Supplementary Material for: Optimal Minimization of the Sum of Three Convex Functions with a Linear Operator

Seyoon Ko
Seoul National University

Joong-Ho Won
Seoul National University

Abstract

This document contains supplementary details for the paper "Optimal Minimization of the Sum of Three Convex Functions with a Linear Operator." All section, equation, table, and figure numbers in this supplementary document are preceded by a capital roman alphabet A, B, C, ... All section, equation, table, and figure numbers without an alphabet prefix refer to the main paper.

A Proof of Proposition 1

Proof of Proposition 1. By the convexity of f and L_f-Lipschitz smoothness of ∇f,

$$\rho_k^{-1} f(x^{k+1}) \leq \rho_k^{-1} f(x_{md}^k) + \rho_k^{-1}\langle \nabla f(x_{md}^k), x^{k+1} - x_{md}^k\rangle + \frac{\rho_k^{-1} L_f}{2}\|x^{k+1} - x_{md}^k\|_2^2.$$

From equation (6b) in Algorithm 1, $x^{k+1} - x_{md}^k = \rho_k(\tilde{x}^{k+1} - \tilde{x}^k)$. Thus,

$$\begin{aligned}
\rho_k^{-1} f(x^{k+1}) &\leq \rho_k^{-1} f(x_{md}^k) + \rho_k^{-1}\langle \nabla f(x_{md}^k), x^{k+1} - x_{md}^k\rangle + \frac{\rho_k L_f}{2}\|\tilde{x}^{k+1} - \tilde{x}^k\|_2^2 \\
&\overset{(6f)}{=} \rho_k^{-1} f(x_{md}^k) + (\rho_k^{-1} - 1)\langle \nabla f(x_{md}^k), x^k - x_{md}^k\rangle + \langle \nabla f(x_{md}^k), \tilde{x}^{k+1} - x_{md}^k\rangle + \frac{\rho_k L_f}{2}\|\tilde{x}^{k+1} - \tilde{x}^k\|_2^2 \\
&= (\rho_k^{-1} - 1)[f(x_{md}^k) + \langle \nabla f(x_{md}^k), x^k - x_{md}^k\rangle] + [f(x_{md}^k) + \langle f(x_{md}^k), \tilde{x}^{k+1} - x_{md}^k\rangle] + \frac{\rho_k L_f}{2}\|\tilde{x}^{k+1} - \tilde{x}^k\|_2^2 \\
&= (\rho_k^{-1} - 1)[f(x_{md}^k) + \langle \nabla f(x_{md}^k), x^k - x_{md}^k\rangle] + [f(x_{md}^k) + \langle f(x_{md}^k), x - x_{md}^k\rangle] \\
&\quad + \langle f(x_{md}^k), \tilde{x}^{k+1} - x\rangle + \frac{\rho_k L_f}{2}\|\tilde{x}^{k+1} - \tilde{x}^k\|_2^2 \\
&\leq (\rho_k^{-1} - 1) f(x^k) + f(x) + \langle \nabla f(x_{md}^k), \tilde{x}^{k+1} - x\rangle + \frac{\rho_k L_f}{2}\|\tilde{x}^{k+1} - \tilde{x}^k\|_2^2,
\end{aligned} \tag{A.1}$$

where the last inequality again uses the convexity of f.

Now using the convexity of g and iteration (6f) in Algorithm 1, we have $g(x^{k+1}) \leq (1 - \rho_k) g(x^k) + \rho_k g(\tilde{x}^{k+1})$. Thus,

$$\rho_k^{-1}[g(x^{k+1}) - g(x)] \leq (\rho_k^{-1} - 1)[g(x^k) - g(x)] + [g(x^{k+1}) - g(x)]. \tag{A.2}$$

Likewise, by the convexity of h^* and iteration (6g) in Algorithm 1,

$$\rho_k^{-1}[h^*(x^{k+1}) - h^*(x)] \leq (\rho_k^{-1} - 1)[h^*(x^k) - h^*(x)] + [h^*(x^{k+1}) - h^*(x)]. \tag{A.3}$$

Combining inequalities (A.1), (A.2), and (A.3), it follows that

$$\begin{aligned}
&\rho_k^{-1} \mathcal{G}(z^{k+1}, z) - (\rho_k^{-1} - 1)\mathcal{G}(z^k, z) \\
&\qquad = \rho_k^{-1}\left\{[f(x^{k+1}) + g(x^{k+1}) + \langle K x^{k+1}, y\rangle - h^*(y)] \right. \\
&\qquad\quad \left. -[f(x) + g(x) - \langle Kx, y^{k+1}\rangle - h^*(y^{k+1})]\right\} \\
&\qquad\quad + (\rho_k^{-1} - 1)\left\{[f(x^k) + g(x^k) + \langle K x^k, y\rangle - h^*(y)]\right.
\end{aligned}$$

$$
\begin{aligned}
&-[f(x) + g(x) - \langle Kx, y^k\rangle - h^*(y^k)]\} \\
&= \rho_k^{-1} f(x^{k+1}) - (\rho_k^{-1} - 1) f(x^k) - f(x) \\
&\quad + \rho_k^{-1}[g(x^{k+1}) - g(x)] - (\rho_k^{-1} - 1)[g(x^k) - g(x)] \\
&\quad + \rho_k^{-1}[h^*(y^{k+1}) - h^*(y)] - (\rho_k^{-1} - 1)[h^*(y^k) - h^*(y)] \\
&\quad + \langle K[\rho_k^{-1} x^{k+1} - (\rho_k^{-1} - 1)x^k], y\rangle - \langle Kx, \rho_k^{-1} y^{k+1} - (\rho_k^{-1} - 1) y^k\rangle \\
&\le f(x) + \langle \nabla f(x^k_{md}, \tilde{x}^{k+1} - x\rangle + \frac{\rho_k L_f}{2}\|x^{k+1} - x^k\|_2^2 \\
&\quad + g(\tilde{x}^{k+1}) - g(x) + h^*(\tilde{y}^{k+1}) - h(y) \\
&\quad + \langle K[\rho_k^{-1} x^{k+1} - (\rho_k^{-1} - 1)x^k], y\rangle - \langle Kx, \rho_k^{-1} y^{k+1} - (\rho_k^{-1} - 1) y^k\rangle \\
&\le f(x) + \langle \nabla f(x^k_{md}, \tilde{x}^{k+1} - x\rangle + \frac{\rho_k L_f}{2}\|x^{k+1} - x^k\|_2^2 \\
&\quad + g(\tilde{x}^{k+1}) - g(x) + h^*(\tilde{y}^{k+1}) - h(y) + \langle Kx^{k+1}, y\rangle - \langle Kx, y^{k+1}\rangle.
\end{aligned}
$$

□

B Optimal deterministic acceleration

In this section, we provide proofs that Algorithm 1 achieves the theoretically optimal rate of convergence in the deterministic settings. We first consider the case in which $\mathcal{X}$ and $\mathcal{Y}$ are both bounded. We then proceed to the unbounded case. Proofs of all the technical lemmas required in the main proofs are deferred in Section D.

B.1 Proof of Theorem 1

To prove Theorem 1, we need the following lemma.

Lemma B.1. *If $z^k = (x^k, y^k)$ is obtained by (6), we have the following under the condition (15):*

$$
\begin{aligned}
\rho_k^{-1}\gamma_k \mathcal{G}(z^{k+1}, z) &\le \mathcal{D}_k(z, \tilde{z}^{[k]}) - \gamma_k\langle \tilde{x}^{k+1} - x, B^T(\tilde{y}^{k+1} - \tilde{y}^k)\rangle - \gamma_k\langle K(\tilde{x}^{k+1} - \tilde{x}^k), \tilde{y}^{k+1} - y\rangle \\
&\quad - \gamma_k\left(\frac{1-q}{2\tau_k} - \frac{L_f\rho_k}{2}\right)\|\tilde{x}^{k+1} - \tilde{x}^k\|_2^2 - \gamma_k\frac{1-r}{2\sigma_k}\|\tilde{y}^{k+1} - \tilde{y}^k\|_2^2,
\end{aligned}
\tag{B.1}
$$

where

$$
\gamma_k = \theta_k^{-1}\gamma_{k-1}, \quad \gamma_1 = 1,
\tag{B.2}
$$

and $\mathcal{D}_k(z, \tilde{z}^{[k]})$ is defined by

$$
\mathcal{D}_k(z, \tilde{z}^{[k]}) := \sum_{i=1}^{k}\left[\frac{\gamma_i}{2\tau_i}(\|x - \tilde{x}^i\|_2^2 - \|x - \tilde{x}^{i+1}\|_2^2) + \frac{\gamma_i}{2\sigma_i}(\|y - \tilde{y}^i\|_2^2 - \|y - \tilde{y}^{i+1}\|_2^2)\right].
\tag{B.3}
$$

Proof of Theorem 1. First we find an upper bound of $\mathcal{D}_k(z, \tilde{z}^{[k]})$.

$$
\begin{aligned}
\mathcal{D}_k(z, \tilde{z}^{[k]}) &= \frac{\gamma_1}{2\tau_1}\|x - \tilde{x}^1\|_2^2 - \sum_{i=1}^{k-1}\frac{1}{2}\left(\frac{\gamma_i}{\tau_i} - \frac{\gamma_{i+1}}{\tau_{i+1}}\right)\|x - \tilde{x}^{i+1}\|_2^2 - \frac{\gamma_k}{2\tau_k}\|x - \tilde{x}^{k+1}\|_2^2 \\
&\quad + \frac{\gamma_1}{2\sigma_1}\|y - \tilde{y}^1\|_2^2 - \sum_{i=1}^{k-1}\frac{1}{2}\left(\frac{\gamma_i}{\sigma_i} - \frac{\gamma_{i+1}}{\sigma_{i+1}}\right)\|y - \tilde{y}^{i+1}\|_2^2 - \frac{\gamma_k}{2\sigma_k}\|y - \tilde{y}^{k+1}\|_2^2 \\
&\le \frac{\gamma_1}{\tau_1}\Omega_X^2 - \sum_{i=1}^{k-1}\left(\frac{\gamma_i}{\tau_i} - \frac{\gamma_{i+1}}{\tau_{i+1}}\right)\Omega_X^2 - \frac{\gamma_k}{2\tau_k}\|x - \tilde{x}^{k+1}\|_2^2 \\
&\quad + \frac{\gamma_1}{\sigma_1}\Omega_Y^2 - \sum_{i=1}^{k-1}\left(\frac{\gamma_i}{\sigma_i} - \frac{\gamma_{i+1}}{\sigma_{i+1}}\right)\Omega_Y^2 - \frac{\gamma_k}{2\sigma_k}\|y - \tilde{y}^{k+1}\|_2^2
\end{aligned}
$$

$$= \frac{\gamma_k}{\tau_k}\Omega_X^2 + \frac{\gamma_k}{\sigma_k}\Omega_Y^2 - \gamma_k\left(\frac{1}{2\tau_k}\|x - \tilde{x}^{k+1}\|_2^2 + \frac{1}{2\sigma_k}\|y - \tilde{y}^{k+1}\|_2^2\right), \tag{B.4}$$

where we used (14) for the inequality.

Consider the following upper bounds of the three inner product terms in (B.1):

$$\begin{aligned} |\gamma_k\langle \tilde{x}^{k+1} - x, B^T(\tilde{y}^{k+1} - \tilde{y}^k)\rangle| &\le \frac{\gamma_k q}{2\tau_k}\|\tilde{x}^{k+1} - x\|_2^2 + \frac{\|B\|_2^2\gamma_k\tau_k}{2q}\|\tilde{y}^{k+1} - \tilde{y}^k\|_2^2 \\ |\gamma_k\langle \tilde{x}^{k+1} - \tilde{x}^k, K^T(\tilde{y}^{k+1} - y)\rangle| &\le \frac{L_K^2\gamma_k\sigma_k}{2r}\|\tilde{x}^{k+1} - \tilde{x}^k\|_2^2 + \frac{\gamma_k r}{2\sigma_k}\|\tilde{y}^{k+1} - y\|_2^2. \end{aligned} \tag{B.5}$$

Then (15a), Lemma B.1, (B.4), and (B.5) imply that

$$\begin{aligned} \rho_k^{-1}\gamma_k\mathcal{G}(z^{k+1}, z) \le{}& \frac{\gamma_k}{\tau_k}\Omega_X^2 + \frac{\gamma_k}{\sigma_k}\Omega_Y^2 - \gamma_k\frac{1-q}{2\tau_k}\|x - \tilde{x}^{k+1}\|_2^2 \\ & - \gamma_k\left(\frac{1-r}{2\sigma_k}\right)\|y - \tilde{y}^{k+1}\|_2^2 \\ & - \gamma_k\left(\frac{1-q}{2\tau_k} - \frac{L_f\rho_k}{2} - \frac{L_K^2\sigma_k}{2}\right)\|\tilde{x}^{k+1} - \tilde{x}^k\|_2^2 \\ & - \gamma_k\left(\frac{1-r}{2\sigma_k} - \frac{\tau_k}{2}\|B\|_2^2\right)\|\tilde{y}^{k+1} - \tilde{y}^k\|_2^2 \\ \le{}& \frac{\gamma_k}{\tau_k}\Omega_X^2 + \frac{\gamma_k}{\sigma_k}\Omega_Y^2. \end{aligned}$$

That is, (17). □

B.2 Proof of Corollary 1

Proof of Corollary 1. First check (7), (9), and (18) satisfy (15):

$$\begin{aligned} \frac{1-q}{\tau_k} - L_f\rho_k - \frac{L_K^2\sigma_k}{r} &\ge \left((1-q)P_2 - \frac{1}{r}\right)\frac{\Omega_X L_K}{\Omega_Y} \ge 0, \\ \frac{1-r}{\sigma_k} - \tau_k\frac{\|B\|_2^2}{q} &\ge \left(1 - r - \frac{b^2/q}{P_2}\right)\frac{\Omega_X L_K}{\Omega_Y} \ge 0, \end{aligned}$$

Then by (17), we have

$$\begin{aligned} \mathcal{G}^\star(z^k) &\le \frac{\rho_{k-1}}{\tau_{k-1}}\Omega_X^2 + \frac{\rho_{k-1}}{\sigma_{k-1}}\Omega_Y^2 \\ &= \frac{4P_1L_f + 2P_2(k-1)L_K\Omega_Y/\Omega_X}{k(k-1)}\Omega_X^2 + \frac{2L_K\Omega_X/\Omega_Y}{k}\Omega_Y^2 \\ &= \frac{4P_1\Omega_X^2}{k(k-1)}L_f + \frac{2\Omega_X\Omega_Y(P_2+1)}{k}L_K. \end{aligned}$$

□

B.3 Proof of Theorem 2

We need the following lemma to prove Theorem 2.

Lemma B.2. *Consider a saddle point $\hat{z} = (\hat{x}, \hat{y})$ of the problem* (PD)*, and the parameters ρ_k, θ_k, τ_k, and σ_k satisfying the conditions for Theorem 2. Then*

$$\|x - \tilde{x}^1\|_2^2 + \frac{\tau_k}{\sigma_k}\|y - \tilde{y}^1\|_2^2 \ge (1-q)\|x - \tilde{x}^{k+1}\|_2^2 + \frac{\tau_k}{\sigma_k}\left(\frac{1}{2} - r\right)\|y - \tilde{y}^{k+1}\|_2^2 \tag{B.6}$$

and

$$\tilde{\mathcal{G}}(\tilde{z}^{k+1}, v^{k+1}) \le \frac{\rho_k}{2\tau_k}\|x^{k+1} - \tilde{x}^1\|_2^2 + \frac{\rho_k}{2\sigma_k}\|y^{k+1} - \tilde{y}^1\|_2^2 =: \delta_{k+1} \tag{B.7}$$

for all $t \ge 1$, *where*

$$v^{k+1} = \left(\frac{\rho_k}{\tau_k}(\tilde{x}^1 - \tilde{x}^{k+1}) - B^T(\tilde{y}^{k+1} - \tilde{y}^k), \frac{\rho_k}{\sigma_k}(\tilde{y}^1 - \tilde{y}^{k+1}) - K(\tilde{x}^{k+1} - \tilde{x}^k)\right). \tag{B.8}$$

Proof of Theorem 2. It is sufficient to find upper bounds of $\|v^{k+1}\|_2$ and δ_{k+1}. From the definition of R and (B.6), we have $\|\hat{x} - \tilde{x}^{k+1}\|_2 \le \mu R$ and $\|\hat{y} - \tilde{y}^{k+1}\|_2 \le \sqrt{\frac{\sigma_k}{\tau_k}}\nu R$. For v^{k+1} defined in (B.8),

$$\begin{aligned}
\|v^{k+1}\|_2 &\le \rho_k(\frac{1}{\tau_k}\|\tilde{x}^1 - \tilde{x}^{k+1}\|_2 + \|B\|_2\|\tilde{y}^{k+1} - \tilde{y}^k\|_2 \\
&\quad + \frac{1}{\sigma_k}\|\tilde{y}^1 - \tilde{y}^{k+1}\|_2 + L_K\|\tilde{x}^{k+1} - \tilde{x}^k\|_2) \\
&\le \rho_k(\frac{1}{\tau_k}(\|\hat{x} - \tilde{x}^1\|_2 + \|\hat{x} - \tilde{x}^{k+1}\|_2) + \frac{1}{\sigma_k}(\|\hat{y} - \tilde{y}^1\|_2 + \|\hat{y} - \tilde{y}^{k+1}\|_2) \\
&\quad + L_K(\|\hat{x} - \tilde{x}^{k+1}\|_2 + \|\hat{x} - \tilde{x}^k\|_2) + \|B\|_2(\|\hat{y} - \tilde{y}^{k+1}\|_2 + \|\hat{y} - \tilde{y}^k\|_2) \\
&\le \frac{\rho_k}{\tau_k}\|\hat{x} - \tilde{x}^1\|_2 + \frac{\rho_k}{\sigma_k}\|\hat{y} - \tilde{y}^1\|_2 \\
&\quad + \rho_k\left(\frac{1}{\tau_k} + 2L_K\right)\mu R + \rho_k\left(\frac{1}{\sigma_k} + 2\|B\|_2\right)\nu R \\
&= \frac{\rho_k}{\tau_k}\|\hat{x} - \tilde{x}^1\|_2 + \frac{\rho_k}{\sigma_k}\|\hat{y} - \tilde{y}^1\|_2 \\
&\quad + R\left[\frac{\rho_k}{\tau_k}\left(\mu + \frac{\tau_1}{\sigma_1}\nu\right) + 2\rho_k\left(L_K\mu + \|B\|_2\nu\right)\right],
\end{aligned}$$

i.e., (22). In the last equality, we used

$$\frac{1}{\sigma_k} = \frac{\tau_k}{\sigma_k}\frac{1}{\tau_k} = \frac{\tau_1}{\sigma_1}\frac{1}{\tau_k}.$$

Next, we find an upper bound for δ_{k+1} defined in Lemma B.2.

$$\begin{aligned}
\delta_{k+1} &= \frac{\rho_k}{2\tau_k}\|x^{k+1} - \tilde{x}^1\|_2^2 + \frac{\rho_k}{2\sigma_k}\|y^{k+1} - \tilde{y}^1\|_2^2 \\
&\le \frac{\rho_k}{\tau_k}\left(\|\hat{x} - x^{k+1}\|_2^2 + \|\hat{x} - \tilde{x}^1\|_2^2\right) + \frac{\rho_k}{\sigma_k}\left(\|\hat{y} - y^{k+1}\|_2^2 + \|\hat{y} - \tilde{y}^1\|_2^2\right) \\
&= \frac{\rho_k}{\tau_k}\Bigg((R^2 + (1-q)\|\hat{x} - x^{k+1}\|_2^2 + \frac{\tau_k}{\sigma_k}(1/2 - r)\|\hat{y} - y^{k+1}\|_2^2 \\
&\quad + q\|\hat{x} - x^{k+1}\|_2^2 + \frac{\tau_k}{\sigma_k}(r + 1/2)\|\hat{y} - y^{k+1}\|_2^2\Bigg) \\
&\le \frac{\rho_k}{\tau_k}\Bigg[R^2 + \frac{\rho_k}{\gamma_k}\sum_{i=1}^{k}\gamma_i\Big[(1-q)\|\hat{x} - \tilde{x}^{i+1}\|_2^2 + \frac{\tau_k}{\sigma_k}(1/2 - r)\|\hat{y} - \tilde{y}^{i+1}\|_2^2 \\
&\quad + q\|\hat{x} - \tilde{x}^{i+1}\|_2^2 + \frac{\tau_k}{\sigma_k}(r + 1/2)\|\hat{y} - \tilde{y}^{i+1}\|_2^2\Big]\Bigg] \\
&\le \frac{\rho_k}{\tau_k}\left[R^2 + \frac{\rho_k}{\gamma_k}\sum_{i=1}^{k}\gamma_i\left[R^2 + q\|\hat{x} - \tilde{x}^{i+1}\|_2^2 + \frac{\tau_k}{\sigma_k}(r + 1/2)\|\hat{y} - \tilde{y}^{i+1}\|_2^2\right]\right] \\
&\le \frac{\rho_k}{\tau_k}\left[2 + q\mu^2 + (r + 1/2)\nu^2\right]R^2 \\
&= \frac{\rho_k}{\tau_k}\left[2 + \frac{q}{1-q} + \frac{r + 1/2}{1/2 - r}\right]R^2,
\end{aligned}$$

i.e., (21). In the second and third inequalities, we used

$$x^{k+1} = \frac{\rho_k}{\gamma_k}\sum_{i=1}^{k}\gamma_i\tilde{x}^{i+1}, \quad y^{k+1} = \frac{\rho_k}{\gamma_k}\sum_{i=1}^{k}\gamma_i\tilde{y}^{i+1}, \quad \text{and} \quad \frac{\rho_k}{\gamma_k}\sum_{i=1}^{k}\gamma_i = 1.$$

□

B.4 Proof of Corollary 2

Proof of Corollary 2. First check if (7), (10), and (24) satisfy (15) and (20). Conditions (20) and (15a) are trivial to see. To prove (15b) and (15c):

$$\begin{aligned}\frac{1-q}{\tau_k} - L_f\rho_k - \frac{L_K^2\sigma_k}{r} &\geq L_K\left((1-q)P_2\frac{N}{k} - \frac{k}{rN}\right)\\ &\geq L_K\left((1-q)P_2 - \frac{1}{r}\right) \geq 0,\end{aligned}$$

and

$$\frac{1-r}{\sigma_k} - \tau_k\frac{\|B\|_2^2}{q} \geq \left(\frac{(1-r)N}{k} - \frac{(b^2/q)k}{NP_2}\right)L_K \geq \left((1-r) - \frac{b^2}{qP_2}\right)L_K.$$

Condition (24) also implies that $\tau_k \leq \sigma_k$.

Note that

$$\begin{aligned}\frac{\rho_N}{\tau_N} &\leq \frac{4P_1L_f}{N^2} + \frac{2P_2L_K}{N}\\ L_K^2\rho_N\tau_N &\leq \frac{2NL_K^2}{(2P_1L_f + P_2NL_K)(N+1)} \leq \frac{2L_K}{P_2N}\\ \rho_NL_K &\leq \frac{2L_K}{N}.\end{aligned} \tag{B.9}$$

When we put $\|B\|_2 \leq bL_K$, $\|v^{k+1}\|_2$ is bounded above by

$$\|v^{k+1}\|_2 \leq \frac{\rho_k}{\tau_k}\left(\|\hat{x} - \tilde{x}^1\|_2 + \|\hat{y} - \tilde{y}^1\|_2\right) + R\left[\frac{\rho_k}{\tau_k}\left(\mu + \frac{\tau_1}{\sigma_1}\nu\right) + 2\rho_kL_K(\mu + b\nu)\right].$$

Thus by (B.9), we have

$$\epsilon_{N+1} \leq \delta_{N+1} \leq \left(\frac{4P_1L_f}{N^2} + \frac{2P_2L_K}{N}\right)\left[2 + \frac{q}{1-q} + \frac{r+1/2}{1/2-r}\right]R^2,$$

which is (25), and

$$\begin{aligned}\|v^{N+1}\|_2 \leq\ & \frac{4P_1L_f}{N^2}\left[\left(\|\hat{x} - \tilde{x}^1\|_2 + \|\hat{y} - \tilde{y}^1\|_2\right) + R\left(\mu + \frac{\tau_1}{\sigma_1}\nu\right)\right]\\ &+ \frac{L_K}{N}\left[2P_2\left(\left(\|\hat{x} - \tilde{x}^1\|_2 + \|\hat{y} - \tilde{y}^1\|_2\right) + R\left(\mu + \frac{\tau_1}{\sigma_1}\nu\right)\right) + 4R(\mu + b\nu)\right],\end{aligned}$$

which is (26). □

C Optimal stochastic acceleration

In this section, we provide proofs that Algorithm 1 achieves the theoretically optimal rate of convergence in the stochastic settings. Proofs of all the technical lemmas are deferred in Section D.

C.1 Proof of Theorem 3

We obtain a bound similar to Lemma B.1 first. The following lemma provides the desired upper bound on $\rho_k^{-1}\gamma_k \mathcal{G}(z^k, z)$.

Lemma C.1. *Assume that $z^k = (x^k, y^k)$ is the iterates generated by the iteration* (6) *with $\widehat{\nabla f}$ satisfying* (27)*. Also assume that the parameters satisfy* (15a), (16), *and* (28)*. Then for any $z \in \mathcal{Z}$, we have*

$$\begin{aligned}\rho_k^{-1}\gamma_k \mathcal{G}(z^{k+1}, z) \le\ & \mathcal{D}_k(z, \tilde{z}^{[k]}) - \gamma_k\langle \tilde{x}^{k+1} - x, B^T(\tilde{y}^{k+1} - \tilde{y}^k)\rangle \\ & - \gamma_k\langle K(\tilde{x}^{k+1} - \tilde{x}^k), \tilde{y}^{k+1} - y\rangle \\ & - \gamma_k\left(\frac{s-q}{2\tau_k} - \frac{\rho_k L_f}{2}\right)\|\tilde{x}^{k+1} - \tilde{x}^k\|_2^2 \\ & - \gamma_k\left(\frac{t-r}{2\sigma_k}\right)\|\tilde{y}^{k+1} - \tilde{y}^k\|_2^2 \\ & + \sum_{i=1}^k \Lambda_i(z),\end{aligned} \tag{C.1}$$

where γ_k and $\mathcal{D}(z, \tilde{z}^{[k]})$ are defined in (B.2) *and* (B.3)*, respectively, and*

$$\Lambda_i(z) := -\frac{(1-s)\gamma_i}{2\tau_i}\|\tilde{x}^{i+1} - x^i\|_2^2 - \frac{(1-t)\gamma_i}{2\sigma_i}\|\tilde{y}^{i+1} - y^i\|_2^2 - \gamma_i\langle \Delta^i, x^{i+1} - x\rangle.$$

We also need the following lemma to prove Theorem 3. For subsequent uses, we define $\Delta^k := \widehat{\nabla f}(x_{md}^k) - \nabla f(x_{md}^k)$.

Lemma C.2 (Lemma 4.5, Chen et al., 2014)**.** *Let τ_i, σ_i, and $\gamma_i > 0$. For any $\tilde{z}^1 \in Z$, define $\tilde{z}_v^1 = \tilde{x}^1$ and*

$$x_v^{i+1} = \arg\min_{x\in\mathcal{X}}\left\{-\tau_i\langle\Delta^i, x\rangle + \frac{1}{2}\|x - x_v^i\|_2^2\right\}, \tag{C.2}$$

then

$$\sum_{i=1}^k \gamma_i\langle -\Delta^i, x_v^i - x\rangle \le \mathcal{D}_k(z, \tilde{z}_v^{[k]}) + \sum_{i=1}^k \frac{\tau_i\gamma_i}{2}\|\Delta^i\|_2^2,$$

where $\tilde{z}_v^{[k]} := \{z_v^i\}_{i=1}^k$.

Proof of Theorem 3. First we use the bounds in (B.5) to obtain

$$\rho_k^{-1}\gamma_k\mathcal{G}(z^{k+1}, z) \le \frac{\gamma_k}{\tau_k}\Omega_X^2 + \frac{\gamma_k}{\sigma_k}\Omega_Y^2 + \sum_{i=1}^k \Lambda_i(z).$$

Then by the definition of $\Lambda_i(z)$, we have

$$\begin{aligned}\Lambda_i(z) &= -\frac{(1-s)\gamma_i}{2\tau_i}\|\tilde{x}^{i+1} - x^i\|_2^2 - \frac{(1-t)\gamma_i}{2\sigma_i}\|\tilde{y}^{i+1} - y^i\|_2^2 + \gamma_i\langle\Delta^i, x - x^{i+1}\rangle \\ &= -\frac{(1-s)\gamma_i}{2\tau_i}\|\tilde{x}^{i+1} - x^i\|_2^2 - \frac{(1-t)\gamma_i}{2\sigma_i}\|\tilde{y}^{i+1} - y^i\|_2^2 + \gamma_i\langle\Delta^i, x^i - x^{i+1}\rangle + \gamma_i\langle\Delta^i, x - x^i\rangle \\ &\le \frac{\tau_i\gamma_i}{2(1-s)}\|\Delta^i\|_2^2 + \gamma_i\langle\Delta^i, x - x^i\rangle,\end{aligned}$$

where the last line is due to Young's inequality. By this result and Lemma C.2, we have

$$\sum_{i=1}^k \Lambda_i(z) \le \sum_{i=1}^k \left[\frac{\tau_i\gamma_i}{2(1-s)}\|\Delta^i\|_2^2 + \gamma_i\langle\Delta^i, x_v^i - x^i\rangle + \gamma_i\langle -\Delta^i, x_v^i - x\rangle\right]$$

$$\leq \mathcal{D}_k(z, \tilde{z}_v^{[k]}) + \frac{1}{2}\sum_{i=1}^{k}\left[\frac{(2-s)\tau_i\gamma_i}{1-s}\|\Delta^i\|_2^2 + \gamma_i\langle\Delta^i, x_v^i - x^i\rangle\right]. \tag{C.3}$$

Let us define U_k as

$$U_k := \frac{1}{2}\sum_{i=1}^{k}\left[\frac{(2-s)\tau_i\gamma_i}{1-s}\|\Delta^i\|_2^2 + \gamma_i\langle\Delta^i, x_v^i - x^i\rangle\right] \tag{C.4}$$

for later use.

Note that Δ^i and x^i are independent by the assumptions of stochastic oracle. Thus,

$$\mathbb{E}[U_k] \leq \frac{1}{2}\sum_{i=1}^{k}\left[\frac{(2-s)\tau_i\gamma_i\chi^2}{1-s}\right]. \tag{C.5}$$

Similar to (B.4), $\mathcal{D}_k(z, \tilde{z}_v^{[k]}) \leq \frac{\Omega_X^2\gamma_k}{\tau_k} + \frac{\Omega_Y^2\gamma_k}{\sigma_k}$. Thus we have:

$$\mathbb{E}[\rho_k^{-1}\gamma_k\mathcal{G}^\star(z^{k+1})] \leq \frac{2\gamma_k}{\tau_k}\Omega_X^2 + \frac{2\gamma_k}{\sigma_k}\Omega_Y^2 + \mathbb{E}[U_k].$$

The above relation along with (C.5) implies the desired result.

□

C.2 Proof of Corollary 3

Proof of Corollary 3. First we check (28):

$$\frac{s-q}{\tau_k} - \rho_k L_f - \frac{L_K^2\sigma_k}{r} \geq \frac{L_K\Omega_Y}{\Omega_X}\left((s-q)P_2 - \frac{1}{r}\right) \geq 0$$

and

$$\frac{t-r}{\sigma_k} - \tau_k\frac{\|B\|_2^2}{q} \geq \left((t-r) - \frac{b^2/q}{P_2}\right)\frac{\Omega_X L_K}{\Omega_Y} \geq 0,$$

where we use (31). Note that $\gamma_k = k$, $\sum_{i=1}^k \sqrt{i} \leq \int_1^{k+1}\sqrt{u}du \leq \frac{2}{3}(k+1)^{3/2} \leq \frac{2\sqrt{2}}{3}(k+1)\sqrt{k}$, so

$$\frac{1}{\gamma_{N-1}}\sum_{i=1}^{k}\tau_i\gamma_i \leq \frac{\Omega_X}{kP_3\chi}\sum_{i=1}^{k}\sqrt{i} \leq \frac{2\sqrt{2}\Omega_X(k+1)\sqrt{k}}{3kP_3\chi},$$

which in turn implies

$$\begin{aligned}\mathbb{E}\left[\mathcal{G}^\star(x^{k+1}, y^{k+1})\right] \leq \frac{2}{k+1}&\left[\frac{2(2P_1L_f\Omega_X + P_2L_K\Omega_Y k + P_3\chi k^{3/2})}{\Omega_X k}\Omega_X^2\right.\\ &\left. + 2\Omega_X\Omega_Y + \frac{2\sqrt{2}(2-s)\Omega_X\chi_x^2(k+1)\sqrt{k}}{6(1-s)P_3\chi k}\right]\\ &\leq \frac{8P_1L_f\Omega_X^2}{k(k+1)} + \frac{4L_K\Omega_X\Omega_Y(P_2+1)}{k+1} + \left(4P_3 + \frac{2\sqrt{2}(2-s)}{3P_3(1-s)}\right)\frac{\chi\Omega_X}{\sqrt{k}}.\end{aligned}$$

□

C.3 Proof of Theorem 4

We need the following lemma to prove Theorem 4.

Lemma C.3. *For a saddle point $\hat{z} = (\hat{x}, \hat{y})$ of* (PD)*, and the parameters ρ_k, θ_k, τ_k, and σ_k satisfy* (15a), (20)*, and* (28)*, then*

$$\begin{aligned}&(1-q)\|\hat{x}-\tilde{x}^{k+1}\|_2^2 + \|\hat{x}-\tilde{x}_v^{k+1}\|_2^2 + \frac{\tau_k(1/2-r)}{\sigma_k}\|\hat{y}-\tilde{y}^{k+1}\|_2^2 + \frac{\tau_k}{\sigma_k}\|\hat{y}-\tilde{y}_v^{k+1}\|_2^2 \\ &\leq 2\|\hat{x}-\tilde{x}^1\|_2^2 + \frac{2\tau_k}{\sigma_k}\|\hat{y}-\tilde{y}^1\|_2^2 + \frac{2\tau_k}{\gamma_k}U_k,\end{aligned} \tag{C.6}$$

where $(\tilde{x}_v^{k+1}, \tilde{y}_v^{k+1})$ is defined in (C.2)*, and U_k is defined by* (C.4)*.*

Furthermore,

$$\tilde{\mathcal{G}}(z^{k+1}, v^{k+1}) \leq \frac{\rho_k}{\tau_k}\|x^{k+1}-\tilde{x}^1\|_2^2 + \frac{\rho_k}{\sigma_k}\|y^{k+1}-\tilde{y}^1\|_2^2 + \frac{\rho_k}{\gamma_k}U_k =: \delta'_{k+1}, \tag{C.7}$$

for $k \geq 1$, where

$$v_{k+1} = \rho_k\left(\frac{1}{\tau_k}(2\tilde{x}^1 - \tilde{x}^{k+1} - \tilde{x}_v^{k+1}) - B^T(\tilde{y}^{k+1}-\tilde{y}^k), \frac{1}{\sigma_k}(2\tilde{y}^1 - \tilde{y}^{k+1} - \tilde{y}_v^{k+1}) - K(\tilde{x}^{k+1}-\tilde{x}^k)\right).$$

Proof of Theorem 4. By the definition of S in (35) and (C.5), we have

$$\mathbb{E}[U_k] \leq \frac{\gamma_k}{2\tau_k}S^2.$$

By the above, Lemma C.3, and (23), we have

$$\mathbb{E}[\|\hat{x}-\tilde{x}^{k+1}\|_2^2] \leq \frac{2R^2+S^2}{1-q} \text{ and } \mathbb{E}[\|\hat{y}-\tilde{y}^{k+1}\|_2^2] \leq \frac{(2R^2+S^2)\sigma_1}{\tau_1(1/2-r)}.$$

Using Jensen's inequality, this leads to

$$\mathbb{E}[\|\hat{x}-\tilde{x}^{k+1}\|_2] \leq \sqrt{\frac{2R^2+S^2}{1-q}} \text{ and } \mathbb{E}[\|\hat{y}-\tilde{y}^{k+1}\|_2] \leq \sqrt{\frac{(2R^2+S^2)\sigma_1}{\tau_1(1/2-r)}}.$$

Similarly, we have

$$\mathbb{E}[\|\hat{x}-\tilde{x}_v^{k+1}\|_2] \leq \sqrt{2R^2+S^2} \text{ and } \mathbb{E}[\|\hat{y}-\tilde{y}_v^{k+1}\|_2] \leq \sqrt{\frac{(2R^2+S^2)\sigma_1}{\tau_1}}.$$

Thus

$$\begin{aligned}\mathbb{E}[\|v^{k+1}\|_2] \leq{}& \rho_k\,\mathbb{E}[\frac{1}{\tau_k}(2\|\hat{x}-\tilde{x}^1\|_2 + \|\hat{x}-\tilde{x}^{k+1}\| + \|\hat{x}-\tilde{x}_v^{k+1}\|_2) \\ &+ \frac{1}{\sigma_k}(2\|\hat{y}-\tilde{y}^1\|_2 + \|\hat{y}-\tilde{y}^{k+1}\|_2 + \|\hat{y}-\tilde{y}_v^{k+1}\|_2) \\ &+ L_K(\|\hat{x}-\tilde{x}^{k+1}\|_2 + \|\hat{x}-\tilde{x}^k\|_2) \\ &+ \|B\|_2(\|\hat{y}-\tilde{y}^{k+1}\|_2 + \|\hat{y}-\tilde{y}^k\|_2)] \\ \leq{}& \frac{2\rho_k\|\hat{x}-\tilde{x}^1\|_2}{\tau_k} + \frac{2\rho_k\|\hat{y}-\tilde{y}^1\|_2}{\sigma_k} \\ &+ \sqrt{2R^2+S^2}\left[\frac{\rho_k}{\tau_k}(1+\mu') + \frac{\rho_k}{\sigma_k}\sqrt{\frac{\sigma_1}{\tau_1}}(1+\nu') \right. \\ &\left. + \rho_k(2L_K\mu' + 2\|B\|_2\nu'\sqrt{\frac{\sigma_1}{\tau_1}})\right],\end{aligned} \tag{C.8}$$

where $\mu' = 1/\sqrt{1-q}$ and $\nu' = 1/\sqrt{1/2-r}$. Now we find an upper bound of $\mathbb{E}[\delta'_{k+1}]$.

$$\begin{aligned}
\mathbb{E}[\delta'_{k+1}] &\leq \mathbb{E}\left[\frac{2\rho_k}{\tau_k}(\|\hat{x}-x^{k+1}\|_2^2 + \|\hat{x}-\tilde{x}^1\|_2^2) + \frac{2\rho_k}{\sigma_k}(\|\hat{y}-y^{k+1}\|_2^2 + \|\hat{y}-\tilde{y}^1\|_2^2)\right] + \frac{\rho_k}{2\tau_k}S^2 \\
&= \frac{\rho_k}{\tau_k}\mathbb{E}\Big[(2R^2 + 2(1-q)\|\hat{x}-x^{k+1}\|_2^2 + \frac{2\tau_k(1/2-r)}{\sigma_k}\|\hat{y}-y^{k+1}\|_2^2) \\
&\quad +2q\|\hat{x}-x^{k+1}\|_2^2 + \frac{2\tau_k(r+1/2)}{\sigma_k}\|\hat{y}-y^{k+1}\|_2^2\Big] + \frac{\rho_k}{2\tau_k}S^2 \\
&\leq \frac{\rho_k}{\tau_k}2R^2 + \frac{2\rho_k}{\gamma_k}\sum_{i=1}^{k}\gamma_i\left[(4R^2+2S^2) + q\mu'^2(2R^2+S^2) + (r+1/2)\nu'^2(2R^2+S^2)\right] + \frac{S^2}{2} \\
&= \frac{\rho_k}{\tau_k}\left[6R^2 + \frac{5}{2}S^2 + \frac{2q}{1-q}(2R^2+S^2) + \frac{2(r+1/2)}{1/2-r}(2R^2+S^2)\right] \\
&= \frac{\rho_k}{\tau_k}\left[\left(6 + \frac{4q}{1-q} + \frac{4(r+1/2)}{1/2-r}\right)R^2 + \left(\frac{5}{2} + \frac{2q}{1-q} + \frac{2(r+1/2)}{1/2-r}\right)S^2\right].
\end{aligned}$$

□

C.4 Proof of Corollary 4

Proof of Corollary 4. First we check (28):

$$\frac{s-q}{\tau_k} - \rho_k L_f - \frac{L_K^2\sigma_k}{r} \geq L_K\left((s-q)P_2 - \frac{1}{r}\right) \geq 0,$$

$$\frac{t-r}{\sigma_k} - \tau_k\frac{b^2L_K^2}{q} \geq L_K\left((t-r) - \frac{b^2}{qP_2}\right) \geq 0,$$

by (36).

Now, when we put $\eta = 2P_1L_f + P_2L_K(N-1) + P_3N\sqrt{N-1}$,

$$\begin{aligned}
S &= \sqrt{\sum_{i=1}^{N-1}\frac{(2-s)\chi^2 i^2}{(1-s)\eta^2}} \\
&\leq \sqrt{\frac{N^2(N-1)}{3\eta^2}\left(\frac{(2-s)\chi^2}{1-s}\right)} = \frac{\chi' N\sqrt{N-1}}{\sqrt{3}\eta} \\
&\leq \frac{\chi' N\sqrt{N-1}}{\sqrt{3}N\sqrt{N-1}\chi'/\tilde{R}} = \frac{\tilde{R}}{\sqrt{3}},
\end{aligned}$$

where we define $\chi' = \sqrt{\frac{2-s}{1-s}}\chi$. Thus ϵ_N is bounded above by

$$\epsilon_N \leq \frac{\rho_{N-1}}{\tau_{N-1}}(\zeta R^2 + \xi S^2) \leq \frac{\rho_{N-1}}{\tau_{N-1}}(\zeta R^2 + \xi\frac{\tilde{R}^2}{3}),$$

where $\zeta = 6 + \frac{4q}{1-q} + \frac{4(r+1/2)}{1/2-r}$ and $\xi = \frac{5}{2} + \frac{2q}{1-q} + \frac{2(r+1/2)}{1/2-r}$.

Note that

$$\frac{\rho_{N-1}}{\tau_{N-1}}\|\hat{x}-\tilde{x}^1\|_2 \leq \frac{\rho_{N-1}}{\tau_{N-1}}R, \quad \frac{\rho_{N-1}}{\sigma_{N-1}}\|\hat{y}-\tilde{y}^1\|_2 \leq \frac{\rho_{N-1}}{\tau_{N-1}}\sqrt{\frac{\sigma_1}{\tau_1}}R,$$

since $\sigma_k \geq \tau_k$, and that

$$
\begin{aligned}
\rho_{N-1} L_K &\leq \frac{2L_K}{N}, \\
\frac{\rho_{N-1}}{\tau_{N-1}} &\leq \frac{2\tau}{N(N-1)} = \frac{4P_1 L_f + 2P_2 L_K (N-1) + 2N\sqrt{N-1}\chi'/\tilde{R}}{N(N-1)} \\
&= \frac{4P_1 L_f}{N(N-1)} + \frac{2P_2 L_K}{N} + \frac{2\chi'/\tilde{R}}{\sqrt{N-1}}
\end{aligned}
$$

Thus

$$
\begin{aligned}
\epsilon_N &\leq \frac{\rho_{N-1}}{\tau_{N-1}} (\zeta R^2 + \xi S^2) \\
&\leq \left(\frac{4P_1 L_f}{N(N-1)} + \frac{2P_2 L_K}{N} + \frac{2\chi'/\tilde{R}}{\sqrt{N-1}} \right) \left(\zeta R^2 + \frac{\xi \tilde{R}^2}{3} \right).
\end{aligned}
$$

Now note that $\sqrt{2R^2 + S^2} \leq \sqrt{2}R + S$. Thus from (C.8),

$$
\begin{aligned}
\mathbb{E}[\|v^N\|_2] &\leq \frac{\rho_{N-1}}{\tau_{N-1}} \left[2R \left(1 + \sqrt{\frac{\sigma_1}{\tau_1}} \right) + (\sqrt{2}R + S) \left(1 + \mu' + \sqrt{\frac{\sigma_1}{\tau_1}} (1 + \nu') \right) \right] + 2\rho_{N-1} L_K (\sqrt{2}R + S) \left(\mu' + b\nu' \sqrt{\frac{\sigma_1}{\tau_1}} \right) \\
&\leq \left(\frac{4P_1 L_f}{N(N-1)} + \frac{2P_2 L_K}{N} + \frac{2\chi'/\tilde{R}}{\sqrt{N-1}} \right) \left[2R \left(1 + \sqrt{\frac{\sigma_1}{\tau_1}} \right) + (\sqrt{2}R + \tilde{R}/\sqrt{3}) \left(1 + \mu' + \sqrt{\frac{\sigma_1}{\tau_1}} (1 + \nu') \right) \right] \\
&\quad + \frac{4L_K}{N} (\sqrt{2}R + \tilde{R}/\sqrt{3}) \left(\mu' + b\nu' \sqrt{\frac{\sigma_1}{\tau_1}} \right),
\end{aligned}
$$

and we obtain the desired order for both ϵ_N and $\mathbb{E}[\|v_N\|_2]$. □

D Proofs of techinical lemmas

Proof of Lemma B.1. Loris and Verhoeven (2011, Lemma 1) state that if $\phi : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ is a convex, closed, and proper, and if $x^+ = \mathbf{prox}_{\sigma\phi}(x^- + \sigma\Delta)$, then for any $x \in \mathbb{R}^n$,

$$
\langle x - x^+, \Delta \rangle - \phi^*(x) + \phi^*(x^+) \leq \frac{1}{2\sigma} \left(\|x - x^-\|_2^2 - \|x - x^+\|_2^2 - \|x^- - x^+\|_2^2 \right). \tag{D.1}
$$

Applying inequality (D.1) to equations (6c) and (6e), we have

$$
\begin{aligned}
\langle y - \tilde{y}^{k+1}, \tilde{u}^{k+1} \rangle + h^*(\tilde{y}^{k+1}) - h^*(y) &\leq \frac{1}{2\sigma_k} \left(\|y - \tilde{y}^k\|_2^2 - \|\tilde{y}^{k+1} - \tilde{y}^k\|_2^2 - \|y - \tilde{y}^{k+1}\|_2^2 \right), \\
\langle \tilde{x}^{k+1} - x, \nabla f(x_{md}^k) + \tilde{v}^{k+1} \rangle + g(\tilde{x}^{k+1}) - g(x) &\leq \frac{1}{2\tau_k} \left(\|x - \tilde{x}^k\|_2^2 - \|\tilde{x}^{k+1} - \tilde{x}^k\|_2^2 - \|x - \tilde{x}^{k+1}\|_2^2 \right).
\end{aligned}
$$

Using the above relationship along with Proposition 1, we obtain the following.

$$
\begin{aligned}
&\rho_k^{-1} \mathcal{G}(z^{k+1}, z) - (\rho_k^{-1} - 1) \mathcal{G}(z^k, z) \\
&\qquad \leq \frac{1}{2\tau_k} \left(\|x - \tilde{x}^k\|_2^2 - \|x - \tilde{x}^{k+1}\|_2^2 \right) - \left(\frac{1}{2\tau_k} - \frac{L_f \rho_k}{2} \right) \|\tilde{x}^{k+1} - \tilde{x}^k\|_2^2 \\
&\qquad + \frac{1}{2\sigma_k} \left(\|y - \tilde{y}^k\|_2^2 - \|y - \tilde{y}^{k+1}\|_2^2 \right) - \frac{1}{2\sigma_k} \|\tilde{y}^{k+1} - \tilde{y}^k\|_2^2 \\
&\qquad - \langle \tilde{x}^{k+1} - x, \tilde{v}^{k+1} \rangle + \langle \tilde{u}^{k+1}, \tilde{y}^{k+1} - y \rangle + \langle K\tilde{x}^{k+1}, y \rangle - \langle Kx, \tilde{y}^{k+1} \rangle.
\end{aligned}
$$

The sum of the four inner products on the last line, namely, $-\langle \tilde{x}^{k+1} - x, \tilde{v}^{k+1} \rangle + \langle \tilde{u}^{k+1}, \tilde{y}^{k+1} - y \rangle + \langle K\tilde{x}^{k+1}, y \rangle - \langle Kx, \tilde{y}^{k+1} \rangle$, multiplied by γ_k can be computed as follows.

$$
\gamma_k [-\langle \tilde{x}^{k+1} - x, \tilde{v}^{k+1} \rangle + \langle \tilde{u}^{k+1}, \tilde{y}^{k+1} - y \rangle + \langle K\tilde{x}^{k+1}, y \rangle - \langle Kx, \tilde{y}^{k+1} \rangle]
$$

$$
\begin{aligned}
&= \gamma_k[-\big(\langle \tilde{x}^{k+1}-x, B^T(\tilde{y}^{k+1}-\tilde{y}^k)\rangle - \theta_k\langle \tilde{x}^{k+1}-x, B^T(\tilde{y}^k-\tilde{y}^{k-1})\rangle\big)\\
&\quad -\big(\langle K(\tilde{x}^{k+1}-\tilde{x}^k), \tilde{y}^{k+1}-y\rangle + \theta_k\langle K(\tilde{x}^k-\tilde{x}^{k-1}), \tilde{y}^{k+1}-y\rangle\big)\\
&= -\big(\gamma_k\langle \tilde{x}^{k+1}-x, B^T(\tilde{y}^{k+1}-\tilde{y}^k)\rangle - \gamma_{k-1}\langle \tilde{x}^k-x, B^T(\tilde{y}^k-\tilde{y}^{k-1})\rangle\big)\\
&\quad -\big(\gamma_k\langle K(\tilde{x}^{k+1}-\tilde{x}^k), \tilde{y}^{k+1}-y\rangle + \gamma_{k-1}\langle K(\tilde{x}^k-\tilde{x}^{k-1}), \tilde{y}^k-y\rangle\big)\\
&\quad + \gamma_{k-1}\langle \tilde{x}^{k+1}-\tilde{x}^k, B^T(\tilde{y}^k-\tilde{y}^{k-1})\rangle + \gamma_{k-1}\langle K(\tilde{x}^k-\tilde{x}^{k-1}), \tilde{y}^{k+1}-\tilde{y}^k\rangle
\end{aligned}
$$

By upper bounding the inner product terms, and noting that $\theta_k = \gamma_{k-1}/\gamma_k = \tau_{k-1}/\tau_k = \sigma_{k-1}/\sigma_k$, we have

$$
\begin{aligned}
|\gamma_{k-1}\langle \tilde{x}^{k+1}-\tilde{x}^k, B^T(\tilde{y}^k-\tilde{y}^{k-1})\rangle| &\le \frac{\gamma_k q}{2\tau_k}\|\tilde{x}^{k+1}-\tilde{x}^k\|_2^2 + \frac{\|B\|_2^2\gamma_{k-1}\tau_{k-1}}{2q}\|\tilde{y}^k-\tilde{y}^{k-1}\|_2^2\\
|\gamma_{k-1}\langle \tilde{x}^k-\tilde{x}^{k-1}, A^T(\tilde{y}^{k+1}-\tilde{y}^k)\rangle| &\le \frac{L_K^2\gamma_{k-1}\sigma_{k-1}}{2r}\|\tilde{x}^k-\tilde{x}^{k-1}\|_2^2 + \frac{\gamma_k r}{2\sigma_k}\|\tilde{y}^{k+1}-\tilde{y}^k\|_2^2
\end{aligned}
$$

for some positive q and r. Thus

$$
\begin{aligned}
&\rho_k^{-1}\gamma_k\mathcal{G}(z^{k+1},z) - (\rho_k^{-1}-1)\gamma_k\mathcal{G}(z^k,z)\\
&\quad\le \frac{\gamma_k}{2\tau_k}\left(\|x-\tilde{x}^k\|_2^2 - \|x-\tilde{x}^{k+1}\|_2^2\right) + \frac{\gamma_k}{2\sigma_k}\left(\|y-\tilde{y}^k\|_2^2 - \|y-\tilde{y}^{k+1}\|_2^2\right)\\
&\qquad -\big(\gamma_k\langle \tilde{x}^{k+1}-x, B^T(\tilde{y}^{k+1}-\tilde{y}^k)\rangle - \gamma_{k-1}\langle \tilde{x}^k-x, B^T(\tilde{y}^k-\tilde{y}^{k-1})\rangle\big)\\
&\qquad -\big(\gamma_k\langle \tilde{x}^{k+1}-\tilde{x}^k, K^T(\tilde{y}^{k+1}-y)\rangle - \gamma_{k-1}\langle \tilde{x}^k-\tilde{x}^{k-1}, K^T(\tilde{y}^k-y)\rangle\big)\\
&\qquad -\gamma_k\left(\frac{1-q}{2\tau_k} - \frac{L_f\rho_k}{2}\right)\|\tilde{x}^{k+1}-\tilde{x}^k\|_2^2 + \frac{L_K^2\gamma_{k-1}\sigma_{k-1}}{2r}\|\tilde{x}^k-\tilde{x}^{k-1}\|_2^2\\
&\qquad -\gamma_k\left(\frac{1-r}{2\sigma_k}\right)\|\tilde{y}^{k+1}-\tilde{y}^k\|_2^2 + \frac{\gamma_{k-1}\tau_{k-1}\|B\|_2^2}{2q}\|\tilde{y}^k-\tilde{y}^{k-1}\|_2^2.
\end{aligned}
$$

Recursively applying the above relation, we obtain the following:

$$
\begin{aligned}
&\rho_k^{-1}\gamma_k\mathcal{G}(z^{k+1},z)\\
&\quad\le \mathcal{D}_k(z,\tilde{z}^{[k]}) - \gamma_k(\langle \tilde{x}^{k+1}-x, B^T(\tilde{y}^{k+1}-\tilde{y}^k)\rangle + \langle \tilde{x}^{k+1}-\tilde{x}^k, K^T(\tilde{y}^{k+1}-y)\rangle\\
&\qquad -\gamma_k\left(\frac{1-q}{2\tau_k} - \frac{L_f\rho_k}{2}\right)\|\tilde{x}^{k+1}-\tilde{x}^k\|_2^2 - \gamma_k\left(\frac{1-r}{2\sigma_k}\right)\|\tilde{y}^{k+1}-\tilde{y}^k\|_2^2\\
&\qquad -\sum_{i=1}^{k-1}\gamma_i\left(\frac{1-q}{2\tau_i} - \frac{L_f\rho_k}{2} - \frac{L_K^2\sigma_i}{2r}\right)\|\tilde{x}^{i+1}-\tilde{x}^i\|_2^2\\
&\qquad -\sum_{i=1}^{k-1}\gamma_i\left(\frac{1-r}{2\sigma_i} - \frac{\tau_i\|B\|_2^2}{2q}\right)\|\tilde{y}^{i+1}-\tilde{y}^i\|_2^2.
\end{aligned}
$$

Thus by the conditions (15), the desired result holds. □

Proof of Lemma B.2. First, let us prove (B.6). The conditions for Lemma B.1 clearly holds. Note that

$$
\begin{aligned}
\mathcal{D}_k(z,\tilde{z}^{[k]}) &= \frac{\gamma_1}{2\tau_1}\|x-\tilde{x}^1\|_2^2 - \sum_{i=1}^{k-1}\left(\frac{\gamma_i}{2\tau_i} - \frac{\gamma_{i+1}}{2\tau_{i+1}}\right)\|x-\tilde{x}^{k+1}\|_2^2 - \frac{\gamma_k}{2\tau_k}\|x-\tilde{x}^{k+1}\|_2^2\\
&\quad + \frac{\gamma_1}{2\sigma_1}\|y-\tilde{y}^1\|_2^2 - \sum_{i=1}^{k-1}\left(\frac{\gamma_i}{2\sigma_i} - \frac{\gamma_{i+1}}{2\sigma_{i+1}}\right)\|y-\tilde{y}^{k+1}\|_2^2 - \frac{\gamma_k}{2\sigma_k}\|y-\tilde{y}^{k+1}\|_2^2.
\end{aligned}
$$

By (20), one may see that

$$
\gamma_k^{-1}\mathcal{D}_k(z,z^{[k]}) = \frac{1}{2\tau_k}\|x-\tilde{x}^1\|_2^2 - \frac{1}{2\tau_k}\|x-\tilde{x}^{k+1}\|_2^2 + \frac{1}{2\sigma_k}\|y-\tilde{y}^1\|_2^2 - \frac{1}{2\sigma_k}\|y-\tilde{y}^{k+1}\|_2^2.
$$

Thus (B.1) in Lemma B.1 is equivalent to

$$\begin{aligned}\rho_k^{-1}\mathcal{G}(\tilde{z}^{k+1},z) \le & \frac{1}{2\tau_k}\|x-\tilde{x}^1\|_2^2 - \frac{1}{2\tau_k}\|x-\tilde{x}^{k+1}\|_2^2 + \frac{1}{2\sigma_k}\|y-\tilde{y}^1\|_2^2 - \frac{1}{2\sigma_k}\|y-\tilde{y}^{k+1}\|_2^2 \\ & - \langle \tilde{x}^{k+1}-x, B^T(\tilde{y}^{k+1}-\tilde{y}^k)\rangle - \gamma_k\langle K(\tilde{x}^{k+1}-\tilde{x}^k), \tilde{y}^{k+1}-y\rangle \\ & - \left(\frac{1-q}{2\tau_k} - \frac{L_f\rho_k}{2}\right)\|\tilde{x}^{k+1}-\tilde{x}^k\|_2^2 \\ & - \left(\frac{1-r}{2\sigma_k}\right)\|\tilde{y}^{k+1}-\tilde{y}^k\|_2^2.\end{aligned}$$

Note that

$$\begin{aligned}\left|\langle K(\tilde{x}^{k+1}-\tilde{x}^k), \tilde{y}^{k+1}-y\rangle\right| &\le \frac{L_K^2\sigma_k}{2r}\|\tilde{x}^{k+1}-\tilde{x}^k\|_2^2 + \frac{r}{2\sigma_k}\|\tilde{y}^{k+1}-y\|_2^2 \\ \left|\langle \tilde{x}^{k+1}-x, B^T(\tilde{y}^{k+1}-\tilde{y}^k)\rangle\right| &\le \frac{q}{2\tau_k}\|\tilde{x}^{k+1}-x\|_2^2 + \frac{\|B\|_2^2\tau_k}{2q}\|\tilde{y}^{k+1}-\tilde{y}^k\|_2^2.\end{aligned} \tag{D.2}$$

Thus

$$\begin{aligned}\rho_k^{-1}\mathcal{G}(z^{k+1},z) \le & \frac{1}{2\tau_k}\|x-\tilde{x}^1\|_2^2 - \frac{1-q}{2\tau_k}\|x-\tilde{x}^{k+1}\|_2^2 \\ & + \frac{1}{2\sigma_k}\|y-\tilde{y}^1\|_2^2 - \frac{1}{2\sigma_k}\left(1-r-\frac{1}{2}\right)\|y-\tilde{y}^{k+1}\|_2^2 \\ & - \left(\frac{1-q}{2\tau_k} - \frac{L_f\rho_k}{2} - \frac{L_K^2\sigma_k}{2r}\right)\|\tilde{x}^{k+1}-\tilde{x}^k\|_2^2 \\ & - \left(\frac{1-r}{2\sigma_k} - \frac{\|B\|_2^2\tau_k}{2q}\right)\|\tilde{y}^{k+1}-\tilde{y}^k\|_2^2.\end{aligned}$$

Hence

$$\rho_k^{-1}\mathcal{G}(z^{k+1},z) \le \frac{1}{2\tau_k}\|x-\tilde{x}^1\|_2^2 - \frac{1-q}{2\tau_k}\|x-\tilde{x}^{k+1}\|_2^2 + \frac{1}{2\sigma_k}\|y-\tilde{y}^1\|_2^2 - \frac{1/2-r}{2\sigma_k}\|y-\tilde{y}^{k+1}\|_2^2.$$

Since $\mathcal{G}(z^{k+1},\hat{z}) \ge 0$, we obtain

$$\|x-\tilde{x}^1\|_2^2 + \frac{\tau_k}{\sigma_k}\|y-\tilde{y}^1\|_2^2 \ge (1-q)\|x-\tilde{x}^{k+1}\|_2^2 + \frac{\tau_k}{\sigma_k}(1/2-r)\|y-\tilde{y}^{k+1}\|_2^2.$$

Next, we prove (B.7). Note that

$$\begin{aligned}\|x-\tilde{x}^1\|_2^2 - \|x-\tilde{x}^{k+1}\|_2^2 &= 2\langle \tilde{x}^{k+1}-\tilde{x}^1, x-x^{k+1}\rangle + \|x^{k+1}-\tilde{x}^1\|_2^2 - \|x^{k+1}-\tilde{x}^{k+1}\|_2^2 \\ \|y-\tilde{y}^1\|_2^2 - \|y-\tilde{y}^{k+1}\|_2^2 &= 2\langle \tilde{y}^{k+1}-\tilde{y}^1, y-y^{k+1}\rangle + \|y^{k+1}-\tilde{y}^1\|_2^2 - \|y^{k+1}-\tilde{y}^{k+1}\|_2^2.\end{aligned} \tag{D.3}$$

From this, we have:

$$\begin{aligned}&\rho_k^{-1}\mathcal{G}(z^{k+1},z) - \frac{1}{\tau_k}\langle \tilde{x}^1-\tilde{x}^{k+1}, x^{k+1}-x\rangle - \frac{1}{\sigma_k}\langle \tilde{y}^1-\tilde{y}^{k+1}, y^{k+1}-y\rangle \\ &\quad - \langle x-x^{k+1}, B^T(\tilde{y}^{k+1}-\tilde{y}^k)\rangle - \langle K(\tilde{x}^{k+1}-\tilde{x}^k), y-y^{k+1}\rangle \\ &\le \frac{1}{2\tau_k}\left(\|x^{k+1}-\tilde{x}^1\|_2^2 - \|x^{k+1}-\tilde{x}^{k+1}\|_2^2\right) + \frac{1}{2\sigma_k}\left(\|y^{k+1}-\tilde{y}^1\|_2^2 - \|y^{k+1}-\tilde{y}^{k+1}\|_2^2\right) \\ &\quad - \left(\frac{1-q}{2\tau_k} - \frac{L_f\rho_k}{2}\right)\|\tilde{x}^{k+1}-\tilde{x}^k\|_2^2 - \left(\frac{1-r}{2\sigma_k}\right)\|\tilde{y}^{k+1}-\tilde{y}^k\|_2^2 \\ &\quad - \langle \tilde{x}^{k+1}-x^{k+1}, B^T(\tilde{y}^{k+1}-\tilde{y}^k)\rangle - \langle K(\tilde{x}^{k+1}-\tilde{x}^k), \tilde{y}^{k+1}-y^{k+1}\rangle \\ &\le \frac{1}{2\tau_k}\|x^{k+1}-\tilde{x}^k\|_2^2 + \frac{1}{2\sigma_k}\|y^{k+1}-\tilde{y}^1\|_2^2 - \frac{1-q}{2\tau_k}\|x^{k+1}-\tilde{x}^{k+1}\|_2^2 - \frac{1/2-r}{2\sigma_k}\|y^{k+1}-\tilde{y}^{k+1}\|_2^2\end{aligned}$$

$$
\begin{aligned}
&- \left(\frac{1-q}{2\tau_k} - \frac{L_f \rho_k}{2} - \frac{L_K^2 \sigma_k}{2r} \right) \|\tilde{x}^{k+1} - \tilde{x}^k\|_2^2 - \left(\frac{1-r}{2\sigma_k} - \frac{\|B\|_2^2 \tau_k}{2q} \right) \|\tilde{y}^{k+1} - \tilde{y}^k\|_2^2 \\
&\le \frac{1}{2\tau_k} \|x^{k+1} - \tilde{x}^1\|_2^2 + \frac{1}{2\sigma_k} \|y^{k+1} - \tilde{y}^1\|_2^2.
\end{aligned}
$$

In the penultimate inequality, the upper bound for inner product terms similar to (D.2) was used. □

Proof of Lemma C.1. Analogous to the proof of Lemma B.1, except for that we start with

$$
\langle -\tilde{u}_{k+1}, \tilde{y}^{k+1} - y \rangle + h^*(\tilde{y}^{k+1}) - h^*(y) \le \frac{1}{2\sigma_k} \|y - \tilde{y}^k\|_2^2 - \frac{1}{2\sigma_k} \|\tilde{y}^{k+1} - \tilde{y}^k\|_2^2 - \frac{1}{2\sigma_k} \|y - \tilde{y}^{k+1}\|_2^2
$$

$$
\langle \hat{\mathcal{F}}(x_{md}^k), \tilde{x}^{k+1} - x \rangle + \langle \tilde{x}^{k+1} - x, \tilde{v}_{k+1} \rangle \le \frac{1}{2\tau_k} \|x - \tilde{x}^k\|_2^2 - \frac{1}{2\tau_k} \|\tilde{x}^{k+1} - \tilde{x}^k\|_2^2 - \frac{1}{2\tau_k} \|x - \tilde{x}^{k+1}\|_2^2.
$$

□

Proof of Lemma C.3. By applying the bounds (D.2) and (C.3) to (C.1), we obtain:

$$
\rho_k^{-1} \gamma_k \mathcal{G}(z^{k+1}, z) \le \bar{\mathcal{D}}_k(z, \tilde{z}^{[k]}) + \frac{q\gamma_k}{2\tau_k} \|x - \tilde{x}^{k+1}\|_2^2 + \frac{(r + 1/2)\gamma_k}{2\sigma_k} \|y - \tilde{y}^{k+1}\|_2^2 + \bar{\mathcal{D}}_k(z, \tilde{z}_v^{[k]}) + U_k,
$$

where

$$
\bar{\mathcal{D}}_k(z, \tilde{z}^{[k]}) = \frac{\gamma_k}{2\tau_k} (\|x - \tilde{x}_1\|_2^2 - \|x - \tilde{x}_{k+1}\|_2^2) + \frac{\gamma_k}{2\sigma_k} (\|y - \tilde{y}_1\|_2^2 - \|y - \tilde{y}_{k+1}\|_2^2).
$$

Letting $z = \hat{z}$ and using $\mathcal{G}(z^{k+1}, \hat{z}) \ge 0$ leads to (C.6). If we only use (C.3) on (C.1), we get:

$$
\begin{aligned}
\rho_k^{-1} \gamma_k \mathcal{G}(z^{k+1}, z) \le \bar{\mathcal{D}}_k(z, \tilde{z}^{[k]}) &- \gamma_k \langle \tilde{x}^{k+1} - x, B^T(\tilde{y}^{k+1} - \tilde{y}^k) \rangle \\
&- \gamma_k \langle K(\tilde{x}^{k+1} - \tilde{x}^k), \tilde{y}^{k+1} - y \rangle
\end{aligned}
$$

Applying (D.3) and following the steps of Lemma B.2 results in (C.7). □

References

Loris, I. and Verhoeven, C. (2011). On a generalization of the iterative soft-thresholding algorithm for the case of non-separable penalty, *Inverse Problems* **27**(12): 125007.