A. Supplement

In this section, we provide the missing proofs.

A.1. Proof of Theorem 2

Proof. An important aspect of the assumptions is that the space of atoms spanned by S is orthogonal to the span of L. Furthermore, $\mathrm{span}(\mathsf{L} \cup \mathsf{S}) \supset \mathrm{span}(\mathsf{S})$. Let $\bar{k} = k + r$. We will first upper bound the denominator in the submodularity ratio. From strong concavity,

$$\frac{m_{\bar{k}}}{2} \|\mathbf{B}^{(\mathsf{L}\cup\mathsf{S})} - \mathbf{B}^{(\mathsf{L})}\|_F^2 \leq \ell(\mathbf{B}^{(\mathsf{L})}) - \ell(\mathbf{B}^{(\mathsf{L}\cup\mathsf{S})}) + \langle \nabla \ell(\mathbf{B}^{(\mathsf{L})}), \mathbf{B}^{(\mathsf{L}\cup\mathsf{S})} - \mathbf{B}^{(\mathsf{L})} \rangle$$

Rearranging

$$\begin{aligned} 0 \leq \ell(\mathbf{B}^{(\mathsf{L}\cup\mathsf{S})}) - \ell(\mathbf{B}^{(\mathsf{L})}) &\leq \langle \nabla \ell(\mathbf{B}^{(\mathsf{L})}), \mathbf{B}^{(\mathsf{L}\cup\mathsf{S})} - \mathbf{B}^{(\mathsf{L})} \rangle - \frac{m_{\bar{k}}}{2} \|\mathbf{B}^{(\mathsf{L}\cup\mathsf{S})} - \mathbf{B}^{(\mathsf{L})}\|_F^2 \\ &\leq \underset{\substack{\mathbf{X}: \\ \mathbf{X} = \mathbf{U}_{\mathsf{L}\cup\mathsf{S}} \mathbf{H} \mathbf{V}_{\mathsf{L}\cup\mathsf{S}} \\ \mathbf{H} \in \mathbb{R}^{|\mathsf{L}\cup\mathsf{S}| \times |\mathsf{L}\cup\mathsf{S}|}}}{\arg\max} \quad \langle \nabla \ell(\mathbf{B}^{(\mathsf{L})}), \mathbf{X} - \mathbf{B}^{(\mathsf{L})} \rangle - \frac{m_{\bar{k}}}{2} \|\mathbf{X} - \mathbf{B}^{(\mathsf{L})}\|_F^2 \\ &= \underset{\substack{\mathbf{X}: \\ \mathbf{X} = \mathbf{U}_{\mathsf{L}\cup\mathsf{S}} \mathbf{H} \mathbf{V}_{\mathsf{L}\cup\mathsf{S}} \\ \mathbf{H} \in \mathbb{R}^{|\mathsf{L}\cup\mathsf{S}| \times |\mathsf{L}\cup\mathsf{S}|}}}{\arg\max} \quad \langle P_{\mathbf{U}_{\mathsf{S}}}(\nabla \ell(\mathbf{B}^{(\mathsf{L})})) P_{\mathbf{V}_{\mathsf{S}}}, \mathbf{X} - \mathbf{B}^{(\mathsf{L})} \rangle - \frac{m_{\bar{k}}}{2} \|\mathbf{X} - \mathbf{B}^{(\mathsf{L})}\|_F^2, \end{aligned}$$

where the last equality holds because $\langle (\nabla \ell(\mathbf{B}^{(\mathsf{L})})), P_{\mathbf{U}_{\mathsf{L}}} \mathbf{X} P_{\mathbf{V}_{\mathsf{L}}} - \mathbf{B}^{(\mathsf{L})} \rangle = 0$. Solving the argmax problem, we get $\mathbf{X} = \mathbf{B}^{(\mathsf{L})} + \frac{1}{m_{\bar{k}}} P_{\mathbf{U}_{\mathsf{S}}}(\nabla \ell(\mathbf{B}^{(\mathsf{L})})) P_{\mathbf{V}_{\mathsf{S}}}$. Plugging in, we get,

$$\ell(\mathbf{B}^{(\mathsf{L}\cup\mathsf{S})}) - \ell(\mathbf{B}^{(\mathsf{L})}) \leq \frac{1}{2 m_{\bar{k}}} \|P_{\mathbf{U}_{\mathsf{S}}}(\nabla \ell(\mathbf{B}^{(\mathsf{L})})) P_{\mathbf{V}_{\mathsf{S}}}\|_F^2$$

We next bound the numerator. Recall that the atoms in S are orthogonal to each other *i.e.* $\mathbf{U}_{\mathsf{S}}$ and $\mathbf{V}_{\mathsf{S}}$ are both orthonormal.

For clarity, we define the shorthand, $\mathbf{B}_{ij}^{(\mathsf{L}\cup\mathsf{S})} = \langle \mathbf{u}_i \mathbf{v}_j^\top, \mathbf{B}^{(\mathsf{L}\cup\mathsf{S})} \rangle \, \mathbf{u}_i \mathbf{v}_j^\top$, for $i, j \in [|\mathsf{L} \cup \mathsf{S}|]$.

With an arbitrary $i \in \mathsf{S}$, and arbitrary scalars $\alpha_{ii}, \alpha_{ij}, \alpha_{ji}$ for $j \in \mathsf{L}$,

$$\begin{aligned} \ell(\mathbf{B}^{(\mathsf{L}\cup\{i\})}) - \ell(\mathbf{B}^{(\mathsf{L})}) &\geq \ell(\mathbf{B}^{(\mathsf{L})} + \alpha_{ii} \mathbf{B}_{ii}^{(\mathsf{L}\cup\mathsf{S})} + \sum_{j \in \mathsf{L}} \alpha_{ij} \mathbf{B}_{ij}^{(\mathsf{L}\cup\mathsf{S})} + \sum_{j \in \mathsf{L}} \alpha_{ji} \mathbf{B}_{ji}^{(\mathsf{L}\cup\mathsf{S})}) - \ell(\mathbf{B}^{(\mathsf{L})}) \\ &\geq \langle \nabla \ell(\mathbf{B}^{(\mathsf{L})}), \alpha_{ii} \mathbf{B}_{ii}^{(\mathsf{L}\cup\mathsf{S})} + \sum_{j \in \mathsf{L}} \alpha_{ij} \mathbf{B}_{ij}^{(\mathsf{L}\cup\mathsf{S})} + \sum_{j \in \mathsf{L}} \alpha_{ji} \mathbf{B}_{ji}^{(\mathsf{L}\cup\mathsf{S})} \rangle \\ &\quad - \frac{\tilde{M}_1}{2} \left[\alpha_{ii}^2 \|\mathbf{B}_{ii}^{(\mathsf{L}\cup\mathsf{S})}\|_F^2 + \sum_{j \in \mathsf{L}} \alpha_{ij}^2 \|\mathbf{B}_{ij}^{(\mathsf{L}\cup\mathsf{S})}\|_F^2 + \sum_{j \in \mathsf{L}} \alpha_{ji}^2 \|\mathbf{B}_{ji}^{(\mathsf{L}\cup\mathsf{S})}\|_F^2 \right]. \\ &\geq \frac{\langle \nabla \ell(\mathbf{B}^{(\mathsf{L})}), \mathbf{B}_{ii}^{(\mathsf{L}\cup\mathsf{S})} \rangle^2}{2 \tilde{M}_1 \|\mathbf{B}_{ii}^{(\mathsf{L}\cup\mathsf{S})}\|_F^2} + \sum_{j \in \mathsf{L}} \left(\frac{\langle \nabla \ell(\mathbf{B}^{(\mathsf{L})}), \mathbf{B}_{ij}^{(\mathsf{L}\cup\mathsf{S})} \rangle^2}{2 \tilde{M}_1 \|\mathbf{B}_{ij}^{(\mathsf{L}\cup\mathsf{S})}\|_F^2} + \frac{\langle \nabla \ell(\mathbf{B}^{(\mathsf{L})}), \mathbf{B}_{ji}^{(\mathsf{L}\cup\mathsf{S})} \rangle^2}{2 \tilde{M}_1 \|\mathbf{B}_{ji}^{(\mathsf{L}\cup\mathsf{S})}\|_F^2} \right), \end{aligned}$$

where the last inequality follows by setting $\alpha_{ij} = \frac{\langle \nabla \ell(\mathbf{B}^{(\mathsf{L})}), \mathbf{B}_{ij}^{(\mathsf{L}\cup\mathsf{S})} \rangle}{\tilde{M}_1 \|\mathbf{B}_{ij}^{(\mathsf{L}\cup\mathsf{S})}\|_F^2}$ for $j \in \mathsf{L}$, and for $j = i$.

Summing up for all $i \in \mathsf{S}$, we get

$$\sum_{i \in \mathsf{S}} \ell(\mathbf{B}^{(\mathsf{L} \cup \{i\})}) - \ell(\mathbf{B}^{(\mathsf{L})}) \geq \sum_{i \in \mathsf{S}} \left[\frac{\langle \nabla \ell(\mathbf{B}^{(\mathsf{L})}), \mathbf{B}_{ii}^{(\mathsf{L} \cup \mathsf{S})} \rangle^2}{2\tilde{M}_1 \|\mathbf{B}_{ii}^{(\mathsf{L} \cup \mathsf{S})}\|_F^2} + \sum_{j \in \mathsf{L}} \left(\frac{\langle \nabla \ell(\mathbf{B}^{(\mathsf{L})}), \mathbf{B}_{ij}^{(\mathsf{L} \cup \mathsf{S})} \rangle^2}{2\tilde{M}_1 \|\mathbf{B}_{ij}^{(\mathsf{L} \cup \mathsf{S})}\|_F^2} + \frac{\langle \nabla \ell(\mathbf{B}^{(\mathsf{L})}), \mathbf{B}_{ji}^{(\mathsf{L} \cup \mathsf{S})} \rangle^2}{2\tilde{M}_1 \|\mathbf{B}_{ji}^{(\mathsf{L} \cup \mathsf{S})}\|_F^2} \right) \right]$$
$$= \frac{1}{2\tilde{M}_1} \|P_{\mathbf{U}_\mathsf{S}} \nabla \ell(\mathbf{B}^{(\mathsf{L})}) P_{\mathbf{V}_\mathsf{S}}\|_F^2$$

□

A.2. Proofs for greedy improvement

Let S_i^G be the support set formed by Algorithm 1 at iteration i. Define $A(i) := f(\mathsf{S}_i^G) - f(\mathsf{S}_{i-1}^G)$ with $A(0) = 0$ as the greedy improvement. We also define $B(i) := f(\mathsf{S}^*) - f(\mathsf{S}_i^G)$ to be the remaining amount to improve, where $\mathsf{S}^\star$ is the optimum k-sized solution. We provide an auxiliary Lemma that uses the submodularity ratio to lower bound the greedy improvement in terms of best possible improvement from step i.

Lemma 1. *At iteration i, the incremental gain of the greedy method (Algorithm 1) is*

$$A(i+1) \geq \frac{\tau \gamma_{\mathsf{S}_i^G, r}}{r} B(i).$$

Proof. Let $\mathsf{S} = \mathsf{S}_i^G$. Let S^R be the sequential orthogonalization of the atoms in S^* relative to S. Thus,

$$\begin{aligned} rA(i+1) \geq |\mathsf{S}^R| A(i+1) &\geq \tau |\mathsf{S}^R| \max_{a \in \mathsf{S}^R} f(\mathsf{S} \cup \{a\}) - f(\mathsf{S}) \\ &\geq \tau \sum_{a \in \mathsf{S}^R} [f(\mathsf{S} \cup \{a\}) - f(\mathsf{S})] \\ &\geq \tau \gamma_{\mathsf{S}, |\mathsf{S}^R|} [f(\mathsf{S} \cup \mathsf{S}^R) - f(\mathsf{S})] \\ &\geq \tau \gamma_{\mathsf{S}, |\mathsf{S}^R|} B(i) \end{aligned}$$

Note that the last inequality follows because $f(\mathsf{S} \cup \mathsf{S}^R) \geq f(\mathsf{S}^*)$. The penultimate inequality follows by the definition of weak submodularity, which applies in this case because the atoms in S^R are orthogonal to eachother and are also orthogonal to S. □

Using Lemma 1, one can prove an approximation guarantee for Algorithm 1.

A.2.1. Proof of Theorem 3

Proof. From the notation used for Lemma 1, $A(i+1) = B(i) - B(i+1)$. Let $C = \frac{\tau \gamma_{\mathsf{S}_i^G, r}}{r}$. From Lemma 1, we have,

$$B(i+1) \leq (1 - C) B(i) \leq (1 - C)^{i+1} B(0).$$

From its definition, $B(0) = f(\mathsf{S}^\star) - f(\emptyset)$. So we get,

$$\begin{aligned} &[f(\mathsf{S}^\star) - f(\emptyset)] - \left[f(\mathsf{S}_i^G) - f(\emptyset)\right] \leq (1 - C)^i \left[f(\mathsf{S}^\star) - f(\emptyset)\right] \\ \implies &\left[f(\mathsf{S}_i^G) - f(\emptyset)\right] \geq (1 - (1 - C)^i) \left[f(\mathsf{S}^\star) - f(\emptyset)\right] \geq \left(1 - \frac{1}{e^{\tau \gamma_{\mathsf{S}_i^G, r} \frac{k}{r}}}\right) \left[f(\mathsf{S}^\star) - f(\emptyset)\right] \end{aligned}$$

from which the result follows. □

A.3. Proof for GECO bounds

Let S_i^O be the support set selected by the GECO procedure (Algorithm 2) at iteration i. Similar to the section on greedy improvement, we define some notation. Let $D(i) := f(\mathsf{S}_i^O) - f(\mathsf{S}_{i-1}^O)$ be the improvement made at step i, and as before we have $B(i) = f(\mathsf{S}^\star) - f(\mathsf{S}_i^O)$ be the remaining amount to improve.

We prove the following auxiliary lemma which lower bounds the gain after adding the atom selected by the subroutine `OMPSel`in terms of operator norm of the gradient of the current iterate and smoothness of the function.

Lemma 2. *Assume that $\ell(\cdot)$ is m_i-strongly concave and M_i-smooth over matrices of in the set $\tilde{\Omega} := \{(\mathbf{X}, \mathbf{Y}) : \text{rank}(\mathbf{X} - \mathbf{Y}) \leq 1\}$. Then,*

$$D(i+1) \geq \frac{\tau m_{r+k}}{r\tilde{M}_1} B(i).$$

Proof. For simplicity, say $\mathsf{L} = \mathsf{S}_i^O$. Recall that for a given support set L, $f(\mathsf{L}) = \ell(\mathbf{B}^{(\mathsf{L})})$ *i.e.* we denote by $\mathbf{B}^{(\mathsf{L})}$ the argmax for $\ell(\cdot)$ for a given support set L. Hence, by the optimality of $\mathbf{B}^{(\mathsf{L}\cup\{i\})}$,

$$\begin{aligned} D(i+1) &= \ell(\mathbf{B}^{(\mathsf{L}\cup\{i\})}) - \ell(\mathbf{B}^{(\mathsf{L})}) \\ &\geq \ell(\mathbf{B}^{(\mathsf{L})} + \alpha \mathbf{u}\mathbf{v}^\top) - \ell(\mathbf{B}^{(\mathsf{L})}) \end{aligned}$$

for an arbitrary $\alpha \in \mathbb{R}$, and the vectors $\mathbf{u}, \mathbf{v}$ selected by `OMPSel`. Using the smoothness of the $\ell(\cdot)$, we get,

$$D(i+1) \geq \alpha \langle \nabla \ell(\mathbf{B}^{(\mathsf{L})}), \mathbf{u}\mathbf{v}^\top \rangle - \alpha^2 \frac{\tilde{M}_1}{2}$$

Putting in $\alpha = \frac{\tau}{\tilde{M}_1} \|\nabla \ell(\mathbf{B}^{(\mathsf{L})})\|_2$, and by τ-optimality of `OMPSel`, we get,

$$D(i+1) \geq \frac{\tau^2}{2\tilde{M}_1} \|\nabla \ell(\mathbf{B}^{(\mathsf{L})})\|_2^2$$

Let S^R be obtained from after sequentially orthogonalizing $\mathsf{S}^\star$ w.r.t. S_i. By definition of the operator norm, we further get,

$$\begin{aligned} D(i+1) &\geq \frac{\tau^2}{2\tilde{M}_1} \|\nabla \ell(\mathbf{B}^{(\mathsf{L})})\|_2^2 \\ &\geq \frac{\tau^2}{2r\tilde{M}_1} \sum_{i \in \mathsf{S}^R} \langle \mathbf{u}_i \mathbf{v}_i^\top, \nabla \ell(\mathbf{B}^{(\mathsf{L})}) \rangle^2 \\ &= \|P_{\mathbf{U}_{\mathsf{S}^R}} \nabla \ell(\mathbf{B}^{(\mathsf{L})}) P_{\mathbf{V}_{\mathsf{S}^R}}\|_F^2 \\ &\geq \frac{\tau^2 m_{r+k}}{r\tilde{M}_1} \left(\ell(\mathbf{B}^{\mathsf{L}\cup\mathsf{S}^R}) - \ell(\mathbf{B}^{(\mathsf{L})}) \right) \\ &\geq \frac{\tau^2 m_{r+k}}{r\tilde{M}_1} \left(\ell(\mathbf{B}^{\mathsf{S}^\star}) - \ell(\mathbf{B}^{(\mathsf{L})}) \right) \\ &= \frac{\tau^2 m_{r+k}}{r\tilde{M}_1} B(i) \end{aligned}$$

□

The proof for Theorem 4 from Lemma 2 now follows using the same steps as for Theorem 3 from Lemma 2.

A.4. Proof for recovery bounds

A.4.1. Proof of Theorem 5

For clarity of representation, let $C = C_{r,k}$, and for an arbitrary $\mathbf{H} \in \mathbb{R}^{r \times r}$, let $\mathbf{B}_{\mathsf{r}} = \mathbf{U}_{\mathsf{S}}^{\top}\mathbf{H}\mathbf{V}_{\mathsf{S}}$, and $\mathbf{\Delta} := \mathbf{B}^{(\mathsf{S}_r)} - \mathbf{B}_{\mathsf{s}}$. Note that $\mathbf{\Delta}$ has rank at most $(k+r)$. Recall that by the m_{k+r} RSC (Definition 3),

$$\ell(\mathbf{B}^{(\mathsf{S}_k)}) - \ell(\mathbf{B}_{\mathsf{r}}) - \langle \nabla \ell(\mathbf{B}_{\mathsf{r}}), \mathbf{\Delta} \rangle \leq \frac{-m_{k+r}}{2} \|\mathbf{\Delta}\|_F^2.$$

From the approximation guarantee, we have,

$$\begin{aligned}
&\ell(\mathbf{B}^{(\mathsf{S}_k)}) - \ell(\mathbf{B}_{\mathsf{r}}) \geq (1-C)[\ell(\mathbf{0}) - \ell(\mathbf{B}_{\mathsf{r}})] \\
\implies &\ell(\mathbf{B}^{(\mathsf{S}_k)}) - \ell(\mathbf{B}_{\mathsf{r}}) - \langle \nabla \ell(\mathbf{B}_{\mathsf{r}}), \mathbf{\Delta} \rangle \geq (1-C)[\ell(\mathbf{0}) - \ell(\mathbf{B}_{\mathsf{r}})] - \langle \nabla \ell(\mathbf{B}_{\mathsf{r}}), \mathbf{\Delta} \rangle \\
\implies &\frac{-m_{k+r}}{2} \|\mathbf{\Delta}\|_F^2 \geq (1-C)[\ell(\mathbf{0}) - \ell(\mathbf{B}_{\mathsf{r}})] - \langle \nabla \ell(\mathbf{B}_{\mathsf{r}}), \mathbf{\Delta} \rangle \\
&\qquad\qquad\qquad \geq (1-C)[\ell(\mathbf{0}) - \ell(\mathbf{B}_{\mathsf{r}})] - (k+r)^{1/2} \|\nabla \ell(\mathbf{B}_{\mathsf{r}})\|_2 \|\mathbf{\Delta}\|_F,
\end{aligned}$$

where the last inequality is due to generalized Holder's inequality. Using $2ab \leq ca^2 + \frac{b^2}{c}$ for any positive numbers a, b, c, we get

$$\frac{m_{k+r}}{2} \|\mathbf{\Delta}\|_F^2 \leq (k+r) \frac{\|\nabla \ell(\mathbf{B}_{\mathsf{r}})\|_2^2}{m_{k+r}} + \frac{m_{k+r} \|\mathbf{\Delta}\|_F^2}{4} + (1-C)[\ell(\mathbf{B}_{\mathsf{r}}) - \ell(\mathbf{0})],$$

which completes the proof.