

## Supplementary Material

### A. Proof of Theorem 1

We begin with three technical lemmas.

**Lemma 2** *Let  $\mathbf{y} \neq \mathbf{0}$  and  $0 < \lambda_1 \leq \|X^T \mathbf{y}\|_\infty$ . We have*

$$\left\langle \frac{\mathbf{y}}{\lambda_1} - \boldsymbol{\theta}_1^*, \boldsymbol{\theta}_1^* \right\rangle \geq 0. \quad (45)$$

**Proof** Since the Euclidean projection of  $\frac{\mathbf{y}}{\lambda_1}$  onto  $\{\boldsymbol{\theta} : \|X^T \boldsymbol{\theta}\|_\infty \leq 1\}$  is  $\boldsymbol{\theta}_1^*$ , it follows from Lemma 1 that

$$\left\langle \boldsymbol{\theta}_1^* - \frac{\mathbf{y}}{\lambda_1}, \boldsymbol{\theta} - \boldsymbol{\theta}_1^* \right\rangle \geq 0, \forall \boldsymbol{\theta} : \|X^T \boldsymbol{\theta}\|_\infty \leq 1. \quad (46)$$

As  $\mathbf{0} \in \{\boldsymbol{\theta} : \|X^T \boldsymbol{\theta}\|_\infty \leq 1\}$ , we have Eq. (45).  $\square$

**Lemma 3** *Let  $\mathbf{y} \neq \mathbf{0}$  and  $0 < \lambda_1 \leq \|X^T \mathbf{y}\|_\infty$ . If  $\boldsymbol{\theta}_1^*$  parallels to  $\mathbf{y}$  in that it can be written as  $\boldsymbol{\theta}_1^* = \gamma \mathbf{y}$  for some  $\gamma$ , then  $\gamma = \frac{1}{\|X^T \mathbf{y}\|_\infty}$ .*

**Proof** Since  $\frac{\mathbf{y}}{\|X^T \mathbf{y}\|_\infty}$  satisfies the condition in Eq. (11), we have

$$\left\langle \gamma \mathbf{y} - \frac{\mathbf{y}}{\lambda_1}, \frac{\mathbf{y}}{\|X^T \mathbf{y}\|_\infty} - \gamma \mathbf{y} \right\rangle = \left( \gamma - \frac{1}{\lambda_1} \right) \left( \frac{1}{\|X^T \mathbf{y}\|_\infty} - \gamma \right) \|\mathbf{y}\|_2^2 \geq 0 \quad (47)$$

which leads to  $\gamma \in \left[ \frac{1}{\|X^T \mathbf{y}\|_\infty}, \frac{1}{\lambda_1} \right]$ . In addition, since  $\|X^T \boldsymbol{\theta}_1^*\|_\infty \leq 1$ , we have  $\gamma = \frac{1}{\|X^T \mathbf{y}\|_\infty}$ . This completes the proof.  $\square$

**Lemma 4** *Let  $\mathbf{y} \neq \mathbf{0}$ . If  $0 < \lambda_1 \leq \|X^T \mathbf{y}\|_\infty$ , we have*

$$\left\langle \frac{\mathbf{y}}{\lambda_1} - \boldsymbol{\theta}_1^*, \mathbf{y} \right\rangle \geq 0, \quad (48)$$

where the equality holds if and only if  $\lambda_1 = \|X^T \mathbf{y}\|_\infty$ .

**Proof** We have

$$\left\langle \frac{\mathbf{y}}{\lambda_1} - \boldsymbol{\theta}_1^*, \frac{\mathbf{y}}{\lambda_1} \right\rangle - \left\langle \frac{\mathbf{y}}{\lambda_1} - \boldsymbol{\theta}_1^*, \boldsymbol{\theta}_1^* \right\rangle = \left\langle \frac{\mathbf{y}}{\lambda_1} - \boldsymbol{\theta}_1^*, \frac{\mathbf{y}}{\lambda_1} - \boldsymbol{\theta}_1^* \right\rangle \geq 0, \quad (49)$$

where the equality holds if and only if  $\frac{\mathbf{y}}{\lambda_1} = \boldsymbol{\theta}_1^*$ . Incorporating Eq. (45) in Lemma 2 and Eq. (49), we have Eq. (48). The equality in Eq. (49) holds if and only if  $\frac{\mathbf{y}}{\lambda_1} = \boldsymbol{\theta}_1^*$ . According to Lemma 3, if  $\boldsymbol{\theta}_1^* = \frac{\mathbf{y}}{\lambda_1}$ , then  $\boldsymbol{\theta}_1^* = \frac{\mathbf{y}}{\|X^T \mathbf{y}\|_\infty}$ , which leads to  $\lambda_1 = \|X^T \mathbf{y}\|_\infty$ . This ends the proof.  $\square$

Now, we are ready to prove Theorem 1. It follows from Eq. (17) and Eq. (48)

$$\langle \mathbf{b}, \mathbf{a} \rangle = \left( \frac{1}{\lambda_2} - \frac{1}{\lambda_1} \right) \left\langle \frac{\mathbf{y}}{\lambda_1} - \boldsymbol{\theta}_1^*, \mathbf{y} \right\rangle + \left\| \frac{\mathbf{y}}{\lambda_1} - \boldsymbol{\theta}_1^* \right\|_2^2 \quad (50)$$

$$\|\mathbf{b}\|_2^2 = \left\| \left( \frac{\mathbf{y}}{\lambda_2} - \frac{\mathbf{y}}{\lambda_1} \right) \right\|_2^2 + 2 \left( \frac{1}{\lambda_2} - \frac{1}{\lambda_1} \right) \left\langle \frac{\mathbf{y}}{\lambda_1} - \boldsymbol{\theta}_1^*, \mathbf{y} \right\rangle + \left\| \frac{\mathbf{y}}{\lambda_1} - \boldsymbol{\theta}_1^* \right\|_2^2 \geq 0. \quad (51)$$

It follows from Lemma 4 that 1)  $\langle \mathbf{b}, \mathbf{a} \rangle \geq 0$  and the equality holds if and only if  $\frac{\mathbf{y}}{\lambda_1} = \boldsymbol{\theta}_1^*$ , and 2)  $\|\mathbf{b}\|_2^2 > 0$ , which leads to  $\mathbf{b} \neq \mathbf{0}$ . According to Lemma 3, if  $\boldsymbol{\theta}_1^*$  parallels to  $\mathbf{y}$ , then  $\boldsymbol{\theta}_1^* = \frac{\mathbf{y}}{\|X^T \mathbf{y}\|_\infty}$ . Therefore, if  $0 < \lambda_1 < \|X^T \mathbf{y}\|_\infty$ , then  $\langle \mathbf{b}, \mathbf{a} \rangle > 0$  and  $\mathbf{a} \neq \mathbf{0}$ .  $\square$

## B. Proof of Theorem 2

If  $\lambda_1 = \|X^T \mathbf{y}\|_\infty$ , the primal and dual optimals can be analytically computed as:  $\beta_1^* = \mathbf{0}$  and  $\theta_1^* = \frac{\mathbf{y}}{\|X^T \theta\|_\infty}$ . Thus, we have  $\mathbf{a} = \mathbf{0}$ . It is easy to get that  $\mathbf{r} = -\frac{\mathbf{x}\|\mathbf{b}\|_2}{\|\mathbf{x}\|_2}$  minimizes Eq. (20) with the minimum function value being

$$\langle \mathbf{x}, \mathbf{r} \rangle = -\|\mathbf{x}\|_2 \|\mathbf{b}\|_2. \quad (52)$$

In our following discussion, we focus on the case  $0 < \lambda_1 < \|X^T \mathbf{y}\|_\infty$  and we have  $\mathbf{a} \neq \mathbf{0}$  according to Theorem 1.

The Lagrangian of Eq. (20) can be written as

$$L(\mathbf{r}, \alpha, \beta) = \langle \mathbf{x}, \mathbf{r} \rangle + \alpha \langle \mathbf{a}, \mathbf{r} + \mathbf{b} \rangle + \frac{\beta}{2} (\|\mathbf{r}\|_2^2 - \|\mathbf{b}\|_2^2), \quad (53)$$

where  $\alpha, \beta \geq 0$  are introduced for the two inequalities, respectively. It is clear that the minimal value of Eq. (20) is lower bounded (the minimum is no less than  $-\|\mathbf{b}\|_2 \|\mathbf{x}\|_2$  by only considering the constraint  $\|\mathbf{r}\|_2^2 \leq \|\mathbf{b}\|_2^2$ ). Therefore, the optimal dual variable  $\beta$  is always positive; otherwise, minimizing Eq. (53) with regard to  $\mathbf{r}$  achieves  $-\infty$ .

Setting the derivative with regard to  $\mathbf{r}$  to zero, we have

$$\mathbf{r} = \frac{-\mathbf{x} - \alpha \mathbf{a}}{\beta}. \quad (54)$$

Plugging Eq. (54) into Eq. (53), we obtain the dual problem of Eq. (20) as:

$$\begin{aligned} \max_{\alpha, \beta} \quad & \alpha \langle \mathbf{a}, \mathbf{b} \rangle - \frac{1}{2\beta} \|\mathbf{x} + \alpha \mathbf{a}\|_2^2 - \frac{\beta}{2} \|\mathbf{b}\|_2^2 \\ \text{subject to} \quad & \alpha \geq 0, \beta \geq 0. \end{aligned} \quad (55)$$

For a given  $\beta$ , we have

$$\alpha = \max \left( \frac{\beta \langle \mathbf{a}, \mathbf{b} \rangle - \langle \mathbf{x}, \mathbf{a} \rangle}{\|\mathbf{a}\|_2^2}, 0 \right). \quad (56)$$

We consider two cases. In the first case, we assume that  $\alpha = 0$ . We have

$$\mathbf{r} = \frac{-\mathbf{x}}{\beta}, \beta \leq \frac{\langle \mathbf{x}, \mathbf{a} \rangle}{\langle \mathbf{a}, \mathbf{b} \rangle}. \quad (57)$$

By using the complementary slackness condition (note that the optimal  $\beta$  does not equal to zero), we have

$$\|\mathbf{r}\|_2 = \left\| \frac{-\mathbf{x}}{\beta} \right\|_2 = \|\mathbf{b}\|_2. \quad (58)$$

Thus, we have

$$\beta = \frac{\|\mathbf{x}\|_2}{\|\mathbf{b}\|_2}. \quad (59)$$

Incorporating Eq. (57) and Eq. (59), we have

$$\frac{\langle \mathbf{b}, \mathbf{a} \rangle}{\|\mathbf{b}\|_2 \|\mathbf{a}\|_2} \leq \frac{\langle \mathbf{x}, \mathbf{a} \rangle}{\|\mathbf{x}\|_2 \|\mathbf{a}\|_2}, \quad (60)$$

so that the angle between  $\mathbf{a}$  and  $\mathbf{b}$  is equal to or larger than the angle between  $\mathbf{x}$  and  $\mathbf{a}$ . Note that  $\langle \mathbf{b}, \mathbf{a} \rangle \geq 0$  according to Theorem 1. In Figure 2, EX<sub>2</sub> and EX<sub>3</sub> illustrate the case that  $\mathbf{x}$  satisfies Eq. (60), while EX<sub>1</sub> and EX<sub>4</sub> show the opposite cases. In addition, we have

$$\langle \mathbf{x}, \mathbf{r} \rangle = -\|\mathbf{x}\|_2 \|\mathbf{b}\|_2. \quad (61)$$

In the second case, Eq. (60) does not hold. We have

$$\alpha = \frac{\beta \langle \mathbf{a}, \mathbf{b} \rangle - \langle \mathbf{x}, \mathbf{a} \rangle}{\|\mathbf{a}\|_2^2}. \quad (62)$$

Plugging Eq. (62) into Eq. (54), we have

$$\mathbf{r} = -\frac{\mathbf{x}\|\mathbf{a}\|_2^2 + \beta\langle\mathbf{a}, \mathbf{b}\rangle\mathbf{a} - \langle\mathbf{x}, \mathbf{a}\rangle\mathbf{a}}{\beta\|\mathbf{a}\|_2^2} \quad (63)$$

Since  $\|\mathbf{r}\|_2^2 = \|\mathbf{b}\|_2^2$ , we have

$$\beta = \sqrt{\frac{\|\mathbf{x}\|_2^2\|\mathbf{a}\|_2^2 - \langle\mathbf{x}, \mathbf{a}\rangle^2}{\|\mathbf{b}\|_2^2\|\mathbf{a}\|_2^2 - \langle\mathbf{b}, \mathbf{a}\rangle^2}} = \frac{\|\mathbf{x}^\perp\|_2}{\sqrt{\|\mathbf{b}\|_2^2 - \frac{\langle\mathbf{b}, \mathbf{a}\rangle^2}{\|\mathbf{a}\|_2^2}}}, \quad (64)$$

where we have used Eq. (21) to get the second equality. In addition, we have

$$\langle\mathbf{x}, \mathbf{r}\rangle = -\|\mathbf{x}^\perp\|_2\sqrt{\|\mathbf{b}\|_2^2 - \frac{\langle\mathbf{b}, \mathbf{a}\rangle^2}{\|\mathbf{a}\|_2^2}} - \frac{\langle\mathbf{a}, \mathbf{b}\rangle\langle\mathbf{x}, \mathbf{a}\rangle}{\|\mathbf{a}\|_2^2}. \quad (65)$$

In summary, Eq. (20) equals to  $-\|\mathbf{x}\|_2\|\mathbf{b}\|_2$ , if  $\frac{\langle\mathbf{b}, \mathbf{a}\rangle}{\|\mathbf{b}\|_2} \leq \frac{\langle\mathbf{x}, \mathbf{a}\rangle}{\|\mathbf{x}\|_2}$ , and  $-\|\mathbf{x}^\perp\|_2\sqrt{\|\mathbf{b}\|_2^2 - \frac{\langle\mathbf{b}, \mathbf{a}\rangle^2}{\|\mathbf{a}\|_2^2}} - \frac{\langle\mathbf{a}, \mathbf{b}\rangle\langle\mathbf{x}, \mathbf{a}\rangle}{\|\mathbf{a}\|_2^2}$  otherwise. This ends the proof of this theorem.  $\square$

### C. Proof of Theorem 3

We prove the four cases one by one as follows.

**Case 1** If  $\mathbf{a} \neq \mathbf{0}$  and  $\frac{\langle\mathbf{b}, \mathbf{a}\rangle}{\|\mathbf{b}\|_2} > \frac{|\langle\mathbf{x}_j, \mathbf{a}\rangle|}{\|\mathbf{x}_j\|_2}$ , i.e., Eq. (60) does not hold with  $\mathbf{x} = \pm\mathbf{x}_j$ . We have

$$\begin{aligned} u_j^+(\lambda_2) &= \max_{\theta: \langle\theta_1^* - \frac{\mathbf{y}}{\lambda_1}, \theta - \theta_1^*\rangle \geq 0, \langle\theta - \frac{\mathbf{y}}{\lambda_2}, \theta_1^* - \theta\rangle \geq 0} \langle\mathbf{x}_j, \theta\rangle \\ &= \frac{1}{2} \max_{\mathbf{r}: \langle\mathbf{a}, \mathbf{r} + \mathbf{b}\rangle \leq 0, \|\mathbf{r}\|_2^2 \leq \|\mathbf{b}\|_2^2} \left[ \langle\mathbf{x}_j, \theta_1^* + \frac{\mathbf{y}}{\lambda_2}\rangle + \langle\mathbf{x}_j, \mathbf{r}\rangle \right] \\ &= \frac{1}{2} \left[ \langle\mathbf{x}_j, \theta_1^* + \frac{\mathbf{y}}{\lambda_2}\rangle + \max_{\mathbf{r}: \langle\mathbf{a}, \mathbf{r} + \mathbf{b}\rangle \leq 0, \|\mathbf{r}\|_2^2 \leq \|\mathbf{b}\|_2^2} \langle\mathbf{x}_j, \mathbf{r}\rangle \right] \\ &= \frac{1}{2} \left[ \langle\mathbf{x}_j, \theta_1^* + \frac{\mathbf{y}}{\lambda_2}\rangle - \min_{\mathbf{r}: \langle\mathbf{a}, \mathbf{r} + \mathbf{b}\rangle \leq 0, \|\mathbf{r}\|_2^2 \leq \|\mathbf{b}\|_2^2} \langle-\mathbf{x}_j, \mathbf{r}\rangle \right] \\ &= \frac{1}{2} \left[ \langle\mathbf{x}_j, 2\theta_1^* + \left(\frac{\mathbf{y}}{\lambda_1} - \theta_1^*\right) + \left(\frac{\mathbf{y}}{\lambda_2} - \frac{\mathbf{y}}{\lambda_1}\right)\rangle \right] \\ &\quad + \frac{1}{2} \left[ \|\mathbf{x}_j^\perp\|_2\sqrt{\|\mathbf{b}\|_2^2 - \frac{\langle\mathbf{b}, \mathbf{a}\rangle^2}{\|\mathbf{a}\|_2^2}} + \frac{\langle\mathbf{a}, \mathbf{b}\rangle\langle-\mathbf{x}_j, \mathbf{a}\rangle}{\|\mathbf{a}\|_2^2} \right] \\ &= \langle\mathbf{x}_j, \theta_1^*\rangle + \frac{\frac{1}{\lambda_2} - \frac{1}{\lambda_1}}{2} [\langle\mathbf{x}_j, \mathbf{y}\rangle - \frac{\langle\mathbf{a}, \mathbf{y}\rangle}{\|\mathbf{a}\|_2^2} \langle\mathbf{x}_j, \mathbf{a}\rangle] \\ &\quad + \frac{\frac{1}{\lambda_2} - \frac{1}{\lambda_1}}{2} \|\mathbf{x}_j^\perp\|_2\sqrt{\|\mathbf{y}\|_2^2 - \frac{\langle\mathbf{y}, \mathbf{a}\rangle^2}{\|\mathbf{a}\|_2^2}}. \end{aligned} \quad (66)$$

The second equality plugs in the notations in Eq. (17). The fifth equality utilizes Eq. (65) which is the result for the case  $\frac{\langle\mathbf{b}, \mathbf{a}\rangle}{\|\mathbf{b}\|_2} > \frac{|\langle\mathbf{x}_j, \mathbf{a}\rangle|}{\|\mathbf{x}_j\|_2} \geq \frac{\langle-\mathbf{x}_j, \mathbf{a}\rangle}{\|\mathbf{x}_j\|_2}$  by setting  $\mathbf{x} = -\mathbf{x}_j$ . To get the last equality, we utilize the following two equalities

$$\|\mathbf{b}\|_2^2 - \frac{\langle\mathbf{b}, \mathbf{a}\rangle^2}{\|\mathbf{a}\|_2^2} = \left(\frac{1}{\lambda_2} - \frac{1}{\lambda_1}\right)^2 (\|\mathbf{y}\|_2^2 - \frac{\langle\mathbf{y}, \mathbf{a}\rangle^2}{\|\mathbf{a}\|_2^2}) \quad (67)$$

and

$$\frac{\langle\mathbf{a}, \mathbf{b}\rangle\langle\mathbf{x}_j, \mathbf{a}\rangle}{\|\mathbf{a}\|_2^2} = \langle\mathbf{x}_j, \mathbf{a}\rangle \left(1 + \frac{\langle\mathbf{a}, \mathbf{y}\rangle \left(\frac{1}{\lambda_2} - \frac{1}{\lambda_1}\right)}{\|\mathbf{a}\|_2^2}\right), \quad (68)$$

which can be derived from Eq. (17). It follows from Eq. (22) and Eq. (23) that

$$\|\mathbf{x}_j^\perp\|_2^2 = \|\mathbf{x}_j\|_2^2 - \frac{\langle \mathbf{x}_j, \mathbf{a} \rangle^2}{\|\mathbf{a}\|_2^2}, \quad (69)$$

$$\|\mathbf{y}^\perp\|_2^2 = \|\mathbf{y}\|_2^2 - \frac{\langle \mathbf{y}, \mathbf{a} \rangle^2}{\|\mathbf{a}\|_2^2}, \quad (70)$$

$$\langle \mathbf{x}_j^\perp, \mathbf{y}^\perp \rangle = \langle \mathbf{x}_j, \mathbf{y} \rangle - \frac{\langle \mathbf{a}, \mathbf{y} \rangle}{\|\mathbf{a}\|_2^2} \langle \mathbf{x}_j, \mathbf{a} \rangle. \quad (71)$$

Incorporating Eq. (66), and Eqs. (70)-(71), we have Eq. (26). Following a similar derivation, we have

$$\begin{aligned} u_j^-(\lambda_2) &= \max_{\boldsymbol{\theta}: \langle \boldsymbol{\theta}_1^* - \frac{\mathbf{y}}{\lambda_1}, \boldsymbol{\theta} - \boldsymbol{\theta}_1^* \rangle \geq 0, \langle \boldsymbol{\theta} - \frac{\mathbf{y}}{\lambda_2}, \boldsymbol{\theta}_1^* - \boldsymbol{\theta} \rangle \geq 0} \langle -\mathbf{x}_j, \boldsymbol{\theta} \rangle \\ &= \frac{1}{2} \max_{\mathbf{r}: \langle \mathbf{a}, \mathbf{r} + \mathbf{b} \rangle \leq 0, \|\mathbf{r}\|_2^2 \leq \|\mathbf{b}\|_2^2} \left[ \langle -\mathbf{x}_j, \boldsymbol{\theta}_1^* + \frac{\mathbf{y}}{\lambda_2} \rangle + \langle -\mathbf{x}_j, \mathbf{r} \rangle \right] \\ &= \frac{1}{2} \left[ \langle -\mathbf{x}_j, \boldsymbol{\theta}_1^* + \frac{\mathbf{y}}{\lambda_2} \rangle + \max_{\mathbf{r}: \langle \mathbf{a}, \mathbf{r} + \mathbf{b} \rangle \leq 0, \|\mathbf{r}\|_2^2 \leq \|\mathbf{b}\|_2^2} \langle -\mathbf{x}_j, \mathbf{r} \rangle \right] \\ &= \frac{1}{2} \left[ \langle -\mathbf{x}_j, \boldsymbol{\theta}_1^* + \frac{\mathbf{y}}{\lambda_2} \rangle - \min_{\mathbf{r}: \langle \mathbf{a}, \mathbf{r} + \mathbf{b} \rangle \leq 0, \|\mathbf{r}\|_2^2 \leq \|\mathbf{b}\|_2^2} \langle \mathbf{x}_j, \mathbf{r} \rangle \right] \\ &= \frac{1}{2} \left[ \langle -\mathbf{x}_j, 2\boldsymbol{\theta}_1^* + \left( \frac{\mathbf{y}}{\lambda_1} - \boldsymbol{\theta}_1^* \right) + \left( \frac{\mathbf{y}}{\lambda_2} - \frac{\mathbf{y}}{\lambda_1} \right) \rangle \right] \\ &\quad + \frac{1}{2} \left[ \|\mathbf{x}_j^\perp\|_2 \sqrt{\|\mathbf{b}\|_2^2 - \frac{\langle \mathbf{b}, \mathbf{a} \rangle^2}{\|\mathbf{a}\|_2^2} + \frac{\langle \mathbf{a}, \mathbf{b} \rangle \langle \mathbf{x}_j, \mathbf{a} \rangle}{\|\mathbf{a}\|_2^2}} \right] \\ &= -\langle \mathbf{x}_j, \boldsymbol{\theta}_1^* \rangle - \frac{\frac{1}{\lambda_2} - \frac{1}{\lambda_1}}{2} [\langle \mathbf{x}_j, \mathbf{y} \rangle - \frac{\langle \mathbf{a}, \mathbf{y} \rangle}{\|\mathbf{a}\|_2^2} \langle \mathbf{x}_j, \mathbf{a} \rangle] \\ &\quad + \frac{\frac{1}{\lambda_2} - \frac{1}{\lambda_1}}{2} \|\mathbf{x}_j^\perp\|_2 \sqrt{\|\mathbf{y}\|_2^2 - \frac{\langle \mathbf{y}, \mathbf{a} \rangle^2}{\|\mathbf{a}\|_2^2}}. \end{aligned} \quad (72)$$

The fifth equality utilizes Eq. (65) which is the result for the case  $\frac{\langle \mathbf{b}, \mathbf{a} \rangle}{\|\mathbf{b}\|_2} > \frac{|\langle \mathbf{x}_j, \mathbf{a} \rangle|}{\|\mathbf{x}_j\|_2} \geq \frac{\langle \mathbf{x}_j, \mathbf{a} \rangle}{\|\mathbf{x}_j\|_2}$  by setting  $\mathbf{x} = \mathbf{x}_j$ . The last equality can be obtained using the similar derivation getting the last equality of Eq. (66). Incorporating Eqs. (70)-(72), we have Eq. (27).

**Case 2** If  $\frac{\langle \mathbf{b}, \mathbf{a} \rangle}{\|\mathbf{b}\|_2} \leq \frac{\langle \mathbf{x}_j, \mathbf{a} \rangle}{\|\mathbf{x}_j\|_2}$  and  $\langle \mathbf{x}_j, \mathbf{a} \rangle > 0$ , we have  $\frac{\langle \mathbf{b}, \mathbf{a} \rangle}{\|\mathbf{b}\|_2} > \frac{\langle -\mathbf{x}_j, \mathbf{a} \rangle}{\|\mathbf{x}_j\|_2}$  since  $\langle \mathbf{b}, \mathbf{a} \rangle \geq 0$  according to Theorem 1. Thus, Eq. (60) does not hold with  $\mathbf{x} = -\mathbf{x}_j$ , and we can get Eq. (66), or equivalently Eq. (26). In addition, Eq. (60) holds with  $\mathbf{x} = \mathbf{x}_j$ , and we have

$$\begin{aligned} u_j^-(\lambda_2) &= \max_{\boldsymbol{\theta}: \langle \boldsymbol{\theta}_1^* - \frac{\mathbf{y}}{\lambda_1}, \boldsymbol{\theta} - \boldsymbol{\theta}_1^* \rangle \geq 0, \langle \boldsymbol{\theta} - \frac{\mathbf{y}}{\lambda_2}, \boldsymbol{\theta}_1^* - \boldsymbol{\theta} \rangle \geq 0} \langle -\mathbf{x}_j, \boldsymbol{\theta} \rangle \\ &= \max_{\mathbf{r}: \langle \mathbf{a}, \mathbf{r} + \mathbf{b} \rangle \leq 0, \|\mathbf{r}\|_2^2 \leq \|\mathbf{b}\|_2^2} \left[ \langle -\mathbf{x}_j, \frac{\boldsymbol{\theta}_1^* + \frac{\mathbf{y}}{\lambda_2}}{2} \rangle + \frac{1}{2} \langle -\mathbf{x}_j, \mathbf{r} \rangle \right] \\ &= \langle -\mathbf{x}_j, \frac{\boldsymbol{\theta}_1^* + \frac{\mathbf{y}}{\lambda_2}}{2} \rangle + \frac{1}{2} \max_{\mathbf{r}: \langle \mathbf{a}, \mathbf{r} + \mathbf{b} \rangle \leq 0, \|\mathbf{r}\|_2^2 \leq \|\mathbf{b}\|_2^2} \langle -\mathbf{x}_j, \mathbf{r} \rangle \\ &= \langle -\mathbf{x}_j, \frac{\boldsymbol{\theta}_1^* + \frac{\mathbf{y}}{\lambda_2}}{2} \rangle - \frac{1}{2} \min_{\mathbf{r}: \langle \mathbf{a}, \mathbf{r} + \mathbf{b} \rangle \leq 0, \|\mathbf{r}\|_2^2 \leq \|\mathbf{b}\|_2^2} \langle \mathbf{x}_j, \mathbf{r} \rangle \\ &= \langle -\mathbf{x}_j, \boldsymbol{\theta}_1^* + \frac{1}{2} \left( \frac{\mathbf{y}}{\lambda_2} - \boldsymbol{\theta}_1^* \right) \rangle + \frac{1}{2} \|\mathbf{x}_j\|_2 \|\mathbf{b}\|_2 \\ &= -\langle \mathbf{x}_j, \boldsymbol{\theta}_1^* \rangle + \frac{1}{2} [\|\mathbf{x}_j\|_2 \|\mathbf{b}\|_2 - \langle \mathbf{x}_j, \mathbf{b} \rangle]. \end{aligned} \quad (73)$$

To get the fifth equality, we utilize Eq. (61) with  $\mathbf{x} = \mathbf{x}_j$ . Therefore, we have Eq. (28).

**Case 3** If  $\frac{\langle \mathbf{b}, \mathbf{a} \rangle}{\|\mathbf{b}\|_2} \leq \frac{-\langle \mathbf{x}_j, \mathbf{a} \rangle}{\|\mathbf{x}_j\|_2}$  and  $\langle \mathbf{x}_j, \mathbf{a} \rangle < 0$ , Eq. (60) holds with  $\mathbf{x} = -\mathbf{x}_j$ , and we have

$$\begin{aligned}
 u_j^+(\lambda_2) &= \max_{\boldsymbol{\theta}: \langle \boldsymbol{\theta}_1^* - \frac{\mathbf{y}}{\lambda_1}, \boldsymbol{\theta} - \boldsymbol{\theta}_1^* \rangle \geq 0, \langle \boldsymbol{\theta} - \frac{\mathbf{y}}{\lambda_2}, \boldsymbol{\theta}_1^* - \boldsymbol{\theta} \rangle \geq 0} \langle \mathbf{x}_j, \boldsymbol{\theta} \rangle \\
 &= \max_{\mathbf{r}: \langle \mathbf{a}, \mathbf{r} + \mathbf{b} \rangle \leq 0, \|\mathbf{r}\|_2^2 \leq \|\mathbf{b}\|_2^2} \left[ \langle \mathbf{x}_j, \frac{\boldsymbol{\theta}_1^* + \frac{\mathbf{y}}{\lambda_2}}{2} \rangle + \frac{1}{2} \langle \mathbf{x}_j, \mathbf{r} \rangle \right] \\
 &= \langle \mathbf{x}_j, \frac{\boldsymbol{\theta}_1^* + \frac{\mathbf{y}}{\lambda_2}}{2} \rangle + \frac{1}{2} \max_{\mathbf{r}: \langle \mathbf{a}, \mathbf{r} + \mathbf{b} \rangle \leq 0, \|\mathbf{r}\|_2^2 \leq \|\mathbf{b}\|_2^2} \langle \mathbf{x}_j, \mathbf{r} \rangle \\
 &= \langle \mathbf{x}_j, \frac{\boldsymbol{\theta}_1^* + \frac{\mathbf{y}}{\lambda_2}}{2} \rangle - \frac{1}{2} \min_{\mathbf{r}: \langle \mathbf{a}, \mathbf{r} + \mathbf{b} \rangle \leq 0, \|\mathbf{r}\|_2^2 \leq \|\mathbf{b}\|_2^2} \langle -\mathbf{x}_j, \mathbf{r} \rangle \\
 &= \langle \mathbf{x}_j, \boldsymbol{\theta}_1^* + \frac{1}{2} \left( \frac{\mathbf{y}}{\lambda_2} - \boldsymbol{\theta}_1^* \right) \rangle + \frac{1}{2} \|-\mathbf{x}_j\|_2 \|\mathbf{b}\|_2 \\
 &= \langle \mathbf{x}_j, \boldsymbol{\theta}_1^* \rangle + \frac{1}{2} [\|\mathbf{x}_j\|_2 \|\mathbf{b}\|_2 + \langle \mathbf{x}_j, \mathbf{b} \rangle],
 \end{aligned} \tag{74}$$

where the fifth equality utilizes Eq. (61) with  $\mathbf{x} = -\mathbf{x}_j$ . Therefore, we have Eq. (29). In addition, we have  $\frac{\langle \mathbf{b}, \mathbf{a} \rangle}{\|\mathbf{b}\|_2} > \frac{\langle \mathbf{x}_j, \mathbf{a} \rangle}{\|\mathbf{x}_j\|_2}$  since  $\langle \mathbf{b}, \mathbf{a} \rangle \geq 0$  according to Theorem 1 and  $\langle \mathbf{x}_j, \mathbf{a} \rangle < 0$ . Thus, Eq. (60) does not hold with  $\mathbf{x} = \mathbf{x}_j$ , and we can get Eq. (72), or equivalently Eq. (27).

**Case 4** If  $\mathbf{a} = \mathbf{0}$ , then we have  $\lambda_1 = \|X^T \mathbf{y}\|_\infty$  according to Theorem 1. Therefore,

$$\begin{aligned}
 u_j^+(\lambda_2) &= \max_{\boldsymbol{\theta}: \langle \boldsymbol{\theta}_1^* - \frac{\mathbf{y}}{\lambda_1}, \boldsymbol{\theta} - \boldsymbol{\theta}_1^* \rangle \geq 0, \langle \boldsymbol{\theta} - \frac{\mathbf{y}}{\lambda_2}, \boldsymbol{\theta}_1^* - \boldsymbol{\theta} \rangle \geq 0} \langle \mathbf{x}_j, \boldsymbol{\theta} \rangle \\
 &= \frac{1}{2} \max_{\mathbf{r}: \langle \mathbf{a}, \mathbf{r} + \mathbf{b} \rangle \leq 0, \|\mathbf{r}\|_2^2 \leq \|\mathbf{b}\|_2^2} \left[ \langle \mathbf{x}_j, \boldsymbol{\theta}_1^* + \frac{\mathbf{y}}{\lambda_2} \rangle + \langle \mathbf{x}_j, \mathbf{r} \rangle \right] \\
 &= \frac{1}{2} \left[ \langle \mathbf{x}_j, \boldsymbol{\theta}_1^* + \frac{\mathbf{y}}{\lambda_2} \rangle + \max_{\mathbf{r}: \langle \mathbf{a}, \mathbf{r} + \mathbf{b} \rangle \leq 0, \|\mathbf{r}\|_2^2 \leq \|\mathbf{b}\|_2^2} \langle \mathbf{x}_j, \mathbf{r} \rangle \right] \\
 &= \frac{1}{2} \left[ \langle \mathbf{x}_j, \boldsymbol{\theta}_1^* + \frac{\mathbf{y}}{\lambda_2} \rangle - \min_{\mathbf{r}: \langle \mathbf{a}, \mathbf{r} + \mathbf{b} \rangle \leq 0, \|\mathbf{r}\|_2^2 \leq \|\mathbf{b}\|_2^2} \langle -\mathbf{x}_j, \mathbf{r} \rangle \right] \\
 &= \frac{1}{2} \left[ \langle \mathbf{x}_j, 2\boldsymbol{\theta}_1^* + \left( \frac{\mathbf{y}}{\lambda_2} - \boldsymbol{\theta}_1^* \right) \rangle + \frac{1}{2} \|-\mathbf{x}_j\|_2 \|\mathbf{b}\|_2 \right]
 \end{aligned} \tag{75}$$

To get the last equality, we utilize Eq. (52) with  $\mathbf{x} = -\mathbf{x}_j$ . Therefore, we have Eq. (74). Similarly,

$$\begin{aligned}
 u_j^-(\lambda_2) &= \max_{\boldsymbol{\theta}: \langle \boldsymbol{\theta}_1^* - \frac{\mathbf{y}}{\lambda_1}, \boldsymbol{\theta} - \boldsymbol{\theta}_1^* \rangle \geq 0, \langle \boldsymbol{\theta} - \frac{\mathbf{y}}{\lambda_2}, \boldsymbol{\theta}_1^* - \boldsymbol{\theta} \rangle \geq 0} \langle -\mathbf{x}_j, \boldsymbol{\theta} \rangle \\
 &= \max_{\mathbf{r}: \langle \mathbf{a}, \mathbf{r} + \mathbf{b} \rangle \leq 0, \|\mathbf{r}\|_2^2 \leq \|\mathbf{b}\|_2^2} \left[ \langle -\mathbf{x}_j, \frac{\boldsymbol{\theta}_1^* + \frac{\mathbf{y}}{\lambda_2}}{2} \rangle + \frac{1}{2} \langle -\mathbf{x}_j, \mathbf{r} \rangle \right] \\
 &= \langle -\mathbf{x}_j, \frac{\boldsymbol{\theta}_1^* + \frac{\mathbf{y}}{\lambda_2}}{2} \rangle + \frac{1}{2} \max_{\mathbf{r}: \langle \mathbf{a}, \mathbf{r} + \mathbf{b} \rangle \leq 0, \|\mathbf{r}\|_2^2 \leq \|\mathbf{b}\|_2^2} \langle -\mathbf{x}_j, \mathbf{r} \rangle \\
 &= \langle -\mathbf{x}_j, \frac{\boldsymbol{\theta}_1^* + \frac{\mathbf{y}}{\lambda_2}}{2} \rangle - \frac{1}{2} \min_{\mathbf{r}: \langle \mathbf{a}, \mathbf{r} + \mathbf{b} \rangle \leq 0, \|\mathbf{r}\|_2^2 \leq \|\mathbf{b}\|_2^2} \langle \mathbf{x}_j, \mathbf{r} \rangle \\
 &= \langle -\mathbf{x}_j, \boldsymbol{\theta}_1^* + \frac{1}{2} \left( \frac{\mathbf{y}}{\lambda_2} - \boldsymbol{\theta}_1^* \right) \rangle + \frac{1}{2} \|\mathbf{x}_j\|_2 \|\mathbf{b}\|_2
 \end{aligned} \tag{76}$$

To get the last equality, we utilize Eq. (52) with  $\mathbf{x} = \mathbf{x}_j$ . Therefore, we have Eq. (73).

This ends the proof of this theorem.  $\square$

## D. Proof of Theorem 4

We begin with a technical lemma. For a geometrical illustration of this lemma, please refer to the first plot of Figure 4.

**Lemma 5** Let  $\mathbf{y} \neq 0$ , and  $\|X^T \mathbf{y}\|_\infty > \lambda_1 > \lambda > 0$ . Suppose that  $\theta_1^* \neq \frac{\mathbf{y}}{\|X^T \mathbf{y}\|_\infty}$ . For the two auxiliary functions defined in Eq. (41) and Eq. (42),  $f(\lambda)$  is strictly increasing with regard to  $\lambda$  in  $(0, \lambda_1]$ .  $g(\lambda)$  is strictly decreasing with regard to  $\lambda$  in  $(0, \lambda_1]$ .

**Proof** Denote  $\gamma = \frac{1}{\lambda} - \frac{1}{\lambda_1}$ . We can rewrite  $f(\lambda)$  as

$$h(\gamma) = \frac{\langle \mathbf{a} + \gamma \mathbf{y}, \mathbf{a} \rangle}{\|\mathbf{a} + \gamma \mathbf{y}\|_2}. \quad (77)$$

The derivative of  $h(\gamma)$  with regard to  $\gamma$  can be computed as

$$h'(\gamma) = \frac{\gamma(\langle \mathbf{a}, \mathbf{y} \rangle^2 - \|\mathbf{y}\|_2^2 \|\mathbf{a}\|_2^2)}{\|\mathbf{a} + \gamma \mathbf{y}\|_2^3} \leq 0 \quad (78)$$

For any  $\gamma > 0$ ,  $h'(\gamma) = 0$  if and only if  $\mathbf{a}$  parallels to  $\mathbf{y}$ . It follows the definition of  $\mathbf{a}$  in Eq. (17) that, if  $\mathbf{a}$  parallels to  $\mathbf{y}$ , then  $\theta_1^*$  parallels  $\mathbf{y}$ . According to Lemma 3, we have  $\theta_1^* = \frac{\mathbf{y}}{\|X^T \mathbf{y}\|_\infty}$ , which contradicts to the assumption  $\theta_1^* \neq \frac{\mathbf{y}}{\|X^T \mathbf{y}\|_\infty}$ . Therefore,  $h'(\gamma) > 0$ ,  $h(\gamma)$  is strictly decreasing  $\forall \gamma > 0$ , and  $f(\lambda)$  is strictly increasing with regard to  $\lambda$  in  $(0, \lambda_1]$ . Following a similar proof, we can show that  $g(\lambda)$  is strictly decreasing with regard to  $\lambda$  in  $(0, \lambda_1]$ .  $\square$

Now, are ready to prove Theorem 4. Firstly, we summarize the  $u_j^+(\lambda_2)$  and  $u_j^-(\lambda_2)$  in unified equations.

Since  $\langle \mathbf{x}_j, \mathbf{a} \rangle \geq 0$ ,  $u_j^+(\lambda_2)$  satisfies Eq. (26) if  $\mathbf{a} \neq 0$ , and Eq. (29) otherwise. Thus, we have

$$u_j^+(\lambda_2) = \begin{cases} \langle \mathbf{x}_j, \theta_1^* \rangle + \frac{\frac{1}{\lambda_2} - \frac{1}{\lambda_1}}{2} [\|\mathbf{x}_j^\perp\|_2 \|\mathbf{y}^\perp\|_2 + \langle \mathbf{x}_j^\perp, \mathbf{y}^\perp \rangle], & \mathbf{a} \neq \mathbf{0} \\ \langle \mathbf{x}_j, \theta_1^* \rangle + \frac{1}{2} [\|\mathbf{x}_j\|_2 \|\mathbf{b}\|_2 + \langle \mathbf{x}_j, \mathbf{b} \rangle], & \mathbf{a} = \mathbf{0} \end{cases} \quad (79)$$

Since  $\langle \mathbf{x}_j, \mathbf{a} \rangle \geq 0$ ,  $u_j^-(\lambda_2)$  satisfies Eq. (28) if  $\frac{\langle \mathbf{b}, \mathbf{a} \rangle}{\|\mathbf{b}\|_2} \leq \frac{\langle \mathbf{x}_j, \mathbf{a} \rangle}{\|\mathbf{x}_j\|_2}$ , and Eq. (27) otherwise. Thus, we have

$$u_j^-(\lambda_2) = \begin{cases} -\langle \mathbf{x}_j, \theta_1^* \rangle + \frac{1}{2} [\|\mathbf{x}_j\|_2 \|\mathbf{b}\|_2 - \langle \mathbf{x}_j, \mathbf{b} \rangle], & \frac{\langle \mathbf{b}, \mathbf{a} \rangle}{\|\mathbf{b}\|_2} \leq \frac{\langle \mathbf{x}_j, \mathbf{a} \rangle}{\|\mathbf{x}_j\|_2} \\ -\langle \mathbf{x}_j, \theta_1^* \rangle + \frac{\frac{1}{\lambda_2} - \frac{1}{\lambda_1}}{2} [\|\mathbf{x}_j^\perp\|_2 \|\mathbf{y}^\perp\|_2 - \langle \mathbf{x}_j^\perp, \mathbf{y}^\perp \rangle], & \frac{\langle \mathbf{b}, \mathbf{a} \rangle}{\|\mathbf{b}\|_2} > \frac{\langle \mathbf{x}_j, \mathbf{a} \rangle}{\|\mathbf{x}_j\|_2} \end{cases} \quad (80)$$

**Case 1** When  $\mathbf{a} = 0$ , we have  $\theta_1^* = \frac{\mathbf{y}}{\lambda_1}$ ,  $\mathbf{b} = \frac{\lambda_2 - \lambda_1}{2} \mathbf{y}$ ,  $\mathbf{x}_j^\perp = \mathbf{x}_j$ , and  $\mathbf{y}^\perp = \mathbf{y}$ . Thus, Eq. (79) can be simplified as:

$$u_j^+(\lambda_2) = \langle \mathbf{x}_j, \theta_1^* \rangle + \frac{\frac{1}{\lambda_2} - \frac{1}{\lambda_1}}{2} [\|\mathbf{x}_j\|_2 \|\mathbf{y}\|_2 + \langle \mathbf{x}_j, \mathbf{y} \rangle]. \quad (81)$$

Since  $\|\mathbf{x}_j\|_2 \|\mathbf{y}\|_2 + \langle \mathbf{x}_j, \mathbf{y} \rangle \geq 0$ ,  $u_j^+(\lambda_2)$  is monotonically decreasing with regard to  $\lambda_2$ .

**Case 2 & Case 3** When  $\frac{\langle \mathbf{b}, \mathbf{a} \rangle}{\|\mathbf{b}\|_2} = \frac{\langle \mathbf{x}_j, \mathbf{a} \rangle}{\|\mathbf{x}_j\|_2}$ , we have

$$\begin{aligned} & \frac{\frac{1}{\lambda_2} - \frac{1}{\lambda_1}}{2} [\|\mathbf{x}_j^\perp\|_2 \|\mathbf{y}^\perp\|_2 - \langle \mathbf{x}_j^\perp, \mathbf{y}^\perp \rangle] \\ &= \frac{\frac{1}{\lambda_2} - \frac{1}{\lambda_1}}{2} \left[ \sqrt{(\|\mathbf{x}_j\|_2^2 - \frac{\langle \mathbf{x}_j, \mathbf{a} \rangle^2}{\|\mathbf{a}\|_2^2})(\|\mathbf{y}\|_2^2 - \frac{\langle \mathbf{y}, \mathbf{a} \rangle^2}{\|\mathbf{a}\|_2^2})} - [\langle \mathbf{x}_j, \mathbf{y} \rangle - \frac{\langle \mathbf{a}, \mathbf{y} \rangle}{\|\mathbf{a}\|_2^2} \langle \mathbf{x}_j, \mathbf{a} \rangle] \right] \\ &= \frac{1}{2} \left[ \sqrt{(\|\mathbf{x}_j\|_2^2 - \frac{\|\mathbf{x}_j\|_2^2 \langle \mathbf{b}, \mathbf{a} \rangle^2}{\|\mathbf{a}\|_2^2 \|\mathbf{b}\|_2^2})(\|\mathbf{b}\|_2^2 - \frac{\langle \mathbf{b}, \mathbf{a} \rangle^2}{\|\mathbf{a}\|_2^2})} - [\langle \mathbf{x}_j, \mathbf{b} - \mathbf{a} \rangle - \frac{\langle \mathbf{a}, \mathbf{b} - \mathbf{a} \rangle}{\|\mathbf{a}\|_2^2} \langle \mathbf{x}_j, \mathbf{a} \rangle] \right] \\ &= \frac{1}{2} \left[ \|\mathbf{x}_j\|_2 \|\mathbf{b}\|_2 \left(1 - \frac{\langle \mathbf{b}, \mathbf{a} \rangle^2}{\|\mathbf{a}\|_2^2 \|\mathbf{b}\|_2^2}\right) - [\langle \mathbf{x}_j, \mathbf{b} - \mathbf{a} \rangle - \frac{\langle \mathbf{a}, \mathbf{b} - \mathbf{a} \rangle}{\|\mathbf{a}\|_2^2} \langle \mathbf{x}_j, \mathbf{a} \rangle] \right] \\ &= \frac{1}{2} [\|\mathbf{x}_j\|_2 \|\mathbf{b}\|_2 - \langle \mathbf{x}_j, \mathbf{b} \rangle] + \frac{1}{2} \left[ -\frac{\|\mathbf{x}_j\|_2 \langle \mathbf{b}, \mathbf{a} \rangle^2}{\|\mathbf{a}\|_2^2 \|\mathbf{b}\|_2} + \frac{\langle \mathbf{a}, \mathbf{b} \rangle}{\|\mathbf{a}\|_2^2} \langle \mathbf{x}_j, \mathbf{a} \rangle \right] \\ &= \frac{1}{2} [\|\mathbf{x}_j\|_2 \|\mathbf{b}\|_2 - \langle \mathbf{x}_j, \mathbf{b} \rangle]. \end{aligned} \quad (82)$$

The first equality plugs in the definition of  $\mathbf{x}_j^\perp$  and  $\mathbf{y}$  in Eq. (22) and Eq. (23). The second equality plugs in  $\frac{\langle \mathbf{b}, \mathbf{a} \rangle}{\|\mathbf{b}\|_2} = \frac{\langle \mathbf{x}_j, \mathbf{a} \rangle}{\|\mathbf{x}_j\|_2}$ , makes use of Eq. (17), and utilizes Eq. (67). The last equality further makes use of  $\frac{\langle \mathbf{b}, \mathbf{a} \rangle}{\|\mathbf{b}\|_2} = \frac{\langle \mathbf{x}_j, \mathbf{a} \rangle}{\|\mathbf{x}_j\|_2}$ . The established equality says that  $u_j^-(\lambda_2)$  is continuous at the  $\lambda_2$  that satisfies  $\frac{\langle \mathbf{b}, \mathbf{a} \rangle}{\|\mathbf{b}\|_2} = \frac{\langle \mathbf{x}_j, \mathbf{a} \rangle}{\|\mathbf{x}_j\|_2}$ .

It follows from the definition of  $\lambda_{2,a}$  that if  $\lambda_2 \in (\lambda_{2,a}, \lambda_1)$  then  $\frac{\langle \mathbf{b}, \mathbf{a} \rangle}{\|\mathbf{b}\|_2} > \frac{\langle \mathbf{x}_j, \mathbf{a} \rangle}{\|\mathbf{x}_j\|_2}$ . Therefore, according to Eq. (80),  $u_j^-(\lambda_2)$  is monotonically decreasing with  $\lambda_2$  in  $(\lambda_{2,a}, \lambda_1)$ . Next, we focus on  $\lambda_2$  in the interval  $(0, \lambda_{2,a}]$ .

Denote  $\gamma = \frac{1}{\lambda_2} - \frac{1}{\lambda_1}$ , and write  $\mathbf{b} = \frac{\mathbf{y}}{\lambda_2} - \boldsymbol{\theta}_1^* = \mathbf{a} + \gamma\mathbf{y}$ . Thus,  $u_j^-(\lambda_2) = -\langle \mathbf{x}_j, \boldsymbol{\theta}_1^* \rangle + \frac{1}{2} [\|\mathbf{x}_j\|_2 \|\mathbf{b}\|_2 - \langle \mathbf{x}_j, \mathbf{b} \rangle]$  can be rewritten as

$$w(\gamma) = \frac{1}{2} [\|\mathbf{x}_j\|_2 \|\mathbf{a} + \gamma\mathbf{y}\|_2 - \langle \mathbf{x}_j, \mathbf{a} + \gamma\mathbf{y} \rangle] \quad (83)$$

The first and second derivatives of  $w(\gamma)$  with regard to  $\gamma$  can be computed as: we have

$$w'(\gamma) = \frac{1}{2} \left[ \frac{\|\mathbf{x}_j\|_2 \langle \mathbf{a} + \gamma\mathbf{y}, \mathbf{y} \rangle}{\|\mathbf{a} + \gamma\mathbf{y}\|_2} - \langle \mathbf{x}_j, \mathbf{y} \rangle \right] \quad (84)$$

$$w''(\gamma) = \frac{\|\mathbf{x}_j\|_2 (\|\mathbf{y}\|_2^2 \|\mathbf{a}\|_2^2 - \langle \mathbf{a}, \mathbf{y} \rangle^2)}{2\|\mathbf{a} + \gamma\mathbf{y}\|_2^3} \geq 0 \quad (85)$$

Therefore, we have

- If  $\frac{\langle \mathbf{a}, \mathbf{y} \rangle}{\|\mathbf{a}\|_2} \geq \frac{\langle \mathbf{x}_j, \mathbf{y} \rangle}{\|\mathbf{x}_j\|_2}$ , i.e., when the angle between  $\mathbf{y}$  and  $\mathbf{a}$  is no larger than the angle between  $\mathbf{y}$  and  $\mathbf{x}_j$ , then  $w'(\gamma) \geq 0$ , and  $u_j^-(\lambda_2)$  is monotonically decreasing with regard to  $\lambda_2$  in  $(0, \lambda_{2,a}]$ . In this case, the  $\lambda_{2,a}$  and  $\lambda_{2,y}$  satisfies  $\lambda_{2,a} \leq \lambda_{2,y}$ .
- If  $\frac{\langle \mathbf{a}, \mathbf{y} \rangle}{\|\mathbf{a}\|_2} < \frac{\langle \mathbf{x}_j, \mathbf{y} \rangle}{\|\mathbf{x}_j\|_2}$ , let  $\gamma_y = \frac{1}{\lambda_{2,y}} - \frac{1}{\lambda_1}$ . Then, 1)  $h'(\gamma_y) = 0$ , 2)  $h'(\gamma_y) < 0, \forall 0 < \gamma < \gamma_y$ , and  $h'(\gamma_y) > 0, \forall \gamma > \gamma_y$ . Therefore,  $u_j^-(\lambda_2)$  is monotonically decreasing with regard to  $\lambda_2$  in  $(0, \lambda_{2,y})$ , and monotonically increasing with regard to  $\lambda_2$  in  $(\lambda_{2,y}, \lambda_{2,a}]$ . In this case, the  $\lambda_{2,a}$  and  $\lambda_{2,y}$  satisfies  $\lambda_{2,a} > \lambda_{2,y}$ .

This ends the proof of this theorem. □