Supplementary Material

8 Proof of Lemma 1

The proof follows along the same lines of the either of the two more general Lemmas proved below: Lemma 2 which considers the multiplicative approximation case, and Lemma 8 which considers the regularized case.

9 Proof of Lemma 2

$$\begin{aligned}\mathcal{L}(w_{t+1}) - \mathcal{L}(w^t) &= -\frac{1}{\kappa_1}|[\nabla\mathcal{L}(w^t)]_{j_t}|^2 \\ &\leq -\frac{c}{\kappa_1}\|\nabla\mathcal{L}(w^t)\|_\infty^2 \\ &\leq -\frac{c}{\kappa_1\|w^0 - w^\star\|_1^2}(\mathcal{L}(w^t) - \mathcal{L}(w^\star))^2 \ , \end{aligned} \tag{9}$$

where we used

$$\begin{aligned}\mathcal{L}(w^t) - \mathcal{L}(w^\star) &\leq \left\langle \nabla\mathcal{L}(w^t), w^t - w^\star \right\rangle \\ &\leq \|\nabla\mathcal{L}(w^t)\|_\infty \cdot \|w^t - w^\star\|_1.\end{aligned}$$

The recursion (9) then gives us the result.

10 Proof of Lemma 8

Lemma 10. The greedy coordinate descent iterates of Algorithm 2 satisfy:

$$\mathcal{L}(w^t) + \mathcal{R}(w^t) - \mathcal{L}(w^*) - \mathcal{R}(w^*) \leq \frac{\kappa_1}{2}\frac{\|w^0 - w^*\|_1^2}{t}.$$

Proof. As shorthand, we use w', w, j for w^{t+1}, w^t, j_t. Note that $|\eta_j| = \|\eta\|_\infty$ by definition of j_t. Now, η_j satisfies $g_j + \kappa_1\eta_j + \rho_j = 0$, for some $\rho \in \partial R(w')$. So,

$$\begin{aligned}R(w') - R(w) &= R_j(w'_j) - R_j(w_j) \\ &\leq \langle \rho_j, \eta_j \rangle \ = \ -\langle g_j, \eta_j \rangle - \kappa_1\eta_j^2 \ .\end{aligned}$$

Using this, we have

$$\begin{aligned}\mathcal{L}(w') + R(w') &\leq \mathcal{L}(w) + g_j\eta_j + \frac{\kappa_1}{2}\eta_j^2 + R(w') \\ &\leq \mathcal{L}(w) + R(w) - \frac{\kappa_1}{2}\eta_j^2 \\ &= \mathcal{L}(w) + R(w) - \frac{\kappa_1}{2}\|\eta\|_\infty^2 \ .\end{aligned} \tag{10}$$

Now let $g' = \nabla\mathcal{L}(w')$ to get,

$$\begin{aligned}\mathcal{L}(w') - \mathcal{L}(w) &\leq \langle g', w' - w \rangle \\ &= \langle g' - g, w' - w \rangle + \langle g, w' - w \rangle \\ &\leq \eta_j^2\kappa_1 + \langle g, w' - w \rangle \ ,\end{aligned}$$

where the last inequality is because $\|g' - g\| \leq \kappa_1\|w' - w\|$ and $\|w' - w\| = |\eta_j|$. Combining this with the fact that $\mathcal{L}(w) - \mathcal{L}(w^\star) \leq \langle g, w - w^\star \rangle$ gives,

$$\mathcal{L}(w') - \mathcal{L}(w^\star) \leq \eta_j^2\kappa_1 + \langle g, w' - w^\star \rangle \ .$$

Addding to this the inequality, $R(w') - R(w) \le \langle \rho, w' - w^\star \rangle$ gives

$$\begin{aligned} \epsilon' &:= \mathcal{L}(w') + R(w') - \mathcal{L}(w^\star) - R(w^\star) \\ &\le \eta_j^2 \kappa_1 + \langle \rho + g, w' - w^\star \rangle \\ &\le \eta_j^2 \kappa_1 + \|\rho + g\|_\infty D \\ &= \|\eta\|_\infty^2 \kappa_1 + \kappa_1 \|\eta\|_\infty D \ , \end{aligned}$$

where $D := \|w^0 - w^\star\|_1$. Assuming that $\eta_j \le D$ (note that $D = O(\sqrt{s})$ is at least lower-bounded by a constant, and the objective can reduce by such a large magnitude $\eta_j > D$ atmost finite number of times), we get the key inequality

$$\epsilon' \le 2\kappa_1 \|\eta\|_\infty D \ .$$

Plugging this back in (10), we get the recurrence

$$\epsilon_{t+1} \le \epsilon_t - \frac{(\epsilon_{t+1})^2}{8\kappa_1 D^2} \ .$$

This yields $\epsilon_t \le O(\kappa_1 D^2 / t)$ as required. □

11 Proof of Lemma 3

Proof. Denote $\bar{r} = r / \|r\|_2$. Suppose $\bar{x}_k$ is a $(1 + \epsilon_{\text{nn}})$ multiplicative factor approximation to the greedy step $\max_j \langle \bar{x}, \bar{r} \rangle$. Then

$$\|\bar{x}_k - \bar{r}\|_2^2 \le (1 + \epsilon_{\text{nn}}) \|\bar{x}_j - \bar{r}\|_2^2,$$

so that $\langle \bar{x}_j, \bar{r} \rangle \le \frac{\epsilon_{\text{nn}}}{(1+\epsilon_{\text{nn}})} + \frac{1}{(1+\epsilon_{\text{nn}})} \langle \bar{x}_k, \bar{r} \rangle$.

Thus if $\langle \bar{x}_k, \bar{r} \rangle > \epsilon$, then

$$\begin{aligned} \langle \bar{x}_j, \bar{r} \rangle &\le \frac{\epsilon_{\text{nn}}}{(1 + \epsilon_{\text{nn}})} \frac{\langle \bar{x}_k, \bar{r} \rangle}{\epsilon} + \frac{1}{(1 + \epsilon_{\text{nn}})} \langle \bar{x}_k, \bar{r} \rangle \\ &= \frac{\epsilon_{\text{nn}}(1/\epsilon) + 1}{(1 + \epsilon_{\text{nn}})} \langle \bar{x}_k, \bar{r} \rangle \ , \end{aligned}$$

which completes the proof. □