Supplementary Material for
" An Algorithmic Framework of Variable Metric Over-Relaxed Hybrid Proximal Extra-Gradient Method "

A. Proof of Theorem 1

Theorem. *Let $\{(x^k, y^k)\}$ be the sequence generated by the VMOR-HPE framework.*
(i) *For any given $x^* \in T^{-1}(0)$, the following approximation contractive sequence of $\|x^k - x^*\|^2_{\mathcal{M}_k}$ holds*

$$\|x^{k+1} - x^*\|^2_{\mathcal{M}_{k+1}} \leq (1+\xi_k)\|x^k - x^*\|^2_{\mathcal{M}_k} - (1-\sigma)(1+\xi_k)(1+\theta_k)\|x^k - y^k\|^2_{\mathcal{M}_k}. \tag{31}$$

(ii) *$\{x^k\}$ and $\{y^k\}$ both converge to a point x^∞ belonging to $T^{-1}(0)$.*

Proof. **(i)** Notice that $v^k \in T^{[\epsilon_k]}(y^k)$ and $x^* \in T^{-1}(0)$. By utilizing the definition of $T^{[\epsilon_k]}$, it holds that $\langle v^k, y^k - x^* \rangle \geq -\epsilon_k$. In combination with this inequality and $x^{k+1} = x^k - (1+\theta_k)c_k\mathcal{M}_k^{-1}v^k$, we obtain that

$$\begin{aligned}
\|x^{k+1} - x^*\|^2_{\mathcal{M}_k} &= \|x^k - x^*\|^2_{\mathcal{M}_k} + (1+\theta_k)^2\|c_k\mathcal{M}_k^{-1}v^k\|^2_{\mathcal{M}_k} - 2(1+\theta_k)\langle c_k v^k, x^k - x^*\rangle \\
&= \|x^k - x^*\|^2_{\mathcal{M}_k} + (1+\theta_k)^2\|c_k\mathcal{M}_k^{-1}v^k\|^2_{\mathcal{M}_k} - 2(1+\theta_k)\langle c_k v^k, x^k - y^k + y^k - x^*\rangle \\
&= \|x^k - x^*\|^2_{\mathcal{M}_k} + (1+\theta_k)^2\|c_k\mathcal{M}_k^{-1}v^k\|^2_{\mathcal{M}_k} - 2(1+\theta_k)\langle c_k v^k, x^k - y^k\rangle - 2(1+\theta_k)c_k\langle v^k, y^k - x^*\rangle \\
&\leq \|x^k - x\|^2_{\mathcal{M}_k} + (1+\theta_k)^2\|c_k\mathcal{M}_k^{-1}v^k\|^2_{\mathcal{M}_k} - 2(1+\theta_k)\langle c_k\mathcal{M}_k^{-1}v^k, \mathcal{M}_k(x^k - y^k)\rangle + 2(1+\theta_k)c_k\epsilon_k \\
&= \|x^k - x\|^2_{\mathcal{M}_k} + (1+\theta_k)\left[\theta_k\|c_k\mathcal{M}_k^{-1}v^k\|^2_{\mathcal{M}_k} + \|c_k\mathcal{M}_k^{-1}v^k + y^k - x^k\|^2_{\mathcal{M}_k} + 2c_k\epsilon_k - \|y^k - x^k\|^2_{\mathcal{M}_k}\right] \\
&\leq \|x^k - x\|^2_{\mathcal{M}_k} - (1-\sigma)(1+\theta_k)\|y^k - x^k\|^2_{\mathcal{M}_k},
\end{aligned} \tag{32}$$

where the last inequality holds according to (7b). Moreover, according to $\mathcal{M}_{k+1} \preceq (1+\xi_k)\mathcal{M}_k$, we obtain $\frac{1}{1+\xi_k}\|z^{k+1} - z^*\|^2_{\mathcal{M}_{k+1}} \leq \|z^{k+1} - z^*\|^2_{\mathcal{M}_k}$. Substituting this inequality into (32) yields the desired approximation contractive sequence

$$\|x^{k+1} - x^*\|^2_{\mathcal{M}_{k+1}} \leq (1+\xi_k)\|x^k - x^*\|^2_{\mathcal{M}_k} - (1-\sigma)(1+\xi_k)(1+\theta_k)\|x^k - y^k\|^2_{\mathcal{M}_k}.$$

(ii) By the inequality (31), $\theta_k \geq \underline{\theta} \geq -1$ and $\sigma < 1$, we obtain $\|x^{k+1} - x^*\|^2_{\mathcal{M}_{k+1}} \leq (1+\xi_k)\|x^k - x^*\|^2_{\mathcal{M}_k}$ and

$$\|x^{k+1} - x^*\|^2_{\mathcal{M}_{k+1}} \leq \prod_{i=1}^{k}(1+\xi_i)\|x^0 - x^*\|^2_{\mathcal{M}_0}. \tag{33}$$

In addition, for any $t \geq 0$, it is easy to verify that $\log(1+t) \leq t$. Hence, $\sum_{i=0}^{\infty}\xi_i < +\infty$ implies

$$\Xi := \prod_{i=0}^{\infty}(1+\xi_i) < \exp\left(\sum_{i=0}^{\infty}\xi_i\right) < +\infty.$$

Combing the above two inequalities implies $\|x^{k+1} - x^*\|^2_{\mathcal{M}_{k+1}} \leq \Xi\|x^0 - x^*\|^2_{\mathcal{M}_0}$. This inequality, in combination with $\mathcal{M}_k \succeq \underline{\omega}\mathcal{I}$, implies the boundedness of sequence $\{x^k\}$. According to (31) again, we obtain

$$\begin{aligned}
(1-\sigma)(1+\xi_k)(1+\theta_k)\|x^k - y^k\|^2_{\mathcal{M}_k} &\leq (1+\xi_k)\|x^k - x^*\|^2_{\mathcal{M}_k} - \|x^{k+1} - x^*\|^2_{\mathcal{M}_{k+1}} \\
&\leq \|x^k - x^*\|^2_{\mathcal{M}_k} - \|x^{k+1} - x^*\|^2_{\mathcal{M}_{k+1}} + \xi_k\Xi\|x^0 - x^*\|^2_{\mathcal{M}_0}.
\end{aligned}$$

Using $\theta_k \geq \underline{\theta} > -1, \sigma < 1$ and taking a summation of both sides of the above inequality, we obtain

$$\begin{aligned}(1-\sigma)(1+\underline{\theta})\sum_{i=1}^{k}\big\|x^i-y^i\big\|^2_{\mathcal{M}_i} &\leq \sum_{i=1}^{k}(1-\sigma)(1+\xi_i)(1+\theta_i)\big\|x^i-y^i\big\|^2_{\mathcal{M}_i}\\ &\leq \big\|x^1-x^*\big\|^2_{\mathcal{M}_1}-\big\|x^{k+1}-x^*\big\|^2_{\mathcal{M}_{k+1}}+\sum_{i=1}^{k}\xi_i\Xi\big\|x^0-x^*\big\|^2_{\mathcal{M}_0}\\ &\leq \big(1+\sum_{i=1}^{k}\xi_i\big)\Xi\big\|x^0-x^*\big\|^2_{\mathcal{M}_0}.\end{aligned} \tag{34}$$

Dividing the term $(1-\sigma)(1+\underline{\theta})$ on both sides of the above inequality, we obtain

$$\sum_{i=1}^{k}\big\|x^i-y^i\big\|^2_{\mathcal{M}_i} \leq \frac{(1+\sum_{i=1}^{k}\xi_i)\Xi}{(1-\sigma)(1+\underline{\theta})}\big\|x^0-x^*\big\|^2_{\mathcal{M}_0}. \tag{35}$$

According to $\sum_{i=1}^{\infty}\xi_i<\infty$, $\mathcal{M}_k \succeq \underline{\omega}\mathcal{I}$, the boundedness of $\{x^k\}$ and inequality (35), sequence $\{y^k\}$ is apparently bounded and has the same limitation points as sequence $\{x^k\}$. To show the convergences of $\{x^k\}$ and $\{y^k\}$, we further need to argue that the accumulated residuals $\sum_{i=1}^{k}\|\mathcal{M}_i^{-1}v^i\|^2_{\mathcal{M}_i}$ and the accumulated error $\sum_{i=1}^{k}\epsilon_i$ are bounded. Expanding the term $\big\|c_k\mathcal{M}_k^{-1}v^k+y^k-x^k\big\|^2_{\mathcal{M}_k}$ in (7b), we acquire $2\langle c_kv^k, x^k-y^k\rangle \geq (1+\theta_k)\big\|c_k\mathcal{M}_k^{-1}v^k\big\|^2_{\mathcal{M}_k}+(1-\sigma)\big\|y^k-x^k\big\|^2_{\mathcal{M}_k}+2c_k\epsilon_k$. In addition, by the Cauchy-Schwartz inequality, it holds that

$$2\langle c_kv^k, x^k-y^k\rangle \leq 2\big\|c_k\mathcal{M}_k^{-1}v^k\big\|_{\mathcal{M}_k}\big\|x^k-y^k\big\|_{\mathcal{M}_k} \leq \frac{1+\theta_k}{2}\big\|c_k\mathcal{M}_k^{-1}v^k\big\|^2_{\mathcal{M}_k}+\frac{2}{1+\theta_k}\big\|x^k-y^k\big\|^2_{\mathcal{M}_k}.$$

Substituting the inequality into the above inequality, we obtain

$$(1+\theta_k)\big\|c_k\mathcal{M}_k^{-1}v^k\big\|^2_{\mathcal{M}_k}+2c_k\epsilon_k-\frac{1+\theta_k}{2}\big\|c_k\mathcal{M}_k^{-1}v^k\big\|^2_{\mathcal{M}_k}-\frac{2}{1+\theta_k}\big\|x^k-y^k\big\|^2_{\mathcal{M}_k} \leq 0, \tag{36}$$

which further indicates $\frac{1+\theta_k}{2}\big\|c_k\mathcal{M}_k^{-1}v^k\big\|^2_{\mathcal{M}_k}+2c_k\epsilon_k \leq \frac{2}{1+\theta_k}\big\|x^k-y^k\big\|^2_{\mathcal{M}_k}$. Hence, we have

$$\big\|c_k\mathcal{M}_k^{-1}v^k\big\|^2_{\mathcal{M}_k} \leq \frac{4}{(1+\theta_k)^2}\big\|x^k-y^k\big\|^2_{\mathcal{M}_k},\ c_k\epsilon_k \leq \frac{1}{1+\theta_k}\big\|x^k-y^k\big\|^2_{\mathcal{M}_k}. \tag{37}$$

Combining (35) and (37) yields the bounds of $\sum_{i=1}^{k}(1+\theta_i)^2\big\|c_i\mathcal{M}_i^{-1}v^i\big\|^2_{\mathcal{M}_i}$ and $\sum_{i=1}^{k}(1+\theta_i)c_i\epsilon_i$, which are

$$\sum_{i=1}^{k}(1+\theta_i)^2\big\|c_i\mathcal{M}_i^{-1}v^i\big\|^2_{\mathcal{M}_i} \leq \frac{4(1+\sum_{i=1}^{k}\xi_i)\Xi}{(1-\sigma)(1+\underline{\theta})}\big\|x^0-x^*\big\|^2_{\mathcal{M}_0}, \tag{38}$$

$$\sum_{i=1}^{k}(1+\theta_i)c_i\epsilon_i \leq \frac{(1+\sum_{i=1}^{k}\xi_i)\Xi}{(1-\sigma)(1+\underline{\theta})}\big\|x^0-x^*\big\|^2_{\mathcal{M}_0}. \tag{39}$$

By $\theta_k \geq \underline{\theta}$ and $c_k \geq \underline{c} > 0$, the upper estimations for $\sum_{i=1}^{k}\big\|\mathcal{M}_i^{-1}v^i\big\|^2_{\mathcal{M}_i}$ and $\sum_{i=1}^{k}\epsilon_i$ are given below:

$$\sum_{i=1}^{k}\big\|\mathcal{M}_i^{-1}v^i\big\|^2_{\mathcal{M}_i} \leq \frac{4(1+\sum_{i=1}^{k}\xi_i)\Xi}{(1-\sigma)\underline{c}^2(1+\underline{\theta})^3}\big\|x^0-x^*\big\|^2_{\mathcal{M}_0},\ \sum_{i=1}^{k}\epsilon_i \leq \frac{(1+\sum_{i=1}^{k}\xi_i)\Xi}{\underline{c}(1-\sigma)(1+\underline{\theta})^2}\big\|x^0-x^*\big\|^2_{\mathcal{M}_0}. \tag{40}$$

By (35), (40) and $\mathcal{M}_k \succeq \underline{\omega}\mathcal{I}$, it holds that $\lim_{k\to\infty}\epsilon_k = \lim_{k\to\infty}\|v^k\| = \lim_{k\to\infty}\|x^k-y^k\| = 0$. In addition, due to the boundedness of $\{x^k\}$ and $\{y^k\}$, there exists a subsequence $\mathcal{K} \subseteq \{1,2,\ldots\}$ such that $\lim_{k\in\mathcal{K},k\to\infty}x^k = \lim_{k\in\mathcal{K},k\to\infty}y^k = x^\infty$. Let $k\in\mathcal{K}$ tend to be infinity in $v^k \in T^{[\epsilon_k]}(y^k)$ in (7a), and then it holds that $0 \in T(x^\infty)$ by verifying the definition of enlargement operator $T^{[\epsilon_k]}$. Hence, x^∞ is a root of inclusion problem (1). Replacing x^* by x^∞ in inequality (31), we derive

$$\big\|x^{k+1}-x^\infty\big\|^2_{\mathcal{M}_{k+1}} \leq (1+\xi_k)\big\|x^k-x^\infty\big\|^2_{\mathcal{M}_k}-(1+\xi_k)(1-\sigma)(1+\theta_k)\big\|x^k-y^k\big\|^2_{\mathcal{M}_k}.$$

Notice that $\lim_{k\in\mathcal{K},k\to\infty} x^k = x^\infty$. Therefore, for any given $\epsilon > 0$, there exists $\overline{k} \in \mathcal{K} > 0$ such that $\|x^{\overline{k}} - x^\infty\|^2_{\mathcal{M}_{\overline{k}}} \leq \frac{\epsilon}{\Xi}$. Then, for all $k \geq \overline{k}$, the above inequality indicates

$$\|x^{k+1} - x^\infty\|^2_{\mathcal{M}_{k+1}} \leq \prod_{i=\overline{k}}^{k}(1+\xi_i)\|x^{\overline{k}} - x^\infty\|^2_{\mathcal{M}_{\overline{k}}} \leq \prod_{i=0}^{k}(1+\xi_i)\frac{\epsilon}{\Xi} \leq \epsilon.$$

Hence, it holds that $\lim_{k\to\infty} x^k = \lim_{k\to\infty} y^k = x^\infty$ by $\mathcal{M}_k \succeq \underline{w}\mathcal{I}$. We complete the proof. □

B. Proof of Theorem 2

Theorem. *Let $\{(x^k, y^k)\}$ be the sequence generated by the VMOR-HPE framework. Assume that the metric subregularity of T at $(x^\infty, 0) \in \operatorname{gph} T$ holds with $\kappa > 0$. Then, there exists $\overline{k} > 0$ such that for all $k \geq \overline{k}$,*

$$\operatorname{dist}^2_{\mathcal{M}_{k+1}}\big(x^{k+1}, T^{-1}(0)\big) \leq \Big(1 - \frac{\varrho_k}{2}\Big)\operatorname{dist}^2_{\mathcal{M}_k}\big(x^k, T^{-1}(0)\big), \tag{41}$$

where $\varrho_k = [(1-\sigma)(1+\theta_k)] \Big/ \left[\Big(1 + \frac{\kappa}{\underline{c}}\sqrt{\frac{\Xi\overline{\omega}}{\underline{\omega}}}\Big)^2\Big(1 + \sqrt{\sigma + \frac{4\max\{-\theta_k,0\}}{(1+\theta_k)^2}}\Big)^2\right] \in (0,1)$.

Proof. Let x^∞ be the limitation point of $\{x^k\}$ and z^k be the point satisfying $0 \in c_k T(z^k) + \mathcal{M}_k(z^k - x^k)$, respectively. By the metric subregularity of T at $(x^\infty, 0) \in \operatorname{gph} T$, there exists $\widetilde{k} \in \mathbb{N}$ such that for all $k \geq \widetilde{k}$,

$$\begin{aligned}\operatorname{dist}_{\mathcal{M}_k}\big(z^k, T^{-1}(0)\big) &\leq \sqrt{\Xi\overline{\omega}}\operatorname{dist}\big(z^k, T^{-1}(0)\big) \leq \sqrt{\Xi\overline{\omega}}\kappa\operatorname{dist}\big(0, T(z^k)\big)\\ &\leq \frac{\sqrt{\Xi\overline{\omega}}\kappa}{\underline{c}}\|\mathcal{M}_k(z^k - x^k)\| \leq \frac{\kappa}{\underline{c}}\sqrt{\frac{\Xi\overline{\omega}}{\underline{\omega}}}\|z^k - x^k\|_{\mathcal{M}_k},\end{aligned} \tag{42}$$

where the third inequality holds due to $-c_k^{-1}\mathcal{M}_k(z^k - x^k) \in T(z^k)$ and $c_k \geq \underline{c}$, and the last inequality holds due to $\big\|\mathcal{M}_k^{\frac{1}{2}}(z^k - x^k)\big\| \geq \lambda_{\min}(\mathcal{M}_k^{\frac{1}{2}})\big\|z^k - x^k\big\|$. By the triangle inequality, inequality (42) indicates

$$\operatorname{dist}_{\mathcal{M}_k}\big(x^k, T^{-1}(0)\big) \leq \big\|x^k - z^k\big\|_{\mathcal{M}_k} + \operatorname{dist}_{\mathcal{M}_k}\big(z^k, T^{-1}(0)\big) \leq \Big(1 + \frac{\kappa}{\underline{c}}\sqrt{\frac{\Xi\overline{\omega}}{\underline{\omega}}}\Big)\big\|z^k - x^k\big\|_{\mathcal{M}_k}. \tag{43}$$

Next, we build the connection between $\|z^k - x^k\|_{\mathcal{M}_k}$ and $\|y^k - x^k\|_{\mathcal{M}_k}$, which is crucial for establishing the linear convergence rate (41). Due to inequality (7a), $0 \in c_k T(z^k) + \mathcal{M}_k(z^k - x^k)$ and the definition of $T^{[\epsilon_k]}$, we obtain $\big\langle c_k v^k - \mathcal{M}_k(x^k - z^k), y^k - z^k\big\rangle \geq -c_k\epsilon_k$. Let $r^k := c_k\mathcal{M}_k^{-1}v^k + y^k - x^k$, and then it holds that $c_k v^k = \mathcal{M}_k r^k + \mathcal{M}_k(x^k - y^k)$. Substituting this equality into the last inequality yields

$$\|z^k - y^k\|^2_{\mathcal{M}_k} - \|r^k\|_{\mathcal{M}_k}\|z^k - y^k\|_{\mathcal{M}_k} - c_k\epsilon_k \leq 0.$$

The above quadratic inequality on the term $\big\|z^k - y^k\big\|_{\mathcal{M}_k}$ directly implies the following result that

$$\big\|z^k - y^k\big\|_{\mathcal{M}_k} \leq \frac{1}{2}\Big[\big\|r^k\big\|_{\mathcal{M}_k} + \sqrt{\big\|r^k\big\|^2_{\mathcal{M}_k} + 4c_k\epsilon_k}\Big] \leq \sqrt{\big\|r^k\big\|^2_{\mathcal{M}_k} + 2c_k\epsilon_k}. \tag{44}$$

Moreover, arranging the terms in (7b), and then using notations r^k and inequality (37), we have

$$\big\|r^k\big\|^2_{\mathcal{M}_k} + 2c_k\epsilon_k \leq \sigma\big\|x^k - y^k\big\|^2_{\mathcal{M}_k} - \theta_k\big\|c_k\mathcal{M}_k^{-1}v^k\big\|^2_{\mathcal{M}_k} \leq \big(\sigma + \max\{-\theta_k, 0\}/(1+\theta_k)^2\big)\big\|x^k - y^k\big\|^2_{\mathcal{M}_k}.$$

Substituting this inequality into (44) and using the triangle inequality, we further obtain

$$\big\|z^k - x^k\big\|_{\mathcal{M}_k} \leq \big\|z^k - y^k\big\|_{\mathcal{M}_k} + \big\|y^k - x^k\big\|_{\mathcal{M}_k} \leq \Big(1 + \sqrt{\sigma + \frac{4\max\{-\theta_k, 0\}}{(1+\theta_k)^2}}\Big)\big\|x^k - y^k\big\|_{\mathcal{M}_k}.$$

Substituting this inequality into inequality (43), for all $k \geq \widetilde{k}$ it holds that

$$\begin{aligned} \mathrm{dist}_{\mathcal{M}_k}\big(x^k, T^{-1}(0)\big) &\leq \Big(1 + \frac{\kappa}{\underline{c}}\sqrt{\frac{\Xi\overline{\omega}}{\underline{\omega}}}\Big)\Big(1 + \sqrt{\sigma + \frac{4\max\{-\theta_k, 0\}}{(1+\theta_k)^2}}\Big)\|x^k - y^k\|_{\mathcal{M}_k} \\ &\leq \Big(1 + \frac{\kappa}{\underline{c}}\sqrt{\frac{\Xi\overline{\omega}}{\underline{\omega}}}\Big)\Big(1 + \sqrt{\sigma + \frac{4\max\{-\theta_k, 0\}}{(1+\theta_k)^2}}\Big)\|x^k - y^k\|_{\mathcal{M}_k}. \end{aligned} \tag{45}$$

According to (31) in Theorem 1, for all $k \in \mathbb{N}$, it holds that

$$\begin{aligned} \mathrm{dist}^2_{\mathcal{M}_{k+1}}\big(x^{k+1}, T^{-1}(0)\big) &= \big\|x^{k+1} - \Pi_{T^{-1}(0)}(x^{k+1})\big\|^2_{\mathcal{M}_{k+1}} \leq \big\|x^{k+1} - \Pi_{T^{-1}(0)}(x^k)\big\|^2_{\mathcal{M}_{k+1}} && (46) \\ &\leq (1+\xi_k)\big\|x^k - \Pi_{T^{-1}(0)}(x^k)\big\|^2_{\mathcal{M}_k} - (1+\xi_k)(1-\sigma)(1+\theta_k)\big\|x^k - y^k\big\|^2_{\mathcal{M}_k} \\ &= (1+\xi_k)\mathrm{dist}^2_{\mathcal{M}_k}\big(x^k, T^{-1}(0)\big) - (1+\xi_k)(1-\sigma)(1+\theta_k)\big\|x^k - y^k\big\|^2_{\mathcal{M}_k}, && (47) \end{aligned}$$

where $\Pi_{T^{-1}(0)}(\cdot) = \arg\inf_{x \in T^{-1}(0)} \big\| \cdot - x\big\|_{\mathcal{M}_{k+1}}$, and the first equality and the first inequality hold due to the definition of $\mathrm{dis}_{\mathcal{M}_{k+1}}(\cdot, T^{-1}(0))$. Utilizing inequalities (45) and (46), we obtain

$$\mathrm{dist}^2_{\mathcal{M}_{k+1}}\big(x^{k+1}, T^{-1}(0)\big) \leq (1+\xi_k)(1-\varrho)\mathrm{dist}^2_{\mathcal{M}_k}\big(x^k, T^{-1}(0)\big), \tag{48}$$

where $\varrho_k = [(1-\sigma)(1+\theta_k)] \Big/ \Big[\Big(1 + \frac{\kappa}{\underline{c}}\sqrt{\frac{\Xi\overline{\omega}}{\underline{\omega}}}\Big)\Big(1 + \sqrt{\sigma + \frac{4\max\{-\theta_k, 0\}}{(1+\theta_k)^2}}\Big)\Big]^2 \in (0,1)$. In addition, recall $\sum_{k=1}^{\infty} \xi_k < \infty$. Hence, there exists $\widehat{k} \in \mathbb{N}$ such that for all $k \geq \widehat{k}$, it holds that $\xi_k \leq \frac{\varrho_k}{2(1-\varrho_k)}$, which means that $(1+\xi_k)(1-\varrho_k) \leq 1 - \frac{\varrho_k}{2} < 1$. Substituting this inequality into (48) and setting $\overline{k} = \max\{\widetilde{k}, \widehat{k}\}$, we acquire the desired result (41). The proof is finished. $\square$

C. Proof of Theorem 3

Theorem. *Let $\{(x^k, y^k, v^k)\}$ and $\{\epsilon_k\}$ be the sequences generated by the VMOR-HPE framework.*
(i) *There exists an integer $k_0 \in \{1, 2, \dots, k\}$ such that $v^{k_0} \in T^{[\epsilon_{k_0}]}(y^{k_0})$ with v^{k_0} and $\epsilon_{k_0} \geq 0$ respectively satisfying*

$$\|v^{k_0}\| \leq \sqrt{\frac{4(1+\sum_{i=1}^k \xi_i)\Xi^2\overline{\omega}}{k(1-\sigma)(1+\underline{\theta})^3\underline{c}^2}}\|x^0 - x^*\|_{\mathcal{M}_0}, \text{ and } \quad \epsilon_{k_0} \leq \frac{(1+\sum_{i=1}^k \xi_i)\Xi}{k(1-\sigma)(1+\underline{\theta})^2\underline{c}}\|x^0 - x^*\|^2_{\mathcal{M}_0}. \tag{49}$$

(ii) *Let $\{\alpha_k\}$ be the nonnegative weight sequence satisfying $\sum_{i=1}^k \alpha_i > 0$. Denote $\tau_i = (1+\theta_i)c_i$, and*

$$\overline{y}^k = \frac{\sum_{i=1}^k \tau_i \alpha_i y^i}{\sum_{i=1}^k \tau_i \alpha_i}\ \overline{v}^k = \frac{\sum_{i=1}^k \tau_i \alpha_i v^i}{\sum_{i=1}^k \tau_i \alpha_i},\ \overline{\epsilon}_k = \frac{\sum_{i=1}^k \tau_i \alpha_i \big(\epsilon_i + \langle y^i - \overline{y}^k, v^i - \overline{v}^k\rangle\big)}{\sum_{i=1}^k \tau_i \alpha_i}. \tag{50}$$

Then, it holds that $\overline{v}^k \in T^{[\overline{\epsilon}_k]}(\overline{y}^k)$ with $\overline{\epsilon}_k \geq 0$. Moreover, if $\mathcal{M}_k \leq (1+\xi_k)\mathcal{M}_{k+1}$, it holds that

$$\|\overline{v}^k\| \leq \frac{\max\limits_{1\leq i\leq k}\{\alpha_{i+1}\}\sum\limits_{i=1}^k \xi_i + \sum\limits_{i=1}^k \big|\alpha_i - \alpha_{i+1}\big| + \alpha_{k+1} + \alpha_1}{\underline{c}(1+\underline{\theta})\sum_{i=1}^k \alpha_i} M, \tag{51}$$

$$\overline{\epsilon}_k = \frac{(10+\underline{\theta})\max\limits_{1\leq i\leq k}\{\alpha_i\}\big(1 + \sum\limits_{i=1}^k \xi_i\big) + (2+\underline{\theta})\sum\limits_{i=1}^k \big|\alpha_{i+1} - \alpha_i\big|}{\underline{c}(1+\underline{\theta})^2\sum_{i=1}^k \alpha_i} B, \tag{52}$$

where M and B are two constants that are respectively defined as $M = \Xi\overline{\omega}\Big[\big\|x^*\big\| + \sqrt{\frac{\Xi}{\underline{\omega}}}\big\|x^0 - x^*\big\|_{\mathcal{M}_0}\Big]$ and

$$B = \max\left\{M,\ \Xi\big\|x^*\big\|^2 + \frac{\Xi^2}{\underline{\omega}}\big\|x^0 - x^*\big\|^2_{\mathcal{M}_0},\ \frac{\Xi^2}{(1-\sigma)\underline{\omega}}\big\|x^0 - x^*\big\|^2_{\mathcal{M}_0},\ \frac{\Xi}{(1-\sigma)}\big\|x^0 - x^*\big\|^2_{\mathcal{M}_0}\right\}.$$

Proof. **(i)** By (35), there exists an integer $k_0 \in \{1, 2, \dots, k\}$ such that the following inequality holds:

$$\|x^{k_0} - y^{k_0}\|^2_{\mathcal{M}_{k_0}} \leq \frac{(1+\sum_{i=1}^k \xi_i)\Xi}{k(1-\sigma)(1+\underline{\theta})}\|x^0 - x^*\|^2_{\mathcal{M}_0}. \tag{53}$$

Combining this inequality with (37) and using $\underline{\omega}\mathcal{I} \preceq \mathcal{M}_{k+1} \preceq (1+\xi_k)\mathcal{M}_k$, $c_k \geq \underline{c}$, we obtain

$$\|v^{k_0}\| \leq \sqrt{\frac{4(1+\sum_{i=1}^k \xi_i)\Xi^2\overline{\omega}}{k(1-\sigma)(1+\underline{\theta})^3\underline{c}^2}}\|x^0 - x^*\|_{\mathcal{M}_0}, \; \epsilon_{k_0} \leq \frac{(1+\sum_{i=1}^k \xi_i)\Xi}{k(1-\sigma)(1+\underline{\theta})^2\underline{c}}\|x^0 - x^*\|^2_{\mathcal{M}_0}.$$

In addition, $v^{k_0} \in T^{[\epsilon_{k_0}]}(y^{k_0})$ holds directly due to (7a). Hence, result **(i)** has been established.

(ii) By (Monteiro & Svaiter, 2010), it holds that $\overline{v}^k \in T^{[\overline{\epsilon}_k]}(\overline{y}^k)$ and $\overline{\epsilon}^k \geq 0$. By (50), it holds that

$$\begin{aligned}
\|\overline{v}^k\| &= \frac{1}{\sum_{i=1}^k c_i\alpha_i(1+\theta_i)}\Big\|\sum_{i=1}^k c_i\alpha_i(1+\theta_i)v^i\Big\| = \frac{1}{\sum_{i=1}^k c_i\alpha_i(1+\theta_i)}\Big\|\sum_{i=1}^k \alpha_i\mathcal{M}_i(x^{i+1} - x^i)\Big\| \\
&= \frac{1}{\sum_{i=1}^k c_i\alpha_i(1+\theta_i)}\Big\|\sum_{i=1}^k\big(\alpha_{i+1}\mathcal{M}_{i+1}x^{i+1} - \alpha_i\mathcal{M}_ix^i\big) + \sum_{i=1}^k\big(\alpha_i\mathcal{M}_i - \alpha_{i+1}\mathcal{M}_{i+1}\big)x^{i+1}\Big\| \\
&\leq \frac{\big\|\sum_{i=1}^k\big(\alpha_{i+1}\mathcal{M}_{i+1}x^{i+1} - \alpha_i\mathcal{M}_ix^i\big)\big\|}{\sum_{i=1}^k c_i\alpha_i(1+\theta_i)} + \frac{\big\|\sum_{i=1}^k\big(\alpha_i\mathcal{M}_i - \alpha_{i+1}\mathcal{M}_{i+1}\big)x^{i+1}\big\|}{\sum_{i=1}^k c_i\alpha_i(1+\theta_i)} \\
&\leq \frac{\big\|\alpha_{k+1}\mathcal{M}_{k+1}x^{k+1} - \alpha_1\mathcal{M}_1x^1\big\|}{\sum_{i=1}^k c_i\alpha_i(1+\theta_i)} + \frac{\sum_{i=1}^k\big\|\alpha_i\mathcal{M}_i - \alpha_{i+1}\mathcal{M}_{i+1}\big\|\max\limits_{1\leq i\leq k}\{\|x^{i+1}\|\}}{\sum_{i=1}^k c_i\alpha_i(1+\theta_i)} \\
&\leq \frac{\alpha_{k+1}\big\|\mathcal{M}_{k+1}x^{k+1}\| + \alpha_1\big\|\mathcal{M}_1x^1\big\|}{\sum_{i=1}^k c_i\alpha_i(1+\theta_i)} + \frac{\sum_{i=1}^k\big\|\alpha_i\mathcal{M}_i - \alpha_{i+1}\mathcal{M}_{i+1}\big\|\max\limits_{1\leq i\leq k}\{\|x^{i+1}\|\}}{\sum_{i=1}^k c_i\alpha_i(1+\theta_i)} \\
&\leq \frac{\alpha_{k+1}\big\|\mathcal{M}_{k+1}\big\| + \alpha_1\big\|\mathcal{M}_1\big\| + \sum_{i=1}^k\big\|\alpha_i\mathcal{M}_i - \alpha_{i+1}\mathcal{M}_{i+1}\big\|}{\sum_{i=1}^k c_i\alpha_i(1+\theta_i)}\max_{1\leq i\leq k}\{\|x^{i+1}\|\},
\end{aligned} \tag{54}$$

where the first and the third inequalities hold by the Cauchy-Schwartz inequality. By using $\mathcal{M}_k \leq (1+\xi_k)\mathcal{M}_{k+1}$ and $\mathcal{M}_{k+1} \leq (1+\xi_k)\mathcal{M}_k$, the following inequality holds that

$$\begin{aligned}
&\sum_{i=1}^k \|\alpha_i\mathcal{M}_i - \alpha_{i+1}\mathcal{M}_{i+1}\| \\
&\leq \sum_{i=1}^k |\alpha_i - \alpha_{i+1}|\max\{\|\mathcal{M}_{i+1}\|, \|\mathcal{M}_i\|\} + \sum_{i=1}^k \xi_i\max\{\alpha_{i+1}\|\mathcal{M}_{i+1}\|, \alpha_i\|\mathcal{M}_i\|\} \\
&\leq \max_{1\leq i\leq k}\{\|\mathcal{M}_{i+1}\|\}\sum_{i=1}^k|\alpha_i - \alpha_{i+1}| + \max_{1\leq i\leq k}\{\alpha_{i+1}\|\mathcal{M}_{i+1}\|\}\sum_{i=1}^k\xi_i \\
&\leq \max_{1\leq i\leq k}\{\|\mathcal{M}_{i+1}\|\}\Big[\sum_{i=1}^k|\alpha_i - \alpha_{i+1}| + \max_{1\leq i\leq k}\{\alpha_{i+1}\}\sum_{i=1}^k\xi_i\Big].
\end{aligned}$$

Substituting this inequality into (54) and using $\|\mathcal{M}_{k+1}\| \leq \Xi\overline{\omega}$ and $c_k \geq \underline{c} > 0$, we obtain

$$\|\overline{v}^k\| \leq \frac{\max\limits_{1\leq i\leq k}\{\alpha_{i+1}\}\sum_{i=1}^k\xi_i + \sum_{i=1}^k|\alpha_i - \alpha_{i+1}| + \alpha_{k+1} + \alpha_1}{\underline{c}\sum_{i=1}^k\alpha_i(1+\theta_i)}\max_{1\leq i\leq k}\{\|x^{i+1}\|\}\Xi\overline{\omega}.$$

By inequality (33), we have $\|x^k\| \leq \|x^*\| + \sqrt{\frac{\Xi}{\underline{\omega}}}\|x^0 - x^*\|_{\mathcal{M}_0}$. By using the notation M and $\theta_k \geq \underline{\theta}$, it holds that

$$\|\overline{v}^k\| \leq \frac{\max\limits_{1\leq i\leq k}\{\alpha_{i+1}\}\sum_{i=1}^k\xi_i + \sum_{i=1}^k|\alpha_i - \alpha_{i+1}| + \alpha_{k+1} + \alpha_1}{\underline{c}(1+\underline{\theta})\sum_{i=1}^k\alpha_i}M.$$

In the following, we estimate the upper bound for $\overline{\epsilon}_k$. By the definition of $\overline{\epsilon}_k$, we obtain

$$
\begin{aligned}
\overline{\epsilon}_k &= \frac{\sum_{i=1}^k \alpha_i c_i (1+\theta_i)\big(\epsilon^i + \langle y^i - \overline{y}^k, v^i\rangle\big)}{\sum_{i=1}^k c_i\alpha_i(1+\theta_i)} = \frac{\sum_{i=1}^k \alpha_i c_i (1+\theta_i)\epsilon^i}{\sum_{i=1}^k c_i\alpha_i(1+\theta_i)} + \frac{\sum_{i=1}^k \alpha_i c_i (1+\theta_i)\langle y^i - \overline{y}^k, v^i\rangle}{\sum_{i=1}^k c_i\alpha_i(1+\theta_i)} \\
&= \frac{\sum_{i=1}^k \alpha_i (1+\theta_i) c_i\epsilon^i}{\sum_{i=1}^k c_i\alpha_i(1+\theta_i)} + \frac{\sum_{i=1}^k \alpha_i c_i (1+\theta_i)\langle x^i - \overline{y}^k, v^i\rangle}{\sum_{i=1}^k c_i\alpha_i(1+\theta_i)} + \frac{\sum_{i=1}^k \alpha_i c_i (1+\theta_i)\langle y^i - x^i, v^i\rangle}{\sum_{i=1}^k c_i\alpha_i(1+\theta_i)} \\
&\le \frac{\max\limits_{1\le i\le k}\{\alpha_i\}\sum_{i=1}^k (1+\theta_i) c_i\epsilon^i}{\sum_{i=1}^k c_i\alpha_i(1+\theta_i)} + \frac{\sum_{i=1}^k \alpha_i c_i (1+\theta_i)\langle x^i - \overline{y}^k, v^i\rangle}{\sum_{i=1}^k c_i\alpha_i(1+\theta_i)} \\
&\quad + \frac{\max\limits_{1\le i\le k}\{\alpha_i\}\sum_{i=1}^k \big((1+\theta_i)^2\|c_i\mathcal{M}_i^{-1}v^i\|^2_{\mathcal{M}_i} + \|y^i - x^i\|^2_{\mathcal{M}_i}\big)}{\sum_{i=1}^k c_i\alpha_i(1+\theta_i)} \\
&\le \frac{6\max\limits_{1\le i\le k}\{\alpha_i\}\sum_{i=1}^k \big\|y^i - x^i\big\|^2_{\mathcal{M}_i}}{\sum_{i=1}^k c_i\alpha_i(1+\theta_i)} + \frac{\sum_{i=1}^k \alpha_i\tau_i\langle x^i - \overline{y}^k, v^i\rangle}{\sum_{i=1}^k c_i\alpha_i(1+\theta_i)},
\end{aligned}
\tag{55}
$$

where the first inequality holds according to the Cauchy-Schwartz inequality and the last inequality holds according to (37). In addition, $\|x^{i+1} - \overline{y}^k\|^2_{\mathcal{M}_i} = \|x^i - \overline{y}^k\|^2_{\mathcal{M}_i} + \|\tau_i\mathcal{M}_i^{-1}v^i\|^2_{\mathcal{M}_i} - 2\langle\tau_i v^i, x^i - \overline{y}^k\rangle$ holds by using $x^{k+1} = x^k - (1+\theta_k)c_k\mathcal{M}_k^{-1}v^k = x^k - \tau_k\mathcal{M}_k^{-1}v^k$. Hence, we obtain

$$
\begin{aligned}
2\alpha_i\langle\tau_i v^i, x^i - \overline{y}^k\rangle &= \alpha_i\|\tau_i\mathcal{M}_i^{-1}v^i\|^2_{\mathcal{M}_i} + \alpha_i\|x^i - \overline{y}^k\|^2_{\mathcal{M}_i} - \alpha_i\|x^{i+1} - \overline{y}^k\|^2_{\mathcal{M}_i} \\
&\le \alpha_i\|\tau_i\mathcal{M}_i^{-1}v^i\|^2_{\mathcal{M}_i} + \alpha_i\|x^i - \overline{y}^k\|^2_{\mathcal{M}_i} - \frac{\alpha_i}{1+\xi_i}\|x^{i+1} - \overline{y}^k\|^2_{\mathcal{M}_{i+1}} \\
&\le \alpha_i\|\tau_i\mathcal{M}_i^{-1}v^i\|^2_{\mathcal{M}_i} + \alpha_i\|x^i - \overline{y}^k\|^2_{\mathcal{M}_i} - \alpha_i\|x^{i+1} - \overline{y}^k\|^2_{\mathcal{M}_{i+1}} + \alpha_i\xi_i\|x^{i+1} - \overline{y}^k\|^2_{\mathcal{M}_{i+1}} \\
&= \alpha_i\|\tau_i\mathcal{M}_i^{-1}v^i\|^2_{\mathcal{M}_i} + \alpha_i\|x^i - \overline{y}^k\|^2_{\mathcal{M}_i} - \alpha_i\|x^{i+1} - \overline{y}^k\|^2_{\mathcal{M}_{i+1}} + \alpha_i\xi_i\|x^{i+1} - \overline{y}^k\|^2_{\mathcal{M}_{i+1}},
\end{aligned}
$$

where the first and the second inequalities hold due to $\mathcal{M}_{i+1} \preceq (1+\xi_i)\mathcal{M}_i$ and $\frac{1}{1+\xi_i} \ge 1 - \xi_i$, respectively. Taking a summation on both sides of the above inequality, it holds that

$$
\begin{aligned}
&2\sum_{i=1}^k \alpha_i\langle\tau_i v^i, x^i - \overline{y}^k\rangle \qquad (56)\\
&\le \sum_{i=1}^k \alpha_i\|\tau_i\mathcal{M}_i^{-1}v^i\|^2_{\mathcal{M}_i} + \sum_{i=1}^k(\alpha_{i+1} - \alpha_i)\|x^{i+1} - \overline{y}^k\|^2_{\mathcal{M}_{i+1}} + \alpha_1\|x^1 - \overline{y}^k\|^2_{\mathcal{M}_1} + \sum_{i=1}^k \alpha_i\xi_i\|x^{i+1} - \overline{y}^k\|^2_{\mathcal{M}_{i+1}} \\
&\le 4\max_{1\le i\le k}\{\alpha_i\}\sum_{i=1}^k\big\|y^i - x^i\big\|^2_{\mathcal{M}_i} + \max_{0\le i\le k}\{\|x^{i+1} - \overline{y}^k\|^2_{\mathcal{M}_{i+1}}\}\Big[\sum_{i=1}^k|\alpha_{i+1} - \alpha_i| + \sum_{i=1}^k\alpha_i\xi_i + \alpha_1\Big] \\
&\le 4\max_{1\le i\le k}\{\alpha_i\}\sum_{i=1}^k\big\|y^i - x^i\big\|^2_{\mathcal{M}_i} + \max_{0\le i\le k}\{\|x^{i+1} - \overline{y}^k\|^2_{\mathcal{M}_{i+1}}\}\Big[\sum_{i=1}^k|\alpha_{i+1} - \alpha_i| + \max_{1\le i\le k}\{\alpha_i\}(\sum_{i=1}^k\xi_i + 1)\Big],
\end{aligned}
$$

where the last inequality holds according to (37). This inequality combined with (55) yields

$$
\overline{\epsilon}_k \le \frac{8\max\limits_{1\le i\le k}\{\alpha_i\}\sum_{i=1}^k\big\|y^i - x^i\big\|^2_{\mathcal{M}_i}}{\sum_{i=1}^k c_i\alpha_i(1+\theta_i)} + \frac{\Big[\sum_{i=1}^k|\alpha_{i+1} - \alpha_i| + \max\limits_{1\le i\le k}\{\alpha_i\}(\sum_{i=1}^k\xi_i + 1)\Big]}{2\sum_{i=1}^k c_i\alpha_i(1+\theta_i)}B_k,
\tag{57}
$$

where $B_k = \max\limits_{0\le i\le k}\{\|x^{i+1} - \overline{y}^k\|^2_{\mathcal{M}_{i+1}}\}$. Moreover, by the definition of $\overline{y}^k$, it holds that

$$
\big\|x^{i+1} - \overline{y}^k\big\|^2_{\mathcal{M}_{i+1}} \le 2\big\|x^{i+1}\big\|^2_{\mathcal{M}_{i+1}} + 2\big\|\overline{y}^k\big\|^2_{\mathcal{M}_{i+1}} \le 2\big\|x^{i+1}\big\|^2_{\mathcal{M}_{i+1}} + 2\max_{0\le j\le k}\{\big\|y^j\big\|^2_{\mathcal{M}_{i+1}}\},
$$

where the second inequality holds according to the convexity of $\|\cdot\|^2_{\mathcal{M}_{i+1}}$. Hence, we obtain

$$
B_k \le 2\Xi\overline{\omega}\max_{0\le i\le k}\big[\big\|x^{i+1}\big\|^2 + \big\|y^{i+1}\big\|^2\big] \le 2\Xi\overline{\omega}\max_{0\le i\le k}\big[2\big\|x^{i+1}\big\|^2 + \big\|x^{i+1} - y^{i+1}\big\|^2\big].
\tag{58}
$$

By (31) and (33), it holds that $\|x^i - y^i\|^2_{\mathcal{M}_i} \leq \frac{\Xi}{(1-\sigma)(1+\underline{\theta})}\|x^0 - x^*\|^2_{\mathcal{M}_0}$. Moreover, by (33), it holds that $\frac{1}{2}\|x^k\|^2 \leq \|x^*\|^2 + \frac{\Xi}{\underline{\omega}}\|x^0 - x^*\|^2_{\mathcal{M}_0}$. Substituting the two inequalities into (58) yields

$$B_k \leq 2\Xi\Big[\|x^*\|^2 + \frac{\Xi}{\underline{\omega}}\|x^0 - x^*\|^2_{\mathcal{M}_0} + \frac{\Xi}{(1-\sigma)\underline{\omega}(1+\underline{\theta})}\|x^0 - x^*\|^2_{\mathcal{M}_0}\Big]. \tag{59}$$

Combining (35),(59) with (57) and using the fact that $c_k \geq \underline{c}$ and $\theta_k \geq \underline{\theta} > -1$, we further obtain

$$\begin{aligned}
\bar{\epsilon}_k \leq\ & \frac{8 \max\limits_{0\leq i\leq k}\{\alpha_i\}}{\sum_{i=1}^k c_i\alpha_i(1+\theta_i)} \frac{(1+\sum_{i=1}^k \xi_i)\Xi}{(1-\sigma)(1+\underline{\theta})}\|x^0 - x^*\|^2_{\mathcal{M}_0} \\
& + \frac{\sum_{i=1}^k|\alpha_{i+1}-\alpha_i| + \max\limits_{1\leq i\leq k}\{\alpha_i\}(\sum_{i=1}^k\xi_i+1)}{\sum_{i=1}^k c_i\alpha_i(1+\theta_i)}\Xi\Big[\|x^*\|^2 + \frac{\Xi}{\underline{\omega}}\Big(1+\frac{1}{(1-\sigma)(1+\underline{\theta})}\Big)\|x^0-x^*\|^2_{\mathcal{M}_0}\Big] \\
\leq\ & \frac{8 \max\limits_{0\leq i\leq k}\{\alpha_i\}\big(1+\sum_{i=1}^k \xi_i\big)}{\underline{c}(1+\underline{\theta})^2\sum_{i=1}^k\alpha_i} \frac{\Xi\|x^0 - x^*\|^2_{\mathcal{M}_0}}{(1-\sigma)} \\
& + \frac{\sum_{i=1}^k|\alpha_{i+1}-\alpha_i| + \max\limits_{1\leq i\leq k}\{\alpha_i\}(\sum_{i=1}^k\xi_i+1)}{\underline{c}(1+\underline{\theta})\sum_{i=1}^k\alpha_i}\Big[\Xi\|x^*\|^2 + \frac{\Xi^2}{\underline{\omega}}\|x^0 - x^*\|^2_{\mathcal{M}_0}\Big] \\
& + \frac{\sum_{i=1}^k|\alpha_{i+1}-\alpha_i| + \max\limits_{1\leq i\leq k}\{\alpha_i\}(\sum_{i=1}^k\xi_i+1)}{\underline{c}(1+\underline{\theta})^2\sum_{i=1}^k\alpha_i}\Big[\frac{\Xi^2}{\underline{\omega}}\frac{\|x^0 - x^*\|^2_{\mathcal{M}_0}}{(1-\sigma)}\Big] \\
\leq\ & \frac{(10+\underline{\theta})\max\limits_{1\leq i\leq k}\{\alpha_i\}\big(1+\sum_{i=1}^k \xi_i\big) + (2+\underline{\theta})\sum_{i=1}^k|\alpha_{i+1}-\alpha_i|}{\underline{c}(1+\underline{\theta})^2\sum_{i=1}^k\alpha_i} B,
\end{aligned}$$

where $B = \max\Big\{\frac{\Xi}{(1-\sigma)}\|x^0-x^*\|^2_{\mathcal{M}_0}, \Xi\|x^*\|^2 + \frac{\Xi^2}{\underline{\omega}}\|x^0-x^*\|^2_{\mathcal{M}_0}, \frac{\Xi^2}{(1-\sigma)\underline{\omega}}\|x^0-x^*\|^2_{\mathcal{M}_0}, M\Big\}$. The proof is finished. □

D. Proof of Proposition 1

Recall that the over-relaxed **F**orward-**B**ackward-**H**alf **F**orward (FBHF) algorithm (Briceño-Arias & Davis, 2018) is defined as

$$\begin{cases} y^k := \mathcal{J}_{\gamma_k A}\big(x^k - \gamma_k(B_1+B_2)x^k\big), & (60a)\\ x^{k+1} := x^k + (1+\theta_k)\big(y^k - x^k + \gamma_k B_2(x^k) - \gamma_k B_2(y^k)\big). & (60b)\end{cases}$$

Proposition. *Let $\{(x^k, y^k)\}$ be the sequence generated by the over-relaxed FBHF algorithm. Denote $\epsilon_k = \|x^k - y^k\|^2/(4\beta)$ and $v^k = \gamma_k^{-1}(x^k - y^k) - B_2(x^k) + B_2(y^k)$. Then,*

$$\begin{cases} (y^k, v^k) \in \operatorname{gph} T^{[\epsilon_k]} = \operatorname{gph}(A+B_1+B_2)^{[\epsilon_k]}, & (61a)\\ \theta_k\|\gamma_k v^k\|^2 + \|\gamma_k v^k + (y^k - x^k)\|^2 + 2\gamma_k\epsilon \leq \sigma\|y^k - x^k\|^2, & (61b)\\ x^{k+1} = x^k - (1+\theta_k)\gamma_k v^k, & (61c)\end{cases}$$

where (γ_k, θ_k) satisfies $\theta_k \leq [\sigma - (\gamma_k L)^2 + \gamma_k/(2\beta))]/[1+(\gamma_k L)^2]$.

Proof. By the definition of resolvent $\mathcal{J}_{\gamma_k A}$, the updating step (60a) of y^k is formulated as follows

$$x^k - \gamma_k(B_1+B_2)(x^k) \in y^k + \gamma_k A(y^k). \tag{62}$$

By (Svaiter, 2014, Lemma 2.2), it holds that $B_1(x^k) \in B_1^{[\epsilon_k]}(y^k)$ with $\epsilon_k = \|x^k - y^k\|^2/(4\beta)$. Then,

$$\begin{aligned}
\gamma_k^{-1}(x^k - y^k) - B_2(x^k) + B_2(y^k) &\in A(y^k) + B_2(y^k) + B_1(x^k)\\
&\subseteq A(y^k) + B_2(y^k) + B_1^{[\epsilon_k]}y^k\\
&\subseteq (A+B_1+B_2)^{[\epsilon_k]}(y^k),
\end{aligned}$$

where the first inclusion holds by (62), and the last inclusion holds by using the additivity property of enlargement operator (Burachik et al., 1998). Hence, utilizing $v^k = \gamma_k^{-1}(x^k - y^k) - B_2(x^k) + B_2(y^k)$, we directly obtain (61a) and (61c) that $(y^k, v^k) \in \operatorname{gph} T^{[\epsilon_k]}$ and $x^{k+1} = x^k - (1+\theta_k)\gamma_k v^k$, respectively. Next, we argue that (61b) holds. By the monotonicity of B_2, it holds that

$$
\begin{aligned}
&\theta_k\|\gamma_k v^k\|^2 + \|\gamma_k v^k + y^k - x^k\|^2 + 2\gamma_k\epsilon_k \\
=&\ \theta_k\big\|y^k - x^k + \gamma_k B_2(x^k) - \gamma_k B_2(y^k)\big\|^2 + \big\|\gamma_k(B_2x^k - B_2y^k)\big\|^2 + 2\gamma_k\epsilon_k \\
\le&\ \theta_k\big[\|y^k - x^k\|^2 + \|\gamma_k B_2(x^k) - \gamma_k B_2(y^k)\|^2\big] + \big\|\gamma_k(B_2x^k - B_2y^k)\big\|^2 + 2\gamma_k\epsilon_k \\
\le&\ \big[\theta_k(1+\gamma_k^2L^2) + \gamma_k^2L^2 + \gamma_k/(2\beta)\big]\|x^k - y^k\|^2 \le \sigma\|x^k - y^k\|^2,
\end{aligned}
$$

where the last inequality holds according to the definition of θ_k. As a consequence, the FBHF algorithm with the iterations (60a) and (60b) is a special case of the VMOR-HPE algorithm. ☐

E. Proof of Proposition 2

Let P be a bounded linear operator and $U = (P+P^*)/2$, $S = (P-P^*)/2$. The over-relaxed **n**on self-adjoint **M**etric **F**orward-**B**ackward-**H**alf **F**orward (nMFBHF) algorithm (Briceño-Arias & Davis, 2018) is defined as

$$
\begin{cases}
y^k := \mathcal{J}_{P^{-1}A}\big(x^k - P^{-1}(B_1+B_2)(x^k)\big), & \text{(63a)}\\
x^{k+1} := x^k + (1+\theta_k)\big(y^k - x^k + U^{-1}[B_2(x^k) - B_2(y^k) - S(x^k - y^k)]\big). & \text{(63b)}
\end{cases}
$$

Proposition. *Let $\{(x^k, y^k)\}$ be the sequence generated by the over-relaxed nMFBHF algorithm. Denote $\epsilon_k = \|x^k - y^k\|^2/(4\beta)$ and $v^k = P(x^k - y^k) + B_2(y^k) - B_2(x^k)$. The step-size θ_k satisfies $\theta_k + \frac{K^2(1+\theta_k)}{\lambda_{\min}^2(U)} + \frac{1}{2\beta\lambda_{\min}(U)} \le \sigma$. Then,*

$$
\begin{cases}
(y^k, v^k) \in \operatorname{gph} T^{[\epsilon_k]} = \operatorname{gph}(A + B_1 + B_2)^{[\epsilon_k]}, & \text{(64a)}\\
\theta_k\big\|U^{-1}v^k\big\|_U^2 + \big\|U^{-1}v^k + (y^k - x^k)\big\|_U^2 + 2\epsilon \le \sigma\big\|y^k - x^k\big\|_U^2, & \text{(64b)}\\
x^{k+1} = x^k - (1+\theta_k)U^{-1}v^k. & \text{(64c)}
\end{cases}
$$

Proof. By the definition of (63a), it holds that $P(x^k - y^k) - (B_1 + B_2)(x^k) \in A(y^k)$, which indicates

$$
\begin{aligned}
P(x^k - y^k) + B_2(y^k) - B_2(x^k) &\in A(y^k) + B_1(x^k) + B_2(y^k) \\
&\subseteq A(y^k) + B_1^{[\epsilon_k]}(y^k) + B_2(y^k) \\
&\subseteq (A + B_1 + B_2)^{[\epsilon_k]}(y^k).
\end{aligned} \tag{65}
$$

By the definition of v^k, we derive (64a) that $(y^k, v^k) \in \operatorname{gph} T^{[\epsilon_k]}$. In addition, recall $U = (P+P^*)/2$ and $S = (P-P^*)/2$. It is easy to check $U^{-1}P - I = U^{-1}S$. Hence, we obtain

$$
\begin{aligned}
x^{k+1} &= x^k + (1+\theta_k)\big(y^k - x^k + U^{-1}[B_2(x^k) - B_2(y^k) - S(x^k - y^k)]\big) \\
&= x^k + (1+\theta_k)\big(y^k - x^k - U^{-1}\big(S(x^k - y^k) + B_2(y^k) - B_2(x^k)\big)\big) \\
&= x^k + (1+\theta_k)\big(y^k - x^k - U^{-1}\big(S(x^k - y^k)\big) - U^{-1}\big(B_2(y^k) - B_2(x^k)\big)\big) \\
&= x^k + (1+\theta_k)\big(y^k - x^k + (I - U^{-1}P)(x^k - y^k) - U^{-1}\big(B_2(y^k) - B_2(x^k)\big)\big) \\
&= x^k + (1+\theta_k)\big(U^{-1}\big(P(y^k - x^k)\big) - U^{-1}\big(B_2(y^k) - B_2(x^k)\big)\big) \\
&= x^k - (1+\theta_k)U^{-1}v^k,
\end{aligned}
$$

which indicates that (64c) holds. In what follows, we argue that (64b) holds. According to the above equality, it clearly holds that $U^{-1}v^k = x^k - y^k - U^{-1}[B_2(x^k) - B_2(y^k) - S(x^k - y^k)]$. Hence

$$\begin{aligned}
&\theta_k\big\|U^{-1}v^k\big\|_U^2 + \big\|U^{-1}v^k + y^k - x^k\big\|_U^2 + 2\epsilon_k \\
&= \theta_k\big\|x^k - y^k - U^{-1}[(B_2-S)(x^k) - (B_2-S)(y^k)]\big\|_U^2 + \big\|U^{-1}\big[(B_2-S)(x^k) - (B_2-S)(y^k)\big]\big\|_U^2 + 2\epsilon_k \\
&\le \theta_k\big\|x^k - y^k\big\|_U^2 + (1+\theta_k)\big\|U^{-1}\big[(B_2-S)(x^k) - (B_2-S)(y^k)\big]\big\|_U^2 + 2\epsilon_k \\
&\le \theta_k\big\|x^k - y^k\big\|_U^2 + (1+\theta_k)\lambda_{\min}^{-1}(U)\|(B_2-S)x^k - (B_2-S)y^k\|^2 + 2\epsilon_k \\
&\le \theta_k\big\|x^k - y^k\big\|_U^2 + \big[(1+\theta_k)\lambda_{\min}^{-1}(U)K^2 + 1/(2\beta)\big]\|x^k - y^k\|^2 \\
&\le \big[\theta_k + [(1+\theta_k)\lambda_{\min}^{-1}(U)K^2 + 1/(2\beta)]\lambda_{\min}^{-1}(U)\big]\big\|x^k - y^k\big\|_U^2 \\
&\le \sigma\|x^k - y^k\|_U^2,
\end{aligned}$$

where the first inequality holds by the monotonicity of $B_2 - S$, the second inequality holds by $\|U^{-1}\cdot\|_U^2 \le \lambda_{\max}(U^{-1})\|\cdot\|^2 = \lambda_{\min}^{-1}(U)\|\cdot\|^2$, the third inequality holds by the Lipschitz continuity of $B_2 - S$, the fourth inequality holds by $\|\cdot\|^2 \le \lambda_{\min}^{-1}(U)\|\cdot\|_U^2$, and the last inequality holds by $\theta_k + [K^2(1+\theta_k)]/[\lambda_{\min}^2(U)] + 1/[2\beta\lambda_{\min}(U)] \le \sigma$. Hence, (64b) holds. In conclusion, the over-relaxed non self-adjoint metric FBHF algorithm with the iterations (63a) and (63b) falls into the framework of VMOR-HPE. The proof is finished. □

F. Proof of Proposition 3

The over-relaxed **P**roximal-**P**roximal-**G**radient (PPG) algorithm (Ryu & Yin, 2017) takes the following iterations:

$$\begin{cases}
x^{k+\frac{1}{2}} := \mathrm{Prox}_{\alpha r}\big(\frac{1}{n}\sum_{i=1}^{n} z_i^k\big), & \text{(66a)} \\
x_i^{k+1} := \mathrm{Prox}_{\alpha g_i}\big(2x^{k+\frac{1}{2}} - z_i^k - \alpha\nabla f_i(x^{k+\frac{1}{2}})\big),\ i = 1,\ldots,n, & \text{(66b)} \\
z_i^{k+1} := z_i^k + (1+\theta_k)(x_i^{k+1} - x^{k+\frac{1}{2}}),\ i = 1,\ldots,n. & \text{(66c)}
\end{cases}$$

To establish Proposition 3, we need the following lemma which characterizes how to calculate the proximal mapping $\mathrm{Prox}_{\alpha\overline{r}}(\cdot)$.

Lemma 1. *Given* $\mathbf{z} \in \mathbb{X}^n$, $\mathrm{Prox}_{\alpha\overline{r}}(z) = \arg\min_{\mathbf{x}\in\mathbb{X}^n} \overline{r}(\mathbf{x}) + \frac{1}{2\alpha}\|\mathbf{x} - \mathbf{z}\|^2$ *can be calculated in parallel with* $\mathrm{Prox}_{\alpha\overline{r}}(\mathbf{z}) = \big(\mathrm{Prox}_{\alpha r}(\frac{1}{n}\sum_{i=1}^n z_i), \mathrm{Prox}_{\alpha r}(\frac{1}{n}\sum_{i=1}^n z_i), \cdots, \mathrm{Prox}_{\alpha r}(\frac{1}{n}\sum_{i=1}^n z_i)\big) \in V$.

Proof. By the definition of $\overline{r}(\mathbf{x})$, it holds that the components of $\mathrm{Prox}_{\alpha\overline{r}}(\mathbf{z})$ are equal to each other. Let $\mathbf{1} = (1, 1, \cdots, 1) \in \mathbb{X}^n$. By definitions of V and $\overline{r}(\mathbf{x})$, the following equalities hold

$$\begin{aligned}
\arg\min_{\mathbf{x}\in\mathbb{X}^n} \overline{r}(\mathbf{x}) + \frac{1}{2\alpha}\|\mathbf{x} - \mathbf{z}\|^2 &= \arg\min_{\mathbf{x}\in\mathbb{X}^n} \mathbf{1}_V(\mathbf{x}) + \frac{1}{n}\sum_{i=1}^n r(x_i) + \frac{1}{2\alpha}\|\mathbf{x} - \mathbf{z}\|^2 \\
&= \arg\min_{\mathbf{x}\in V} \frac{1}{n}\sum_{i=1}^n r(x_i) + \frac{1}{2\alpha}\|\mathbf{x} - \mathbf{z}\|^2. \qquad (67)
\end{aligned}$$

Let $\mathrm{Prox}_{\alpha r}(\frac{1}{n}\sum_{i=1}^n z_i) = \arg\min_{x\in\mathbb{X}} r(x) + \frac{1}{2\alpha}\|x\mathbf{1} - \mathbf{z}\|^2$. By the definition of V, we obtain

$$\min_{\mathbf{x}\in V} \frac{1}{n}\sum_{i=1}^n r(x_i) + \frac{1}{2\alpha}\|\mathbf{x} - \mathbf{z}\|^2 = \min_{x\in\mathbb{X}} r(x) + \frac{1}{2\alpha}\|x\mathbf{1} - \mathbf{z}\|^2,$$

and that $\mathrm{Prox}_{\alpha r}(\frac{1}{n}z\mathbf{1}^T)$ solves (67). Hence, $\mathrm{Prox}_{\alpha r}(\frac{1}{n}z\mathbf{1}^T)\mathbf{1} = \mathrm{Prox}_{\alpha\overline{r}}(z)$. The proof is completed. □

Proposition. *Let* $(x^{k+\frac{1}{2}}, x_i^k, z_i^k)$ *be the sequence generated by the over-relaxed PPG algorithm. Denote* $\mathbf{x}^k = (x_1^k, \cdots, x_n^k)$, $\mathbf{z}^k = (z_1^k, \cdots, z_n^k)$, $\mathbf{1} = (1, \cdots, 1) \in \mathbb{X}^n$, $\mathbf{y}^k = \mathbf{z}^k + \mathbf{x^{k+1}} - x^{k+\frac{1}{2}}\mathbf{1}$, $\mathbf{v}^k = x^{k+\frac{1}{2}}\mathbf{1} - \mathbf{x}^{k+1}$, *and* $\epsilon_k = L\sum_{i=1}^n \|x_i^{k+1} -$

$x^{k+\frac{1}{2}}\|/4$. *Parameters* (θ_k, α) *are constrained by* $\theta_k + L\alpha/2 \leq \sigma$. *Then, it holds that*

$$\begin{cases} (\mathbf{y}^k, \mathbf{v}^k) \in \operatorname{gph} \mathcal{S}^{[\alpha\epsilon_k]}_{\alpha,\nabla\overline{f}+\partial\overline{g},\partial\overline{r}} = \operatorname{gph} T^{[\alpha\epsilon_k]}, & (68a) \\ \theta_k\|\mathbf{v}^k\|^2 + \|\mathbf{v}^k + (\mathbf{y}^k - \mathbf{z}^k)\|^2 + 2\alpha\epsilon_k \leq \sigma\|\mathbf{y}^k - \mathbf{z}^k\|^2, & (68b) \\ \mathbf{z}^{k+1} = \mathbf{z}^k - (1+\theta_k)\mathbf{v}^k. & (68c) \end{cases}$$

Proof. By Lemma 1 and equation (66a), we derive $x^{k+\frac{1}{2}}\mathbf{1} = \operatorname{Prox}_{\alpha\overline{r}}(\mathbf{z}^k)$. Hence,

$$\alpha^{-1}\big(\mathbf{z}^k - x^{k+\frac{1}{2}}\mathbf{1}\big) \in \partial\overline{r}(x^{k+\frac{1}{2}}\mathbf{1}) \tag{69}$$

Unitizing $\overline{g}$ and $\overline{f}$, (66b) is reformulated as $\mathbf{x}^{k+1} = \operatorname{Prox}_{\alpha\overline{g}}\big(2x^{k+\frac{1}{2}}\mathbf{1} - \mathbf{z}^k - \alpha\nabla\overline{f}(x^{k+\frac{1}{2}}\mathbf{1})\big)$. Then,

$$\begin{aligned} \alpha^{-1}\big(2x^{k+\frac{1}{2}}\mathbf{1} - \mathbf{x}^{k+1} - \mathbf{z}^k\big) &\in \partial\overline{g}(\mathbf{x}^{k+1}) + \nabla\overline{f}(x^{k+\frac{1}{2}}\mathbf{1}) \\ &\subseteq \partial\overline{g}(\mathbf{x}^{k+1}) + \big[\nabla\overline{f}\big]^{[\epsilon_k]}(\mathbf{x^{k+1}}) \\ &\subseteq \big[\partial\overline{g} + \nabla\overline{f}\big]^{[\epsilon_k]}(\mathbf{x^{k+1}}), \end{aligned} \tag{70}$$

where $\epsilon_k = L\|\mathbf{x}^{k+1} - x^{k+\frac{1}{2}}\mathbf{1}\|/4 = L\sum_{i=1}^n \|x_i^{k+1} - x^{k+\frac{1}{2}}\|/4$ and the second inclusion holds by (Svaiter, 2014, Lemma 2.2). Combining (69), (70) and using simple calculations, we obtain

$$\begin{aligned} x^{k+\frac{1}{2}}\mathbf{1} - \mathbf{x^{k+1}} &\in \mathcal{S}_{\alpha,[\nabla\overline{f}+\partial\overline{g}]^{[\epsilon_k]},\partial\overline{r}}\big(\mathbf{x^{k+1}} + \alpha[\alpha^{-1}\big(\mathbf{z}^k - x^{k+\frac{1}{2}}\mathbf{1}\big)]\big) \\ &= \mathcal{S}_{\alpha,[\nabla\overline{f}+\partial\overline{g}]^{[\epsilon_k]},\partial\overline{r}}\big(\mathbf{z}^k + \mathbf{x^{k+1}} - x^{k+\frac{1}{2}}\mathbf{1}\big) \\ &\subseteq \mathcal{S}^{[\alpha\epsilon_k]}_{\alpha,[\nabla\overline{f}+\partial\overline{g}],\partial\overline{r}}\big(\mathbf{z}^k + \mathbf{x^{k+1}} - x^{k+\frac{1}{2}}\mathbf{1}\big) = \mathcal{S}^{[\alpha\epsilon_k]}_{\alpha,[\nabla\overline{f}+\partial\overline{g}],\partial\overline{r}}\big(\mathbf{y}^k\big), \end{aligned}$$

where the first inclusion holds by $\mathbf{x^{k+1}} + \alpha[\alpha^{-1}\big(2x^{k+\frac{1}{2}}\mathbf{1} - \mathbf{x}^{k+1} - \mathbf{z}^k\big)] = x^{k+\frac{1}{2}}\mathbf{1} - \alpha[\alpha^{-1}\big(\mathbf{z}^k - x^{k+\frac{1}{2}}\mathbf{1}\big)]$ and using the definition of $\mathcal{S}_{\alpha,\nabla\overline{f}+\partial\overline{g},\partial\overline{r}}$, and the last inclusion holds by (Shen, 2017). By using the notation $\mathbf{v}^k$, (68a) directly holds. In addition, (66c) can also be equivalently reformulated as $\mathbf{z}^{k+1} = \mathbf{z}^k + (1+\theta_k)(\mathbf{x}^{k+1} - x^{k+\frac{1}{2}}\mathbf{1})$, which is equivalent to $\mathbf{z}^{k+1} = \mathbf{z}^k - (1+\theta_k)\mathbf{v}^k$ by utilizing the definition of $\mathbf{v}^k$. Hence, (68c) holds. Next, using the definition of $\mathbf{v}^k$, it holds that

$$\begin{aligned} &\theta_k\|\mathbf{v}^k\|^2 + \|\mathbf{v}^k + (\mathbf{y}^k - \mathbf{z}^k)\|^2 + 2\alpha\epsilon_k \\ &= \theta_k\|x^{k+\frac{1}{2}}\mathbf{1} - \mathbf{x}^{k+1}\|^2 + \|x^{k+\frac{1}{2}}\mathbf{1} - \mathbf{x}^{k+1} + (\mathbf{z}^k + \mathbf{x^{k+1}} - x^{k+\frac{1}{2}}\mathbf{1} - \mathbf{z}^k)\|^2 + 2\alpha\epsilon_k \\ &= \big(\theta_k + L\alpha/2\big)\|x^{k+\frac{1}{2}}\mathbf{1} - \mathbf{x}^{k+1}\|^2 \\ &\leq \sigma\|\mathbf{y}^k - \mathbf{z}^k\|^2, \end{aligned}$$

where the first equality holds due to the definitions of $\mathbf{v}^k$ and $\mathbf{y}^k$, the second equality holds due to the definition of ϵ_k, and the last inequality holds due to $\theta_k + L\alpha/2 \leq \sigma$, which indicates that (68b) holds. In conclusion, the over-relaxed PPG algorithm with the iterations (66a),(66b),(66c) falls into the framework of VMOR-HPE. The proof is finished. ☐

G. Proof of Proposition 4

The **A**symmetric **F**orward **B**ackward **A**djoint **S**plitting (AFBAS) algorithm (Latafat & Patrinos, 2017) is defined as:

$$\begin{cases} \overline{x}^k := (H+A)^{-1}\big(H - M - C\big)x^k & (71a) \\ x^{k+1} := x^k + \alpha_k S^{-1}(H+M^*)(\overline{x}^k - x^k), & (71b) \end{cases}$$

where $\alpha_k = \big[\lambda_k\|\overline{z}^k - z^k\|_P^2\|\big] \,/\, \big[\|(H+M^*)(\overline{z}^k - z^k)\|_{S^{-1}}^2\big]$ and $\lambda_k \in [\underline{\lambda}, \overline{\lambda}] \leq [0, (2 - 1/(2\beta)]$.

Proposition. *Let* $(x^k, \overline{x}^k)$ *be the sequence generated by the AFBAS algorithm. Denote* $\theta_k = \alpha_k - 1$, $v^k = (H+M^*)(x^k) - (H+M^*)(\overline{x}^k)$, *and* $\epsilon_k = \frac{\|\overline{z}^k - z^k\|_P^2}{4\beta}$. *Then,*

$$\begin{cases} (\overline{x}^k, v^k) \in \operatorname{gph}(A+M+C)^{[\epsilon_k]}, & (72a) \\ \theta_k\|S^{-1}v^k\|_S^2 + \|S^{-1}v + (\overline{x}^k - x^k)\|_S^2 + 2\epsilon \leq \sigma\|\overline{x}^k - x^k\|_S^2, & (72b) \\ x^{k+1} = x^k - (1+\theta_k)S^{-1}v^k. & (72c) \end{cases}$$

Proof. We first argue that $C(z) \in C^{[\epsilon]}(x)$ with $\epsilon = \|x-z\|_P^2/(4\beta)$ for any $x, z \in \mathbb{X}$. Notice that for any $y \in \mathbb{X}$,

$$\begin{aligned}\langle x-y, C(z)-C(y)\rangle &= \langle x-z, C(z)-C(y)\rangle + \langle z-y, C(z)-C(y)\rangle \\ &\geq \langle x-z, C(z)-C(y)\rangle + \beta\|C(z)-C(y)\|_{P^{-1}}^2 \\ &\geq -\|x-z\|_P\|C(z)-C(y)\|_{P^{-1}} + \beta\|C(z)-C(y)\|_{P^{-1}}^2 \\ &\geq \inf_{t\geq 0} \beta t^2 - \|x-z\|_P t = -\|x-z\|_P^2/(4\beta),\end{aligned}$$

where the first inequality holds by $\langle x-x', C(x)-C(x')\rangle \geq \beta\|C(x)-C(x')\|_{P^{-1}}^2$, which implies $C(z) \in C^{[\epsilon]}(x)$ with $\epsilon = \|x-z\|_P^2/(4\beta)$ by the definition of $C^{[\epsilon]}(x)$. Specifying (x, z) as $(x^k, \overline{x}^k)$, it holds that $C(x^k) \in C^{[\epsilon_k]}(\overline{x}^k)$ with $\epsilon_k = \|x^k - \overline{x}^k\|_P^2/(4\beta)$. This inclusion equation, in combination with (71a), yields

$$\begin{aligned}(H-M)(x^k) - (H-M)(\overline{x}^k) &\in A(\overline{x}^k) + M(\overline{x}^k) + C(x^k) \\ &\subseteq A(\overline{x}^k) + M(\overline{x}^k) + C^{[\epsilon_k]}(\overline{x}^k) \\ &\subseteq (A+M+C)^{[\epsilon_k]}(\overline{x}^k).\end{aligned}$$

Due to the definition of v^k and the operator M being skew-adjoint, the above inequality indicates $v^k \in (A+M+C)^{[\epsilon_k]}(\overline{x}^k)$, *i.e.*, (72a) holds. Next, we argue that (72b) holds. Utilizing the formula of v^k, we obtain

$$\begin{aligned}&\theta\|S^{-1}v^k\|_S^2 + \|S^{-1}v^k + \overline{z}^k - z^k\|_S^2 + 2\epsilon_k \\ &= \theta\|(H+M^*)(x^k - \overline{x}^k)\|_{S^{-1}}^2 + \|(H+M^*-S)(x^k - \overline{z}^k)\|_{S^{-1}}^2 + \|x^k - \overline{x}^k\|_P^2/(2\beta) \\ &= \|x^k - \overline{x}^k\|^2_{\theta_k(H-M)S^{-1}(H+M^*)+(H-M-S)S^{-1}(H+M^*-S)+P/(2\beta)} \\ &= \|x^k - \overline{x}^k\|^2_{(\theta_k+1)(H-M)S^{-1}(H+M^*)-2H+S+P/(2\beta)} \\ &= \|x^k - \overline{x}^k\|^2_{(\theta_k+1)(H-M)S^{-1}(H+M^*)-(2-1/(2\beta)P+S} \\ &\leq \sigma\|x^k - \overline{x}^k\|_S^2,\end{aligned}$$

where the first equality holds by using the definition of ϵ_k, the second and the third equalities hold according to M being skew-adjoint, the fourth equality holds by $H = P + K$ and K being skew-adjoint, and the last inequality holds by the condition on $\theta_k = \alpha_k - 1$, which implies that (72b) holds. At last, $x^{k+1} = x^k + \alpha_k S^{-1}(H+M^*)(\overline{x}^k - x^k) = x^k - (1+\theta_k)S^{-1}v^k$ holds by utilizing the definitions of v^k and θ_k. Hence, (72c) holds. By now, we have shown that the AFBAS algorithm with the iterations (71a)-(71b) falls into the framework of VMOR-HPE. The proof is finished. □

H. Proof of Proposition 5

The Condat-Vu **P**rimal-**D**ual **S**plitting (Condat-Vu PDS) algorithm (Vũ, 2013; Condat, 2013) takes the following iterations:

$$\begin{cases} \widetilde{x}^{k+1} := \mathrm{Prox}_{r^{-1}g}\big(x^k - r^{-1}\nabla f(x^k) - r^{-1}B^* y^k\big), & (73a) \\ \widetilde{y}^{k+1} := \mathrm{Prox}_{s^{-1}h^*}\big(y^k + s^{-1}B(2\widetilde{x}^{k+1} - x^k)\big), & (73b) \\ (x^{k+1}, y^{k+1}) := (x^k, y^k) + (1+\theta_k)\big((\widetilde{x}^{k+1}, \widetilde{y}^{k+1}) - (x^k, y^k)\big). & (73c)\end{cases}$$

Proposition. *Let $(x^k, y^k, \widetilde{x}^k, \widetilde{y}^k)$ be the sequence generated by the Condat-Vu PDS algorithm. Let $z^k = (x^k, y^k)$, and $w^k = (\widetilde{x}^{k+1}, \widetilde{y}^{k+1})$. Parameters (r, s, θ_k) satisfy $s - r^{-1}\|\mathcal{B}\|^2 > 0$, and $\theta_k + L/[2(s - r^{-1}\|\mathcal{B}\|^2)] \leq \sigma$. Denote $v^k = \mathcal{M}(z^k - w^k)$ and $\epsilon_k = L\|x^k - \widetilde{x}^{k+1}\|^2/4$. Then,*

$$\begin{cases} v^k \in T^{[\epsilon_k]}(w^k), & (74a) \\ \theta_k\|\mathcal{M}^{-1}v^k\|_{\mathcal{M}}^2 + \|\mathcal{M}^{-1}v^k + w^k - z^k\|_{\mathcal{M}}^2 + 2\epsilon_k \leq \sigma\|w^k - z^k\|_{\mathcal{M}}^2, & (74b) \\ z^{k+1} = z^k - (1+\theta_k)\mathcal{M}^{-1}v^k. & (74c)\end{cases}$$

Proof. By the definition of $\mathrm{Prox}_{r^{-1}g}$, (73a) yields $r(x^k - \widetilde{x}^{k+1}) - B^*y^k \in \partial g(\widetilde{x}^{k+1}) + \nabla f(x^k)$. Using (Svaiter, 2014, Lemma 2.2), we obtain $\nabla f(x^k) \in (\nabla f)^{[\epsilon_k]}(\widetilde{x}^{k+1})$ with $\epsilon_k = L\|x^k - \widetilde{x}^{k+1}\|^2/4$. Combining the above two inclusions and performing simple calculations yield

$$r(x^k - \widetilde{x}^{k+1}) - B^*(y^k - \widetilde{y}^{k+1}) \in \partial g(\widetilde{x}^{k+1}) + (\nabla f)^{[\epsilon_k]}(\widetilde{x}^{k+1}) + B^*\widetilde{y}^{k+1}. \quad (75)$$

Using the definition of $\mathrm{Prox}_{s^{-1}h^*}$ and performing similar operations on $\widetilde{y}^{k+1}$ as $\widetilde{x}^{k+1}$, we obtain

$$s(y^k - \widetilde{y}^{k+1}) - B(x^k - \widetilde{x}^{k+1}) \in \partial h^*(\widetilde{y}^{k+1}) - B\widetilde{x}^{k+1}. \tag{76}$$

By the definitions of $\mathcal{M}, z^k, w^k, T$ and $T^{[\epsilon]}$, (75) and (76) indicate that $\mathcal{M}(z^k - w^k) \in T^{[\epsilon_k]}(w^k)$. Thus, (73a) holds by utilizing $v^k = \mathcal{M}(z^k - w^k)$. In addition, (73c) can be equivalently reformulated as $z^{k+1} = z^k + (1+\theta_k)(w^k - z^k) = z^k - (1+\theta_k)\mathcal{M}^{-1}v^k$ by using the definitions of z^k, w^k and v^k. Hence, (74c) holds. Below, we argue that (74b) holds. By the definition of v^k, it holds that

$$\begin{aligned}
\theta_k\big\|\mathcal{M}v^k\big\|_{\mathcal{M}}^2 + \big\|\mathcal{M}^{-1}v^k + w^k - z^k\big\|_{\mathcal{M}}^2 + 2\epsilon_k &\le \theta_k\big\|w^k - z^k\big\|_{\mathcal{M}}^2 + L\|x^k - \widetilde{x}^{k+1}\|^2/2 \\
&\le \big(\theta_k + L/(2\lambda_{\min}(\mathcal{M}))\big)\|z^k - w^k\|_{\mathcal{M}}^2 \\
&\le \big[\theta_k + L/[2(s - r^{-1}\|\mathcal{B}\|^2)]\big]\big\|w^k - z^k\big\|_{\mathcal{M}}^2 \\
&\le \sigma\big\|w^k - z^k\big\|_{\mathcal{M}}^2,
\end{aligned}$$

where the first and the second inequalities hold by using ϵ_k and $\|x^k - \widetilde{x}^{k+1}\|^2 \le \|z^k - w^k\|^2 \le \|z^k - w^k\|_{\mathcal{M}}^2/\lambda_{\min}(\mathcal{M})$, respectively. Hence, (74b) holds. In conclusion, the Condat-Vu PDS algorithm with the iterations (73a)-(73c) falls into the framework of VMOR-HPE. The proof is finished. □

I. Proof of Proposition 6

The **A**symmetric **F**orward **B**ackward **A**djoint **S**plitting **P**rimal-**D**ual (AFBAS-PD) algorithm (Latafat & Patrinos, 2017) is defined as

$$\begin{cases}
\overline{x}^k := \mathrm{Prox}_{\gamma_1 g}\big(x^k - \gamma_1 B^* y^k - \gamma_1 \nabla f(x^k)\big), & \text{(77a)}\\
\overline{y}^k := \mathrm{Prox}_{\gamma_2 h^*}\big(y^k + \gamma_2 B((1-\theta)x^k + \theta\overline{x}^k)\big), & \text{(77b)}\\
x^{k+1} := x^k + \alpha_k\big((\overline{x}^k - x^k) - \mu\gamma_1(2-\theta)B^*(\overline{y}^k - y^k)\big), & \text{(77c)}\\
y^{k+1} := y^k + \alpha_k\big(\gamma_2(1-\mu)(2-\theta)B(\overline{x}^k - x^k) + (\overline{y}^k - y^k)\big), & \text{(77d)}
\end{cases}$$

where $\alpha_k = \big[\lambda_k(\gamma_1^{-1}\|\overline{x}^k - x^k\|^2 + \gamma_2^{-1}\|\overline{y}^k - y^k\|^2 - \theta\langle\overline{x}^k - x^k, B^*(\overline{y}^k - y^k)\rangle)\big]/V(\overline{x}^k - x^k, \overline{y}^k - y^k)$, $\lambda_k \in [\underline{\lambda}, \overline{\lambda}] \subseteq (0, \delta)$, and δ and $V(x,y)$ are defined as $\delta = 2 - L(\gamma_1^{-1} - \gamma_2\theta^2\|B\|^2/4)^{-1}/2$ and $V(x,y) = \gamma_1^{-1}\|x\|^2 + \gamma_2^{-1}\|y\|^2 + (1-\mu)\gamma_2(1-\theta)(2-\theta)\|Bx\|^2 + \mu\gamma_1(2-\theta)\|B^*y\|^2 + 2((1-\mu)(1-\theta) - \mu)\langle x, B^*y\rangle$ which requires $\gamma_1^{-1} - \gamma_2\theta^2\|B\|^2/4 > L/4$ and $\mu \in [0,1], \theta \in [0, \infty)$.

Denote a linear operator $M : \mathbb{Z} \to \mathbb{Z}$ that $M = RS^{-1}$, where $R, S : \mathbb{Z} \to \mathbb{Z}$ are defined as below

$$R = \begin{bmatrix} \gamma_1^{-1} & -B^* \\ (1-\theta)B & \gamma_2^{-1} \end{bmatrix},\ S = \begin{bmatrix} 1 & -\mu\gamma_1(2-\theta)B^* \\ \gamma_2(1-\mu)(2-\theta)B & 1 \end{bmatrix}. \tag{78}$$

By the block matrix inversion formula (Horn & Johnson, 1990), R^{-1} and M^{-1} are derived as below

$$R^{-1} = \begin{bmatrix} \gamma_2^{-1}\Xi & \Xi B^* \\ -(1-\theta)B\Xi & \gamma_2 - \gamma_2(1-\theta)B\Xi B^* \end{bmatrix},\ \Xi = \big[\gamma_1^{-1}\gamma_2^{-1} + (1-\theta)B^*B\big]^{-1},$$

$$M^{-1} = SR^{-1} = \begin{bmatrix} \gamma_1\mu(2-\theta) + \gamma_2^{-1}[1-\mu(2-\theta)]\Xi & [1-\mu(2-\theta)]\Xi B^* \\ [1-\mu(2-\theta)]B\Xi & \gamma_2 + \gamma_2[1-\mu(2-\theta)]B\Xi B^* \end{bmatrix}.$$

Here, we claim that $\Xi = \big[\gamma_1^{-1}\gamma_2^{-1} + (1-\theta)B^*B\big]^{-1} \succ 0$. In fact, if $\theta \le 1$, it is obvious that $\Xi \succ 0$, otherwise, $\gamma_1^{-1} - \gamma_2\theta^2\|B\|^2/4 > L/4 > 0$ indicates $\gamma_1^{-1}\gamma_2^{-1} > \theta^2\|B\|^2/4 > (\theta-1)\|B\|^2 \succeq (\theta-1)B^*B$. Hence, $\Xi \succ 0$ holds for $\theta \ge 0$. In addition, M is a self-adjoint positive definite linear operator by Schur complement theorem (Horn & Johnson, 1990).

Proposition. *Let $\{(\overline{x}^k, \overline{y}^k, x^k, y^k)\}$ be the sequence generated by the AFBAS-PD algorithm. Denote $w^k = (\overline{x}^k, \overline{y}^k)$, $z^k = (x^k, y^k)$, $v^k = R(z^k - w^k)$, $\epsilon_k = L\|x^k - \overline{x}^k\|^2/4$, and $\theta_k = \alpha_k - 1$. Then, it holds that*

$$\begin{cases}
v^k \in T^{[\epsilon_k]}(w^k), & \text{(79a)}\\
\theta_k\big\|\mathcal{M}^{-1}v^k\big\|_{\mathcal{M}}^2 + \big\|\mathcal{M}^{-1}v^k + w^k - z^k\big\|_{\mathcal{M}}^2 + 2\epsilon_k \le \sigma\big\|w^k - z^k\big\|_{\mathcal{M}}^2, & \text{(79b)}\\
z^{k+1} = z^k - (1+\theta_k)\mathcal{M}^{-1}v^k. & \text{(79c)}
\end{cases}$$

Proof. By the definition of $\mathrm{Prox}_{\gamma_1 g}$, (77a) indicates $x^k - \gamma_1 B^* y^k - \gamma_1 \nabla f(x^k) \in \overline{x}^k + \gamma_1 \partial g(\overline{x}^k)$, *i.e.*,

$$\gamma_1^{-1}(x^k - \overline{x}^k) - B^*(y^k - \overline{y}^k) \in \partial g(\overline{x}^k) + (\nabla f)^{[\epsilon_k]}(\overline{x}^k) + B^* \overline{y}^k \tag{80}$$

by using $\nabla f(x^k) \in (\nabla f)^{[\epsilon_k]}(\overline{x}^k)$. Similarly, by the definition of $\mathrm{Prox}_{\gamma_2 h^*}$, (77a) indicates that

$$(1-\theta)B(x^k - \overline{x}^k) + \gamma_2^{-1}(y^k - \overline{y}^k) \in \partial g(\overline{y}^k) - B\overline{x}^k. \tag{81}$$

By the definitions of (z^k, w^k, v^k, T^k) and using the additivity property of enlargement operator (Burachik et al., 1998), the two inclusions (80)-(81) indicate that $v^k = R(z^k - w^k) \in T^{[\epsilon_k]}(w^k)$. Hence, (79a) holds. By using (z^k, w^k), (77c)-(77d) can be reformulated as a compact form that

$$z^{k+1} = z^k - \alpha_k S(z^k - w^k) = z^k - \alpha_k M^{-1} R(z^k - w^k) = z^k - \alpha_k M^{-1} v^k, \tag{82}$$

which indicates that (79c) holds. At last, we verify (79b). By the definition of (M, ϵ_k, v^k), it holds

$$\begin{aligned}
&\theta_k \big\|\mathcal{M}^{-1} v^k\big\|_{\mathcal{M}}^2 + \big\|\mathcal{M}^{-1} v^k + w^k - z^k\big\|_{\mathcal{M}}^2 + 2\epsilon_k - \sigma \|w^k - z^k\|_{\mathcal{M}}^2 \\
&= \|w^k - z^k\|^2_{(\theta_k+1)S^*MS - S^*M - MS + (1-\sigma)M} + L\|x^k - \overline{x}^k\|^2/2 \\
&= \|w^k - z^k\|^2_{\alpha_k S^* R - R^* - R + (1-\sigma)M} + L\|x^k - \overline{x}^k\|^2/2,
\end{aligned}$$

where the first equality holds due to $\mathcal{M}^{-1} v^k = S(z^k - w^k)$, and the second equality holds due to $MS = R$. Hence, $\theta_k \big\|\mathcal{M}^{-1} v^k\big\|_{\mathcal{M}}^2 + \big\|\mathcal{M}^{-1} v^k + w^k - z^k\big\|_{\mathcal{M}}^2 + 2\epsilon_k < \sigma \big\|w^k - z^k\big\|_{\mathcal{M}}^2$, *i.e.*, (79b) holds if it can be shown that $\alpha_k < \big[\|w^k - z^k\|^2_{R^*+R} - L\|x^k - \overline{x}^k\|^2/2\big]/\|w^k - z^k\|^2_{S^*R}$. Notice

$$S^* R = \begin{bmatrix} \gamma_1^{-1} + \gamma_2(1-\mu)(2-\theta)(1-\theta)B^*B & [(1-\mu)(1-\theta) - \mu]B^* \\ [(1-\mu)(1-\theta) - \mu]B & \gamma_2^{-1} + \mu\gamma_1(2-\theta)BB^* \end{bmatrix}.$$

Simple algebraic manipulations yield $\|w^k - z^k\|^2_{S^*R} = V(x^k - \overline{x}^k, y^k - \overline{y}^k)$. In addition,

$$\begin{aligned}
&\|w^k - z^k\|^2_{R^*+R} - L\|x^k - \overline{x}^k\|^2/2 \\
&= 2\big[\gamma_1^{-1}\|x^k - \overline{x}^k\|^2 + \gamma_2^{-1}\|y^k - \overline{y}^k\|^2 - \theta\langle x^k - \overline{x}^k, B^*(y^k - \overline{y}^k)\rangle\big] - L\|x^k - \overline{x}^k\|^2/2 \\
&\geq \big[2 - L/[2(\gamma_1^{-1} - \gamma_2\theta^2\|B\|^2/4)]\big]\big[\gamma_1^{-1}\|x^k - \overline{x}^k\|^2 + \gamma_2^{-1}\|y^k - \overline{y}^k\|^2 - \theta\langle x^k - \overline{x}^k, B^*(y^k - \overline{y}^k)\rangle\big],
\end{aligned}$$

where the first equality holds by using the definition of R, and the second inequality holds by the fact that

$$\|x^k - \overline{x}^k\|^2 \leq \|x^k - \overline{x}^k\|_P^2 \lambda_{\max}(P^{-1}) \leq \|x^k - \overline{x}^k\|_P^2 \lambda_{\min}^{-1}(P) \leq (\gamma_1^{-1} - \gamma_2\theta^2\|B\|^2/4)^{-1}\|x^k - \overline{x}^k\|_P^2,$$

where $P = \begin{pmatrix} \gamma_1^{-1} & -\theta B^*/2 \\ -\theta B/2 & \gamma_2^{-1} \end{pmatrix} \succ 0$. Hence, we have that $\theta_k = \alpha_k < \big[\|w^k - z^k\|^2_{R^*+R} - L\|x^k - \overline{x}^k\|^2/2\big]/\|w^k - z^k\|^2_{S^*R}$ holds. In conclusion, the AFBAS-PD algorithm with the iterations (77a)-(77d) falls into the framework of VMOR-HPE algorithm. The proof is finished. □

J. Proof of Theorem 4

Theorem. *Let $(\widetilde{x}^k, \widetilde{y}^k, x^k, y^k)$ be the sequence generated by the PADMM-EBB algorithm. Denote $v^k = U^k(z^k - w^k)$, $\epsilon_k = \|x^k - \widetilde{x}^{k+1}\|_{\mathcal{D}}/4$, and operator T as (25). Then, it holds that*

$$\begin{cases}
v^k \in T^{[\epsilon_k]}(w^k), & (83a)\\
\theta_k \big\|\mathcal{M}_k^{-1} v^k\big\|_{\mathcal{M}_k}^2 + \big\|\mathcal{M}_k^{-1} v^k + w^k - z^k\big\|_{\mathcal{M}_k}^2 + 2\epsilon_k \leq \sigma \big\|w^k - z^k\big\|_{\mathcal{M}_k}^2, & (83b)\\
z^{k+1} = z^k - (1+\theta_k)\mathcal{M}_k^{-1} v^k. & (83c)
\end{cases}$$

Besides, **(i)** $(x^k, \widetilde{x}^k)$ *and* $(y^k, \widetilde{y}^k)$ *converge to* x^∞ *and* y^∞*, respectively, belonging to the optimal primal-dual solution set of* (6).
(ii) *There exists an integer* $\overline{k} \in \{1, 2, \ldots, k\}$ *such that*

$$\sum_{i=1}^p \mathrm{dist}\big((\partial g_i + \nabla f_i)(\widetilde{x}^{\overline{k}}) + \mathcal{A}_i \widetilde{y}^{\overline{k}}, 0\big) + \big\|b - \sum_{i=1}^p \mathcal{A}_i^* \widetilde{x}_i^{\overline{k}}\big\| \leq \mathcal{O}(\frac{1}{\sqrt{k}}).$$

(iii) *Let* $\alpha_i = 1$ or i. *There exists* $0 \leq \overline{\epsilon}_k^{x_i} \leq \mathcal{O}(\frac{1}{k})$ *such that*

$$\sum_{i=1}^{p} \mathrm{dist}\big((\partial g_i + \nabla f_i)_{\overline{\epsilon}_k^{x_i}}(\overline{x}^k) + \mathcal{A}_i \overline{y}^k, 0\big) + \big\|b - \sum_{i=1}^{p} \mathcal{A}_i^* \overline{x}_i^k\big\| \leq \mathcal{O}(\frac{1}{k}),$$

where $\overline{x}^k = \frac{\sum_{i=1}^k (1+\theta_i)\alpha_i \widetilde{x}^{i+1}}{\sum_{i=1}^k (1+\theta_i)\alpha_i}$ *and* $\overline{y}^k = \frac{\sum_{i=1}^k (1+\theta_i)\alpha_i \widetilde{y}^{i+1}}{\sum_{i=1}^k (1+\theta_i)\alpha_i}$.

(iv) *If* T *satisfies metric subregularity at* $\big((x^\infty, y^\infty), 0\big) \in \mathrm{gph} T$ *with modulus* $\kappa > 0$. *Then, there exists* $\overline{k} > 0$ *such that*

$$\mathrm{dist}_{\mathcal{M}_{k+1}}\big((x^{k+1}, y^{k+1}), T^{-1}(0)\big) \leq \Big(1 - \frac{\varrho_k}{2}\Big) \mathrm{dist}_{\mathcal{M}_k}\big((x^k, y^k), T^{-1}(0)\big), \ \forall k \geq \overline{k},$$

where $\varrho_k := [(1-\sigma)(1+\theta_k)] \Big/ \Big[\big(1 + \kappa\sqrt{\frac{\Xi\overline{\omega}}{\underline{\omega}}}\big)^2 \big(1 + \sqrt{\sigma + \frac{4\max\{-\theta_k, 0\}}{(1+\theta_k)^2}}\big)^2\Big] \in (0, 1)$.

Proof. By the optimality condition of the subproblem of $\widetilde{x}_i^{k+1}$, the following inclusion directly holds for $i = 1, \ldots, p$ that

$$0 \in \nabla f_i(x^k) + \partial g_i(\widetilde{x}_i^{k+1}) + \mathcal{A}_i y^k + \beta_k \mathcal{A}_i \big(\sum_{j=1}^{i} \mathcal{A}_j^* \widetilde{x}_j^{k+1} + \sum_{j=i+1}^{p} \mathcal{A}_j^* x_j^k - b\big) + \big(\widehat{\Sigma}_i + P_i^k\big)(\widetilde{x}_i^{k+1} - x_i^k).$$

Substituting $y^k = \widetilde{y}^{k+1} - \beta_k\big(\mathcal{A}_1^* \widetilde{x}_1^{k+1} + \sum_{i=2}^p \mathcal{A}_i^* x_i^k - b\big)$ into the above inclusion, we obtain

$$\big(\widehat{\Sigma}_i + P_i^k\big)(x_i^k - \widetilde{x}_i^{k+1}) + \beta_k \mathcal{A}_i \sum_{j=2}^{i} \mathcal{A}_j^* (x_j^k - \widetilde{x}_j^{k+1}) \in \nabla f_i(x^k) + \partial g_i(\widetilde{x}_i^{k+1}) + \mathcal{A}_i \widetilde{y}^{k+1}. \tag{84}$$

Stacking (84) for $i = 1, 2, \ldots, p$ and $y^k = \widetilde{y}^{k+1} - \beta_k\big(\mathcal{A}_1^* \widetilde{x}_1^{k+1} + \sum_{i=2}^p \mathcal{A}_p^* x_p^k - b\big)$, we obtain

$$\begin{bmatrix} (\widehat{\Sigma}_1 + P_1^k)(x_1^k - \widetilde{x}_1^{k+1}) \\ \vdots \\ \big(\widehat{\Sigma}_i + P_i^k\big)(x_i^k - \widetilde{x}_i^{k+1}) + \beta_k \mathcal{A}_i \sum_{j=2}^{i} \mathcal{A}_j^* (x_j^k - \widetilde{x}_j^{k+1}) \\ \vdots \\ \big(\widehat{\Sigma}_p + P_p^k\big)(x_p^k - \widetilde{x}_p^{k+1}) + \beta_k \mathcal{A}_p \sum_{j=2}^{p} \mathcal{A}_j^* (x_j^k - \widetilde{x}_j^{k+1}) \\ \beta_k^{-1}(y^k - \widetilde{y}^{k+1}) + \sum_{i=2}^p \mathcal{A}_p^* (x_i^k - \widetilde{x}_i^{k+1}) \end{bmatrix} \in \begin{bmatrix} \partial g_1(\widetilde{x}_1^{k+1}) \\ \vdots \\ \partial g_i(\widetilde{x}_i^{k+1}) \\ \vdots \\ \partial g_p(\widetilde{x}_p^{k+1}) \\ b \end{bmatrix} + \begin{bmatrix} \nabla f_1(x^k) + \mathcal{A}_1 \widetilde{y}^{k+1} \\ \vdots \\ \nabla f_i(x^k) + \mathcal{A}_i \widetilde{y}^{k+1} \\ \vdots \\ \nabla f_p(x^k) + \mathcal{A}_p \widetilde{y}^{k+1} \\ -\sum_{i=1}^p \mathcal{A}_i^* \widetilde{x}_i^{k+1} \end{bmatrix}.$$

By utilizing the notations U^k, z^k, w^k and T, the above inclusion is further reformulated as:

$$\begin{aligned} U^k(z^k - w^k) &\in \begin{bmatrix} \partial g(\widetilde{x}^{k+1}) \\ b \end{bmatrix} + \begin{bmatrix} \nabla f(x^k) \\ 0 \end{bmatrix} + \begin{bmatrix} \mathcal{A}^* \widetilde{y}^k \\ -\sum_{i=1}^p \mathcal{A}_i^* \widetilde{x}_i^{k+1} \end{bmatrix} \\ &\subseteq \begin{bmatrix} \partial g(\widetilde{x}^{k+1}) \\ b \end{bmatrix} + \begin{bmatrix} \nabla f^{[\epsilon_k]}(\widetilde{x}^{k+1}) \\ 0 \end{bmatrix} + \begin{bmatrix} \mathcal{A}^* \widetilde{y}^k \\ -\sum_{i=1}^p \mathcal{A}_i^* \widetilde{x}_i^{k+1} \end{bmatrix}, \end{aligned} \tag{85}$$

where $g(x) = \sum_{i=1}^p g_i(x_i)$, and $\mathcal{A} = [\mathcal{A}_1 \ \ \mathcal{A}_2 \ \ \cdots \ \ \mathcal{A}_p]$. Using the additivity property of enlargement operator (Burachik et al., 1998) and the definition of T, the above inclusion indicates

$$v^k = U^k(z^k - w^k) \in T^{[\epsilon_k]}(w^k).$$

Besides, by utilizing the updating step of (x^{k+1}, y^{k+1}) and the definition of (v^k, w^k, z^k) , it holds that

$$\begin{aligned} z^{k+1} = (x^{k+1}, y^{k+1}) &= (x^k, y^k) + (1+\theta_k)\mathcal{M}_k^{-1} U^k \big(\widetilde{x}^{k+1} - x^k, \widetilde{y}^{k+1} - y^k\big) \\ &= z^k + (1+\theta_k)\mathcal{M}_k^{-1} U^k (w^k - z^k) \\ &= z^k - (1+\theta_k)\mathcal{M}_k^{-1} v^k. \end{aligned}$$

Hence, (83a) and (83c) hold. At last, we check (83b). By the definition of (v^k, ϵ_k), it holds that

$$\begin{aligned}&\theta_k\big\|\mathcal{M}_k^{-1}v^k\big\|_{\mathcal{M}_k}^2 + \big\|\mathcal{M}_k^{-1}v^k + w^k - z^k\big\|_{\mathcal{M}_k}^2 + 2\epsilon_k - \sigma\big\|w^k - z^k\big\|_{\mathcal{M}_k}^2\\ &= \big\|w^k - z^k\big\|^2_{(1+\theta_k)(U^k)^*\mathcal{M}_k^{-1}U^k - (U^k)^* - U^k + (1-\sigma)\mathcal{M}_k + \mathcal{D}/2}\\ &\le 0,\end{aligned}$$

where the last inequality holds by the setting of over-relaxed step-size θ_k. Hence, PADMM-EBB is equivalently reformulated as (83a)-(83c), *i.e.*, it falls into the framework of VMOR-HPE. By Theorem 1, **(i)** directly holds that (x^k, y^k) and $(\widetilde{x}^k, \widetilde{y}^k)$ simultaneously converge to a point (x^∞, y^∞) belonging to $T^{-1}(0)$ which is exactly the primal-dual optimal solution set of (6). In the following, we argue that **(ii)** and **(iii)** hold by utilizing Theorem 3. In fact, using (85), we have

$$v^k + \begin{bmatrix} \nabla f(\widetilde{x}^{k+1}) \\ 0 \end{bmatrix} - \begin{bmatrix} \nabla f(x^k) \\ 0 \end{bmatrix} \in \begin{bmatrix} \partial g(\widetilde{x}^{k+1}) \\ b \end{bmatrix} + \begin{bmatrix} \nabla f(\widetilde{x}^{k+1}) \\ 0 \end{bmatrix} + \begin{bmatrix} \mathcal{A}^*\widetilde{y}^k \\ -\sum_{i=1}^p \mathcal{A}_i^*\widetilde{x}_i^{k+1} \end{bmatrix} = T(w^k).$$

Hence, $\mathrm{dist}\big(T(w^k), 0\big) \le \|v^k\| + L\|x^k - \widetilde{x}^{k+1}\| = \|v^k\| + 4\epsilon_k$. This, in combination with (49), yields the desired result **(i)**, *i.e.*, there exists an integer $\overline{k} \in \{1, 2, \ldots, k\}$ such that

$$\sum_{i=1}^p \mathrm{dist}\big(\partial g_i(\widetilde{x}_i^{\overline{k}}) + \nabla f_i(\widetilde{x}^{\overline{k}}) + \mathcal{A}_i\widetilde{y}^{\overline{k}}, 0\big) + \big\|b - \sum_{i=1}^p \mathcal{A}_i^*\widetilde{x}_i^{\overline{k}}\big\| = \mathrm{dist}\big(T(w^{\overline{k}-1}), 0\big) \le \mathcal{O}(\frac{1}{\sqrt{k}}).$$

Next, we claim that $\overline{\epsilon}_k^{x_j} = \frac{\sum_{i=1}^k (1+\theta_i)\alpha_i\left(\epsilon_k^{x_j} + \langle \widetilde{x}_j^{i+1} - \overline{x}_j^k, G_{x_j}^i - \overline{G}_{x_j}^k\rangle\right)}{\sum_{i=1}^k (1+\theta_i)\alpha_i}$, where $\epsilon_k^{x_j} = \frac{L_j\|x_j^k - \widetilde{x}_j^{k+1}\|^2}{4}$,

$$\begin{aligned} G_{x_1}^i &= (\widehat{\Sigma}_1 + \beta_k P_1^k)\big(x_1^k - \widetilde{x}_1^{i+1}\big) - \mathcal{A}_1\widetilde{y}^{i+1},\\ G_{x_2}^i &= \big(\widehat{\Sigma}_2 + \beta_k(\mathcal{A}_2\mathcal{A}_2^* + P_2^k)\big)\big(x_2^i - \widetilde{x}_2^{i+1}\big) - \mathcal{A}_2\widetilde{y}^{i+1},\\ &\vdots \\ G_{x_p}^i &= \big(\widehat{\Sigma}_p + \beta_k(\mathcal{A}_p\mathcal{A}_p^* + P_p^k)\big)\big(x_p^i - \widetilde{x}_p^{i+1}\big) + \sum_{j=2}^{p-1}\sigma\mathcal{A}_i\mathcal{A}_j^*\big(x_j^i - \widetilde{x}_j^{i+1}\big) - \mathcal{A}_p\widetilde{y}^{i+1},\end{aligned}$$

$$\overline{G}_{x_1}^k = \frac{\sum_{i=1}^k (1+\theta_i)\alpha_i G_{x_1}^i}{\sum_{i=1}^k (1+\theta_i)\alpha_i},\ \overline{G}_{x_2}^k = \frac{\sum_{i=1}^k (1+\theta_i)\alpha_i G_{x_2}^i}{\sum_{i=1}^k (1+\theta_i)\alpha_i},\ \cdots,\ \overline{G}_{x_p}^k = \frac{\sum_{i=1}^k (1+\theta_i)\alpha_i G_{x_p}^i}{\sum_{i=1}^k (1+\theta_i)\alpha_i},$$

$$G_y^i = \beta_k^{-1}(y^i - \widetilde{y}^{i+1}) + \sum_{j=2}^p \mathcal{A}_p^*(x_j^i - \widetilde{x}_j^{i+1}),\ \text{ and } \overline{G}_y^k = \frac{\sum_{i=1}^k (1+\theta_i)\alpha_i G_y^i}{\sum_{i=1}^k (1+\theta_i)\alpha_i}.$$

Define $\overline{w}^k = \frac{\sum_{i=1}^k (1+\theta_i)\alpha_i w^i}{\sum_{i=1}^k (1+\theta_i)\alpha_i}$, $\overline{v}^k = \frac{\sum_{i=1}^k (1+\theta_i)\alpha_i v^i}{\sum_{i=1}^k (1+\theta_i)\alpha_i}$ and $\overline{\epsilon}^k = \frac{\sum_{i=1}^k (1+\theta_i)\alpha_i\left(\epsilon_i + \langle w^i - \overline{w}^k, v^i - \overline{v}^k\rangle\right)}{\sum_{i=1}^k (1+\theta_i)\alpha_i}$ as Theorem 3. Hence, utilizing (51)-(52) and (83a)-(83a), we obtain $\|\overline{v}^k\| \le \frac{1}{k}$ and $\overline{\epsilon}^k \le \frac{1}{k}$ by setting $\alpha_i = 1$ or $\alpha_i = i$. Using (85) and the definitions of $G_{x_1}^k, \cdots, G_{x_p}^k$ and $\overline{G}_{x_1}^k, \cdots, \overline{G}_{x_p}^k$, we have

$$\begin{pmatrix} G_{x_1}^k + \mathcal{A}_1\widetilde{y}^{k+1} \\ \vdots \\ G_{x_p}^k + \mathcal{A}_p\widetilde{y}^{k+1} \end{pmatrix} \in \begin{pmatrix} \big(\partial g_1 + \nabla f_1\big)_{[\epsilon_k^{x_1}]}(\widetilde{x}_1^{k+1}) + \mathcal{A}_1\widetilde{y}^{k+1} \\ \vdots \\ \big(\partial g_p + \nabla f_p\big)_{[\epsilon_k^{x_p}]}(\widetilde{x}_p^{k+1}) + \mathcal{A}_p\widetilde{y}^{k+1} \end{pmatrix}.$$

By utilizing (Burachik et al., 1998, theorem 2.3), it holds that $\overline{\epsilon}_k^{x_i} \ge 0$ for all $i \in \{1, \cdots, p\}$ and

$$\begin{pmatrix} \overline{G}_{x_1}^k + \mathcal{A}_1\overline{y}^k \\ \vdots \\ \overline{G}_{x_p}^k + \mathcal{A}_p\overline{y}^k \end{pmatrix} \subseteq \begin{pmatrix} \big(\partial g_1 + \nabla f_1\big)_{[\overline{\epsilon}_k^{x_1}]}(\overline{x}_1^k) + \mathcal{A}_1\overline{y}^k \\ \vdots \\ \big(\partial g_p + \nabla f_p\big)_{[\overline{\epsilon}_k^{x_p}]}(\overline{x}_p^k) + \mathcal{A}_p\overline{y}^k \end{pmatrix}.$$

By (85) and $G_y^k = \beta_k^{-1}(y^k - \widetilde{y}^{k+1}) + \sum_{i=2}^p \mathcal{A}_p^*(x_i^k - \widetilde{x}_i^{k+1}) = b - \sum_{i=1}^p \mathcal{A}_i^* \widetilde{x}_i^{k+1}$, we get that

$$\overline{v}^k = \begin{pmatrix} \overline{G}_{x_1}^k + \mathcal{A}_1 \overline{y}^k \\ \vdots \\ \overline{G}_{x_p}^k + \mathcal{A}_p \overline{y}^k \\ \overline{G}_y^k \end{pmatrix} \subseteq \begin{pmatrix} \left(\partial g_1 + \nabla f_1\right)_{[\overline{\epsilon}_k^{x_1}]}(\overline{x}_1^k) + \mathcal{A}_1 \overline{y}^k \\ \vdots \\ \left(\partial g_p + \nabla f_p\right)_{[\overline{\epsilon}_k^{x_p}]}(\overline{x}_p^k) + \mathcal{A}_p \overline{y}^k \\ b - \sum_{i=1}^p \mathcal{A}_i^* \overline{x}_i^k \end{pmatrix} \subseteq T^{[\overline{\epsilon}_k^{x_1} + \ldots + \overline{\epsilon}_k^{x_p}]}(w^i).$$

Hence, we obtain $\sum_{i=1}^p \operatorname{dist}\big((\partial g_i + \nabla f_i)_{\overline{\epsilon}_k^{x_i}}(\overline{x}^k) + \mathcal{A}_i \overline{y}^k, 0\big) + \big\|b - \sum_{i=1}^p \mathcal{A}_i^* \overline{x}_i^k\big\| \leq \|\overline{v}^k\| \leq \mathcal{O}(\frac{1}{k})$. Next, we show that $0 \leq \overline{\epsilon}_k^{x_i} \leq \mathcal{O}(\frac{1}{k})$ for all $i = 1, 2, \ldots, p$. Notice

$$\begin{aligned} \overline{\epsilon}_k^{x_1} + \cdots + \overline{\epsilon}_k^{x_p} &= \sum_{j=1}^p \left\{ \frac{1}{\sum_{i=1}^k (1+\theta_i)\alpha_i} \sum_{i=1}^k (1+\theta_i)\alpha_i \big(\epsilon_k^{x_i} + \langle \widetilde{x}_j^{i+1} - \overline{x}_j^k, G_{x_j}^i - \overline{G}_{x_j}^k \rangle\big) \right\} \\ &= \frac{1}{\sum_{i=1}^k (1+\theta_i)\alpha_i} \sum_{i=1}^k (1+\theta_i)\alpha_i \big(\sum_{j=1}^p \epsilon_i^{x^j} + \sum_{j=1}^p \langle \widetilde{x}_j^{i+1} - \overline{x}_j^k, G_{x_j}^i - \overline{G}_{x_j}^k \rangle \big) \\ &= \frac{1}{\sum_{i=1}^k (1+\theta_i)\alpha_i} \sum_{i=1}^k (1+\theta_i)\alpha_i \big(\epsilon_i + \langle \widetilde{x}^{i+1} - \overline{x}^k, G_x^i - \overline{G}_x^k \rangle\big), \end{aligned} \tag{86}$$

where the third equality holds according to $\epsilon_i = \sum_{j=1}^p \epsilon_i^{x_j}$, and $(\widetilde{x}^{i+1}, \overline{x}^k, G_x^i, \overline{G}_x^k)$ are defined as

$$\widetilde{x}^{i+1} = \begin{pmatrix} \widetilde{x}_1^{i+1} \\ \vdots \\ \widetilde{x}_p^{i+1} \end{pmatrix}, \overline{x}^k = \begin{pmatrix} \overline{x}_1^k \\ \vdots \\ \overline{x}_p^k \end{pmatrix}, G_x^i = \begin{pmatrix} G_{x_1}^i \\ \vdots \\ G_{x_p}^i \end{pmatrix}, \overline{G}_x^k = \begin{pmatrix} \overline{G}_{x_1}^k \\ \vdots \\ \overline{G}_{x_p}^k \end{pmatrix}.$$

Let $v_i^k = G_{x_i}^k + \mathcal{A}_i \widetilde{y}^{k+1}$ be the i-th component of v^k. Using $\widetilde{x}^{i+1}$, $\overline{x}^k$, G_x^i, $\overline{G}_x^k$, we obtain

$$\begin{aligned} &\sum_{i=1}^k (1+\theta_i)\alpha_i \langle \widetilde{x}^{i+1} - \overline{x}^k, G_x^i - \overline{G}_x^k \rangle = \sum_{i=1}^k (1+\theta_i)\alpha_i \langle \widetilde{x}^{i+1} - \overline{x}^k, G_x^i \rangle \qquad (87) \\ &= \sum_{i=1}^k (1+\theta_i)\alpha_i \langle \widetilde{x}^{i+1} - \overline{x}^k, [v_1^i - \mathcal{A}_1 \widetilde{y}^{i+1}, \cdots, v_p^i - \mathcal{A}_p \widetilde{y}^{i+1}]^T \rangle \\ &= \sum_{i=1}^k (1+\theta_i)\alpha_i \langle \widetilde{x}^{i+1} - \overline{x}^k, [v_1^i, \cdots, v_p^i]^T \rangle - \sum_{i=1}^k (1+\theta_i)\alpha_i \langle \widetilde{x}^{i+1} - \overline{x}^k, [\mathcal{A}_1 \widetilde{y}^{i+1}, \cdots, \mathcal{A}_p \widetilde{y}^{i+1}]^T \rangle \\ &= \sum_{i=1}^k (1+\theta_i)\alpha_i \langle \widetilde{x}^{i+1} - \overline{x}^k, [v_1^i, \cdots, v_p^i]^T \rangle - \sum_{i=1}^k (1+\theta_i)\alpha_i (\widetilde{y}^{i+1})^\top \sum_{j=1}^p \mathcal{A}_j^* (x_j^{i+1} - \overline{x}_j^k) \\ &= -\sum_{i=1}^k (1+\theta_i)\alpha_i \langle \widetilde{y}^{i+1}, G_y^i - \overline{G}_y^i) \rangle - \sum_{i=1}^k (1+\theta_i)\alpha_i (\widetilde{y}^{i+1})^\top \sum_{j=1}^p \mathcal{A}_j^* (\widetilde{x}_j^{i+1} - \overline{x}_j^k) + \sum_{i=1}^k (1+\theta_i)\alpha_i \langle w^i - \overline{w}^k, v^i \rangle, \end{aligned}$$

where the last equality holds by using the definitions of v^k, w^k and $\overline{v}^k, \overline{w}^k$. In addition,

$$\begin{aligned} &\sum_{i=1}^k (1+\theta_i)\alpha_i (\widetilde{y}^{i+1})^\top \sum_{j=1}^p \mathcal{A}_j^* (\widetilde{x}_j^{i+1} - \overline{x}_j^k) + \sum_{i=1}^k (1+\theta_i)\alpha_i \langle \widetilde{y}^{i+1}, G_y^i - \overline{G}_y^i) \rangle \\ &= \sum_{i=1}^k (1+\theta_i)\alpha_i (\widetilde{y}^{i+1})^\top \left\{ \sum_{j=1}^p \mathcal{A}_j^* \widetilde{x}_j^{i+1} - b - \big(\sum_{j=1}^p \mathcal{A}_j^* \overline{x}_j^k - b \big) \right\} + \sum_{i=1}^k (1+\theta_i)\alpha_i \langle \widetilde{y}^{i+1}, G_y^i - \overline{G}_y^i) \rangle \\ &= \sum_{i=1}^k (1+\theta_i)\alpha_i \langle \widetilde{y}^{i+1}, \overline{G}_y^i - G_y^i \rangle + \sum_{i=1}^k (1+\theta_i)\alpha_i \langle \widetilde{y}^{i+1}, G_y^i - \overline{G}_y^i) \rangle = 0. \end{aligned}$$

By the definition of ϵ_k and combining the above equality with (86) and (87), it directly holds that

$$\bar{\epsilon}_k^{x_1} + \cdots + \bar{\epsilon}_k^{x_p} = \epsilon_k^x \leq \mathcal{O}(\frac{1}{k}).$$

Thus, **(iii)** has been established. At last, **(iv)** is directly derived according to Theorem 2 by setting $c_k = \underline{c} = 1$. As a consequence, the proof is completed. □

K. More Experiments

Actually, to make the subproblems of PADMM-EBB, PLADMM-PSAP (Liu et al., 2013; Lin et al., 2015), PGSADMM and M-GSJADMM (Lu et al., 2017) have closed-form solutions, we equivalently reformulate problem (29) as the following form by introducing two slack variables (H, F) to separate the sparsity and nonnegativity of (Z, G):

$$\begin{aligned} &\min\ \|H\|_* + \|F\|_* + \lambda\|E\|_1 + \frac{\mu}{2}\|Z\|_{L_Z}^2 + \frac{\gamma}{2}\|G\|_{L_G}^2 \\ &\text{s.t. } X = XZ + GX + E, Z \geq 0,\ G \geq 0, Z = H, G = F. \end{aligned} \tag{88}$$

In the implementation, we measure the performance of the four solvers of PADMM-EBB, PLADMM-PSAP (Liu et al., 2013; Lin et al., 2015), PGSADMM and M-GSJADMM (Lu et al., 2017) in terms of the proximal KKT residual defined as (25), objective value, and feasibility of (29) over iterations and runtime. Below, we report the performance on $X = \text{randn}(200, 200)$ and PIE_pose27 of PADMM-EBB, PLADMM-PSAP, PGSADMM and M-GSJADMM with new hyperparameters $(\lambda, \mu, \gamma) = (10^2, 10^4, 10^4)$. In addition, we conduct experiments on two extra real datasets (COIL20, YaleB_32x32)[2] with hyperparameters $(\lambda, \mu, \gamma) = (10^2, 10^4, 10^4)$ and $(\lambda, \mu, \gamma) = (10^3, 10^4, 10^4)$. In the implementation of PLADMM-PSAP, PGSADMM and M-GSJADMM, the penalty parameters β_k are all updated via the suggestions from (Lu et al., 2017), *i.e.*, $\beta_{k+1} = \min(\rho\beta_k, 1.0e10)$ where $\rho = 1.1$ and $\beta_0 = 1.0e-4$.

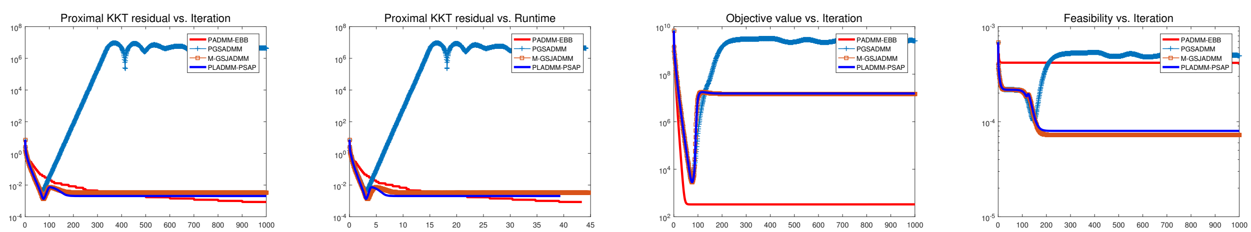

Figure 3. The above four figures illustrate the proximal KKT residual vs. iteration, proximal KKT residual vs. runtime, objective value vs. iteration, and feasibility vs. iteration on the synthetic dataset with parameters $(\lambda, \mu, \gamma) = (10^2, 10^4, 10^4)$, respectively.

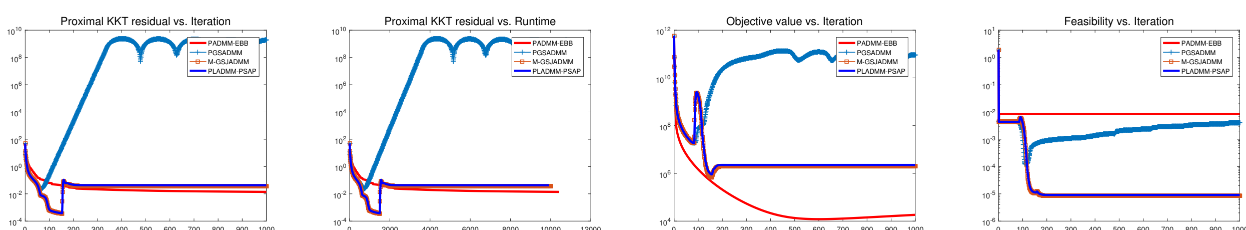

Figure 4. The above four figures illustrate the proximal KKT residual vs. iteration, proximal KKT residual vs. runtime, objective value vs. iteration, and feasibility vs. iteration on the real dataset PIE_pose27 with parameters $(\lambda, \mu, \gamma) = (10^2, 10^4, 10^4)$, respectively.

[2]http://dengcai.zjulearning.org:8081/Data/FaceDataPIE.html

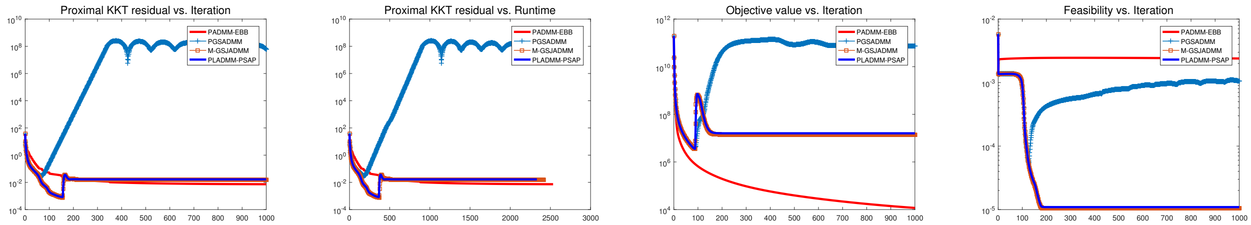

Figure 5. The above four figures illustrate the proximal KKT residual vs. iteration, proximal KKT residual vs. runtime, objective value vs. iteration, and feasibility vs. iteration on the real dataset COIL20 with parameters $(\lambda, \mu, \gamma) = (10^2, 10^4, 10^4)$, respectively.

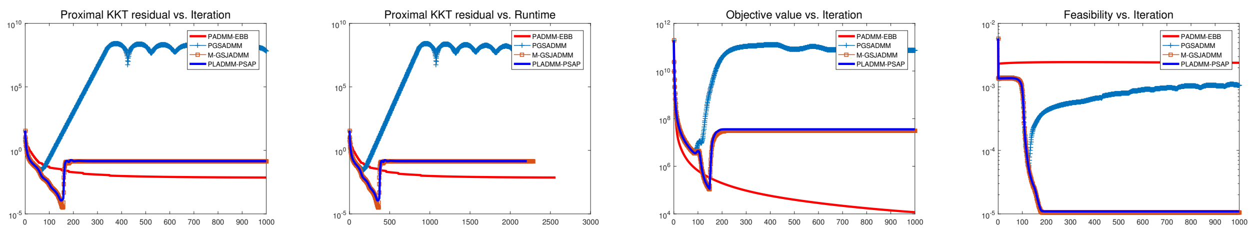

Figure 6. The above four figures illustrate the proximal KKT residual vs. iteration, proximal KKT residual vs. runtime, objective value vs. iteration, and feasibility vs. iteration on the real dataset COIL20 with parameters $(\lambda, \mu, \gamma) = (10^3, 10^4, 10^4)$, respectively.

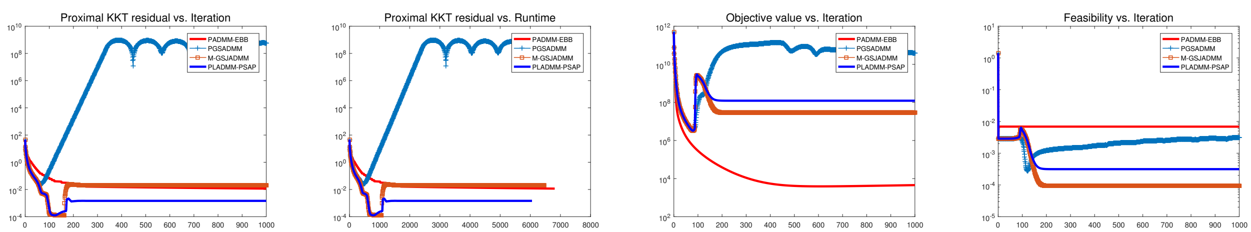

Figure 7. The above four figures illustrate the proximal KKT residual vs. iteration, proximal KKT residual vs. runtime, objective value vs. iteration, and feasibility vs. iteration on the real dataset YaleB_32x32 with parameters $(\lambda, \mu, \gamma) = (10^2, 10^4, 10^4)$, respectively.

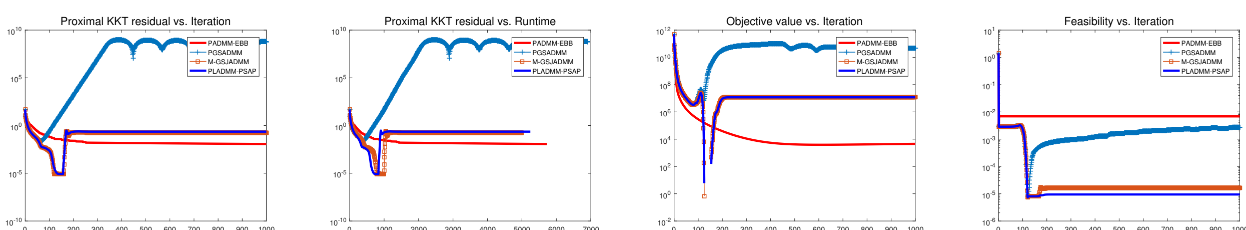

Figure 8. The above four figures illustrate the proximal KKT residual vs. iteration, proximal KKT residual vs. runtime, objective value vs. iteration, and feasibility vs. iteration on the real dataset YaleB_32x32 with parameters $(\lambda, \mu, \gamma) = (10^3, 10^4, 10^4)$, respectively.