A One shot application of the Fractional Away-step Frank Wolfe

Running once Fractional Away-step Frank-Wolfe with a large value of γ allows to find an approximate minimizer with the desired precision. The following lemma explains the rate of convergence. Importantly the rate does not depend on r. Hence there is no hope of observing linear convergence for the strongly convex case.

Lemma A.1. *Let f be a smooth convex function, $\epsilon > 0$ be a target accuracy, and $x_0 \in \mathcal{C}$ be an initial point. Then for any $\gamma > \ln \frac{w(x_0)}{\epsilon}$, Algorithm 1 satisfies:*

$$f(x_T) - f(x^*) \leq \epsilon,$$

for $T \geq \frac{2C_f^{\mathcal{A}}}{\epsilon}$.

Proof. We can stop the algorithm as soon as the criterion $w(x_t) < \epsilon$ in step 2 is met or we observe an away step, whichever comes first. In former case we have $f(x_t) - f^* \leq w(t) < \epsilon$, in the latter it holds

$$f(x_t) - f^* \leq -\nabla f(x_t)(d_t^{FW}) \leq \epsilon/2 < \epsilon.$$

Thus, when the algorithms stops, we have achieved the target accuracy and it suffices to bound the number of iterations required to achieve that accuracy. Moreover, while running, the algorithm only executes Frank-Wolfe and we drop the FW superscript in the directions; otherwise we would have stopped.

From the proof of Proposition 4.1, we have each Frank-Wolfe step ensures progress of the form

$$f(x_t) - f(x_{t+1}) \geq \begin{cases} \frac{(r_t^T d_t)^2}{2C_f^{\mathcal{A}}} & \text{if } r_t^T d_t \leq C_f^{\mathcal{A}} \\ r_t^T d_t - C_f^{\mathcal{A}}/2 & \text{otherwise.} \end{cases}$$

For convenience, let $h_t \triangleq f(x_t) - f^*$. By convexity we have $h_t \leq \langle r_t; d_t \rangle$, so that the above becomes

$$f(x_t) - f(x_{t+1}) \geq \begin{cases} \frac{h_t^2}{2C_f^{\mathcal{A}}} & \text{if } h_t \leq C_f^{\mathcal{A}} \\ h_t - C_f^{\mathcal{A}}/2 & \text{otherwise.} \end{cases},$$

and moreover observe that the second case can only happen in the very first step: $h_1 \leq h_0 - (h_0 - C_f^{\mathcal{A}}/2) = C_f^{\mathcal{A}}/2 \leq 2C_f^{\mathcal{A}}/t$ for $t = 1$ providing the start of the following induction: we claim $h_t \leq \frac{2C_f^{\mathcal{A}}}{t}$.

Suppose we have established the bound for t, then for $t+1$, we have

$$h_{t+1} \leq \left(1 - \frac{h_t}{2C_f^{\mathcal{A}}}\right) h_t \leq \frac{2C_f^{\mathcal{A}}}{t} - \frac{2C_f^{\mathcal{A}}}{t^2} \leq \frac{2C_f^{\mathcal{A}}}{t+1}.$$

Therefore the induction is complete and it follows that the algorithm requires $T \geq \frac{2C_f^{\mathcal{A}}}{\epsilon}$ to reach ϵ-accuracy. ■