Supplementary Material: Sparse Learning for Large-scale and High-dimensional Data

Lijun Zhang[1], Tianbao Yang[2], Rong Jin[3], and Zhi-Hua Zhou[1]

[1] National Key Laboratory for Novel Software Technology, Nanjing University, Nanjing 210023, China
{zhanglj, zhouzh}@lamda.nju.edu.cn
[2] Department of Computer Science, The University of Iowa, Iowa City 52242, USA
tianbao-yang@uiowa.edu
[3] Alibaba Group, Seattle, USA
jinrong.jr@alibaba-inc.com

A Proofs of Lemmas for Supporting Theorem 1

A.1 Proof of Lemma 4

Let Ω_w include the subset of non-zeros entries in $\mathbf{w}_*$ and $\bar{\Omega}_w = [d] \setminus \Omega_w$. Define

$$\begin{aligned} \mathcal{G}(\mathbf{w}) &= g(\mathbf{w}) + \max_{\boldsymbol{\lambda} \in \Delta} -h(\boldsymbol{\lambda}) - \mathbf{w}^\top A \boldsymbol{\lambda}, \\ \widehat{\mathcal{G}}(\mathbf{w}) &= g(\mathbf{w}) + \gamma_w \|\mathbf{w}\|_1 + \max_{\boldsymbol{\lambda} \in \Delta} -h(\boldsymbol{\lambda}) - \mathbf{w}^\top \widehat{A} R^\top \boldsymbol{\lambda} - \gamma_\lambda \|\boldsymbol{\lambda}\|_1. \end{aligned}$$

Let $\mathbf{v} \in \partial \|\mathbf{w}_*\|_1$ be any subgradient of $\|\cdot\|_1$ at $\mathbf{w}_*$. Then, we have

$$\mathbf{u} = \nabla g(\mathbf{w}_*) - ARR^\top \widetilde{\boldsymbol{\lambda}} + \gamma_w \mathbf{v} \in \partial \widehat{\mathcal{G}}(\mathbf{w}_*). \text{ }^4$$

Using the fact that $\widehat{\mathbf{w}}$ minimizes $\widehat{\mathcal{G}}(\cdot)$ over the domain Ω and $g(\cdot)$ is α-strongly convex, we have

$$\begin{aligned} 0 &\geq \widehat{\mathcal{G}}(\widehat{\mathbf{w}}) - \widehat{\mathcal{G}}(\mathbf{w}_*) \geq \langle \widehat{\mathbf{w}} - \mathbf{w}_*, \mathbf{u} \rangle + \frac{\alpha}{2} \|\widehat{\mathbf{w}} - \mathbf{w}_*\|_2^2 \\ &= \left\langle \widehat{\mathbf{w}} - \mathbf{w}_*, \nabla g(\mathbf{w}_*) - ARR^\top \widetilde{\boldsymbol{\lambda}} + \gamma_w \mathbf{v} \right\rangle + \frac{\alpha}{2} \|\widehat{\mathbf{w}} - \mathbf{w}_*\|_2^2. \end{aligned} \tag{21}$$

By setting $v_i = \operatorname{sign}(\widehat{w}_i)$, $\forall i \in \bar{\Omega}_w$, we have $\langle \widehat{\mathbf{w}}_{\bar{\Omega}_w}, \mathbf{v}_{\bar{\Omega}_w} \rangle = \|\widehat{\mathbf{w}}_{\bar{\Omega}_w}\|_1$. As a result,

$$\langle \widehat{\mathbf{w}} - \mathbf{w}_*, \mathbf{v} \rangle = \langle \widehat{\mathbf{w}}_{\bar{\Omega}_w}, \mathbf{v}_{\bar{\Omega}_w} \rangle + \langle \widehat{\mathbf{w}}_{\Omega_w} - \mathbf{w}_*, \mathbf{v}_{\Omega_w} \rangle \geq \|\widehat{\mathbf{w}}_{\bar{\Omega}_w}\|_1 - \|\widehat{\mathbf{w}}_{\Omega_w} - \mathbf{w}_*\|_1. \tag{22}$$

Combining (21) with (22), we have

$$\left\langle \widehat{\mathbf{w}} - \mathbf{w}_*, \nabla g(\mathbf{w}_*) - ARR^\top \widetilde{\boldsymbol{\lambda}} \right\rangle + \frac{\alpha}{2} \|\widehat{\mathbf{w}} - \mathbf{w}_*\|_2^2 + \gamma_w \|\widehat{\mathbf{w}}_{\bar{\Omega}_w}\|_1 \leq \gamma_w \|\widehat{\mathbf{w}}_{\Omega_w} - \mathbf{w}_*\|_1. \tag{23}$$

[4] In the case that $g(\cdot)$ is non-smooth, $\nabla g(\ _*)$ refers to a subgradient of $g(\cdot)$ at $\ _*$. In particular, we choose the subgradient that satisfies (24).

From the fact that $\mathbf{w}_*$ minimizes $\mathcal{G}(\cdot)$ over the domain Ω, we have

$$\langle \nabla \mathcal{G}(\mathbf{w}_*), \mathbf{w} - \mathbf{w}_* \rangle = \langle \nabla g(\mathbf{w}_*) - A\boldsymbol{\lambda}_*, \mathbf{w} - \mathbf{w}_* \rangle \geq 0, \ \forall \mathbf{w} \in \Omega. \tag{24}$$

Then,

$$\begin{aligned}
&\left\langle \widehat{\mathbf{w}} - \mathbf{w}_*, \nabla g(\mathbf{w}_*) - ARR^\top \widetilde{\boldsymbol{\lambda}} \right\rangle \\
= &\langle \widehat{\mathbf{w}} - \mathbf{w}_*, \nabla g(\mathbf{w}_*) - A\boldsymbol{\lambda}_* \rangle + \left\langle \widehat{\mathbf{w}} - \mathbf{w}_*, A(I - RR^\top)\boldsymbol{\lambda}_* \right\rangle + \left\langle \widehat{\mathbf{w}} - \mathbf{w}_*, ARR^\top(\boldsymbol{\lambda}_* - \widetilde{\boldsymbol{\lambda}}) \right\rangle \\
\overset{(24)}{\geq} & - \|\widehat{\mathbf{w}} - \mathbf{w}_*\|_1 \left(\left\| A(I - RR^\top)\boldsymbol{\lambda}_* \right\|_\infty + \left\| ARR^\top(\boldsymbol{\lambda}_* - \widetilde{\boldsymbol{\lambda}}) \right\|_\infty \right) \\
\overset{(14)}{=} & - \rho_w \|\widehat{\mathbf{w}} - \mathbf{w}_*\|_1 = -\rho_w \left(\|\widehat{\mathbf{w}}_{\bar{\Omega}_w}\|_1 + \|\widehat{\mathbf{w}}_{\Omega_w} - \mathbf{w}_*\|_1 \right).
\end{aligned} \tag{25}$$

From (23) and (25), we have

$$\frac{\alpha}{2}\|\widehat{\mathbf{w}} - \mathbf{w}_*\|_2^2 + (\gamma_w - \rho_w)\|\widehat{\mathbf{w}}_{\bar{\Omega}_w}\|_1 \leq (\gamma_w + \rho_w)\|\widehat{\mathbf{w}}_{\Omega_w} - \mathbf{w}_*\|_1.$$

Since $\gamma_w \geq 2\rho_w$, we have

$$\frac{\alpha}{2}\|\widehat{\mathbf{w}} - \mathbf{w}_*\|_2^2 + \frac{\gamma_w}{2}\|\widehat{\mathbf{w}}_{\bar{\Omega}_w}\|_1 \leq \frac{3\gamma_w}{2}\|\widehat{\mathbf{w}}_{\Omega_w} - \mathbf{w}_*\|_1.$$

And thus,

$$\begin{aligned}
\frac{\alpha}{2}\|\widehat{\mathbf{w}} - \mathbf{w}_*\|_2^2 \leq \frac{3\gamma_w}{2}\|\widehat{\mathbf{w}}_{\Omega_w} - \mathbf{w}_*\|_1 \leq \frac{3\gamma_w\sqrt{s_w}}{2}\|\widehat{\mathbf{w}}_{\Omega_w} - \mathbf{w}_*\|_2 \Rightarrow \|\widehat{\mathbf{w}} - \mathbf{w}_*\|_2 \leq \frac{3\gamma_w\sqrt{s_w}}{\alpha} \\
\frac{\alpha}{2s_w}\|\widehat{\mathbf{w}}_{\Omega_w} - \mathbf{w}_*\|_1^2 \leq \frac{\alpha}{2}\|\widehat{\mathbf{w}} - \mathbf{w}_*\|_2^2 \leq \frac{3\gamma_w}{2}\|\widehat{\mathbf{w}}_{\Omega_w} - \mathbf{w}_*\|_1 \Rightarrow \|\widehat{\mathbf{w}}_{\Omega_w} - \mathbf{w}_*\|_1 \leq \frac{3\gamma_w s_w}{\alpha} \\
\frac{\gamma_w}{2}\|\widehat{\mathbf{w}}_{\bar{\Omega}_w}\|_1 \leq \frac{3\gamma_w}{2}\|\widehat{\mathbf{w}}_{\Omega_w} - \mathbf{w}_*\|_1 \Rightarrow \|\widehat{\mathbf{w}}_{\bar{\Omega}_w}\|_1 \leq 3\|\widehat{\mathbf{w}}_{\Omega_w} - \mathbf{w}_*\|_1 \Rightarrow \|\widehat{\mathbf{w}} - \mathbf{w}_*\|_1 \leq \frac{12\gamma_w s_w}{\alpha} \\
\frac{\|\widehat{\mathbf{w}} - \mathbf{w}_*\|_1}{\|\widehat{\mathbf{w}} - \mathbf{w}_*\|_2} = \frac{\|\widehat{\mathbf{w}}_{\Omega_w} - \mathbf{w}_*\|_1 + \|\widehat{\mathbf{w}}_{\bar{\Omega}_w}\|_1}{\|\widehat{\mathbf{w}} - \mathbf{w}_*\|_2} \leq \frac{4\|\widehat{\mathbf{w}}_{\Omega_w} - \mathbf{w}_*\|_1}{\|\widehat{\mathbf{w}} - \mathbf{w}_*\|_2} \leq \frac{4\sqrt{s_w}\|\widehat{\mathbf{w}}_{\Omega_w} - \mathbf{w}_*\|_2}{\|\widehat{\mathbf{w}} - \mathbf{w}_*\|_2} \leq 4\sqrt{s_w}.
\end{aligned}$$

A.2 Proof of Lemma 6

First, we assume $\|\mathbf{u}\|_2 = \|$

Thus, with a probability at least $1-\delta$, we have

$$\left|\mathbf{u}^\top RR^\top \mathbf{v} - \mathbf{u}^\top \mathbf{v}\right| \leq \sqrt{\frac{c}{m}\log\frac{4}{\delta}}$$

provided (10) holds.

We complete the proof by noticing

$$\left|\mathbf{u}^\top RR^\top \mathbf{v} - \mathbf{u}^\top \mathbf{v}\right| = \|\mathbf{u}\|_2\|\mathbf{v}\|_2 \left|\frac{\mathbf{u}^\top}{\|\mathbf{u}\|_2} RR^\top \frac{\mathbf{v}}{\|\mathbf{v}\|_2} - \frac{\mathbf{u}^\top \mathbf{v}}{\|\mathbf{u}\|_2\|\mathbf{v}\|_2}\right|.$$

A.3 Proof of Lemma 7

First, we define

$$\mathcal{S}_{n,16s_\lambda} = \left\{\mathbf{x} \in \mathbb{R}^n : \|\mathbf{x}\|_2 \leq 1, \|\mathbf{x}\|_0 \leq 16s_\lambda\right\}.$$

Using Lemma 3.1 from [25], we have $\mathcal{K}_{n,16s_\lambda} \subset 2\,\mathrm{conv}(\mathcal{S}_{n,16s_\lambda})$ and therefore

$$U_4 \leq 2 \sup_{\mathbf{z}\in \mathrm{conv}(\mathcal{S}_{n,16s_\lambda})} \left\|A(RR^\top - I)\mathbf{z}\right\|_\infty = 2 \underbrace{\sup_{\mathbf{z}\in \mathcal{S}_{n,16s_\lambda}} \left\|A(RR^\top - I)\mathbf{z}\right\|_\infty}_{:=\theta} \tag{28}$$

where the last equality follows from the fact that the maximum of a convex function over a convex set generally occurs at some extreme point of the set [27].

Let $\mathcal{S}_{n,s}(\epsilon)$ be a proper ϵ-net for $\mathcal{S}_{n,s}$ with the smallest cardinality, and $|\mathcal{S}_{n,s}(\epsilon)|$ be the covering number for $\mathcal{S}_{n,s}$. We have the following lemma for bounding $|\mathcal{S}_{n,s}(\epsilon)|$.

Lemma 8 **[25, Lemma 3.3] For $\epsilon \in (0,1)$ and $s \leq n$, we have**

$$\log|\mathcal{S}_{n,s}(\epsilon)| \leq s \log\left(\frac{9n}{\epsilon s}\right).$$

Let $\mathcal{S}_{n,16s_\lambda}(\epsilon)$ be a ϵ-net of $\mathcal{S}_{n,16s_\lambda}$ with smallest cardinality. With the help of $\mathcal{S}_{n,16s_\lambda}(\epsilon)$, we define a discretized version of θ in (28) as

$$\theta(\epsilon) = \sup\left\{\left\|A(RR^\top - I)\mathbf{z}\right\|_\infty : \mathbf{z} \in \mathcal{S}_{n,16s_\lambda}(\epsilon)\right\}.$$

The following lemma relates θ with $\theta(\epsilon)$.

Lemma 9 **[17, Lemma 9.2] For $\epsilon \in (0, 1/\sqrt{2})$, we have**

$$\theta \leq \frac{\theta(\epsilon)}{1-\sqrt{2}\epsilon}.$$

By choosing $\epsilon = 1/2$, we have $\theta \leq (2+\sqrt{2})\theta(1/2)$. Combining with (28), we obtain

$$U_4 \leq 2(2+\sqrt{2}) \underbrace{\sup\left\{\left\|A(RR^\top - I)\mathbf{z}\right\|_\infty : \mathbf{z} \in \mathcal{S}_{n,16s_\lambda}(1/2)\right\}}_{\theta(1/2)}$$

Furthermore, Lemma 8 implies

$$\log |\mathcal{S}_{n,16s_\lambda}(1/2)| \leq 16 s_\lambda \log \left(\frac{9n}{8s_\lambda} \right).$$

We proceed by providing an upper bound for $\theta(1/2)$. Following the arguments for bounding U_1 in the proof of Lemma 5, we have with a probability at least $1-\delta$,

$$\left\| A \left(RR^\top - I \right) \mathbf{z} \right\|_\infty \leq \sqrt{\frac{c}{m} \log \frac{4d}{\delta}}$$

for each $\mathbf{z} \in \mathcal{S}_{n,16s_\lambda}(1/2)$. We complete the proof by taking the union bound over all $\mathbf{z} \in \mathcal{S}_{n,16s_\lambda}(1/2)$.

B Proof of Theorem 2

The analysis here is similar to that for Lemma 1. Recall that in the proof of Theorem 1, we have proved that

$$\gamma_\lambda \geq 2\|A^\top \mathbf{w}_*\|_2 \sqrt{\frac{c}{m} \log \frac{4n}{\delta}} \geq 2 \left\| (RR^\top - I) A^\top \mathbf{w}_* \right\|_\infty \tag{29}$$

holds with a probability at least $1-\delta$.

Define

$$\widehat{\mathcal{L}}(\boldsymbol{\lambda}) = -h(\boldsymbol{\lambda}) - \widehat{\mathbf{w}}^\top \widehat{A} R^\top \boldsymbol{\lambda} - \gamma_\lambda \|\boldsymbol{\lambda}\|_1.$$

Using the fact that $\widehat{\boldsymbol{\lambda}}$ maximizes $\widehat{\mathcal{L}}(\cdot)$ over the domain Δ and $h(\cdot)$ is β-strongly convex, we have

$$\left\langle \widehat{\boldsymbol{\lambda}} - \boldsymbol{\lambda}_*, \nabla h(\boldsymbol{\lambda}_*) + RR^\top A^\top \widehat{\mathbf{w}} \right\rangle + \frac{\beta}{2} \|\boldsymbol{\lambda}_* - \widehat{\boldsymbol{\lambda}}\|_2^2 + \gamma_\lambda \|\widehat{\boldsymbol{\lambda}}_{\bar{\Omega}_\lambda}\|_1 \leq \gamma_\lambda \|\widehat{\boldsymbol{\lambda}}_{\Omega_\lambda} - \boldsymbol{\lambda}_*\|_1. \tag{30}$$

On the other hand, we have β $\lambda_*\|$ ∀∈∀⇒∈∈∈∈∀∀∈∀⇒∀{∈∀⇒∈∀∈{∀∈∀⇒∈∈∀∈∀⇒∀∀⇒

$$\left\langle \widehat{\boldsymbol{\lambda}} - \boldsymbol{\lambda}_*, \nabla h(\boldsymbol{\lambda}_*) + RR^\top A^\top \widehat{\mathbf{w}} \right\rangle \quad {}^* \quad)A^\top$$

‖ 556018 [(A)0.[()-0.249957]TJ /R209.9626156f -0.5

$$=\langle \widehat{\boldsymbol{\lambda}} - \boldsymbol{\lambda}_*, \nabla h(\boldsymbol{\lambda}_*) + A \quad {}_* \nabla \quad {}^\top$$

From (30) and (31), we have

$$
\begin{aligned}
&\frac{\beta}{2}\|\boldsymbol{\lambda}_* - \widehat{\boldsymbol{\lambda}}\|_2^2 + \frac{\gamma_\lambda}{2}\|\widehat{\boldsymbol{\lambda}}_{\bar{\Omega}_\lambda}\|_1 \\
\leq & \frac{3\gamma_\lambda}{2}\|\widetilde{\boldsymbol{\lambda}}_{\Omega_\lambda} - \boldsymbol{\lambda}_*\|_1 + \|\widehat{\boldsymbol{\lambda}} - \boldsymbol{\lambda}_*\|_2 \left\|RR^\top A^\top(\widehat{\mathbf{w}} - \mathbf{w}_*)\right\|_2 \\
\leq & \frac{3\gamma_\lambda\sqrt{s_\lambda}}{2}\|\widetilde{\boldsymbol{\lambda}}_{\Omega_\lambda} - \boldsymbol{\lambda}_*\|_2 + \|\widehat{\boldsymbol{\lambda}} - \boldsymbol{\lambda}_*\|_2 \left\|RR^\top A^\top(\widehat{\mathbf{w}} - \mathbf{w}_*)\right\|_2 \\
\leq & \|\widehat{\boldsymbol{\lambda}} - \boldsymbol{\lambda}_*\|_2 \left(\frac{3\gamma_\lambda\sqrt{s_\lambda}}{2} + \left\|RR^\top A^\top(\widehat{\mathbf{w}} - \mathbf{w}_*)\right\|_2\right)
\end{aligned}
$$

which implies

$$
\begin{aligned}
&\|\boldsymbol{\lambda}_* - \widehat{\boldsymbol{\lambda}}\|_2 \\
\leq & \frac{2}{\beta}\left(\frac{3\gamma_\lambda\sqrt{s_\lambda}}{2} + \left\|RR^\top A^\top(\widehat{\mathbf{w}} - \mathbf{w}_*)\right\|_2\right) \\
\leq & \frac{2}{\beta}\left(\frac{3\gamma_\lambda\sqrt{s_\lambda}}{2} + \|A^\top(\widehat{\mathbf{w}} - \mathbf{w}_*)\|_2 + \left\|(RR^\top - I)A^\top(\widehat{\mathbf{w}} - \mathbf{w}_*)\right\|_2\right) \\
\leq & \frac{2}{\beta}\left(\frac{3\gamma_\lambda\sqrt{s_\lambda}}{2} + \left(1 + \|RR^\top - I\|_2\right)\|A^\top(\widehat{\mathbf{w}} - \mathbf{w}_*)\|_2\right).
\end{aligned}
$$

C Proof of Theorem 4

The proof is almost identical to that of Theorem 1. We just need to replace Lemmas 1 and 4 with the following ones.

Lemma 10 **Denote**

$$\rho_\lambda = \left\|(RR^\top - I)A^\top\mathbf{w}_*\right\|_\infty + \varsigma. \tag{32}$$

By choosing $\gamma_\lambda \geq 2\rho_\lambda$**, we have**

$$\|\widetilde{\boldsymbol{\lambda}} - \boldsymbol{\lambda}_*\|_2 \leq \frac{3\gamma_\lambda\sqrt{s_\lambda}}{\beta},\ \|\widetilde{\boldsymbol{\lambda}} - \boldsymbol{\lambda}_*\|_1 \leq \frac{12\gamma_\lambda s_\lambda}{\beta},\ \textbf{and}\ \frac{\|\widetilde{\boldsymbol{\lambda}} - \boldsymbol{\lambda}_*\|_1}{\|\widetilde{\boldsymbol{\lambda}} - \boldsymbol{\lambda}_*\|_2} \leq 4\sqrt{s_\lambda}.$$

Lemma 11 **Denote**

$$\rho_w = \left\|A\left(I - RR^\top\right)\boldsymbol{\lambda}_*\right\|_\infty + \left\|ARR^\top(\boldsymbol{\lambda}_* - \widetilde{\boldsymbol{\lambda}})\right\|_\infty + \varsigma. \tag{33}$$

By choosing $\gamma_w \geq 2\rho_w$**, we have**

$$\|\widehat{\mathbf{w}} - \mathbf{w}_*\|_2 \leq \frac{3\gamma_w\sqrt{s_w}}{\alpha},\ \|\widehat{\mathbf{w}} - \mathbf{w}_*\|_1 \leq \frac{12\gamma_w s_w}{\alpha},\ \textbf{and}\ \frac{\|\widehat{\mathbf{w}} - \mathbf{w}_*\|_1}{\|\widehat{\mathbf{w}} - \mathbf{w}_*\|_2} \leq 4\sqrt{s_w}.$$

C.1 Proof of Lemma 10

From the assumption, we have

$$\begin{aligned}
&\left\langle \widetilde{\boldsymbol{\lambda}} - \boldsymbol{\lambda}_*, \nabla h(\boldsymbol{\lambda}_*) + RR^\top A^\top \mathbf{w}_* \right\rangle \\
=&\left\langle \widetilde{\boldsymbol{\lambda}} - \boldsymbol{\lambda}_*, \nabla h(\boldsymbol{\lambda}_*) + A^\top \mathbf{w}_* \right\rangle + \left\langle \widetilde{\boldsymbol{\lambda}} - \boldsymbol{\lambda}_*, (RR^\top - I) A^\top \mathbf{w}_* \right\rangle \\
\overset{(12)}{\geq} & - \|\widetilde{\boldsymbol{\lambda}} - \boldsymbol{\lambda}_*\|_1 \left(\left\|(RR^\top - I)A^\top \mathbf{w}_*\right\|_\infty + \varsigma \right) \\
\overset{(32)}{=} & - \rho_\lambda \|\widetilde{\boldsymbol{\lambda}} - \boldsymbol{\lambda}_*\|_1 = -\rho_\lambda \left(\|\widetilde{\boldsymbol{\lambda}}_{\bar{\Omega}_\lambda}\|_1 + \|\widetilde{\boldsymbol{\lambda}}_{\Omega_\lambda} - \boldsymbol{\lambda}_*\|_1 \right).
\end{aligned}$$

Substituting the above inequality into (17), and the rest proof is identical to that of Lemma 1.

C.2 Proof of Lemma 11

Similarly, we have

$$\begin{aligned}
&\left\langle \widehat{\mathbf{w}} - \mathbf{w}_*, \nabla g(\mathbf{w}_*) - ARR^\top \widetilde{\boldsymbol{\lambda}} \right\rangle \\
=&\langle \widehat{\mathbf{w}} - \mathbf{w}_*, \nabla g(\mathbf{w}_*) - A\boldsymbol{\lambda}_* \rangle + \langle \widehat{\mathbf{w}} - \mathbf{w}_*, A(I - RR^\top)\boldsymbol{\lambda}_* \rangle + \left\langle \widehat{\mathbf{w}} - \mathbf{w}_*, ARR^\top(\boldsymbol{\lambda}_* - \widetilde{\boldsymbol{\lambda}}) \right\rangle \\
\overset{(11)}{\geq} & - \|\widehat{\mathbf{w}} - \mathbf{w}_*\|_1 \left(\left\|A(I - RR^\top)\boldsymbol{\lambda}_*\right\|_\infty + \left\|ARR^\top(\boldsymbol{\lambda}_* - \widetilde{\boldsymbol{\lambda}})\right\|_\infty + \varsigma \right) \\
\overset{(33)}{=} & - \rho_w \|\widehat{\mathbf{w}} - \mathbf{w}_*\|_1 = -\rho_w \left(\|\widehat{\mathbf{w}}_{\bar{\Omega}_w}\|_1 + \|\widehat{\mathbf{w}}_{\Omega_w} - \mathbf{w}_*\|_1 \right).
\end{aligned}$$

Substituting the above inequality into (23), and the rest proof is identical to that of Lemma 4.