Appendix A Proof of $\mathbf{g}_t^\top(\mathbf{x}_t^* - \mathbf{x}_{t+1}^*) \le 2\eta\|\mathbf{g}_t\|^2$

Lemma 6 (Theorem 5.1 in (Hazan, 2016))**.** *Let* $\mathbf{x}_t^* = \arg\min_{\mathbf{x}\in(1-\alpha)\mathcal{K}} F_t(\mathbf{x})$. *We have* $\mathbf{g}_t^\top(\mathbf{x}_t^* - \mathbf{x}_{t+1}^*) \le 2\eta\|\mathbf{g}_t\|^2$.

Proof. We denote the regularizer in line 6 of Algorithm 1 by $R(\mathbf{x}) \triangleq \|\mathbf{x}-\mathbf{x}_1\|^2$ and define the Bregman divergence with respect the function F by

$$B_F(\mathbf{x}\|\mathbf{y}) = F(\mathbf{x}) - F(\mathbf{y}) - \nabla F(\mathbf{y})^\top(\mathbf{x}-\mathbf{y}). \tag{14}$$

Since $\mathbf{x}_{t+1}^*$ is a minimizer of F_{t+1} and F_{t+1} is convex, we have

$$\begin{aligned} F_{t+1}(\mathbf{x}_t^*) &= F_{t+1}(\mathbf{x}_{t+1}^*) + (\mathbf{x}_t^* - \mathbf{x}_{t+1}^*)^\top \nabla F_{t+1}(\mathbf{x}_{t+1}^*) \\ &\quad + B_{F_{t+1}}(\mathbf{x}_t^*\|\mathbf{x}_{t+1}^*) \\ &\ge F_{t+1}(\mathbf{x}_{t+1}^*) + B_{F_{t+1}}(\mathbf{x}_t^*\|\mathbf{x}_{t+1}^*) \\ &= F_{t+1}(\mathbf{x}_{t+1}^*) + B_R(\mathbf{x}_t^*\|\mathbf{x}_{t+1}^*) \end{aligned}$$

In the last equation, we use the fact that the Bregman divergence is not influenced by the linear terms in F. Using again the fact that $\mathbf{x}_t^*$ is the minimizer of F_t, we further deduce

$$\begin{aligned} B_R(\mathbf{x}_t^*\|\mathbf{x}_{t+1}^*) &\le F_{t+1}(\mathbf{x}_t^*) - F_{t+1}(\mathbf{x}_{t+1}^*) \\ &= (F_t(\mathbf{x}_t^*) - F_t(\mathbf{x}_{t+1}^*)) + \eta\mathbf{g}_t^\top(\mathbf{x}_t^* - \mathbf{x}_{t+1}^*) \\ &\le \eta\mathbf{g}_t^\top(\mathbf{x}_t^* - \mathbf{x}_{t+1}^*). \end{aligned}$$

On the other hand, applying Taylor's theorem in several variables with the remainder given in Lagrange's form, we know that there exists $\boldsymbol{\xi}_t \in [\mathbf{x}_t^*, \mathbf{x}_{t+1}^*] \triangleq \{\lambda\mathbf{x}_t^* + (1-\lambda)\mathbf{x}_{t+1}^* : \lambda \in [0,1]\}$ such that

$$B_R(\mathbf{x}_t^*\|\mathbf{x}_{t+1}^*) = \frac{1}{2}(\mathbf{x}_t^* - \mathbf{x}_{t+1}^*)^\top \mathbf{H}(\boldsymbol{\xi}_t)(\mathbf{x}_t^* - \mathbf{x}_{t+1}^*),$$

where $\mathbf{H}(\boldsymbol{\xi}_t)$ denotes the Hessian matrix of R at point $\boldsymbol{\xi}_t$. Notice that the Hessian matrix of R is the identity matrix everywhere. Therefore $B_R(\mathbf{x}_t^*\|\mathbf{x}_{t+1}^*) = \frac{1}{2}\|\mathbf{x}_t^* - \mathbf{x}_{t+1}^*\|^2$. By Cauchy-Schwarz inequality, we obtain

$$\begin{aligned} \mathbf{g}_t^\top(\mathbf{x}_t^* - \mathbf{x}_{t+1}^*) &\le \|\mathbf{g}_t\|\cdot\|\mathbf{x}_t^* - \mathbf{x}_{t+1}^*\| \\ &= \|\mathbf{g}_t\|\cdot\sqrt{2B_R(\mathbf{x}_t^*\|\mathbf{x}_{t+1}^*)} \\ &\le \|\mathbf{g}_t\|\cdot\sqrt{2\eta\mathbf{g}_t^\top(\mathbf{x}_t^* - \mathbf{x}_{t+1}^*)} \end{aligned},$$

which immediately yields

$$\mathbf{g}_t^\top(\mathbf{x}_t^* - \mathbf{x}_{t+1}^*) \le 2\eta\|\mathbf{g}_t\|^2$$

□

Appendix B Proof of Lemma 4

Proof. We verify the inequality when $t = 1$ or 2. When $t \ge 3$, we have

$$(1+1/t)^{2/5} \ge 1 \ge \frac{8}{5}t^{-3/5}.$$

Since $2(1+1/t)^{2/5} \ge 2(1+1/t)^{1/5}$, we obtain

$$3(1+1/t)^{2/5} \ge 2(1+1/t)^{1/5} + \frac{8}{5}t^{-3/5}.$$

Therefore, we have

$$3(1+1/t)^{2/5} - 2(1+1/t)^{1/5} - \frac{8}{5}t^{-3/5} \ge 0. \tag{15}$$

Let $g(t) = t^{2/5}$. Since $g(t)$ is concave, we have $g(t+1) - g(t) \leq g'(t)$, which gives $(t+1)^{2/5} - t^{2/5} \leq \frac{2}{5}t^{-3/5}$. Combining the above inequality with (15), we see

$$3(1+1/t)^{2/5} - 2(1+1/t)^{1/5} + 4t^{2/5} - 4(t+1)^{2/5} \geq 0.$$

Multiplying both sides with $t^{2/5}$, we complete the proof. □

Appendix C Proof of Lemma 5

Proof. By the definition of $\mathbf{g}_{t+1}$, we have $\|\mathbf{g}_{t+1}\| \leq nM/\delta$. It suffices to show $\sqrt{2D^2\sigma_{t+1}} \geq n\eta M/(2\delta)$. By the definition of σ_{t+1}, η, and δ, it is equivalent to $4T^{3/5} - (t+1)^{1/5} \geq 0$. Since $1 \leq t \leq T$, we only need to show $4T^{3/5} - (T+1)^{1/5} \geq 0$. We define $f(T) = 4T^{3/5} - (T+1)^{1/5}$. Its derivative is $f'(T) = \frac{12(T+1)^{4/5} - T^{2/5}}{5T^{2/5}(T+1)^{4/5}}$. We have

$$\frac{12(T+1)^{4/5}}{T^{2/5}} = 12\left(T + \frac{1}{T} + 2\right)^{2/5} \geq 12 \cdot 4^{2/5} \geq 1$$

if $T \geq 1$. Therefore, we know that $f'(T) \geq 0$ if $T \geq 1$. Thus f is non-decreasing on $[1, \infty]$. This immediately yields $f(T) \geq f(1) \geq 0$, which completes the proof. □

Appendix D Proof of Theorem 2

Proof. The regret of Algorithm 2 by the end of the t-th iteration is at most

$$\begin{aligned}\sum_{m=0}^{\lceil \log_2(t+1) \rceil - 1} \beta (2^m)^{4/5} &= \beta \frac{\left(2^{\lceil \log_2(t+1) \rceil}\right)^{4/5} - 1}{2^{4/5} - 1} \\ &\leq \frac{\beta}{1 - 2^{-4/5}} (t+1)^{4/5}.\end{aligned}$$

□