} \\
=&E\Big(E\{(w_i - w_j)I(w_i - w_j > 0)[1 - \boldsymbol{\beta}^*(\boldsymbol{X}_i^T - \boldsymbol{X}_j^T)]_+|\boldsymbol{X}_i^T - \boldsymbol{X}_j^T\}\Big)
\end{aligned}
$$

$$
\begin{aligned}
L(\boldsymbol{\beta}) =& \int E\Big[(w_i - w_j)I(w_i - w_j > 0)|\boldsymbol{z}\Big](1 - \boldsymbol{z}^T\boldsymbol{\beta})_+ f^*(\boldsymbol{z})d\boldsymbol{z} \\
\geq& \int I(\boldsymbol{z}^T\boldsymbol{\beta} \leq 0)E\Big[(w_i - w_j)I(w_i - w_j > 0)|\boldsymbol{z}\Big](1 - \boldsymbol{z}^T\boldsymbol{\beta})f^*(\boldsymbol{z})d\boldsymbol{z} \\
\geq& \int I(\boldsymbol{z}^T\boldsymbol{\beta} \leq 0)E\Big[(w_i - w_j)I(w_i - w_j > 0)|\boldsymbol{z}\Big](-\boldsymbol{z}^T\boldsymbol{\beta})f^*(\boldsymbol{z})d\boldsymbol{z} \\
\geq& C_3 \int_{B(0,\delta_0)} I(\boldsymbol{z}^T\boldsymbol{\beta} \leq 0)(-\boldsymbol{z}^T\boldsymbol{\beta})d\boldsymbol{z} \\
=& C_3||\boldsymbol{\beta}|| \int_{B(0,\delta_0)} I(\boldsymbol{z}^T w \leq 0)(-\boldsymbol{z}^T w)d\boldsymbol{z} \\
\geq& C_3||\boldsymbol{\beta}||vol\Big(B(0,\delta_0)\bigcap\{-\boldsymbol{z}^T w \geq \epsilon\}\Big)\epsilon,
\end{aligned}
$$

where $w = \boldsymbol{\beta}/||\boldsymbol{\beta}||$ and vol is short for volume. Note that $vol\Big(B(0,\delta_0)\bigcap\{-\boldsymbol{z}^T w \geq \epsilon\}\Big) > 0$ for some $\epsilon < \delta_0$ and $vol\Big(B(0,\delta_0)\bigcap\{-\boldsymbol{z}^T w \geq \epsilon\}\Big)$ is independent of $\boldsymbol{\beta}$. Thus $L(\beta) \to \infty$ as $||\boldsymbol{\beta}|| \to \infty$. Since $L(\boldsymbol{\beta})$ is convex in $\boldsymbol{\beta}$, the solution $\boldsymbol{\beta}^*$ exists.

Lemma 0.2: *Condition (A3) implies $\boldsymbol{\beta}^* \neq 0$.*

Proof. of Lemma 0.2:

Without loss of generality, suppose $\int E[(w_i - w_j)I(w_i - w_j > 0)|\boldsymbol{z}]z_k f^*(\boldsymbol{z})d\boldsymbol{z} > 0$, then for $\beta_k^* > 0$,

$$
\begin{aligned}
L(0,...,\beta_k^*,0,...0) =& \int E[(w_i - w_j)I(w_i - w_j > 0)|\boldsymbol{z}](1 - \beta_k^* z_k)I(1 - \beta_k^* z_k > 0)f^*(\boldsymbol{z})d\boldsymbol{z} \\
=& \int E[(w_i - w_j)I(w_i - w_j > 0)|\boldsymbol{z}]I(z_k < \frac{1}{\beta_k^*})f^*(\boldsymbol{z})d\boldsymbol{z} \\
&-\beta_k^* \int E[(w_i - w_j)I(w_i - w_j > 0)|\boldsymbol{z}]z_k I(z_k < \frac{1}{\beta_k^*})f^*(\boldsymbol{z})d\boldsymbol{z}.
\end{aligned}
$$

The second term is non-negative for a sufficient small $\beta_k^* > 0$.
Therefore,

$$L(0,...,\beta_k^*,0,...0) < \int E[(w_i - w_j)I(w_i - w_j > 0)|\boldsymbol{z}]f^*(\boldsymbol{z})d\boldsymbol{z} = L(0,...,0).$$

Now we prove all lemmas and theorems from Section 3.

Proof of Lemma 1: Recall that

$$\hat{S}(\boldsymbol{\beta}) = \frac{-2}{n(n-1)}\sum_{i=1}^{n}\sum_{j=i+1}^{n}(w_i - w_j)I[1 - (\boldsymbol{X}_i^T - \boldsymbol{X}_j^T)\boldsymbol{\beta} \geq 0](\boldsymbol{X}_i - \boldsymbol{X}_j).$$

Note that

$$\frac{2}{n(n-1)}\sum_{i=1}^{n}\sum_{j=i+1}^{n}(w_i - w_j)I[1 - (\boldsymbol{X}_i^T - \boldsymbol{X}_j^T)\boldsymbol{\beta} \geq 0](X_{ik} - X_{jk}) = \frac{1}{n(n-1)}\sum_{i\neq j} q_1\Big((w_i, \boldsymbol{X}_i), (w_j, \boldsymbol{X}_j)\Big),$$

where

$$q_1\Big((w_i, \boldsymbol{X}_i), (w_j, \boldsymbol{X}_j)\Big) = \begin{cases} (w_i - w_j)I[1 - (\boldsymbol{X}_i^T - \boldsymbol{X}_j^T)\boldsymbol{\beta} \geq 0](\boldsymbol{X}_{ik} - \boldsymbol{X}_{jk}) & if\ i < j, \\ q_1\Big((w_j, \boldsymbol{X}_j), (w_i, \boldsymbol{X}_i)\Big) & if\ i > j. \end{cases}$$

By Hoeffding's inequality for u-statistics of order 2, which can be found in Peel et al. (2010):

$$\begin{aligned} P\Big(\sqrt{32A(\alpha)(\log p)^3/n} \leq & \frac{2}{n(n-1)}\Big|\sum_{i=1}^{n}\sum_{j=i+1}^{n}(w_i - w_j)I[1 - (\boldsymbol{X}_i^T - \boldsymbol{X}_j^T)\boldsymbol{\beta} \geq 0](X_{ik} - X_{jk})\Big|\Big) \\ & \leq 2\exp\Big(-\frac{32A(\alpha)n(\log p)^3}{2n(4M_0w_M)^2}\Big) + n\exp(-\frac{w_M}{M_3}). \end{aligned}$$

Let $w_M = A(\alpha)^{1/3}\log p M_0^{-2/3}M_3^{1/3}$, then

$$\exp\Big(-\frac{32A(\alpha)n(\log p)^3}{32nM_0^2w_M^2}\Big) = \exp(-\frac{w_M}{M_3}),$$

$$\begin{aligned} P\Big(\sqrt{32A(\alpha)(\log p)^3/n} \leq & \frac{2}{n(n-1)}\Big|\sum_{i=1}^{n}\sum_{j=i+1}^{n}(w_i - w_j)I[1 - (\boldsymbol{X}_i^T - \boldsymbol{X}_j^T)\boldsymbol{\beta} \geq 0](X_{ik} - X_{jk})\Big|\Big) \\ & \leq (2+n)p^{-A(\alpha)^{1/3}M_0^{-2/3}M_3^{-2/3}}, \end{aligned}$$

$$\begin{aligned} & P\Big(c\sqrt{32A(\alpha)(\log p)^3/n} \leq c||\hat{S}(\boldsymbol{\beta}^*)||_\infty\Big) \\ \leq & \sum_{k=1}^{p} P\Big(\sqrt{32A(\alpha)(\log p)^3/n} \leq \frac{2}{n(n-1)}\Big|\sum_{i=1}^{n}\sum_{j=i+1}^{n}(w_i - w_j)I[1 - (\boldsymbol{X}_i^T - \boldsymbol{X}_j^T)\boldsymbol{\beta} \geq 0](\boldsymbol{X}_{ik} - \boldsymbol{X}_{jk})\Big|\Big) \\ \leq & (2+n)p^{-A(\alpha)^{1/3}M_0^{-2/3}M_3^{-2/3}+1} \leq \alpha. \end{aligned}$$

Proof of Lemma 2: Let

$$q_2\Big((w_i,\boldsymbol{X}_i),(w_j,\boldsymbol{X}_j)\Big)=\begin{cases}(w_i-w_j)[1-(\boldsymbol{X}_i^T-\boldsymbol{X}_j^T)(\boldsymbol{\beta}^*-h)]_+-(w_i-w_j)[1-(X_i^T-X_j^T)\boldsymbol{\beta}^*]_+ & if\ i<j,\\ q_2\Big((w_j,\boldsymbol{X}_j),(w_i,\boldsymbol{X}_i)\Big) & if\ i>j.\end{cases}$$

and $\hat{U}_{q_2}(\mathcal{W}_n,\boldsymbol{\mathcal{X}}_n)=\frac{1}{n(n-1)}\sum_{i\neq j}q_2\Big((w_i,\boldsymbol{X}_i),(w_j,\boldsymbol{X}_j)\Big)$. It is a u-statistics of order 2.

$$\Big|(w_i-w_j)[1-(\boldsymbol{X}_i^T-\boldsymbol{X}_j^T)(\boldsymbol{\beta}^*-h)]_+-(w_i-w_j)[1-(\boldsymbol{X}_i^T-\boldsymbol{X}_j^T)\boldsymbol{\beta}^*]_+\Big|\leq\Big|2w_M(\boldsymbol{X}_i^T-\boldsymbol{X}_j^T)h\Big|$$

holds with at least probability $1-ne^{-\frac{w_M}{M_3}}$, $\forall\ 1\leq i,j\leq n$.

By Hoeffding's inequality,

$$P\Big(\sup_{h\in R^p:||\boldsymbol{h}||_0\leq q,||\boldsymbol{h}||_2\neq 0}\frac{B(\boldsymbol{h})}{||\boldsymbol{h}||_2}\geq\frac{t}{\sqrt{n}}|\boldsymbol{X}\Big)\leq 2\exp(-\frac{t^2}{32M_0^2w_M^2q})$$

holds with at least probability $1-ne^{-\frac{w_M}{M_3}}$, $\forall\ 1\leq i,j\leq n$. Let $t=C\sqrt{32\log p}qw_M$, then

$$P\Big(\sup_{h\in R^p:||\boldsymbol{h}||_0\leq q,||\boldsymbol{h}||_2\neq 0}\frac{B(\boldsymbol{h})}{||\boldsymbol{h}||_2}\geq C\sqrt{\frac{32\log p}{n}}qw_M\Big)\leq 2p^{-\frac{qC^2}{M_0^2}},$$
$$P\Big(\sup_{\boldsymbol{h}\in N}\frac{B(\boldsymbol{h})}{||\boldsymbol{h}||_2}\geq C\sqrt{\frac{32\log p}{n}}qw_M\Big)\leq 2(\frac{3}{\epsilon}p^{1-\frac{C^2}{M_0^2}})^q,$$

where N is a ϵ-net to cover $\{\boldsymbol{h}\in R^p,||\boldsymbol{h}||_0\leq q,||\boldsymbol{h}||_2\neq 0\}$, for any $\boldsymbol{h}_1$, $\boldsymbol{h}_2$ within the same ϵ-ball and $||\boldsymbol{h}_1||_2\neq 0$ and $||\boldsymbol{h}_2||_2\neq 0$, $\Big|\frac{\boldsymbol{h}_1}{||\boldsymbol{h}_1||_2}-\frac{\boldsymbol{h}_2}{||\boldsymbol{h}_2||_2}\Big|\leq\epsilon$ holds.

We also have

$$\sup_{\boldsymbol{h}_1,\boldsymbol{h}_2\in R^p,||\boldsymbol{h}_1-\boldsymbol{h}_2||_0\leq 2q,||\boldsymbol{h}_1||_2\neq 0,||\boldsymbol{h}_2||_2\neq 0}\Big|\frac{B(\boldsymbol{h}_1)}{||\boldsymbol{h}_1||_2}-\frac{B(\boldsymbol{h}_2)}{||\boldsymbol{h}_2||_2}\Big|\leq\frac{8w_M}{n(n-1)}\Big|\Big|D(\frac{\boldsymbol{h}_1}{||\boldsymbol{h}_1||_2}-\frac{\boldsymbol{h}_2}{||\boldsymbol{h}_2||_2})\Big|\Big|_1$$
$$\leq\frac{8w_M}{\sqrt{\frac{n(n-1)}{2}}}\Big|\Big|D(\frac{\boldsymbol{h}_1}{||\boldsymbol{h}_1||_2}-\frac{\boldsymbol{h}_2}{||\boldsymbol{h}_2||_2})\Big|\Big|_2\leq 8w_M\sqrt{M_1}\epsilon$$

holds with probability at least $1-ne^{-\frac{w_M}{M_3}}$.

$$\sup_{||\boldsymbol{h}||_0\leq q,||\boldsymbol{h}||_2\neq 0}\frac{B(\boldsymbol{h})}{||\boldsymbol{h}||_2}\leq\sup_{\boldsymbol{h}\in N}\frac{B(\boldsymbol{h})}{||\boldsymbol{h}||_2}+8w_M\sqrt{M_1}\epsilon$$

holds with probability at least $1-ne^{-\frac{w_M}{M_3}}$.

Let $\epsilon = q\sqrt{\frac{32\log p}{n}}\frac{1}{8\sqrt{M_1}}$, we have that

$$P\Big(\sup_{||\boldsymbol{h}||_0\le q, ||\boldsymbol{h}||_2\neq 0}\frac{B(\boldsymbol{h})}{||\boldsymbol{h}||_2} \ge Cq\sqrt{\frac{32\log p}{n}}w_M\Big) \le P\Big(\sup_{\boldsymbol{h}\in N}\frac{B(\boldsymbol{h})}{||\boldsymbol{h}||_2} \ge (C-1)q\sqrt{\frac{32\log p}{n}}w_M\Big) + 2n\exp(-\frac{w_M}{M_3}).$$

Since $p > n$ and take $C = 1 + C_2M_0$ for some $C_2 \ge \sqrt{2}$, for sufficiently large n:

$$P\Big(\sup_{||\boldsymbol{h}||_0\le q, ||\boldsymbol{h}||_2\neq 0}\frac{B(\boldsymbol{h})}{||\boldsymbol{h}||_2} \ge (1+C_2M_0)q\sqrt{\frac{32\log p}{n}}w_M\Big) \le 2p^{-q(C_2^2-2)} + \exp(\log 2n - \frac{w_M}{M_3}).$$

Take $w_M = M_3q(C_2^2-2)\log p + M_3\log 2n$, we have that $P\Big(\sup_{||\boldsymbol{h}||_0\le q, ||\boldsymbol{h}||_2\neq 0}\frac{B(\boldsymbol{h})}{||\boldsymbol{h}||_2} \ge (1 + C_2M_0)q\sqrt{\frac{32\log p}{n}}[M_3q(C_2^2-2)\log p + M_3\log 2n]\Big) \le 3p^{-q(C_2^2-2)}$.

Proof of Lemma 3.
By the definition of $\boldsymbol{\beta}^*$, we have

$$\begin{aligned}
&\frac{2}{n(n-1)}\sum_{i=1}^n\sum_{j=i+1}^n (w_i-w_j)[1-(\boldsymbol{X}_i^T-\boldsymbol{X}_j^T)\hat{\boldsymbol{\beta}}]_+ + \lambda||\hat{\boldsymbol{\beta}}||_1 \le \\
&\frac{2}{n(n-1)}\sum_{i=1}^n\sum_{j=i+1}^n (w_i-w_j)[1-(\boldsymbol{X}_i^T-\boldsymbol{X}_j^T)\boldsymbol{\beta}^*]_+ + \lambda||\boldsymbol{\beta}^*||_1 \\
&\frac{2}{n(n-1)}\sum_{i=1}^n\sum_{j=i+1}^n (w_i-w_j)[1-(\boldsymbol{X}_i^T-\boldsymbol{X}_j^T)\hat{\boldsymbol{\beta}}]_+ \\
&-\frac{2}{n(n-1)}\sum_{i=1}^n\sum_{j=i+1}^n (w_i-w_j)[1-(\boldsymbol{X}_i^T-\boldsymbol{X}_j^T)\boldsymbol{\beta}^*]_+ \le \lambda||\boldsymbol{\beta}^*||_1 - \lambda||\hat{\boldsymbol{\beta}}||_1.
\end{aligned} \tag{4}$$

We can also show that

$$||\boldsymbol{\beta}^*||_1 - ||\hat{\boldsymbol{\beta}}||_1 \le ||\boldsymbol{h}_T||_1 - ||\boldsymbol{h}_{T^c}||_1, \tag{5}$$

$$\begin{aligned}
&\frac{2}{n(n-1)}\sum_{i=1}^n\sum_{j=i+1}^n (w_i-w_j)[1-(\boldsymbol{X}_i^T-\boldsymbol{X}_j^T)\hat{\boldsymbol{\beta}}]_+ - \frac{2}{n(n-1)}\sum_{i=1}^n\sum_{j=i+1}^n (w_i-w_j)[1-(\boldsymbol{X}_i^T-\boldsymbol{X}_j^T)\boldsymbol{\beta}^*]_+ \\
&\ge \hat{S}^T(\boldsymbol{\beta}^*)\boldsymbol{h} \ge -||\boldsymbol{h}||_1||\hat{S}(\boldsymbol{\beta}^*)||_\infty \ge -\frac{\lambda}{c}(||\boldsymbol{h}_T||_1 + ||\boldsymbol{h}_{T^c}||_1).
\end{aligned} \tag{6}$$

Based on (4),(6) and (6), we have

$$||\boldsymbol{h}_T||_1 \ge \frac{c-1}{c+1}||\boldsymbol{h}_{T^c}||_1.$$

Proof. of Theorem 4:
Assume $|h_1| \geq |h_2| \geq ... \geq |h_p|$, then for the partition: $S_0 = \{1, 2, .., q\}, S_1 = \{q+1, q+2, .., 2q\}, ...,$

$$||\boldsymbol{h}_{S_0}||_1 \geq ||\boldsymbol{h}_T||_1 \geq \bar{c}||\boldsymbol{h}_{T^C}||_1 \geq \bar{c}||\boldsymbol{h}_{S_0^C}||_1$$

holds. We have

$$\frac{2}{n(n-1)}\sum_{i=1}^{n}\sum_{j=i+1}^{n}(w_i - w_j)[1 - (\boldsymbol{X}_i^T - \boldsymbol{X}_j^T)(\boldsymbol{\beta}^* - \boldsymbol{h})]_+ - \sum_{i=1}^{n}\sum_{j=i+1}^{n}(w_i - w_j)[1 - (\boldsymbol{X}_i^T - \boldsymbol{X}_j^T)\boldsymbol{\beta}^*]_+$$
$$= \frac{2}{n(n-1)}\sum_{i=1}^{n}\sum_{j=i+1}^{n}(w_i - w_j)[1 - (\boldsymbol{X}_i^T - \boldsymbol{X}_j^T)(\boldsymbol{\beta}^* - \boldsymbol{h}_{S_0})]_+ - \sum_{i=1}^{n}\sum_{j=i+1}^{n}(w_i - w_j)[1 - (\boldsymbol{X}_i^T - \boldsymbol{X}_j^T)\boldsymbol{\beta}^*]_+$$
$$+ \frac{2}{n(n-1)}\sum_{l\geq 1}\Big\{\sum_{i=1}^{n}\sum_{j=i+1}^{n}(w_i - w_j)[1 - (\boldsymbol{X}_i^T - \boldsymbol{X}_j^T)(\boldsymbol{\beta}^* - \sum_{k=0}^{l}\boldsymbol{h}_{S_k})]_+$$
$$- \sum_{i=1}^{n}\sum_{j=i+1}^{n}(w_i - w_j)[1 - (\boldsymbol{X}_i^T - \boldsymbol{X}_j^T)(\boldsymbol{\beta}^* - \sum_{k=0}^{l-1}\boldsymbol{h}_{S_k})]_+\Big\}.$$

By Lemma 2 we have with at least probability $1 - 3p^{-q(C_2^2-2)}$,

$$E\Big\{(w_i - w_j)I(w_i - w_j > 0)[1 - (\boldsymbol{X}_i^T - \boldsymbol{X}_j^T)(\boldsymbol{\beta}^* - \sum_{k=1}^{l}\boldsymbol{h}_{S_k})]_+\Big\}$$
$$- E\Big\{(w_i - w_j)I(w_i - w_j > 0)[1 - (\boldsymbol{X}_i^T - \boldsymbol{X}_j^T)(\boldsymbol{\beta}^* - \sum_{k=1}^{l-1}\boldsymbol{h}_{S_k})]_+\Big\}$$
$$\leq \frac{2}{n(n-1)}\sum_{i=1}^{n}\sum_{j=i+1}^{n}(w_i - w_j)[1 - (\boldsymbol{X}_i^T - \boldsymbol{X}_j^T)(\boldsymbol{\beta}^* - \sum_{k=1}^{l}\boldsymbol{h}_{S_k})]_+$$
$$- \frac{2}{n(n-1)}\sum_{i=1}^{n}\sum_{j=i+1}^{n}(w_i - w_j)[1 - (\boldsymbol{X}_i^T - \boldsymbol{X}_j^T)(\boldsymbol{\beta}^* - \sum_{k=1}^{l-1}\boldsymbol{h}_{S_k})]_+ + C_4 q^2\sqrt{\frac{(\log p)^3}{n}}||\boldsymbol{h}_{S_l}||_2.$$

It holds for every l, therefore

$$\frac{2}{n(n-1)}E\Big\{\sum_{i=1}^{n}\sum_{j=i+1}^{n}(w_i-w_j)[1-(\boldsymbol{X}_i^T-\boldsymbol{X}_j^T)(\boldsymbol{\beta}^*-\boldsymbol{h})]_+-\sum_{i=1}^{n}\sum_{j=i+1}^{n}(w_i-w_j)[1-(\boldsymbol{X}_i^T-\boldsymbol{X}_j^T)\boldsymbol{\beta}^*]_+\Big\}$$
$$\leq \frac{2}{n(n-1)}\sum_{i=1}^{n}\sum_{j=i+1}^{n}(w_i-w_j)[1-(\boldsymbol{X}_i^T-\boldsymbol{X}_j^T)(\boldsymbol{\beta}^*-\boldsymbol{h})]_+$$
$$-\frac{2}{n(n-1)}\sum_{i=1}^{n}\sum_{j=i+1}^{n}(w_i-w_j)[1-(\boldsymbol{X}_i^T-\boldsymbol{X}_j^T)\boldsymbol{\beta}^*]_+ + C_4q^2\sqrt{\frac{(\log p)^3}{n}}\sum_{j\geq 0}||\boldsymbol{h}_{S_j}||_2$$
$$\leq \lambda(||\boldsymbol{h}_T||_1-||\boldsymbol{h}_{T^C}||_1)+C_4q^2\sqrt{\frac{(\log p)^3}{n}}||\boldsymbol{h}_{S_0}||_2+\sum_{j\geq 1}C_4q^2\sqrt{\frac{(\log p)^3}{n}}||\boldsymbol{h}_{S_j}||_2$$
$$\leq \lambda\sqrt{q}||\boldsymbol{h}_{S_0}||_2+C_4q^2\sqrt{\frac{(\log p)^3}{n}}(\frac{5}{4}+\frac{1}{\bar{c}})||\boldsymbol{h}_{S_0}||_2$$

holds with probability at least $1-3p^{-q(C_2^2-2)+1}$.

The last inequality holds by,

$$\begin{aligned}\sum_{j\geq 1}||\boldsymbol{h}_{S_j}||_2 \leq & \sum_{j\geq 1}\frac{||\boldsymbol{h}_{S_j}||_1}{\sqrt{q}}+\frac{\sqrt{q}}{4}|h_q| \leq \frac{||\boldsymbol{h}_{S_0^C}||_1}{\sqrt{q}}+\frac{||\boldsymbol{h}_{S_0}||_1}{4\sqrt{q}}\\ \leq & (\frac{1}{\sqrt{q}\bar{c}}+\frac{1}{4\sqrt{q}})||\boldsymbol{h}_{S_0}||_1 \leq (\frac{1}{4}+\frac{1}{\bar{c}})||\boldsymbol{h}_{S_0}||_2.\end{aligned}$$

And more details can be found in Cai et al. (2010).
By Taylor expansion of $L(\boldsymbol{\beta})$ around $\boldsymbol{\beta}^*$:

$$\frac{2}{n(n-1)}E\Big\{\sum_{i=1}^{n}\sum_{j=i+1}^{n}(w_i-w_j)[1-(\boldsymbol{X}_i^T-\boldsymbol{X}_j^T)(\boldsymbol{\beta}^*-\boldsymbol{h})]_+-\sum_{i=1}^{n}\sum_{j=i+1}^{n}(w_i-w_j)[1-(\boldsymbol{X}_i^T-\boldsymbol{X}_j^T)\boldsymbol{\beta}^*]_+\Big\}$$
$$=\frac{1}{2}\boldsymbol{h}^T H(\boldsymbol{\beta}^*)\boldsymbol{h}+o_p(||\boldsymbol{h}||_2^2) \geq \frac{1}{2}M_2||\boldsymbol{h}||_2^2+o_p(||\boldsymbol{h}||_2^2).$$

Then we have

$$\frac{1}{2}M_2||\boldsymbol{h}||_2^2+o_p(||\boldsymbol{h}||_2^2) \leq \lambda\sqrt{q}||\boldsymbol{h}_{S_0}||_2+C_4q^2\sqrt{\frac{(\log p)^3}{n}}(\frac{5}{4}+\frac{1}{\bar{c}})||\boldsymbol{h}_{S_0}||_2.$$

On the other hand,

$$\begin{aligned}||\boldsymbol{h}||_2^2 =&||\mathbf{h}_{S_0}||_2^2 + \sum_{j\geq 1}||\boldsymbol{h}_{S_j}||_2^2 \leq ||\boldsymbol{h}_{S_0}||_2^2 + |h_q|\sum_{j\geq 1}||\boldsymbol{h}_{S_j}||_1\\ \leq&||\boldsymbol{h}_{S_0}||_2^2 + \frac{1}{\bar{c}}||\boldsymbol{h}_{S_0}||_1|h_q| \leq (1+\frac{1}{\bar{c}})||\boldsymbol{h}_{S_0}||_2^2.\end{aligned}$$

Therefore,

$$\begin{aligned}&||\boldsymbol{h}_{S_0}||_2^2 \leq ||\boldsymbol{h}||_2^2 \leq (1+\frac{1}{\bar{c}})||\boldsymbol{h}_{S_0}||_2^2,\\ &o(||\boldsymbol{h}||_2^2) = o(||\boldsymbol{h}_{S_0}||_2^2).\end{aligned}$$

Hence

$$\begin{aligned}&\frac{1}{2}M_2||\boldsymbol{h}_{S_0}||_2 + o_p(||\boldsymbol{h}_{S_0}||_2) \leq \lambda\sqrt{q} + C_4q^2\sqrt{\frac{(\log p)^3}{n}}(\frac{5}{4}+\frac{1}{\bar{c}}),\\ &||\boldsymbol{h}||_2 + o_p(||\boldsymbol{h}||_2) \leq \sqrt{1+\frac{1}{\bar{c}}}[\frac{2\lambda\sqrt{q}}{M_2} + \frac{2C_4}{M_2}q^2\sqrt{\frac{(\log p)^3}{n}}(\frac{5}{4}+\frac{1}{\bar{c}})].\end{aligned}$$

That is,

$$||\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}^*||_2 \leq \sqrt{1+\frac{1}{\bar{c}}}[\frac{2\lambda\sqrt{q}}{M_2} + \frac{2C_4}{M_2}q^2\sqrt{\frac{(\log p)^3}{n}}(\frac{5}{4}+\frac{1}{\bar{c}})]$$

with probability at least $1-3p^{-q(C_2^2-2)+1}$.

Appendix B.

We now show that

$$\begin{aligned}&min\sum_{i\neq j}(w_i - w_j)I(\boldsymbol{\beta}^T\boldsymbol{X}_i < \boldsymbol{\beta}^T\boldsymbol{X}_j),\\ &\text{subject to } ||\boldsymbol{\beta}|| = 1.\end{aligned}$$

is equivalent to

$$\begin{aligned}&min\sum_{w_i>w_j}(w_i - w_j)I(\boldsymbol{\beta}^T\boldsymbol{X}_i < \boldsymbol{\beta}^T\boldsymbol{X}_j),\\ &\text{subject to } ||\boldsymbol{\beta}|| = 1.\end{aligned}$$

Suppose $\boldsymbol{\beta}^T\boldsymbol{X}_i \neq \boldsymbol{\beta}^T\boldsymbol{X}_j, \forall i, j$, we have

$$\sum_{i\neq j}(w_i - w_j)I(\boldsymbol{\beta}^T\boldsymbol{X}_i < \boldsymbol{\beta}^T\boldsymbol{X}_j)$$
$$\sum_{w_i>w_j}(w_i - w_j)I(\boldsymbol{\beta}^T\boldsymbol{X}_i < \boldsymbol{\beta}^T\boldsymbol{X}_j) + \sum_{w_i>w_j}(w_j - w_i)I(\boldsymbol{\beta}^T\boldsymbol{X}_j < \boldsymbol{\beta}^T\boldsymbol{X}_i)$$
$$\sum_{w_i>w_j}(w_i - w_j)[I(\boldsymbol{\beta}^T\boldsymbol{X}_i - \boldsymbol{\beta}^T\boldsymbol{X}_j < 0) - I(\boldsymbol{\beta}^T\boldsymbol{X}_i - \boldsymbol{\beta}^T\boldsymbol{X}_j > 0)]$$
$$\sum_{w_i>w_j}(w_i - w_j)[I(\boldsymbol{\beta}^T\boldsymbol{X}_i - \boldsymbol{\beta}^T\boldsymbol{X}_j < 0) - 1 + I(\boldsymbol{\beta}^T\boldsymbol{X}_i - \boldsymbol{\beta}^T\boldsymbol{X}_j < 0)]$$
$$\sum_{w_i>w_j}(w_i - w_j)[2I(\boldsymbol{\beta}^T\boldsymbol{X}_i - \boldsymbol{\beta}^T\boldsymbol{X}_j < 0) - 1].$$

Appendix C.

We now show that given $\boldsymbol{X}_i$, w_i is an unbiased estimator of $D(\boldsymbol{X}_i)$. Specifically,

$$\begin{aligned} & E\{\frac{[Y_i - \nu(\boldsymbol{X}_i)][A_i - \pi(\boldsymbol{X}_i)]}{\pi(\boldsymbol{X}_i)[1-\pi(\boldsymbol{X}_i)]}|\boldsymbol{X}_i\} \\ = \; & E\left\{\frac{[Y_i - \nu(\boldsymbol{X}_i)][1 - \pi(\boldsymbol{X}_i)]}{\pi(\boldsymbol{X}_i)[1-\pi(\boldsymbol{X}_i)]}|A_i = 1, \boldsymbol{X}_i\right\}\pi(\boldsymbol{X}_i) + E\left\{\frac{[Y_i - \nu(\boldsymbol{X}_i)][-\pi(\boldsymbol{X}_i)]}{\pi(\boldsymbol{X}_i)[1-\pi(\boldsymbol{X}_i)]}|A_i = 0, \boldsymbol{X}_i\right\}[1 - \pi(\boldsymbol{X}_i)] \\ = \; & E[Y_i|A_i = 1, \boldsymbol{X}_i] - E[Y_i|A_i = 0, \boldsymbol{X}_i] = D(\boldsymbol{X}_i). \end{aligned}$$

References

Gnu linear programming kit, version 4.58,. `http://www.gnu.org/software/glpk/glpk.html`, February 2016.

Peter J Bickel, Ya'acov Ritov, and Alexandre B Tsybakov. Simultaneous analysis of lasso and dantzig selector. *The Annals of Statistics*, pages 1705–1732, 2009.

T. Cai, L. Wang, and G. Xu. New bounds for restricted isometry constants. *Information Theory, IEEE Transactions on*, 56(9):4388–4394, 2010.

C. Cavanagh and R. Sherman. Rank estimators for monotonic index models. *Journal of Econometrics*, 84(2):351–381, 1998.

C. Cortes and V. Vapnik. Support-vector networks. *Machine learning*, 20(3):273–297, 1995.

Jim Douglas and Henry H Rachford. On the numerical solution of heat conduction problems in two and three space variables. *Transactions of the American mathematical Society*, 82 (2):421–439, 1956.

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)*, 2016.

J. Fan and R. Li. Variable selection via nonconcave penalized likelihood and its oracle properties. *Journal of the American statistical Association*, 10(1):1348–1360, 2001.

M. Fava, J. Rush, M. H Trivedi, A. Nierenberg, M. Thase, F. Sackeim, H.and Quitkin, S. Wisniewski, P. Lavori, J. Rosenbaum, et al. Background and rationale for the sequenced treatment alternatives to relieve depression (star d) study. *Psychiatric Clinics of North America*, 26(2):457–494, 2003.

G. Gordon. Support vector machines and kernel methods. *Lecture*, 2004.

A. Han. Non-parametric analysis of a generalized regression model: the maximum rank correlation estimator. *Journal of Econometrics*, 35(2):303–316, 1987.

M. Kendall. A new measure of rank correlation. *Biometrika*, 30(1/2):81–93, 1938.

J. Koo, Y. Lee, Y. Kim, and C. Park. A bahadur representation of the linear support vector machine. *J. Mach. Learn. Res.*, 9:1343–1368, June 2008. ISSN 1532-4435. URL `http://dl.acm.org/citation.cfm?id=1390681.1442777`.

W. Lu, H. Zhang, and D. Zeng. Variable selection for optimal treatment decision. *Statistical methods in medical research*, page 0962280211428383, 2011.

G. Michael and B. Stephen. Graph implementations for nonsmooth convex programs. In V. Blondel, S. Boyd, and H. Kimura, editors, *Recent Advances in Learning and Control*, Lecture Notes in Control and Information Sciences, pages 95–110. Springer-Verlag Limited, 2008. `http://stanford.edu/~boyd/graph_dcp.html`.

G. Michael and B. Stephen. CVX: Matlab software for disciplined convex programming, version 2.1. `http://cvxr.com/cvx`, March 2014.

S. Murphy. Optimal dynamic treatment regimes. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 65(2):331–355, 2003.

T. Peel, S. Anthoine, and L. Ralaivola. Empirical bernstein inequalities for u-statistics. In *Advances in Neural Information Processing Systems*, pages 1903–1911, 2010.

B. Peng, L. Wang, and Y. Wu. An error bound for l1-norm support vector machine coefficients in ultra-high dimension. *Manuscript*, 2015.

M. Qian and S. Murphy. Performance guarantees for individualized treatment rules. *Annals of statistics*, 39(2):1180, 2011.

M. Qian, I. Nahum-Shani, and S. Murphy. Dynamic treatment regimes. In *Modern Clinical Trial Analysis*, pages 127–148. Springer, 2012.

D. Rubin. Randomization analysis of experimental data: The fisher randomization test comment. *Journal of the American Statistical Association*, 75(371):591–593, 1980.

R. Song, W. Wang, D. Zeng, and M. Kosorok. Penalized q-learning for dynamic treatment regimens. *Statistica Sinica*, (25):901–920, 2015.

Jonathan E Spingarn. Applications of the method of partial inverses to convex programming: decomposition. *Mathematical Programming*, 32(2):199–223, 1985.

R. Tibshirani. Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society. Series B (Methodological)*, pages 267–288, 1996.

C. Watkins and P. Dayan. Q-learning. *Machine learning*, 8(3-4):279–292, 1992.

B. Zhang, A. Tsiatis, E. Laber, and M. Davidian. A robust method for estimating optimal treatment regimes. *Biometrics*, 68(4):1010–1018, 2012.

X. Zhang, Y. Wu, L. Wang, and R. Li. Variable selection for support vector machines in moderately high dimensions. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 78(1):53–76, 2016.

Y. Zhao, D. Zeng, J. Rush, and M. Kosorok. Estimating individualized treatment rules using outcome weighted learning. *Journal of the American Statistical Association*, 107 (499):1106–1118, 2012.

J. Zhu, S. Rosset, T. Hastie, and R. Tibshirani. 1-norm support vector machines. *Advances in neural information processing systems*, 16(1):49–56, 2004.