A Proof of Lemma 1

Proof. Using the assumption that F is G-Lipschitz continuous, we have

$$\begin{aligned}|\tilde{F}(x)-\tilde{F}(y)| &= |\mathbb{E}_{v\sim B^d}[F(x+\delta v)-F(y+\delta v)]| &(15)\\ &\le \mathbb{E}_{v\sim B^d}[|F(x+\delta v)-F(y+\delta v)|] &(16)\\ &\le \mathbb{E}_{v\sim B^d}[G\|(x+\delta v)-(y+\delta v)\|] &(17)\\ &= G\|x-y\|, &(18)\end{aligned}$$

and

$$\begin{aligned}|\tilde{F}(x)-F(x)| &= |\mathbb{E}_{v\sim B^d}[F(x+\delta v)-F(x)]| &(19)\\ &\le \mathbb{E}_{v\sim B^d}[|F(x+\delta v)-F(x)|] &(20)\\ &\le \mathbb{E}_{v\sim B^d}[G\delta\|v\|] &(21)\\ &\le \delta G. &(22)\end{aligned}$$

If F is G-Lipschitz continuous and monotone continuous DR-submodular, then F is differentiable. For $\forall x \le y$, we also have

$$\nabla F(x) \ge \nabla F(y), \tag{23}$$

and

$$F(x) \le F(y). \tag{24}$$

By definition of $\tilde{F}$, we have $\tilde{F}$ is differentiable and for $\forall x \le y$,

$$\begin{aligned}\nabla\tilde{F}(x)-\nabla\tilde{F}(y) &= \nabla\mathbb{E}_{v\sim B^d}[F(x+\delta v)]-\nabla\mathbb{E}_{v\sim B^d}[F(y+\delta v)] &(25)\\ &= \mathbb{E}_{v\sim B^d}[\nabla F(x+\delta v)-\nabla F(y+\delta v)] &(26)\\ &\ge \mathbb{E}_{v\sim B^d}[0] &(27)\\ &= 0, &(28)\end{aligned}$$

and

$$\begin{aligned}\tilde{F}(x)-\tilde{F}(y) &= \mathbb{E}_{v\sim B^d}[F(x+\delta v)]-\mathbb{E}_{v\sim B^d}[F(y+\delta v)] &(29)\\ &= \mathbb{E}_{v\sim B^d}[F(x+\delta v)-F(y+\delta v)] &(30)\\ &\le \mathbb{E}_{v\sim B^d}[0] &(31)\\ &= 0, &(32)\end{aligned}$$

i.e., $\nabla\tilde{F}(x) \ge \nabla\tilde{F}(y), \tilde{F}(x) \le \tilde{F}(y)$. So $\tilde{F}$ is also a monotone continuous DR-submodular function. □

B Proof of Theorem 1

In order to prove Theorem 1, we need the following variance reduction lemmas [Shamir, 2017, Chen et al., 2018b], where the second one is a slight improvement of Lemma 2 in [Mokhtari et al., 2018a] and Lemma 5 in [Mokhtari et al., 2018b].

Lemma 4 (Lemma 10 of [Shamir, 2017])**.** *It holds that*

$$\mathbb{E}_{u\sim S^{d-1}}[\frac{d}{2\delta}(F(z+\delta u)-F(z-\delta u))u|z] = \nabla\tilde{F}(z), \tag{33}$$

$$\mathbb{E}_{u\sim S^{d-1}}[\|\frac{d}{2\delta}(F(z+\delta u)-F(z-\delta u))u-\nabla\tilde{F}(z)\|^2|z] \le cdG^2, \tag{34}$$

where c is a constant.

Lemma 5 (Theorem 3 of [Chen et al., 2018b])**.** *Let $\{a_t\}_{t=0}^T$ be a sequence of points in $\mathbb{R}^n$ such that $\|a_t - a_{t-1}\| \leq G_0/(t+s)$ for all $1 \leq t \leq T$ with fixed constants $G_0 \geq 0$ and $s \geq 3$. Let $\{\tilde{a}_t\}_{t=1}^T$ be a sequence of random variables such that $\mathbb{E}[\tilde{a}_t|\mathcal{F}_{t-1}] = a_t$ and $\mathbb{E}[\|\tilde{a}_t - a_t\|^2|\mathcal{F}_{t-1}] \leq \sigma^2$ for every $t \geq 0$, where $\mathcal{F}_{t-1}$ is the σ-field generated by $\{\tilde{a}_i\}_{i=1}^t$ and $\mathcal{F}_0 = \varnothing$. Let $\{d_t\}_{t=0}^T$ be a sequence of random variables where d_0 is fixed and subsequent d_t are obtained by the recurrence*

$$d_t = (1-\rho_t)d_{t-1} + \rho_t \tilde{a}_t \tag{35}$$

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

$$\mathbb{E}[\|a_t - d_t\|^2] \leq \frac{Q}{(t+s+1)^{2/3}}, \tag{36}$$

where $Q \triangleq \max\{\|a_0 - d_0\|^2(s+1)^{2/3}, 4\sigma^2 + 3G_0^2/2\}$.

Now we turn to prove Theorem 1.

Proof of Theorem 1. First of all, we note that technically we need the iteration number $T \geq 4$, which always holds in practical applications.

Then we show that $\forall t = 1, \ldots, T+1$, $x_t \in \mathcal{D}_\delta$. By the definition of x_t, we have $x_t = \sum_{i=1}^{t-1} \frac{v_i}{T}$. Since v_t's are non-negative vectors, we know that x_t's are also non-negative vectors and that $0 = x_1 \leq x_2 \leq \ldots \leq x_{T+1}$. It suffices to show that $x_{T+1} \in \mathcal{D}_\delta$. Since x_{T+1} is a convex combination of $v_1, \ldots, v_T$ and v_t's are in $\mathcal{D}_\delta$, we conclude that $x_{T+1} \in \mathcal{D}_\delta$. In addition, since v_t's are also in $\mathcal{K} - \delta\mathbf{1}$, x_{T+1} is also in $\mathcal{K} - \delta\mathbf{1}$. Therefore our final choice $x_{T+1} + \delta\mathbf{1}$ resides in the constraint $\mathcal{K}$.

Let $z_t \triangleq x_t + \delta\mathbf{1}$ and the shrunk domain (without translation) $\mathcal{D}'_\delta \triangleq \mathcal{D}_\delta + \delta\mathbf{1} = \prod_{i=1}^d [\delta, a_i - \delta] \subseteq \mathcal{D}$. By Jensen's inequality and the fact F has L-Lipschitz continuous gradients, we have

$$\|\nabla\tilde{F}(x) - \nabla\tilde{F}(y)\| \leq L\|x - y\|. \tag{37}$$

Thus,

$$\begin{aligned}
\tilde{F}(z_{t+1}) - \tilde{F}(z_t) &= \tilde{F}(z_t + \frac{v_t}{T}) - \tilde{F}(z_t) && (38)\\
&\geq \frac{1}{T}\nabla\tilde{F}(z_t)^\top v_t - \frac{L}{2T^2}\|v_t\|^2 && (39)\\
&\geq \frac{1}{T}\nabla\tilde{F}(z_t)^\top v_t - \frac{L}{2T^2}D_1^2 && (40)\\
&= \frac{1}{T}\left(\bar{g}_t^\top v_t + (\nabla\tilde{F}(z_t) - \bar{g}_t)^\top v_t\right) - \frac{L}{2T^2}D_1^2. && (41)
\end{aligned}$$

Let $x_\delta^* \triangleq \arg\max_{x \in \mathcal{D}'_\delta \cap \mathcal{K}} \tilde{F}(x)$. Since $x_\delta^*, z_t \in \mathcal{D}'_\delta$, we have $v_t^* \triangleq (x_\delta^* - z_t) \vee 0 \in \mathcal{D}_\delta$. We know $z_t + v_t^* = x_\delta^* \vee z_t \in \mathcal{D}'_\delta$ and

$$v_t^* + \delta\mathbf{1} = (x_\delta^* - x_t) \vee \delta\mathbf{1} \leq x_\delta^*. \tag{42}$$

Since we assume that F is monotone continuous DR-submodular, by Lemma 1, $\tilde{F}$ is also monotone continuous DR-submodular. As a result, $\tilde{F}$ is concave along non-negative directions, and $\nabla\tilde{F}$ is entry-wise non-negative. Thus we have

$$\begin{aligned}
\tilde{F}(z_t + v_t^*) - \tilde{F}(z_t) &\leq \nabla\tilde{F}(z_t)^\top v_t^* && (43)\\
&\leq \nabla\tilde{F}(z_t)^\top (x_\delta^* - \delta\mathbf{1}). && (44)
\end{aligned}$$

Since $x_\delta^* - \delta\mathbf{1} \in \mathcal{K}'$, we deduce

$$\begin{aligned}
\bar{g}_t^\top v_t &\geq \bar{g}_t^\top (x_\delta^* - \delta\mathbf{1}) && (45)\\
&= \nabla\tilde{F}(z_t)^\top (x_\delta^* - \delta\mathbf{1}) + (\bar{g}_t - \nabla\tilde{F}(z_t))^\top (x_\delta^* - \delta\mathbf{1}) && (46)\\
&\geq \tilde{F}(z_t + v_t^*) - \tilde{F}(z_t) + (\bar{g}_t - \nabla\tilde{F}(z_t))^\top (x_\delta^* - \delta\mathbf{1}) && (47)\\
&\geq \tilde{F}(x_\delta^*) - \tilde{F}(z_t) + (\bar{g}_t - \nabla\tilde{F}(z_t))^\top (x_\delta^* - \delta\mathbf{1}). && (48)
\end{aligned}$$

Therefore, we obtain

$$\bar{g}_t^\top v_t + (\nabla \tilde{F}(z_t) - \bar{g}_t)^\top v_t \geq \tilde{F}(x_\delta^*) - \tilde{F}(z_t) + (\nabla \tilde{F}(z_t) - \bar{g}_t)^\top (v_t - (x_\delta^* - \delta \mathbf{1})). \tag{49}$$

By plugging Eq. (49) into Eq. (41), after re-arrangement of the terms, we obtain

$$h_{t+1} \leq (1 - \frac{1}{T})h_t + \frac{1}{T}(\nabla \tilde{F}(z_t) - \bar{g}_t)^\top ((x_\delta^* - \delta \mathbf{1}) - v_t) + \frac{L}{2T^2} D_1^2, \tag{50}$$

where $h_t \triangleq \tilde{F}(x_\delta^*) - \tilde{F}(z_t)$. Next we derive an upper bound for $(\nabla \tilde{F}(z_t) - \bar{g}_t)^\top ((x_\delta^* - \delta \mathbf{1}) - v_t)$. By Young's inequality, it can be deduced that for any $\beta_t > 0$,

$$(\nabla \tilde{F}(z_t) - \bar{g}_t)^\top ((x_\delta^* - \delta \mathbf{1}) - v_t) \leq \frac{\beta_t}{2} \|\nabla \tilde{F}(z_t) - \bar{g}_t\|^2 + \frac{1}{2\beta_t} \|(x_\delta^* - \delta \mathbf{1}) - v_t\|^2 \tag{51}$$

$$\leq \frac{\beta_t}{2} \|\nabla \tilde{F}(z_t) - \bar{g}_t\|^2 + \frac{1}{2\beta_t} D_1^2. \tag{52}$$

Now let $\mathcal{F}_1 \triangleq \varnothing$ and $\mathcal{F}_t$ be the σ-field generate by $\{\bar{g}_1, \ldots, \bar{g}_{t-1}\}$, then by Lemma 4, we have

$$\mathbb{E}[\frac{d}{2\delta}(F(y_{t,i}^+) - F(y_{t,i}^-))u_{t,i} | \mathcal{F}_{t-1}] = \nabla \tilde{F}(z_t), \tag{53}$$

and

$$\mathbb{E}[\|\frac{d}{2\delta}(F(y_{t,i}^+) - F(y_{t,i}^-))u_{t,i} - \nabla \tilde{F}(z_t)\|^2 | \mathcal{F}_{t-1}] \leq cdG^2. \tag{54}$$

Therefore,

$$\mathbb{E}[g_t | \mathcal{F}_{t-1}] = \mathbb{E}[\frac{1}{B_t} \sum_{i=1}^{B_t} \frac{d}{2\delta}(F(y_{t,i}^+) - F(y_{t,i}^-))u_{t,i} | \mathcal{F}_{t-1}] \tag{55}$$

$$= \nabla \tilde{F}(z_t), \tag{56}$$

and

$$\mathbb{E}[\|g_t - \nabla \tilde{F}(z_t)\|^2 | \mathcal{F}_{t-1}] = \frac{1}{B_t^2} \sum_{i=1}^{B_t} \mathbb{E}[\|\frac{d}{2\delta}(F(y_{t,i}^+) - F(y_{t,i}^-))u_{t,i} - \nabla \tilde{F}(z_t)\|^2 | \mathcal{F}_{t-1}] \tag{57}$$

$$\leq \frac{cdG^2}{B_t}. \tag{58}$$

By Jensen's inequality and the assumption F is L-smooth, we have

$$\|\nabla \tilde{F}(z_t) - \nabla \tilde{F}(z_{t-1})\| \leq L\frac{D_1}{T} \leq \frac{2LD_1}{t+3}. \tag{59}$$

Then by Lemma 5 with $s = 3, d_t = \bar{g}_t, \forall t \geq 0, \tilde{a}_t = g_t, a_t = \nabla \tilde{F}(z_t), \forall t \geq 1, a_0 = \nabla \tilde{F}(z_1), G_0 = 2LD_1$, we have

$$\mathbb{E}[\|\nabla \tilde{F}(z_t) - \bar{g}_t\|^2] \leq \frac{Q}{(t+4)^{2/3}}, \tag{60}$$

where $Q \triangleq \max\{\|\nabla \tilde{F}(x_1 + \delta \mathbf{1})\|^2 4^{2/3}, \frac{4cdG^2}{B_t} + 6L^2 D_1^2\}$. Note that by Lemma 1, we have $\|\nabla \tilde{F}(x)\| \leq G$, thus we can re-define $Q = \max\{4^{2/3}G^2, \frac{4cdG^2}{B_t} + 6L^2 D_1^2\}$.

Using Eqs. (50), (52) and (60) and taking expectation, we obtain

$$\mathbb{E}[h_{t+1}] \leq (1 - \frac{1}{T})\mathbb{E}[h_t] + \frac{1}{T}\left(\frac{\beta_t}{2} \cdot \frac{Q}{(t+4)^{2/3}} + \frac{D_1^2}{2\beta_t}\right) + \frac{L}{2T^2} D_1^2 \leq (1 - \frac{1}{T})\mathbb{E}[h_t] + \frac{D_1 Q^{1/2}}{T(t+4)^{1/3}} + \frac{L}{2T^2} D_1^2, \tag{61}$$

where we set $\beta_t = \frac{D_1(t+4)^{1/3}}{Q^{1/2}}$. Using the above inequality recursively, we have

$$
\begin{aligned}
\mathbb{E}[h_{T+1}] \leq & (1-\frac{1}{T})^T(\tilde{F}(x_\delta^*)-\tilde{F}(\delta\mathbf{1}))+\sum_{t=1}^{T}\frac{D_1Q^{1/2}}{T(t+4)^{1/3}}+\frac{L}{2T}D_1^2 && (62)\\
\leq & e^{-1}(\tilde{F}(x_\delta^*)-\tilde{F}(\delta\mathbf{1}))+\frac{D_1Q^{1/2}}{T}\int_0^T\frac{\mathrm{d}x}{(x+4)^{1/3}}+\frac{L}{2T}D_1^2 && (63)\\
\leq & e^{-1}(\tilde{F}(x_\delta^*)-\tilde{F}(\delta\mathbf{1}))+\frac{D_1Q^{1/2}}{T}\frac{3}{2}(T+4)^{2/3}+\frac{L}{2T}D_1^2 && (64)\\
\leq & e^{-1}(\tilde{F}(x_\delta^*)-\tilde{F}(\delta\mathbf{1}))+\frac{D_1Q^{1/2}}{T}\frac{3}{2}(2T)^{2/3}+\frac{L}{2T}D_1^2 && (65)\\
\leq & e^{-1}(\tilde{F}(x_\delta^*)-\tilde{F}(\delta\mathbf{1}))+\frac{3D_1Q^{1/2}}{T^{1/3}}+\frac{LD_1^2}{2T}. && (66)
\end{aligned}
$$

By re-arranging the terms, we conclude

$$
\begin{aligned}
(1-\frac{1}{e})\tilde{F}(x_\delta^*)-\mathbb{E}[\tilde{F}(z_{T+1})] \leq & -e^{-1}\tilde{F}(\delta\mathbf{1})+\frac{3D_1Q^{1/2}}{T^{1/3}}+\frac{LD_1^2}{2T} && (67)\\
\leq & \frac{3D_1Q^{1/2}}{T^{1/3}}+\frac{LD_1^2}{2T}, && (68)
\end{aligned}
$$

where the second inequality holds since the image of F is in $\mathbb{R}_+$.

By Lemma 1, we have $\tilde{F}(z_{T+1}) \leq F(z_{T+1})+\delta G$ and

$$
\tilde{F}(x_\delta^*) \geq \tilde{F}(x^*)-\delta G\sqrt{d} \geq F(x^*)-\delta G(\sqrt{d}+1). \tag{69}
$$

Therefore,

$$
(1-\frac{1}{e})F(x^*)-\mathbb{E}[F(z_{T+1})] \leq \frac{3D_1Q^{1/2}}{T^{1/3}}+\frac{LD_1^2}{2T}+\delta G(1+(\sqrt{d}+1)(1-\frac{1}{e})). \tag{70}
$$

□

C Proof of Theorem 2

Proof. By the unbiasedness of $\hat{F}$ and Lemma 4, we have

$$
\begin{aligned}
\mathbb{E}[\frac{d}{2\delta}(\hat{F}(y_{t,i}^+)-\hat{F}(y_{t,i}^-))u_{t,i}|\mathcal{F}_{t-1}] = & \mathbb{E}[\mathbb{E}[\frac{d}{2\delta}(\hat{F}(y_{t,i}^+)-\hat{F}(y_{t,i}^-))u_{t,i}|\mathcal{F}_{t-1},u_{t,i}]|\mathcal{F}_{t-1}] && (71)\\
= & \mathbb{E}[\frac{d}{2\delta}(F(y_{t,i}^+)-F(y_{t,i}^-))u_{t,i}|\mathcal{F}_{t-1}] && (72)\\
= & \nabla\tilde{F}(z_t), && (73)
\end{aligned}
$$

where $z_t = x_t + \delta\mathbf{1}$, and

$$\begin{aligned}
&\mathbb{E}[\|\frac{d}{2\delta}(\hat{F}(y_{t,i}^+) - \hat{F}(y_{t,i}^-))u_{t,i} - \nabla\tilde{F}(z_t)\|^2|\mathcal{F}_{t-1}] && (74)\\
=&\mathbb{E}[\mathbb{E}[\|\frac{d}{2\delta}(F(y_{t,i}^+) - F(y_{t,i}^-))u_{t,i} - \nabla\tilde{F}(z_t) && (75)\\
&+ \frac{d}{2\delta}(\hat{F}(y_{t,i}^+) - F(y_{t,i}^+))u_{t,i} && (76)\\
&- \frac{d}{2\delta}(\hat{F}(y_{t,i}^-) - F(y_{t,i}^-))u_{t,i}\|^2|\mathcal{F}_{t-1}, u_{t,i}]|\mathcal{F}_{t-1}] && (77)\\
=&\mathbb{E}[\mathbb{E}[\|\frac{d}{2\delta}(F(y_{t,i}^+) - F(y_{t,i}^-))u_{t,i} - \nabla\tilde{F}(z_t)\|^2|\mathcal{F}_{t-1}, u_{t,i}]|\mathcal{F}_{t-1}] && (78)\\
&+ \mathbb{E}[\mathbb{E}[\|\frac{d}{2\delta}(\hat{F}(y_{t,i}^+) - F(y_{t,i}^+))u_{t,i}\|^2|\mathcal{F}_{t-1}, u_{t,i}]|\mathcal{F}_{t-1}] && (79)\\
&+ \mathbb{E}[\mathbb{E}[\|\frac{d}{2\delta}(\hat{F}(y_{t,i}^-) - F(y_{t,i}^-))u_{t,i}\|^2|\mathcal{F}_{t-1}, u_{t,i}]|\mathcal{F}_{t-1}] && (80)\\
\le&\mathbb{E}[\|\frac{d}{2\delta}(F(y_{t,i}^+) - F(y_{t,i}^-))u_{t,i} - \nabla\tilde{F}(z_t)\|^2|\mathcal{F}_{t-1}] && (81)\\
&+ \frac{d^2}{4\delta^2}\mathbb{E}[\mathbb{E}[|\hat{F}(y_{t,i}^+) - F(y_{t,i}^+)|^2\cdot\|u_{t,i}\|^2|\mathcal{F}_{t-1}, u_{t,i}]|\mathcal{F}_{t-1}] && (82)\\
&+ \frac{d^2}{4\delta^2}\mathbb{E}[\mathbb{E}[|\hat{F}(y_{t,i}^-) - F(y_{t,i}^-)|^2\cdot\|u_{t,i}\|^2|\mathcal{F}_{t-1}, u_{t,i}]|\mathcal{F}_{t-1}] && (83)\\
\le&cdG^2 + \frac{d^2}{4\delta^2}\sigma_0^2 + \frac{d^2}{4\delta^2}\sigma_0^2 && (84)\\
=&cdG^2 + \frac{d^2}{2\delta^2}\sigma_0^2. && (85)
\end{aligned}$$

Then we have

$$\begin{aligned}
\mathbb{E}[g_t|\mathcal{F}_{t-1}] =&\mathbb{E}[\frac{1}{B_t}\sum_{i=1}^{B_t}\frac{d}{2\delta}(\hat{F}(y_{t,i}^+) - \hat{F}(y_{t,i}^-))u_{t,i}|\mathcal{F}_{t-1}] && (86)\\
=&\nabla\tilde{F}(z_t), && (87)
\end{aligned}$$

and

$$\begin{aligned}
\mathbb{E}[\|g_t - \nabla\tilde{F}(z_t)\|^2|\mathcal{F}_{t-1}] =&\frac{1}{B_t^2}\sum_{i=1}^{B_t}\mathbb{E}[\|\frac{d}{2\delta}(\hat{F}(y_{t,i}^+) - \hat{F}(y_{t,i}^-))u_{t,i} - \nabla\tilde{F}(z_t)\|^2|\mathcal{F}_{t-1}] && (88)\\
\le&\frac{cdG^2 + \frac{d^2}{2\delta^2}\sigma_0^2}{B_t}. && (89)
\end{aligned}$$

Similar to the proof of Theorem 1, we have

$$\mathbb{E}[\|\nabla\tilde{F}(z_t) - \bar{g}_t\|^2] \le \frac{Q}{(t+4)^{2/3}}, \tag{90}$$

where $Q = \max\{4^{2/3}G^2, 6L^2D_1^2 + \frac{4cdG^2 + 2d^2\sigma_0^2/\delta^2}{B_t}\}$. Thus we conclude

$$(1 - \frac{1}{e})F(x^*) - \mathbb{E}[F(z_{T+1})] \le \frac{3D_1Q^{1/2}}{T^{1/3}} + \frac{LD_1^2}{2T} + \delta G(1 + (\sqrt{d} + 1)(1 - \frac{1}{e})). \tag{91}$$

□

D Proof of Lemma 3

Proof. Recall that $F(x) = \mathbb{E}_{X\sim x}[f(X)] = \sum_{S\subseteq\Omega} f(S)\prod_{i\in S} x_i \prod_{j\notin S}(1 - x_j)$, then for any fixed $i \in [d]$, where $d = |\Omega|$, we have

$$\begin{aligned}|\frac{\partial F(x)}{\partial x_i}| &= |\sum_{\substack{S\subseteq\Omega\\ i\in S}} f(S)\prod_{\substack{j\in S\\ j\neq i}} x_j \prod_{\substack{k\notin S\\ k\neq i}}(1-x_k) - \sum_{\substack{S\subseteq\Omega\\ i\notin S}} f(S)\prod_{\substack{j\in S\\ j\neq i}} x_j \prod_{\substack{k\notin S\\ k\neq i}}(1-x_k)| && (92)\\ &\leq M[\sum_{\substack{S\subseteq\Omega\\ i\in S}} \prod_{\substack{j\in S\\ j\neq i}} x_j \prod_{\substack{k\notin S\\ k\neq i}}(1-x_k) + \sum_{\substack{S\subseteq\Omega\\ i\notin S}} \prod_{\substack{j\in S\\ j\neq i}} x_j \prod_{\substack{k\notin S\\ k\neq i}}(1-x_k)] && (93)\\ &= 2M. && (94)\end{aligned}$$

So we have

$$\|\nabla F(x)\| \leq 2M\sqrt{d}. \quad (95)$$

Then F is $2M\sqrt{d}$-Lipschitz.

Now we turn to prove that F has Lipschitz continuous gradients. Thanks to the multilinearity, we have

$$\frac{\partial F}{\partial x_i} = F(x|x_i = 1) - F(x|x_i = 0). \quad (96)$$

Since

$$F(x|x_i = 1) = \sum_{\substack{S\subseteq\Omega\\ i\in S}} f(S)\prod_{\substack{j\in S\\ j\neq i}} x_j \prod_{\substack{k\notin S\\ k\neq i}}(1-x_k), \quad (97)$$

we have

$$\frac{\partial F(x|x_i = 1)}{\partial x_i} = 0, \quad (98)$$

and for any fixed $j \neq i$,

$$\begin{aligned}|\frac{\partial F(x|x_i = 1)}{\partial x_j}| &= |\sum_{\substack{S\subseteq\Omega\\ i,j\in S}} f(S)\prod_{\substack{l\in S\\ l\notin\{i,j\}}} x_l \prod_{\substack{k\notin S\\ k\notin\{i,j\}}}(1-x_k) - \sum_{\substack{S\subseteq\Omega\\ i\in S, j\notin S}} f(S)\prod_{\substack{l\in S\\ l\notin\{i,j\}}} x_l \prod_{\substack{k\notin S\\ k\notin\{i,j\}}}(1-x_k)| && (99)\\ &\leq M[\sum_{\substack{S\subseteq\Omega\\ i,j\in S}} \prod_{\substack{l\in S\\ l\notin\{i,j\}}} x_l \prod_{\substack{k\notin S\\ k\notin\{i,j\}}}(1-x_k) + \sum_{\substack{S\subseteq\Omega\\ i\in S, j\notin S}} \prod_{\substack{l\in S\\ l\notin\{i,j\}}} x_l \prod_{\substack{k\notin S\\ k\notin\{i,j\}}}(1-x_k)] && (100)\\ &= 2M. && (101)\end{aligned}$$

Similarly, we have $\frac{\partial F(x|x_i=0)}{\partial x_i} = 0$, and $|\frac{\partial F(x|x_i=0)}{\partial x_j}| \leq 2M$ for $j \neq i$. So we conclude that

$$|\frac{\partial^2 F}{\partial x_j \partial x_i}| \leq \begin{cases} 0, & \text{if } j = i,\\ 4M, & \text{if } j \neq i. \end{cases} \quad (102)$$

Then $\|\nabla \frac{\partial F}{\partial x_i}\| \leq 4M\sqrt{d-1}$, *i.e.*, $\frac{\partial F}{\partial x_i}$ is $4M\sqrt{d-1}$-Lipschitz.

Then we deduce that

$$\begin{aligned}\|\nabla F(z_1) - \nabla F(z_2)\| &= \left[\sum_{i=1}^{d}\left(\frac{\partial F(z_1)}{\partial x_i} - \frac{\partial F(z_2)}{\partial x_i}\right)^2\right]^{1/2} && (103)\\ &\leq \left[\sum_{i=1}^{d}(4M\sqrt{d-1})^2\|z_1 - z_2\|^2\right]^{1/2} && (104)\\ &= \sqrt{\sum_{i=1}^{d}(4M\sqrt{d-1})^2} \cdot \|z_1 - z_2\| && (105)\\ &= 4M\sqrt{d(d-1)}\|z_1 - z_2\|. && (106)\end{aligned}$$

So F is $4M\sqrt{d(d-1)}$-smooth. □

E Proof of Theorem 3

Proof. Recall that we define $z_t = x_t + \delta\mathbf{1}$. Then we have

$$\begin{aligned}\mathbb{E}[\|g_t - \nabla\tilde{F}(z_t)\|^2|\mathcal{F}_{t-1}] =& \frac{1}{B_t^2}\sum_{i=1}^{B_t}\mathbb{E}[\|\frac{d}{2\delta}(\bar{f}_{t,i}^{+} - \bar{f}_{t,i}^{-})u_{t,i} - \nabla\tilde{F}(z_t)\|^2|\mathcal{F}_{t-1}] & (107)\\ =& \frac{1}{B_t^2}\sum_{i=1}^{B_t}\mathbb{E}[\|[\frac{d}{2\delta}(F(y_{t,i}^{+}) - F(y_{t,i}^{-}))u_{t,i} - \nabla\tilde{F}(z_t)] & (108)\\ &+ \frac{d}{2\delta}[\bar{f}_{t,i}^{+} - F(y_{t,i}^{+})]u_{t,i} - \frac{d}{2\delta}[\bar{f}_{t,i}^{-} - F(y_{t,i}^{-})]u_{t,i}\|^2|\mathcal{F}_{t-1}] & (109)\\ =& \frac{1}{B_t^2}\sum_{i=1}^{B_t}\mathbb{E}[\|[\frac{d}{2\delta}(F(y_{t,i}^{+}) - F(y_{t,i}^{-}))u_{t,i} - \nabla\tilde{F}(z_t)]\|^2|\mathcal{F}_{t-1}] & (110)\\ &+ \frac{1}{B_t^2}\sum_{i=1}^{B_t}\mathbb{E}[|\frac{d}{2\delta}[\bar{f}_{t,i}^{+} - F(y_{t,i}^{+})]|^2|\mathcal{F}_{t-1}] & (111)\\ &+ \frac{1}{B_t^2}\sum_{i=1}^{B_t}\mathbb{E}[|\frac{d}{2\delta}[\bar{f}_{t,i}^{-} - F(y_{t,i}^{-})]|^2|\mathcal{F}_{t-1}], & (112)\end{aligned}$$

where we used the independence of $\bar{f}_{t,i}^{\pm}$ and the facts that $\mathbb{E}[\bar{f}_{t,i}^{\pm}] = F(y_{t,i}^{\pm}), \mathbb{E}[\frac{d}{2\delta}(F(y_{t,i}^{+}) - F(y_{t,i}^{-}))u_{t,i}] = \nabla\tilde{F}(z_t)$.

Then same to Eq. (58) and by Lemma 3, the first item is no more than $\frac{4cd^2M^2}{B_t}$. To upper bound the last two items, we have for every $i \in [B_t]$,

$$\begin{aligned}\mathbb{E}[|\frac{d}{2\delta}[\bar{f}_{t,i}^{+} - F(y_{t,i}^{+})]|^2|\mathcal{F}_{t-1}] &= \frac{d^2}{4\delta^2}\mathbb{E}[[\sum_{j=1}^{l}[f(Y_{t,i,j}^{+}) - F(y_{t,i}^{+})]/l]^2|\mathcal{F}_{t-1}]\\ &\le \frac{d^2}{4\delta^2}\cdot l\cdot\frac{M^2}{l^2}\\ &= \frac{d^2M^2}{4l\delta^2}.\end{aligned} \tag{113}$$

Similarly, we have

$$\mathbb{E}[|\frac{d}{2\delta}[\bar{f}_{t,i}^{-} - F(y_{t,i}^{-})]|^2|\mathcal{F}_{t-1}] \le \frac{d^2M^2}{4l\delta^2}. \tag{114}$$

As a result, we have

$$\begin{aligned}\mathbb{E}[\|g_t - \nabla\tilde{F}(z_t)\|^2|\mathcal{F}_{t-1}] &\le \frac{4cd^2M^2}{B_t} + \frac{1}{B_t^2}\cdot B_t\cdot\frac{d^2M^2}{4l\delta^2} + \frac{1}{B_t^2}\cdot B_t\cdot\frac{d^2M^2}{4l\delta^2}\\ &= \frac{4cd^2M^2}{B_t} + \frac{d^2M^2}{2B_tl\delta^2}.\end{aligned} \tag{115}$$

Then same to the proof for Theorem 1, we have

$$(1-\frac{1}{e})F(x^*) - \mathbb{E}[F(z_{T+1})] \le \frac{3D_1Q^{1/2}}{T^{1/3}} + \frac{2M\sqrt{d(d-1)}D_1^2}{T} + 2M\delta\sqrt{d}(1 + (\sqrt{d}+1)(1-\frac{1}{e})). \tag{116}$$

where $D_1 \triangleq \mathrm{diam}(\mathcal{K}')$, $Q = \max\{4^{5/3}dM^2, \frac{2d^2M^2(8c+\frac{1}{l\delta^2})}{B_t} + 96d(d-1)M^2D_1^2\}$, x^* is the global maximizer of F on $\mathcal{K}$.

Note that since the rounding scheme is lossless, we have

$$(1-\frac{1}{e})f(X^*) - \mathbb{E}[f(X_{T+1})] \le (1-\frac{1}{e})F(x^*) - \mathbb{E}[F(z_{T+1})]. \tag{117}$$

Combine Eqs. (116) and (117), we have

$$(1-\frac{1}{e})f(X^*)-\mathbb{E}[f(X_{T+1})]\leq\frac{3D_1Q^{1/2}}{T^{1/3}}+\frac{2M\sqrt{d(d-1)}D_1^2}{T}+2M\delta\sqrt{d}(1+(\sqrt{d}+1)(1-\frac{1}{e})). \tag{118}$$

□