A. Variance Reduction Theorem

Each of our results relies on a recent variance reduction technique, proposed by (Mokhtari et al., 2018a;b). We now present Theorem 3, which appears as Lemma 2 in (Mokhtari et al., 2018a). Although the proof is essentially the same, we present it here so that it is self-contained. When we apply Theorem 3 in the analysis of our algorithms, we will have that $\{\mathbf{a}_t\}$ are a sequence of gradients, $\{\tilde{\mathbf{a}}_t\}$ are stochastic gradient estimates, and $\{\mathbf{d}_t\}$ are the sequence of averaged gradient estimates. Moreover, the upper bound on the norm of the difference of gradients $\|\mathbf{a}_t - \mathbf{a}_{t-1}\|$ comes from the iterate update procedure and smoothness of the objective function.

Theorem 3. *Let $\{\mathbf{a}_t\}_{t=0}^T$ be a sequence of points in $\mathbb{R}^n$ such that $\|\mathbf{a}_t - \mathbf{a}_{t-1}\| \leq G/(t+s)$ for all $1 \leq t \leq T$ with fixed constants $G \geq 0$ and $s \geq 3$. Let $\{\tilde{\mathbf{a}}_t\}_{t=1}^T$ be a sequence of random variables such that $\mathbb{E}[\tilde{\mathbf{a}}_t|\mathcal{F}_{t-1}] = \mathbf{a}_t$ and $\mathbb{E}[\|\tilde{\mathbf{a}}_t - \mathbf{a}_t\|^2|\mathcal{F}_{t-1}] \leq \sigma^2$ for every $t \geq 0$, where $\mathcal{F}_{t-1}$ is the σ-field generated by $\{\tilde{\mathbf{a}}_i\}_{i=1}^t$ and $\mathcal{F}_0 = \varnothing$. Let $\{\mathbf{d}_t\}_{t=0}^T$ be a sequence of random variables where $\mathbf{d}_0$ is fixed and subsequent $\mathbf{d}_t$ are obtained by the recurrence*

$$\mathbf{d}_t = (1-\rho_t)\mathbf{d}_{t-1} + \rho_t \tilde{\mathbf{a}}_t$$

with $\rho_t = \frac{2}{(t+s)^{2/3}}$. Then, we have

$$\mathbb{E}[\|\mathbf{a}_t - \mathbf{d}_t\|^2] \leq \frac{Q}{(t+s+1)^{2/3}},$$

where $Q \triangleq \max\{\|\mathbf{a}_0 - \mathbf{d}_0\|^2(s+1)^{2/3}, 4\sigma^2 + 3G^2/2\}$.

We remark that we only need $s \geq 2^{3/2} \approx 2.83$ in the statement of Theorem 3.

Proof. Let $\Delta_t = \|\mathbf{a}_t - \mathbf{d}_t\|^2$. We have the following identity

$$\Delta_t = \|\rho_t(\mathbf{a}_t - \tilde{\mathbf{a}}_t) + (1-\rho_t)(\mathbf{a}_t - \mathbf{a}_{t-1}) + (1-\rho_t)(\mathbf{a}_{t-1} - \mathbf{d}_{t-1})\|^2.$$

Expanding the square and taking the expectation with respect to $\mathcal{F}_{t-1}$ gives

$$\mathbb{E}[\Delta_t|\mathcal{F}_{t-1}] \leq \rho_t^2\sigma^2 + (1-\rho_t)^2\frac{G^2}{(t+s)^2} + (1-\rho_t)^2\Delta_{t-1} + 2(1-\rho_t)^2\mathbb{E}[\langle \mathbf{a}_t - \mathbf{a}_{t-1}, \mathbf{a}_{t-1} - \mathbf{d}_{t-1}\rangle|\mathcal{F}_{t-1}].$$

Taking the expectation again gives

$$\mathbb{E}[\Delta_t] \leq \rho_t^2\sigma^2 + (1-\rho_t)^2\frac{G^2}{(t+s)^2} + (1-\rho_t)^2\mathbb{E}[\Delta_{t-1}] + 2(1-\rho_t)^2\mathbb{E}[\langle \mathbf{a}_t - \mathbf{a}_{t-1}, \mathbf{a}_{t-1} - \mathbf{d}_{t-1}\rangle].$$

By Young's inequality, we have

$$2\langle \mathbf{a}_t - \mathbf{a}_{t-1}, \mathbf{a}_{t-1} - \mathbf{d}_{t-1}\rangle \leq \beta_t\|\mathbf{a}_{t-1} - \mathbf{d}_{t-1}\|^2 + (1/\beta_t)\frac{G^2}{(t+s)^2}.$$

Therefore we deduce

$$\begin{aligned}\mathbb{E}[\Delta_t] &\leq \rho_t^2\sigma^2 + (1-\rho_t)^2\frac{G^2}{(t+s)^2} + (1-\rho_t)^2\mathbb{E}[\Delta_{t-1}] + (1-\rho_t)^2\left(\beta_t\mathbb{E}[\Delta_{t-1}] + (1/\beta_t)\frac{G^2}{(t+s)^2}\right)\\ &\leq \rho_t^2\sigma^2 + \frac{G^2}{(t+s)^2}(1-\rho_t)^2(1+\frac{1}{\beta_t}) + \mathbb{E}[\Delta_{t-1}](1-\rho_t)^2(1+\beta_t).\end{aligned}$$

We write z_t for $\mathbb{E}[\Delta_t]$. Notice that $(1-\rho_t)(1+\rho_t/2) \leq 1$ as long as $\rho_t \geq 0$. If we assume $\rho_t \in [0,1]$, setting $\beta_t = \rho_t/2$ yields

$$\begin{aligned}z_t &\leq \rho_t^2\sigma^2 + \frac{G^2}{(t+s)^2}(1-\rho_t)^2(1+\frac{2}{\rho_t}) + z_{t-1}(1-\rho_t)^2(1+\frac{\rho_t}{2})\\ &\leq \rho_t^2\sigma^2 + \frac{G^2}{(t+s)^2}(1+\frac{2}{\rho_t}) + z_{t-1}(1-\rho_t).\end{aligned}$$

We set $\rho_t = \frac{2}{(t+s)^{2/3}}$, where $s^{2/3} \geq 2$. Since $(t+s)^2 = (t+s)^{4/3}(t+s)^{2/3} \geq 2(t+s)^{4/3}$, we have

$$
\begin{aligned}
z_t &\leq (1 - \frac{2}{(t+s)^{2/3}})z_{t-1} + \frac{4\sigma^2}{(t+s)^{4/3}} + \frac{G^2}{(t+s)^2} + \frac{G^2}{(t+s)^{4/3}} \\
&\leq (1 - \frac{2}{(t+s)^{2/3}})z_{t-1} + \frac{4\sigma^2}{(t+s)^{4/3}} + \frac{3G^2}{2(t+s)^{4/3}} \\
&\leq (1 - \frac{2}{(t+s)^{2/3}})z_{t-1} + \frac{4\sigma^2 + 3G^2/2}{(t+s)^{4/3}} \\
&\leq (1 - \frac{2}{(t+s)^{2/3}})z_{t-1} + \frac{Q}{(t+s)^{4/3}}.
\end{aligned}
$$

We claim $z_t \leq \frac{Q}{(t+s+1)^{2/3}}$ for $\forall 0 \leq t \leq T$ and show this by induction. It holds for $t = 0$ due to the definition of Q. Now we assume that it is true for $t = k-1$. We have

$$
\begin{aligned}
z_k &\leq (1 - \frac{2}{(k+s)^{2/3}})z_{k-1} + \frac{Q}{(k+s)^{4/3}} \\
&\leq (1 - \frac{2}{(k+s)^{2/3}})\frac{Q}{(k+s)^{2/3}} + \frac{Q}{(k+s)^{4/3}} \\
&= Q\frac{(k+s)^{2/3} - 1}{(k+s)^{4/3}}.
\end{aligned}
$$

In order to show that $z_k \leq \frac{Q}{(k+s+1)^{2/3}}$, it suffices to show that

$$((k+s)^{2/3} - 1)(k+s+1)^{2/3} \leq (k+s)^{4/3}.$$

The above inequality holds since $(k+s+1)^{2/3} \leq (k+s)^{2/3} + 1$. □

B. Proof of Theorem 1: Convex Case

We begin by examining the sequence of iterates $\mathbf{x}_t^{(1)}, \mathbf{x}_t^{(2)}, \ldots, \mathbf{x}_t^{(K+1)}$ produced in Algorithm 1 for a fixed t. By definition of the update and because f_t is L-smooth, we have

$$
\begin{aligned}
f_t(\mathbf{x}_t^{(k+1)}) - f_t(\mathbf{x}^*) &= f_t(\mathbf{x}_t^{(k)} + \eta_k(\mathbf{v}_t^{(k)} - \mathbf{x}_t^{(k)})) - f_t(\mathbf{x}^*) \\
&\leq f_t(\mathbf{x}_t^{(k)}) - f_t(\mathbf{x}^*) + \eta_k\langle \nabla f_t(\mathbf{x}_t^{(k)}), \mathbf{v}_t^{(k)} - \mathbf{x}_t^{(k)}\rangle + \eta_k^2 \frac{L}{2}\|\mathbf{v}_t^{(k)} - \mathbf{x}_t^{(k)}\|^2 \\
&\leq f_t(\mathbf{x}_t^{(k)}) - f_t(\mathbf{x}^*) + \eta_k\langle \nabla f_t(\mathbf{x}_t^{(k)}), \mathbf{v}_t^{(k)} - \mathbf{x}_t^{(k)}\rangle + \eta_k^2 \frac{LD^2}{2}.
\end{aligned}
$$

Now, observe that the dual pairing may be decomposed as

$$\langle \nabla f_t(\mathbf{x}_t^{(k)}), \mathbf{v}_t^{(k)} - \mathbf{x}_t^{(k)}\rangle = \langle \nabla f_t(\mathbf{x}_t^{(k)}) - \mathbf{d}_t^{(k)}, \mathbf{v}_t^{(k)} - \mathbf{x}^*\rangle + \langle \nabla f_t(\mathbf{x}_t^{(k)}), \mathbf{x}^* - \mathbf{x}_t^{(k)}\rangle + \langle \mathbf{d}_t^{(k)}, \mathbf{v}_t^{(k)} - \mathbf{x}^*\rangle.$$

We can bound the first term using Young's Inequality to get

$$
\begin{aligned}
\langle \nabla f_t(\mathbf{x}_t^{(k)}) - \mathbf{d}_t^{(k)}, \mathbf{v}_t^{(k)} - \mathbf{x}^*\rangle &\leq \frac{1}{2\beta_k}\|f_t(\mathbf{x}_t^{(k)}) - \mathbf{d}_t^{(k)}\|^2 + 2\beta_k\|\mathbf{v}_t^{(k)} - \mathbf{x}^*\|^2 \\
&\leq \frac{1}{2\beta_k}\|f_t(\mathbf{x}_t^{(k)}) - \mathbf{d}_t^{(k)}\|^2 + 2\beta_k D^2
\end{aligned}
$$

for any $\beta_k > 0$, which will be chosen later in the proof. We may also bound the second term in the decomposition of the dual pairing using convexity of f_t, i.e. $\langle \nabla f_t(\mathbf{x}_t^{(k)}), \mathbf{x}^* - \mathbf{x}_t^{(k)} \rangle \le f_t(\mathbf{x}^*) - f_t(\mathbf{x}_t^{(k)})$. Using these upper bounds, we get that

$$\langle \nabla f_t(\mathbf{x}_t^{(k)}), \mathbf{v}_t^{(k)} - \mathbf{x}_t^{(k)} \rangle \le \frac{1}{2\beta_k} \| f_t(\mathbf{x}_t^{(k)}) - \mathbf{d}_t^{(k)} \|^2 + 2\beta_k D^2 + f_t(\mathbf{x}^*) - f_t(\mathbf{x}_t^{(k)}) + \langle \mathbf{d}_t^{(k)}, \mathbf{v}_t^{(k)} - \mathbf{x}^* \rangle.$$

Using this upper bound on the dual pairing in the first inequality, we get that

$$f_t(\mathbf{x}_t^{(k+1)}) - f_t(\mathbf{x}^*) \le (1-\eta_k)(f_t(\mathbf{x}_t^{(k)}) - f_t(\mathbf{x}^*)) + \eta_k \left[\frac{1}{2\beta_k} \| f_t(\mathbf{x}_t^{(k)}) - \mathbf{d}_t^{(k)} \|^2 + 2\beta_k D^2 + \langle \mathbf{d}_t^{(k)}, \mathbf{v}_t^{(k)} - \mathbf{x}^* \rangle + \eta_k \frac{LD^2}{2} \right].$$

Now we will apply the variance reduction technique. Note that

$$\| \nabla f_t(\mathbf{x}_t^{(k+1)} - \nabla f_t(\mathbf{x}_t^{(k)}) \| \le L \| \mathbf{x}_t^{(k+1)} - \mathbf{x}_t^{(k)} \| \le L\eta_k \| \mathbf{x}_t^{(k)} - \mathbf{v}_t^{(k)} \| \le \frac{LD}{k+3}$$

Where we have used that f_t is L-smooth, the convex update, and that the step size is $\eta_k = \frac{1}{k+3}$. Now, using Theorem 3 with $G = LD$ and $s = 3$, we have that

$$\mathbb{E}[\| f_t(\mathbf{x}_t^{(k)}) - \mathbf{d}_t^{(k)} \|^2] \le \frac{Q_t}{(k+4)^{2/3}} \le \frac{Q}{(k+4)^{2/3}}.$$

Where $Q_t \triangleq \max\{ \| \nabla f_t(\mathbf{x}_1) \|^2 4^{2/3}, 4\sigma^2 + 3(LD)^2/2 \}$ and $Q \triangleq \max\{ 4^{2/3} \max_{1 \le t \le T} \| \nabla f_t(\mathbf{x}_1) \|^2, 4\sigma^2 + 3(LD)^2/2 \}$ Thus, taking expectation of both sides of the optimality gap and setting $\beta_k = \frac{Q^{1/2}}{2D(k+4)^{1/3}}$ yields

$$\mathbb{E}[f_t(\mathbf{x}_t^{(k+1)})] - f_t(\mathbf{x}^*) \le (1-\eta_k)(\mathbb{E}[f_t(\mathbf{x}_t^{(k)})] - f_t(\mathbf{x}^*)) + \eta_k \left[\frac{2Q^{1/2}D}{(k+4)^{1/3}} + \langle \mathbf{d}_t^{(k)}, \mathbf{v}_t^{(k)} - \mathbf{x}^* \rangle + \eta_k \frac{LD^2}{2} \right].$$

Now we have obtained an upper bound on the expected optimality gap $\mathbb{E}[f_t(\mathbf{x}_t^{(k+1)})] - f_t(\mathbf{x}^*)$ in terms of the expected optimality gap $\mathbb{E}[f_t(\mathbf{x}_t^{(k)})] - f_t(\mathbf{x}^*)$ in the previous iteration. By induction on k, we get that the final iterate in the sequence, $\mathbf{x}_t \triangleq \mathbf{x}_t^{(K+1)}$, satisfies the following expected optimality gap

$$\mathbb{E}[f_t(\mathbf{x}_t)] - f_t(\mathbf{x}^*) \le \prod_{k=1}^{K} (1-\eta_k) \left[f_t(\mathbf{x}_1) - f_t(\mathbf{x}^*) \right] + \sum_{k=1}^{K} \eta_k \prod_{j=k+1}^{K} (1-\eta_j) \left[\frac{2Q^{1/2}D}{(k+4)^{1/3}} + \langle \mathbf{d}_t^{(k)}, \mathbf{v}_t^{(k)} - \mathbf{x}^* \rangle + \eta_k \frac{LD^2}{2} \right] \tag{2}$$

Recall that the Frank Wolfe step sizes are $\eta_k = \frac{1}{k+3}$. We may obtain upper bounds on product of the form $\prod_{k=r}^{K}(1-\eta_k)$ by

$$\prod_{k=r}^{K} (1-\eta_k) = \prod_{k=r}^{K} \left(1 - \frac{1}{k+3} \right) \le \exp\left(-\sum_{k=r}^{K} \frac{1}{x+3} \right) \le \exp\left(-\int_{x=r}^{K+1} \frac{1}{x+3} dx \right) = \frac{r+3}{K+4} \le \frac{r+3}{K}$$

Substituting step sizes $\eta_k = \frac{1}{k+3}$ into Eq (2) and using this upper bound yields

$$\mathbb{E}[f_t(\mathbf{x}_t)] - f_t(\mathbf{x}^*) \le \frac{4}{K} \left[f_t(\mathbf{x}_1) - f_t(\mathbf{x}^*) \right] + \sum_{k=1}^{K} \left(\frac{1}{k+3} \cdot \frac{k+4}{K} \right) \left[\frac{2Q^{1/2}D}{(k+4)^{1/3}} + \langle \mathbf{d}_t^{(k)}, \mathbf{v}_t^{(k)} - \mathbf{x}^* \rangle + \frac{LD^2}{2(k+3)} \right] \tag{3}$$

Which may be further simplified by using $\left(\frac{1}{k+3} \cdot \frac{k+4}{K} \right) \le \frac{4}{3K}$ to obtain

$$\mathbb{E}[f_t(\mathbf{x}_t)] - f_t(\mathbf{x}^*) \le \frac{4}{K} \left[f_t(\mathbf{x}_1) - f_t(\mathbf{x}^*) \right] + \frac{4}{3K} \sum_{k=1}^{K} \left[\frac{2Q^{1/2}D}{(k+3)^{1/3}} + \langle \mathbf{d}_t^{(k)}, \mathbf{v}_t^{(k)} - \mathbf{x}^* \rangle + \frac{LD^2}{2(k+3)} \right],$$

As before, we can obtain the following upper bounds using integral methods:

$$\sum_{k=1}^{K} \frac{1}{k+3} \leq \log\left(\frac{K+3}{3}\right) \leq \log(K+1) \quad \text{and} \quad \sum_{k=1}^{K} \frac{1}{(k+3)^{1/3}} \leq \frac{3}{2}\left((K+3)^{2/3} - 3^{2/3}\right) \leq \frac{3}{2}K^{2/3}$$

Substituting these bounds into Eq (3) yields

$$\mathbb{E}[f_t(\mathbf{x}_t)] - f_t(\mathbf{x}^*) \leq \frac{4}{K}\left[f_t(\mathbf{x}_1) - f_t(\mathbf{x}^*)\right] + \frac{4Q^{1/2}D}{K^{1/3}} + \frac{4LD^2\log(K+1)}{3K} + \frac{4}{3K}\sum_{k=1}^{K}\langle \mathbf{d}_t^{(k)}, \mathbf{v}_t^{(k)} - \mathbf{x}^*\rangle.$$

Now, we can begin to bound regret by summing over all $t = 1 \ldots T$ to obtain

$$\begin{aligned}\sum_{t=1}^{T} \mathbb{E}[f_t(\mathbf{x}_t)] - \sum_{t=1}^{T} f_t(\mathbf{x}^*) &\leq \frac{4}{K}\sum_{t=1}^{T}\left[f_t(\mathbf{x}_1) - f_t(\mathbf{x}^*)\right] + \frac{4TQ^{1/2}D}{K^{1/3}} \\ &\quad + \frac{4TLD^2\log(K+1)}{3K} + \frac{4}{3K}\sum_{t=1}^{T}\sum_{k=1}^{K}\langle \mathbf{d}_t^{(k)}, \mathbf{v}_t^{(k)} - \mathbf{x}^*\rangle\end{aligned}$$

Recall that for a fixed k, the sequence $\{\mathbf{v}_t^{(k)}\}_{t=1}^{T}$ is produced by a online linear minimization oracle with regret $\mathcal{R}_T^{\mathcal{E}}$ so that

$$\sum_{t=1}^{T}\langle \mathbf{d}_t^{(k)}, \mathbf{v}_t^{(k)} - \mathbf{x}^*\rangle \leq \sum_{t=1}^{T}\langle \mathbf{d}_t^{(k)}, \mathbf{v}_t^{(k)}\rangle - \min_{\mathbf{x}\in\mathcal{K}}\sum_{t=1}^{T}\langle \mathbf{d}_t^{(k)}, \mathbf{x}\rangle \leq \mathcal{R}_T^{\mathcal{E}}.$$

Substituting this into the upper bound and using $M = \max_{1\leq t\leq T}\left[f_t(\mathbf{x}_1) - f_t(\mathbf{x}^*)\right]$ yields

$$\sum_{t=1}^{T} \mathbb{E}[f_t(\mathbf{x}_t)] - \sum_{t=1}^{T} f_t(\mathbf{x}^*) \leq \frac{4TDQ^{1/2}}{K^{1/3}} + \frac{4T}{K}\left(M + \frac{LD^2}{3}\log(K+1)\right) + \frac{4}{3}\mathcal{R}_T^{\mathcal{E}}$$

Now, setting $K = T^{3/2}$ and using a linear oracle with $\mathcal{R}_T^{\mathcal{E}} = O(\sqrt{T})$ yields

$$\begin{aligned}\sum_{t=1}^{T} \mathbb{E}[f_t(\mathbf{x}_t)] - \sum_{t=1}^{T} f_t(\mathbf{x}^*) &\leq 4\sqrt{T}DQ^{1/2} + \frac{4}{\sqrt{T}}\left(M + \frac{LD^2}{3}(\log T^{3/2} + 1)\right) + \frac{4}{3}\mathcal{R}_T^{\mathcal{E}} \\ &= O(\sqrt{T}).\end{aligned}$$

C. Proof of Theorem 1: DR-Submodular Case

Using the smoothness of f_t and recalling $\mathbf{x}_t^{(k+1)} - \mathbf{x}_t^{(k)} = \frac{1}{K}\mathbf{v}_t^{(k)}$, we have

$$\begin{aligned}f_t(\mathbf{x}_t^{(k+1)}) &\geq f_t(\mathbf{x}_t^{(k)}) + \langle \nabla f_t(\mathbf{x}_t^{(k)}), \mathbf{x}_t^{(k+1)} - \mathbf{x}_t^{(k)}\rangle - \frac{L}{2}\|\mathbf{x}_t^{(k+1)} - \mathbf{x}_t^{(k)}\|^2 \\ &= f_t(\mathbf{x}_t^{(k)}) + \langle \frac{1}{K}\nabla f_t(\mathbf{x}_t^{(k)}), \mathbf{v}_t^{(k)}\rangle - \frac{L}{2K^2}\|\mathbf{v}_t^{(k)}\|^2 \\ &\geq f_t(\mathbf{x}_t^{(k)}) + \frac{1}{K}\langle \nabla f_t(\mathbf{x}_t^{(k)}), \mathbf{v}_t^{(k)}\rangle - \frac{LD^2}{2K^2}.\end{aligned} \tag{4}$$

We can re-write the term $\langle \nabla f_t(\mathbf{x}_t^{(k)}), \mathbf{v}_t^{(k)}\rangle$ as

$$\begin{aligned}\langle \nabla f_t(\mathbf{x}_t^{(k)}), \mathbf{v}_t^{(k)}\rangle &= \langle \nabla f_t(\mathbf{x}_t^{(k)}) - \mathbf{d}_t^{(k)}, \mathbf{v}_t^{(k)}\rangle + \langle \mathbf{d}_t^{(k)}, \mathbf{v}_t^{(k)}\rangle \\ &= \langle \nabla f_t(\mathbf{x}_t^{(k)}) - \mathbf{d}_t^{(k)}, \mathbf{v}_t^{(k)}\rangle + \langle \mathbf{d}_t^{(k)}, \mathbf{x}^*\rangle + \langle \mathbf{d}_t^{(k)}, \mathbf{v}_t^{(k)} - \mathbf{x}^*\rangle \\ &= \langle \nabla f_t(\mathbf{x}_t^{(k)}) - \mathbf{d}_t^{(k)}, \mathbf{v}_t^{(k)} - \mathbf{x}^*\rangle + \langle \nabla f_t(\mathbf{x}_t^{(k)}), \mathbf{x}^*\rangle + \langle \mathbf{d}_t^{(k)}, \mathbf{v}_t^{(k)} - \mathbf{x}^*\rangle.\end{aligned} \tag{5}$$

We claim $\langle \nabla f_t(\mathbf{x}_t^{(k)}), \mathbf{x}^* \rangle \geq f_t(\mathbf{x}^*) - f_t(\mathbf{x}_t^{(k)})$. Indeed, using monotonicity of f_t and concavity along non-negative directions, we have

$$\begin{aligned} f_t(\mathbf{x}^*) - f_t(\mathbf{x}_t^{(k)}) &\leq f_t(\mathbf{x}^* \vee \mathbf{x}_t^{(k)}) - f_t(\mathbf{x}_t^{(k)}) \\ &\leq \langle \nabla f_t(\mathbf{x}_t^{(k)}), \mathbf{x}^* \vee \mathbf{x}_t^{(k)} - \mathbf{x}_t^{(k)} \rangle \\ &= \langle \nabla f_t(\mathbf{x}_t^{(k)}), (\mathbf{x}^* - \mathbf{x}_t^{(k)}) \vee 0 \rangle \\ &\leq \langle \nabla f_t(\mathbf{x}_t^{(k)}), \mathbf{x}^* \rangle. \end{aligned} \tag{6}$$

Plugging Eq. (6) into Eq. (5), we obtain

$$\langle \nabla f_t(\mathbf{x}_t^{(k)}), \mathbf{v}_t^{(k)} \rangle \geq \langle \nabla f_t(\mathbf{x}_t^{(k)}) - \mathbf{d}_t^{(k)}, \mathbf{v}_t^{(k)} - \mathbf{x}^* \rangle + \langle \mathbf{d}_t^{(k)}, \mathbf{v}_t^{(k)} - \mathbf{x}^* \rangle + (f_t(\mathbf{x}^*) - f_t(\mathbf{x}_t^{(k)})). \tag{7}$$

Using Young's inequality, we can show that

$$\begin{aligned} \langle \nabla f_t(\mathbf{x}_t^{(k)}) - \mathbf{d}_t^{(k)}, \mathbf{v}_t^{(k)} - \mathbf{x}^* \rangle &\geq -\frac{1}{2\beta^{(k)}} \|\nabla f_t(\mathbf{x}_t^{(k)}) - \mathbf{d}_t^{(k)}\|^2 - \frac{\beta^{(k)}}{2} \|\mathbf{v}_t^{(k)} - \mathbf{x}^*\|^2 \\ &\geq -\frac{1}{2\beta^{(k)}} \|\nabla f_t(\mathbf{x}_t^{(k)}) - \mathbf{d}_t^{(k)}\|^2 - \beta^{(k)} D^2/2 \end{aligned} \tag{8}$$

Then we plug Eqs. (7) and (8) into Eq. (4), we deduce

$$f_t(\mathbf{x}_t^{(k+1)}) \geq f_t(\mathbf{x}_t^{(k)}) + \frac{1}{K} \left[-\frac{1}{2\beta^{(k)}} \|\nabla f_t(\mathbf{x}_t^{(k)}) - \mathbf{d}_t^{(k)}\|^2 - \beta^{(k)} D^2/2 + \langle \mathbf{d}_t^{(k)}, \mathbf{v}_t^{(k)} - \mathbf{x}^* \rangle + (f_t(\mathbf{x}^*) - f_t(\mathbf{x}_t^{(k)})) \right] - \frac{LD^2}{2K^2}.$$

Equivalently, we have

$$\begin{aligned} f_t(\mathbf{x}^*) - f_t(\mathbf{x}_t^{(k+1)}) &\leq (1 - 1/K)[f_t(\mathbf{x}^*) - f_t(\mathbf{x}_t^{(k)})] \\ &\quad - \frac{1}{K} \left[-\frac{1}{2\beta^{(k)}} \|\nabla f_t(\mathbf{x}_t^{(k)}) - \mathbf{d}_t^{(k)}\|^2 - \beta^{(k)} D^2/2 + \langle \mathbf{d}_t^{(k)}, \mathbf{v}_t^{(k)} - \mathbf{x}^* \rangle \right] + \frac{LD^2}{2K^2}. \end{aligned} \tag{9}$$

Applying Eq. (9) recursively for $1 \leq k \leq K$ immediately yields

$$\begin{aligned} f_t(\mathbf{x}^*) - f_t(\mathbf{x}_t^{(k+1)}) &\leq (1 - 1/K)^K [f_t(\mathbf{x}^*) - f_t(\mathbf{x}_t^{(1)})] \\ &\quad + \frac{1}{K} \sum_{k=1}^{K} \left[\frac{1}{2\beta^{(k)}} \|\nabla f_t(\mathbf{x}_t^{(k)}) - \mathbf{d}_t^{(k)}\|^2 + \beta^{(k)} D^2/2 + \langle \mathbf{d}_t^{(k)}, \mathbf{x}^* - \mathbf{v}_t^{(k)} \rangle \right] + \frac{LD^2}{2K}. \end{aligned}$$

Recall that the point played in round t is $\mathbf{x}_t \triangleq \mathbf{x}_t^{(K+1)}$, the first iterate in the sequence is $\mathbf{x}_t^{(1)} = 0$, and that $(1 - 1/K)^K \leq 1/e$ for all $K \geq 1$ so that

$$f_t(\mathbf{x}^*) - f_t(\mathbf{x}_t) \leq \frac{1}{e}[f_t(\mathbf{x}^*) - f_t(0)] + \frac{1}{K} \sum_{k=1}^{K} \left[\frac{1}{2\beta^{(k)}} \|\nabla f_t(\mathbf{x}_t^{(k)}) - \mathbf{d}_t^{(k)}\|^2 + \beta^{(k)} D^2/2 + \langle \mathbf{d}_t^{(k)}, \mathbf{x}^* - \mathbf{v}_t^{(k)} \rangle \right] + \frac{LD^2}{2K}.$$

Since $f_t(0) \geq 0$, we obtain

$$(1 - 1/e) f_t(\mathbf{x}^*) - f_t(\mathbf{x}_t) \leq \frac{1}{K} \sum_{k=1}^{K} \left[\frac{1}{2\beta^{(k)}} \|\nabla f_t(\mathbf{x}_t^{(k)}) - \mathbf{d}_t^{(k)}\|^2 + \beta^{(k)} D^2/2 + \langle \mathbf{d}_t^{(k)}, \mathbf{x}^* - \mathbf{v}_t^{(k)} \rangle \right] + \frac{LD^2}{2K}. \tag{10}$$

If we sum Eq. (10) over $t = 1, 2, 3, \ldots, T$, we obtain

$$(1-1/e)\sum_{t=1}^{T} f_t(\mathbf{x}^*) - \sum_{t=1}^{T} f_t(\mathbf{x}_t) \leq \frac{1}{K}\sum_{k=1}^{K}\left[\frac{1}{2\beta^{(k)}}\sum_{t=1}^{T}\|\nabla f_t(\mathbf{x}_t^{(k)}) - \mathbf{d}_t^{(k)}\|^2 + \beta^{(k)} D^2 T/2 + \sum_{t=1}^{T}\langle \mathbf{d}_t^{(k)}, \mathbf{x}^* - \mathbf{v}_t^{(k)}\rangle\right] + \frac{LD^2T}{2K}.$$

By the definition of the regret, we have

$$\sum_{t=1}^{T}\langle \mathbf{d}_t^{(k)}, \mathbf{x}^* - \mathbf{v}_t^{(k)}\rangle \leq \mathcal{R}_T^{\mathcal{E}}.$$

Therefore, we deduce

$$(1-1/e)\sum_{t=1}^{T} f_t(\mathbf{x}^*) - \sum_{t=1}^{T} f_t(\mathbf{x}_t) \leq \frac{1}{K}\sum_{k=1}^{K}\left[\frac{1}{2\beta^{(k)}}\sum_{t=1}^{T}\|\nabla f_t(\mathbf{x}_t^{(k)}) - \mathbf{d}_t^{(k)}\|^2 + \beta^{(k)} D^2 T/2\right] + \frac{LD^2T}{2K} + \mathcal{R}_T^{\mathcal{E}}.$$

Taking the expectation in both sides, we obtain

$$(1-1/e)\sum_{t=1}^{T} \mathbb{E}[f_t(\mathbf{x}^*)] - \sum_{t=1}^{T} \mathbb{E}[f_t(\mathbf{x}_t)] \leq \frac{1}{K}\sum_{k=1}^{K}\left[\frac{1}{2\beta^{(k)}}\sum_{t=1}^{T}\mathbb{E}[\|\nabla f_t(\mathbf{x}_t^{(k)}) - \mathbf{d}_t^{(k)}\|^2] + \beta^{(k)} D^2 T/2\right] + \frac{LD^2T}{2K} + \mathcal{R}_T^{\mathcal{E}}. \tag{11}$$

Notice that $\|\nabla f_t(\mathbf{x}_t^{(k)}) - \nabla f_t(\mathbf{x}_t^{(k-1)})\| \leq L\|\mathbf{v}_t^{(k)}\|/T \leq LR/T \leq 2LR/(k+3)$. By Theorem 3, if we set $\rho_k = \frac{2}{(k+3)^{2/3}}$, we have

$$\mathbb{E}[\|\nabla f_t(\mathbf{x}_t^{(k)}) - \mathbf{d}_t^{(k)}\|^2] \leq \frac{Q_t}{(k+4)^{2/3}} \leq \frac{Q}{(k+4)^{2/3}}, \tag{12}$$

where $Q_t \triangleq \max\{\|\nabla f_t(0)\|^2 4^{2/3}, 4\sigma^2 + 6L^2R^2\}$ and $Q \triangleq \max\{\max_{1\leq t\leq T}\|\nabla f_t(\mathbf{x}_1)\|^2 4^{2/3}, 4\sigma^2 + 6L^2R^2\}$. Plugging Eq. (12) into Eq. (11) and setting $\beta^{(k)} = (Q^{1/2})/(D(k+3)^{1/3})$, we deduce

$$(1-1/e)\sum_{t=1}^{T} \mathbb{E}[f_t(\mathbf{x}^*)] - \sum_{t=1}^{T} \mathbb{E}[f_t(\mathbf{x}_t)] \leq \frac{TDQ^{1/2}}{K}\sum_{k=1}^{K}\frac{1}{(k+4)^{1/3}} + \frac{LD^2T}{2K} + \mathcal{R}_T^{\mathcal{E}}$$

Since $\sum_{k=1}^{K}\frac{1}{(k+4)^{1/3}} \leq \int_0^K \frac{dx}{(x+4)^{1/3}} = \frac{3}{2}[(K+4)^{2/3} - 9^{2/3}] \leq \frac{3}{2}K^{2/3}$, we have

$$(1-1/e)\sum_{t=1}^{T} \mathbb{E}[f_t(\mathbf{x}^*)] - \sum_{t=1}^{T} \mathbb{E}[f_t(\mathbf{x}_t)] \leq \frac{3TDQ^{1/2}}{2K^{1/3}} + \frac{LD^2T}{2K} + \mathcal{R}_T^{\mathcal{E}}.$$

D. Proof of Theorem 2: Convex Case

Let $f(\mathbf{x}) = \mathbb{E}_{f_t \sim \mathcal{D}}[f_t(\mathbf{x})]$ denote the expected function. Because f is L-smooth and convex, we have

$$\begin{aligned} f(\mathbf{x}_{t+1}) - f(\mathbf{x}^*) &= f(\mathbf{x}_t + \eta_t(\mathbf{v}_t - \mathbf{x}_t)) - f(\mathbf{x}^*) \\ &\leq f(\mathbf{x}_t) - f(\mathbf{x}^*) + \eta_t \langle \nabla f(\mathbf{x}_t), \mathbf{v}_t - \mathbf{x}_t \rangle + \eta_t^2 \frac{L}{2} \|\mathbf{v}_t - \mathbf{x}_t\|^2 \\ &\leq f(\mathbf{x}_t) - f_t(\mathbf{x}^*) + \eta_t \langle \nabla f(\mathbf{x}_t), \mathbf{v}_t - \mathbf{x}_t \rangle + \eta_t^2 \frac{LD^2}{2}. \end{aligned}$$

As before, the dual pairing may be decomposed as

$$\langle \nabla f(\mathbf{x}_t), \mathbf{v}_t - \mathbf{x}_t \rangle = \langle \nabla f(\mathbf{x}_t) - \mathbf{d}_t, \mathbf{v}_t - \mathbf{x}^* \rangle + \langle \nabla f(\mathbf{x}_t), \mathbf{x}^* - \mathbf{x}_t \rangle + \langle \mathbf{d}_t, \mathbf{v}_t - \mathbf{x}^* \rangle.$$

We can bound the first term using Young's Inequality to get

$$\begin{aligned} \langle \nabla f(\mathbf{x}_t) - \mathbf{d}_t, \mathbf{v}_t - \mathbf{x}^* \rangle &\leq \frac{1}{2\beta} \|f(\mathbf{x}_t) - \mathbf{d}_t\|^2 + 2\beta \|\mathbf{v}_t - \mathbf{x}^*\|^2 \\ &\leq \frac{1}{2\beta} \|f(\mathbf{x}_t) - \mathbf{d}_t\|^2 + 2\beta D^2. \end{aligned}$$

for any $\beta > 0$, which will be chosen later in the proof. We may also bound the second term in the decomposition of the dual pairing using convexity of f, i.e. $\langle \nabla f(\mathbf{x}_t), \mathbf{x}^* - \mathbf{x}_t \rangle \leq f_t(\mathbf{x}^*) - f(\mathbf{x}_t)$. Finally, the third term is nonpositive, by the choice of $\mathbf{v}_t$, namely $\mathbf{v}_t = \arg\min_{\mathbf{v} \in \mathcal{K}} \langle \mathbf{d}_t, \mathbf{v} \rangle$. Using these inequalities, we now have that

$$f(\mathbf{x}_{t+1}) - f(\mathbf{x}^*) \leq (1 - \eta_t) \left(f(\mathbf{x}_t) - f(\mathbf{x}^*) \right) + \eta_t \left(\frac{1}{2\beta} \|f(\mathbf{x}_t) - \mathbf{d}_t\|^2 + 2\beta D^2 \right) + \eta_t^2 \frac{LD^2}{2}.$$

Taking expectation over the randomness in the iterates (i.e. the stochastic gradient estimates), we have that

$$\mathbb{E}[f(\mathbf{x}_{t+1})] - f(\mathbf{x}^*) \leq (1 - \eta_t) \left(\mathbb{E}[f(\mathbf{x}_t)] - f(\mathbf{x}^*) \right) + \eta_t \left(\frac{1}{2\beta} \mathbb{E}[\|f(\mathbf{x}_t) - \mathbf{d}_t\|^2] + 2\beta D^2 \right) + \eta_t^2 \frac{LD^2}{2}. \tag{13}$$

Now we will apply the variance reduction technique. Note that

$$\|\nabla f(\mathbf{x}_{t+1}) - \nabla f(\mathbf{x}_t)\| \leq L \|\mathbf{x}_{t+1} - \mathbf{x}_t\| \leq L\eta_t \|\mathbf{x}_t - \mathbf{v}_t\| \leq L\eta_t D$$

where we have used that f is L-smooth, the convex update, and the diameter. Now, using Theorem 3 with $G = LD$ and $s = 3$, we have that

$$\mathbb{E}[\|f(\mathbf{x}_t) - \mathbf{d}_t\|^2] \leq \frac{Q}{(t+4)^{2/3}},$$

where $Q \triangleq \max\{4^{2/3} \|\nabla f(\mathbf{x}_1)\|^2, 4\sigma^2 + 3(LD)^2/2\}$. Using this bound in Eq (13) and setting $\beta = \frac{Q^{1/2}}{2D(t+4)^{1/3}}$ yields

$$\mathbb{E}[f(\mathbf{x}_{t+1})] - f(\mathbf{x}^*) \leq (1 - \eta_t) \left(\mathbb{E}[f(\mathbf{x}_t)] - f(\mathbf{x}^*) \right) + \eta_t \frac{2Q^{1/2} D}{(t+4)^{1/3}} + \eta_t^2 \frac{LD^2}{2}.$$

By induction, we have

$$\mathbb{E}[f(\mathbf{x}_{t+1})] - f(\mathbf{x}^*) \leq \prod_{k=1}^{t} (1 - \eta_k) M + \sum_{k=1}^{t} \eta_k \prod_{j=k+1}^{t} (1 - \eta_j) \left(\frac{2Q^{1/2} D}{(k+4)^{1/3}} + \eta_k \frac{LD^2}{2} \right),$$

where $M = f(\mathbf{x}_1) - f(\mathbf{x}^*)$. Recall that the step size is set to be $\eta_t = \frac{1}{t+3}$. As in Appendix B, we can obtain the bounds $\prod_{k=1}^t (1-\eta_k) = \prod_{k=1}^t (1 - \frac{1}{k+3}) \le \exp(-\sum_{k=1}^t \frac{1}{k+3}) \le \exp(-\int_1^{t+1} \frac{dx}{x+3}) = 4/(t+4)$ and similarly $\prod_{j=k+1}^t (1 - \frac{1}{j+3}) \le \frac{k+4}{t+4}$. Using these bounds as well as the choice of step size $\eta_t = \frac{1}{t+3}$ in the above yields

$$\begin{aligned}
\mathbb{E}[f(\mathbf{x}_{t+1})] - f(\mathbf{x}^*) &\le \frac{4M}{t+4} + \sum_{k=1}^t \left(\frac{1}{k+3} \cdot \frac{k+4}{t+4}\right) \left(\frac{2Q^{1/2}D}{(k+4)^{1/3}} + \frac{1}{k+3}\frac{LD^2}{2}\right) \\
&= \frac{4M}{t+4} + \frac{4}{3(t+4)} \sum_{k=1}^t \left(\frac{2Q^{1/2}D}{(k+4)^{1/3}} + \frac{1}{k+3}\frac{LD^2}{2}\right)
\end{aligned}$$

where the second inequality used $\left(\frac{1}{k+3} \cdot \frac{k+4}{(t+4)}\right) < \frac{4}{3(t+4)}$. As before in Section B, using the inequalities $\sum_{k=1}^t \frac{1}{k+3} \le \log(t+1)$ and $\sum_{k=1}^t \frac{1}{(k+3)^{1/3}} \le \frac{3}{2} t^{2/3}$ in the above yields

$$\mathbb{E}[f(\mathbf{x}_{t+1})] - f(\mathbf{x}^*) \le \frac{4M}{t+4} + 4Q^{1/2}D\frac{t^{2/3}}{t+4} + \frac{4}{3}LD^2\frac{\log(t+1)}{t+4}. \tag{14}$$

To obtain a regret bound, we sum over rounds $t = 1, \dots T$ to obtain

$$\sum_{t=1}^T \mathbb{E}[f(\mathbf{x}_t)] - Tf(\mathbf{x}^*) \le 4M \left(\sum_{t=1}^T \frac{1}{t+4}\right) + 4Q^{1/2}D\left(\sum_{t=1}^T \frac{t^{2/3}}{t+4}\right) + \frac{4}{3}LD^2\left(\sum_{t=1}^T \frac{\log(t+1)}{t+4}\right)$$

Using the integral trick again, we obtain the upper bounds $\sum_{t=1}^T \frac{1}{t+4} \le \log(T+1)$, $\sum_{t=1}^T \frac{t^{2/3}}{t+4} \le \frac{3}{2}T^{2/3}$, and $\sum_{t=1}^T \frac{\log(t+3)}{t+4} \le \log^2(T+3)$. Substituting these bounds in the regret bound above yields

$$\sum_{t=1}^T \mathbb{E}[f(\mathbf{x}_t)] - Tf(\mathbf{x}^*) \le 4M\log(T+1) + 6Q^{1/2}DT^{2/3} + \frac{4}{3}LD^2\log^2(T+3) = O\left(T^{2/3}\right)$$

E. Proof of Theorem 2: DR-Submodular Case

Since f is L-smooth, we obtain

$$\begin{aligned}
f(\mathbf{x}_{t+1}) &\ge f(\mathbf{x}_t) + \langle \nabla f(\mathbf{x}_t), \frac{1}{T}\mathbf{v}_t\rangle - \frac{L}{2}\|\frac{1}{T}\mathbf{v}_t\|^2 \\
&\ge f(\mathbf{x}_t) + \frac{1}{T}\langle \nabla f(\mathbf{x}_t), \mathbf{v}_t\rangle - \frac{LD^2}{2T^2} \\
&= f(\mathbf{x}_t) + \frac{1}{T}\langle \mathbf{d}_t, \mathbf{v}_t\rangle + \frac{1}{T}\langle \nabla f(\mathbf{x}_t) - \mathbf{d}_t, \mathbf{v}_t\rangle - \frac{LD^2}{2T^2} \\
&\ge f(\mathbf{x}_t) + \frac{1}{T}\langle \mathbf{d}_t, \mathbf{x}^*\rangle + \frac{1}{T}\langle \nabla f(\mathbf{x}_t) - \mathbf{d}_t, \mathbf{v}_t\rangle - \frac{LD^2}{2T^2} \\
&= f(\mathbf{x}_t) + \frac{1}{T}\langle \nabla f(\mathbf{x}_t) - \mathbf{d}_t, \mathbf{v}_t - \mathbf{x}^*\rangle + \frac{1}{T}\langle f(\mathbf{x}_t), \mathbf{x}^*\rangle - \frac{LD^2}{2T^2}.
\end{aligned}$$

In the last inequality, we used the fact that $\mathbf{v}_t = \arg\max_{\mathbf{v}\in\mathcal{K}} \langle \mathbf{d}_t, \mathbf{v}\rangle$. Similar to Eq. (6) in Appendix C, we have $\langle f(\mathbf{x}_t), \mathbf{x}^*\rangle \ge f(\mathbf{x}^*) - f(\mathbf{x}_t)$. Again, Young's inequality gives $\langle \nabla f(\mathbf{x}_t) - \mathbf{d}_t, \mathbf{v}_t - \mathbf{x}^*\rangle \ge -\frac{1}{2}(\beta_t\|\mathbf{v}_t - \mathbf{x}^*\|^2 + \|f(\mathbf{x}_t) - \mathbf{d}_t\|^2/\beta_t)$. Therefore, we deduce

$$\begin{aligned}
f(\mathbf{x}_{t+1}) &\ge f(\mathbf{x}_t) - \frac{1}{2T}(\beta_t\|\mathbf{v}_t - \mathbf{x}^*\|^2 + \|f(\mathbf{x}_t) - \mathbf{d}_t\|^2/\beta_t) + \frac{1}{T}(f(\mathbf{x}^*) - f(\mathbf{x}_t)) - \frac{LD^2}{2T^2} \\
&\ge f(\mathbf{x}_t) - \frac{1}{2T}(\beta_t D^2 + \|f(\mathbf{x}_t) - \mathbf{d}_t\|^2/\beta_t) + \frac{1}{T}(f(\mathbf{x}^*) - f(\mathbf{x}_t)) - \frac{LD^2}{2T^2}.
\end{aligned}$$

Re-arrangement of the terms yields

$$f(\mathbf{x}^*) - f(\mathbf{x}_{t+1}) \le (1 - 1/T)(f(\mathbf{x}^*) - f(\mathbf{x}_t)) + \frac{1}{2T}(\beta_t D^2 + \|f(\mathbf{x}_t) - \mathbf{d}_t\|^2/\beta_t) + \frac{LD^2}{2T^2}.$$

Recalling that $(1 - 1/T)^T \le 1/e$ and $f(\mathbf{x}_1) = f(0) \ge 0$, we have

$$\begin{aligned} f(\mathbf{x}^*) - f(\mathbf{x}_{t+1}) &\le (1 - 1/T)^t (f(\mathbf{x}^*) - f(\mathbf{x}_1)) + \frac{1}{2T}\sum_{i=1}^{t}(\beta_i D^2 + \|f(\mathbf{x}_i) - \mathbf{d}_i\|^2/\beta_i) + \frac{LD^2}{2T} \\ &\le \frac{1}{e} f(\mathbf{x}^*) + \frac{1}{2T}\sum_{i=1}^{t}(\beta_i D^2 + \|f(\mathbf{x}_i) - \mathbf{d}_i\|^2/\beta_i) + \frac{LD^2}{2T}, \end{aligned}$$

which in turn yields

$$(1 - 1/e) f(\mathbf{x}^*) - f(\mathbf{x}_{t+1}) \le \frac{1}{2T}\sum_{i=1}^{t}(\beta_i D^2 + \|f(\mathbf{x}_i) - \mathbf{d}_i\|^2/\beta_i) + \frac{LD^2}{2T}.$$

Taking expectation in both sides gives

$$(1 - 1/e)\mathbb{E}[f(\mathbf{x}^*)] - \mathbb{E}[f(\mathbf{x}_{t+1})] \le \frac{1}{2T}\sum_{i=1}^{t}(\beta_i D^2 + \mathbb{E}[\|f(\mathbf{x}_i) - \mathbf{d}_i\|^2]/\beta_i) + \frac{LD^2}{2T}.$$

Notice that $\|\nabla f(\mathbf{x}_t) - \nabla f(\mathbf{x}_{t-1})\| \le L\|\mathbf{v}_t\|/T \le LR/T \le 2LR/(k+3)$. By Theorem 3, if we set $\rho_i = \frac{2}{(i+3)^{2/3}}$, we have

$$\mathbb{E}[\|f(\mathbf{x}_i) - \mathbf{d}_i\|^2] \le \frac{Q}{(i+4)^{2/3}}$$

for every $i \le T$ and $Q = \max\{\|\nabla f(0)\|^2 4^{2/3}, 4\sigma^2 + 6L^2R^2\}$. If we set $\beta_i = \frac{Q^{1/2}}{D(i+4)^{1/3}}$, we have

$$(1 - 1/e)\mathbb{E}[f(\mathbf{x}^*)] - \mathbb{E}[f(\mathbf{x}_{t+1})] \le \sum_{i=1}^{t}\frac{DQ^{1/2}}{(i+4)^{1/3}T} + \frac{LD^2}{2T} \le \frac{3DQ^{1/2}t^{2/3}}{2T} + \frac{LD^2}{2T}$$

since $\sum_{i=1}^{t}\frac{1}{(i+4)^{1/3}} \le \int_0^t \frac{1}{(x+4)^{1/3}}dx = \frac{3}{2}[(x+4)^{2/3}]_0^t \le \frac{3}{2}[x^{2/3}]_0^t = \frac{3}{2}t^{2/3}$.

Therefore we have

$$\begin{aligned} &(1 - 1/e)T\mathbb{E}[f(\mathbf{x}^*)] - \sum_{t=1}^{T}\mathbb{E}[f(\mathbf{x}_t)] \\ &= (1 - 1/e)\mathbb{E}[f(\mathbf{x}^*)] - f(0) + \sum_{t=1}^{T-1}[(1 - 1/e)\mathbb{E}[f(\mathbf{x}^*)] - \mathbb{E}[f(\mathbf{x}_t)]] \\ &\le (1 - 1/e)\mathbb{E}[f(\mathbf{x}^*)] - f(0) + \sum_{t=1}^{T-1}\left[\frac{3DQ^{1/2}t^{2/3}}{2T} + \frac{LD^2}{2T}\right]. \end{aligned}$$

Since $\sum_{t=1}^{T-1} t^{2/3} = 1 + \sum_{t=2}^{T-1} t^{2/3} \le 1 + \int_1^T t^{2/3}dt = \frac{3}{5}T^{5/3} + \frac{2}{5}$, we conclude

$$(1 - 1/e)T\mathbb{E}[f(\mathbf{x}^*)] - \sum_{t=1}^{T}\mathbb{E}[f(\mathbf{x}_t)] \le (1 - 1/e)\mathbb{E}[f(\mathbf{x}^*)] - f(0) + \frac{3DQ^{1/2}}{10}(3T^{2/3} + 2T^{-1}) + \frac{LD^2}{2} = O(T^{2/3}).$$