7. Appendix

7.1. Proof of Lemma 1

Proof. Invoking the optimality condition for (6), we have
\[
(g(x^*) + s\nabla D(x^*, u), x - x^*) \geq 0, \ \forall x \in X,
\]
which is equivalent to
\[
\langle g(x^*), x^* - x \rangle \leq s \langle \nabla D(x^*, u), x - x^* \rangle = s \langle \nabla \omega(x^*) - \nabla \omega(u), x - x^* \rangle = s [D(x, u) - D(x, x^*) - D(x^*, u)].
\]
\[
\square
\]

7.2. Proof of Lemma 2

Proof. Due to the convexity of \( \theta_1 \) and using the definition of \( \delta_k \), we have
\[
\theta_1(x_k) - \theta_1(x) \leq \langle \theta'_1(x_k), x_k - x \rangle = \langle \theta'_1(x_k, \xi_{k+1}), x_{k+1} - x \rangle + \langle \delta_{k+1}, x - x_k \rangle + \langle \theta'_1(x_k, \xi_{k+1}), x_k - x_{k+1} \rangle.
\]
Applying Lemma 1 to Line 1 of Alg. 2 and taking \( D(u, v) = \frac{1}{2}||v - u||^2 \), we have
\[
\langle \theta'_1(x_k, \xi_{k+1}) + A^T [\beta(Ax_{k+1} + By_k - b) - \lambda_k], x_{k+1} - x \rangle \leq \frac{1}{2\eta_{k+1}} (||x_k - x||^2 - ||x_{k+1} - x||^2 - ||x_k - x_{k+1}||^2)
\]
(24)

Combining (23) and (24) we have
\[
\begin{align*}
\theta_1(x_k) &- \theta_1(x) + \langle x_{k+1} - x, -A^T \lambda_{k+1} \rangle \\
& \leq \langle \theta'_1(x_k, \xi_{k+1}), x_{k+1} - x \rangle + \langle \delta_{k+1}, x - x_k \rangle + \langle \theta'_1(x_k, \xi_{k+1}), x_k - x_{k+1} \rangle + \\
& + \langle x_{k+1} - x, A^T [\beta(Ax_{k+1} + By_k - b) - \lambda_k], x_{k+1} - x \rangle + \\
& = \langle \theta'_1(x_k, \xi_{k+1}) + A^T [\beta(Ax_{k+1} + By_k - b) - \lambda_k], x_{k+1} - x \rangle + \\
& + \langle \delta_{k+1}, x - x_k \rangle + \langle x - x_{k+1}, \beta A^T B(y_k - y_{k+1}) \rangle + \langle \theta'_1(x_k, \xi_{k+1}), x_k - x_{k+1} \rangle + \\
& \leq \frac{1}{2\eta_{k+1}} (||x_k - x||^2 - ||x_{k+1} - x||^2 - ||x_k - x_{k+1}||^2) + \langle \delta_{k+1}, x - x_k \rangle + \\
& + \langle x - x_{k+1}, \beta A^T B(y_k - y_{k+1}) \rangle + \langle \theta'_1(x_k, \xi_{k+1}), x_k - x_{k+1} \rangle
\end{align*}
\]

(25)

We handle the last two terms separately:
\[
\langle x - x_{k+1}, \beta A^T B(y_k - y_{k+1}) \rangle = \beta \langle Ax - Ax_{k+1}, By_k - By_{k+1} \rangle
\]
\[
= \frac{\beta}{2} \left( \langle Ax + By_k - b \rangle^2 - \langle Ax + By_{k+1} - b \rangle^2 \right) + \left( \langle Ax_{k+1} + By_k - b \rangle^2 - \langle Ax_{k+1} + By_{k+1} - b \rangle^2 \right)
\]
\[
\leq \frac{\beta}{2} \left( ||Ax + By_k - b||^2 - ||Ax + By_{k+1} - b||^2 \right) + \frac{1}{2\beta} ||\lambda_{k+1} - \lambda_k||^2
\]
(26)

and
\[
\langle \theta'_1(x_k, \xi_{k+1}), x_k - x_{k+1} \rangle \leq \frac{\eta_{k+1}}{2} ||\theta'_1(x_k, \xi_{k+1})||^2 + \frac{\eta_{k+1} ||\theta'_1(x_k, \xi_{k+1})||^2}{2} + \langle \delta_{k+1}, x - x_k \rangle
\]
(27)

where the last step is due to Young’s inequality. Inserting (26) and (27) into (25), we have
\[
\begin{align*}
\theta_1(x_k) &- \theta_1(x) + \langle x_{k+1} - x, -A^T \lambda_{k+1} \rangle \\
& \leq \frac{1}{2\eta_{k+1}} (||x_k - x||^2 - ||x_{k+1} - x||^2) + \frac{\eta_{k+1}}{2} ||\theta'_1(x_k, \xi_{k+1})||^2 + \langle \delta_{k+1}, x - x_k \rangle + \\
& + \frac{\beta}{2} \left( ||Ax + By_k - b||^2 - ||Ax + By_{k+1} - b||^2 \right) + \frac{1}{2\beta} ||\lambda_{k+1} - \lambda_k||^2,
\end{align*}
\]

(28)
Due to the optimality condition of Line 2 in Alg. 2 and the convexity of $\theta_2$, we have

$$\theta_2(y_{k+1}) - \theta_2(y) + \left(y_{k+1} - y, -B^T \lambda_{k+1}\right) \leq 0. \tag{29}$$

Using Line 3 in Alg. 2, we have

$$(\lambda_{k+1} - \lambda, Ax_{k+1} + By_{k+1} - b)$$

$$= \frac{1}{\beta} \left(\lambda_{k+1} - \lambda, \lambda_k - \lambda_{k+1}\right)$$

$$= \frac{1}{2\beta} \left(\|\lambda - \lambda_k\|^2 - \|\lambda - \lambda_{k+1}\|^2 - \|\lambda_{k+1} - \lambda_k\|^2\right) \tag{30}$$

Taking the summation of inequalities (28) (29) and (30), we obtain the result as desired. \qed

7.3. Proof of Theorem 1

Proof. (i). Invoking convexity of $\theta_1(\cdot)$ and $\theta_2(\cdot)$ and the monotonicity of operator $F(\cdot)$, we have $\forall w \in \mathcal{W}$:

$$\theta(u_t) - \theta(u) + (w_t - w)^T F(w_t) = \frac{1}{t} \sum_{k=1}^t \left[\theta_1(x_{k-1}) + \theta_2(y_{k-1}) - \theta(u) + (w_{k-1} - w)^T F(w_{k-1})\right]$$

$$= \frac{1}{t} \sum_{k=0}^{t-1} \left[\theta_1(x_k) + \theta_2(y_{k+1}) - \theta(u) + (w_{k+1} - w)^T F(w_{k+1})\right] \tag{31}$$

Applying Lemma 2 at the optimal solution $(x, y) = (x_*, y_*)$, we can derive from (31) that, $\forall \lambda$

$$\theta(u_t) - \theta(u) + (\tilde{x}_t - x_*)^T (-A^T \tilde{\lambda}_t) + (\tilde{y}_t - y_*)^T (-B^T \tilde{\lambda}_t) + (\tilde{\lambda}_t - \lambda)^T (A\tilde{x}_t + B\tilde{y}_t - b)$$

$$\leq \frac{1}{t} \sum_{k=0}^{t-1} \left[\eta_{k+1} \|\theta_1'(x_k, \xi_{k+1})\|^2\right] + \frac{1}{2t \eta_{k+1}} \left(\|x_k - x_*\|^2 - \|x_{k+1} - x_*\|^2 + \alpha_k^1 \|x_k - x_*\|^2\right)$$

$$+ \frac{1}{t} \left(\frac{D^2}{2t^2} + \frac{1}{2} \lambda_t \|\lambda - \lambda_0\|^2\right) \tag{32}$$

The above inequality is true for all $\lambda \in \mathbb{R}^m$, hence it also holds in the ball $\mathcal{B}_0 = \{\lambda : \|\lambda\|_2 \leq \rho\}$. Combing with the fact that the optimal solution must also be feasible, it follows that

$$\max_{\lambda \in \mathcal{B}_0} \left\{\theta(u_t) - \theta(u) + (\tilde{x}_t - x_*)^T (-A^T \tilde{\lambda}_t) + (\tilde{y}_t - y_*)^T (-B^T \tilde{\lambda}_t) + (\tilde{\lambda}_t - \lambda)^T (A\tilde{x}_t + B\tilde{y}_t - b)\right\}$$

$$= \max_{\lambda \in \mathcal{B}_0} \left\{\theta(u_t) - \lambda^T (A\tilde{x}_t + B\tilde{y}_t - b)\right\} - \lambda^T (A\tilde{x}_t + B\tilde{y}_t - b)$$

$$= \max_{\lambda \in \mathcal{B}_0} \left\{\theta(u_t) - \lambda^T (A\tilde{x}_t + B\tilde{y}_t - b)\right\}$$

$$= \theta(u_t) - \theta(u) + \rho \|A\tilde{x}_t + B\tilde{y}_t - b\|_2 \tag{33}$$
Taking an expectation over (33) and using (32) we have:

\[
E[\theta(u_t) - \theta(u_0) + \rho \|Ax_t + By_t - b\|]
\leq E \left[ \frac{1}{t} \sum_{k=0}^{t-1} \left( \eta_{k+1} \|\theta'_1(x_k, \xi_{k+1})\|^2 + \langle \delta_{k+1}, x_* - x_k \rangle \right) + \frac{1}{t} \left( \frac{D_x^2}{2t} + \frac{1}{2} \beta^2 \|y_* - b\| \right) \right]
+ E \left[ \max_{\lambda \in \mathcal{B}} \left\{ \frac{1}{2\beta t} \|\lambda - \lambda_0\|^2 \right\} \right]
\leq \frac{1}{t} \left( \frac{M^2}{2} \sum_{k=1}^{t} \eta_k + \frac{D_x^2}{2t} \right) + \frac{\beta D^2 y_* - b}{2t} + \frac{\rho^2}{2\beta t} + \frac{1}{t} \sum_{k=0}^{t-1} \left( \|\delta_{k+1}, x_* - x_k \| \right)
\leq \frac{\sqrt{2D_x M}}{\sqrt{t}} + \frac{\beta D^2 y_* - b}{2t} + \frac{\rho^2}{2\beta t}
\]

In the second last step, we use the fact that \(x_k\) is independent of \(\xi_{k+1}\), hence \(E_{\xi_{k+1}} (\xi_{k+1}, x_* - x_k) = \langle E_{\xi_{k+1}} (\delta_{k+1}, x_* - x_k) \rangle = 0\).

(ii) From the steps in the proof of part (i), it follows that,

\[
\theta(u_0) - \theta(u_t) + \rho \|Ax_t + By_t - b\|
\leq \frac{1}{t} \sum_{k=0}^{t-1} \eta_{k+1} \|\theta'_1(x_k, \xi_{k+1})\|^2 + \frac{1}{t} \sum_{k=0}^{t-1} \langle \delta_{k+1}, x_* - x_k \rangle + \frac{1}{t} \left( \frac{D_x^2}{2t} + \frac{\beta^2}{2} \|y_* - b\| \right) \]
\equiv A_t + B_t + C_t

Note that random variables \(A_t\) and \(B_t\) are dependent on \(\xi_0\).

**Claim 1.** For \(\Omega_1 > 0\),

\[
\text{Prob} \left( A_t \geq (1 + \Omega_1) \frac{M^2}{2t} \sum_{k=1}^{t} \eta_k \right) \leq \exp\{-\Omega_1\}. \tag{35}
\]

Let \(\alpha_k \equiv \frac{\eta_k}{\sum_{k=1}^{t} \eta_k} \quad \forall k = 1, \ldots, t\), then \(0 \leq \alpha_k \leq 1\) and \(\sum_{k=1}^{t} \alpha_k = 1\). Using the fact that \(\{\delta_k, \forall k\}\) are independent and applying Assumption 2, one has

\[
E \left[ \exp \left\{ \sum_{k=1}^{t} \alpha_k \|\theta'_1(x_k, \xi_{k+1})\|^2 / M^2 \right\} \right] = \prod_{k=1}^{t} E \left[ \exp \left\{ \alpha_k \|\theta'_1(x_k, \xi_{k+1})\|^2 / M^2 \right\} \right]
\leq \prod_{k=1}^{t} \left( E \left[ \exp \left\{ \|\theta'_1(x_k, \xi_{k+1})\|^2 / M^2 \right\} \right] \right)^{\alpha_k} \quad \text{(Jensen's Inequality)}
\leq \prod_{k=1}^{t} \left( \exp\{1\} \right)^{\alpha_k} = \exp \left\{ \sum_{k=1}^{t} \alpha_k \right\} = \exp\{1\}
\]

Hence, by Markov's Inequality, we can get

\[
\text{Prob} \left( A_t \geq (1 + \Omega_1) \frac{M^2}{2t} \sum_{k=1}^{t} \eta_k \right) \leq \exp\{- (1 + \Omega_1) \} E \left[ \exp \left\{ \sum_{k=1}^{t} \alpha_k \|\theta'_1(x_k, \xi_{k+1})\|^2 / M^2 \right\} \right] \leq \exp\{-\Omega_1\}.
\]

We have therefore proved Claim 1.

**Claim 2.** For \(\Omega_2 > 0\),

\[
\text{Prob} \left( B_t > 2\Omega_2 \frac{D_x M}{\sqrt{t}} \right) \leq \exp\left\{ - \frac{\Omega_2^2}{4} \right\}. \tag{36}
\]

In order to prove this claim, we adopt the following facts in Nemirovski’s paper (Nemirovski et al., 2009).
Lemma 3. Given that for all $k = 1, \ldots, t$, $\zeta_k$ is a deterministic function of $\xi_{[k]}$ with $E[\zeta_k|\xi_{[k-1]}] = 0$ and $E[\exp(\zeta_k^2/\sigma_k^2)|\xi_{[k-1]}] \leq \exp(1)$, we have

(a) For $\gamma \geq 0$, $E[\exp(\gamma \zeta_k)|\xi_{[k-1]}] \leq \exp(\gamma^2 \sigma_k^2)$, $\forall k = 1, \ldots, t$

(b) Let $S_t = \sum_{k=1}^t \zeta_k$, then $Prob(S_t > \Omega \sqrt{\sum_{k=1}^t \sigma_k^2}) \leq \exp\left\{-\frac{\Omega^2}{4}\right\}$.

Using this result by setting $\zeta_k = \langle \delta_k, x_k - x_{k-1} \rangle$, $S_t = \sum_{k=1}^t \zeta_k$, and $\sigma_k = 2D_xM$, $\forall k$, we can verify that $E[\zeta_k|\xi_{[k-1]}] = 0$ and $E[\exp(\zeta_k^2/\sigma_k^2)|\xi_{[k-1]}] \leq \exp(1)$, since $|\zeta_k|^2 \leq \|x_k - x_{k-1}\|^2||\delta_k||^2 \leq D_x^2 (2||\theta'_k(x_k, \xi_{k-1})||^2 + 2M^2)$.

Implementing the above results, it follows that

$$Prob\left(S_t > 2\Omega_x D_x M \sqrt{t}\right) \leq \exp\left\{-\frac{\Omega^2}{4}\right\}.$$

Since $S_t = tB_t$, we have

$$Prob\left(B_t > 2\Omega_x D_x M \sqrt{t}\right) \leq \exp\left\{-\frac{\Omega^2}{4}\right\}$$

as desired.

Combining (34), (35) and (36), we obtain

$$Prob\left(\left|\text{Err}_p(\tilde{u}_t) - (1 + \Omega_1) M^2 \frac{t}{2t} \sum_{k=1}^t \eta_k + 2\Omega_x D_x M \sqrt{t} + C_t\right| \leq \exp\left\{-\Omega_1\right\} + \exp\left\{-\frac{\Omega_2}{4}\right\}\right),$$

where $\text{Err}_p(\tilde{u}_t) \equiv \theta(\tilde{u}_t) - \theta(\tilde{u}_*) + \rho\|A\tilde{x}_t + BY_t - b\|_2$. Substituting $\Omega_1 = \Omega, \Omega_2 = 2\sqrt{\Omega}$ and plugging in $\eta_k = \frac{D_x}{3\sqrt{2k}}$, we obtain (10) as desired.

7.4. Proof of Theorem 2

Proof. By the strong-convexity of $\theta_1$ we have $\forall x$:

$$\theta_1(x_k) - \theta_1(x) \leq \langle \theta'_1(x_k), x_k - x \rangle - \frac{\mu}{2}\|x - x_k\|^2$$

$$= \langle \theta'_1(x_k, \xi_{k+1}), x_{k+1} - x \rangle + \langle \delta_{k+1}, x - x_k \rangle + \langle \theta'_1(x_k, \xi_{k+1}), x_k - x_{k+1} \rangle - \frac{\mu}{2}\|x - x_k\|^2.$$

Following the same derivations as in Lemma 2 and Theorem 1 (i), we have

$$E[\theta(\tilde{u}_t) - \theta(\tilde{u}_*) + \rho\|A\tilde{x}_t + BY_t - b\|_2]$$

$$\leq E\left\{\frac{1}{t} \sum_{k=1}^{t-1} \frac{\eta_k}{\lambda_k} ||\theta'_1(x_k, \xi_{k+1})||^2 + \left(\frac{1}{2\eta_{k+1}} - \frac{\mu}{2}\right) \|x_k - x_*\|^2 - \frac{\|x_{k+1} - x_*\|^2}{2\eta_{k+1}}\right\}$$

$$+ \frac{\beta D_x^2 \mu \rho}{2t} + E\left[\max_{\lambda \in \mathcal{B}_0} \left\{\frac{1}{2\beta t}\|\lambda - \lambda_0\|_0^2\right\}\right]$$

$$\leq M^2 \frac{1}{2t} \sum_{k=0}^{t-1} \frac{1}{\mu} + \frac{1}{t} \sum_{k=0}^{t-1} \frac{\mu k}{2} ||x_k - x_*||^2 - \frac{\mu (k + 1)}{2} ||x_{k+1} - x_*||^2 + \frac{\beta D_x^2 \mu \rho}{2t} + \frac{\rho^2}{2\beta t}$$

$$\leq \frac{M^2 \log t}{\mu t} + \frac{\mu D_x^2 \rho}{2t} + \frac{\beta D_x^2 \mu \rho}{2t} + \frac{\rho^2}{2\beta t}.$$

7.5. Proof of Theorem 3

Proof. The Lipschitz smoothness of $\theta_1$ implies that $\forall k \geq 0$:

$$\theta_1(x_{k+1}) \leq \theta_1(x_k) + \langle \nabla \theta_1(x_k), x_{k+1} - x_k \rangle + \frac{L}{2} ||x_{k+1} - x_k||^2$$

$$\overset{(2)}{=} \theta_1(x_k) + \langle \nabla \theta_1(x_k, \xi_{k+1}), x_{k+1} - x_k \rangle - \langle \delta_{k+1}, x_{k+1} - x_k \rangle + \frac{L}{2} ||x_{k+1} - x_k||^2.$$
It follows that $\forall x \in \mathcal{X}$:
\[
\theta_1(x_{k+1}) - \theta_1(x) + \left\langle x_{k+1} - x, -A^T\lambda_{k+1} \right\rangle \\
\leq \theta_1(x_k) - \theta_1(x) + \langle \nabla \theta_1(x_k, \xi_{k+1}), x_{k+1} - x_k \rangle - \langle \delta_{k+1}, x_{k+1} - x_k \rangle + \frac{L}{2}||x_{k+1} - x_k||^2 + \left\langle x_{k+1} - x, -A^T\lambda_{k+1} \right\rangle \\
= \theta_1(x_k) - \theta_1(x) + \langle \nabla \theta_1(x_k, \xi_{k+1}), x - x_k \rangle - \langle \delta_{k+1}, x_{k+1} - x_k \rangle + \frac{L}{2}||x_{k+1} - x_k||^2 \\
\quad + \left[ \nabla \theta_1(x_k, \xi_{k+1}), x_{k+1} - x \right] + \left\langle x_{k+1} - x, -A^T\lambda_{k+1} \right\rangle \\
\leq \langle \nabla \theta_1(x_k), x_k - x \rangle + \langle \nabla \theta_1(x_k, \xi_{k+1}), x - x_k \rangle - \langle \delta_{k+1}, x_{k+1} - x_k \rangle + \frac{L}{2}||x_{k+1} - x_k||^2 \\
\quad + \left[ \nabla \theta_1(x_k, \xi_{k+1}), x_{k+1} - x \right] + \left\langle x_{k+1} - x, -A^T\lambda_{k+1} \right\rangle \\
= \langle \delta_{k+1}, x - x_{k+1} \rangle + \frac{L}{2}||x_{k+1} - x_k||^2 + \left[ \nabla \theta_1(x_k, \xi_{k+1}), x_{k+1} - x \right] + \left\langle x_{k+1} - x, -A^T\lambda_{k+1} \right\rangle \\
= \langle \delta_{k+1}, x - x_{k+1} \rangle + \frac{L}{2}||x_{k+1} - x_k||^2 + \left( x - x_{k+1}, \beta A^TB(y_k - y_{k+1}) \right) \\
\quad + \left\langle x_{k+1} - x, -A^T\lambda_{k+1} \right\rangle \\
\leq \frac{1}{2\eta_{k+1}} \left( \|x - x_k\|^2 - ||x - x_{k+1}||^2 \right) - \frac{1}{2\eta_{k+1} - L}||x_{k+1} - x_k||^2 \\
\quad + \left\langle x - x_{k+1}, \beta A^TB(y_k - y_{k+1}) \right\rangle + \langle \delta_{k+1}, x - x_{k+1} \rangle.
\]

The last inner product can be bounded as below using Young’s inequality, given that $\eta_{k+1} \leq \frac{1}{2}$:
\[
\langle \delta_{k+1}, x - x_{k+1} \rangle = \langle \delta_{k+1}, x - x_k \rangle + \langle \delta_{k+1}, x_k - x_{k+1} \rangle \\
\leq \langle \delta_{k+1}, x - x_k \rangle + \frac{1}{2(1/\eta_{k+1} - L)}||\delta_{k+1}||^2 + \frac{1}{2\eta_{k+1} - L}||x_k - x_{k+1}||^2.
\]

Combining this with inequalities \((26, 29)\) and \((30)\), we can get a similar statement as that of Lemma 2:
\[
\theta(u_{k+1}) - \theta(u) + (w_{k+1} - w)^T F(w_{k+1}) \leq \frac{||\delta_{k+1}||^2}{2(1/\eta_{k+1} - L)} \\
\quad + \frac{1}{2\eta_{k+1}} \left( \|x_k - x\|^2 - ||x_{k+1} - x||^2 \right) + \frac{\beta}{2} \left( \|Ax + By_k - b\|^2 - \|Ax + By_{k+1} - b\|^2 \right) \\
\quad + \langle \delta_{k+1}, x - x_k \rangle + \frac{1}{2\beta} \left( \|\lambda - \lambda_k\|^2 - \|\lambda - \lambda_{k+1}\|^2 \right).
\]

The rest of the proof are essentially the same as Theorem 1 (i), except that we use the new definition of $\bar{u}_k$ in \((12)\).