High Dimensional A-learning for Optimal Dynamic Treatment Regime

Chengchun Shi

Department of Statistics
North Carolina State University

Joint work with Ailin Fan, Rui Song and Wenbin Lu

March 6, 2016

A few words on causal inference

On dynamic treatment regime

Data

- $A^{(1)}$: first treatment received at time t_1 (0 or 1)
- $S^{(1)}$: patient's baseline covariates prior to t_1
- $A^{(2)}$: second treatment received at time t_2 (0 or 1)
- $S^{(2)}$: intermediate covariates collected between t_1 and t_2
- Y: patient's final outcome (usually the larger the better)

On dynamic treatment regime

Data

- $A^{(1)}$: first treatment received at time t_1 (0 or 1)
- $S^{(1)}$: patient's baseline covariates prior to t_1
- $A^{(2)}$: second treatment received at time t_2 (0 or 1)
- $S^{(2)}$: intermediate covariates collected between t_1 and t_2
- Y: patient's final outcome (usually the larger the better)

Objective

Identify the optimal regime d_1^{opt}, d_2^{opt} to reach the best clinical outcome

- d_1, d_2: Maximize Y

$$d_1 : S^{(1)} \rightarrow \{0, 1\}$$
$$d_2 : (S^{(1)}, A^{(1)}, S^{(2)}) \rightarrow \{0, 1\}$$

Statistical model

Denote X_i, vector of covariates, $[(S_i^{(1)})^T, A_i^{(1)}, (S_i^{(2)})^T]^T$

$$Y_i = h^{(2)}(X_i) + A_i^{(2)} \beta_2^T X_i + \varepsilon_i,$$
$$\mathrm{E}(V_i | S_i, A_i^{(1)}) = h^{(1)}(S_i^{(1)}) + A_i^{(1)} C(S_i^{(1)})$$

where the V-function $V_i = \max_{A_i^{(2)}} Q(X_i, A_i^{(2)})$ and Q-function

$$Q(X_i, A_i^{(2)}) = \mathrm{E}(Y_i | X_i, A_i^{(2)}) = h^{(2)}(X_i) + A_i^{(2)} I(X_i^T \beta_2 > 0).$$

Statistical model

Denote X_i, vector of covariates, $[(S_i^{(1)})^T, A_i^{(1)}, (S_i^{(2)})^T]^T$

$$Y_i = h^{(2)}(X_i) + A_i^{(2)} \beta_2^T X_i + \varepsilon_i,$$
$$\mathrm{E}(V_i | S_i, A_i^{(1)}) = h^{(1)}(S_i^{(1)}) + A_i^{(1)} C(S_i^{(1)})$$

where the V-function $V_i = \max_{A_i^{(2)}} Q(X_i, A_i^{(2)})$ and Q-function

$$Q(X_i, A_i^{(2)}) = \mathrm{E}(Y_i | X_i, A_i^{(2)}) = h^{(2)}(X_i) + A_i^{(2)} I(X_i^T \beta_2 > 0).$$

Optimal treatment regime

- SUTVA, no unmeasured confounders, positivity assumption
- optimal dynamic regime

$$d_2^{opt} = I(X_i^T \beta_2 > 0), \quad d_1^{opt} = I(C(S_i^{(1)}) > 0)$$

Existing literature

- Q-learning (Watkins and Dayan, 1992; Chakraborty et al., 2010)
- A-learning (Murphy, 2003; Robins, 2004)
- Value search method (Zhao et al., 2012; Zhang et al., 2012)

Existing literature

- Q-learning (Watkins and Dayan, 1992; Chakraborty et al., 2010)
- A-learning (Murphy, 2003; Robins, 2004)
- Value search method (Zhao et al., 2012; Zhang et al., 2012)

Notation

- $A^{(j)} = (A_1^{(j)}, \ldots, A_n^{(j)})^T, \;\; j = 1, 2, \qquad Y = (Y_1, \ldots, Y_n)^T,$
- $X = (X_1^T, \ldots, X_n^T)^T, \qquad S = [(S_1^{(1)})^T, \ldots, (S_n^{(1)})^T]^T,$
- $\pi^{(2)}(x) = \Pr(A_i^{(2)} = 1 | X_i = x), \qquad \pi^{(1)}(s) = \Pr(A_i^{(1)} = 1 | S_i = s),$
- $V = (V_1, \ldots, V_n)^T.$

A learning estimating equation

- Estimate β_2:

$$X^T \mathrm{diag}(A^{(2)} - \hat{\pi}^{(2)})[Y - \hat{h}^{(2)} - A^{(2)} \circ (X\hat{\beta}_2)] = 0,$$

A learning estimating equation

- Estimate β_2:

$$X^T \text{diag}(A^{(2)} - \hat{\pi}^{(2)})[Y - \hat{h}^{(2)} - A^{(2)} \circ (X\hat{\beta}_2)] = 0,$$

- Estimate V_i using advantage function (Murphy, 2003):

$$\hat{V}_i = Y_i + X_i^T \hat{\beta}_2 [I(X_i^T \hat{\beta}_2 > 0) - A_i^{(2)}],$$

A learning estimating equation

- Estimate β_2:

$$X^T \text{diag}(A^{(2)} - \hat{\pi}^{(2)})[Y - \hat{h}^{(2)} - A^{(2)} \circ (X\hat{\beta}_2)] = 0,$$

- Estimate V_i using advantage function (Murphy, 2003):

$$\hat{V}_i = Y_i + X_i^T \hat{\beta}_2 [I(X_i^T \hat{\beta}_2 > 0) - A_i^{(2)}],$$

- Estimate β_1:

$$\frac{\partial C(S, \hat{\beta}_1)}{\partial \beta}^T \text{diag}(A^{(1)} - \hat{\pi}^{(1)})[\hat{V} - \hat{h}^{(1)} - A^{(1)} \circ C(S, \hat{\beta}_1)] = 0.$$

A learning estimating equation

- Estimate β_2:

$$X^T \text{diag}(A^{(2)} - \hat{\pi}^{(2)})[Y - \hat{h}^{(2)} - A^{(2)} \circ (X\hat{\beta}_2)] = 0,$$

- Estimate V_i using advantage function (Murphy, 2003):

$$\hat{V}_i = Y_i + X_i^T \hat{\beta}_2 [I(X_i^T \hat{\beta}_2 > 0) - A_i^{(2)}],$$

- Estimate β_1:

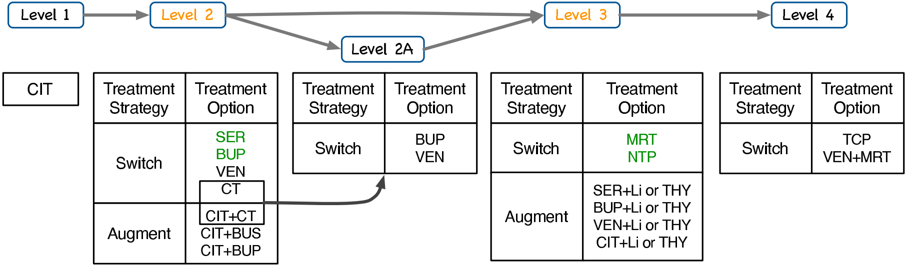
$$\frac{\partial C(S, \hat{\beta}_1)}{\partial \beta}^T \text{diag}(A^{(1)} - \hat{\pi}^{(1)})[\hat{V} - \hat{h}^{(1)} - A^{(1)} \circ C(S, \hat{\beta}_1)] = 0.$$

- Double robustness of $\hat{\beta}_2$:
  consistency either $\pi^{(2)}$ or $h^{(2)}$ is correct

High Dimensional A-learning for Optimal Dynamic Treatment Regime

Motivation

- Sequenced Treatment Alternatives to Relieve Depression (STAR*D)
- Patients with major depression disorder (MDD)
- 4041 patients, 381 covariates available at Level 3, 305 at Level 2
- 73 patients BUP or SER at Level 2, MRT or NTP at Level 3



Penalized A-learning

- Step 1: Estimate $\pi^{(2)}$ and $h^{(2)}$ using nonconcave penalized regression (Fan and Li, 2001; Fan and Lv, 2011)

Penalized A-learning

- Step 1: Estimate $\pi^{(2)}$ and $h^{(2)}$ using nonconcave penalized regression (Fan and Li, 2001; Fan and Lv, 2011)
- Step 2: Estimate β_2 using penalized A-learning estimating equation

Penalized A-learning

- Step 1: Estimate $\pi^{(2)}$ and $h^{(2)}$ using nonconcave penalized regression (Fan and Li, 2001; Fan and Lv, 2011)
- Step 2: Estimate β_2 using penalized A-learning estimating equation
- Step 3: Move backward to estimate V_i using advantage function

Penalized A-learning

- Step 1: Estimate $\pi^{(2)}$ and $h^{(2)}$ using nonconcave penalized regression (Fan and Li, 2001; Fan and Lv, 2011)
- Step 2: Estimate β_2 using penalized A-learning estimating equation
- Step 3: Move backward to estimate V_i using advantage function
- Step 4: Estimate $\pi^{(1)}$ and $h^{(1)}$ using nonconcave penalized regression

Penalized A-learning

- Step 1: Estimate $\pi^{(2)}$ and $h^{(2)}$ using nonconcave penalized regression (Fan and Li, 2001; Fan and Lv, 2011)
- Step 2: Estimate β_2 using penalized A-learning estimating equation
- Step 3: Move backward to estimate V_i using advantage function
- Step 4: Estimate $\pi^{(1)}$ and $h^{(1)}$ using nonconcave penalized regression
- Step 5: Estimate β_1 using penalized A-learning estimating equation

Step 1

- Logistic model for $\pi^{(2)}$ and linear model for $h^{(2)}$

$$\pi^{(2)}(x) = \frac{\exp(x^T \alpha_2)}{1 + \exp(x^T \alpha_2)}, \quad h^{(2)}(x) = x^T \theta_2,$$

Step 1

- Logistic model for $\pi^{(2)}$ and linear model for $h^{(2)}$

$$\pi^{(2)}(x) = \frac{\exp(x^T\alpha_2)}{1+\exp(x^T\alpha_2)}, \quad h^{(2)}(x) = x^T\theta_2,$$

- Estimate α_2 and θ_2 using non-concave penalized regression

$$\begin{aligned}\hat{\alpha}_2 &= \arg\min_{\alpha_2\in\mathbb{R}^p} \frac{1}{n}\sum_{i=1}^{n}[\log\{1+\exp(X_i^T\alpha_2)\} - A_i^{(2)}X_i^T\alpha_2] \\ &+ \sum_{j=1}^{p}\rho_1^{(2)}(|\alpha_2^j|, \lambda_{1n}^{(2)}), \\ \hat{\theta}_2 &= \arg\min_{\theta_2\in\mathbb{R}^p} \frac{1}{n}\sum_{i=1}^{n}(1-A_i^{(2)})(Y_i - X_i^T\theta_2)^2 + \sum_{j=1}^{p}\rho_2^{(2)}(|\theta_2^j|, \lambda_{2n}^{(2)})\end{aligned}$$

Step 2

Add Dantzig selector (Candès and Tao, 2007) on A-learning equation

$$\hat{\beta}_2 = \arg\min_{\beta_2 \in \Lambda^{(2)}} ||\beta_2||_1,$$

where

$$\Lambda^{(2)} = \left\{ \beta_2 : ||\frac{1}{n} X^T \text{diag}(A^{(2)} - \hat{\pi}^{(2)})\{Y - X\hat{\theta}_2 - A^{(2)} \circ (X\beta_2)\}||_\infty \leq \lambda_{3n}^{(2)} \right\},$$

Step 2

Add Dantzig selector (Candès and Tao, 2007) on A-learning equation

$$\hat{\beta}_2 = \arg \min_{\beta_2 \in \Lambda^{(2)}} ||\beta_2||_1,$$

where

$$\Lambda^{(2)} = \left\{ \beta_2 : ||\frac{1}{n} X^T \text{diag}(A^{(2)} - \hat{\pi}^{(2)})\{Y - X\hat{\theta}_2 - A^{(2)} \circ (X\beta_2)\}||_\infty \leq \lambda_{3n}^{(2)} \right\},$$

Step 3

Estimate V-function

$$\hat{V}_i = Y_i + X_i^T \hat{\beta}_2 \{I(X_i^T \hat{\beta}_2 > 0) - A_i^{(2)}\}.$$

Step 4

- Logistic model for $\pi^{(1)}$ and linear model for $h^{(1)}$

$$\pi^{(1)}(s) = \frac{\exp(s^T\alpha_1)}{1+\exp(s^T\alpha_1)}, \quad h^{(1)}(s) = s^T\theta_1,$$

Step 4

- Logistic model for $\pi^{(1)}$ and linear model for $h^{(1)}$

$$\pi^{(1)}(s) = \frac{\exp(s^T\alpha_1)}{1+\exp(s^T\alpha_1)}, \quad h^{(1)}(s) = s^T\theta_1,$$

- Estimate α_1 and θ_1:

$$\begin{aligned} \hat{\alpha}_1 &= \arg\min_{\alpha_1 \in \mathbb{R}^q} \frac{1}{n}\sum_{i=1}^{n}[\log\{1+\exp(S_i^T\alpha_1)\} - A_i^{(1)} S_i^T \alpha_1] \\ &+ \sum_{j=1}^{q} \rho_1^{(1)}(|\alpha_1^j|, \lambda_{1n}^{(1)}), \\ \hat{\theta}_1 &= \arg\min_{\theta_1 \in \mathbb{R}^q} \frac{1}{n}\sum_{i=1}^{n}(1-A_i^{(1)})(\hat{V}_i - S_i^T\theta_1)^2 + \sum_{j=1}^{q} \rho_2^{(1)}(|\theta_1^j|, \lambda_{2n}^{(1)}), \end{aligned}$$

Finally...

- A linear model for $C(s) = s^T \beta_1$,

Finally...

- A linear model for $C(s) = s^T \beta_1$,
- Estimate β_1:

$$\hat{\beta}_1 = \arg \min_{\beta_1 \in \Lambda^{(1)}} ||\beta_1||_1,$$

where

$$\Lambda^{(1)} = \left\{ \beta_1 : ||\frac{1}{n} S^T \text{diag}(A^{(1)} - \hat{\pi}^{(1)}) \{ \hat{V} - S\hat{\theta}_1 - A^{(1)} \circ (S\beta_1) \} ||_\infty \leq \lambda_{3n}^{(1)}$$

Finally...

- A linear model for $C(s) = s^T \beta_1$,
- Estimate β_1:

$$\hat{\beta}_1 = \arg \min_{\beta_1 \in \Lambda^{(1)}} ||\beta_1||_1,$$

where

$$\Lambda^{(1)} = \left\{ \beta_1 : || \frac{1}{n} S^T \text{diag}(A^{(1)} - \hat{\pi}^{(1)}) \{ \hat{V} - S\hat{\theta}_1 - A^{(1)} \circ (S\beta_1) \} ||_\infty \leq \lambda_{3n}^{(1)}$$

Estimated optimal regime

$$\hat{d}_1(S_i) = I(\hat{\beta}_1^T S_i > 0) \quad \text{and} \quad \hat{d}_2(X_i) = I(\hat{\beta}_2^T X_i > 0).$$

Theoretical performance guarantees

A non-asymptotic upper bound for difference of value function

$$\mathrm{E}Y_0^\star(d_1^{opt}, d_2^{opt}) - \mathrm{E}Y_0^\star(\hat{d}_1, \hat{d}_2),$$

$$\mathrm{E}Y_0^\star(d_1, d_2) = \mathrm{E}[Y_0 + X_0^T \beta_2 (d_2 - A_0^{(2)}) + C(S_0^{(1)})(d_1 - A_0^{(1)})].$$

Theoretical performance guarantees

A non-asymptotic upper bound for difference of value function

$$\mathrm{E}Y_0^{\star}(d_1^{opt}, d_2^{opt}) - \mathrm{E}Y_0^{\star}(\hat{d}_1, \hat{d}_2),$$

$$\mathrm{E}Y_0^{\star}(d_1, d_2) = \mathrm{E}[Y_0 + X_0^T \beta_2 (d_2 - A_0^{(2)}) + C(S_0^{(1)})(d_1 - A_0^{(1)})].$$

How to get there

- Step 1: Weak oracle non-asymptotic bound for $\hat{\alpha}_2$, $\hat{\theta}_2$

Theoretical performance guarantees

A non-asymptotic upper bound for difference of value function

$$\mathrm{E}Y_0^{\star}(d_1^{opt}, d_2^{opt}) - \mathrm{E}Y_0^{\star}(\hat{d}_1, \hat{d}_2),$$

$$\mathrm{E}Y_0^{\star}(d_1, d_2) = \mathrm{E}[Y_0 + X_0^T\beta_2(d_2 - A_0^{(2)}) + C(S_0^{(1)})(d_1 - A_0^{(1)})].$$

How to get there

- Step 1: Weak oracle non-asymptotic bound for $\hat{\alpha}_2$, $\hat{\theta}_2$
- Step 2: Error bound for $||\hat{\beta}_2 - \beta_2||_2$

Theoretical performance guarantees

A non-asymptotic upper bound for difference of value function

$$\mathrm{E}Y_0^{\star}(d_1^{opt}, d_2^{opt}) - \mathrm{E}Y_0^{\star}(\hat{d}_1, \hat{d}_2),$$

$$\mathrm{E}Y_0^{\star}(d_1, d_2) = \mathrm{E}[Y_0 + X_0^T\beta_2(d_2 - A_0^{(2)}) + C(S_0^{(1)})(d_1 - A_0^{(1)})].$$

How to get there

- Step 1: Weak oracle non-asymptotic bound for $\hat{\alpha}_2$, $\hat{\theta}_2$
- Step 2: Error bound for $||\hat{\beta}_2 - \beta_2||_2$
- Step 3: Weak oracle non-asymptotic bound for $\hat{\alpha}_1$, $\hat{\theta}_1$

Theoretical performance guarantees

A non-asymptotic upper bound for difference of value function

$$\mathrm{E} Y_0^\star(d_1^{opt}, d_2^{opt}) - \mathrm{E} Y_0^\star(\hat{d}_1, \hat{d}_2),$$

$$\mathrm{E} Y_0^\star(d_1, d_2) = \mathrm{E}[Y_0 + X_0^T \beta_2(d_2 - A_0^{(2)}) + C(S_0^{(1)})(d_1 - A_0^{(1)})].$$

How to get there

- Step 1: Weak oracle non-asymptotic bound for $\hat{\alpha}_2$, $\hat{\theta}_2$
- Step 2: Error bound for $||\hat{\beta}_2 - \beta_2||_2$
- Step 3: Weak oracle non-asymptotic bound for $\hat{\alpha}_1$, $\hat{\theta}_1$
- Step 4: Error bound for $||\hat{\beta}_1 - \beta_1^\star||_2$

Theoretical performance guarantees

A non-asymptotic upper bound for difference of value function

$$\mathrm{E}Y_0^\star(d_1^{opt}, d_2^{opt}) - \mathrm{E}Y_0^\star(\hat{d}_1, \hat{d}_2),$$

$$\mathrm{E}Y_0^\star(d_1, d_2) = \mathrm{E}[Y_0 + X_0^T \beta_2(d_2 - A_0^{(2)}) + C(S_0^{(1)})(d_1 - A_0^{(1)})].$$

How to get there

- Step 1: Weak oracle non-asymptotic bound for $\hat{\alpha}_2$, $\hat{\theta}_2$
- Step 2: Error bound for $||\hat{\beta}_2 - \beta_2||_2$
- Step 3: Weak oracle non-asymptotic bound for $\hat{\alpha}_1$, $\hat{\theta}_1$
- Step 4: Error bound for $||\hat{\beta}_1 - \beta_1^\star||_2$
- Step 5: Upper bound for $\mathrm{E}Y_0^\star(d_1^{opt}, d_2^{opt}) - \mathrm{E}Y_0^\star(\hat{d}_1, \hat{d}_2)$

Some technical challenges

- Deal with NP-dimensionality ($\log p = O(n^a), 0 < a < 1$)

Some technical challenges

- Deal with NP-dimensionality ($\log p = O(n^a), 0 < a < 1$)
- Dantzig selector rarely studied in random design

Some technical challenges

- Deal with NP-dimensionality ($\log p = O(n^a), 0 < a < 1$)
- Dantzig selector rarely studied in random design
- Nonconcave penalized regression never studied in random design

Some technical challenges

- Deal with NP-dimensionality ($\log p = O(n^a), 0 < a < 1$)
- Dantzig selector rarely studied in random design
- Nonconcave penalized regression never studied in random design
- Nonconcave penalized regression not well studied in the present of model misspecification, even in fixed design

Some technical challenges

- Deal with NP-dimensionality ($\log p = O(n^a), 0 < a < 1$)
- Dantzig selector rarely studied in random design
- Nonconcave penalized regression never studied in random design
- Nonconcave penalized regression not well studied in the present of model misspecification, even in fixed design
- Deal with model misspecification back to the first stage

Some technical challenges

- Deal with NP-dimensionality ($\log p = O(n^a), 0 < a < 1$)
- Dantzig selector rarely studied in random design
- Nonconcave penalized regression never studied in random design
- Nonconcave penalized regression not well studied in the present of model misspecification, even in fixed design
- Deal with model misspecification back to the first stage

Restricted eigenvalue (RE) condition in penalized A-learning

- Linear models: RE on $X^T X$

Some technical challenges

- Deal with NP-dimensionality ($\log p = O(n^a), 0 < a < 1$)
- Dantzig selector rarely studied in random design
- Nonconcave penalized regression never studied in random design
- Nonconcave penalized regression not well studied in the present of model misspecification, even in fixed design
- Deal with model misspecification back to the first stage

Restricted eigenvalue (RE) condition in penalized A-learning

- Linear models: RE on $X^T X$
- In our setting: RE on $X^T \text{diag}(A - \hat{\pi})X$
- Substantiate difficulty due to the plug-in estimator $\hat{\pi}$!

Nonconcave penalized regression in random design

- Need to establish concentration inequality for random variable and random matrix
- For example, need the following regularity condition:

$$\max_{j=1}^{p} \lambda_{\max}[X_M^T \text{diag}(|X^j|) X_M] = O(n),$$

for some $M \subseteq [1, 2, \ldots, p]$.

Nonconcave penalized regression in random design

- Need to establish concentration inequality for random variable and random matrix
- For example, need the following regularity condition:

$$\max_{j=1}^{p} \lambda_{\max}[X_M^T \text{diag}(|X^j|) X_M] = O(n),$$

for some $M \subseteq [1, 2, \ldots, p]$.

Model misspecification and least false parameter $\beta_1^\star$

$$\beta_1^\star = \arg\min_{\beta_1 \in \Lambda^*} ||\beta_1||_1,$$

where

$$\Lambda^* = \left\{ \beta_1 \in \mathbb{R}^q : ||\mathrm{E}[S_i A_i (1 - \pi_i^{(1)}) \{ C(S_i) - S_i^T \beta_1 \}]||_\infty \leq \kappa_0 \right\}.$$

Weak oracle property in the presence of model misspecification

- $\hat{\alpha} \to \alpha^\star$, $\hat{\theta} \to \theta^\star$ (omit the superscript (2))

Weak oracle property in the presence of model misspecification

- $\hat{\alpha} \to \alpha^\star$, $\hat{\theta} \to \theta^\star$ (omit the superscript (2))
- when π or h is correct, $\alpha^\star$, $\theta^\star$ the true parameter

Weak oracle property in the presence of model misspecification

- $\hat{\alpha} \to \alpha^\star$, $\hat{\theta} \to \theta^\star$ (omit the superscript (2))
- when π or h is correct, $\alpha^\star$, $\theta^\star$ the true parameter
- when π or h is misspecified, $\alpha^\star$, $\theta^\star$ some least false parameter

Weak oracle property in the presence of model misspecification

- $\hat{\alpha} \to \alpha^\star$, $\hat{\theta} \to \theta^\star$ (omit the superscript (2))
- when π or h is correct, $\alpha^\star$, $\theta^\star$ the true parameter
- when π or h is misspecified, $\alpha^\star$, $\theta^\star$ some least false parameter
- $M_\alpha = \text{supp}(\alpha^\star)$, $M_\theta = \text{supp}(\theta^\star)$, $s_\alpha = |M_\alpha| = O(n^{l_4})$, $s_\theta = |M_\theta| = O(n^{l_5})$, l_1, $l_2 \in (0, 1/2)$, $\log p = O(n^{a_2})$, $a_2 \in (0, 1)$,

Weak oracle property in the presence of model misspecification

- $\hat{\alpha} \to \alpha^\star$, $\hat{\theta} \to \theta^\star$ (omit the superscript (2))
- when π or h is correct, $\alpha^\star$, $\theta^\star$ the true parameter
- when π or h is misspecified, $\alpha^\star$, $\theta^\star$ some least false parameter
- $M_\alpha = \text{supp}(\alpha^\star)$, $M_\theta = \text{supp}(\theta^\star)$, $s_\alpha = |M_\alpha| = O(n^{l_4})$, $s_\theta = |M_\theta| = O(n^{l_5})$, $l_1, l_2 \in (0, 1/2)$, $\log p = O(n^{a_2})$, $a_2 \in (0, 1)$,

Theorem (Weak oracle property of $\hat{\alpha}_2$ and $\hat{\theta}_2$)

Under certain conditions, there exists some constant $\bar{c}$, $\gamma_{\alpha_2} \in (0, 1/2]$, $\gamma_{\theta_2} \in (0, 1/2]$, *such that with prob. at least* $1 - \bar{c}/(n+p)$,

- $\hat{\alpha}_{M_\alpha^c} = 0$, $\hat{\theta}_{M_\theta^c} = 0$,
- $||\hat{\alpha}_{M_\alpha} - \alpha^\star_{M_\alpha}||_\infty = O(n^{-\gamma_{\alpha_2}} \log n)$, $||\hat{\theta}_{M_\theta} - \theta^\star_{M_\theta}||_\infty = O(n^{-\gamma_{\theta_2}} \log n)$

Theorem (Error bound for $\hat{\beta}_2$)

Under certain conditions, if $\lambda_{3n}^{(2)} = E_1 + E_2 + E_3 + E_4$ *defined below, then as long as either* $\pi^{(2)}$ *or* $h^{(2)}$ *is correct, then for any fixed* $0 < \theta_s < 1$*, with prob. at least* $1 - \bar{c}/(n+p)$ *for some constant* $\bar{c}$*,*

$$||\hat{\beta}_2 - \beta_2||_2 \le \frac{12\lambda_{3n}^{(2)}\sqrt{s_{\beta_2}}}{(1-\theta_s)\inf_{\alpha_2 \in H_{\alpha_2}} K^2(s_{\beta_2}, 1, \Omega^{(2)}(\alpha_2))},$$

where

$$E_1 = O(\sqrt{\log p/n}), E_2 = O(s_{\alpha_2} n^{-2\gamma_{\alpha_2}} \log^2 n + s_{\theta_2} n^{-2\gamma_{\theta_2}} \log^2 n),$$
$$E_3 = O(\sigma_3(\sqrt{s_{\alpha_2}\log n/n} + \sqrt{s_{\alpha_2}}\lambda_{1n}^{(2)}\rho_1^{(2)}(d_{n\alpha_2}))),$$
$$E_4 = O(\sigma_4(\sqrt{s_{\theta_2}\log n/n} + \sqrt{s_{\theta_2}}\lambda_{2n}^{(2)}\rho_2^{(2)}(d_{n\theta_2}))),$$

and $\sigma_3^2 = E[h_2(X_i) - X_i^T\theta_2^\star]^2$, $\sigma_4^2 = E[\pi^{(2)}(X_i) - \pi_i^{2*}]^2$.

Theorem (Weak oracle property of $\hat{\alpha}_1$ and $\hat{\theta}_1$)

Under certain regularity condition, there exists some $\gamma_{\alpha_1}, \gamma_{\theta_1} \in (0, 1/2]$, *with probability at least* $1 - \bar{c}/(n+q+p)$ *for some constant* $\bar{c}$, *the estimators* $\hat{\alpha}_1$ *and* $\hat{\theta}_1$ *must satisfy*

- $\hat{\alpha}_1^{M_{\alpha_1}^c} = 0$, $\hat{\theta}_1^{M_{\theta_1}^c} = 0$,
- $||\hat{\alpha}_1^{M_{\alpha_1}} - \alpha_1^{\star M_{\alpha_1}}||_\infty = O(n^{-\gamma_{\alpha_1}} \log n)$,
  $||\hat{\theta}_1^{M_{\theta_1}} - \theta_1^{\star M_{\theta_1}}||_\infty = O(n^{-\gamma_{\theta_1}} \log n)$.

Theorem (Error bound for $\hat{\beta}_1$)

Assume $\lambda_{3n}^{(1)} = \sum_{k=5}^{10} E_k$ *defined below. If either* $\pi^{(1)}$ *or* $h^{(1)}$ *is correctly specified, then there exists a constant* $\bar{c}$*, such that for sufficiently large* n *and some fixed* $0 < \theta_s < 1$*, with probability at least* $1 - \bar{c}/(n+p+q)$*,*

$$E_5 = O(\sqrt{\log q}\log^2 n/n), \quad E_6 = O(s_{\alpha_1} n^{-2\gamma_{\alpha_1}} \log^2 n + s_{\theta_1} n^{-2\gamma_{\theta_1}} \log^2 n),$$

$$E_7 = O\{\sigma_1(\sqrt{s_{\alpha_1}\log n/n} + \sqrt{s_{\alpha_1}}\lambda_{1n}^{(1)}\rho_1^{(1)}(d_{n\alpha_1}))\},$$

$$E_8 = O\{\sigma_2(\sqrt{s_{\theta_1}\log n/n} + \sqrt{s_{\theta_1}}\lambda_{2n}^{(1)}\rho_2^{(1)}(d_{n\theta_1}))\},$$

$$E_9 = O\{\sigma_0(\sqrt{s_{\alpha_1}\log n/n} + \sqrt{s_{\alpha_1}}\lambda_{1n}^{(1)}\rho_1^{(1)}(d_{n\alpha_1})) + \tau_0 + \kappa_0^*\},$$

$$E_{10} = O(n^{\mu_1}\log n),$$

where $\sigma_0^2 = E\{C(S_i) - S_i^T \beta_1^\star\}^2$, $\sigma_1^2 = E(h^{(1)} - S_i^T\theta_1^\star)^2$*, and* $\sigma_2^2 = E\{\pi_i^{(1)*} - \pi^{(1)}(S_i)\}^2$.

Theorem (Error bound for $EY_0^\star(d_1^{opt}, d_2^{opt}) - EY_0^\star(\hat{d}_1, \hat{d}_2)$)

Under certain conditions, if the probability density function of $S_0^T \beta_1^\star$ exists and is bounded. For some fixed $0 < \theta_s < 1$ and sufficiently large n, there exists some constants $\bar{c}$, c_1 and c_2 such that

$$
\begin{aligned}
&0 \le EY_0^\star(d_1^{opt}, d_2^{opt}) - EY_0^\star(\hat{d}_1, \hat{d}_2) \le \\
&\bar{c}\sigma_0^{4/3} + \frac{\bar{c}\omega}{n}\sqrt{\lambda_{\max}(\Sigma_{M_{\beta_2} M_{\beta_2}})}||\beta_2||_2 + \frac{\bar{c}\zeta}{n}\sqrt{\lambda_{\max}(\Sigma_{M_{\beta_1} M_{\beta_1}})}||\beta_1^\star||_2 \\
&+\frac{c_1\omega^2\rho_{\max}^{s_{\beta_2}}(\Sigma){\lambda_{3n}^{(2)}}^2 s_{\beta_2}\log^2 n}{(1-\theta_s)^2 \inf\limits_{\alpha_1 \in H_{\alpha_1}} K^4(s,1,\Omega^{(2)}(\alpha_1))} + \frac{c_2\zeta^2\rho_{\max}^{s_{\beta_1}}(\Psi){\lambda_{3n}^{(1)}}^2 s_{\beta_1}\log^2 n}{(1-\theta_s)^2 \inf\limits_{\alpha_1 \in H_{\alpha_1}} K^4(s,1,\Omega^{(1)}(\alpha_1))}
\end{aligned}
$$

where

$$\sigma_0^2 = E[\{C(S_i) - S_i^T \beta_1^\star\}^2].$$

STAR*D study

- Consider patients receiving BUP or SER at Level 2, and randomized to MIRT or NTP at Level 3.
- 73 patients that had complete record of 381 covariates at Level 3
- Penalized A-learning: 3 variables at Level 2, 3 variables at Level 3
- Examination of the method

$$V = \frac{1}{n}\sum_{i=1}^{n}\left[Y_i + X_i^T\hat{\beta}_2\{d_2(X_i) - A_i^{(2)}\} + S_i^T\hat{\beta}_1\{d_1(S_i) - A_i^{(1)}\}\right].$$

Table: Estimated Values of Different Treatment Regimes and CIs

Treatment Regime	Estimated Value	95% CI on Diff
estimated optimal regime	-10.04	
BUP + NTP	-13.41	[0.95,7.14]
BUP + MIRT	-12.75	[0.62,5.96]
SER + NTP	-12.63	[0.34,6.50]
SER + MIRT	-11.97	[0.25,4.70]

Reference I

Candès, E. and Tao, T. (2007). Rejoinder: "The Dantzig selector: statistical estimation when p is much larger than n" [Ann. Statist. **35** (2007), no. 6, 2313–2351; mr2382644]. *Ann. Statist.*, 35(6):2392–2404.

Chakraborty, B., Murphy, S., and Strecher, V. (2010). Inference for non-regular parameters in optimal dynamic treatment regimes. *Stat. Methods Med. Res.*, 19(3):317–343.

Fan, J. and Li, R. (2001). Variable selection via nonconcave penalized likelihood and its oracle properties. *J. Amer. Statist. Assoc.*, 96(456):1348–1360.

Fan, J. and Lv, J. (2011). Nonconcave penalized likelihood with NP-dimensionality. *IEEE Trans. Inform. Theory*, 57(8):5467–5484.

Murphy, S. A. (2003). Optimal dynamic treatment regimes. *J. R. Stat. Soc. Ser. B Stat. Methodol.*, 65(2):331–366.

Reference II

Robins, J. M. (2004). Optimal structural nested models for optimal sequential decisions. In *Proceedings of the Second Seattle Symposium in Biostatistics*, volume 179 of *Lecture Notes in Statist.*, pages 189–326. Springer, New York.

Watkins, C. and Dayan, P. (1992). Q-learning. *Mach. Learn.*, 8:279–292.

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

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