Appendix

A   Proofs

Proof of Theorem 2. Fix $\epsilon > 0$ and let $\mathbf{x}^\star(\epsilon) \in \text{conv}(\mathcal{A})$ be such that

$$f(\mathbf{x}^\star(\epsilon)) \leq \inf_{\mathbf{x} \in \text{conv}(\mathcal{A})} f(\mathbf{x}) + \epsilon .$$

Using smoothness we get, for any α,

$$f(\mathbf{x}_t + \alpha(\mathbf{a}_t - \mathbf{x}_t)) \leq f(\mathbf{x}_t) + \alpha \langle \nabla f(\mathbf{x}_t), \mathbf{a}_t - \mathbf{x}_t \rangle + \frac{L}{r} \alpha^r \|\mathbf{a}_t - \mathbf{x}_t\|^r$$

$$\leq f(\mathbf{x}_t) + \alpha \langle \nabla f(\mathbf{x}_t), \mathbf{a}_t - \mathbf{x}_t \rangle + \frac{L}{r} \alpha^r (\|\mathbf{a}_t\| + \|\mathbf{x}_t\|)^r$$

$$\leq f(\mathbf{x}_t) - \alpha(\langle \nabla f(\mathbf{x}_t), \mathbf{a}_t \rangle + \langle \nabla f(\mathbf{x}_t), \mathbf{x}_t \rangle) + \frac{L}{r} (2\,R\,\alpha)^r . \tag{9}$$

The second inequality follows by triangle inequality. The last inequality follows because $\|\mathbf{a}_t\|, \|\mathbf{x}_t\| \leq R$. Now, by convexity of f,

$$\delta_t := f(\mathbf{x}_t) - f(\mathbf{x}^\star(\epsilon)) \leq \langle \nabla f(\mathbf{x}_t), \mathbf{x}_t - \mathbf{x}^\star(\epsilon) \rangle$$

$$\leq \langle \nabla f(\mathbf{x}_t), \mathbf{x}_t \rangle + \langle -\nabla f(\mathbf{x}_t), \mathbf{x}^\star(\epsilon) \rangle$$

$$\leq \langle \nabla f(\mathbf{x}_t), \mathbf{x}_t \rangle + \langle -\nabla f(\mathbf{x}_t), \mathbf{a}_t \rangle . \tag{10}$$

The last inequality holds because $\mathbf{a}_t$ maximized the linear functional $\langle -\nabla f(\mathbf{x}_t), \cdot \rangle$ over $\mathcal{A}$ and hence also over $\text{conv}(\mathcal{A})$. Thus, $\langle -\nabla f(\mathbf{x}_t), \mathbf{a}_t \rangle \geq \langle -\nabla f(\mathbf{x}_t), \mathbf{x}^\star(\epsilon) \rangle$ as $\mathbf{x}^\star(\epsilon) \in \text{conv}(\mathcal{A})$. Plugging (10) into (9), we have, for any $\alpha \geq 0$,

$$f(\mathbf{x}_t + \alpha(\mathbf{a}_t - \mathbf{x}_t)) \leq f(\mathbf{x}_t) - \alpha\delta_t + \frac{L}{r} (2\,R\,\alpha)^r .$$

Since $\mathbf{x}_{t+1}$ is chosen by minimizing the LHS over $\alpha \in [0, 1]$, we have

$$f(\mathbf{x}_{t+1}) \leq f(\mathbf{x}_t) + \min_{\alpha \in [0,1]} \left(-\alpha\delta_t + \frac{L}{r} (2\,R\,\alpha)^r \right) .$$

Thus, we have, for all $t \geq 0$,

$$\delta_{t+1} \leq \delta_t + \min_{\alpha \in [0,1]} \left(-\alpha\delta_t + \frac{L}{r} (2\,R\,\alpha)^r \right) .$$

For $t = 0$, choose $\alpha = 1$ on the RHS to get $\delta_1 \leq L(2\,R)^r/r$. Since δ_t's are monotonically non-increasing, this shows $\delta_t \leq L(2\,R)^r/r$ for all $t \geq 1$. Hence, for $t \geq 1$, we can choose α such that $\alpha^{r-1} = \delta_t/(L2^r R^r) \in [0, \frac{1}{r}]$ on the RHS to get

$$\forall t \geq 1, \ \ \delta_{t+1} \leq \delta_t - \left(1 - \frac{1}{r} \right) \frac{\delta_t^{\frac{r}{r-1}}}{L^{\frac{1}{r-1}} (2\,R)^{\frac{r}{r-1}}} .$$

Solving this recursion easily gives, for all $t \geq 1$,

$$f(\mathbf{x}_{t+1}) - f(\mathbf{x}^\star) \leq \frac{K_r L R^r}{t^{r-1}} ,$$

for some K_r that depends only on r. $\qquad\square$

Proof of Corollary 3. Since $h \in \text{conv}(\mathcal{A})$ and $\|\cdot\|_\#$ is equivalent to $\|\cdot\|$, we have

$$\inf_{g \in \text{conv}(\mathcal{A})} \|h - g\|_\#^p = 0 .$$

Thus, using p-uniform smoothness of $\|\cdot - h\|_\#^p$, Theorem 2 gives

$$\|g_{t+1} - h\|_\#^p = O\left(t^{-p+1} \right) .$$

The corollary now follows by again noting the equivalence of the two norms. $\qquad\square$