Appendices

A COST-BENEFIT GREEDY

Algorithm 1 Cost-benefit greedy

$\mathsf{U} \leftarrow [d]$, $\mathsf{S} \leftarrow \emptyset$
while $\mathsf{U} \neq \emptyset$ **do**
  $j \leftarrow \operatorname{argmax}_{j' \in \mathsf{U}} \frac{F(j'|\mathsf{S})}{G(j'|\mathsf{S})}$
  if $G(\mathsf{S} \cup \{j\}) \leq c$ **then**
    $\mathsf{S} \leftarrow \mathsf{S} \cup \{j\}$
  $\mathsf{U} \leftarrow \mathsf{U} \backslash \{j\}$
return S

We prove the approximation guarantee of the CBG algorithm (Algorithm 1) for the following problem:

$$\underset{\mathsf{S} \subseteq [d]}{\text{maximize}} \ F(\mathsf{S}) \quad \text{subject to} \ G(\mathsf{S}) \leq c.$$

To obtain the main theorem, we use the following definitions and lemmas.

Let S^* be any subset of $[d]$ such that $G(\mathsf{S}^*) \leq c^*$ and $|\mathsf{S}^*| = k^*$. As in the theorem of the main paper, we assume that $\min\{c, c^*\} \geq \rho$ holds. We suppose that $t+1$ elements are added when $G(\mathsf{S}) > c - \rho$ occurs in the loops of Algorithm 1 for the first time. We let j_i be the i-th element added to S for $i \in [t+1]$. We define $\mathsf{S}_i \coloneqq \{j_1, \dots, j_i\}$ for $i \in [t+1]$ and $\mathsf{S}_0 \coloneqq \emptyset$. Thanks to the monotonicity of $G(\cdot)$, the definition of ρ, and $G(\mathsf{S}_t) \leq c - \rho$, we have

$$G(\mathsf{S}_i) = G(\mathsf{S}_{i-1}) + G(j_i \mid \mathsf{S}_{i-1}) \leq G(\mathsf{S}_t) + G(j_i \mid \mathsf{S}_{i-1}) \leq c - \rho + \rho = c$$

for $i \in [t+1]$; in particular, we have $c - \rho \leq G(\mathsf{S}_{t+1}) \leq c$. Namely, $G(\mathsf{S}_1), \dots, G(\mathsf{S}_{t+1})$ do not exceed the budget value, which means that j_i ($i \in [t+1]$) is the element added in the i-th iteration of the algorithm. Moreover, the output, S, satisfies $\mathsf{S}_{t+1} \subseteq \mathsf{S}$.

Lemma 1. *For $i = 1, \dots, t+1$, we have*

$$F(\mathsf{S}_i) - F(\mathsf{S}_{i-1}) \geq \theta \beta_{k^*} \gamma_{\mathsf{S}_t, k^*} \cdot \frac{G(j_i \mid \mathsf{S}_{i-1})}{c^*} (F(\mathsf{S}^*) - F(\mathsf{S}_{i-1})).$$

Proof. Thanks to the weak submodularity of $F(\cdot)$, we have

$$\begin{aligned} \sum_{j \in \mathsf{S}^* \backslash \mathsf{S}_{i-1}} F(j \mid \mathsf{S}_{i-1}) &\geq \gamma_{\mathsf{S}_{i-1}, k^*} F(\mathsf{S}^* \backslash \mathsf{S}_{i-1} \mid \mathsf{S}_{i-1}) \\ &= \gamma_{\mathsf{S}_{i-1}, k^*} F(\mathsf{S}^* \mid \mathsf{S}_{i-1}) \geq \gamma_{\mathsf{S}_t, k^*} F(\mathsf{S}^* \mid \mathsf{S}_{i-1}). \end{aligned}$$

Since j_i is chosen greedily, $j_i = \operatorname{argmax}_{j \notin \mathsf{S}_{i-1}} \frac{F(j|\mathsf{S}_{i-1})}{G(j|\mathsf{S}_{i-1})}$ holds, and hence $\frac{F(j_i|\mathsf{S}_{i-1})}{G(j_i|\mathsf{S}_{i-1})} \geq \frac{F(j|\mathsf{S}_{i-1})}{G(j|\mathsf{S}_{i-1})}$ for any $j \in \mathsf{S}^* \backslash \mathsf{S}_{i-1}$. Using this fact and the above inequality, we obtain

$$\begin{aligned} F(j_i \mid \mathsf{S}_{i-1}) \sum_{j \in \mathsf{S}^* \backslash \mathsf{S}_{i-1}} G(j \mid \mathsf{S}_{i-1}) &\geq G(j_i \mid \mathsf{S}_{i-1}) \sum_{j \in \mathsf{S}^* \backslash \mathsf{S}_{i-1}} F(j \mid \mathsf{S}_{i-1}) \\ &\geq G(j_i \mid \mathsf{S}_{i-1}) \times \gamma_{\mathsf{S}_t, k^*} F(\mathsf{S}^* \mid \mathsf{S}_{i-1}). \end{aligned}$$

We consider bounding from above $\sum_{j \in \mathsf{S}^* \backslash \mathsf{S}_{i-1}} G(j \mid \mathsf{S}_{i-1})$ in LHS. By using the definition of restricted inverse curvature and superadditivity ratio of $G(\cdot)$, we obtain

$$\sum_{j \in \mathsf{S}^* \backslash \mathsf{S}_{i-1}} G(j \mid \mathsf{S}_{i-1}) \leq \frac{1}{\theta} \sum_{j \in \mathsf{S}^* \backslash \mathsf{S}_{i-1}} G(j) \leq \frac{1}{\theta \beta_{k^*}} G(\mathsf{S}^* \backslash \mathsf{S}_{i-1}) \leq \frac{c^*}{\theta \beta_{k^*}},$$

where the last inequality comes from $G(\mathsf{S}^*\backslash\mathsf{S}_{i-1}) \leq G(\mathsf{S}^*) \leq c^*$. Hence we obtain

$$\frac{c^*}{\theta\beta_{k^*}}(F(\mathsf{S}_i) - F(\mathsf{S}_{i-1})) = \frac{c^*}{\theta\beta_{k^*}}F(j_i \mid \mathsf{S}_{i-1}) \geq G(j_i \mid \mathsf{S}_{i-1}) \times \gamma_{\mathsf{S}_t,k^*}F(\mathsf{S}^* \mid \mathsf{S}_{i-1}).$$

The lemma is obtained by using $F(\mathsf{S}^* \mid \mathsf{S}_{i-1}) = F(\mathsf{S}^* \cup \mathsf{S}_{i-1}) - F(\mathsf{S}_{i-1}) \geq F(\mathsf{S}^*) - F(\mathsf{S}_{i-1})$ and rearranging terms. □

Lemma 2. *For $i = 1, \ldots, t+1$, we have*

$$F(\mathsf{S}_i) \geq \left(1 - \prod_{i'=1}^{i}\left(1 - \theta\beta_{k^*}\gamma_{\mathsf{S}_t,k^*} \cdot \frac{G(j_{i'} \mid \mathsf{S}_{i'-1})}{c^*}\right)\right)F(\mathsf{S}^*).$$

Proof. We prove the lemma by induction on $i = 1, \ldots, t+1$. First, if $i = 1$, the target inequality holds thanks to Lemma 1. We then assume that the target inequality holds for $\mathsf{S}_1, \ldots, \mathsf{S}_{i-1}$ and prove it for S_i. Combining Lemma 1 and the assumption, we obtain

$$\begin{aligned}
F(\mathsf{S}_i) &= F(\mathsf{S}_{i-1}) + (F(\mathsf{S}_i) - F(\mathsf{S}_{i-1})) \\
&\geq F(\mathsf{S}_{i-1}) + \theta\beta_{k^*}\gamma_{\mathsf{S}_t,k^*} \cdot \frac{G(j_i \mid \mathsf{S}_{i-1})}{c^*}(F(\mathsf{S}^*) - F(\mathsf{S}_{i-1})) \\
&= \left(1 - \theta\beta_{k^*}\gamma_{\mathsf{S}_t,k^*} \cdot \frac{G(j_i \mid \mathsf{S}_{i-1})}{c^*}\right)F(\mathsf{S}_{i-1}) + \theta\beta_{k^*}\gamma_{\mathsf{S}_t,k^*} \cdot \frac{G(j_i \mid \mathsf{S}_{i-1})}{c^*}F(\mathsf{S}^*) \\
&\geq \left(1 - \prod_{i'=1}^{i}\left(1 - \theta\beta_{k^*}\gamma_{\mathsf{S}_t,k^*} \cdot \frac{G(j_{i'} \mid \mathsf{S}_{i'-1})}{c^*}\right)\right)F(\mathsf{S}^*).
\end{aligned}$$

Thus the lemma holds by induction. □

Theorem 1. *Let S be the output of CBG and S^* be any subset that satisfies $G(\mathsf{S}^*) \leq c^*$ and $|\mathsf{S}^*| = k^*$. If $F(\cdot)$ has submodularity ratio γ_{S,k^*}, $G(\cdot)$ has superadditivity ratio β_{k^*} and restricted inverse curvature θ, and $\min\{c, c^*\} \geq \rho$ holds, then we have*

$$F(\mathsf{S}) \geq \left(1 - \exp\left(-\theta\beta_{k^*}\gamma_{\mathsf{S},k^*} \cdot \frac{c-\rho}{c^*}\right)\right)F(\mathsf{S}^*).$$

Proof. We define $x \coloneqq \theta\beta_{k^*}\gamma_{\mathsf{S}_t,k^*} \cdot \frac{c-\rho}{c^*}$ and $y_i \coloneqq \frac{G(j_i|\mathsf{S}_{i-1})}{G(\mathsf{S}_{t+1})}$ for $i \in [t+1]$. Thanks to $G(\mathsf{S}_{t+1}) \geq c - \rho$, $G(j_i \mid \mathsf{S}_{i-1}) \leq \rho \leq c^*$, and $\theta\beta_{k^*}\gamma_{\mathsf{S}_t,k^*} \leq 1$, we obtain

$$xy_i = \theta\beta_{k^*}\gamma_{\mathsf{S}_t,k^*} \cdot \frac{c-\rho}{c^*} \cdot y_i \leq \theta\beta_{k^*}\gamma_{\mathsf{S}_t,k^*} \cdot \frac{G(\mathsf{S}_{t+1})}{G(j_i \mid \mathsf{S}_{i-1})} \cdot y_i = \theta\beta_{k^*}\gamma_{\mathsf{S}_t,k^*} \leq 1.$$

Hence $1 - xy_i \geq 0$. Since $\sum_{i=1}^{t+1} y_i = \sum_{i=1}^{t+1} \frac{G(j_i|\mathsf{S}_{i-1})}{G(\mathsf{S}_{t+1})} = 1$ holds, $\prod_{i=1}^{t+1}(1 - xy_i)$ attains its maximum value when we have $y_1 = \cdots = y_{t+1} = \frac{1}{t+1}$. Thus we obtain

$$\prod_{i=1}^{t+1}\left(1 - \theta\beta_{k^*}\gamma_{\mathsf{S}_t,k^*} \cdot \frac{c-\rho}{c^*} \cdot \frac{G(j_i \mid \mathsf{S}_{i-1})}{G(\mathsf{S}_{t+1})}\right) \leq \left(1 - \theta\beta_{k^*}\gamma_{\mathsf{S}_t,k^*} \cdot \frac{c-\rho}{c^*} \cdot \frac{1}{t+1}\right)^{t+1}. \tag{A1}$$

By using Lemma 2, inequality (A1), and $G(\mathsf{S}_{t+1}) \geq c - \rho$, we obtain

$$
\begin{aligned}
F(\mathsf{S}_{t+1}) &\geq \left(1 - \prod_{i=1}^{t+1}\left(1 - \theta\beta_{k^*}\gamma_{\mathsf{S}_t,k^*} \cdot \frac{G(j_i \mid \mathsf{S}_{i-1})}{c^*}\right)\right) F(\mathsf{S}^*) && \because \text{Lemma 2} \\
&= \left(1 - \prod_{i=1}^{t+1}\left(1 - \theta\beta_{k^*}\gamma_{\mathsf{S}_t,k^*} \cdot \frac{c-\rho}{c^*} \cdot \frac{G(j_i \mid \mathsf{S}_{i-1})}{c-\rho}\right)\right) F(\mathsf{S}^*) \\
&\geq \left(1 - \prod_{i=1}^{t+1}\left(1 - \theta\beta_{k^*}\gamma_{\mathsf{S}_t,k^*} \cdot \frac{c-\rho}{c^*} \cdot \frac{G(j_i \mid \mathsf{S}_{i-1})}{G(\mathsf{S}_{t+1})}\right)\right) F(\mathsf{S}^*) && \because G(\mathsf{S}_{t+1}) \geq c-\rho \\
&\geq \left(1 - \left(1 - \theta\beta_{k^*}\gamma_{\mathsf{S}_t,k^*} \cdot \frac{c-\rho}{c^*} \cdot \frac{1}{t+1}\right)^{t+1}\right) F(\mathsf{S}^*) && \because \text{inequality (A1)} \\
&\geq \left(1 - \exp\left(-\theta\beta_{k^*}\gamma_{\mathsf{S}_t,k^*} \cdot \frac{c-\rho}{c^*}\right)\right) F(\mathsf{S}^*).
\end{aligned}
$$

Since we have $\mathsf{S}_t \subseteq \mathsf{S}_{t+1} \subseteq \mathsf{S}$, we obtain

$$
F(\mathsf{S}) \geq F(\mathsf{S}_{t+1}) \geq \left(1 - \exp\left(-\theta\beta_{k^*}\gamma_{\mathsf{S},k^*} \cdot \frac{c-\rho}{c^*}\right)\right) F(\mathsf{S}^*).
$$

□

B IHT WITH CBG PROJECTION

Algorithm 2 IHT with CBG projection

1: Initialize $\mathbf{x}_0 \in \mathbb{R}^{[d]}$
2: **for** $t = 0, 1, \ldots, T-1$ **do**
3: $\quad \mathbf{g}_t \leftarrow \mathbf{x}_t - \eta \nabla l(\mathbf{x}_t)$
4: $\quad \mathbf{x}_{t+1} \leftarrow \mathcal{P}_c(\mathbf{g}_t)$ ▷ CBG projection
5: **return** $\mathbf{x}_T$

We prove the theorem that guarantees the performance of Algorithm 2 for the following problem:

$$
\underset{\mathbf{x}\in\mathbb{R}^d}{\text{minimize}} \;\; l(\mathbf{x}) \quad \text{subject to} \;\; G(\text{supp}(\mathbf{x})) \leq c.
$$

We first explain the CBG projection (Step 4) in detail. Given any $\mathbf{z} \in \mathbb{R}^{[d]}$, $\hat{\mathbf{z}} = \mathcal{P}_c(\mathbf{z})$ is obtained as follows: We perform Algorithm 1 with objective function $F(\mathsf{S}) \coloneqq \|\mathbf{z}_\mathsf{S}\|_2^2 = \sum_{j\in\mathsf{S}} |\mathbf{z}_j|^2$, cost function $G(\mathsf{S})$, and budget value c; the resulting solution, S, satisfies $c - \rho \leq G(\mathsf{S}) \leq c$. We set $\hat{\mathbf{z}}_j$ to $\mathbf{z}_j$ if $j \in \mathsf{S}$ and 0 otherwise. Note that $F(\cdot)$ is monotone and modular, which implies that its submodularity ratio is equal to 1. We first prove a key lemma that guarantees the performance of the CBG projection.

Lemma 3. *Assume $c^* \geq \rho$ and let S^* be an arbitrary subset such that $G(\mathsf{S}^*) \leq c^*$ and $|\mathsf{S}^*| = k^*$. Given any $\mathbf{z} \in \mathbb{R}^{[d]}$, we let $\mathsf{S} \coloneqq \text{supp}(\mathcal{P}_c(\mathbf{z}))$. If $G(\cdot)$ has superadditivity ratio β_{k^*} and restricted inverse curvature θ, then the following inequality holds for any $\tilde{c} \geq \frac{c^*}{\theta\beta_{k^*}} \log(\|\mathbf{z}_{\mathsf{S}^*}\|_2^2/\epsilon) + \rho$ and $c > \tilde{c} + \rho$:*

$$
\frac{\|\mathbf{z}_{\mathsf{S}^*\setminus\mathsf{S}}\|_2^2}{c^*} \leq \frac{\|\mathbf{z}_{\mathsf{S}\setminus\mathsf{S}^*}\|_2^2 + \epsilon}{\theta\beta_{k^*}(c - \tilde{c} - \rho)}.
$$

Proof. We define $\tilde{\mathsf{S}}$ as the first subset, S, in the loops of Algorithm 1 that satisfy $G(\mathsf{S}) > \tilde{c} - \rho$; note that $\tilde{c} - \rho < G(\tilde{\mathsf{S}}) \leq \tilde{c}$ holds. As in the proof of Theorem 1, we have

$$
F(\tilde{\mathsf{S}}) \geq \left(1 - \exp\left(-\theta\beta_{k^*} \cdot \frac{\tilde{c}-\rho}{c^*}\right)\right) F(\mathsf{S}^*).
$$

Therefore, from $\tilde{c} \geq \frac{c^*}{\theta\beta_{k^*}} \log(\|\mathbf{z}_{\mathsf{S}^*}\|_2^2/\epsilon) + \rho = \frac{c^*}{\theta\beta_{k^*}} \log(F(\mathsf{S}^*)/\epsilon) + \rho$, we obtain

$$F(\tilde{\mathsf{S}}) \geq F(\mathsf{S}^*) - \epsilon. \tag{A2}$$

We then suppose that $t+1$ elements are added to $\tilde{\mathsf{S}}$ in the loops of Algorithm 1 when $G(\mathsf{S}) > c - \rho$ occurs for the first time. We let j_i be the i-th element added to $\tilde{\mathsf{S}}$. We define $\hat{\mathsf{S}}_i \coloneqq \{j_1, \ldots, j_i\}$ for $i \in [t+1]$ and $\hat{\mathsf{S}}_0 \coloneqq \emptyset$. As discussed in the proof of Theorem 1, we have

$$\tilde{c} - \rho \leq G(\hat{\mathsf{S}}_0 \cup \tilde{\mathsf{S}}) \leq \cdots \leq G(\hat{\mathsf{S}}_{t+1} \cup \tilde{\mathsf{S}}) = G(\hat{\mathsf{S}}_t \cup \tilde{\mathsf{S}}) + G(j_{t+1} \mid \hat{\mathsf{S}}_t \cup \tilde{\mathsf{S}}) \leq c - \rho + \rho = c.$$

This inequality means that the budget constraint is not violated for $i \in [t+1]$, and thus j_i is added in the $(|\tilde{\mathsf{S}}| + i)$-th iteration of Algorithm 1. In particular, we have $c - \rho \leq G(\hat{\mathsf{S}}_{t+1} \cup \tilde{\mathsf{S}}) \leq c$. Furthermore, the output, S, satisfies $\hat{\mathsf{S}}_i \cup \tilde{\mathsf{S}} \subseteq \mathsf{S}$ for $i = 0, \ldots, t+1$. Since j_i is chosen greedily w.r.t. the cost-benefit ratio, we have $\frac{F(j_i|\hat{\mathsf{S}}_{i-1}\cup\tilde{\mathsf{S}})}{G(j_i|\hat{\mathsf{S}}_{i-1}\cup\tilde{\mathsf{S}})} \geq \frac{F(j|\hat{\mathsf{S}}_{i-1}\cup\tilde{\mathsf{S}})}{G(j|\hat{\mathsf{S}}_{i-1}\cup\tilde{\mathsf{S}})}$ for any $j \in \mathsf{S}^* \backslash \{\hat{\mathsf{S}}_{i-1} \cup \tilde{\mathsf{S}}\}$. Therefore, we obtain

$$F(j_i \mid \hat{\mathsf{S}}_{i-1} \cup \tilde{\mathsf{S}}) \sum_{j \in \mathsf{S}^* \backslash \{\hat{\mathsf{S}}_{i-1} \cup \tilde{\mathsf{S}}\}} G(j \mid \hat{\mathsf{S}}_{i-1} \cup \tilde{\mathsf{S}}) \geq G(j_i \mid \hat{\mathsf{S}}_{i-1} \cup \tilde{\mathsf{S}}) \sum_{j \in \mathsf{S}^* \backslash \{\hat{\mathsf{S}}_{i-1} \cup \tilde{\mathsf{S}}\}} F(j \mid \hat{\mathsf{S}}_{i-1} \cup \tilde{\mathsf{S}}). \tag{A3}$$

We can bound from below $\sum_{j \in \mathsf{S}^* \backslash \{\hat{\mathsf{S}}_{i-1} \cup \tilde{\mathsf{S}}\}} F(j \mid \hat{\mathsf{S}}_{i-1} \cup \tilde{\mathsf{S}})$ in RHS as follows

$$\sum_{j \in \mathsf{S}^* \backslash \{\hat{\mathsf{S}}_{i-1} \cup \tilde{\mathsf{S}}\}} F(j \mid \hat{\mathsf{S}}_{i-1} \cup \tilde{\mathsf{S}}) = \sum_{j \in \mathsf{S}^* \backslash \{\hat{\mathsf{S}}_{i-1} \cup \tilde{\mathsf{S}}\}} |\mathbf{z}_j|^2 \geq \|\mathbf{z}_{\mathsf{S}^* \backslash \mathsf{S}}\|_2^2.$$

On the other hand, $\sum_{j \in \mathsf{S}^* \backslash \{\hat{\mathsf{S}}_{i-1} \cup \tilde{\mathsf{S}}\}} G(j \mid \hat{\mathsf{S}}_{i-1} \cup \tilde{\mathsf{S}})$ in LHS of (A3) can be bounded from above as follows:

$$\begin{aligned}
&\sum_{j \in \mathsf{S}^* \backslash \{\hat{\mathsf{S}}_{i-1} \cup \tilde{\mathsf{S}}\}} G(j \mid \hat{\mathsf{S}}_{i-1} \cup \tilde{\mathsf{S}}) \\
&\leq \frac{1}{\theta} \sum_{j \in \mathsf{S}^* \backslash \{\hat{\mathsf{S}}_{i-1} \cup \tilde{\mathsf{S}}\}} G(j) && \because \text{ definition of restricted inverse curvature} \\
&\leq \frac{1}{\theta\beta_{k^*}} G(\mathsf{S}^* \backslash \{\hat{\mathsf{S}}_{i-1} \cup \tilde{\mathsf{S}}\}) && \because \text{definition of superadditivity ratio} \\
&\leq \frac{c^*}{\theta\beta_{k^*}}. && \because \; G(\mathsf{S}^* \backslash \{\hat{\mathsf{S}}_{i-1} \cup \tilde{\mathsf{S}}\}) \leq c^*
\end{aligned}$$

Consequently, we have

$$\begin{aligned}
F(\hat{\mathsf{S}}_i \cup \tilde{\mathsf{S}}) - F(\hat{\mathsf{S}}_{i-1} \cup \tilde{\mathsf{S}}) &= F(j_i \mid \hat{\mathsf{S}}_{i-1} \cup \tilde{\mathsf{S}}) \\
&\geq \theta\beta_{k^*} \cdot \frac{G(j_i \mid \hat{\mathsf{S}}_{i-1} \cup \tilde{\mathsf{S}})}{c^*} \|\mathbf{z}_{\mathsf{S}^* \backslash \mathsf{S}}\|_2^2 = \theta\beta_{k^*} \cdot \frac{G(\hat{\mathsf{S}}_i \cup \tilde{\mathsf{S}}) - G(\hat{\mathsf{S}}_{i-1} \cup \tilde{\mathsf{S}})}{c^*} \|\mathbf{z}_{\mathsf{S}^* \backslash \mathsf{S}}\|_2^2.
\end{aligned}$$

Taking the summation of both sides for $i = 1, \ldots, t+1$, and using $G(\hat{\mathsf{S}}_{t+1} \cup \tilde{\mathsf{S}}) - G(\tilde{\mathsf{S}}) \geq (c - \rho) - \tilde{c}$ and $\hat{\mathsf{S}}_{t+1} \cup \tilde{\mathsf{S}} \subseteq \mathsf{S}$, we obtain

$$\begin{aligned}
F(\mathsf{S}) - F(\tilde{\mathsf{S}}) &\geq F(\hat{\mathsf{S}}_{t+1} \cup \tilde{\mathsf{S}}) - F(\tilde{\mathsf{S}}) \\
&\geq \theta\beta_{k^*} \cdot \frac{G(\hat{\mathsf{S}}_{t+1} \cup \tilde{\mathsf{S}}) - G(\tilde{\mathsf{S}})}{c^*} \|\mathbf{z}_{\mathsf{S}^* \backslash \mathsf{S}}\|_2^2 \geq \theta\beta_{k^*} \cdot \frac{c - \tilde{c} - \rho}{c^*} \|\mathbf{z}_{\mathsf{S}^* \backslash \mathsf{S}}\|_2^2.
\end{aligned} \tag{A4}$$

Combining inequalities (A2) and (A4), we obtain the target inequality as follows:

$$\frac{\|\mathbf{z}_{\mathsf{S}^* \backslash \mathsf{S}}\|_2^2}{c^*} \leq \frac{F(\mathsf{S}) - F(\tilde{\mathsf{S}})}{\theta\beta_{k^*}(c - \tilde{c} - \rho)} \leq \frac{F(\mathsf{S}) - F(\mathsf{S}^*) + \epsilon}{\theta\beta_{k^*}(c - \tilde{c} - \rho)} \leq \frac{\|\mathbf{z}_{\mathsf{S} \backslash \mathsf{S}^*}\|_2^2 + \epsilon}{\theta\beta_{k^*}(c - \tilde{c} - \rho)},$$

where the last inequality comes from $F(\mathsf{S}) - F(\mathsf{S}^*) \leq F(\mathsf{S} \cup \mathsf{S}^*) - F(\mathsf{S}^*) = \|\mathbf{z}_{\mathsf{S} \backslash \mathsf{S}^*}\|_2^2$. □

Using the above lemma, we obtain the main theorem:

Theorem 2. *Let* $k \coloneqq \max_{t:0\le t\le T}\|\mathbf{x}_t\|_0$ *and* $\omega \coloneqq \max_{t:0\le t\le T}\|\mathbf{g}_t\|_2$. *Assume that* $l(\cdot)$ *is continuously twice differentiable,* μ_{2k+k^*}*-RSC, and* ν_{2k+k^*}*-RSM, and that* $G(\cdot)$ *has superadditivity ratio* β_{k^*} *and restricted inverse curvature* θ. *Set* $\eta = \frac{1}{\nu_{2k+k^*}}$. *If* $c^* \ge \rho$ *and* $c \ge \frac{4c^*}{\theta\beta_{k^*}}\left(\frac{\nu_{2k+k^*}}{\mu_{2k+k^*}}\right)^2 + \frac{2c^*}{\theta\beta_{k^*}}\log\left(\frac{\omega}{2\epsilon}\right) + 2\rho$ *hold, then we have*

$$\|\mathbf{x}_{t+1}-\mathbf{x}^*\|_2 \le \left(1-\frac{1}{2}\cdot\frac{\mu_{2k+k^*}}{\nu_{2k+k^*}}\right)\|\mathbf{x}_t-\mathbf{x}^*\|_2 + \zeta + \frac{\mu_{2k+k^*}}{\nu_{2k+k^*}}\cdot\epsilon,$$

where $\zeta \coloneqq \frac{1}{\nu_{2k+k^*}}\left(1+\frac{1}{2}\cdot\frac{\mu_{2k+k^*}}{\nu_{2k+k^*}}\right)\max_{\mathsf{S}\in\mathcal{F}}\|\nabla l(\mathbf{x}^*)_{\mathsf{S}}\|_2$. *Specifically, after*

$$T \ge 2\cdot\frac{\nu_{2k+k^*}}{\mu_{2k+k^*}}\log\frac{\|\mathbf{x}_0-\mathbf{x}^*\|_2}{\epsilon}$$

steps, we have

$$\|\mathbf{x}_T-\mathbf{x}^*\|_2 \le 3\epsilon + 2\zeta\cdot\frac{\nu_{2k+k^*}}{\mu_{2k+k^*}}.$$

Proof. Let $\mathsf{S}^* \coloneqq \mathrm{supp}(\mathbf{x}^*)$, $\mathsf{S}_t \coloneqq \mathrm{supp}(\mathbf{x}_t)$, $\mathsf{S}_{t+1} \coloneqq \mathrm{supp}(\mathbf{x}_{t+1})$, and $\mathsf{U} \coloneqq \mathsf{S}_{t+1}\cup\mathsf{S}^*$. In what follows, given any $\mathsf{A},\mathsf{B}\subseteq[d]$, if the inequality of RSC holds with $\Omega = \{(\mathbf{x},\mathbf{y}) \mid \mathrm{supp}(\mathbf{x})\subseteq\mathsf{A}, \mathrm{supp}(\mathbf{y})\subseteq\mathsf{B}\}$, then we say $l(\cdot)$ is $\mu_{\mathsf{A},\mathsf{B}}$-RSC.

We first prove an inequality for later use. Note that $l(\cdot)$ is assumed to be twice differentiable. We let $\mathbf{H}(\mathbf{x})$ denote the Hessian of $l(\cdot)$ evaluated at $\mathbf{x}$. Given $\mathsf{A},\mathsf{B}\subseteq[d]$ such that $\mathsf{A}\subseteq\mathsf{B}$, if $l(\cdot)$ is $\mu_{\mathsf{A},\mathsf{B}}$-RSC, then the inequality of RSC implies that function $l(\mathbf{x}) - \frac{\mu_{\mathsf{A},\mathsf{B}}}{2}\|\mathbf{x}\|_2^2$ is convex at $\mathbf{x}$ w.r.t. direction $\mathbf{d}$, where $\mathrm{supp}(\mathbf{x})\subseteq\mathsf{A}$ and $\mathrm{supp}(\mathbf{d})\subseteq\mathsf{B}$; i.e., $\mathbf{H}(\mathbf{x})_{\mathsf{B},\mathsf{B}} \succeq \mu_{\mathsf{A},\mathsf{B}}\mathbf{I}_{\mathsf{B},\mathsf{B}}$ holds. This fact means that, for any $\mathbf{y}\in\mathbb{R}^{[d]}$ such that $\mathrm{supp}(\mathbf{y})\subseteq\mathsf{S}^*\cup\mathsf{S}_t$, the spectral norm of $(\mathbf{I}-\eta\mathbf{H}(\mathbf{y}))_{(\mathsf{U}\cup\mathsf{S}_t),(\mathsf{U}\cup\mathsf{S}_t)}$ is bounded from above by $1-\eta\mu_{(\mathsf{S}^*\cup\mathsf{S}_t),(\mathsf{U}\cup\mathsf{S}_t)}$. Therefore, by using the mean value inequality, we obtain

$$\begin{aligned}
\|(\mathbf{x}^*-\mathbf{x}_t-\eta(\nabla l(\mathbf{x}^*)-\nabla l(\mathbf{x}_t)))_{\mathsf{U}}\|_2 &\le \|(\mathbf{x}^*-\mathbf{x}_t-\eta(\nabla l(\mathbf{x}^*)-\nabla l(\mathbf{x}_t)))_{\mathsf{U}\cup\mathsf{S}_t}\|_2 \\
&\le (1-\eta\mu_{(\mathsf{S}^*\cup\mathsf{S}_t),(\mathsf{U}\cup\mathsf{S}_t)})\|(\mathbf{x}_t-\mathbf{x}^*)\|_2 \\
&\le (1-\eta\mu_{2k+k^*})\|(\mathbf{x}_t-\mathbf{x}^*)\|_2 \\
&= \left(1-\frac{\mu_{2k+k^*}}{\nu_{2k+k^*}}\right)\|(\mathbf{x}_t-\mathbf{x}^*)\|_2.
\end{aligned} \tag{A5}$$

We then evaluate the performance of the CBG projection by bounding $\|(\mathbf{x}_{t+1}-\mathbf{g}_t)_{\mathsf{U}}\|_2^2$ from above. Since $\mathbf{x}_{t+1} = \mathcal{P}_c(\mathbf{g}_t)$, we have $\|(\mathbf{x}_{t+1}-\mathbf{g}_t)_{\mathsf{U}}\|_2^2 = \|(\mathbf{g}_t)_{\mathsf{S}^*\setminus\mathsf{S}_{t+1}}\|_2^2$. From Lemma 3 with $\mathbf{z}=\mathbf{g}_t$, $\mathsf{S}^* = \mathrm{supp}(\mathbf{x}^*)$, $\mathsf{S}=\mathsf{S}_{t+1}$, and $\tilde{c} = \frac{2c^*}{\theta\beta_{k^*}}\log\left(\frac{\omega}{2\epsilon}\right)+\rho \ge \frac{c^*}{\theta\beta_{k^*}}\log\left(\|\mathbf{g}_t\|_2^2/(4\epsilon^2)\right)+\rho$, the following inequality holds for $c > \tilde{c}+\rho$:

$$\begin{aligned}
\|(\mathbf{x}_{t+1}-\mathbf{g}_t)_{\mathsf{U}}\|_2^2 = \|(\mathbf{g}_t)_{\mathsf{S}^*\setminus\mathsf{S}_{t+1}}\|_2^2 &\le \frac{c^*}{\theta\beta_{k^*}(c-\tilde{c}-\rho)}\|(\mathbf{g}_t)_{\mathsf{S}_{t+1}\setminus\mathsf{S}^*}\|_2^2 + \frac{4c^*\epsilon^2}{\theta\beta_{k^*}(c-\tilde{c}-\rho)} \\
&\le \frac{c^*}{\theta\beta_{k^*}(c-\tilde{c}-\rho)}(\|(\mathbf{g}_t)_{\mathsf{S}_{t+1}\setminus\mathsf{S}^*}\|_2^2 + \|(\mathbf{x}^*-\mathbf{g}_t)_{\mathsf{S}^*}\|_2^2) + \frac{4c^*\epsilon^2}{\theta\beta_{k^*}(c-\tilde{c}-\rho)} \\
&= \frac{c^*}{\theta\beta_{k^*}(c-\tilde{c}-\rho)}\|(\mathbf{x}^*-\mathbf{g}_t)_{\mathsf{U}}\|_2^2 + \frac{4c^*\epsilon^2}{\theta\beta_{k^*}(c-\tilde{c}-\rho)}.
\end{aligned} \tag{A6}$$

Setting $c \ge \frac{4c^*}{\theta\beta_{k^*}}\left(\frac{\nu_{2k+k^*}}{\mu_{2k+k^*}}\right)^2 + \frac{2c^*}{\theta\beta_{k^*}}\log\left(\frac{\omega}{2\epsilon}\right) + 2\rho = \frac{4c^*}{\theta\beta_{k^*}}\left(\frac{\nu_{2k+k^*}}{\mu_{2k+k^*}}\right)^2 + \tilde{c}+\rho$, we obtain the target inequality as follows:

$$
\begin{aligned}
&\|\mathbf{x}_{t+1} - \mathbf{x}^*\|_2 \\
&= \|(\mathbf{x}_{t+1} - \mathbf{x}^*)_{\mathsf{U}}\|_2 && \because \mathrm{supp}(\mathbf{x}_{t+1} - \mathbf{x}^*) \subseteq \mathsf{U} \\
&\le \|(\mathbf{x}_{t+1} - \mathbf{g}_t)_{\mathsf{U}}\|_2 + \|(\mathbf{x}^* - \mathbf{g}_t)_{\mathsf{U}}\|_2 && \because \text{triangle inequality} \\
&\le \left(1 + \sqrt{\frac{c^*}{\theta\beta_{k^*}(c - \tilde{c} - \rho)}}\right) \|(\mathbf{x}^* - \mathbf{g}_t)_{\mathsf{U}}\|_2 + 2\epsilon\sqrt{\frac{c^*}{\theta\beta_{k^*}(c - \tilde{c} - \rho)}} && \because \text{(A6) and } \sqrt{x^2 + y^2} \le |x| + |y| \\
&\le \left(1 + \frac{1}{2} \cdot \frac{\mu_{2k+k^*}}{\nu_{2k+k^*}}\right) \|(\mathbf{x}^* - \mathbf{x}_t + \eta\nabla l(\mathbf{x}_t))_{\mathsf{U}}\|_2 + \frac{\mu_{2k+k^*}}{\nu_{2k+k^*}} \cdot \epsilon && \because \text{definitions of } c \text{ and } \mathbf{g}_t \\
&\le \left(1 + \frac{1}{2} \cdot \frac{\mu_{2k+k^*}}{\nu_{2k+k^*}}\right) \|(\mathbf{x}^* - \mathbf{x}_t - \eta(\nabla l(\mathbf{x}^*) - \nabla l(\mathbf{x}_t)))_{\mathsf{U}}\|_2 \\
&\quad + \eta\left(1 + \frac{1}{2} \cdot \frac{\mu_{2k+k^*}}{\nu_{2k+k^*}}\right) \|\nabla l(\mathbf{x}^*)_{\mathsf{S}_{t+1}}\|_2 + \frac{\mu_{2k+k^*}}{\nu_{2k+k^*}} \cdot \epsilon && \because \text{triangle inequality and } \nabla l(\mathbf{x}^*)_{\mathsf{S}^*} = 0 \\
&\le \left(1 + \frac{1}{2} \cdot \frac{\mu_{2k+k^*}}{\nu_{2k+k^*}}\right) \|(\mathbf{x}^* - \mathbf{x}_t - \eta(\nabla l(\mathbf{x}^*) - \nabla l(\mathbf{x}_t)))_{\mathsf{U}}\|_2 + \zeta + \frac{\mu_{2k+k^*}}{\nu_{2k+k^*}} \cdot \epsilon && \because \text{definition of } \zeta \\
&\le \left(1 + \frac{1}{2} \cdot \frac{\mu_{2k+k^*}}{\nu_{2k+k^*}}\right)\left(1 - \frac{\mu_{2k+k^*}}{\nu_{2k+k^*}}\right) \|\mathbf{x}_t - \mathbf{x}^*\|_2 + \zeta + \frac{\mu_{2k+k^*}}{\nu_{2k+k^*}} \cdot \epsilon && \because \text{(A5)} \\
&\le \left(1 - \frac{1}{2} \cdot \frac{\mu_{2k+k^*}}{\nu_{2k+k^*}}\right) \|\mathbf{x}_t - \mathbf{x}^*\|_2 + \zeta + \frac{\mu_{2k+k^*}}{\nu_{2k+k^*}} \cdot \epsilon. && \because \left(1 + \frac{a}{2}\right)(1 - a) \le 1 - \frac{a}{2} \text{ for } a \ge 0
\end{aligned}
$$

We turn to the inequality obtained after T steps. Using the inequality proved above, we obtain

$$
\begin{aligned}
&\|\mathbf{x}_T - \mathbf{x}^*\|_2 \\
&\le \left(1 - \frac{1}{2} \cdot \frac{\mu_{2k+k^*}}{\nu_{2k+k^*}}\right)^T \|\mathbf{x}_0 - \mathbf{x}^*\|_2 + \left(\zeta + \frac{\mu_{2k+k^*}}{\nu_{2k+k^*}} \cdot \epsilon\right) \sum_{t=0}^{T-1} \left(1 - \frac{1}{2} \cdot \frac{\mu_{2k+k^*}}{\nu_{2k+k^*}}\right)^t \\
&\le \left(1 - \frac{1}{2} \cdot \frac{\mu_{2k+k^*}}{\nu_{2k+k^*}}\right)^T \|\mathbf{x}_0 - \mathbf{x}^*\|_2 + 2\zeta \cdot \frac{\nu_{2k+k^*}}{\mu_{2k+k^*}} + 2\epsilon.
\end{aligned}
$$

Setting $T \ge 2 \cdot \frac{\nu_{2k+k^*}}{\mu_{2k+k^*}} \log \frac{\|\mathbf{x}_0 - \mathbf{x}^*\|_2}{\epsilon} \ge \log \frac{\|\mathbf{x}_0 - \mathbf{x}^*\|_2}{\epsilon} \Big/ \log \left(1 - \frac{1}{2} \cdot \frac{\mu_{2k+k^*}}{\nu_{2k+k^*}}\right)^{-1}$, we obtain

$$
\|\mathbf{x}_T - \mathbf{x}^*\|_2 \le 3\epsilon + 2\zeta \cdot \frac{\nu_{2k+k^*}}{\mu_{2k+k^*}}.
$$

□