Appendix A  Proof of $g_t^T (x_t^* - x_{t+1}^*) \leq 2\eta \| g_t \|^2$

Lemma 6 (Theorem 5.1 in [Hazar 2016]). Let $x_t^* = \arg \min_{x \in (1-\alpha)c} F_t(x)$. We have $g_t^T (x_t^* - x_{t+1}^*) \leq 2\eta \| g_t \|^2$.

Proof. We denote the regularizer in line 4 of Algorithm 1 by $R(x) \triangleq \| x - x_1 \|^2$ and define the Bregman divergence with respect to the function $F$ by

$$B_F(x\|y) = F(x) - F(y) - \nabla F(y)^T (x - y). \quad (14)$$

Since $x_{t+1}^*$ is a minimizer of $F_{t+1}$ and $F_{t+1}$ is convex, we have

$$F_{t+1}(x_t^*) = F_{t+1}(x_{t+1}^*) + (x_t^* - x_{t+1}^*)^T \nabla F_{t+1}(x_{t+1}^*) + B_{F_{t+1}}(x_t^*\|x_{t+1}^*) \geq F_{t+1}(x_{t+1}^*) + B_{F_{t+1}}(x_t^*\|x_{t+1}^*) = F_{t+1}(x_{t+1}^*) + B_R(x_t^*\|x_{t+1}^*)$$

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 $x_t^*$ is the minimizer of $F_t$, we further deduce

$$B_R(x_t^*\|x_{t+1}^*) \leq F_{t+1}(x_t^*) - F_{t+1}(x_{t+1}^*) = (F_t(x_t^*) - F_t(x_{t+1}^*)) + \eta g_t^T (x_t^* - x_{t+1}^*) \leq \eta \| g_t \| (x_t^* - x_{t+1}^*).$$

On the other hand, applying Taylor’s theorem in several variables with the remainder given in Lagrange’s form, we know that there exists $\xi_t \in [x_t^*, x_{t+1}^*] \triangleq \{ \lambda x_t^* + (1 - \lambda)x_{t+1}^* : \lambda \in [0, 1] \}$ such that

$$B_R(x_t^*\|x_{t+1}^*) = \frac{1}{2} (x_t^* - x_{t+1}^*)^T H(\xi_t)(x_t^* - x_{t+1}^*),$$

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

$$g_t^T (x_t^* - x_{t+1}^*) \leq \| g_t \| \| x_t^* - x_{t+1}^* \|$$

which immediately yields

$$g_t^T (x_t^* - x_{t+1}^*) \leq 2\eta \| g_t \|^2$$

Appendix B  Proof of Lemma 4

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

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

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

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

Therefore, we have

$$3(1 + 1/t)^{2/5} - 2(1 + 1/t)^{1/5} - \frac{8}{5} t^{-3/5} \geq 0. \quad (15)$$
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 \( g_{t+1} \), we have \( \|g_{t+1}\| \leq nM/\delta \). It suffices to show \( \sqrt{2D^2\sigma_{r+1}} \geq n\eta M/(2\delta) \). By the definition of \( \sigma_{t+1} \), \( \eta \), and \( \delta \), 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 \( f'(T) = \frac{12(T + 1)^{4/5} - T^{2/5}}{T^{2/5}(T + 1)^{1/5}} \). 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
\[
\sum_{m=0}^{[\log_2(t+1)]-1} \beta m^{4/5} = \beta \frac{2^{[\log_2(t+1)]} - 1}{2^{4/5} - 1} \leq \frac{\beta}{1 - 2^{-4/5}(t + 1)^{4/5}}.
\]

\[\square\]