A Preliminaries

There are several preliminaries we will use in the following section. The first one is a convergence result from [22, Lemma 2.2.2] of a special sequence which appears in B.2.

Lemma 2. *Let* $a_k \geq 0$ *and let*

$$\begin{aligned} &a_{k+1} \leq (1+\nu_k)a_k + \zeta_k, \quad \nu_k \geq 0,\ \zeta_k \geq 0, \\ &\sum_{k\in\mathbb{N}} \nu_k < \infty, \quad \sum_{k\in\mathbb{N}} \zeta_k < \infty. \end{aligned} \tag{22}$$

Then, $a_k \to A \geq 0$ *for some* $A < +\infty$*.*

The following identity is called (cosine rule), which proves to be very useful.

$$2\langle a-b, c-a\rangle = \|b-c\|^2 - \|a-b\|^2 - \|a-c\|^2 \quad \forall a,b,c \in X\,. \tag{23}$$

Another inequality appears many times in B.2 is the characteristic property of the proximal operator with respect to a symmetric positive definite matrix M:

$$\hat{x} = \operatorname{prox}_g^M(\bar{x}) \iff \langle \hat{x}-\bar{x}, y-\hat{x}\rangle_M \geq g(\bar{x}) - g(y) \quad \forall y \in X\,. \tag{24}$$

If $M = I$ is an identity matrix, then (24) is the characteristic property of the standard proximal operator. Assume $(\hat{x},\hat{y})$ is a saddle point which solves (1). Then we obtain

$$\begin{aligned} &P_{\hat{x},\hat{y}}(x) = g(x) + h(x) - g(\hat{x}) - h(\hat{x}) + \langle K^*\hat{y}, x-\hat{x}\rangle \geq 0 \quad \forall x \in X\,, \\ &D_{\hat{x},\hat{y}}(y) = f^*(y) - f^*(\hat{y}) - \langle K\hat{x}, y-\hat{y}\rangle \geq 0 \quad \forall y \in Y\,, \end{aligned} \tag{25}$$

where $P_{\hat{x},\hat{y}}(x)$ and $D_{\hat{x},\hat{y}}(y)$ are convex. Then $\mathcal{G}_{\hat{x},\hat{y}}(x,y) := P_{\hat{x},\hat{y}}(x) + D_{\hat{x},\hat{y}}(y)$ is the primal-dual gap. Without ambiguity, in the proofs, we may omit the subscript in P and D.

B Collection of Proofs

B.1 Proof of Lemma 1

It is a similar argument with the one in [20].

(i)&(ii) σ_k is decreased by $\mu \in (0,1)$ and the inequality (6) is satisfied as long as $\sigma_k < \underline{\sigma}_k := \frac{-1+\sqrt{(4\delta\alpha)/\beta_k+1}}{2\hat{L}}$ where $\hat{L} = \max\{L, L_K\}$. We introduce a notation $\underline{\sigma} := \frac{-1+\sqrt{(4\delta\alpha)/\beta+1}}{2\hat{L}}$. Since $\beta_k < \beta$, we have $\underline{\sigma}_k \geq \underline{\sigma}$. We argument by induction. We assume $\sigma_0 > \mu\underline{\sigma}_0$ and $\sigma_{k-1} > \mu\underline{\sigma}_{k-1}$. For the case $\sigma_k = \bar{\sigma}_k$, then $\sigma_k \geq (\frac{\beta_{k-1}}{\beta_k})\sigma_{k-1} > \mu(\frac{\beta_{k-1}}{\beta_k})\underline{\sigma}_{k-1} > \mu\underline{\sigma}_k > \mu\underline{\sigma}$. For the case $\sigma_k = \mu^i\bar{\sigma}_k$, $\sigma'_k = \mu^{i-1}\bar{\sigma}_k$ does not satisfy (6). It follows $\sigma'_k > \underline{\sigma}_k$. Thus, $\sigma_k = \mu\sigma'_k > \mu\underline{\sigma}_k \geq \mu\underline{\sigma}$.

(iii) By $\sigma_k \leq \sigma_{k-1}\sqrt{1+\theta_{k-1}}$, we get $\theta_k \leq \sqrt{1+\theta_{k-1}}$. Thus, θ_k is bounded from above. □

B.2 Proof of Theorem 1

The following proof is adapted from [20]. Assume $(\hat{x}, \hat{y})$ is a saddle point of problem 1 and $\beta_k \equiv \beta$. By using (24), we obtain the following two inequalities:

$$\left\langle y^{k+1} - y^k - \sigma_k K x^{k+1}, \hat{y} - y^{k+1} \right\rangle \geq \sigma_k (f^*(y^{k+1}) - f^*(\hat{y})) \tag{26}$$

$$\left\langle x^{k+1} - x^k + \tau_k M_k^{-1} K^* \bar{y}^k + \tau_k M_k^{-1} \nabla h(x^k), \hat{x} - x^{k+1} \right\rangle_{M_k} \geq \tau_k (g(x^{k+1}) - g(\hat{x})) \tag{27}$$

By using $\tau_k = \beta \sigma_k$

$$\begin{aligned} &\left\langle \frac{1}{\beta}(x^{k+1} - x^k) + \sigma_k M_k^{-1} K^* \bar{y}^k + \sigma_k M_k^{-1} \nabla h(x^k), \hat{x} - x^{k+1} \right\rangle_{M_k} \\ &\geq \sigma_k (g(x^{k+1}) - g(\hat{x})) \end{aligned} \tag{28}$$

Similarly, we apply (24) on y^k and obtain

$$\left\langle y^k - y^{k-1} - \sigma_{k-1} K x^k, y - y^k \right\rangle \geq \sigma_{k-1} (f^*(y^k) - f^*(y)) \quad \forall y \in Y\,. \tag{29}$$

Setting $y = y^{k+1}$ and $y = y^{k-1}$ respectively, we obtain

$$\left\langle y^k - y^{k-1} - \sigma_{k-1} K x^k, y^{k+1} - y^k \right\rangle \geq \sigma_{k-1} (f^*(y^k) - f^*(y^{k+1})) \quad \forall y \in Y\,, \tag{30}$$

$$\left\langle y^k - y^{k-1} - \sigma_{k-1} K x^k, y^{k-1} - y^k \right\rangle \geq \sigma_{k-1} (f^*(y^k) - f^*(y^{k-1})) \quad \forall y \in Y\,. \tag{31}$$

We deduce from (30) $\times\, \theta_k$ and $\theta_k = \frac{\sigma_k}{\sigma_{k-1}}$ that:

$$\left\langle \theta_k (y^k - y^{k-1}) - \sigma_k K x^k, y^{k+1} - y^k \right\rangle \geq \sigma_k (f^*(y^k) - f^*(y^{k+1}))\,. \tag{32}$$

By (31) $\times\, \theta_k^2$, we also get:

$$\left\langle \theta_k (y^k - y^{k-1}) - \sigma_k K x^k, \theta_k (y^{k-1} - y^k) \right\rangle \geq \sigma_k (\theta_k f^*(y^k) - \theta_k f^*(y^{k-1}))\,. \tag{33}$$

Summing (32) and (33) together, by using $\bar{y}^k = y^k + \theta_k (y^k - y^{k-1})$, we obtain

$$\left\langle \bar{y}^k - y^k - \sigma_k K x^k, y^{k+1} - \bar{y}^k \right\rangle \geq \sigma_k ((1 + \theta_k) f^*(y^k) - \theta_k f^*(y^{k-1}) - f^*(y^{k+1}))\,. \tag{34}$$

To sum up inequalties (26), (28) and (34) , we obtain

$$\begin{aligned} &\left\langle y^{k+1} - y^k - \sigma_k K x^{k+1}, \hat{y} - y^{k+1} \right\rangle \\ &+ \left\langle \frac{1}{\beta}(x^{k+1} - x^k) + \sigma_k M_k^{-1} K^* \bar{y}^k + \sigma_k M_k^{-1} \nabla h(x^k), \hat{x} - x^{k+1} \right\rangle_{M_k} \\ &+ \left\langle \bar{y}^k - y^k - \sigma_k K x^k, y^{k+1} - \bar{y}^k \right\rangle \\ &\geq \sigma_k (f^*(y^{k+1}) - f^*(\hat{y})) + \sigma_k (g(x^{k+1}) - g(\hat{x})) + \sigma_k ((1 + \theta_k) f^*(y^k) - \theta_k f^*(y^{k-1}) \\ &- f^*(y^{k+1}))\,, \end{aligned} \tag{35}$$

Reorganizing the above inequality and using $\tau_k = \beta\sigma_k$, we have

$$\begin{aligned}
&\left\langle y^{k+1} - y^k, \hat{y} - y^{k+1} \right\rangle + \frac{1}{\beta} \left\langle x^{k+1} - x^k, \hat{x} - x^{k+1} \right\rangle_{M_k} + \left\langle \bar{y}^k - y^k, y^{k+1} - \bar{y}^k \right\rangle \\
&+ \left\langle -\sigma_k K x^k, y^{k+1} - \bar{y}^k \right\rangle + \left\langle -\sigma_k K x^{k+1}, \hat{y} - y^{k+1} \right\rangle \\
&+ \left\langle \sigma_k K^* \bar{y}^k + \sigma_k \nabla h(x^k), \hat{x} - x^{k+1} \right\rangle \\
&\geq \sigma_k (g(x^{k+1}) - g(\hat{x})) + \sigma_k ((1+\theta_k) f^*(y^k) - \theta_k f^*(y^{k-1}) - f^*(\hat{y})) ,
\end{aligned} \tag{36}$$

As in [20], we still have:

$$\begin{aligned}
&\left\langle -\sigma_k K x^k, y^{k+1} - \bar{y}^k \right\rangle + \left\langle -\sigma_k K x^{k+1}, \hat{y} - y^{k+1} \right\rangle + \left\langle \sigma_k K^* \bar{y}^k, \hat{x} - x^{k+1} \right\rangle \\
&= \sigma_k \left\langle K x^k - K x^{k+1}, \bar{y}^k - y^{k+1} \right\rangle + \sigma_k \left\langle K\hat{x}, \bar{y}^k - \hat{y} \right\rangle - \sigma_k \left\langle K^* \hat{y}, x^{k+1} - \hat{x} \right\rangle
\end{aligned} \tag{37}$$

Adding $\sigma_k h(x^{k+1}) - \sigma_k h(\hat{x})$ on both sides of (36), we obtain:

$$\begin{aligned}
&\left\langle y^{k+1} - y^k, \hat{y} - y^{k+1} \right\rangle + \frac{1}{\beta} \left\langle x^{k+1} - x^k, \hat{x} - x^{k+1} \right\rangle_{M_k} + \left\langle \bar{y}^k - y^k, y^{k+1} - \bar{y}^k \right\rangle \\
&+ \left\langle -\sigma_k K x^k, y^{k+1} - \bar{y}^k \right\rangle + \left\langle -\sigma_k K x^{k+1}, \hat{y} - y^{k+1} \right\rangle \\
&+ \left\langle \sigma_k K^* \bar{y}^k + \sigma_k \nabla h(x^k), \hat{x} - x^{k+1} \right\rangle + \sigma_k h(x^{k+1}) - \sigma_k h(\hat{x}) \\
&\geq \sigma_k (g(x^{k+1}) - g(\hat{x}) + (1+\theta_k) f^*(y^k) - \theta_k f^*(y^{k-1}) - f^*(\hat{y}) + h(x^{k+1}) - h(\hat{x})) .
\end{aligned} \tag{38}$$

Combining (37) and (38), we have

$$\begin{aligned}
&\left\langle y^{k+1} - y^k, \hat{y} - y^{k+1} \right\rangle + \frac{1}{\beta} \left\langle x^{k+1} - x^k, \hat{x} - x^{k+1} \right\rangle_{M_k} + \left\langle \bar{y}^k - y^k, y^{k+1} - \bar{y}^k \right\rangle \\
&\sigma_k \left\langle K x^k - K x^{k+1}, \bar{y}^k - y^{k+1} \right\rangle + \sigma_k \left\langle K\hat{x}, \bar{y}^k - \hat{y} \right\rangle - \sigma_k \left\langle K^* \hat{y}, x^{k+1} - \hat{x} \right\rangle \\
&+ \left\langle \sigma_k \nabla h(x^k), \hat{x} - x^{k+1} \right\rangle + \sigma_k h(x^{k+1}) - \sigma_k h(\hat{x}) \\
&\geq \sigma_k (g(x^{k+1}) - g(\hat{x}) + (1+\theta_k) f^*(y^k) - \theta_k f^*(y^{k-1}) - f^*(\hat{y}) + h(x^{k+1}) - h(\hat{x})) .
\end{aligned} \tag{39}$$

By the definition of $D(y)$ (25) and $\bar{y}^k = y^k + \theta_k (y^k - y^{k-1})$, we have

$$\begin{aligned}
&(1+\theta_k) f^*(y^k) - \theta_k f^*(y^{k-1}) - f^*(\hat{y}) - \left\langle K\hat{x}, \bar{y}^k - \hat{y} \right\rangle \\
&= (1+\theta_k)(f^*(y^k) - f^*(\hat{y}) - \left\langle K\hat{x}, y^k - \hat{y} \right\rangle) - \theta_k (f^*(y^{k-1}) - f^*(\hat{y}) \\
&- \left\langle K\hat{x}, y^{k-1} - \hat{y} \right\rangle) \\
&= (1+\theta_k) D(y^k) - \theta_k D(y^{k-1}) .
\end{aligned} \tag{40}$$

Using (40) and the definition of $P(x)$, we deduce from (39) that

$$\begin{aligned}
&\left\langle y^{k+1}-y^k, \hat{y}-y^{k+1}\right\rangle+\frac{1}{\beta}\left\langle x^{k+1}-x^k, \hat{x}-x^{k+1}\right\rangle_{M_k}+\left\langle \bar{y}^k-y^k, y^{k+1}-\bar{y}^k\right\rangle \\
&+\sigma_k\left\langle Kx^k-Kx^{k+1}, \bar{y}^k-y^{k+1}\right\rangle+\left\langle \sigma_k \nabla h(x^k), \hat{x}-x^{k+1}\right\rangle \\
&+\sigma_k h(x^{k+1})-\sigma_k h(\hat{x}) \\
&\geq \sigma_k(P(x^{k+1})+(1+\theta_k)D(y^k)-\theta_k D(y^{k-1}))\,.
\end{aligned} \tag{41}$$

From the line search condition (6), we have

$$\begin{aligned}
&\sigma_k(h(x^{k+1})-h(x^k)-\left\langle \nabla h(x^k), x^{k+1}-x^k\right\rangle) \\
&\leq \frac{\delta}{2\beta}\|x^{k+1}-x^k\|_{M_k}^2-\frac{1}{2}\sigma_k^2\|Kx^{k+1}-Kx^k\|^2\,.
\end{aligned} \tag{42}$$

Additionally, by the convexity of $h(x)$, we also have

$$h(x^k)-h(\hat{x})+\left\langle \nabla h(x^k), \hat{x}-x^k\right\rangle \leq 0\,. \tag{43}$$

Combining (42) and $\sigma_k \times$ (43), we get

$$\begin{aligned}
&\sigma_k(h(x^{k+1})-h(\hat{x})-\left\langle \nabla h(x^k), x^{k+1}-\hat{x}\right\rangle) \\
&\leq \frac{\delta}{2\beta}\|x^{k+1}-x^k\|_{M_k}^2-\frac{1}{2}\sigma_k^2\|Kx^{k+1}-Kx^k\|^2\,.
\end{aligned} \tag{44}$$

Thus, it follows from (41) and (44) that

$$\begin{aligned}
&\left\langle y^{k+1}-y^k, \hat{y}-y^{k+1}\right\rangle+\frac{1}{\beta}\left\langle x^{k+1}-x^k, \hat{x}-x^{k+1}\right\rangle_{M_k}+\left\langle \bar{y}^k-y^k, y^{k+1}-\bar{y}^k\right\rangle \\
&+\sigma_k\left\langle Kx^k-Kx^{k+1}, \bar{y}^k-y^{k+1}\right\rangle+\frac{\delta}{2\beta}\|x^{k+1}-x^k\|_{M_k}^2-\frac{1}{2}\sigma_k^2\|Kx^{k+1}-Kx^k\|^2 \\
&\geq \sigma_k(P(x^{k+1})+(1+\theta_k)D(y^k)-\theta_k D(y^{k-1}))\,.
\end{aligned} \tag{45}$$

Using Cauchy-Schwarz inequality, we obtain

$$\begin{aligned}
&\left\langle y^{k+1}-y^k, \hat{y}-y^{k+1}\right\rangle+\frac{1}{\beta}\left\langle x^{k+1}-x^k, \hat{x}-x^{k+1}\right\rangle_{M_k}+\left\langle \bar{y}^k-y^k, y^{k+1}-\bar{y}^k\right\rangle \\
&+\frac{1}{2}\sigma_k^2\|Kx^k-Kx^{k+1}\|^2+\frac{1}{2}\|\bar{y}^k-y^{k+1}\|^2+\frac{\delta}{2\beta}\|x^{k+1}-x^k\|_{M_k}^2 \\
&-\frac{1}{2}\sigma_k^2\|Kx^{k+1}-Kx^k\|^2 \\
&\geq \sigma_k(P(x^{k+1})+(1+\theta_k)D(y^k)-\theta_k D(y^{k-1}))\,.
\end{aligned} \tag{46}$$

Applying (23), we deduce from (46) that

$$
\begin{aligned}
&(\frac{1}{2}\|y^k-\hat{y}\|^2-\frac{1}{2}\|y^{k+1}-y^k\|^2-\frac{1}{2}\|\hat{y}-y^{k+1}\|^2) \\
&+(\frac{1}{2\beta}\|x^k-\hat{x}\|^2_{M_k}-\frac{1}{2\beta}\|x^{k+1}-x^k\|^2_{M_k}-\frac{1}{2\beta}\|\hat{x}-x^{k+1}\|^2_{M_k}) \\
&+(\frac{1}{2}\|y^k-y^{k+1}\|^2-\frac{1}{2}\|\bar{y}^k-y^k\|^2-\frac{1}{2}\|y^{k+1}-\bar{y}^k\|^2) \\
&+\frac{1}{2}\|\bar{y}^k-y^{k+1}\|^2+\frac{\delta}{2\beta}\|x^{k+1}-x^k\|^2_{M_k} \\
&\geq \sigma_k(P(x^{k+1})+(1+\theta_k)D(y^k)-\theta_k D(y^{k-1}))\,.
\end{aligned}
\tag{47}
$$

Reorganizing the above inequalities, we obtain

$$
\begin{aligned}
&\frac{1}{2}\|y^k-\hat{y}\|^2+\frac{1}{2\beta}\|x^k-\hat{x}\|^2_{M_k}-\frac{1-\delta}{2\beta}\|x^{k+1}-x^k\|^2_{M_k} \\
&+\sigma_k\theta_k D(y^{k-1})-\frac{1}{2}\|\bar{y}^k-y^k\|^2 \\
&\geq \sigma_k(P(x^{k+1})+(1+\theta_k)D(y^k))+\frac{1}{2}\|\hat{y}-y^{k+1}\|^2+\frac{1}{2\beta}\|\hat{x}-x^{k+1}\|^2_{M_k}\,.
\end{aligned}
\tag{48}
$$

It follows from $\bar{\sigma}_k \leq \sqrt{1+\theta_{k-1}}\sigma_{k-1}$ that $\sigma_k\theta_k \leq \frac{\sigma_k^2}{\sigma_{k-1}} \leq \frac{\bar{\sigma}_k^2}{\sigma_{k-1}} \leq (1+\theta_{k-1})\sigma_{k-1}$. Thus,

$$
\begin{aligned}
&\frac{1}{2}\|y^k-\hat{y}\|^2+\frac{1}{2\beta}\|x^k-\hat{x}\|^2_{M_k}-\frac{1-\delta}{2\beta}\|x^{k+1}-x^k\|^2_{M_k} \\
&+\sigma_{k-1}(1+\theta_{k-1})D(y^{k-1})-\frac{1}{2}\|\bar{y}^k-y^k\|^2 \\
&\geq \sigma_k(1+\theta_k)D(y^k)+\frac{1}{2}\|\hat{y}-y^{k+1}\|^2+\frac{1}{2\beta}\|\hat{x}-x^{k+1}\|^2_{M_k}\,.
\end{aligned}
\tag{49}
$$

Since $(1+\eta_k)M_k \succeq M_{k+1}$, we can obtain the following key inequality:

$$
\begin{aligned}
&\frac{1}{2}\|y^k-\hat{y}\|^2+\frac{1}{2\beta}\|x^k-\hat{x}\|^2_{M_k}-\frac{1-\delta}{2\beta}\|x^{k+1}-x^k\|^2_{M_k} \\
&+\sigma_{k-1}(1+\theta_{k-1})D(y^{k-1})-\frac{1}{2}\|\bar{y}^k-y^k\|^2 \\
&\geq \sigma_k(1+\theta_k)D(y^k)+\frac{1}{2}\|\hat{y}-y^{k+1}\|^2+\frac{1}{2\beta(1+\eta_k)}\|\hat{x}-x^{k+1}\|^2_{M_{k+1}}\,.
\end{aligned}
\tag{50}
$$

Set $A_k \coloneqq \frac{1}{2}\|y^k-\hat{y}\|^2+\sigma_{k-1}(1+\theta_{k-1})D(y^{k-1})+\frac{1}{2\beta}\|x^k-\hat{x}\|^2_{M_k}$. Then, we deduce from (50) that

$$
A_{k+1} \leq (1+\eta_k)A_k\,. \tag{51}
$$

By Lemma 2, A_k is bounded from above by some constant C. Thus, $\|y^k-\hat{y}\|$ and $\|x^k-\hat{x}\|_{M_k}$ are both bounded. By the assumption that M_k is uniformly bounded, $\|x^k - \hat{x}\|$ is also bounded. As a result, we deduce from (50) that

$$\begin{aligned}\sum_k \big(\frac{1-\delta}{2\beta}\|x^{k+1} - x^k\|^2_{M_k} + \frac{1}{2}\|\bar{y}^k - y^k\|^2\big) &\leq \sum_k \big((1+\eta_k)A_k - A_{k+1}\big) \\ &\leq C\sum_k \eta_k + A_0 < +\infty\,.\end{aligned} \tag{52}$$

It implies that $\|x^{k+1} - x^k\|_{M_k} \to 0$ and $\|\bar{y}^k - y^k\| \to 0$. So does $\|x^{k+1} - x^k\| \to 0$, since $(M_k)_{k\in\mathbb{N}} \subset \mathcal{S}_\alpha(X)$. Since $\sigma_k > \sigma$ for some σ which is shown in Lemma 1 and $\beta > 0$ is fixed,

$$\begin{aligned}&\frac{y^{k+1} - y^k}{\sigma_k} = \frac{\bar{y}^{k+1} - y^{k+1}}{\sigma_{k+1}} \to 0 \quad \text{as}\, k \to +\infty\,, \\ &\frac{\|x^{k+1} - x^k\|^2_{M_k}}{\tau_k} \to 0 \quad \text{as}\, k \to +\infty\,.\end{aligned} \tag{53}$$

Since $(x^k, y^k)_{k\in\mathbb{N}}$ is bounded, we can extract a subsequence $(x^{k_i}, y^{k_i})_{i\in\mathbb{N}}$ converging to some cluster point (x^*, y^*). As in [20], similarly, by using the lower semi-continuity of functions g and f^* and the continuity of function h, we can pass the following two inequalities to the limit:

$$\begin{aligned}&\left\langle \frac{y^{k_i+1} - y^{k_i}}{\sigma_{k_i}} - Kx^{k_i+1}, y - y^{k_i+1}\right\rangle \geq (f^*(y^{k_i+1}) - f^*(y)) \quad \forall y \in Y\,, \\ &\left\langle \frac{x^{k_i+1} - x^{k_i}}{\tau_{k_i}} + M_{k_i}^{-1}K^*\bar{y}^{k_i} + M_{k_i}^{-1}\nabla h(x^{k_i}), x - x^{k_i+1}\right\rangle_{M_{k_i}} \\ &= \left\langle \frac{M_{k_i}(x^{k_i+1} - x^{k_i})}{\tau_{k_i}}, x - x^{k_i+1}\right\rangle + \left\langle K^*\bar{y}^{k_i} + \nabla h(x^{k_i}), x - x^{k_i+1}\right\rangle \\ &\geq (g(x^{k_i+1}) - g(x)) \quad \forall x \in X\,.\end{aligned} \tag{54}$$

Thus, (x^*, y^*) is the saddle point of (1). If, additionally, $f^*(y)|_{dom_{f^*}}$ is continuous, then $f^*(y^{k_i}) \to f^*(y^*)$ and $D(y^{k_i}) \to 0$ as $i \to +\infty$. From (50), we have $\frac{1}{\varPi_{j=1}^k(1+\eta_j)}A_k$ is monotone. Setting $\hat{x} = x^*$ and $\hat{y} = y^*$ in (50), by the boundedness of σ_k and θ_k, it follows that

$$\lim_{k\to\infty} \frac{A_k}{\varPi_{i=1}^\infty(1+\eta_i)} \leq \lim_{k\to\infty} \frac{A_k}{\varPi_{i=1}^k(1+\eta_i)} = \lim_{i\to\infty} \frac{A_{k_i}}{\varPi_{j=1}^{k_i}(1+\eta_j)} \leq \lim_{i\to\infty} A_{k_i} = 0 \tag{55}$$

Since $\varPi_{i=1}^\infty(1+\eta_i) < +\infty$, we have $\lim_{k\to+\infty} A_k \to 0$ which means $x^k \to x^*$ and $y^k \to y^*$ as $k \to +\infty$. $\square$

B.3 Proof of Theorem 2

We adapt the corresponding proof in [20]. Let $\epsilon_k \coloneqq \sigma_k\big(P(x^{k+1}) + (1+\theta_k)D(y^k) - \theta_k D(y^{k-1})\big)$. Then we obtain the following inequality from (47),

$$\frac{1}{2}\|y^k - \hat{y}\|^2 - \frac{1}{2}\|y^{k+1} - \hat{y}\|^2 + \frac{1}{2\beta}\|x^k - \hat{x}\|^2_{M_k} - \frac{1}{2\beta}\|x^{k+1} - \hat{x}\|^2_{M_k} - \frac{1}{2}\|\bar{y}^k - y^k\|^2 \geq \epsilon_k\,. \tag{56}$$

By the assumption 1, we get

$$\frac{1}{2}\|y^k-\hat{y}\|^2-\frac{1}{2}\|y^{k+1}-\hat{y}\|^2+\frac{1}{2\beta}\|x^k-\hat{x}\|^2_{M_k}-\frac{1}{2\beta}\frac{\|x^{k+1}-\hat{x}\|^2_{M_{k+1}}}{(1+\eta_k)}-\frac{1}{2}\|\bar{y}^k-y^k\|^2 \geq \epsilon_k\,. \tag{57}$$

Since $(1+\eta_k)\geq 1$, it follows

$$\frac{1}{2}\|y^k-\hat{y}\|^2-\frac{1}{2}\frac{\|y^{k+1}-\hat{y}\|^2}{(1+\eta_k)}+\frac{1}{2\beta}\|x^k-\hat{x}\|^2_{M_k}-\frac{1}{2\beta}\frac{\|x^{k+1}-\hat{x}\|^2_{M_{k+1}}}{(1+\eta_k)} \geq \epsilon_k\,. \tag{58}$$

Let both sides of the above inequality be divided by $\Pi_{i=1}^{k-1}(1+\eta_i)$ and it is common to assume that an empty product yields identity i.e. $\Pi_{i=1}^{0}(1+\eta_i)=1$. Thus,

$$\begin{aligned}&\frac{1}{2}\frac{\|y^k-\hat{y}\|^2}{\Pi_{i=1}^{k-1}(1+\eta_i)}-\frac{1}{2}\frac{\|y^{k+1}-\hat{y}\|^2}{\Pi_{i=1}^{k}(1+\eta_i)}+\frac{1}{2\beta}\frac{\|x^k-\hat{x}\|^2_{M_k}}{\Pi_{i=1}^{k-1}(1+\eta_i)}-\frac{1}{2\beta}\frac{\|x^{k+1}-\hat{x}\|^2_{M_{k+1}}}{\Pi_{i=1}^{k}(1+\eta_i)}\\ &\geq \frac{\epsilon_k}{\Pi_{i=1}^{k}(1+\eta_i)}\,.\end{aligned} \tag{59}$$

Summing up (59) for $k=1,\cdots,N$, we obtain

$$\frac{1}{2}\|y^1-\hat{y}\|^2+\frac{1}{2\beta}\|x^1-\hat{x}\|^2_{M_1}\geq \sum_{k=1}^{N}\frac{\epsilon_k}{\Pi_{i=1}^{k}(1+\eta_i)}\geq \sum_{k=1}^{N}\frac{\epsilon_k}{C}\,. \tag{60}$$

Here, we used the $C=\sum_{k\in\mathbb{N}}(1+\eta_k)<+\infty$.

The following steps are similar with the ones in [20].

$$\begin{aligned}\sum_{k=1}^{N}\epsilon_k =&\sigma_N(1+\theta_N)D(y^k)+\sum_{k=2}^{N}[(1+\theta_{k-1})\sigma_{k-1}-\theta_k\sigma_k]D(y^{k-1})\\ &-\theta_1\sigma_1 D(y^0)+\sum_{k=1}^{N}\sigma_k P(x^{k+1})\,.\end{aligned} \tag{61}$$

Since D is convex,

$$\begin{aligned}\sigma_N(1+\theta_N)D(y^N)+&\sum_{k=2}^{N}[(1+\theta_{k-1})\sigma_{k-1}-\theta_k\sigma_k]D(y^{k-1})\\ &\geq(\sigma_1\theta_1+s_N)D(\frac{\sigma_1(1+\theta_1)y^1+\sum_{k=2}^{N}\sigma_k\bar{y}^k}{\sigma_1\theta_1+s_N})\\ &=(\sigma_k\theta_1+s_N)D(\frac{\sigma_1\theta_1 y^0+\sum_{k=1}^{N}\sigma_k\bar{y}^k}{\sigma_1\theta_1+s_N})\\ &\geq s_N D(\bar{Y}^N)\,,\end{aligned} \tag{62}$$

where $s_N = \sum_{k=1}^N \sigma_k$. Similarly,

$$\sum_{k=1}^{N} \sigma_k P(x^{k+1}) \geq s_N P(\frac{\sum_{k=1}^N \sigma_k x^{k+1}}{s_N}) = s_N P(\bar{X}^N)\,. \tag{63}$$

As a result,

$$\mathcal{G}(\bar{X}^N, \bar{Y}^N) = P(\bar{X}^N) + D(\bar{Y}^N) \leq \frac{C}{s_N}\big(\frac{1}{2\beta}\|x^1 - \hat{x}\|^2_{M_1} + \frac{1}{2}\|y^1 - \hat{y}\|^2 + \sigma_1\theta_1 D(y^0)\big)\,. \tag{64}$$

□

B.4 Proof of Theorem 3

The proof is also adapted from [20]. From the update formula of β_k, it follows that β_k is decreasing. First, we are going to prove that θ_k is bounded from above. It is not difficult but tedious. We know that if there exists a $C \in \mathbb{R}_+$ s.t $\theta_k \leq C\sqrt{1+\theta_{k-1}}$ then θ_k is bounded. From this, it is sufficient to prove that $\frac{\beta_{k-1}}{\beta_k}$ is uniformly bounded from above by some C_θ. According to

$$\beta_k = \frac{\beta_{k-1}}{\min\{1 + \frac{\gamma}{C_M}\beta_{k-1}\sigma_{k-1}, C_\theta\}}\,, \ \forall k \in \mathbb{N}\,, \quad \text{and} \quad \beta_0 > 0\,, \tag{65}$$

we have that $\frac{\beta_{k-1}}{\beta_k} = \min\{1 + \frac{\gamma}{C_M}\beta_{k-1}\sigma_{k-1}, C_\theta\} \leq C_\theta$.

Second part, we are going to show the convergence rate. Since g is strongly convex, we obtain:

$$\begin{aligned} &\left\langle \frac{x^{k+1} - x^k}{\tau_k} + M_k^{-1}K^*\bar{y}^k + M_k^{-1}\nabla h(x^k), \hat{x} - x^{k+1} \right\rangle_{M_k} \\ &\geq (g(x^{k+1}) - g(\hat{x})) + \frac{\gamma}{2}\|x^{k+1} - \hat{x}\|^2\,. \end{aligned} \tag{66}$$

From Assumption 1, it follows that for any $k \in \mathbb{N}$,

$$\frac{\gamma}{2}\|x^{k+1} - \hat{x}\|^2 \geq \frac{\gamma}{2C_M}\|x^{k+1} - \hat{x}\|^2_{M_{k+1}}\,. \tag{67}$$

Following the same way in which we got equation (48), by equation (66) and the assumption that $(1+\eta_k)M_k \succeq M_{k+1}$, we obtain

$$\begin{aligned} &\frac{1}{2}\|y^k - \hat{y}\|^2 - \frac{1}{2}\|y^{k+1} - \hat{y}\|^2 + \frac{1}{2\beta_k}\|x^k - \hat{x}\|^2_{M_k} - \frac{1-\delta}{2\beta_k}\|x^{k+1} - x^k\|^2_{M_k} \\ &\qquad - \frac{1}{2\beta_k}\frac{\|x^{k+1} - \hat{x}\|^2_{M_{k+1}}}{(1+\eta_k)} - \frac{1}{2}\|\bar{y}^k - y^k\|^2 \geq \epsilon_k + \frac{\gamma\sigma_k}{2}\|x^{k+1} - \hat{x}\|^2\,. \end{aligned} \tag{68}$$

In order to obtain the following inequality, it is sufficient to assume $\delta \leq 1$. Thus,

$$\frac{1}{2}\|y^k - \hat{y}\|^2 - \frac{1}{2}\|y^{k+1} - \hat{y}\|^2 + \frac{1}{2\beta_k}\|x^k - \hat{x}\|^2_{M_k} - \frac{1}{2\beta_k}\frac{\|x^{k+1} - \hat{x}\|^2_{M_{k+1}}}{(1+\eta_k)} - \frac{1}{2}\|\bar{y}^k - y^k\|^2 \geq \epsilon_k + \frac{\gamma\sigma_k}{2}\|x^{k+1} - \hat{x}\|^2 . \tag{69}$$

Since $\delta \leq 1$, by dividing the above inequality with σ_k, we have

$$\frac{1}{2\sigma_k}\|y^k - \hat{y}\|^2 - \frac{1}{2\sigma_k}\|y^{k+1} - \hat{y}\|^2 + \frac{1}{2\tau_k}\|x^k - \hat{x}\|^2_{M_k} - \frac{1}{2\tau_k}\frac{\|x^{k+1} - \hat{x}\|^2_{M_{k+1}}}{(1+\eta_k)} - \frac{1}{2\sigma_k}\|\bar{y}^k - y^k\|^2 \geq \frac{\epsilon_k}{\sigma_k} + \frac{\gamma}{2}\|x^{k+1} - \hat{x}\|^2 , \tag{70}$$

where, we used $\tau_k = \beta_k\sigma_k$. By using (67), from the above inequality, we obtain that

$$\frac{1}{2\sigma_k}\|y^k - \hat{y}\|^2 - \frac{1}{2\sigma_k}\|y^{k+1} - \hat{y}\|^2 + \frac{1}{2\tau_k}\|x^k - \hat{x}\|^2_{M_k} - \frac{1}{2\tau_k}\frac{\|x^{k+1} - \hat{x}\|^2_{M_{k+1}}}{(1+\eta_k)} - \frac{1}{2\sigma_k}\|\bar{y}^k - y^k\|^2 \geq \frac{\epsilon_k}{\sigma_k} + \frac{\gamma}{2C_M}\|x^{k+1} - \hat{x}\|^2_{M_{k+1}} . \tag{71}$$

It follows from the above inequality that

$$\begin{aligned}
\frac{1}{2\sigma_k}\|y^k - \hat{y}\|^2 - \frac{1}{2\sigma_k}\|y^{k+1} - \hat{y}\|^2 + \frac{1}{2\tau_k}\|x^k - \hat{x}\|^2_{M_k} - \frac{1}{2\sigma_k}\|\bar{y}^k - y^k\|^2 \\
\geq \frac{\epsilon_k}{\sigma_k} + \frac{1 + (1+\eta_k)\tau_k\gamma/C_M}{2\tau_k(1+\eta_k)}\|x^{k+1} - \hat{x}\|^2_{M_{k+1}} , \\
\frac{1}{2\sigma_k}\|y^k - \hat{y}\|^2 - \frac{1}{2\sigma_k}\|y^{k+1} - \hat{y}\|^2 + \frac{1}{2\tau_k}\|x^k - \hat{x}\|^2_{M_k} - \frac{1}{2\sigma_k}\|\bar{y}^k - y^k\|^2 \\
\geq \frac{\epsilon_k}{\sigma_k} + \frac{\tau_{k+1}(1 + \tau_k\gamma/C_M)}{\tau_k}\frac{\|x^{k+1} - \hat{x}\|^2_{M_{k+1}}}{2\tau_{k+1}(1+\eta_k)} ,
\end{aligned} \tag{72}$$

For convenience, we set $\tilde{\gamma} = \gamma/C_M$. From the update step of β_k, it follows that

$$\frac{\tau_{k+1}(1+\tilde{\gamma}\tau_k)}{\tau_k} \geq \frac{\tau_{k+1}\min\{C_\theta, (1+\tilde{\gamma}\tau_k)\}}{\tau_k} = \frac{\sigma_{k+1}}{\sigma_k} \tag{73}$$

Set $B_k \coloneqq \frac{1}{2\tau_k}\|x^k - \hat{x}\|^2_{M_k} + \frac{1}{2\sigma_k}\|y^k - \hat{y}\|^2$ and $\tilde{B}_k \coloneqq \frac{B_k}{\Pi_{i=1}^{k-1}(1+\eta_i)}$. From (72), we have:

$$\frac{\sigma_{k+1}}{\sigma_k(1+\eta_k)}B_{k+1} + \frac{\epsilon_k}{\sigma_k} \leq B_k - \frac{1}{2\sigma_k}\|\bar{y}^k - y^k\|^2 \tag{74}$$

By dividing the above inequality by $\Pi_{i=1}^{k-1}(1+\eta_i) \geq 1$, we obtain

$$\frac{\sigma_{k+1}}{\sigma_k}\tilde{B}_{k+1} + \frac{\epsilon_k}{\sigma_k \Pi_{i=1}^{k-1}(1+\eta_i)} \leq \tilde{B}_k - \frac{1}{2\sigma_k \Pi_{i=1}^{k-1}(1+\eta_i)}\|\bar{y}^k - y^k\|^2 \tag{75}$$

By multiplying σ_k on both sides, we have

$$\sigma_{k+1}\tilde{B}_{k+1} + \frac{\epsilon_k}{\Pi_{i=1}^{k-1}(1+\eta_i)} \leq \sigma_k\tilde{B}_k - \frac{1}{2\Pi_{i=1}^{k-1}(1+\eta_i)}\|\bar{y}^k - y^k\|^2 \,. \tag{76}$$

By Assumption 1, $C = \Pi_{i\in\mathbb{N}}(1+\eta_i) < +\infty$, we have

$$\sigma_{k+1}\tilde{B}_{k+1} + \frac{\epsilon_k}{C} \leq \sigma_k\tilde{B}_k - \frac{1}{2C}\|\bar{y}^k - y^k\|^2 \,. \tag{77}$$

Summing up (77) from $k = 1, \cdots, N$, we obtain

$$\sigma_{N+1}\tilde{B}_{N+1} + \sum_{k=1}^{N}\frac{\epsilon_k}{C} \leq \sigma_1\tilde{B}_1 - \frac{1}{2C}\sum_{k=1}^{N}\|\bar{y}^k - y^k\|^2 \,. \tag{78}$$

Since σ_k is bounded by some σ for any $k \in \mathbb{N}$, $\tilde{B}_k$ is bounded from above. Since $C = \Pi_{i\in\mathbb{N}}(1+\eta_i) < +\infty$, B_k is also bounded from above. So, y^k is also bounded with $\lim_{k\to\infty}\|\bar{y}^k - y^k\|^2 = 0$. Thus, using the similar argument and notations in the proof B.2, we retrieve the same key inequality as the one in [20]:

$$\begin{aligned} \mathcal{G}(\bar{X}^N, \bar{Y}^N) &\leq \frac{C}{s_N}(\sigma_1 B_1 + \theta_1\sigma_1 P(x^0)), \\ \|x^{N+1} - \hat{x}\|_{M_{N+1}}^2 &\leq \frac{C\tau_{N+1}}{\sigma_{N+1}}(\sigma_1 A_1 + \theta_1\tau_1 P(x^0)) = C\beta_{N+1}\,, \end{aligned} \tag{79}$$

Using the same argument from [20], we know from B.1 that σ_k is bounded by $\mu\underline{\sigma}_k = \mu(\frac{-1+\sqrt{(4\delta\alpha)/\beta_k+1}}{2\hat{L}})$ where $\hat{L} = \max\{L, L_K\}$. We claim that there exists a constant C_β such that, $\beta_k = C_\beta(1/k^2)$.

i If $\alpha\delta/(\beta_k) \leq 1$, by $\sigma_k \geq \mu\underline{\sigma}_k \geq \mu\underline{\sigma}$, we have

$$\beta_{k+1} = \frac{\beta_k}{\min\{C_\theta, 1+\tilde{\gamma}\beta_k\sigma_k\}} \leq \frac{\beta_k}{\min\{C_\theta, 1+\mu\underline{\sigma}\delta\alpha\tilde{\gamma}\}} \,. \tag{80}$$

In this case, β_k decreases linearly. Thus, $\beta_{k+1} \leq C_\beta/(k+1)^2$ for k sufficiently large.

ii If $\alpha\delta/(\beta_k) \geq 1$, then $\sigma_k > \mu\underline{\sigma}_k > \frac{\mu}{2\hat{L}}\sqrt{\frac{\delta\alpha}{\beta_k}}$. Therefore, for k large enough, we have

$$\beta_{k+1} = \frac{\beta_k}{\min\{C_\theta, 1+\tilde{\gamma}\beta_k\sigma_k\}} \leq \frac{\beta_k}{\min\{C_\theta, 1+\frac{\mu\sqrt{\delta\alpha}\tilde{\gamma}}{2\hat{L}}\sqrt{\beta_k}\}} = \frac{\beta_k}{1+\frac{\mu\sqrt{\delta\alpha}\tilde{\gamma}}{2\hat{L}}\sqrt{\beta_k}} \,. \tag{81}$$

In this case, by induction $\beta_k \leq \frac{C_\beta}{k^2}$ for some constant $C_\beta > 0$.

From $\sigma_k > \mu\underline{\sigma}_k > \mu\underline{\sigma}$, we have $s_N = \sum_{k=1}^N \sigma_k > \sum_{k=1}^N \underline{\sigma}_k > \sum_{k=1}^N O(k) \sim N^2$ since $\beta_k \leq C_\beta/k^2$ for k sufficiently large. Then, we conclude the results. □