A. Supplementary Material

Lemma 6. *Consider $\mathbb{R}^d$ endowed with the standard inner product. For any convex set $\mathcal{W} \subset \mathbb{R}^d$ and the associated projection operator $\Pi_{\mathcal{W}}$, we have:*

$$\|\Pi_{\mathcal{W}}(a) - \Pi_{\mathcal{W}}(b)\| \leq \|a - b\|$$

For all $a, b \in \mathbb{R}^d$

Proof. By Lemma 3.1.4 in (Nesterov, 2013), we conclude:

$$\langle a - \Pi_{\mathcal{W}}(a), \Pi_{\mathcal{W}}(b) - \Pi_{\mathcal{W}}(a)\rangle \leq 0\,.$$

Similarly,

$$\langle b - \Pi_{\mathcal{W}}(b), \Pi_{\mathcal{W}}(a) - \Pi_{\mathcal{W}}(b)\rangle \leq 0\,.$$

Adding the equations above, we conclude:

$$\|\Pi_{\mathcal{W}}(a) - \Pi_{\mathcal{W}}(b)\|^2 \leq \langle a - b, \Pi_{\mathcal{W}}(a) - \Pi_{\mathcal{W}}(b)\rangle$$

Using Cauchy-Schwarz inequality on the RHS, we conclude the result. □

A.1. Proof of Theorem 2

We have chosen $\alpha_{k,i} = \alpha = \min\left(\frac{2}{L}, 4l\frac{\log nK}{\mu n K}\right)$. By definition: $x^k_{i+1} = \Pi_{\mathcal{W}}\left(x^k_i - \alpha\nabla f(x^k_i;\sigma_k(i+1))\right)$.

Taking norm squared and using Lemma 6

$$\begin{aligned}
&\|x^k_{i+1} - x^*\|^2 \\
&\leq \|x^k_i - x^*\|^2 - 2\alpha\langle\nabla f(x^k_i;\sigma_k(i+1)), x^k_i - x^*\rangle \\
&\quad + \alpha^2\|\nabla f(x^k_i;\sigma_k(i+1))\|^2 \\
&\leq \|x^k_i - x^*\|^2 - 2\alpha\langle\nabla f(x^k_i;\sigma_k(i+1)), x^k_i - x^*\rangle \\
&\quad + \alpha^2 G^2 \\
&\leq \|x^k_i - x^*\|^2 - 2\alpha\langle\nabla F(x^k_i), x^k_i - x^*\rangle \\
&\quad + 2\alpha\langle\nabla F(x^k_i) - \nabla f(x^k_i;\sigma_k(i+1)), x^k_i - x^*\rangle + \alpha^2 G^2 \\
&\leq \|x^k_i - x^*\|^2(1-\alpha\mu) - 2\alpha\left[F(x^k_i) - F(x^*)\right] \\
&\quad + 2\alpha R_{i,k} + \alpha^2 G^2 \qquad (13)
\end{aligned}$$

We have used strong convexity of $F(\cdot)$ in the fourth step. Here $R_{i,k} := \langle\nabla F(x^k_i) - \nabla f(x^k_i;\sigma_k(i+1)), x^k_i - x^*\rangle$. We will bound $\mathbb{E}[R_{i,k}]$.

Clearly,

$$\begin{aligned}
R_{i,k} = \frac{1}{n}\sum_{r=1}^{n}&\langle\nabla f(x^k_i;r), x^k_i - x^*\rangle \\
&- \langle\nabla f(x^k_i;\sigma_k(i+1)), x^k_i - x^*\rangle
\end{aligned}$$

Recall the definition of $\mathcal{D}_{i,k}$ and $\mathcal{D}^{(r)}_{i,k}$ from Section 4. Let $Y \sim \mathcal{D}_{i,k}$ and $Z_r \sim \mathcal{D}^{(r)}_{i,k}$, with any arbitrary coupling. Taking expecation in the expression for $R_{i,k}$, we have:

$$\begin{aligned}
\mathbb{E}[R_{i,k}] &= \frac{1}{n}\sum_{r=1}^{n}\mathbb{E}\left[\langle\nabla f(x^k_i;r), x^k_i - x^*\rangle\right] \\
&\quad - \frac{1}{n}\sum_{r=1}^{n}\mathbb{E}\left[\langle\nabla f(x^k_i;r), x^k_i - x^*\rangle\middle|\sigma_k(i+1) = r\right] \\
&= \frac{1}{n}\sum_{r=1}^{n}\mathbb{E}\left[\langle\nabla f(Y;r), Y - x^*\rangle - \langle\nabla f(Z_r;r), Z_r - x^*\rangle\right] \\
&= \frac{1}{n}\sum_{r=1}^{n}\mathbb{E}\Big[\langle\nabla f(Y;r) - \nabla f(Z_r;r), Y - x^*\rangle \\
&\qquad + \langle\nabla f(Z_r;r), Y - Z_r\rangle\Big] \\
&\leq \frac{1}{n}\sum_{r=1}^{n}\mathbb{E}[L\|Y - x^*\|.\|Z_r - Y\| + G\|Z_r - Y\|] \\
&\leq \frac{1}{n}\sum_{r=1}^{n} L\sqrt{\mathbb{E}[\|Y - x^*\|^2]}\sqrt{\mathbb{E}[\|Z_r - Y\|^2]} + G\mathbb{E}[\|Z_r - Y\|]
\end{aligned}$$

We have used smoothness of $f(;r)$ and Cauchy-Schwarz inequality in the fourth step and Cauchy-Schwarz inequality in the fifth step. Since the inequality above holds for every coupling between Y and Z_r, we conclude:

$$\begin{aligned}
\mathbb{E}[R_{i,k}] &\leq \frac{1}{n}\sum_{r=1}^{n} L\mathsf{D}^{(2)}_{\mathsf{W}}\left(\mathcal{D}_{i,k}, \mathcal{D}^{(r)}_{i,k}\right)\sqrt{\mathbb{E}[\|x^k_i - x^*\|^2]} \\
&\quad + G\mathsf{D}^{(2)}_{\mathsf{W}}\left(\mathcal{D}_{i,k}, \mathcal{D}^{(r)}_{i,k}\right) \\
&\leq \frac{1}{n}\sum_{r=1}^{n}\frac{L^2}{\mu}\left[\mathsf{D}^{(2)}_{\mathsf{W}}\left(\mathcal{D}_{i,k}, \mathcal{D}^{(r)}_{i,k}\right)\right]^2 + \frac{\mu}{4}\mathbb{E}[\|x^k_i - x^*\|^2] \\
&\quad + G\mathsf{D}^{(2)}_{\mathsf{W}}\left(\mathcal{D}_{i,k}, \mathcal{D}^{(r)}_{i,k}\right) \qquad (14)
\end{aligned}$$

by our hypothesis we have $\alpha \leq \frac{2}{L}$. So we can apply Lemma 4. Equation (14) along with equation (13) implies:

$$\begin{aligned}
&\mathbb{E}\|x^k_{i+1} - x^*\|^2 \\
&\leq \mathbb{E}\|x^k_i - x^*\|^2(1-\alpha\mu) - 2\alpha\mathbb{E}\left[F(x^k_i) - F(x^*)\right] \\
&\quad + 2\alpha\mathbb{E}R_{i,1} + \alpha^2 G^2 \\
&\leq \mathbb{E}[\|x^k_i - x^*\|^2]\left(1 - \tfrac{\alpha\mu}{2}\right) - 2\alpha\mathbb{E}\left[F(x^k_i) - F(x^*)\right] \\
&\quad 3G^2\alpha^2 + \frac{4L^2G^2\alpha^3}{\mu}
\end{aligned}$$

We use the fact that $F(x^k_i) - F(x^*) \geq 0$ and unroll the recursion above to conclude:

$$\mathbb{E}[\|x_0^{k+1} - x^*\|^2] \leq \left(1 - \tfrac{\alpha\mu}{2}\right)^{nk} \|x_0^1 - x^*\|^2 + \sum_{t=0}^{\infty} \left(1 - \tfrac{\alpha\mu}{2}\right)^t \left[3G^2\alpha^2 + \tfrac{4L^2G^2\alpha^3}{\mu}\right]$$
$$= \left(1 - \tfrac{\alpha\mu}{2}\right)^{nk} \|x_0^1 - x^*\|^2 + \left[\tfrac{6G^2\alpha}{\mu} + \tfrac{8L^2G^2\alpha^2}{\mu^2}\right]$$
$$\leq e^{-\frac{n\alpha k\mu}{2}} \|x_0^1 - x^*\|^2 + \left[\tfrac{6G^2\alpha}{\mu} + \tfrac{8L^2G^2\alpha^2}{\mu^2}\right]$$

Using the fact that $\alpha = \min\left(\frac{2}{L}, 4l\frac{\log nK}{\mu nK}\right)$, we conclude that when $k \geq \frac{K}{2}$,

$$\mathbb{E}[\|x_0^{k+1} - x^*\|^2] \leq \frac{\|x_0^1 - x^*\|^2}{(nK)^l} + \left[\tfrac{6G^2\alpha}{\mu} + \tfrac{8L^2G^2\alpha^2}{\mu^2}\right] \tag{15}$$

We can easily verify that equation 12 also holds in this case (because all other assumptions hold). Therefore, for $k \geq \frac{K}{2}$,

$$\mathbb{E}[\|x_{i+1}^k - x^*\|^2] \leq \mathbb{E}[\|x_i^k - x^*\|^2] - 2\alpha\mathbb{E}[F(x_i^k) - F(x^*)] + 5\alpha^2G^2$$

Summing this equation for $0 \leq i \leq n-1$, $\lceil\frac{K}{2}\rceil \leq k \leq K$, we conclude:

$$\frac{1}{n(K - \lceil\frac{K}{2}\rceil + 1)} \sum_{k=\lceil\frac{K}{2}\rceil}^{K} \sum_{i=0}^{n-1} \mathbb{E}(F(x_i^k) - F(x^*))$$
$$\leq \frac{1}{2n\alpha(K - \lceil\frac{K}{2}\rceil + 1)} \mathbb{E}\left\|x_0^{\lceil\frac{K}{2}\rceil} - x^*\right\|^2 + \frac{5}{2}\alpha G^2$$
$$= O\left(\mu\tfrac{\|x_0^1 - x^*\|^2}{(nK)^l} + L\tfrac{\|x_0^1 - x^*\|^2}{(nK)^{(l+1)}}\right) + O\left(\tfrac{G^2\log nK}{\mu nK} + \tfrac{L^2G^2\log nK}{\mu^3n^2K^2}\right)$$

In the last step we have used Equation (15) and the fact that $\alpha \leq \frac{4l\log nK}{\mu nK}$ and $\frac{1}{\alpha} \leq \frac{L}{2} + \frac{nK\mu}{4l\log nK}$. Using convexity of F, we conclude that:

$$F(\hat{x}) \leq \frac{1}{n(K - \lceil\frac{K}{2}\rceil + 1)} \sum_{k=\lceil\frac{K}{2}\rceil}^{K} \sum_{i=0}^{n-1} F(x_i^k)\,.$$

This proves the result.

B. Proofs of useful lemmas

Proof of Lemma 2. For simplicity of notation, we denote $y_i \overset{\text{def}}{=} x_i(\sigma_k)$ and $z_i \overset{\text{def}}{=} x_i(\sigma'_k)$. We know that $\|y_0 - z_0\| = 0$ almost surely by definition. Let $j < i$. First we Suppose $\tau_y(j+1) = r \neq s = \tau_z(j+1)$. Then, by Lemma 6

$$\|y_{j+1} - z_{j+1}\|$$
$$= \big\|\Pi_{\mathcal{W}}\left(y_j - \alpha_{k,j}\nabla f(y_j; r)\right) - \Pi_{\mathcal{W}}\left(z_j - \alpha_{k,j}\nabla f(z_j; s)\right)\big\|$$
$$\leq \|y_j - z_j - \alpha_{k,j}\left(\nabla f(y_j; r) - \nabla f(z_j; s)\right)\|$$
$$\leq \|y_j - z_j\| + \alpha_{k,j}\|\nabla f(y_j; r)\| + \alpha_{k,j}\|\nabla f(z_j; s)\|$$
$$\leq 2G\alpha_{k,j} + \|y_j - z_j\|$$
$$\leq 2G\alpha_{k,0} + \|y_j - z_j\|$$

In the last step above, we have used monotonicity of α_t. Now, suppose $\tau_y(j+1) = \tau_z(j+1) = r$. Then,

$$\|y_{j+1} - z_{j+1}\|^2$$
$$= \big\|\Pi_{\mathcal{W}}\left(y_j - \alpha_{k,j}\nabla f(y_j; r)\right) - \Pi_{\mathcal{W}}\left(z_j - \alpha_{k,j}\nabla f(z_j; r)\right)\big\|^2$$
$$\leq \|\left(y_j - \alpha_{k,j}\nabla f(y_j; r)\right) - \left(z_j - \alpha_{k,j}\nabla f(z_j; r)\right)\|^2$$
$$= \|y_j - z_j\|^2 - 2\alpha_{k,i}\langle\nabla f(y_j; r) - \nabla f(z_j; r), y_j - z_j\rangle + \alpha_{k,j}^2\|\nabla f(y_j; r) - \nabla f(z_j; r)\|^2$$
$$\leq \|y_j - z_j\|^2 - (2\alpha_{k,j} - L\alpha_{k,j}^2)\langle\nabla f(y_j; r) - \nabla f(z_j; r), y_j - z_j\rangle$$
$$\leq \|y_j - z_j\|^2$$

In the second equation we have used Lemma 3 and in the third equation we have used the fact that when $\alpha_{k,0} \leq \frac{2}{L}$, $2\alpha_{k,i} - L\alpha_{k,i}^2 \geq 0$ and $\langle\nabla f(y_i; r) - \nabla f(z_i; r), y_i - z_i\rangle \geq 0$ by convexity. This proves the lemma. □

Proof of Lemma 5. For the sake of clarity of notation, in this proof we take $R_j := \sigma_k(j)$ for all $j \in [n]$. By defintion, $x_{j+1}^k - x_0^k = \Pi_{\mathcal{W}}\left(x_j^k - \alpha_{k,j}\nabla f(x_j^k; R_{j+1})\right) - x_0^k$. Taking norm squared on both sides, we have:

$$\|x_{j+1}^k - x_0^k\|^2$$
$$\leq \|x_j^k - x_0^k\|^2 - 2\alpha_{k,j}\langle f(x_j^k; R_{j+1}, x_j^k - x_0^k\rangle + \alpha_{k,j}^2G^2$$
$$\leq \|x_j^k - x_0^k\|^2 + 2\alpha_{k,j}\left(f(x_0^k; R_{j+1}) - f(x_j^k; R_{j+1})\right) + \alpha_{k,j}^2G^2$$

Taking expectation on both sides, we have:

$$
\begin{aligned}
&\mathbb{E}[\|x_{j+1}^k - x_0^k\|^2] \\
&\leq \mathbb{E}[\|x_j^k - x_0^k\|^2] + \alpha_{k,j}^2 G^2 \\
&\quad + 2\alpha_{k,j}\mathbb{E}\left[f(x_0^k; R_{j+1}) - f(x_j^k; R_{j+1})\right] \\
&= \mathbb{E}[\|x_j^k - x_0^k\|^2] + 2\alpha_{k,j}\mathbb{E}\left[F(x_0^k) - f(x_j^k; R_{j+1})\right] \\
&\quad + \alpha_{k,j}^2 G^2 \\
&= \mathbb{E}[\|x_j^k - x_0^k\|^2] + 2\alpha_{k,j}\mathbb{E}\left[F(x_0^k) - F(x_j^k)\right] \\
&\quad + 2\alpha_{k,j}\mathbb{E}\left[F(x_j^k) - f(x_j^k; R_{j+1})\right] + \alpha_{k,j}^2 G^2 \\
&\leq \mathbb{E}[\|x_j^k - x_0^k\|^2] + 2\alpha_{k,j}\mathbb{E}\left[F(x_0^k) - F(x_j^k)\right] \\
&\quad + 4\alpha_{k,j}\alpha_{k,0} G^2 + \alpha_{k,j}^2 G^2 \\
&\leq \mathbb{E}[\|x_j^k - x_0^k\|^2] + 2\alpha_{k,j}\mathbb{E}\left[F(x_0^k) - F(x^*)\right] \\
&\quad + 4\alpha_{k,j}\alpha_{k,0} G^2 + \alpha_{k,j}^2 G^2
\end{aligned}
$$

In the fourth step we have used Lemma 4 and in the fifth step, we have used the fact that x^* is the minimizer of F. We sum the equation above from $j = 0$ to $j = i - 1$ and use the fact that $\alpha_{k,0} \geq \alpha_{k,j}$ and that $\|x_j^k - x_0^k\| = 0$ when $j = 0$ to conclude the result. For the proof of the second equation in the lemma, we use x^* instead of x_0^k above and go through similar steps. □