Localizing Changes in High-Dimensional Regression Models: Supplementary Materials

1 Proof of Theorem 1

1.1 Sketch of the Proofs

In this subsection, we first sketch the proof of Theorem 1, which serves as a general template to derive upper bounds on the localization error change point problems in the general regression framework described in Model 1.

Theorem 1 is an immediate consequence of Propositions 1 and 2.

Proposition 1. *Under the same conditions in Theorem 1 and letting $\widehat{\mathcal{P}}$ being the solution to* (1)*, the following hold with probability at least $1 - C(n \vee p)^{-c}$.*

(i) For each interval $\widehat{I} = (s, e] \in \widehat{\mathcal{P}}$ containing one and only one true change point η, it must be the case that

$$\min\{e - \eta, \eta - s\} \leq C_\epsilon \left(\frac{d_0 \lambda^2 + \gamma}{\kappa^2} \right),$$

where $C_\epsilon > 0$ is an absolute constant;

(ii) for each interval $\widehat{I} = (s, e] \in \widehat{\mathcal{P}}$ containing exactly two true change points, say $\eta_1 < \eta_2$, it must be the case that

$$\max\{e - \eta_2, \eta_1 - s\} \leq C_\epsilon \left(\frac{d_0 \lambda^2 + \gamma}{\kappa^2} \right),$$

where $C_\epsilon > 0$ is an absolute constant;

(iii) for all consecutive intervals $\widehat{I}$ and $\widehat{J}$ in $\widehat{P}$, the interval $\widehat{I} \cup \widehat{J}$ contains at least one true change point; and

(iv) no interval $\widehat{I} \in \widehat{\mathcal{P}}$ contains strictly more than two true change points.

Proposition 2. *Under the same conditions in Theorem 1, with $\widehat{\mathcal{P}}$ being the solution to* (1)*, satisfying $K \leq |\widehat{\mathcal{P}}| \leq 3K$, then with probability at least $1 - C(n \vee p)^{-c}$, it holds that $|\widehat{\mathcal{P}}| = K$.*

Proof of Theorem 1. It follows from Proposition 1 that, $K \leq |\widehat{\mathcal{P}}| \leq 3K$. This combined with Proposition 2 completes the proof. $\square$

The key ingredients of the proofs of both Propositions 1 and 2 are two types of deviation inequalities.

- **Restricted eigenvalues.** In the literature on sparse regression, there are several versions of the restricted eigenvalue conditions (see, e.g. Bühlmann & van de Geer, 2011). In our analysis, such conditions amount to controlling the probability of the event

$$\mathcal{E}_I = \left\{ \sqrt{\sum_{t \in I} \left(x_t^\top v\right)^2} \geq \frac{c_x \sqrt{|I|}}{4} \|v\|_2 - 9 C_x \sqrt{\log(p)} \|v\|_1, \quad \forall v \in \mathbb{R}^p \right\},$$

which is done in Lemma 3.

- **Deviations bounds of scaled noise.** In addition, we need to control the deviations of the quantities of the form
$$\left\|\sum_{t \in I} \varepsilon_t x_t\right\|_\infty. \tag{1}$$
See Lemma 4.

In standard analyses of the performance of the Lasso estimator, as detailed e.g. in Section 6.2 of Bühlmann & van de Geer (2011), the combination of restricted eigenvalues conditions and large probability bounds on the noise lead to oracle inequalities for the estimation and prediction errors in situations in which there exists no change point and the data are independent. We have extended this line of arguments to the present, more challenging settings, to derive analogous oracle inequalities. We emphasize a few points in this regard.

- In standard analyses of the Lasso estimator, where there is one and only one true coefficient vector, the magnitude of λ is determined as a high-probability upper bound to (1). However in our situation, in order to control the ℓ_1- and ℓ_2-loss of the estimators $\widehat{\beta}_I^\lambda$, where the interval I contains more than one true coefficient vectors, the value of λ needs to be inflated by a factor of $\sqrt{d_0}$. This is detailed in Lemma 7; see, in particular, (12).
- The magnitude of the tuning parameter γ is determined based on an appropriate oracle inequality for the Lasso and on the number of true change points; more precisely, γ can be derived as a high-probability bound for
$$\left|\sum_{t \in I}\left\{(y_t - x_t^\top \widehat{\beta}_I^\lambda)^2 - (y_t - x_t^\top \beta_t^*)^2\right\}\right|.$$
See Lemma 6 for details.

  The fact that γ is linear in the number of change point K is to prompt the consistency. This is shown in (32) in the proof of Proposition 2.
- The final localization error is obtained by the following calculations. Assume that there exists one and only one true change point $\eta \in I = (s, e]$. Define $I_1 = (s, \eta]$ and $I_2 = (\eta_1, e]$. Let $\beta_{I_1}^*$ and $\beta_{I_2}^*$ be the two true coefficient vectors in I_1 and I_2, respectively. For readability, below we will omit all constants here and use the symbol $\lessapprox$ to denote an inequality up to hidden universal constants. We first assume by contradiction that
$$\min\{|I_1|, |I_2|\} \gtrapprox d_0 \log(n \vee p), \tag{2}$$
then use oracle inequalities to establish that
$$\begin{aligned}
&\sum_{t \in I_1}\{x_t^\top(\widehat{\beta}_I^\lambda - \beta_{I_1}^*)\}^2 + \sum_{t \in I_2}\{x_t^\top(\widehat{\beta}_I^\lambda - \beta_{I_2}^*)\}^2 \\
\lessapprox & \lambda\sqrt{\max\{|I_1|, \log(n \vee p)\}}\left\{\sqrt{d_0}\|(\widehat{\beta}_I^\lambda - \beta_{I_1}^*)(S)\|_2 + \|\widehat{\beta}_I^\lambda(S^c)\|_1\right\} \\
& \qquad + \lambda\sqrt{\max\{|I_2|, \log(n \vee p)\}}\left\{\sqrt{d_0}\|(\widehat{\beta}_I^\lambda - \beta_{I_2}^*)(S)\|_2 + \|\widehat{\beta}_I^\lambda(S^c)\|_1\right\} + \gamma \\
\lessapprox & \lambda\sqrt{|I_1|}\left\{\sqrt{d_0}\|(\widehat{\beta}_I^\lambda - \beta_{I_1}^*)(S)\|_2 + \|\widehat{\beta}_I^\lambda(S^c)\|_1\right\} \\
& \qquad + \lambda\sqrt{|I_2|}\left\{\sqrt{d_0}\|(\widehat{\beta}_I^\lambda - \beta_{I_2}^*)(S)\|_2 + \|\widehat{\beta}_I^\lambda(S^c)\|_1\right\} + \gamma \\
\lessapprox & \frac{\lambda^2 d_0}{c_x^2} + |I_1|\|\widehat{\beta}_I^\lambda - \beta_{I_1}^*\|_2^2 + |I_2|\|\widehat{\beta}_I^\lambda - \beta_{I_2}^*\|_2^2 + \lambda^2 + (|I_1|^2 + |I_2|^2)\|\widehat{\beta}_I^\lambda(S^c)\|_1^2 + \gamma,
\end{aligned} \tag{3}$$
where the second inequality follows (2) and the third inequality follows from $2ab \leq a^2 + b^2$ and from setting
$$a = \lambda\sqrt{d_0} \quad \text{and} \quad b = \sqrt{|I_1|}\|\widehat{\beta}_I^\lambda - \beta_{I_1}^*\|_2.$$
Next we apply the restricted eigenvalue conditions along with standard arguments from the Lasso literature to establish that
$$\sum_{t \in I_1}\{x_t^\top(\widehat{\beta}_I^\lambda - \beta_{I_1}^*)\}^2 + \sum_{t \in I_2}\{x_t^\top(\widehat{\beta}_I^\lambda - \beta_{I_2}^*)\}^2$$

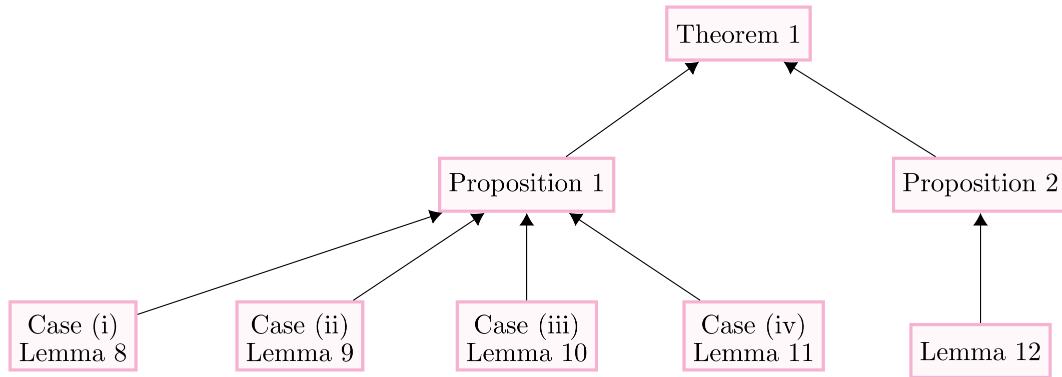

Figure 1: Road map to complete the proof of Theorem 1. The directed edges mean the heads of the edges are used in the tails of the edges.

$$\geq c_x^2 |I_1| \|\widehat{\beta}_I^{\lambda} - \beta_{I_1}^*\|^2 + c_x^2 |I_2| \|\widehat{\beta}_I^{\lambda} - \beta_{I_2}^*\|^2 \geq c_x^2 \kappa^2 \epsilon, \tag{4}$$

where ϵ is an upper bound on the localization error. Combining (3) and (4) leads to

$$\epsilon \lesssim \frac{\lambda^2 d_0 + \gamma}{\kappa^2}.$$

- Finally, the signal-to-noise ratio condition that one needs to assume in order to obtain consistent localization rates is determined by setting $\epsilon \lesssim \Delta$.

The proofs related with Algorithm 1 and Corollary 2 are all based on an oracle inequality of the group Lasso estimator. Once it is established that

$$\sum_{t=s+1}^{e} \|\widehat{\beta}_t - \beta_t^*\|_2^2 \leq \delta \leq \kappa\sqrt{\Delta}, \tag{5}$$

where $\delta \asymp d_0 \log(n \vee p)$ and where there is one and only one change point in the interval $(s, e]$ for both the sequence $\{\widehat{\beta}_t\}$ and $\{\beta_t^*\}$, then the final claim follows immediately that the refined localization error ϵ satisfies

$$\epsilon \leq \delta/\kappa^2.$$

The group Lasso penalty is deployed to prompt (5) and the designs of the algorithm guarantee the desirability of each working interval.

The proof of Theorem 1 proceeds through several steps. For convenience, Figure 1 provides a roadmap for the entire proof. Throughout this section, with some abuse of notation, for any interval $I \subset (0, n]$, we denote with $\beta_I^* = |I|^{-1} \sum_{t \in I} \beta_t^*$.

1.2 Large Probability Events

Lemma 3. *For Model 1, under Assumption 1(**c**), for any interval $I \subset (0, n]$, it holds that*

$$\mathbb{P}\{\mathcal{E}_I\} \geq 1 - c_1 \exp(-c_2|I|),$$

where $c_1, c_2 > 0$ are absolute constants only depending on the distributions of covariants $\{x_t\}$, and

$$\mathcal{E}_I = \left\{ \sqrt{\sum_{t \in I} \left(x_t^\top v\right)^2} \geq \frac{c_x\sqrt{|I|}}{4} \|v\|_2 - 9C_x\sqrt{\log(p)}\|v\|_1, \quad v \in \mathbb{R}^p \right\}.$$

This follows from the same proof as Theorem 1 in Raskutti *et al.* (2010), therefore we omit the proof of Lemma 3. For interval I satisfying $|I| > Cd_0 \log(p)$, an immediate consequence of Lemma 3 is a restricted eigenvalue condition (e.g. van de Geer & Bühlmann, 2009; Bickel *et al.* , 2009). It will be used repeatedly in the rest of this paper.

It will become clearer in the rest of the paper, we only deal with intervals satisfying $|I| \gtrsim d_0 \log(n \vee p)$ when considering the events $\mathcal{E}_I$.

Lemma 4. *For Model 1, under Assumption 1(**c**), for any interval $I \subset (0, n]$, it holds that for any*

$$\lambda \geq \lambda_1 := C_\lambda \sigma_\varepsilon \sqrt{\log(n \vee p)},$$

where $C_\lambda > 0$ is a large enough absolute constant such that, we have

$$\mathbb{P}\{\mathcal{B}_I(\lambda)\} > 1 - 2(n \vee p)^{-c_3},$$

where

$$\mathcal{B}_I(\lambda) = \left\{ \left\| \sum_{t \in I} \varepsilon_t x_t \right\|_\infty \leq \lambda \sqrt{\max\{|I|, \log(n \vee p)\}}/8 \right\},$$

where $c_3 > 0$ is an absolute constant depending only on the distributions of covariants $\{x_t\}$ and $\{\varepsilon_t\}$.

For notational simplicity, we drop the dependence on λ in the notation $\mathcal{B}_I(\lambda)$.

Proof. Since ε_t's are sub-Gaussian random variables and x_t's are sub-Gaussian random vectors, we have that $\varepsilon_t x_t$'s are sub-Exponential random vectors with parameter $C_x \sigma_\varepsilon$ (see e.g. Lemma 2.7.7 in Vershynin, 2018). It then follows from Bernstein's inequality (see e.g. Theorem 2.8.1 in Vershynin, 2018) that for any $t > 0$,

$$\mathbb{P}\left\{ \left\| \sum_{t \in I} \varepsilon_t x_t \right\|_\infty > t \right\} \leq 2p \exp\left\{ -c \min\left\{ \frac{t^2}{|I| C_x^2 \sigma_\varepsilon^2}, \frac{t}{C_x \sigma_\varepsilon} \right\} \right\}.$$

Taking

$$t = C_\lambda C_x / 4\sigma_\varepsilon \sqrt{\log(n \vee p)} \sqrt{\max\{|I|, \log(n \vee p)\}}$$

yields that

$$\mathbb{P}\{\mathcal{B}_I\} > 1 - 2(n \vee p)^{-c_3},$$

where $c_3 > 0$ is an absolute constant depending on $C_\lambda, C_x, \sigma_\varepsilon$. □

1.3 Auxiliary Lemmas

Lemma 5. *For Model 1, under Assumption 1(**a**) and (**c**), if there exists no true change point in $I = (s, e]$, with $|I| > 288^2 C_x^2 d_0 \log(n \vee p)/c_x^2$ and*

$$\lambda \geq \lambda_1 := C_\lambda \sigma_\varepsilon \sqrt{\log(n \vee p)},$$

where $C_\lambda > 0$ being an absolute constant, it holds that

$$\mathbb{P}\left\{ \left\|\widehat{\beta}_I^\lambda - \beta_I^*\right\|_2 \leq \frac{C_3 \lambda \sqrt{d_0}}{\sqrt{|I|}}, \quad \left\|\widehat{\beta}_I^\lambda - \beta_I^*\right\|_1 \leq \frac{C_3 \lambda d_0}{\sqrt{|I|}} \right\}$$
$$\geq 1 - c_1 (n \vee p)^{-288^2 C_x^2 d_0 c_2 / c_x^2} - 2(n \vee p)^{-c_3},$$

where $C_3 > 0$ is an absolute constant depending on all the other absolute constants, c_1, c_2, c_3 are absolute constants defined in Lemmas 3 and 4.

Proof. Let $v = \widehat{\beta}_I^\lambda - \beta_I^*$. Since $|I| > \log(n \vee p)$, it follows from the definition of $\widehat{\beta}_I^\lambda$ that

$$\sum_{t \in I} (y_t - x_t^\top \widehat{\beta}_I^\lambda)^2 + \lambda \sqrt{|I|} \|\widehat{\beta}_I^\lambda\|_1 \leq \sum_{t \in I} (y_t - x_t^\top \beta_I^*)^2 + \lambda \sqrt{|I|} \|\beta_I^*\|_1,$$

which leads to

$$\sum_{t \in I} (x_t^\top v)^2 + \lambda \sqrt{|I|} \|\widehat{\beta}_I^\lambda\|_1 \leq \lambda \sqrt{|I|} \|\beta_I^*\|_1 + 2 \sum_{t \in I} \varepsilon_t x_t^\top v \leq \lambda \sqrt{|I|} \|\beta_I^*\|_1 + \frac{\lambda}{2} \sqrt{|I|} \|v\|_1, \tag{6}$$

where the last inequality holds on the event $\mathcal{B}_I$, with the choice of λ and due to Lemma 4. Note that

$$\|\widehat{\beta}_I^\lambda\|_1 \geq \|\beta_I^*(S)\|_1 - \|v(S)\|_1 + \|\widehat{\beta}_I^\lambda(S^c)\|_1 \tag{7}$$

and

$$\|v\|_1 = \|v(S)\|_1 + \|\widehat{\beta}_I^\lambda(S^c)\|_1. \tag{8}$$

Combining (6), (7) and (8) yields

$$\sum_{t\in I}(x_t^\top v)^2 + \frac{\lambda}{2}\sqrt{|I|}\|\widehat{\beta}_I^\lambda(S^c)\|_1 \leq \frac{3\lambda}{2}\sqrt{|I|}\|\widehat{\beta}_I^\lambda(S)\|_1, \tag{9}$$

which in turn implies

$$\|\widehat{\beta}_I^\lambda(S^c)\|_1 \leq 3\|\widehat{\beta}_I^\lambda(S)\|_1.$$

On the event of $\mathcal{E}_I$, it holds that

$$\begin{aligned}
\sqrt{\sum_{t\in I}\left(x_t^\top v\right)^2} &\geq \frac{c_x\sqrt{|I|}}{4}\|v\|_2 - 9C_x\sqrt{\log(p)}\|v\|_1 \\
&= \frac{c_x\sqrt{|I|}}{4}\|v\|_2 - 9C_x\sqrt{\log(p)}\|v(S)\|_1 - 9C_x\sqrt{\log(p)}\|v(S^c)\|_1 \\
&\geq \frac{c_x\sqrt{|I|}}{4}\|v\|_2 - 36C_x\sqrt{\log(p)}\|v(S)\|_1 \geq \frac{c_x\sqrt{|I|}}{4}\|v\|_2 - 36C_x\sqrt{d_0\log(p)}\|v(S)\|_2 \\
&\geq \left(\frac{c_x\sqrt{|I|}}{4} - 36C_x\sqrt{d_0\log(p)}\right)\|v\|_2 > \frac{c_x\sqrt{|I|}}{8}\|v\|_2,
\end{aligned} \tag{10}$$

where the second inequality follows from (9), the third inequality follows from Assumption 1(**a**) and the last inequality follows from the choice of $|I|$.

Combining (9) and (10) leads to

$$\frac{c_x^2|I|}{64}\|v\|_2^2 \leq \frac{3\lambda}{2}\sqrt{|I|}\|v(S)\|_1 \leq \frac{3\lambda}{2}\sqrt{|I|d_0}\|v\|_2,$$

therefore

$$\|v\|_2 \leq \frac{96\lambda\sqrt{d_0}}{\sqrt{|I|}c_x^2}$$

and

$$\|v\|_1 = \|v(S)\|_1 + \|v(S^c)\|_1 \leq 4\|v(S)\|_1 \leq 4\sqrt{d_0}\|v\|_2 \leq \frac{384\lambda d_0}{\sqrt{|I|}c_x^2}.$$

□

Lemma 6. *For Model 1, under Assumption 1(**a**) and (**c**), if there exists no true change point in $I = (s, e]$, and*

$$\lambda \geq \lambda_1 := C_\lambda\sigma_\varepsilon\sqrt{\log(n\vee p)},$$

where $C_\lambda > 0$ being an absolute constant, it holds that if $|I| \geq 288^2C_x^2d_0\log(n\vee p)/c_x^2$, then

$$\mathbb{P}\left\{\left|\sum_{t\in I}\left\{(y_t - x_t^\top\widehat{\beta})^2 - (y_t - x_t^\top\beta^*)^2\right\}\right| \leq \lambda^2 d_0\right\}$$
$$\geq 1 - c_1(n\vee p)^{-288^2C_x^2d_0c_2/c_x^2} - 2(n\vee p)^{-c_3};$$

if $|I| < 288^2C_x^2d_0\log(n\vee p)/c_x^2$, then

$$\mathbb{P}\left\{\left|\sum_{t\in I}\left\{(y_t - x_t^\top\widehat{\beta})^2 - (y_t - x_t^\top\beta^*)^2\right\}\right| \leq C_4\lambda\sqrt{\log(n\vee p)}d_0^{3/2}\right\} \geq 1 - 2(n\vee p)^{-c_3},$$

where $C_4 > 0$ is an absolute constant depending on all the other constants.

Proof. To ease notation, in this proof, let $\widehat{\beta} = \widehat{\beta}_I^\lambda$ and $\beta^* = \beta_I^*$.

Case 1. If $|I| \geq 288^2 C_x^2 d_0 \log(n \vee p)/c_x^2$, then $|I| > \log(n \vee p)$. With probability at least $1 - c_1 \exp(-c_2|I|) - 2(n \vee p)^{-c_3}$, we have that

$$\sum_{t \in I} \big\{(y_t - x_t^\top \widehat{\beta})^2 - (y_t - x_t^\top \beta^*)^2\big\} \leq \lambda\sqrt{|I|}\|\beta^*\|_1 - \lambda\sqrt{|I|}\|\widehat{\beta}\|_1 \leq \lambda\sqrt{|I|}\|\widehat{\beta} - \beta^*\|_1 \leq C_3 \lambda^2 d_0,$$

where the fist inequality follows from the definition of $\widehat{\beta}$ and the second is due to Lemma 5.

Case 2. If $|I| < 288^2 C_x^2 d_0 \log(n \vee p)/c_x^2$, then

$$\sum_{t \in I} \big\{(y_t - x_t^\top \widehat{\beta})^2 - (y_t - x_t^\top \beta^*)^2\big\} \leq \lambda\sqrt{\max\{|I|, \log(n \vee p)\}}\|\beta^*\|_1 \leq C_4 \lambda \sqrt{\log(n \vee p)} d_0^{3/2},$$

since $\|\beta^*\|_1 \leq C_\beta d_0$. In addition, it holds with probability at least $1 - 2(n \vee p)^{-c_3}$ that

$$\begin{aligned}
&\sum_{t \in I} \big\{(y_t - x_t^\top \beta^*)^2 - (y_t - x_t^\top \widehat{\beta})^2\big\} = -\sum_{t \in I} (x_t^\top \beta^* - x_t^\top \widehat{\beta})^2 + 2\sum_{t \in I} \varepsilon_t x_t^\top (\widehat{\beta} - \beta^*) \\
\leq & -\sum_{t \in I} (x_t^\top \beta^* - x_t^\top \widehat{\beta})^2 + \sum_{t \in I} (x_t^\top \beta^* - x_t^\top \widehat{\beta})^2 + \sum_{t \in I} \varepsilon_t^2 \leq \sum_{t \in I} \varepsilon_t^2 \\
\leq & \max\{\sqrt{|I| \log(n \vee p)}, \log(n \vee p)\} \leq C_4 \lambda \sqrt{\log(n \vee p)} d_0^{3/2},
\end{aligned}$$

where the first inequality follow from $2ab \leq a^2 + b^2$ and letting $a = \varepsilon_t$, $b = x_t^\top(\widehat{\beta} - \beta^*)$, the third inequality follows from the sub-Gaussianity of $\{\varepsilon_t\}$. $\square$

Lemma 7. *For Model 1, under Assumption 1(**a**)-(**c**), for any interval $I = (s, e]$ and*

$$\lambda \geq \lambda_2 := C_\lambda \sigma_\varepsilon \sqrt{d_0 \log(n \vee p)},$$

where $C_\lambda > 8 C_\beta C_x/\sigma_\varepsilon$, it holds with probability at least of $1 - 2(n \vee p)^{-c}$ that,

$$\|\widehat{\beta}_I^\lambda(S^c)\|_1 \leq 3\|\widehat{\beta}_I^\lambda(S)\|_1.$$

If in addition, the interval I satisfies $|I| > 288^2 C_x^2 d_0 \log(n \vee p)/c_x^2$, it holds with probability at least $1 - c_1(n \vee p)^{-288^2 C_x^2 d_0 c_2/c_x^2} - 2(n \vee p)^{-c_3}$ that

$$\left\|\widehat{\beta}_I^\lambda - \frac{1}{|I|}\sum_{t \in I} \beta_t^*\right\|_2 \leq \frac{C_5 \lambda \sqrt{d_0}}{\sqrt{|I|}} \quad \text{and} \quad \left\|\widehat{\beta}_I^\lambda - \frac{1}{|I|}\sum_{t \in I} \beta_t^*\right\|_1 \leq \frac{C_5 \lambda d_0}{\sqrt{|I|}},$$

where $C_5 > 0$ is an absolute constant depending on other constants.

Proof. Denote $\widehat{\beta} = \widehat{\beta}_I^\lambda$ and $\beta^* = (|I|)^{-1}\sum_{t \in I} \beta_t^*$. It follows from the definition of $\widehat{\beta}$ that

$$\sum_{t \in I} (y_t - x_t^\top \widehat{\beta})^2 + \lambda\sqrt{\max\{|I|, \log(n \vee p)\}}\big\|\widehat{\beta}\big\|_1 \leq \sum_{t \in I} (y_t - x_t^\top \beta^*)^2 + \lambda\sqrt{\max\{|I|, \log(n \vee p)\}}\big\|\beta^*\big\|_1,$$

which leads to

$$\begin{aligned}
&\sum_{t \in I} \big\{x_t^\top(\widehat{\beta} - \beta^*)\big\}^2 + 2\sum_{t \in I} (y_t - x_t^\top \beta^*) x_t^\top (\beta^* - \widehat{\beta}) + \lambda\sqrt{\max\{|I|, \log(n \vee p)\}}\big\|\widehat{\beta}\big\|_1 \\
&\qquad \leq \lambda\sqrt{\max\{|I|, \log(n \vee p)\}}\big\|\beta^*\big\|_1,
\end{aligned}$$

therefore

$$\sum_{t \in I} \big\{x_t^\top(\widehat{\beta} - \beta^*)\big\}^2 + 2(\widehat{\beta} - \beta^*)^\top \sum_{t \in I} x_t x_t^\top (\beta^* - \beta_t^*)$$

$$\leq 2\sum_{t\in I}\varepsilon_t x_t^\top(\widehat{\beta}-\beta^*)+\lambda\sqrt{\max\{|I|,\,\log(n\vee p)\}}\left(\left\|\beta^*\right\|_1-\left\|\widehat{\beta}\right\|_1\right). \tag{11}$$

We bound

$$\left\|\sum_{t\in I} x_t x_t^\top(\beta^*-\beta_t^*)\right\|_\infty.$$

For any $k\in\{1,\ldots,p\}$, the kth entry of $\sum_{t\in I} x_t x_t^\top(\beta^*-\beta_t^*)$ satisfies that

$$\mathbb{E}\left\{\sum_{t\in I}\sum_{j=1}^p x_t(k)x_t(j)(\beta^*(j)-\beta_t^*(j))\right\}=\sum_{t\in I}\sum_{j=1}^p\mathbb{E}\{x_t(k)x_t(j)\}\{\beta^*(j)-\beta_t^*(j)\}$$
$$=\sum_{j=1}^p\mathbb{E}\{x_1(k)x_1(j)\}\sum_{t\in I}\{\beta^*(j)-\beta_t^*(j)\}=0.$$

Note that $x_t^\top(\beta^*-\beta_t^*)$'s are sub-Gaussian random variables with a common parameter $2C_\beta C_x\sqrt{d_0}$, and x_t's are sub-Gaussian random vectors with parameter C_x. Therefore due to sub-Exponential inequalities (e.g. Proposition 2.7.1 in Vershynin, 2018), it holds with probability at least of $1-2(n\vee p)^{-c}$ that,

$$\left\|\sum_{t\in I} x_t x_t^\top(\beta^*-\beta_t^*)\right\|_\infty\leq 2C_xC_\beta\sqrt{d_0}\max\{\sqrt{|I|\log(n\vee p)},\,\log(n\vee p)\}$$
$$\leq\lambda\sqrt{\max\{|I|,\,\log(n\vee p)\}}/4. \tag{12}$$

On the event $\mathcal{B}_I$, combining (11) and (12) yields

$$\sum_{t\in I}\left\{x_t^\top(\widehat{\beta}-\beta^*)\right\}^2+\lambda\sqrt{\max\{|I|,\,\log(n\vee p)\}}\left\|\widehat{\beta}\right\|_1$$
$$\leq\lambda/2\sqrt{\max\{|I|,\,\log(n\vee p)\}}\left\|\beta^*-\widehat{\beta}\right\|_1+\lambda\sqrt{\max\{|I|,\,\log(n\vee p)\}}\left\|\beta^*\right\|_1.$$

The final claims follow from the same arguments as in Lemma 5. □

1.4 All cases in Proposition 1

Lemma 8 (Case (i))**.** *With the conditions and notation in Proposition 1, assume that $I=(s,e]\in\widehat{\mathcal{P}}$ has one and only one true change point η. Denote $I_1=(s,\eta]$, $I_2=(\eta,e]$ and $\|\beta^*_{I_1}-\beta^*_{I_2}\|_2=\kappa$. If, in addition, it holds that*

$$\sum_{t\in I}(y_t-x_t^\top\widehat{\beta}_I^\lambda)^2\leq\sum_{t\in I_1}(y_t-x_t^\top\widehat{\beta}_{I_1}^\lambda)^2+\sum_{t\in I_2}(y_t-x_t^\top\widehat{\beta}_{I_2}^\lambda)^2+\gamma, \tag{13}$$

then with

$$\lambda\geq\lambda_2=C_\lambda\sigma_\varepsilon\sqrt{d_0\log(n\vee p)},$$

where $C_\lambda>8C_\beta C_x/\sigma_\varepsilon$, it holds with probability at least $1-2c_1(n\vee p)^{-288^2C_x^2d_0c_2/c_x^2}-2(n\vee p)^{-c_3}$ that, that

$$\min\{|I_1|,\,|I_2|\}\leq C_\epsilon\left(\frac{\lambda^2 d_0+\gamma}{\kappa^2}\right).$$

Proof. First we notice that with the choice of λ, it holds that

$$\lambda\geq\max\{\lambda_1,\,\lambda_2\},$$

and therefore we can apply Lemmas 5, 6 and 7 when needed.

We prove by contradiction, assuming that

$$\min\{|I_1|,\,|I_2|\}>C_\epsilon\left(\frac{\lambda^2 d_0+\gamma}{\kappa^2}\right)>288^2C_x^2d_0\log(n\vee p)/c_x^2, \tag{14}$$

where the second inequality follows from the observation that $\kappa^2 \leq 4d_0 C_\beta^2$. Therefore we also have

$$\min\{|I_1|,\, |I_2|\} > \log(n \vee p).$$

It follows from Lemma 6 and (13) that, with probability at least $1 - 2c_1(n \vee p)^{-288^2 C_x^2 d_0 c_2 / c_x^2} - 2(n \vee p)^{-c_3}$ that, that

$$\begin{aligned}
&\sum_{t \in I_1} (y_t - x_t^\top \widehat{\beta}_I^\lambda)^2 + \sum_{t \in I_2} (y_t - x_t^\top \widehat{\beta}_I^\lambda)^2 = \sum_{t \in I} (y_t - x_t^\top \widehat{\beta}_I^\lambda)^2 \\
\leq & \sum_{t \in I_1} (y_t - x_t^\top \widehat{\beta}_{I_1}^\lambda)^2 + \sum_{t \in I_2} (y_t - x_t^\top \widehat{\beta}_{I_2}^\lambda)^2 + \gamma \\
\leq & \sum_{t \in I_1} (y_t - x_t^\top \beta_{I_1}^*)^2 + \sum_{t \in I_2} (y_t - x_t^\top \beta_{I_2}^*)^2 + \gamma + 2C_3 \lambda^2 d_0. \qquad (15)
\end{aligned}$$

Denoting $\Delta_i = \widehat{\beta}_I^\lambda - \beta_{I_i}^*$, $i = 1, 2$, (15) leads to that

$$\begin{aligned}
&\sum_{t \in I_1} (x_t^\top \Delta_1)^2 + \sum_{t \in I_2} (x_t^\top \Delta_2)^2 \leq 2 \sum_{t \in I_1} \varepsilon_t x_t^\top \Delta_1 + 2 \sum_{t \in I_2} \varepsilon_t x_t^\top \Delta_2 + \gamma + 2C_3 \lambda^2 d_0 \\
\leq & 2 \left\| \sum_{t \in I_1} \varepsilon_t x_t \right\|_\infty \|\Delta_1\|_1 + 2 \left\| \sum_{t \in I_2} \varepsilon_t x_t \right\|_\infty \|\Delta_2\|_1 + \gamma + 2C_3 \lambda^2 d_0 \\
\leq & 2 \left\| \sum_{t \in I_1} \varepsilon_t x_t \right\|_\infty \big(\|\Delta_1(S)\|_1 + \|\Delta_1(S^c)\|_1\big) + 2 \left\| \sum_{t \in I_2} \varepsilon_t x_t \right\|_\infty \big(\|\Delta_2(S)\|_1 + \|\Delta_2(S^c)\|_1\big) \\
& + \gamma + 2C_3 \lambda^2 d_0 \\
\leq & 2 \left\| \sum_{t \in I_1} \varepsilon_t x_t \right\|_\infty \big(\sqrt{d_0} \|\Delta_1(S)\|_2 + \|\Delta_1(S^c)\|_1\big) + 2 \left\| \sum_{t \in I_2} \varepsilon_t x_t \right\|_\infty \big(\sqrt{d_0} \|\Delta_2(S)\|_2 + \|\Delta_2(S^c)\|_1\big) \\
& + \gamma + 2C_3 \lambda^2 d_0. \qquad (16)
\end{aligned}$$

On the events $\mathcal{B}_{I_1} \cap \mathcal{B}_{I_2}$, it holds that

$$\begin{aligned}
(16) \leq & \lambda/2 \big(\sqrt{|I_1| d_0} \|\Delta_1(S)\|_2 + \sqrt{|I_1|} \|\Delta_1(S^c)\|_1 + \sqrt{|I_2| d_0} \|\Delta_2(S)\|_2 \\
& + \sqrt{|I_2|} \|\Delta_2(S^c)\|_1\big) + \gamma + 2C_3 \lambda^2 d_0 \\
\leq & \frac{32\lambda^2 d_0}{c_x^2} + \frac{c_x^2 |I_1| \|\Delta_1\|_2^2}{256} + \frac{c_x^2 |I_2| \|\Delta_2\|_2^2}{256} + \frac{\lambda(\sqrt{|I_1|} + \sqrt{|I_2|})}{2} \|\widehat{\beta}_I^\lambda(S^c)\|_1 + \gamma + 2C_3 \lambda^2 d_0 \\
\leq & \frac{32\lambda^2 d_0}{c_x^2} + \frac{c_x^2 |I_1| \|\Delta_1\|_2^2}{256} + \frac{c_x^2 |I_2| \|\Delta_2\|_2^2}{256} + \gamma + 4C_3 \lambda^2 d_0, \qquad (17)
\end{aligned}$$

where the second inequality follows from $2ab \leq a^2 + b^2$, letting

$$a = 4\lambda \sqrt{d_0} / c_x \quad \text{and} \quad b = c_x \sqrt{|I_j|} \|\Delta_1\|_2 / 16, \quad j = 1, 2,$$

and the last inequality follows from Lemma 7.

Note that

$$\|\Delta_1\|_1 \leq \|\Delta_1(S)\|_1 + \|\Delta_1(S^c)\|_1 \leq \sqrt{d_0} \|\Delta_1\|_2 + \frac{C_5 \lambda d_0}{\sqrt{|I_1|}},$$

which combines with (14), on the event $\mathcal{E}_{I_1}$, leads to

$$\sqrt{\sum_{t \in I_1} (x_t^\top \Delta_1)^2} > \frac{c_x \sqrt{|I_1|}}{4} \|\Delta_1\|_2 - 9C_x \sqrt{\log(p)} \|\Delta_1\|_1 > \frac{c_x \sqrt{|I_1|}}{8} \|\Delta_1\|_2 - \frac{9C_5 C_x \lambda d_0 \sqrt{\log(p)}}{c_x^2 \sqrt{|I_1|}}.$$

Moreover, we have

$$
\begin{aligned}
&\sqrt{|I_1|}\|\Delta_1\|_2 + \sqrt{|I_2|}\|\Delta_2\|_2 \geq \sqrt{|I_1|\|\Delta_1\|_2^2 + |I_2|\|\Delta_2\|_2^2} \\
\geq &\sqrt{\inf_{v \in \mathbb{R}^p} \{|I_1|\|\beta^*_\eta - v\|^2 + |I_2|\|\beta^*_{\eta+1} - v\|^2\}} = \kappa\sqrt{\frac{|I_1||I_2|}{|I|}} \geq \frac{\kappa}{\sqrt{2}} \min\{\sqrt{|I_1|}, \sqrt{|I_2|}\}. \qquad (18)
\end{aligned}
$$

Therefore, on the event $\mathcal{E}_{I_1} \cap \mathcal{E}_{I_2} \cap \mathcal{B}_{I_1} \cap \mathcal{B}_{I_2}$, combining (16) and (17), we have that

$$
\begin{aligned}
\sqrt{|I_1|}\|\Delta_1\|_2 + \sqrt{|I_2|}\|\Delta_2\|_2 \leq & \frac{8}{c_x}\left(\sqrt{\sum_{t \in I_1}(x_t^\top \Delta_1)^2} + \sqrt{\sum_{t \in I_2}(x_t^\top \Delta_2)^2}\right) \\
& + \frac{8}{c_x}\left(\frac{9C_5 C_x \lambda d_0 \sqrt{\log(p)}}{c_x^2\sqrt{|I_1|}} + \frac{9C_5 C_x \lambda d_0 \sqrt{\log(p)}}{c_x^2\sqrt{|I_2|}}\right) \\
\leq \frac{8\sqrt{2}}{c_x}\sqrt{\frac{32\lambda^2 d_0}{c_x^2} + \frac{c_x^2|I_1|\|\Delta_1\|_2^2}{256} + \frac{c_x^2|I_2|\|\Delta_2\|_2^2}{256} + \gamma + 4C_3\lambda^2 d_0}& \\
& + \frac{8}{c_x}\left(\frac{9C_5 C_x \lambda d_0 \sqrt{\log(p)}}{c_x^2\sqrt{|I_1|}} + \frac{9C_5 C_x \lambda d_0 \sqrt{\log(p)}}{c_x^2\sqrt{|I_2|}}\right) \\
\leq \frac{64\lambda\sqrt{d_0}}{c_x^2} + \frac{\sqrt{2}\sqrt{|I_1|}\|\Delta_1\|_2}{2} + \frac{\sqrt{2}\sqrt{|I_2|}\|\Delta_2\|_2}{2} &+ \frac{8\sqrt{2\gamma}}{c_x} + \frac{16\sqrt{2C_3}\lambda\sqrt{d_0}}{c_x} + \frac{C_5\lambda\sqrt{d_0}}{2c_x^2},
\end{aligned}
$$

which implies that

$$
\frac{2-\sqrt{2}}{2}\left(\sqrt{|I_1|}\|\Delta_1\|_2 + \sqrt{|I_2|}\|\Delta_2\|_2\right) \leq \frac{128 + 32\sqrt{2}c_x\sqrt{C_3} + C_5}{2c_x^2}\lambda\sqrt{d_0} + \frac{8\sqrt{2\gamma}}{c_x}. \qquad (19)
$$

Combining (18) and (19) yields

$$
\frac{2-\sqrt{2}}{2\sqrt{2}}\kappa\sqrt{\min\{|I_1|,\,|I_2|\}} \leq \frac{128 + 32\sqrt{2}c_x\sqrt{C_3} + C_5}{2c_x^2}\lambda\sqrt{d_0} + \frac{8\sqrt{2\gamma}}{c_x},
$$

therefore

$$
\min\{|I_1|,\,|I_2|\} \leq C_\epsilon\left(\frac{\lambda^2 d_0 + \gamma}{\kappa^2}\right),
$$

which is a contradiction with (14). □

Lemma 9 (Case (ii))**.** *For Model 1, under Assumption 1, with*

$$
\lambda \geq \lambda_2 = C_\lambda \sigma_\varepsilon \sqrt{d_0 \log(n \vee p)},
$$

where $C_\lambda > 8C_\beta C_x/\sigma_\varepsilon$, $I = (s, e]$ containing exactly two change points η_1 and η_2. Denote $I_1 = (s, \eta_1]$, $I_2 = (\eta_1, \eta_2]$, $I_3 = (\eta_2, e]$, $\|\beta^*_{I_1} - \beta^*_{I_2}\|_2 = \kappa_1$ and $\|\beta^*_{I_2} - \beta^*_{I_3}\|_2 = \kappa_2$. If in addition it holds that

$$
\sum_{t \in I}(y_t - x_t^\top \widehat{\beta}^\lambda_I)^2 \leq \sum_{t \in I_1}(y_t - x_t^\top \widehat{\beta}^\lambda_{I_1})^2 + \sum_{t \in I_2}(y_t - x_t^\top \widehat{\beta}^\lambda_{I_2})^2 + \sum_{t \in I_3}(y_t - x_t^\top \widehat{\beta}^\lambda_{I_3})^2 + 2\gamma,
$$

then

$$
\max\{|I_1|,\,|I_3|\} \leq C_\epsilon\left(\frac{\lambda^2 d_0 + \gamma}{\kappa^2}\right),
$$

with probability at least $1 - 3c_1(n \vee p)^{-288^2 C_x^2 d_0 c_2/c_x^2} - 2(n \vee p)^{-c_3}$.

Proof. First we notice that with the choice of λ, it holds that

$$\lambda \geq \max\{\lambda_1,\, \lambda_2\},$$

and therefore we can apply Lemmas 5, 6 and 7 when needed.

By symmetry, it suffices to show that

$$|I_1| \leq C_\epsilon \left(\frac{\lambda^2 d_0 + \gamma}{\kappa^2}\right).$$

We prove by contradiction, assuming that

$$|I_1| > C_\epsilon \left(\frac{\lambda^2 d_0 + \gamma}{\kappa^2}\right) > 288^2 C_x^2 d_0 \log(n \vee p)/c_x^2, \tag{20}$$

where the second inequality follows from the observation that $\kappa^2 \leq 4 d_0 C_\beta^2$. Therefore we have $|I_1| > \log(n \vee p)$. Denote $\Delta_i = \widehat{\beta}_I^\lambda - \beta_{I_i}^*$, $i = 1, 2, 3$. We then consider the following two cases.

Case 1. If

$$|I_3| > 288^2 C_x^2 d_0 \log(n \vee p)/c_x^2,$$

then $|I_3| > \log(n \vee p)$. It follows from Lemma 6 that the following holds with probability at least $1 - 3c_1(n \vee p)^{-288^2 C_x^2 d_0 c_2/c_x^2} - 2(n \vee p)^{-c_3}$ that,

$$\sum_{t \in I} (y_t - x_t^\top \widehat{\beta}_I^\lambda)^2 \leq \sum_{t \in I_1} (y_t - x_t^\top \widehat{\beta}_{I_1}^\lambda)^2 + \sum_{t \in I_2} (y_t - x_t^\top \widehat{\beta}_{I_2}^\lambda)^2 + \sum_{t \in I_3} (y_t - x_t^\top \widehat{\beta}_{I_3}^\lambda)^2 + 2\gamma$$
$$\leq \sum_{t \in I_1} (y_t - x_t^\top \beta_{I_1}^*)^2 + \sum_{t \in I_2} (y_t - x_t^\top \beta_{I_2}^*)^2 + \sum_{t \in I_3} (y_t - x_t^\top \beta_{I_3}^*)^2 + 3C_3 \lambda^2 d_0 + 2\gamma,$$

which implies that

$$\sum_{i=1}^3 \sum_{t \in I_i} (x_t^\top \Delta_i)^2 \leq 2 \sum_{i=1}^3 \sum_{t \in I_i} \varepsilon_t x_t^\top \Delta_i + 3C_3 \lambda^2 d_0 + 2\gamma$$
$$\leq 2 \sum_{i=1}^3 \left\| \frac{1}{\sqrt{|I_i|}} \sum_{t \in I_i} \varepsilon_t x_t \right\|_\infty \|\sqrt{|I_i|} \Delta_i\|_1 + 3C_3 \lambda^2 d_0 + 2\gamma$$
$$\leq \lambda/2 \sum_{i=1}^3 \left(\sqrt{d_0 |I_i|} \|\Delta_i(S)\|_2 + \sqrt{|I_i|} \|\Delta_i(S^c)\|_1 \right) + 3C_3 \lambda^2 d_0 + 2\gamma,$$

where the last inequality follows from Lemma 4.

It follows from identical arguments in Lemma 8 that, with probability at least $1 - 3c_1(n \vee p)^{-288^2 C_x^2 d_0 c_2/c_x^2} - 2(n \vee p)^{-c_3}$,

$$\min\{|I_1|,\, |I_2|\} \leq C_\epsilon \left(\frac{\lambda^2 d_0 + \gamma}{\kappa^2}\right).$$

Since $|I_2| \geq \Delta$ by assumption, it follows from Assumption 1(**d**) that

$$|I_1| \leq C_\epsilon \left(\frac{\lambda^2 d_0 + \gamma}{\kappa^2}\right),$$

which contradicts (20).

Case 2. If

$$|I_3| \leq 288^2 C_x^2 d_0 \log(n \vee p)/c_x^2,$$

then it follows from Lemma 6 that the following holds with probability at least $1 - 2c_1(n \vee p)^{-288^2 C_x^2 d_0 c_2/c_x^2} - 2(n \vee p)^{-c_3}$ that,

$$\sum_{t \in I} (y_t - x_t^\top \widehat{\beta}_I^\lambda)^2 \leq \sum_{t \in I_1} (y_t - x_t^\top \widehat{\beta}_{I_1}^\lambda)^2 + \sum_{t \in I_2} (y_t - x_t^\top \widehat{\beta}_{I_2}^\lambda)^2 + \sum_{t \in I_3} (y_t - x_t^\top \widehat{\beta}_{I_3}^\lambda)^2 + 2\gamma$$

$$\leq \sum_{t\in I_1}(y_t - x_t^\top \beta^*_{I_1})^2 + \sum_{t\in I_2}(y_t - x_t^\top \beta^*_{I_2})^2 + \sum_{t\in I_3}(y_t - x_t^\top \beta^*_{I_3})^2 + 2C_3\lambda^2 d_0 + C_4\lambda\sqrt{\log(p)}d_0^{3/2} + 2\gamma,$$

which implies that

$$\begin{aligned}
&\sum_{i=1}^{3}\sum_{t\in I_i}(x_t^\top \Delta_i)^2 \leq 2\sum_{i=1}^{3}\sum_{t\in I_i}\varepsilon_t x_t^\top \Delta_i + 2C_3\lambda^2 d_0 + C_4\lambda\sqrt{\log(p)}d_0^{3/2} + 2\gamma \\
\leq & 2\sum_{i=1}^{2}\left\|\frac{1}{\sqrt{|I_i|}}\sum_{t\in I_i}\varepsilon_t x_t\right\|_\infty \|\sqrt{|I_i|}\Delta_i\|_1 + 2C_3\lambda^2 d_0 + C_4\lambda\sqrt{\log(p)}d_0^{3/2} \\
&\qquad + 2\gamma + \sum_{t\in I_3}(x_t^\top \Delta_3)^2 + \sum_{t\in I_3}\varepsilon_t^2 \\
\leq & \lambda/2\sum_{i=1}^{2}\left(\sqrt{d_0|I_i|}\|\Delta_i(S)\|_2 + \sqrt{|I_i|}\|\Delta_i(S^c)\|_1\right) + 2C_3\lambda^2 d_0 + C_4\lambda\sqrt{\log(p)}d_0^{3/2} \\
&\qquad + 2\gamma + \sum_{t\in I_3}(x_t^\top \Delta_3)^2 + \sum_{t\in I_3}\varepsilon_t^2.
\end{aligned}$$

The rest follows from the same arguments as in **Case 1.**

□

Lemma 10 (Case (iii) in Proposition 1)**.** *For Model 1, under Assumption 1, if there exists no true change point in $I = (s, e]$, with*

$$\lambda \geq \lambda_2 = C_\lambda \sigma_\varepsilon \sqrt{d_0 \log(n \vee p)},$$

where $C_\lambda > \max\{8C_1C_x, 8C_\beta C_x/\sigma_\varepsilon\}$, and $\gamma = C_\gamma \sigma_\varepsilon^2 d_0^2 \log(n \vee p)$, where $C_\gamma > \max\{3C_3/c_x^2, 3C_4/c_x\}$, it holds with probability at least $1 - 3c_1(n \vee p)^{-288^2 C_x^2 d_0 c_2/c_x^2} - 2(n \vee p)^{-c_3}$ that

$$\sum_{t\in I}(y_t - x_t^\top \widehat{\beta}^\lambda_I)^2 < \min_{b=s+1,\ldots,e-1}\left\{\sum_{t\in(s,b]}(y_t - x_t^\top \widehat{\beta}^\lambda_{(s,b]})^2 + \sum_{t\in(b,e]}(y_t - x_t^\top \widehat{\beta}^\lambda_{(b,e]})^2\right\} + \gamma.$$

Proof. First we notice that with the choice of λ, it holds that $\lambda > \lambda_1$, therefore we can apply Lemma 6 when needed.

For any $b = s+1, \ldots, e-1$, let $I_1 = (s, b]$ and $I_2 = (b, e]$. It follows from Lemma 6 that with probability at least $1 - 3c_1(n \vee p)^{-288^2 C_x^2 d_0 c_2/c_x^2} - 2(n \vee p)^{-c_3}$,

$$\max_{J\in\{I_1,I_2,I\}}\left|\sum_{t\in J}(y_t - x_t^\top \widehat{\beta}^\lambda_J)^2 - \sum_{t\in J}(y_t - x_t^\top \beta^*_J)^2\right| \leq \max\left\{C_3\lambda^2 d_0,\, C_4\lambda\sqrt{\log(n \vee p)}d_0^{3/2}\right\} < \gamma/3.$$

Since $\beta^*_I = \beta^*_{I_1} = \beta^*_{I_2}$, the final claim holds automatically.

□

Lemma 11 (Case (iv) in Proposition 1)**.** *For Model 1, under Assumption 1, if $I = (s, e]$ contains J true change points $\{\eta_k\}_{j=1}^J$, where $|J| \geq 3$, if*

$$\lambda \geq \lambda_2 = C_\lambda \sigma_\varepsilon \sqrt{d_0}\log(n \vee p),$$

where $C_\lambda > 8C_\beta C_x/\sigma_\varepsilon$, then with probability at least $1 - nc_1(n \vee p)^{-288^2 C_x^2 d_0 c_2/c_x^2} - 2(n \vee p)^{-c_3}$,

$$\sum_{t\in I}(y_t - x_t^\top \widehat{\beta}^\lambda_I)^2 > \sum_{j=1}^{J+1}\sum_{t\in I_j}(y_t - x_t^\top \widehat{\beta}^\lambda_{I_j})^2 + J\gamma,$$

where $I_1 = (s, \eta_1]$, $I_j = (\eta_j, \eta_{j+1}]$ for any $2 \leq j \leq J$ and $I_{J+1} = (\eta_J, e]$.

Proof. First we notice that with the choice of λ, it holds that

$$\lambda \geq \max\{\lambda_1,\, \lambda_2\},$$

and therefore we can apply Lemmas 5, 6 and 7 when needed.

We prove the claim by contradiction, assuming that

$$\sum_{t\in I}(y_t - x_t^\top \widehat{\beta}_I^\lambda)^2 \leq \sum_{j=1}^{J+1}\sum_{t\in I_j}(y_t - x_t^\top \widehat{\beta}_{I_j}^\lambda)^2 + J\gamma.$$

Let $\Delta_i = \widehat{\beta}_I^\lambda - \beta_{I_i}^*$, $i = 1, \ldots, J+1$. It then follows from Lemma 6 that with probability at least $1 - nc_1(n \vee p)^{-288^2 C_x^2 d_0 c_2/c_x^2} - 2(n\vee p)^{-c_3}$,

$$\begin{aligned}\sum_{t\in I}(y_t - x_t^\top \widehat{\beta}_I^\lambda)^2 &\leq \sum_{j=1}^{J+1}\sum_{t\in I_j}(y_t - x_t^\top \widehat{\beta}_{I_j}^\lambda)^2 + J\gamma \\ &\leq \sum_{j=1}^{J+1}\sum_{t\in I_j}(y_t - x_t^\top \beta_{I_j}^*)^2 + J\gamma + (J+1)C_\gamma\sigma_\varepsilon^2 d_0^2\log(n\vee p),\end{aligned}$$

which implies that

$$\sum_{j=1}^{J+1}\sum_{t\in I_j}(x_t^\top\Delta_j)^2 \leq 2\sum_{j=1}^{J+1}\sum_{t\in I_j}\varepsilon_t x_t^\top\Delta_j + J\gamma + (J+1)C_\gamma\sigma_\varepsilon^2 d_0^2\log(n\vee p). \tag{21}$$

Step 1. For any $j \in \{2, \ldots, J\}$, it follows from Assumption 1 that

$$|I_j| \geq \Delta \geq 288^2 C_x^2 d_0 \log(n\vee p)/c_x^2. \tag{22}$$

Due to Lemma 4, on the event $\mathcal{B}_{(0,n]}$, it holds that

$$\begin{aligned}&\sum_{t\in I_j}\varepsilon_t x_t^\top\Delta_j \leq \left\|\frac{1}{\sqrt{|I_j|}}\sum_{t\in I_j}\varepsilon_t x_t\right\|_\infty \|\sqrt{|I_j|}\Delta_j\|_1 \leq \lambda/4\left(\sqrt{d_0|I_j|}\|\Delta_j(S)\|_2 + \sqrt{|I_j|}\|\Delta_j(S^c)\|_1\right)\\ \leq&\frac{4\lambda^2 d_0}{c_x^2} + \frac{c_x^2|I_j|}{256}\|\Delta_j\|_2^2 + \lambda/4\sqrt{|I_j|}\|\widehat{\beta}_I^\lambda(S^c)\|_1\\ =&\frac{4\lambda^2 d_0}{c_x^2} + \frac{c_x^2|I_j|}{256}\|\Delta_j\|_2^2 + \lambda/4\sqrt{|I_j|}\|(\widehat{\beta}_I^\lambda - (|I|)^{-1}\sum_{t\in I}\beta_t^*)(S^c)\|_1\\ \leq&\frac{4\lambda^2 d_0}{c_x^2} + \frac{c_x^2|I_j|}{256}\|\Delta_j\|_2^2 + \lambda/4\sqrt{|I_j|}\|\widehat{\beta}_I^\lambda - (|I|)^{-1}\sum_{t\in I}\beta_t^*\|_1\\ \leq&\frac{4\lambda^2 d_0}{c_x^2} + \frac{c_x^2|I_j|}{256}\|\Delta_j\|_2^2 + C_5/4\lambda^2 d_0,\end{aligned} \tag{23}$$

where the third inequality follows from $2ab \leq a^2 + b^2$, letting

$$a = 2\lambda\sqrt{d_0}/c_x \quad \text{and]} \quad b = c_x\sqrt{|I_j|}\|\Delta_j\|_2/16,$$

and the last inequality follows from Lemma 7. In addition, on the event of $\mathcal{E}_{I_j}$, due to Lemma 3, it holds that

$$\sqrt{\sum_{t\in I_j}(x_t^\top\Delta_j)^2} \geq \frac{c_x\sqrt{|I_j|}}{4}\|\Delta_j\|_2 - 9C_x\sqrt{\log(p)}\|\Delta_j\|_1$$

$$
\begin{aligned}
&\geq \frac{c_x\sqrt{|I_j|}}{4}\|\Delta_j\|_2 - 9C_x\sqrt{d_0\log(p)}\|\Delta_j\|_2 - 9C_x\sqrt{\log(p)}\|\Delta_j(S^c)\|_1 \\
&\geq \frac{c_x\sqrt{|I_j|}}{8}\|\Delta_j\|_2 - 9C_x\sqrt{\log(p)}\|\Delta_j(S^c)\|_1 \geq \frac{c_x\sqrt{|I_j|}}{8}\|\Delta_j\|_2 - \frac{9C\lambda d_0\sqrt{\log(p)}}{\sqrt{|I|}},
\end{aligned} \tag{24}
$$

where the third inequality follows from (22) and the last follows from Lemma 7.

Step 2. We then discuss the intervals I_1 and I_{J+1}. These two will be treated in the same way, and therefore for $L \in \{I_1, I_{J+1}\}$ and $l \in \{1, J+1\}$, we have the following arguments. If $|L| \geq 288^2 C_x^2 d_0 \log(n \vee p)/c_x^2$, then due to the same arguments in **Step 1**, (23) and (24) hold. If instead, $|L| < 288^2 C_x^2 d_0 \log(n \vee p)/c_x^2$ holds, then

$$
\sum_{t\in L} \varepsilon_t x_t^\top \Delta_l \leq 2^{-1}\sum_{t\in L}(x_t^\top \Delta_l)^2 + 4\sum_{t\in L}\varepsilon_t^2.
$$

Therefore, it follows from (21) that

$$
\sum_{j=2}^{J} |I_j| c_x^2 \|\Delta_j\|_2^2 \leq JC\max\left\{\lambda^2 d_0,\, \lambda\sqrt{\log(n\vee p)}d_0^{3/2}\right\} + J\gamma.
$$

Step 3. Since for any $j \in \{2, \ldots, J-1\}$, it holds that

$$
\begin{aligned}
|I_j|\|\Delta_j\|_2^2 + |I_{j+1}|\|\Delta_{j+1}\|_2^2 &\geq \inf_{v\in\mathbb{R}^p}\left\{|I_j|\|\beta^*_{I_j} - v\|_2^2 + |I_{j+1}|\|\beta^*_{I_{j+1}} - v\|_2^2\right\} \\
&\geq \frac{|I_j||I_{j+1}|}{|I_j| + |I_{j+1}|}\kappa^2 \geq \min\{|I_j|,\, |I_{j+1}|\}\kappa^2/2.
\end{aligned}
$$

It then follows from the same arguments in Lemma 8 that

$$
\min_{j=2,\ldots,J-1} |I_j| \leq C_\epsilon\left(\frac{\lambda^2 d_0 + \gamma}{\kappa^2}\right),
$$

which is a contradiction to (22). □

1.5 Proof of Proposition 2

Lemma 12. *Under the assumptions and notation in Proposition 1, suppose there exists no true change point in the interval I. For any interval $J \supset I$, with*

$$
\lambda \geq \lambda_2 = C_\lambda \sigma_\varepsilon\sqrt{d_0\log(n\vee p)},
$$

where $C_\lambda > \max\{8C_1C_x,\, 8C_\beta C_x/\sigma_\varepsilon\}$, it holds that with probability at least $1 - c_1(n\vee p)^{-288^2C_x^2 d_0 c_2/c_x^2} - 2(n\vee p)^{-c_3}$,

$$
\sum_{t\in I}(y_t - x_t^\top \beta_I^*)^2 - \sum_{t\in I}(y_t - x_t^\top \widehat{\beta}_J^\lambda)^2 \leq C_6\lambda^2 d_0.
$$

Proof. **Case 1.** If

$$
|I| \geq 288^2 C_x^2 d_0 \log(n\vee p)/c_x^2, \tag{25}
$$

then letting $\Delta_I = \beta_I^* - \widehat{\beta}_J^\lambda$, on the event $\mathcal{E}_I$, we have

$$
\begin{aligned}
\sqrt{\sum_{t\in I}(x_t^\top\Delta_I)^2} &\geq \frac{c_x\sqrt{|I|}}{4}\|\Delta_I\|_2 - 9C_x\sqrt{\log(p)}\|\Delta_I\|_1 \\
&= \frac{c_x\sqrt{|I|}}{4}\|\Delta_I\|_2 - 9C_x\sqrt{\log(p)}\|\Delta_I(S)\|_1 - 9C_x\sqrt{\log(p)}\|\Delta_I(S^c)\|_1 \\
&\geq \frac{c_x\sqrt{|I|}}{4}\|\Delta_I\|_2 - 9C_x\sqrt{d_0\log(p)}\|\Delta_I\|_2 - 9C_x\sqrt{\log(p)}\|\Delta_I(S^c)\|_1
\end{aligned}
$$

$$\geq \frac{c_x\sqrt{|I|}}{8}\|\Delta_I\|_2 - 9C_x\sqrt{\log(p)}\|\widehat{\beta}^{\lambda}_J(S^c)\|_1 \geq \frac{c_x\sqrt{|I|}}{8}|\Delta_I\|_2 - 9C_5C_xd_0\lambda\log^{1/2}(p), \tag{26}$$

where the last inequality follows from Lemma 7. We then have on the event $\mathcal{B}_I$,

$$\begin{aligned}
&\sum_{t\in I}(y_t - x_t^\top\beta^*_I)^2 - \sum_{t\in I}(y_t - x_t^\top\widehat{\beta}^{\lambda}_J)^2 = 2\sum_{t\in I}\varepsilon_t x_t^\top\Delta_I - \sum_{t\in I}(x_t^\top\Delta_I)^2\\
\leq & 2\left\|\sum_{t\in I}x_t\varepsilon_t\right\|_\infty\left(\sqrt{d_0}\|\Delta_I(S)\|_2 + \|\widehat{\beta}^{\lambda}_J(S^c)\|_1\right)\\
&\qquad - \frac{c_x^2|I|}{64}\|\Delta_I\|_2^2 - \frac{81C_5^2C_x^2\lambda^2d_0^2\log(p)}{c_x^4|I|} + \frac{9C_5C_xd_0\lambda\log^{1/2}(p)\|\Delta_I\|_2}{4}\\
\leq & \frac{\lambda}{2}\sqrt{d_0}\|\Delta_I\|_2 + \frac{\lambda^2d_0C_5}{2c_x^2\sqrt{|I|}} - \frac{c_x^2|I|}{64}\|\Delta_I\|_2^2 + \frac{9C_5C_xd_0\lambda\log^{1/2}(p)\|\Delta_I\|_2}{4}\\
\leq & \frac{\lambda}{2}\sqrt{d_0}\|\Delta_I\|_2 + \frac{\lambda^2\sqrt{d_0}C_5}{576c_x\sqrt{\log(n\vee p)}C_x} - 36^2C_x^2d_0\log(n\vee p)\|\Delta_I\|_2^2 + \frac{9C_5C_xd_0\lambda\log^{1/2}(p)\|\Delta_I\|_2}{4}\\
\leq & \frac{\lambda^2}{16C_x^2} + d_0C_x^2\|\Delta_I\|_2^2 + \frac{\lambda^2\sqrt{d_0}C_5}{576c_x\sqrt{\log(n\vee p)}C_x} - 36^2C_x^2d_0\log(n\vee p)\|\Delta_I\|_2^2\\
&\qquad + d_0\log(p)\|\Delta_I\|_2^2C_x^2 + \frac{81C_5^2d_0\lambda^2}{64}\\
\leq & C_6\lambda^2d_0.
\end{aligned}$$

where the first inequality follows from (26), the second inequality follows from event $\mathcal{B}_I$ and Lemma 7, the third follows from the (25), the fourth follows from $2ab \leq a^2 + b^2$, first letting

$$a = \lambda/(4C_x) \quad \text{and} \quad b = \sqrt{d_0}C_x\|\Delta_I\|_2,$$

then letting

$$a = C_x\sqrt{d_0\log(p)}\|\Delta_I\|_2 \quad \text{and} \quad b = 9C_5\sqrt{d_0}\lambda/8,$$

and the last inequality follows from Lemma 7.

Case 2. If $|I| \leq 288^2C_x^2d_0\log(n\vee p)/c_x^2$, then with probability at least $1 - 2(n\vee p)^{-c}$,

$$\begin{aligned}
&\sum_{t\in I}(y_t - x_t^\top\beta^*_I)^2 - \sum_{t\in I}(y_t - x_t^\top\widehat{\beta}^{\lambda}_J)^2 = 2\sum_{t\in I}\varepsilon_t x_t^\top(\widehat{\beta}^{\lambda}_J - \beta^*_I) - \sum_{t\in I}\{x_t^\top(\beta^*_I - \widehat{\beta}^{\lambda}_J)\}^2\\
\leq & \sum_{t\in I}\varepsilon_t^2 \leq \max\{\sqrt{|I|\log(n\vee p)},\, \log(n\vee p)\} \leq C_6\lambda^2d_0.
\end{aligned}$$

□

Proof of Proposition 2. Denote $S_n^* = \sum_{t=1}^n(y_t - x_t^\top\beta_t^*)^2$. Given any collection $\{t_1,\ldots,t_m\}$, where $t_1 < \cdots < t_m$, and $t_0 = 0$, $t_{m+1} = n$, let

$$S_n(t_1,\ldots,t_m) = \sum_{k=1}^m\sum_{t=t_k+1}^{t_{k+1}}\left(y_t - x_t^\top\widehat{\beta}^{\lambda}_{(t_k,t_{k+1}]}\right)^2. \tag{27}$$

For any collection of time points, when defining (27), the time points are sorted in an increasing order.

Let $\{\widehat{\eta}_k\}_{k=1}^{\widehat{K}}$ denote the change points induced by $\widehat{\mathcal{P}}$. If one can justify that

$$\begin{aligned}
S_n^* + K\gamma \geq & S_n(\eta_1,\ldots,\eta_K) + K\gamma - C_3(K+1)d_0\lambda^2 && (28)\\
\geq & S_n(\widehat{\eta}_1,\ldots,\widehat{\eta}_{\widehat{K}}) + \widehat{K}\gamma - C_3(K+1)d_0\lambda^2 && (29)\\
\geq & S_n(\widehat{\eta}_1,\ldots,\widehat{\eta}_{\widehat{K}},\eta_1,\ldots,\eta_K) + \widehat{K}\gamma - 2C(K+1)d_0\lambda^2 - C_3(K+1)d_0\lambda^2 && (30)
\end{aligned}$$

and that

$$S_n^* - S_n(\widehat{\eta}_1, \ldots, \widehat{\eta}_{\widehat{K}}, \eta_1, \ldots, \eta_K) \le C(K + \widehat{K} + 2)\lambda^2 d_0, \tag{31}$$

then it must hold that $|\widehat{\mathcal{P}}| = K$, as otherwise if $\widehat{K} \ge K + 1$, then

$$\begin{aligned} C(K + \widehat{K} + 2)\lambda^2 d_0 &\ge S_n^* - S_n(\widehat{\eta}_1, \ldots, \widehat{\eta}_{\widehat{K}}, \eta_1, \ldots, \eta_K) \\ &\ge -3C(K+1)\lambda^2 d_0 + (\widehat{K} - K)\gamma \ge C_\gamma (K+1)\lambda^2 d_0. \end{aligned}$$

Therefore due to the assumption that $|\widehat{\mathcal{P}}| = \widehat{K} \le 3K$, it holds that

$$C(5K + 3)\lambda^2 d_0 \ge (\widehat{K} - K)\gamma \ge \gamma, \tag{32}$$

Note that (32) contradicts the choice of γ.

Note that (28) is implied by

$$|S_n^* - S_n(\eta_1, \ldots, \eta_K)| \le C_3(K+1)d_0\lambda^2, \tag{33}$$

which is immediate consequence of Lemma 6. Since $\{\widehat{\eta}_k\}_{k=1}^{\widehat{K}}$ are the change points induced by $\widehat{\mathcal{P}}$, (29) holds because $\widehat{\mathcal{P}}$ is a minimiser.

For every $\widehat{I} = (s, e] \in \widehat{\mathcal{P}}$ denote

$$\widehat{I} = (s, \eta_{p+1}] \cup \ldots \cup (\eta_{p+q}, e] = J_1 \cup \ldots \cup J_{q+1},$$

where $\{\eta_{p+l}\}_{l=1}^{q+1} = \widehat{I} \cap \{\eta_k\}_{k=1}^K$. Then (30) is an immediate consequence of the following inequality

$$\sum_{t \in \widehat{I}} (y_t - x_t^\top \widehat{\beta}_{\widehat{I}}^\lambda)^2 \ge \sum_{l=1}^{q+1} \sum_{t \in J_l} (y_t - x_t^\top \widehat{\beta}_{J_l}^\lambda)^2 - C(q+1)\lambda^2 d_0. \tag{34}$$

By Lemma 6, it holds that

$$\begin{aligned} \sum_{l=1}^{q+1} \sum_{t \in J_l} (y_t - x_t^\top \widehat{\beta}_{J_l}^\lambda)^2 &\le \sum_{l=1}^{q+1} \sum_{t \in J_l} (y_t - x_t^\top \beta_t^*)^2 + (q+1)\max\left\{C_3 d_0 \lambda^2,\, C_4 \lambda \sqrt{\log(n \vee p)} d_0^{3/2}\right\} \\ &= \sum_{t \in \widehat{I}} (y_t - x_t^\top \beta_t^*)^2 + (q+1)\max\left\{C_3 d_0 \lambda^2,\, C_4 \lambda \sqrt{\log(n \vee p)} d_0^{3/2}\right\}. \end{aligned} \tag{35}$$

Then for each $l \in \{1, \ldots, q+1\}$,

$$\sum_{t \in J_l} (y_t - x_t^\top \widehat{\beta}_{\widehat{I}}^\lambda)^2 \ge \sum_{t \in J_l} (y_t - x_t^\top \beta_t^*)^2 - C_6 \lambda^2 d_0,$$

where the inequality follows from Lemma 12. Therefore the above inequality implies that

$$\sum_{t \in \widehat{I}} (y_t - x_t^\top \widehat{\beta}_{\widehat{I}}^\lambda)^2 \ge \sum_{t \in \widehat{I}} (y_t - x_t^\top \beta_t^*)^2 - C_6(q+1)\lambda^2 d_0. \tag{36}$$

Note that (35) and (36) implies (34).

Finally, to show (31), observe that from (33), it suffices to show that

$$S_n(\eta_1, \ldots, \eta_K) - S_n(\widehat{\eta}_1, \ldots, \widehat{\eta}_{\widehat{K}}, \eta_1, \ldots, \eta_K) \le C(K + \widehat{K})\lambda^2,$$

the analysis of which follows from a similar but simpler argument as above. $\square$

2 Proof of Corollary 2

Lemma 13. *Let $\mathcal{S}$ be any linear subspace in $\mathbb{R}^n$ and $\mathcal{N}_{1/4}$ be a 1/4-net of $\mathcal{S} \cap B(0,1)$, where $B(0,1)$ is the unit ball in $\mathbb{R}^n$. For any $u \in \mathbb{R}^n$, it holds that*

$$\sup_{v \in \mathcal{S} \cap B(0,1)} \langle v, u \rangle \leq 2 \sup_{v \in \mathcal{N}_{1/4}} \langle v, u \rangle,$$

where $\langle \cdot, \cdot \rangle$ denotes the inner product in $\mathbb{R}^n$.

Proof. Due to the definition of $\mathcal{N}_{1/4}$, it holds that for any $v \in \mathcal{S} \cap B(0,1)$, there exists a $v_k \in \mathcal{N}_{1/4}$, such that $\|v - v_k\|_2 < 1/4$. Therefore,

$$\langle v, u \rangle = \langle v - v_k + v_k, u \rangle = \langle x_k, u \rangle + \langle v_k, u \rangle \leq \frac{1}{4}\langle v, u \rangle + \frac{1}{4}\langle v^{\perp}, u \rangle + \langle v_k, u \rangle,$$

where the inequality follows from $x_k = v - v_k = \langle x_k, v \rangle v + \langle x_k, v^{\perp} \rangle v^{\perp}$. Then we have

$$\frac{3}{4}\langle v, u \rangle \leq \frac{1}{4}\langle v^{\perp}, u \rangle + \langle v_k, u \rangle.$$

It follows from the same argument that

$$\frac{3}{4}\langle v^{\perp}, u \rangle \leq \frac{1}{4}\langle v, u \rangle + \langle v_l, u \rangle,$$

where $v_l \in \mathcal{N}_{1/4}$ satisfies $\|v^{\perp} - v_l\|_2 < 1/4$. Combining the previous two equation displays yields

$$\langle v, u \rangle \leq 2 \sup_{v \in \mathcal{N}_{1/4}} \langle v, u \rangle,$$

and the final claims holds. $\square$

Lemma 14 is an adaptation of Lemma 3 in Wang *et al.* (2019).

Lemma 14. *For data generated from Model 1, for any interval $I = (s, e] \subset \{1, \ldots, n\}$, it holds that for any $\delta > 0$, $i \in \{1, \ldots, p\}$,*

$$\mathbb{P}\left\{ \sup_{\substack{v \in \mathbb{R}^{(e-s)}, \, \|v\|_2 = 1 \\ \sum_{t=1}^{e-s-1} \mathbb{1}\{v_i \neq v_{i+1}\} = m}} \left| \sum_{t=s+1}^{e} v_t \varepsilon_t x_t(i) \right| > \delta \right\} \leq C(e-s-1)^m 9^{m+1} \exp\left\{ -c \min\left\{ \frac{\delta^2}{4C_x^2}, \frac{\delta}{2C_x \|v\|_\infty} \right\} \right\}.$$

Proof. For any $v \in \mathbb{R}^{(e-s)}$ satisfying $\sum_{t=1}^{e-s-1} \mathbb{1}\{v_i \neq v_{i+1}\} = m$, it is determined by a vector in $\mathbb{R}^{m+1}$ and a choice of m out of $(e-s-1)$ points. Therefore we have,

$$\begin{aligned}
&\mathbb{P}\left\{ \sup_{\substack{v \in \mathbb{R}^{(e-s)}, \, \|v\|_2 = 1 \\ \sum_{t=1}^{e-s-1} \mathbb{1}\{v_i \neq v_{i+1}\} = m}} \left| \sum_{t=s+1}^{e} v_t \varepsilon_t x_t(i) \right| > \delta \right\} \\
\leq & \binom{e-s-1}{m} 9^{m+1} \sup_{v \in \mathcal{N}_{1/4}} \mathbb{P}\left\{ \left| \sum_{t=s+1}^{e} v_t \varepsilon_t x_t(i) \right| > \delta/2 \right\} \\
\leq & \binom{e-s-1}{m} 9^{m+1} C \exp\left\{ -c \min\left\{ \frac{\delta^2}{4C_x^2}, \frac{\delta}{2C_x \|v\|_\infty} \right\} \right\} \\
\leq & C(e-s-1)^m 9^{m+1} \exp\left\{ -c \min\left\{ \frac{\delta^2}{4C_x^2}, \frac{\delta}{2C_x \|v\|_\infty} \right\} \right\}.
\end{aligned}$$

$\square$

Proof of Corollary 2. For each $k \in \{1, \dots, K\}$, let

$$\widehat{\beta}_t = \begin{cases} \widehat{\beta}_1, & t \in \{s_k + 1, \dots, \widehat{\eta}_k\}, \\ \widehat{\beta}_2, & t \in \{\widehat{\eta}_k + 1, \dots, e_k\}. \end{cases}$$

Without loss of generality, we assume that $s_k < \eta_k < \widehat{\eta}_k < e_k$. We proceed the proof discussing two cases.

Case (i). If

$$\widehat{\eta}_k - \eta_k < \max\{288^2 C_x^2 d_0 \log(n \vee p)/c_x^2,\, C_\varepsilon \log(n \vee p)/\kappa^2\},$$

then the result holds.

Case (ii). If

$$\widehat{\eta}_k - \eta_k \geq \max\{288^2 C_x^2 d_0 \log(n \vee p)/c_x^2,\, C_\varepsilon \log(n \vee p)/\kappa^2\}, \tag{37}$$

then we first to prove that with probability at least $1 - C(n \vee p)^{-c}$,

$$\sum_{t=s_k+1}^{e_k} \|\widehat{\beta}_t - \beta_t^*\|_2^2 \leq C_1 d_0 \zeta^2 = \delta.$$

Due to (4), it holds that

$$\sum_{t=s_k+1}^{e_k} \|y_t - x_t^\top \widehat{\beta}_t\|_2^2 + \zeta \sum_{i=1}^{p} \sqrt{\sum_{t=s_k+1}^{e_k} \left(\widehat{\beta}_t\right)_i^2} \leq \sum_{t=s_k+1}^{e_k} \|y_t - x_t^\top \beta_t^*\|_2^2 + \zeta \sum_{i=1}^{p} \sqrt{\sum_{t=s_k+1}^{e_k} \left(\beta_t^*\right)_i^2}. \tag{38}$$

Let $\Delta_t = \widehat{\beta}_t - \beta_t^*$. It holds that

$$\sum_{t=s_k+1}^{e_k-1} \mathbb{1}\left\{\Delta_t \neq \Delta_{t+1}\right\} = 2.$$

Eq.(38) implies that

$$\sum_{t=s_k+1}^{e_k} \|\Delta_t^\top x_t\|_2^2 + \zeta \sum_{i=1}^{p} \sqrt{\sum_{t=s_k+1}^{e_k} \left(\widehat{\beta}_t\right)_i^2} \leq 2 \sum_{t=s_k+1}^{e_k} (y_t - x_t^\top \beta_t^*) \Delta_t^\top x_t + \zeta \sum_{i=1}^{p} \sqrt{\sum_{t=s_k+1}^{e_k} \left(\beta_t^*\right)_i^2}. \tag{39}$$

Note that

$$\begin{aligned} &\sum_{i=1}^{p} \sqrt{\sum_{t=s_k+1}^{e_k} \left(\beta_t^*\right)_i^2} - \sum_{i=1}^{p} \sqrt{\sum_{t=s_k+1}^{e_k} \left(\widehat{\beta}_t\right)_i^2} = \sum_{i \in S} \sqrt{\sum_{t=s_k+1}^{e_k} \left(\beta_t^*\right)_i^2} - \sum_{i \in S} \sqrt{\sum_{t=s_k+1}^{e_k} \left(\widehat{\beta}_t\right)_i^2} - \sum_{i \in S^c} \sqrt{\sum_{t=s_k+1}^{e_k} \left(\widehat{\beta}_t\right)_i^2} \\ \leq &\sum_{i \in S} \sqrt{\sum_{t=s_k+1}^{e_k} \left(\Delta_t\right)_i^2} - \sum_{i \in S^c} \sqrt{\sum_{t=s_k+1}^{e_k} \left(\Delta_t\right)_i^2}. \end{aligned} \tag{40}$$

We then examine the cross term, with probability at least $1 - C(n \vee p)^{-c}$, which satisfies the following

$$\begin{aligned} &\left| \sum_{t=s_k+1}^{e_k} (y_t - x_t^\top \beta_t^*) \Delta_t^\top x_t \right| = \left| \sum_{t=s_k+1}^{e_k} \varepsilon_t \Delta_t^\top x_t \right| = \sum_{i=1}^{p} \left\{ \left| \frac{\sum_{t=s_k+1}^{e_k} \varepsilon_t \Delta_t(i) x_t(i)}{\sqrt{\sum_{t=s_k+1}^{e_k} \left(\Delta_t(i)\right)^2}} \right| \sqrt{\sum_{t=s_k+1}^{e_k} \left(\Delta_t(i)\right)^2} \right\} \\ \leq &\sup_{i=1,\dots,p} \left| \frac{\sum_{t=s_k+1}^{e_k} \varepsilon_t \Delta_t(i) X_t(i)}{\sqrt{\sum_{t=s_k+1}^{e_k} \left(\Delta_t(i)\right)^2}} \right| \sum_{i=1}^{p} \sqrt{\sum_{t=s_k+1}^{e_k} \left(\Delta_t(i)\right)^2} \leq (\zeta/4) \sum_{i=1}^{p} \sqrt{\sum_{t=s_k+1}^{e_k} \left(\Delta_t(i)\right)^2}, \end{aligned} \tag{41}$$

where the second inequality follows from Lemma 14 and (37).

Combining (38), (39), (40) and (41) yields

$$\sum_{t=s_k+1}^{e_k} \|\Delta_t^\top x_t\|_2^2 + \frac{\zeta}{2}\sum_{i\in S^c}\sqrt{\sum_{t=s_k+1}^{e_k}(\Delta_t)_i^2} \le \frac{3\zeta}{2}\sum_{i\in S}\sqrt{\sum_{t=s_k+1}^{e_k}(\Delta_t)_i^2}. \tag{42}$$

Now we are to explore the restricted eigenvalue inequality. Let

$$I_1 = (s_k, \eta_k], \quad I_2 = (\eta_k, \widehat{\eta}_k], \quad I_3 = (\widehat{\eta}_k, e_k].$$

We have that with probability at least $1 - C(n\vee p)^{-c}$, on the event $\cap_{i=1,3}\mathcal{E}_{I_i}$,

$$\begin{aligned}
&\sum_{t=s_k+1}^{e_k} \|\Delta_t^\top x_t\|_2^2 = \sum_{i=1}^{3}\sum_{t\in I_i}\|\Delta_{I_i}^\top x_t\|_2^2 \ge \sum_{i=1,3}\sum_{t\in I_i}\|\Delta_{I_i}^\top x_t\|_2^2 \\
\ge &\sum_{i=1,3}\left(\frac{c_x\sqrt{|I_i|}}{4}\|\Delta_{I_i}\|_2 - 9C_x\sqrt{\log(p)}\|\Delta_{I_i}\|_1\right)^2 \\
\ge &\sum_{i=1,3}\left(\frac{c_x\sqrt{|I_i|}}{8}\|\Delta_{I_i}\|_2 - 9C_x\sqrt{\log(p)}\|\Delta_{I_i}(S^c)\|_1\right)^2,
\end{aligned}$$

where the last inequality follows from (8) and Assumption 1, that

$$\min\{|I_1|,\, |I_3|\} > (1/3)\Delta > 288^2 C_x^2 d_0 \log(n\vee p)/c_x^2.$$

Since $|I_2| > 288^2 C_x^2 d_0 \log(n\vee p)/c_x^2$, we have

$$\sqrt{\sum_{t\in I_2}\|\Delta_{I_2}^\top x_t\|_2^2} \ge \frac{c_x\sqrt{|I_2|}}{8}\|\Delta_{I_2}\|_2 - 9C_x\sqrt{\log(p)}\|\Delta_{I_2}(S^c)\|_1.$$

Note that

$$\begin{aligned}
&\sqrt{\sum_{i=1}^{3}\left(\sum_{j\in S^c}|\Delta_{I_i}(j)|\right)^2} \le \sqrt{\sum_{i=1}^{3}\left(\sqrt{\frac{|I_i|}{I_0}}\sum_{j\in S^c}|\Delta_{I_i}(j)|\right)^2} \\
\le &\sum_{j\in S^c} I_0^{-1/2}\sqrt{\sum_{t=s_k+1}^{e_k}(\Delta_t(i))^2} \le 3\sum_{j\in S} I_0^{-1/2}\sqrt{\sum_{t=s_k+1}^{e_k}(\Delta_t(i))^2} \\
\le &I_0^{-1/2} 3\sqrt{d_0\sum_{j\in S}\sum_{t=s_k+1}^{e_k}(\Delta_t(i))^2} \le \frac{c_x}{96C_x\sqrt{\log(n\vee p)}}\sqrt{\sum_{t=s_k+1}^{e_k}\|\Delta_t\|_2^2}.
\end{aligned}$$

Therefore,

$$\begin{aligned}
&\frac{c_x}{8}\sqrt{\sum_{t=s_k+1}^{e_k}\|\Delta_t\|_2^2} - \frac{3c_x}{32C_x\sqrt{\log(n\vee p)}}\sqrt{\sum_{t=s_k+1}^{e_k}\|\Delta_t\|_2^2} \\
\le &\sum_{i=1}^{3}\frac{c_x\sqrt{|I_i|}}{8}\|\Delta_{I_i}\|_2 - \frac{3c_x}{32C_x\sqrt{\log(n\vee p)}}\sqrt{\sum_{t=s_k+1}^{e_k}\|\Delta_t\|_2^2} \le \sqrt{3}\sqrt{\sum_{t=s_k+1}^{e_k}\|\Delta_t^\top x_t\|_2^2} \\
\le &\frac{3\sqrt{\zeta}}{\sqrt{2}} d_0^{1/4}\left(\sum_{t=s_k+1}^{e_k}\|\Delta_t\|_2^2\right)^{1/4} \le \frac{18\zeta d_0^{1/2}}{c_x} + \frac{c_x}{16}\sqrt{\sum_{t=s_k+1}^{e_k}\|\Delta_t\|_2^2}
\end{aligned}$$

where the last inequality follows from (42) and which implies

$$\frac{c_x}{32}\sqrt{\sum_{t=s_k+1}^{e_k}\|\Delta_t\|_2^2} \leq \frac{18\zeta d_0^{1/2}}{c_x}$$

Therefore,

$$\sum_{t=s_k+1}^{e_k}\|\widehat{\beta}_t - \beta_t^*\|_2^2 \leq 576^2\zeta^2 d_0/c_x^4.$$

Let $\beta_1^* = \beta_{\eta_k}^*$ and $\beta_2^* = \beta_{\eta_k+1}^*$. We have that

$$\sum_{t=s_k+1}^{e_k}\|\widehat{\beta}_t - \beta_t^*\|_2^2 = |I_1|\|\beta_1^* - \widehat{\beta}_1\|_2^2 + |I_2|\|\beta_2^* - \widehat{\beta}_1\|_2^2 + |I_3|\|\beta_2^* - \widehat{\beta}_2\|_2^2.$$

Since

$$\begin{aligned}
&\eta_k - s_k = \eta_k - \frac{2}{3}\widetilde{\eta}_k - \frac{1}{3}\widetilde{\eta}_k \\
=&\frac{2}{3}(\eta_k - \eta_{k-1}) + \frac{2}{3}(\widetilde{\eta}_k - \eta_k) - \frac{2}{3}(\widetilde{\eta}_{k-1} - \eta_{k-1}) + (\eta_k - \widetilde{\eta}_k) \\
\geq&\frac{2}{3}\Delta - \frac{1}{3}\Delta = \frac{1}{3}\Delta,
\end{aligned}$$

where the inequality follows from Assumption 1 and (8), we have that

$$\Delta\|\beta_1^* - \widehat{\beta}_1\|_2^2/3 \leq |I_1|\|\beta_1^* - \widehat{\beta}_1\|_2^2 \leq \delta \leq \frac{C_1 C_\zeta^2 \Delta\kappa^2}{C_{\mathrm{SNR}} d_0 K \sigma_\epsilon^2 \log^\xi(n \vee p)} \leq c_1\Delta\kappa^2,$$

where $1/4 > c_1 > 0$ is an arbitrarily small positive constant. Therefore we have

$$\|\beta_1^* - \widehat{\beta}_1\|_2^2 \leq c_1\kappa^2.$$

In addition we have

$$\|\beta_2^* - \widehat{\beta}_1\|_2 \geq \|\beta_2^* - \beta_1^*\|_2 - \|\beta_1^* - \widehat{\beta}_1\|_2 \geq \kappa/2.$$

Therefore, it holds that

$$\kappa^2|I_2|/4 \leq |I_2|\|\beta_2^* - \widehat{\beta}_1\|_2^2 \leq \delta,$$

which implies that

$$|\widehat{\eta}_k - \eta_k| \leq \frac{4C_1 d_0\zeta^2}{\kappa^2}.$$

$\square$

3 Lower bounds

Proof of Lemma 3. For any vector β, if $x \sim \mathcal{N}(0, I_p)$, $\epsilon \sim \mathcal{N}(0, \sigma^2)$ and $y = x^\top\beta + \epsilon$, then we denote

$$\begin{pmatrix} y \\ x \end{pmatrix} \sim \mathcal{N}(0, \Sigma_\beta), \quad \text{where} \quad \Sigma_\beta = \begin{pmatrix} \beta^\top\beta + \sigma^2 & \beta^\top \\ \beta & I \end{pmatrix}.$$

Now for a fixed $S \subset \{1, \ldots, p\}$ satisfying $|S| = d$, define

$$\mathcal{S} = \left\{u \in \mathbb{R}^p : u_i = 0, i \notin S; u_i = \kappa/\sqrt{d} \text{ or } -\kappa/\sqrt{d}, i \in S\right\}.$$

Define

$$P_0 = \mathcal{N}\left(0, \Sigma_0\right) \quad \text{and} \quad P_u = \mathcal{N}\left(0, \Sigma_u\right), \quad \forall u \in \mathcal{S},$$

where

$$\Sigma_0 = \begin{pmatrix} \sigma^2 & 0 \\ 0 & I_p \end{pmatrix} \quad \text{and} \quad \Sigma_u = \begin{pmatrix} \sigma^2 + \kappa^2 & u^\top \\ u & I_p \end{pmatrix}.$$

Step 1. Let $P_{0,u}^T$ denote the joint distribution of independent random vectors $\{Z_i = (y_i, x_i^\top)^\top\}_{i=1}^T \subset \mathbb{R}^{p+1}$ such that

$$Z_1, \ldots, Z_\Delta \overset{\text{iid}}{\sim} \mathcal{N}(0, \Sigma_u) \quad \text{and} \quad Z_{\Delta+1}, \ldots, Z_T \overset{\text{iid}}{\sim} \mathcal{N}(0, \Sigma_0).$$

Let $P_{1,u}^T$ denote the joint distribution of independent random vectors $\{Z_i = (y_i, x_i^\top)^\top\}_{i=1}^T \subset \mathbb{R}^{p+1}$ such that

$$Z_1, \ldots, Z_{T-\Delta} \overset{\text{iid}}{\sim} \mathcal{N}(0, \Sigma_0) \quad \text{and} \quad Z_{T-\Delta+1}, \ldots, Z_T \overset{\text{iid}}{\sim} \mathcal{N}(0, \Sigma_u).$$

For $i \in \{0, 1\}$, let

$$P_i = 2^{-d} \sum_{u \in \mathcal{S}} P_{i,u}^T.$$

Let $\eta(P)$ denote the change point location of a distribution P. Then since $\eta(P_{0,u}) = \Delta$ and $\eta(P_{1,u}) = T - \Delta$ for any $u \in \mathcal{S}$, we have that

$$|\eta(P_0) - \eta(P_1)| = T - 2\Delta \geq T/2,$$

due to the fact that $\Delta \leq T/4$. It follows from Le Cam's lemma (Yu, 1997) that

$$\inf_{\widehat{\eta}} \sup_{P \in \mathcal{P}} \mathbb{E}_P(|\widehat{\eta} - \eta|) \geq T/2(1 - d_{\mathrm{TV}}(P_0, P_1)),$$

where $d_{\mathrm{TV}}(P_0, P_1) = \|P_0 - P_1\|_1/2$, with $\|P_0 - P_1\|_1$ denoting the L_1 distance between the Lebesgue densities of the distributions P_0 and P_1. Then we have that

$$\inf_{\widehat{\eta}} \sup_{P \in \mathcal{P}} \mathbb{E}_P(|\widehat{\eta} - \eta|) \geq T/2(1 - 2^{-1}\|P_0 - P_1\|_1).$$

Step 2. Let P_0^Δ be the joint distribution of

$$Z_1, \ldots, Z_\Delta \overset{\text{iid}}{\sim} \mathcal{N}(0, \Sigma_0)$$

and $P_1^\Delta = 2^{-d} \sum_{u \in \mathcal{S}} P_{1,u}^\Delta$, where P_1^Δ is the joint distribution of

$$Z_1, \ldots, Z_\Delta \overset{\text{iid}}{\sim} \mathcal{N}(0, \Sigma_u).$$

It follows from Step 2 in the proof of Lemma 3.1 in Wang *et al.* (2017) that

$$\|P_0 - P_1\|_1 \leq 2\|P_0^\Delta - P_1^\Delta\|_1,$$

which leads to

$$\inf_{\widehat{\eta}} \sup_{P \in \mathcal{P}} \mathbb{E}_P(|\widehat{\eta} - \eta|) \geq T/2(1 - \|P_0^\Delta - P_1^\Delta\|_1) \geq T/2(1 - \sqrt{\chi^2(P_1^\Delta, P_0^\Delta)}),$$

where the last inequality follows from Tsybakov (2008).

Note that

$$\chi^2(P_1^\Delta, P_0^\Delta) = \mathbb{E}_{P_p^\Delta}\left\{\left(\frac{dP_1^\Delta}{dP_0^\Delta} - 1\right)^2\right\} = \frac{1}{4^d} \sum_{u,v \in \mathcal{S}} \mathbb{E}_{P_0^\Delta}\left(\frac{dP_u^\Delta dP_v^\Delta}{dP_0^\Delta dP_0^\Delta}\right) - 1$$
$$= \frac{1}{4^d} \sum_{u,v \in \mathcal{S}} \left\{\mathbb{E}_{P_0}\left(\frac{dP_u dP_v}{dP_0 dP_0}\right)\right\}^\Delta - 1.$$

Step 3. For any $u, v \in \mathcal{S}$, we have that

$$
\begin{aligned}
&\mathbb{E}_{P_0}\left(\frac{dP_u dP_v}{dP_0 dP_0}\right) \\
=&\frac{|\Sigma_u|^{-1/2}|\Sigma_v|^{-1/2}}{|\Sigma_0|^{-1/2}}(2\pi)^{-\frac{p+1}{2}}\int_{\mathbb{R}^{p+1}}\exp\left\{-\frac{z^\top(\Sigma_u^{-1}+\Sigma_v^{-1}-\Sigma_0^{-1})z}{2}\right\}dz \\
=&\frac{|\Sigma_u|^{-1/2}|\Sigma_v|^{-1/2}}{|\Sigma_0|^{-1/2}}|\Sigma_u^{-1}+\Sigma_v^{-1}-\Sigma_0^{-1}|^{-1/2}.
\end{aligned}
$$

In addition, we have that

$$
|\Sigma_u| = |\Sigma_v| = |\Sigma_0| = \sigma^2, \quad \Sigma_0^{-1} = \begin{pmatrix} \sigma^{-2} & 0 \\ 0 & I \end{pmatrix},
$$

$$
\Sigma_u^{-1} = \begin{pmatrix} \sigma^{-2} & -\sigma^{-2}u^\top \\ -\sigma^{-2}u & I+\sigma^{-2}uu^\top \end{pmatrix} \quad \text{and} \quad \Sigma_v^{-1} = \begin{pmatrix} \sigma^{-2} & -\sigma^{-2}v^\top \\ -\sigma^{-2}v & I+\sigma^{-2}vv^\top \end{pmatrix}.
$$

Then

$$
\mathbb{E}_{P_0}\left(\frac{dP_u dP_v}{dP_0 dP_0}\right) = \sigma^p\left|\begin{pmatrix} 1 & -(u+v)^\top \\ -(u+v) & \sigma^2 I_p + uu^\top + vv^\top \end{pmatrix}\right|^{-1/2} = \sigma^p|M|^{-1/2}.
$$

Note that

$$
|M| = \left|\left\{1-(u+v)^\top\left(\sigma^2 I_p + uu^\top + vv^\top\right)^{-1}(u+v)\right\}\right|\left|\sigma^2 I_p + uu^\top + vv^\top\right|.
$$

As for the matrix $M_1 = \sigma^2 I_p + uu^\top + vv^\top$, since $u, v \neq 0$, there are two cases. Let

$$
\rho_{u,v} = \frac{u^\top v}{\kappa^2}.
$$

- The dimension of the linear space spanned by u and v is one, i.e. $|\rho| = 1$. In this case, for any $w \perp \text{span}\{u\}$, $\|w\|_2 = 1$, it holds that

  $$
  M_1 w = \sigma^2 w.
  $$

  There are $p-1$ such linearly independent w. For any $w \in \text{span}\{u\}$, $\|w\|_2 = 1$, it holds that

  $$
  M_1 w = (\sigma^2 + 2\kappa^2)w.
  $$

  Then $|M_1| = \sigma^{2p-2}(\sigma^2 + 2\kappa^2)$.
  If $\rho_{u,v} = -1$, then $|M| = |M_1| = \sigma^{2p-2}(\sigma^2 + 2\kappa^2)$.
  If $\rho_{u,v} = 1$, then

  $$
  \begin{aligned}
  |M| &= \left|1 - 4u^\top\frac{u}{\kappa}\frac{1}{\sigma^2+2\kappa^2}\frac{u^\top}{\kappa}u\right|\sigma^{2p-2}(\sigma^2+2\kappa^2) \\
  &= \frac{|\sigma^2 - 2\kappa^2|}{\sigma^2+2\kappa^2}\sigma^{2p-2}(\sigma^2+2\kappa^2) = \sigma^{2p-2}|\sigma^2 - 2\kappa^2|.
  \end{aligned}
  $$

  Therefore in this case

  $$
  |M| = \sigma^{2p-2}|\sigma^2 - 2\rho_{u,v}\kappa^2|.
  $$

- The dimension of the linear space spanned by u and v is two, i.e. $|\rho| < 1$. In this case, for any $w \perp \text{span}\{u\}$, $\|w\|_2 = 1$, it holds that

  $$
  M_1 w = \sigma^2 w.
  $$

  There are $p-2$ such linearly independent w.

  We also have

  $$
  M_1\frac{u+v}{\|u+v\|} = (\sigma^2 + \kappa^2 + \rho_{u,v}\kappa^2)\frac{u+v}{\|u+v\|}
  $$

and

$$M_1 \frac{u-v}{\|u-v\|} = (\sigma^2 + \kappa^2 - \rho_{u,v}\kappa^2)\frac{u-v}{\|u-v\|}.$$

Then

$$|M_1| = \sigma^{2p-4}(\sigma^2 + \kappa^2 + \rho_{u,v}\kappa^2)(\sigma^2 + \kappa^2 - \rho_{u,v}\kappa^2)$$

In addition,

$$\begin{aligned}
&(u+v)^\top \left(\sigma^2 I_p + uu^\top + vv^\top\right)^{-1}(u+v) \\
=&(u+v)^\top \frac{u+v}{\|u+v\|}\frac{1}{\sigma^2+\kappa^2+\rho_{u,v}\kappa^2}\left(\frac{u+v}{\|u+v\|}\right)^\top (u+v) \\
&\quad + (u+v)^\top \frac{u-v}{\|u-v\|}\frac{1}{\sigma^2+\kappa^2-\rho_{u,v}\kappa^2}\left(\frac{u-v}{\|u-v\|}\right)^\top (u+v) \\
=&\frac{2\kappa^2 + 2\kappa^2\rho_{u,v}}{\sigma^2+\kappa^2+\rho_{u,v}\kappa^2}.
\end{aligned}$$

Then,

$$|M| = \sigma^{2p-4}|\sigma^2 - \kappa^2 - \rho_{u,v}\kappa^2|(\sigma^2 + \kappa^2 - \rho_{u,v}\kappa^2),$$

which is consistent with the case when $|\rho_{u,v}| = 1$.

We then have

$$\mathbb{E}_{P_0}\left(\frac{dP_u dP_v}{dP_0 dP_0}\right) = \left|1 - \frac{\kappa^2}{\sigma^2} - \frac{u^\top v}{\sigma^2}\right|^{-1/2}\left|1 + \frac{\kappa^2}{\sigma^2} - \frac{u^\top v}{\sigma^2}\right|^{-1/2}.$$

Due to the fact that $cd/\Delta < 1/4$, we have that

$$1 - \frac{\kappa^2}{\sigma^2} - \frac{u^\top v}{\sigma^2} \geq 1 - \frac{2\kappa^2}{\sigma^2} \geq 1 - \frac{2cd}{\Delta} > 0,$$

then

$$\begin{aligned}
\mathbb{E}_{P_0}\left(\frac{dP_u dP_v}{dP_0 dP_0}\right) &= \left(1 - \frac{\kappa^2}{\sigma^2} - \frac{u^\top v}{\sigma^2}\right)^{-1/2}\left(1 + \frac{\kappa^2}{\sigma^2} - \frac{u^\top v}{\sigma^2}\right)^{-1/2} \\
= \left(1 - \frac{2u^\top v}{\sigma^2} - \frac{\kappa^4}{\sigma^4} + \frac{(u^\top v)^2}{\sigma^4}\right)^{-1/2} &\leq \left(1 - \frac{2u^\top v}{\sigma^2} - \frac{\kappa^4}{\sigma^4}\right)^{-1/2}
\end{aligned}$$

Then we have

$$\begin{aligned}
&\chi^2(P_1^\Delta, P_0^\Delta) \leq \frac{1}{4^d}\sum_{u,v\in\mathcal{S}}\left(1 - \frac{2u^\top v}{\sigma^2} - \frac{\kappa^4}{\sigma^4}\right)^{-\Delta/2} - 1 \\
=&\mathbb{E}_{U,V}\left\{1 - \frac{\kappa^2}{\sigma^2}(U^\top V/d)^2 - \frac{\kappa^4}{\sigma^4}\right\}^{-\Delta/2} - 1 = \mathbb{E}_V\left\{1 - \frac{\kappa^2}{\sigma^2}(1^\top V/d)^2 - \frac{\kappa^4}{\sigma^4}\right\}^{-\Delta/2} - 1 \\
\leq&\mathbb{E}\left\{\exp\left(\frac{\kappa^2\Delta}{\sigma^2}\varepsilon_d + \frac{\kappa^4\Delta}{\sigma^4}\right)\right\} - 1,
\end{aligned}$$

where U and V are two independent d-dimensional Radamacher random vectors, $\varepsilon_d = (1^\top V/d)^2$, and the last inequality follows from $(1-t)^{-\Delta/2} \leq \exp(\Delta t)$, for any $t \leq 1/2$.

Due to the Hoeffding inequality, it holds that for any $\lambda > 0$,

$$\mathbb{P}(\varepsilon_d \geq \lambda) \leq 2e^{-2d\lambda}.$$

Then

$$\mathbb{E}\left\{\exp\left(\frac{\kappa^2\Delta}{\sigma^2}\varepsilon_d + \frac{\kappa^4\Delta}{\sigma^4}\right)\right\} = \int_0^\infty \mathbb{P}\left\{\exp\left(\frac{\kappa^2\Delta}{\sigma^2}\varepsilon_d + \frac{\kappa^4\Delta}{\sigma^4}\right) \geq u\right\} du$$

$$
\begin{aligned}
&\leq 1+\int_1^\infty \mathbb{P}\left\{\frac{\kappa^2\Delta}{\sigma^2}\varepsilon_d+\frac{\kappa^4\Delta}{\sigma^4}\geq \log(u)\right\}du = 1+\int_1^\infty \mathbb{P}\left\{\varepsilon_d \geq \frac{\log(u)-\frac{\kappa^4\Delta}{\sigma^4}}{\frac{\kappa^2\Delta}{\sigma^2}}\right\}du\\
&\leq 1+2\int_1^\infty \exp\left\{-\frac{2d\sigma^2}{\kappa^2\Delta}\log(u)+\frac{2d\kappa^2}{\sigma^2}\right\}du\\
&=1+\frac{2\exp(2d\kappa^2\sigma^{-2})}{\frac{2d\sigma^2}{\kappa^2\Delta}-1}\leq 1+\frac{2e}{2/c-1}\leq 5/4,
\end{aligned}
$$

where the last two inequalities hold due to

$$
2cd^2 \leq \Delta \quad \text{and} \quad c<\frac{2}{8e+1}.
$$

We then complete the proof.

□

Proof of Lemma 4. For any vector β, if $x \sim \mathcal{N}(0, I_p)$, $\epsilon \sim \mathcal{N}(0,\sigma^2)$ and $y = x^\top \beta + \epsilon$, then we denote

$$
\begin{pmatrix} y \\ x \end{pmatrix} \sim \mathcal{N}\left(0, \Sigma_\beta\right), \quad \text{where} \quad \Sigma_\beta = \begin{pmatrix} \beta^\top\beta+\sigma^2 & \beta^\top \\ \beta & I \end{pmatrix}.
$$

Now for a fixed $S \subset \{1,\ldots,p\}$ satisfying $|S| = d$, define

$$
\mathcal{S} = \left\{u \in \mathbb{R}^p : \, u_i = 0, i \notin S; u_i = \kappa/\sqrt{d} \text{ or } -\kappa/\sqrt{d}, i \in S\right\}.
$$

Define

$$
P_0 = \mathcal{N}\left(0, \Sigma_0\right) \quad \text{and} \quad P_u = \mathcal{N}\left(0, \Sigma_u\right), \quad \forall u \in \mathcal{S},
$$

where

$$
\Sigma_0 = \begin{pmatrix} \sigma^2 & 0 \\ 0 & I_p \end{pmatrix} \quad \text{and} \quad \Sigma_u = \begin{pmatrix} \sigma^2+\kappa^2 & u^\top \\ u & I_p \end{pmatrix}.
$$

Step 1. Let $P_{0,u}^T$ denote the joint distribution of independent random vectors $\{Z_i = (y_i, x_i^\top)^\top\}_{i=1}^T \subset \mathbb{R}^{p+1}$ such that

$$
Z_1,\ldots,Z_\Delta \overset{\text{iid}}{\sim} \mathcal{N}(0,\Sigma_u) \quad \text{and} \quad Z_{\Delta+1},\ldots,Z_T \overset{\text{iid}}{\sim} \mathcal{N}(0,\Sigma_0).
$$

Let $P_{1,u}^T$ denote the joint distribution of independent random vectors $\{Z_i = (y_i, x_i^\top)^\top\}_{i=1}^T \subset \mathbb{R}^{p+1}$ such that

$$
Z_1,\ldots,Z_{\Delta+\delta} \overset{\text{iid}}{\sim} \mathcal{N}(0,\Sigma_u) \quad \text{and} \quad Z_{\Delta+\delta+1},\ldots,Z_T \overset{\text{iid}}{\sim} \mathcal{N}(0,\Sigma_0).
$$

For $i \in \{0,1\}$, let

$$
P_i = 2^{-d}\sum_{u\in\mathcal{S}} P_{i,u}^T.
$$

Then we have that

$$
\inf_{\widehat{\eta}} \sup_{P\in\mathcal{P}} \mathbb{E}_P(|\widehat{\eta}-\eta|) \geq \delta(1-2^{-1}\|P_0-P_1\|_1).
$$

Step 2. Let P_0^δ be the joint distribution of

$$
Z_1,\ldots,Z_\delta \overset{\text{iid}}{\sim} \mathcal{N}(0,\Sigma_0)
$$

and $P_1^\delta = 2^{-d}\sum_{u\in\mathcal{S}} P_{1,u}^\delta$, where $P_{1,u}^\delta$ is the joint distribution of

$$
Z_1,\ldots,Z_\delta \overset{\text{iid}}{\sim} \mathcal{N}(0,\Sigma_u).
$$

It follows from the identical arguments in the proof of Lemma 3 that

$$\inf_{\widehat{\eta}} \sup_{P \in \mathcal{P}} \mathbb{E}_P(|\widehat{\eta} - \eta|) \geq \delta(1 - \|P_0^\delta - P_1^\delta\|_1) \geq \delta(1 - \sqrt{\chi^2(P_1^\delta, P_0^\delta)})$$

and

$$\chi^2(P_1^\delta, P_0^\delta) = \frac{1}{4^d} \sum_{u,v \in \mathcal{S}} \left\{ \mathbb{E}_{P_0} \left(\frac{dP_u dP_v}{dP_0 dP_0} \right) \right\}^\delta - 1 \leq \frac{2\exp(2d\kappa^2\sigma^{-2})}{\frac{2d\sigma^2}{\kappa^2\delta} - 1}.$$

Step 3. Let

$$\delta = \frac{Cd\sigma^2}{\kappa^2}.$$

We have that

$$\chi^2(P_1^\delta, P_0^\delta) = 1/4,$$

provided that $d^2\zeta_T\Delta^{-1} < 1$ and with $C = 2/(8e+1)$. Then we conclude the proof. $\square$

3.1 Numerical Results

In Table 1, we provide a detailed summery of the numerical results for the simulated experiments conducted in Section 4.2.

Setting	Cases	DP	DP.LR	EBSA	EBSA.LR
$\kappa = 4, d_0 = 10$		0.023(0.015)	0.008(0.004)	0.104(0.031)	0.034(0.045)
$\kappa = 4, d_0 = 15$	All	0.031(0.020)	0.017(0.047)	0.104(0.029)	0.038(0.050)
$\kappa = 4, d_0 = 20$		0.038(0.032)	0.019(0.042)	0.104(0.027)	0.036(0.051)
$\kappa = 4, d_0 = 10$		0.022(0.015)	0.008(0.004)	0.061(0.047)	0.008(0.008)
$\kappa = 4, d_0 = 15$	$\hat{K} = K$	0.025(0.018)	0.008(0.007)	0.071(0.045)	0.010(0.016)
$\kappa = 4, d_0 = 20$		0.028(0.020)	0.014(0.012)	0.076(0.048)	0.010(0.011)
$\kappa = 5, d_0 = 10$		0.022(0.022)	0.007(0.004)	0.102(0.033)	0.033(0.046)
$\kappa = 5, d_0 = 15$	All	0.025(0.023)	0.015(0.025)	0.102(0.030)	0.027(0.042)
$\kappa = 5, d_0 = 20$		0.030(0.027)	0.016(0.030)	0.102(0.030)	0.041(0.048)
$\kappa = 5, d_0 = 10$		0.020(0.015)	0.007(0.004)	0.068(0.073)	0.007(0.008)
$\kappa = 5, d_0 = 15$	$\hat{K} = K$	0.021(0.012)	0.010(0.006)	0.075(0.049)	0.007(0.008)
$\kappa = 5, d_0 = 20$		0.025(0.018)	0.010(0.007)	0.076(0.065)	0.010(0.012)
$\kappa = 6, d_0 = 10$		0.009(0.010)	0.007(0.004)	0.100(0.028)	0.034(0.049)
$\kappa = 6, d_0 = 15$	All	0.022(0.017)	0.009(0.005)	0.101(0.029)	0.037(0.049)
$\kappa = 6, d_0 = 20$		0.023(0.017)	0.010(0.006)	0.102(0.031)	0.028(0.043)
$\kappa = 6, d_0 = 10$		0.009(0.010)	0.007(0.004)	0.061(0.064)	0.006(0.010)
$\kappa = 6, d_0 = 15$	$\hat{K} = K$	0.022(0.017)	0.009(0.005)	0.064(0.050)	0.007(0.007)
$\kappa = 6, d_0 = 20$		0.023(0.017)	0.010(0.006)	0.076(0.041)	0.009(0.013)

Table 1: Scaled Hausdorff Distance. The numbers in the brackets indicate the corresponding standard errors of the scaled Hausdorff distance.

References

Bickel, Peter J, Ritov, Ya'acov, & Tsybakov, Alexandre B. 2009. Simultaneous analysis of Lasso and Dantzig selector. *The Annals of Statistics*, **37**(4), 1705–1732.

Bühlmann, Peter, & van de Geer, Sara. 2011. *Statistics for high-dimensional data: methods, theory and applications*. Springer Science & Business Media.

Raskutti, Garvesh, Wainwright, Martin J, & Yu, Bin. 2010. Restricted eigenvalue properties for correlated Gaussian designs. *Journal of Machine Learning Research*, **11**(Aug), 2241–2259.

Tsybakov, Alexandre B. 2008. *Introduction to nonparametric estimation*. Springer Science & Business Media.

van de Geer, Sara A, & Bühlmann, Peter. 2009. On the conditions used to prove oracle results for the Lasso. *Electronic Journal of Statistics*, **3**, 1360–1392.

Vershynin, Roman. 2018. *High-dimensional probability: An introduction with applications in data science*. Vol. 47. Cambridge University Press.

Wang, Daren, Yu, Yi, & Rinaldo, Alessandro. 2017. Optimal Covariance Change Point Localization in High Dimension. *arXiv preprint arXiv:1712.09912*.

Wang, Daren, Lin, Kevin, & Willett, Rebecca. 2019. Statistically and Computationally Efficient Change Point Localization in Regression Settings. *arXiv preprint arXiv:1906.11364*.

Yu, Bin. 1997. *Festschrift for Lucien Le Cam*. Vol. 423. Springer Science & Business Media. Chap. Assouad, Fano, and Le Cam, page 435.