For completeness, let us re-state the main iteration of hard thresholding gradient descent methods:

$$\bar{\mathbf{x}}_i = \mathbf{x}_i - \frac{\mu_i}{2}\nabla f(\mathbf{x}_i), \quad \mathbf{x}_{i+1} = \mathcal{P}_{\Sigma_k}\left(\bar{\mathbf{x}}_i\right), \tag{1}$$

where we assume that $\mu_i > 0$ is an iteration dependent step size selection and $\mathcal{P}_{\Sigma_k}(\cdot)$ is the hard-thresholding operation. By the definition of the hard thresholding operation $\mathcal{P}_{\Sigma_k}(\cdot)$, at the i-th iteration, $\mathbf{x}_{i+1}$ is a better k-sparse approximation to $\bar{\mathbf{x}}_i$ than $\mathbf{x}^\star$. This translates into:

$$\begin{aligned}\|\mathbf{x}_{i+1} - \bar{\mathbf{x}}_i\|_2^2 \leq \|\mathbf{x}^\star - \bar{\mathbf{x}}_i\|_2^2 \Rightarrow \\ \|\left(\mathbf{x}_{i+1} - \mathbf{x}^\star\right) + \left(\mathbf{x}^\star - \bar{\mathbf{x}}_i\right)\|_2^2 \leq \|\mathbf{x}^\star - \bar{\mathbf{x}}_i\|_2^2 \Rightarrow \\ \|\mathbf{x}_{i+1} - \mathbf{x}^\star\|_2^2 + \|\mathbf{x}^\star - \bar{\mathbf{x}}_i\|_2^2 + 2\langle \mathbf{x}_{i+1} - \mathbf{x}^\star,\ \mathbf{x}^\star - \bar{\mathbf{x}}_i\rangle \leq \|\mathbf{x}^\star - \bar{\mathbf{x}}_i\|_2^2 \Rightarrow \\ \|\mathbf{x}_{i+1} - \mathbf{x}^\star\|_2^2 \leq 2\langle \mathbf{x}_{i+1} - \mathbf{x}^\star,\ \bar{\mathbf{x}}_i - \mathbf{x}^\star\rangle\end{aligned} \tag{2}$$

However, we observe that:

$$\begin{aligned}\bar{\mathbf{x}}_i := \mathbf{x}_i - \frac{\mu_i}{2}\nabla f(\mathbf{x}_i) &= \mathbf{x}_i + \mu_i \mathbf{\Phi}^\top\left(\boldsymbol{y} - \mathbf{\Phi}\mathbf{x}_i\right) && \text{(By definition of } \nabla f(\mathbf{x}_i)\text{)} \\ &= \mathbf{x}_i + \mu_i \mathbf{\Phi}^\top\left(\mathbf{\Phi}\mathbf{x}^\star + \mathbf{w} - \mathbf{\Phi}\mathbf{x}_i\right) && \text{(by } \boldsymbol{y} = \mathbf{\Phi}\mathbf{x}^\star + \mathbf{w}\text{)} \\ &= \mathbf{x}_i + \mu_i \mathbf{\Phi}^\top\mathbf{\Phi}\left(\mathbf{x}^\star - \mathbf{x}_i\right) + \mu_i\mathbf{\Phi}^\top\mathbf{w}\end{aligned} \tag{3}$$

Combining (2) and (3), we obtain:

$$\|\mathbf{x}_{i+1} - \mathbf{x}^\star\|_2^2 \leq 2\langle \mathbf{x}_{i+1} - \mathbf{x}^\star,\ \mathbf{x}_i + \mu_i\mathbf{\Phi}^\top\mathbf{\Phi}\left(\mathbf{x}^\star - \mathbf{x}_i\right) + \mu_i\mathbf{\Phi}^\top\mathbf{w} - \mathbf{x}^\star\rangle \tag{4}$$

"Massaging" the right hand side of (4) further, observe that the following two applications of the linear map are present in the inequality above:

$$\langle \mathbf{\Phi}\left(\mathbf{x}_{i+1} - \mathbf{x}^\star\right), \mu_i\mathbf{\Phi}\left(\mathbf{x}_i - \mathbf{x}^\star\right)\rangle + \mu_i\langle \mathbf{\Phi}\left(\mathbf{x}_{i+1} - \mathbf{x}^\star\right), \mathbf{w}\rangle \tag{5}$$

Let $\mathcal{S}^\star := \text{supp}(\mathbf{x}^\star)$, $\mathcal{S}_{i+1} := \text{supp}(\mathbf{x}_{i+1})$ and $\mathcal{S}_i := \text{supp}(\mathbf{x}_i)$; in all cases, $|\mathcal{S}^\star| \leq k$, $|\mathcal{S}_{i+1}| \leq k$ and, $|\mathcal{S}_i| \leq k$. Thus, the above can be equivalently written as:

$$\langle \mathbf{\Phi}_{\mathcal{S}_{i+1}\cup\mathcal{S}^\star}\left(\mathbf{x}_{i+1} - \mathbf{x}^\star\right), \mu_i\mathbf{\Phi}_{\mathcal{S}_i\cup\mathcal{S}^\star}\left(\mathbf{x}_i - \mathbf{x}^\star\right)\rangle + \mu_i\langle \mathbf{\Phi}_{\mathcal{S}_{i+1}\cup\mathcal{S}^\star}\left(\mathbf{x}_{i+1} - \mathbf{x}^\star\right), \mathbf{w}\rangle$$

where $\mathbf{\Phi}_{\mathcal{S}}$ is the submatrix in $\mathbf{\Phi}$, restricted in the columns indexed in $\mathcal{S}$. Let $\mathcal{A} := \mathcal{S}^\star \cup \mathcal{S}_{i+1} \cup \mathcal{S}_i$ which satisfies $|\mathcal{A}| \leq 3k$. Then, one can easily observe that in the inequality above, we can restrict the "active" columns in $\mathbf{\Phi}$ to those indexed by $\mathcal{A}$, such that (5) is equal to:

$$\langle \mathbf{\Phi}_{\mathcal{A}}\left(\mathbf{x}_{i+1} - \mathbf{x}^\star\right), \mu_i\mathbf{\Phi}_{\mathcal{A}}\left(\mathbf{x}_i - \mathbf{x}^\star\right)\rangle + \mu_i\langle \mathbf{\Phi}_{\mathcal{A}}\left(\mathbf{x}_{i+1} - \mathbf{x}^\star\right), \mathbf{w}\rangle$$

Combining the above with (4) and applying the Cauchy-Schwarz inequality iteratively, we obtain:

$$\begin{aligned}\|\mathbf{x}_{i+1} - \mathbf{x}^\star\|_2^2 &\leq 2\langle \mathbf{x}_{i+1} - \mathbf{x}^\star, \left(\mathbf{I} - \mu_i\mathbf{\Phi}_{\mathcal{A}}^\top\mathbf{\Phi}_{\mathcal{A}}\right)\left(\mathbf{x}_i - \mathbf{x}^\star\right)\rangle + 2\mu_i\langle \mathbf{\Phi}_{\mathcal{A}}\left(\mathbf{x}_{i+1} - \mathbf{x}^\star\right), \mathbf{w}\rangle \\ &\leq 2\|\mathbf{x}_{i+1} - \mathbf{x}^\star\|_2 \cdot \|\left(\mathbf{I} - \mu_i\mathbf{\Phi}_{\mathcal{A}}^\top\mathbf{\Phi}_{\mathcal{A}}\right)\left(\mathbf{x}_i - \mathbf{x}^\star\right)\|_2 + 2\mu_i\|\mathbf{\Phi}_{\mathcal{A}}\left(\mathbf{x}_{i+1} - \mathbf{x}^\star\right)\|_2\|\mathbf{w}\|_2 \\ &\leq 2\|\mathbf{x}_{i+1} - \mathbf{x}^\star\|_2 \cdot \|\mathbf{I} - \mu_i\mathbf{\Phi}_{\mathcal{A}}^\top\mathbf{\Phi}_{\mathcal{A}}\|_2\|\mathbf{x}_i - \mathbf{x}^\star\|_2 + 2\mu_i\|\mathbf{\Phi}_{\mathcal{A}}\left(\mathbf{x}_{i+1} - \mathbf{x}^\star\right)\|_2\|\mathbf{w}\|_2\end{aligned} \tag{6}$$

and, thus,

$$\|\mathbf{x}_{i+1} - \mathbf{x}^\star\|_2 \leq 2\|\mathbf{I} - \mu_i\mathbf{\Phi}_{\mathcal{A}}^\top\mathbf{\Phi}_{\mathcal{A}}\|_2\|\mathbf{x}_i - \mathbf{x}^\star\|_2 + 2\mu_i\sqrt{\beta_{2k}}\|\mathbf{w}\|_2.$$

due to non-symmetric RIP. Observe that

$$\|\mathbf{I} - \mu_i\mathbf{\Phi}_{\mathcal{A}}^T\mathbf{\Phi}_{\mathcal{A}}\|_2 \leq \max\left\{\mu_i\lambda_{\text{max}}(\mathbf{\Phi}_{\mathcal{A}}^\top\mathbf{\Phi}_{\mathcal{A}}) - 1,\ \ 1 - \mu_i\lambda_{\text{min}}(\mathbf{\Phi}_{\mathcal{A}}^\top\mathbf{\Phi}_{\mathcal{A}})\right\}. \tag{7}$$