Appendix A Proof of Theorem 1

Before presenting the proof of Theorem 1, we need two lemmas first. Lemma 1 shows that a β-smooth function can be bounded by quadratic functions from above and below. Lemma 2 shows the concavity of continuous DR-submodular functions along non-negative and non-positive directions.

Lemma 1. *If f is β-smooth, then we have for any $\mathbf{x}$ and $\mathbf{y}$,*

$$\left|f(\mathbf{x}) - f(\mathbf{y}) - \nabla f(\mathbf{x})^\top (\mathbf{x}-\mathbf{y})\right| \le \frac{\beta}{2}\|\mathbf{x}-\mathbf{y}\|^2.$$

$$\left|f(\mathbf{x}) - f(\mathbf{y}) - \nabla f(\mathbf{y})^\top (\mathbf{x}-\mathbf{y})\right| \le \frac{\beta}{2}\|\mathbf{x}-\mathbf{y}\|^2.$$

Proof. Let us define an auxiliary function $g(t) = f(\mathbf{x}+t(\mathbf{y}-\mathbf{x}))$. We observe that $g(0) = f(\mathbf{x})$ and $g(1) = f(\mathbf{y})$. The derivative of $g(t)$ is

$$g'(t) = \nabla f(\mathbf{x}+t(\mathbf{y}-\mathbf{x}))^\top (\mathbf{y}-\mathbf{x}).$$

We have

$$f(\mathbf{y}) - f(\mathbf{x}) = g(1) - g(0) = \int_0^1 g'(t)dt = \int_0^1 \nabla f(\mathbf{x}+t(\mathbf{y}-\mathbf{x}))^\top (\mathbf{y}-\mathbf{x})dt.$$

The left-hand side of the first inequality is equal to

$$\begin{aligned}
&\left|\int_0^1 \nabla f(\mathbf{x}+t(\mathbf{y}-\mathbf{x}))^\top (\mathbf{x}-\mathbf{y})dt - \nabla f(\mathbf{x})^\top(\mathbf{x}-\mathbf{y})\right| \\
&= \left|\int_0^1 (\nabla f(\mathbf{x}+t(\mathbf{y}-\mathbf{x})) - \nabla f(\mathbf{x}))^\top (\mathbf{x}-\mathbf{y})\right| dt \\
&\le \int_0^1 \left|(\nabla f(\mathbf{x}+t(\mathbf{y}-\mathbf{x})) - \nabla f(\mathbf{x}))^\top (\mathbf{x}-\mathbf{y})\right| dt \\
&\le \int_0^1 \|\nabla f(\mathbf{x}+t(\mathbf{y}-\mathbf{x})) - \nabla f(\mathbf{x})\|\|\mathbf{x}-\mathbf{y}\| dt \\
&\le \int_0^1 \beta t\|\mathbf{x}-\mathbf{y}\|^2 dt \\
&= \frac{\beta}{2}\|\mathbf{x}-\mathbf{y}\|.
\end{aligned}$$

Exchanging $\mathbf{x}$ and $\mathbf{y}$ in the first inequality, we obtain the second one immediately. □

Lemma 2 (Proposition 4 in Bian et al. (2017))**.** *A continuous DR-submodular function is concave along any non-negative direction and any non-positive direction.*

Lemma 2 implies that if f is continuous DR-submodular, fixing any x in its domain, $g(z) \triangleq f(\mathbf{x}+z\mathbf{v})$ is concave in z as long as $\mathbf{v} \ge 0$ holds elementwise. Now we present the proof of Theorem 1.

Proof. As the first step, let us fix t and k. Since f_t is β-smooth, by Lemma 1, for any $\xi \ge 0$ and $\mathbf{x}, \mathbf{v} \in \mathbb{R}^n_{\ge 0}$, we have

$$f_t(\mathbf{x}+\xi\mathbf{v}) - f_t(\mathbf{x}) - \nabla f_t(\mathbf{x})^\top (\xi \mathbf{v}) \ge -\frac{\beta}{2}\|\xi\mathbf{v}\|^2$$

Let $L \triangleq \beta R^2$. We deduce

$$f_t(\mathbf{x}_t(k+1)) - f_t(\mathbf{x}_t(k)) = f_t(\mathbf{x}_t(k) + \frac{1}{K}\mathbf{v}_t^k) - f_t(\mathbf{x}_t(k)) \ge \frac{1}{K}\langle \mathbf{v}_t^k, \nabla f_t(\mathbf{x}_t(k))\rangle - \frac{L}{2K^2}.$$

We sum the above equation over t and obtain

$$\sum_{t=1}^{T} f_t(\mathbf{x}_t(k+1)) - f_t(\mathbf{x}_t(k)) \geq \sum_{t=1}^{T} \frac{1}{K}\langle \mathbf{v}_t^k, \nabla f_t(\mathbf{x}_t(k))\rangle - \frac{LT}{2K^2}.$$

The RFTL algorithm instance $\mathcal{E}^k$ finds $\{\mathbf{v}_t^k : 1 \leq t \leq T\}$ such that

$$\sum_{t=1}^{T}\langle \mathbf{v}^{k*}, \nabla f_t(\mathbf{x}_t(k))\rangle - \sum_{t=1}^{T}\langle \mathbf{v}_t^k, \nabla f_t(\mathbf{x}_t(k))\rangle \leq r^k \leq 2DG\sqrt{T},$$

where

$$\mathbf{v}^{k*} = \arg\max_{\mathbf{v}\in\mathcal{P}} \sum_{t=1}^{T}\langle \mathbf{v}, \nabla f_t(\mathbf{x}_t(k))\rangle$$

and r^k is the total regret that the RFTL instance suffers by the end of the Tth iteration. According to the regret bound of the RFTL, we know that $r^k \leq 2DG\sqrt{T}$. Therefore,

$$\sum_{t=1}^{T} f_t(\mathbf{x}_t(k+1)) - f_t(\mathbf{x}_t(k)) \geq \frac{1}{K}\left(\sum_{t=1}^{T}\langle \mathbf{v}^{k*}, \nabla f_t(\mathbf{x}_t(k))\rangle - r^k\right) - \frac{LT}{2K^2}.$$

We define $\mathbf{x}^* \triangleq \arg\max_{\mathbf{v}\in\mathcal{P}} \sum_{t=1}^{T} f_t(\mathbf{v})$ and $\mathbf{w}_t^k = (\mathbf{x}^* - \mathbf{x}_t(k)) \vee 0$. For every t, we have $\mathbf{w}_t^k = (\mathbf{x}^* - \mathbf{x}_t(k)) \vee 0 \leq \mathbf{x}^*$. It is obvious that $\mathbf{w}_t^k \geq 0$. Therefore we deduce that $\mathbf{w}_t^k \in \mathcal{X}$. Due to the concavity of f_t along any non-negative direction (see Lemma 2), we have

$$f_t(\mathbf{x}_t(k) + \mathbf{w}_t^k) - f_t(\mathbf{x}_t(k)) \leq \langle \mathbf{w}_t^k, \nabla f_t(\mathbf{x}_t(k))\rangle.$$

In light of the above equation, we obtain a lower bound for $\sum_{t=1}^{T}\langle \mathbf{v}^{k*}, \nabla f_t(\mathbf{x}_t(k))\rangle$:

$$\begin{aligned}
\sum_{t=1}^{T}\langle \mathbf{v}^{k*}, \nabla f_t(\mathbf{x}_t(k))\rangle &\geq \sum_{t=1}^{T}\langle \mathbf{x}^*, \nabla f_t(\mathbf{x}_t(k))\rangle \\
&\overset{(a)}{\geq} \sum_{t=1}^{T}\langle \mathbf{w}_t^k, \nabla f_t(\mathbf{x}_t(k))\rangle \\
&\geq \sum_{t=1}^{T}(f_t(\mathbf{x}_t(k) + \mathbf{w}_t^k) - f_t(\mathbf{x}_t(k))) \\
&= \sum_{t=1}^{T}(f_t(\mathbf{x}^* \vee \mathbf{x}_t(k)) - f_t(\mathbf{x}_t(k))) \\
&\geq \sum_{t=1}^{T}(f_t(\mathbf{x}^*) - f_t(\mathbf{x}_t(k))).
\end{aligned}$$

We use the fact that $\nabla f_t(\mathbf{x}_t(k)) \geq 0$ and $\mathbf{x}^* \geq \mathbf{w}_t^k$ entrywise in the inequality (a).

$$\sum_{t=1}^{T} f_t(\mathbf{x}_t(k+1)) - f_t(\mathbf{x}_t(k)) \geq \frac{1}{K}\left(\sum_{t=1}^{T}(f_t(\mathbf{x}^*) - f_t(\mathbf{x}_t(k))) - r^k\right) - \frac{LT}{2K^2}.$$

After rearrangement,

$$\sum_{t=1}^{T}\left(f_t(\mathbf{x}_t(k+1)) - f_t(\mathbf{x}^*)\right) \geq (1 - \frac{1}{K})\sum_{t=1}^{T}\left(f_t(\mathbf{x}_t(k)) - f_t(\mathbf{x}^*)\right) - \frac{1}{K}r^k - \frac{LT}{2K^2}.$$

Therefore,

$$\begin{aligned}\sum_{t=1}^{T}(f_t(\mathbf{x}_t(K))-f_t(\mathbf{x}^*)) &\geq (1-\frac{1}{K})^K\sum_{t=1}^{T}(f_t(\mathbf{x}_t(0))-f_t(\mathbf{x}^*))-\frac{1}{K}\sum_{k=0}^{K-1}r^k-\frac{LT}{2K}\\ &=(1-\frac{1}{K})^K\sum_{t=1}^{T}(f_t(0)-f_t(\mathbf{x}^*))-\frac{1}{K}\sum_{k=0}^{K-1}r^k-\frac{LT}{2K}.\end{aligned}$$

Since $(1-\frac{1}{K})^K \leq e^{-1}$, we have

$$\begin{aligned}\sum_{t=1}^{T}(f_t(\mathbf{x}^*)-f_t(\mathbf{x}_t(K))) &\leq (1-\frac{1}{K})^K\sum_{t=1}^{T}(f_t(\mathbf{x}^*)-f_t(0))+\frac{1}{K}\sum_{k=0}^{K-1}r^k+\frac{LT}{2K}\\ &\leq e^{-1}\sum_{t=1}^{T}(f_t(\mathbf{x}^*)-f_t(0))+\frac{1}{K}\sum_{k=0}^{K-1}r^k+\frac{LT}{2K}.\end{aligned}$$

After rearrangement, we have

$$\sum_{t=1}^{T}f_t(\mathbf{x}_t(K)) \geq (1-1/e)\sum_{t=1}^{T}f_t(\mathbf{x}^*)+e^{-1}\sum_{t=1}^{T}f_t(0)-\frac{1}{K}\sum_{k=0}^{K-1}r^k-\frac{LT}{2K^2}.$$

Plugging in the definition of r^k gives

$$\sum_{t=1}^{T}f_t(\mathbf{x}_t)=\sum_{t=1}^{T}f_t(\mathbf{x}_t(K)) \geq (1-1/e)\sum_{t=1}^{T}f_t(\mathbf{x}^*)+e^{-1}\sum_{t=1}^{T}f_t(0)-2DG\sqrt{T}-\frac{LT}{2K}.$$

Recall that $\mathbf{x}_t(K)$ is exactly $\mathbf{x}_t$. Thus equivalently, we have

$$(1-1/e)\sum_{t=1}^{T}f_t(\mathbf{x}^*)-\sum_{t=1}^{T}f_t(\mathbf{x}_t) \leq -e^{-1}\sum_{t=1}^{T}f_t(0)+2DG\sqrt{T}+\frac{\beta R^2 T}{2K}.$$

□

Appendix B Proof of Theorem 2

B.1 Gradient Ascent Case

The theoretical guarantee of gradient ascent methods applied to concave functions relies on a pivotal property that characterizes concavity: if F is concave, then $F(\mathbf{y})-F(\mathbf{x}) \leq \langle \nabla F(\mathbf{x}), \mathbf{y}-\mathbf{x}\rangle$. Fortunately, there is a similar property that holds for monotone weakly DR-submodular functions, which is presented in Lemma 3.

Lemma 3. *Let $F:\mathcal{X}\to\mathbb{R}_+$ be a monotone and weakly DR-submodular function with parameter γ. For any two vector $\mathbf{x},\mathbf{y}\in\mathcal{X}$, we have*

$$F(\mathbf{y})-\left(1+\frac{1}{\gamma^2}\right)F(\mathbf{x}) \leq \frac{1}{\gamma}\left\langle \nabla F(\mathbf{x}), \mathbf{y}-\mathbf{x}\right\rangle.$$

The proof of Lemma 3 can be found in the proof of Theorem 4.2 in (Hassani et al., 2017). Now we can prove Theorem 2 in the gradient ascent case.

Proof. Let $\mathbf{x}^* = \arg\max_{\mathbf{x}\in\mathcal{P}}\sum_{t=1}^{T}f_t(\mathbf{x})$. We define $\nabla_t \triangleq \nabla f_t(\mathbf{x}_t)$. By the definition of $\mathbf{x}_{t+1}$ and properties of the projection operator for a convex set, we have

$$\|\mathbf{x}_{t+1}-\mathbf{x}^*\|^2 = \|\Pi_{\mathcal{P}}(\mathbf{x}_t+\eta_t\nabla_t)-\mathbf{x}^*\|^2 \leq \|\mathbf{x}_t+\eta_t\nabla_t-\mathbf{x}^*\|^2 \leq \|\mathbf{x}_t-\mathbf{x}^*\|^2+\eta_t^2\|\nabla_t\|^2-2\eta_t\nabla_t^\top(\mathbf{x}^*-\mathbf{x}_t).$$

Therefore we deduce

$$\begin{aligned}\nabla_t^\top(\mathbf{x}^*-\mathbf{x}_t) &\le \frac{\|\mathbf{x}_t-\mathbf{x}^*\|^2-\|\mathbf{x}_{t+1}-\mathbf{x}^*\|^2+\eta_t^2\|\nabla_t\|^2}{2\eta_t}\\ &\le \frac{\|\mathbf{x}_t-\mathbf{x}^*\|^2-\|\mathbf{x}_{t+1}-\mathbf{x}^*\|^2}{2\eta_t}+\frac{\eta_t G^2}{2}\end{aligned}$$

By Lemma 3, we obtain that

$$f_t(\mathbf{x}^*)-\left(1+\frac{1}{\gamma^2}\right)f(\mathbf{x}_t)\le\frac{1}{\gamma}\langle\nabla_t,\mathbf{x}^*-\mathbf{x}_t\rangle.$$

If we define $\frac{1}{\eta_0}\triangleq 0$, it can be deduced that

$$\begin{aligned}\sum_{t=1}^T\left[f_t(\mathbf{x}^*)-\left(1+\frac{1}{\gamma^2}\right)f_t(\mathbf{x}_t)\right] &\le \frac{1}{\gamma}\sum_{t=1}^T\nabla_t^\top(\mathbf{x}^*-\mathbf{x}_t)\\ &\le \frac{1}{\gamma}\left[\frac{1}{2\eta_t}\sum_{t=1}^T\left(\|\mathbf{x}_t-\mathbf{x}^*\|^2-\|\mathbf{x}_{t+1}-\mathbf{x}^*\|^2\right)+\frac{G^2}{2}\sum_{t=1}^T\eta_t\right]\\ &\le \frac{1}{\gamma}\left[\frac{1}{2}\left(\sum_{t=1}^T\|\mathbf{x}_t-\mathbf{x}^*\|^2\left(\frac{1}{\eta_t}-\frac{1}{\eta_{t-1}}\right)\right)+\frac{G^2}{2}\sum_{t=1}^T\eta_t\right]\\ &\le \frac{1}{\gamma}\left[\frac{D^2}{2\eta_T}+\frac{G^2}{2}\sum_{t=1}^T\eta_t\right]\\ &\le \frac{3}{2\gamma}DG\sqrt{T}.\end{aligned}$$

After rearrangement, it is clear that

$$\frac{\gamma^2}{\gamma^2+1}\sum_{t=1}^T f_t(\mathbf{x}^*)-\sum_{t=1}^T f_t(\mathbf{x}_t)\le\frac{3\gamma DG\sqrt{T}}{2(\gamma^2+1)}.$$

□

B.2 Stochastic Gradient Ascent Case

Proof. The strategy for the stochastic gradient ascent case is similar to that of the gradient ascent case. Again, by the definition of $\mathbf{x}_{t+1}$, we have

$$\|\mathbf{x}_{t+1}-\mathbf{x}^*\|^2=\|\Pi_{\mathcal{P}}(\mathbf{x}_t+\eta_t\mathbf{g}_t)-\mathbf{x}^*\|^2\le\|\mathbf{x}_t+\eta_t\mathbf{g}_t-\mathbf{x}^*\|^2\le\|\mathbf{x}_t-\mathbf{x}^*\|^2+\eta_t^2\|\mathbf{g}_t\|^2-2\eta_t\mathbf{g}_t^\top(\mathbf{x}^*-\mathbf{x}_t)$$

Therefore we deduce

$$\mathbf{g}_t^\top(\mathbf{x}^*-\mathbf{x}_t)\le\frac{\|\mathbf{x}_t-\mathbf{x}^*\|^2-\|\mathbf{x}_{t+1}-\mathbf{x}^*\|^2+\eta_t^2\|\mathbf{g}_t\|^2}{2\eta_t}\le\frac{\|\mathbf{x}_t-\mathbf{x}^*\|^2-\|\mathbf{x}_{t+1}-\mathbf{x}^*\|^2}{2\eta_t}+\frac{\eta_t G^2}{2}$$

Similarly, if we define $\frac{1}{\eta_0} \triangleq 0$ and in light of Lemma 3, it can be deduced that

$$\begin{aligned}
\sum_{t=1}^{T} \mathbb{E}\left[f_t(\mathbf{x}^*) - \left(1 + \frac{1}{\gamma^2}\right) f_t(\mathbf{x}_t)\right] &\leq \frac{1}{\gamma} \sum_{t=1}^{T} \mathbb{E}\left[\nabla_t^\top (\mathbf{x}^* - \mathbf{x}_t)\right] \\
&= \frac{1}{\gamma} \sum_{t=1}^{T} \mathbb{E}\left[\mathbb{E}\left[\nabla_t^\top (\mathbf{x}^* - \mathbf{x}_t) | \mathbf{x}_t\right]\right] \\
&= \frac{1}{\gamma} \sum_{t=1}^{T} \mathbb{E}\left[\mathbb{E}\left[\mathbf{g}_t^\top (\mathbf{x}^* - \mathbf{x}_t) | \mathbf{x}_t\right]\right] \\
&\leq \frac{1}{\gamma} \left[\frac{1}{2\eta_t} \sum_{t=1}^{T} \mathbb{E}\left[\|\mathbf{x}_t - \mathbf{x}^*\|^2 - \|\mathbf{x}_{t+1} - \mathbf{x}^*\|^2\right] + \frac{G^2}{2} \sum_{t=1}^{T} \eta_t\right] \\
&\leq \frac{1}{\gamma} \left[\frac{1}{2} \left(\sum_{t=1}^{T} \mathbb{E}\left[\|\mathbf{x}_t - \mathbf{x}^*\|^2\right] \left(\frac{1}{\eta_t} - \frac{1}{\eta_{t-1}}\right)\right) + \frac{G^2}{2} \sum_{t=1}^{T} \eta_t\right] \\
&\leq \frac{1}{\gamma} \left[\frac{D^2}{2\eta_T} + \frac{G^2}{2} \sum_{t=1}^{T} \eta_t\right] \\
&\leq \frac{3}{2\gamma} DG\sqrt{T}.
\end{aligned}$$

After rearrangement, it is clear that

$$\frac{\gamma^2}{\gamma^2+1} \sum_{t=1}^{T} f_t(\mathbf{x}^*) - \sum_{t=1}^{T} \mathbb{E}\left[f_t(\mathbf{x}_t)\right] \leq \frac{3\gamma DG\sqrt{T}}{2(\gamma^2+1)}.$$

$\square$