A PROOF OF LEMMA 2

Proof. The proof technique is standard, and can be found in Zinkevich (2003); Hazan et al. (2016).

First, we prove the regret bound of (21). Note that by Definition 2, $s_t^\eta(\mathbf{x})$ is $2\eta^2 G^2$-strongly convex. For convince, we denote $\alpha_{t+1} = 1/(2\eta^2 G^2 t)$, $\lambda^s = 2\eta^2 G^2$, and define the upper bound of the gradients of $s_t^\eta(\mathbf{x})$ as

$$\max_{\mathbf{x}\in\mathcal{D}} \|\nabla s_t^\eta(\mathbf{x})\| = \max_{\mathbf{x}\in\mathcal{D}} \|\eta \mathbf{g}_t + 2\eta^2 G^2(\mathbf{x}-\mathbf{x}_t)\| \le G\eta + 2\eta^2 G^2 D =: G^s.$$

By the update rule of $\mathbf{x}_{t+1}^{\eta,s}$, we have

$$\begin{aligned}
\|\mathbf{x}_{t+1}^{\eta,s} - \mathbf{u}\| =& \left\|\Pi_{\mathcal{D}}^{I_d}\left(\mathbf{x}_t^{\eta,s} - \alpha_{t+1}\nabla s_t^\eta(\mathbf{x}_t^{\eta,s})\right) - \mathbf{u}\right\| \\
\le& \|\mathbf{x}_t^{\eta,s} - \alpha_{t+1}\nabla s_t^\eta(\mathbf{x}_t^{\eta,s}) - \mathbf{u}\| \\
=& \|\mathbf{x}_t^{\eta,s} - \mathbf{u}\|^2 + \alpha_{t+1}^2 \|\nabla s_t^\eta(\mathbf{x}_t^{\eta,s})\|^2 - 2\alpha_{t+1}(\mathbf{x}_t^{\eta,s} - \mathbf{u})^\top \nabla s_t^\eta(\mathbf{x}_t^{\eta,s}).
\end{aligned} \tag{28}$$

Hence,

$$2(\mathbf{x}_t^{\eta,s} - \mathbf{u})^\top \nabla s_t^\eta(\mathbf{x}_t^{\eta,s}) \le \frac{\|\mathbf{x}_t^{\eta,s} - \mathbf{u}\| - \|\mathbf{x}_{t+1}^{\eta,s} - \mathbf{u}\|^2}{\alpha_{t+1}} + \alpha_{t+1}(G^s)^2. \tag{29}$$

Summing over 1 to T and applying definition 2, we get

$$\begin{aligned}
2\sum_{t=1}^T s_t^\eta(\mathbf{x}_t^{\eta,s}) - 2\sum_{t=1}^T s_t^\eta(\mathbf{u}) \le& \sum_{t=1}^T \|\mathbf{x}_t^{\eta,s} - \mathbf{u}\|^2 \left(\frac{1}{\alpha_{t+1}} - \frac{1}{\alpha_t} - \lambda^s\right) + (G^s)^2 \sum_{t=1}^T \alpha_{t+1} \\
\le& \frac{(G^s)^2}{\lambda^s}(1+\log T).
\end{aligned} \tag{30}$$

Note that $\eta \le \frac{1}{5DG}$. We have

$$(G^s)^2 = G^2\eta^2 + 4\eta^3 G^3 D + 4\eta^4 G^4 D^2 \le G^2\eta^2 + \frac{4\eta^2 G^2}{5} + \frac{4\eta^2 G^2}{25} \le 2\eta^2 G^2 = \lambda^s. \tag{31}$$

Next, we prove the regret bound of (22). We start with the following inequality

$$\begin{aligned}
\nabla \ell_t^\eta(\mathbf{x})(\nabla \ell_t^\eta(\mathbf{x}))^\top =& \eta^2 \mathbf{g}_t \mathbf{g}_t^\top + 4\eta^3 \mathbf{g}_t (\mathbf{x}-\mathbf{x}_t)^\top \mathbf{g}_t \mathbf{g}_t^\top + 4\eta^4 \mathbf{g}_t \mathbf{g}_t^\top (\mathbf{x}-\mathbf{x}_t)(\mathbf{x}-\mathbf{x}_t)^\top \mathbf{g}_t \mathbf{g}_t^\top \\
=& \eta^2 \mathbf{g}_t \mathbf{g}_t^\top + \mathbf{g}_t \left(4\eta^3 (\mathbf{x}-\mathbf{x}_t)^\top \mathbf{g}_t + 4\eta^4 \left((\mathbf{x}-\mathbf{x}_t)^\top \mathbf{g}_t\right)^2\right) \mathbf{g}_t^\top \\
\preceq& 2\eta^2 \mathbf{g}_t \mathbf{g}_t^\top = \nabla^2 \ell_t^\eta(\mathbf{x})
\end{aligned} \tag{32}$$

where $\nabla^2 \ell_t^\eta(\mathbf{x})$ denotes the Hessian matrix. The inequality implies that $\nabla^2 \ell_t^\eta(\mathbf{x}) \succeq \nabla \ell_t^\eta(\mathbf{x})(\nabla \ell_t^\eta(\mathbf{x}))^\top$. According to Lemma 4.1 in Hazan et al. (2016), $\ell_t^\eta(\mathbf{x})$ is 1-exp-concave. Next, we prove that the gradient of $\ell_t^\eta(\mathbf{x})$ can be upper bounded as follows

$$\max_{\mathbf{x}\in\mathcal{D}} \|\nabla \ell_t^\eta(\mathbf{x})\| \le \eta G + 2\eta^2 G^2 D \le \frac{7}{25D} = G^\ell. \tag{33}$$

By Theorem 4.3 in Hazan et al. (2016), we have

$$\sum_{t=1}^T \ell_t^\eta(\mathbf{x}_t^{\eta,\ell}) - \sum_{t=1}^T \ell_t^\eta(\mathbf{u}) \le 5(1 + G^\ell D) d\log T \le 10 d \log T. \tag{34}$$

Finally, we prove the regret bound of (23). Note that the gradient of $c_t(\mathbf{x})$ is upper bounded by $\max_{\mathbf{x}\in\mathcal{D}} \|\nabla c_t(\mathbf{x})\| \le \eta^c G$. Define $m_t = \frac{D}{\eta^c G\sqrt{t}}$. By the convexity of $c_t(\mathbf{x})$, we have $\forall \mathbf{u} \in \mathcal{D}$,

$$c_t(\mathbf{x}_t^c) - c_t(\mathbf{u}) \le (\mathbf{x}_t^c - \mathbf{u})^\top \nabla c_t(\mathbf{x}_t^c). \tag{35}$$

On the other hand, according to the update rule of $\mathbf{x}_{t+1}^c$, we have

$$\begin{aligned}
\|\mathbf{x}_{t+1}^c - \mathbf{u}\|^2 =& \|\Pi_{\mathcal{D}}^{I_d}\left(\mathbf{x}_t^c - m_t \nabla c_t(\mathbf{x}_t^c)\right) - \mathbf{u}\|^2 \\
\leq& \|\mathbf{x}_t^c - m_t \nabla c_t\left(\mathbf{x}_t^c\right) - \mathbf{u}\|^2 \\
=& \|\mathbf{x}_t^c - \mathbf{u}\|^2 + m_t^2 \|\nabla c_t\left(\mathbf{x}_t^c\right)\|^2 - 2m_t\left(\mathbf{x}_t^c - \mathbf{u}\right)^\top \nabla c_t\left(\mathbf{x}_t^c\right)
\end{aligned} \tag{36}$$

where the inequality follows from Theorem 2.1 in Hazan et al. (2016). Hence,

$$\begin{aligned}
& 2\left(\mathbf{x}_t^c - \mathbf{u}\right)^\top \nabla c_t\left(\mathbf{x}_t^c\right) \\
\leq& \frac{\|\mathbf{x}_t^c - \mathbf{u}\|^2 - \|\mathbf{x}_{t+1}^c - \mathbf{u}\|^2}{m_t} + m_t \|\nabla c_t\left(\mathbf{x}_t^c\right)\|^2 \\
\leq& \frac{\|\mathbf{x}_t^c - \mathbf{u}\|^2 - \|\mathbf{x}_{t+1}^c - \mathbf{u}\|^2}{m_t} + m_t (\eta^c G)^2
\end{aligned} \tag{37}$$

Substituting the above inequality into (35) and summing over T, we have

$$\begin{aligned}
\sum_{t=1}^T c_t(\mathbf{x}_t^c) - c_t(\mathbf{u}) \overset{(2)}{\leq}& \sum_{t=1}^T \left(\mathbf{x}_t^c - \mathbf{u}\right)^\top \nabla c_t\left(\mathbf{x}_t^c\right) \\
\leq& \frac{1}{2}\sum_{t=1}^T \|\mathbf{x}_t^c - \mathbf{u}\|^2 \left(\frac{1}{m_t} - \frac{1}{m_{t-1}}\right) + \frac{(\eta^c G)^2}{2} \sum_{t=1}^T m_t \\
\leq& D^2 \frac{1}{2m_T} + \frac{(\eta^c G)^2}{2} \sum_{t=1}^T m_t \\
\leq& \frac{3}{2} \eta^c G D \sqrt{T} \leq \frac{3}{4}
\end{aligned} \tag{38}$$

where the last inequality is due to $\eta^c = \frac{1}{2GD\sqrt{T}}$. $\square$