A Offline Proofs

A.1 Proof of the Scaling Inequalities

Proof of Lemma 2.4. By definition of local uniform convexity between x^* and x, we have that for any $z \in \mathbb{R}^d$ of unit norm $(x^* + x)/2 + \alpha/4\|x^* - x\|^q z \in \mathcal{C}$. Then, by optimality of x^*, *i.e.* $x^* \in \text{argmax}_{v\in\mathcal{C}}\langle -\nabla f(x^*), v\rangle$, we have $\langle -\nabla f(x^*), x^*\rangle \geq \langle -\nabla f(x^*), (x^* + x)/2\rangle + \alpha/4\|x^* - x\|^q z\rangle$. Choosing the best z implies

$$\langle -\nabla f(x^*), x^* - x\rangle \geq \alpha/2\|x^* - x\|^q \|\nabla f(x^*)\|_*.$$

■

Proof of Corollary 2.7. Here, because $-\nabla f(x^*) \in N_\Gamma(x^*)$, we have that $x^* \in \text{argmax}_{v\in\Gamma}\langle -\nabla f(x^*), v\rangle$. Also, for $x \in \mathcal{C} \subset \Gamma$, by (α, q)-uniform convexity of Γ, we also have that for any $z \in \mathbb{R}^d$ of unit norm that $(x^* + x)/2 + \alpha/4\|x^* - x\|^q z \in \Gamma$. Then, by optimality of x^*, we have $\langle -\nabla f(x^*), x^*\rangle \geq \langle -\nabla f(x^*), (x^* + x)/2\rangle + \alpha/4\|x^* - x\|^q z\rangle$. Choosing the best z implies (for any $x \in \mathcal{C}$) $\langle -\nabla f(x^*), x^* - x\rangle \geq \alpha/2\|x^* - x\|^q \|\nabla f(x^*)\|_*$. ■

A.2 Recursive Lemma

The proofs of Theorems 2.2, 2.5, and 2.9 involve finding explicit bounds for sequences (h_t) satisfying recursive inequalities of the form,

$$h_{t+1} \leq h_t \cdot \max\{1/2, 1 - C h_t^\eta\}. \tag{5}$$

with $\eta < 1$. An explicit solution with $\eta = 1/2$ is given in [Garber and Hazan, 2015] and corresponds to $h_t = \mathcal{O}(1/T^2)$, while for $\eta = 1$ we recover the classical sublinear Frank-Wolfe regime of $\mathcal{O}(1/T)$. For a $\eta \in]0, 1]$, we have $\mathcal{O}(1/T^{1/\eta})$ (see for instance [Temlyakov, 2011] or [Nguyen and Petrova, 2017, Lemma 4.2.]), which can be guessed via $h(t) = (C\eta)^{1/\eta} t^{-1/\eta}$ the solution of the differential equation $h'(t) = -Ch(t)^{\eta+1}$ for $t > 0$. A quantitative statement is, for instance, given in [Xu and Yang, 2018, proof of Theorem 1.] that we reproduce here.

Lemma A.1 (Recurrence and sub-linear rates)**.** *Consider a sequence $(h_t)_{t\in\mathbb{N}}$ of non-negative numbers satisfying* (5) *with $0 < \eta \leq 1$, then $h_T = \mathcal{O}\big(1/T^{1/\eta}\big)$. More precisely for all $t \geq 0$,*

$$h_t \leq \frac{M}{(t + k)^{1/\eta}}$$

with $k \triangleq (2 - 2^\eta)/(2^\eta - 1)$ and $M \triangleq max\{h_0 k^{1/\eta}, 2/((\eta - (1 - \eta)(2^\eta - 1))C)^{1/\eta}\}$.

A.3 Convergence Rates with Local Scaling Inequality

The local scaling inequality expresses a property between x^* and any $x \in \mathcal{C}$. In Lemma A.2, we show that albeit we only have access to a local scaling inequality, it is still possible to control the variation of the distance of the iterate to its Frank-Wolfe vertex $\|x_t - v_t\|$ in terms of a power of the primal gap, see beginning of Section 2 for a qualitative explanation.

Lemma A.2. *Consider f a L-smooth convex function and a compact convex set $\mathcal{C}$. Assume $inf_{x\in\mathcal{C}}\|\nabla f(x)\|_* > c > 0$ and write $x^* \in \partial\mathcal{C}$ the solution of* (OPT)*. Assume that $\mathcal{C}$ satisfies a local scaling inequality at x^* for problem* (OPT) *with $\alpha > 0$ and $q \geq 2$, i.e. for all $x \in \mathcal{C}$*

$$\langle -\nabla f(x^*), x^* - x\rangle \geq \alpha/2\|\nabla f(x^*)\|_* \cdot \|x^* - x\|^q. \tag{6}$$

Write $v_t \triangleq argmax_{v\in\mathcal{C}}\langle -\nabla f(x_t), v\rangle$ the Frank-Wolfe vertex. Assume that $h_t = f(x_t) - f(x^*) \leq 1$ (a simple burn-in phase). Then, we have

$$\|x_t - v_t\| \leq H h_t^{1/(q(q-1))}, \tag{7}$$

with $H \triangleq 2 \cdot max\Big\{\Big(\frac{2L}{c\alpha}\Big)^{1/(q-1)}\Big(\frac{2}{c\alpha}\Big)^{1/(q(q-1))}, \Big(\frac{2}{c\alpha}\Big)^{1/q}\Big\}$.

Proof. We apply the local scaling inequality (6) with $x = v_t$ and $x = x_t$. We obtain two important inequalities: one (7) that upper bounds $\|x_t - v_t\|$ in terms of $f(x_t) - f(x^*)$ and another (8) that upper bounds $\|v_t - x^*\|$ in terms of $\|x^* - x_t\|$, where v_t is the Frank-Wolfe vertex related to iterate x_t. These two inequalities rely of convexity, L-smoothness and (6), but do not rely on strong convexity of the function f.

By optimality of the Frank-Wolfe vertex v_t, we have $\nabla f(x_t)^T v_t \leq \nabla f(x_t)^T x^*$. Hence, combining that with the local scaling inequality evaluated at v_t andCauchy-Schwartz, we get

$$\begin{aligned}\|\nabla f(x^*) - \nabla f(x_t)\| \, \|v_t - x^*\| &\geq \langle \nabla f(x^*) - \nabla f(x_t), \, v_t - x^* \rangle + \underbrace{\langle \nabla f(x_t), \, v_t - x^* \rangle}_{\leq 0} \\ &\geq \langle \nabla f(x^*), \, v_t - x^* \rangle \geq c\alpha/2 \|v_t - x^*\|^q.\end{aligned}$$

Then, L-smoothness applied to the left hand side leaves us with

$$\|x_t - x^*\| \geq \frac{c\alpha}{2L} \|v_t - x^*\|^{q-1}. \tag{8}$$

Then, the triangular inequality gives

$$\begin{aligned}\|x_t - v_t\| &\leq \|v_t - x^*\| + \|x^* - x_t\| \\ \|x_t - v_t\| &\leq \Big(\frac{2L}{c\alpha}\Big)^{1/(q-1)} \|x_t - x^*\|^{1/(q-1)} + \|x^* - x_t\|.\end{aligned}$$

Finally applying (6) with $x = x_t$ with $\inf_{x\in\mathcal{C}} \|\nabla f(x)\|_* > c > 0$ and using the convexity of f (*i.e.*, $f(x_t) \geq f(x^*) + \langle \nabla f(x^*); x_t - x^* \rangle$), we have $\|x_t - x^*\| \leq \big(\frac{2}{c\alpha}\big)^{1/q} h_t^{1/q}$ which leads to

$$\|x_t - v_t\| \leq \Big(\frac{2L}{c\alpha}\Big)^{1/(q-1)} \Big(\frac{2}{c\alpha}\Big)^{1/(q(q-1))} h_t^{1/(q(q-1))} + \Big(\frac{2}{c\alpha}\Big)^{1/q} h_t^{1/q} \ .$$

We can simplify this previous expression, and we assumed without loss of generality (*i.e.* up to a burning-phase) that $h_t \leq 1$, which implies for $q \geq 2$ that $h_t^{1/(q(q-1))} \geq h_t^{1/q}$. With $H \triangleq 2 \cdot \max\Big\{ \big(\frac{2L}{c\alpha}\big)^{1/(q-1)} \big(\frac{2}{c\alpha}\big)^{1/(q(q-1))}, \big(\frac{2}{c\alpha}\big)^{1/q} \Big\}$, we then have

$$\|x_t - v_t\| \leq H h_t^{1/(q(q-1))} \ .$$

■

We now give the proof of Theorem 2.5. Recall that this theorem gives the convergence rates of the Frank-Wolfe algorithm when the set $\mathcal{C}$ satisfies a local scaling inequality (that is, for instance, with Lemma 2.4 a consequence of local uniform convexity of the constraint set $\mathcal{C}$). It does not require uniform convexity of the function f, but assumes that f is a convex L-smooth function with $\inf_{x\in\mathcal{C}} \|\nabla f(x)\|_* > 0$.

Theorem 2.5′. *Consider f a L-smooth convex function and a compact convex set $\mathcal{C}$. Assume $\|\nabla f(x)\|_* > c > 0$ for all $x \in \mathcal{C}$ and write $x^* \in \partial\mathcal{C}$ a solution of* (OPT). *Further, assume that $\mathcal{C}$ satisfies a local scaling inequality at x^* with the (α, q) parameters. Then the iterates of the Frank-Wolfe algorithm, with short step satisfy*

$$\begin{cases} f(x_T) - f(x^*) \leq M/(T+k)^{\frac{1}{1-2/(q(q-1))}} & \text{when } q > 2 \\ f(x_T) - f(x^*) \leq \big(1-\rho\big)^T h_0 & \text{when } q = 2, \end{cases} \tag{9}$$

with $\rho = max\big\{\frac{1}{2}, 1 - c\alpha/L\big\}$, $k \triangleq (2 - 2^\eta)/(2^\eta - 1)$ and $M \triangleq max\{h_0 k^{1/\eta}, 2/((\eta - (1-\eta)(2^\eta - 1))C)^{1/\eta}\}$, where $\eta \triangleq 1 - 2/(q(q-1))$ and $C \triangleq 1/(2LH^2)$. Note that H depends only on C, α, L and q (see Lemma A.2).

Proof of Theorem 2.5. With Lemma A.2, which satisfies the assumption of Theorem 2.5, we have

$$\|x_t - v_t\| \leq H h_t^{1/(q(q-1))},$$

with $H \triangleq 2 \cdot \max\Big\{\Big(\frac{2L}{c\alpha}\Big)^{1/(q-1)}\Big(\frac{2}{c\alpha}\Big)^{1/(q(q-1))}, \Big(\frac{2}{c\alpha}\Big)^{1/q}\Big\}$. We plug this last expression in the classical descent guarantee given by L-smoothness

$$
\begin{aligned}
h_{t+1} &\leq (1-\gamma)h_t + \gamma^2 \frac{L}{2}\|v_t - x_t\|^2 \\
h_{t+1} &\leq (1-\gamma)h_t + \gamma^2 \frac{L}{2} H^2 h_t^{2/(q(q-1))} .
\end{aligned}
$$

The optimal decrease $\gamma \in [0,1]$ is $\gamma^* = \min\Big\{\frac{h_t^{1-2/(q(q-1))}}{LH^2}, 1\Big\}$. When $\gamma^* = 1$, or equivalently $h_t \geq \big(LH^2\big)^{2/(q(q-1))}$, we have $h_{t+1} \leq h_t/2$. In other words, for the very first iterations, there is a brief linear convergence regime. Otherwise, when $\gamma^* \leq 1$, we have

$$
h_{t+1} \leq h_t\Big(1 - \frac{1}{2LH^2} h_t^{1-2/(q(q-1))}\Big) . \tag{10}
$$

When $q = 2$, this corresponds to the strongly convex case and we recover the classical linear-convergence regime. We conclude using Lemma A.1 that the rate is $\mathcal{O}\Big(1/T^{1/(1-2/(q(q-1)))}\Big)$. ■

A.4 Proof without restriction on the location of the optimum.

We regroup here the proofs of the lemma and the theorem contained in Section 2.4.

Proof of Lemma 2.8. By Lemma 2.1 we have $g(x_t) = \langle -\nabla f(x_t); v_t - x_t\rangle \geq \alpha/2\|x_t - v_t\|^q \|\nabla f(x_t)\|_*$. Then, by combining the convexity of f, Cauchy-Schwartz and (μ,θ)-Hölderian Error Bound, we have

$$
f(x_t) - f(x^*) \leq \langle \nabla f(x_t), x - x^*\rangle \leq \|\nabla f(x_t)\|_* \cdot \|x_t - x^*\| \leq \mu \|\nabla f(x_t)\|_* \cdot \big(f(x_t) - f(x^*)\big)^\theta,
$$

so that $\big(f(x) - f(x^*)\big)^{1-\theta} \leq \mu\|\nabla f(x)\|_*$ and finally $g(x_t) \geq \alpha/(2\mu)\|x_t - v_t\|^q h_t^{1-\theta}$. ■

Proof of Theorem 2.9. From the proof of Theorem 2.2, L-smoothness and the step size decision we have

$$
h(x_{t+1}) \leq h(x_t) - \frac{g(x_t)}{2} \cdot \min\Big\{1; \frac{g(x_t)}{L\|x_t - v_t\|^2}\Big\}.
$$

From Lemma 2.8, we have $g(x_t) \geq \alpha/(2\mu)\|x_t - v_t\|^q h_t^{1-\theta}$. Hence, we can rewrite

$$
\frac{g(x_t)}{\|x_t - v_t\|^2} = \Big(\frac{g(x_t)^{q/2-1} g(x_t)}{\|x_t - v_t\|^q}\Big)^{2/q} \geq \Big(\alpha/(2\mu)\Big)^{2/q} g(x_t)^{1-2/q} h_t^{(1-\theta)2/q}.
$$

And because $g(x_t) \geq h_t$, we have

$$
\frac{g(x_t)}{\|x_t - v_t\|^2} \geq \Big(\alpha/(2\mu)\Big)^{2/q} h_t^{1-2\theta/q}.
$$

We finally end up with the following recursion

$$
h(x_{t+1}) \leq h(x_t) \cdot \max\Big\{\frac{1}{2}; 1 - \Big(\alpha/(2\mu)\Big)^{2/q} h_t^{1-2\theta/q}/L\Big\},
$$

and we conclude with Lemma A.1. ■

B Proofs in Online Optimization

The following is the generalization of [Huang et al., 2017, (6)] when the set is uniformly convex (see Definition 1.1). Note that in our version $\mathcal{C}$ can be uniformly convex with respect to any norm.

Proposition B.1. *Assume $\mathcal{C} \subset \mathbb{R}^d$ is a (α, q)-uniformly convex set with respect to $\|\cdot\|$, with $\alpha > 0$ and $q \geq 2$. Consider the non-zero vectors $\phi_1, \phi_2 \in \mathbb{R}^d$ and $v_{\phi_1} \in \operatorname{argmax}_{v\in\mathcal{C}} \langle \phi_1, v\rangle$ and $v_{\phi_2} \in \operatorname{argmax}_{v\in\mathcal{C}} \langle \phi_2, v\rangle$. Then*

$$\langle v_{\phi_1} - v_{\phi_2}, \phi_1\rangle \leq \Big(\frac{1}{\alpha}\Big)^{1/(q-1)} \frac{\|\phi_1 - \phi_2\|_*^{1+1/(q-1)}}{(max\{\|\phi_1\|_*, \|\phi_2\|_*\})^{1/(q-1)}}, \tag{11}$$

where $\|\cdot\|_*$ is the dual norm to $\|\cdot\|$.

Proof. By definition of uniform convexity, for any z of unit norm, $v_\gamma(z) \in \mathcal{C}$ where

$$v_\gamma(z) \triangleq \gamma v_{\phi_1} + (1-\gamma)v_{\phi_2} + \gamma(1-\gamma)\alpha\|v_{\phi_1} - v_{\phi_2}\|^q z.$$

By optimality of v_{ϕ_1} and v_{ϕ_2}, we have $\langle v_\gamma(z), \phi_1\rangle \leq \langle v_1, \phi_1\rangle$ and $\langle v_\gamma(z), \phi_2\rangle \leq \langle v_2, \phi_2\rangle$, so that

$$\langle v_\gamma(z), \gamma\phi_1 + (1-\gamma)\phi_2\rangle \leq \gamma\langle v_1, \phi_1\rangle + (1-\gamma)\langle v_2, \phi_2\rangle.$$

Write $\phi_\gamma = \gamma\phi_1 + (1-\gamma)\phi_2$. Then, when developing the left hand side, we get

$$\gamma(1-\gamma)\alpha\|v_{\phi_1} - v_{\phi_2}\|^q\langle z, \phi_\gamma\rangle \leq \gamma(1-\gamma)\langle v_{\phi_1} - v_{\phi_2}, \phi_1 - \phi_2\rangle$$

Choosing the best z of unit norm we get

$$\alpha\|v_{\phi_1} - v_{\phi_2}\|^q\|\phi_\gamma\|_* \leq \langle v_{\phi_1} - v_{\phi_2}, \phi_1 - \phi_2\rangle$$

and for $\gamma = 0$ and $\gamma = 1$ and via generalized Cauchy-Schwartz we get

$$\alpha\|v_{\phi_1} - v_{\phi_2}\|^q \cdot \max\{\|\phi_1\|_*, \|\phi_2\|_*\} \leq \|v_{\phi_1} - v_{\phi_2}\| \cdot \|\phi_1 - \phi_2\|_*,$$

Hence

$$\|v_{\phi_1} - v_{\phi_2}\| \leq \Big(\frac{1}{\alpha}\Big)^{1/(q-1)} \frac{\|\phi_1 - \phi_2\|_*^{1/(q-1)}}{(\max\{\|\phi_1\|_*, \|\phi_2\|_*\})^{1/(q-1)}}.$$

Then, since $\langle v_{\phi_2} - v_{\phi_1}; \phi_2\rangle \geq 0$ by optimality of v_{ϕ_2}, we have

$$\langle v_{\phi_1} - v_{\phi_2}, \phi_1\rangle \leq \|v_{\phi_1} - v_{\phi_2}\| \cdot \|\phi_1 - \phi_2\|_* \leq \Big(\frac{1}{\alpha}\Big)^{1/(q-1)} \frac{\|\phi_1 - \phi_2\|_*^{1+1/(q-1)}}{(\max\{\|\phi_1\|_*, \|\phi_2\|_*\})^{1/(q-1)}},$$

and we finally obtain (11). ■

We now provide a proof of Theorem 3.1.

Theorem 3.1′. *Let $\mathcal{C}$ be a compact and (α, q)-uniformly convex set with respect to $\|\cdot\|$. Assume that $L_T = min_{1\leq t\leq T}\|\frac{1}{t}\sum_{\tau=1}^t c_\tau\|_* > 0$. Then the regret R_T of FTL (4) for online linear optimization satisfies*

$$\begin{cases} R_T \leq 2M\Big(\dfrac{2M}{\alpha L_T}\Big)^{1/(q-1)}\Big(\dfrac{q-1}{q-2}\Big)T^{1-1/(q-1)} & \text{when } q > 2 \\ R_T \leq \dfrac{4M^2}{\alpha L_T}(1 + \log(T)) & \text{when } q = 2, \end{cases} \tag{12}$$

where $M = sup_{c\in\mathcal{W}}\|c\|_*$, with the losses $l_t(x) = \langle c_t, x\rangle$ and (c_t) belong to the bounded set $\mathcal{W}$.

Proof of Theorem 3.1. The proof follows exactly that of [Huang et al., 2017, Theorem 5]. Write $M = \sup_{c\in\mathcal{W}}\|c\|_*$, $F_t(x) = \frac{1}{t}\sum_{\tau=1}^t \langle c_t, x\rangle$ and short cut $\nabla F_t \triangleq \frac{1}{t}\sum_{\tau=1}^t c_t$ the gradient of the linear function $F_t(x)$. Recall that with FTL, x_t is defined as

$$x_t \in \operatorname{argmin}_{x\in\mathcal{C}}\langle\sum_{\tau=1}^{t-1} c_t, x\rangle.$$

As in [Huang et al., 2017, Theorem 5] we have

$$\|\nabla F_t - \nabla F_{t-1}\|_* \leq \frac{2M}{t}.$$

Using [Huang et al., 2017, Proposition 2] and Proposition B.1 we get the following upper bound on the regret

$$R_T = \sum_{t=1}^{T} t\langle x_{t+1} - x_t, \nabla F_t\rangle \leq \Big(\frac{1}{\alpha}\Big)^{1/(q-1)} \sum_{t=1}^{T} t \frac{\|\nabla F_t - \nabla F_{t-1}\|_*^{1+1/(q-1)}}{(\max\{\|\nabla F_t\|_*, \|\nabla F_{t-1}\|_*\})^{1/(q-1)}}.$$

Hence, with $L_T = \min_{1\leq t\leq T}\|\nabla F_t\|_* > 0$ and $\|\nabla F_t - \nabla F_{t-1}\|_* \leq (2M)/t$, we have

$$R_T \leq 2M\Big(\frac{2M}{\alpha L_T}\Big)^{1/(q-1)} \sum_{t=1}^{T} t^{-1/(q-1)}.$$

Then we have for $q > 2$

$$\sum_{t=1}^{T} t^{-1/(q-1)} = 1 + \sum_{t=2}^{T} t^{-1/(q-1)} \leq 1 + \int_{x=1}^{T-1} x^{-1/(q-1)} dx = 1 + \Big[\frac{t^{1-1/(q-1)}}{1 - 1/(q-1)}\Big]_1^{T-1},$$

so that finally

$$R_T \leq 2M\Big(\frac{2M}{\alpha L_T}\Big)^{1/(q-1)}\Big(\frac{q-1}{q-2}\Big)T^{1-1/(q-1)}.$$

■

C Uniform Convexity

C.1 Uniformly Convex Spaces

Proof of Lemma 4.2. The argument is similar to that in [Molinaro, 2020, Appendix A], we repeat it for completeness. Assume $(\mathbb{X}, \|\cdot\|)$ is uniformly convex with modulus of convexity $\delta(\cdot)$. Let us write $\mathcal{C} \triangleq B_{\|\cdot\|}(1)$. Then for any $(x, y, z) \in \mathcal{C}$, we have by definition $1 - \frac{\|x+y\|}{2} \geq \delta(\|x - y\|)$ and then

$$\Big\|\frac{x+y}{2} + \delta(\|x-y\|)z\Big\| \leq \Big\|\frac{x+y}{2}\Big\| + \delta(\|x-y\|) \leq 1 \ .$$

Hence, $\frac{x+y}{2} + \delta(\|x - y\|)z \in \mathcal{C}$. Without loss of generality, consider $\eta \in]0; 1/2]$. We need to show that $\eta x + (1-\eta)y + \delta(\|x-y\|)z \in \mathcal{C}$ for any z with norm lesser than 1. First, note that $\eta x + (1-\eta)y = (1-2\eta)y + (2\eta)(x+y)/2$. Note also that because $1 - 2\eta \in [0, 1]$, we have for any z of norm lesser than 1

$$(1 - 2\eta)x + (2\eta)\big[(x + y)/2 + \delta(\|x - y\|)z\big] \in \mathcal{C}.$$

Hence, for any z of norm lesser than 1, we have

$$\eta x + (1 - \eta)y + 2\eta\delta(\|x - y\|)z \in \mathcal{C}.$$

In particular, choosing $z' = (1 - \eta)z$ gives that for any z of norm less than 1

$$\eta x + (1 - \eta)y + 2\eta(1 - \eta)\delta(\|x - y\|)z \in \mathcal{C},$$

which proves that $\mathcal{C}$ is $\delta(\cdot)$-uniformly convex with respect to $\|\cdot\|$.

Let us now assume that $B_{\|\cdot\|}(1)$ is (α, q)-uniformly convex w.r.t. $\|\|$. Let us show that $B_{\|\cdot\|}(r)$ is $(\alpha/r^{q-1}, q)$-uniformly convex. Consider $(x, y) \in B_{\|\cdot\|}(r)$ and z with $\|z\| \leq 1$, we have (with $x' = x/r$ and $y' = y/r$)

$$\begin{aligned} I &\triangleq \big\|\eta x + (1 - \eta)y + \eta(1 - \eta)\alpha r^{-(q-1)}\|x - y\|^q z\big\| \\ &= r\big\|\eta x' + (1 - \eta)y' + \eta(1 - \eta)\alpha\|x' - y'\|^q z\big\| \leq r. \end{aligned}$$

■

C.2 Uniformly Convex Functions

Uniform convexity is also a property of convex functions and defined as follows.

Definition C.1. *A differentiable function f is (μ, r)-uniformly convex on a convex set $\mathcal{C}$ w.r.t. $\|\cdot\|$ if there exists $r \geq 2$ and $\mu > 0$ such that for all $(x, y) \in \mathcal{C}$*

$$f(y) \geq f(x) + \langle \nabla f(x), y - x \rangle + \frac{\mu}{2}\|x - y\|^r .$$

We now state the equivalent of [Journée et al., 2010, Theorem 12] for the level sets of uniformly convex functions. This was already used in [Garber and Hazan, 2015] in the case of strongly-convex sets.

Lemma C.2. *Let $f : \mathbb{R}^d \to \mathbb{R}^+$ be a non-negative, L-smooth and (μ, r)-uniformly convex function on $\mathbb{R}^d$ w.r.t. $\|\cdot\|_2$, with $r \geq 2$. Then for any $w > 0$, the level set*

$$\mathcal{L}_w = \Big\{x \mid f(x) \leq w\Big\},$$

is (α, r)-uniformly convex w.r.t. with $\alpha = \frac{\mu}{2^{r-1}\sqrt{2wL}}$.

Proof. The proof follows exactly that of [Journée et al., 2010, Theorem 12], replacing $\|x-y\|^2$ with $\|x-y\|^r$. We state it for the sake of completeness. Consider $w_0 > 0$, $(x, y) \in \mathcal{L}_w$ and $\gamma \in [0, 1]$. We denote $z = \gamma x + (1 - \gamma)y$. For $u \in \mathbb{R}^d$, by L-smoothness applied at z and at x^* (the unconstrained optimum of f), we have

$$\begin{aligned} f(z+u) &\leq f(z) + \langle \nabla f(z), u \rangle + \frac{L}{2}\|u\|_2^2 \\ &\leq f(z) + \|\nabla f(z)\|_2 \cdot \|u\|_2 + \frac{L}{2}\|u\|_2^2 \\ &\leq f(z) + \sqrt{2Lf(z)}\|u\|_2 + \frac{L}{2}\|u\|_2^2 = \Big(\sqrt{f(z)} + \sqrt{\frac{L}{2}}\|u\|_2\Big)^2 . \end{aligned}$$

Note that (μ, r)-uniform convexity of f w.r.t. $\|\cdot\|_2$ implies that (see, *e.g.*, [Kerdreux et al., 2021, Definition 3.2.])

$$f(z) \leq \gamma f(x) + (1-\gamma)f(y) - \frac{\mu}{2^{r-1}}\gamma(1-\gamma)\|x-y\|^r.$$

In particular then, because $x, y \in \mathcal{L}_w$, we have $f(z) \leq w - \frac{\mu}{2^{r-1}}\gamma(1-\gamma)\|x-y\|^r$ so that

$$f(z+u) \leq \Big(\sqrt{w - \frac{\mu}{2^{r-1}}\gamma(1-\gamma)\|x-y\|^r} + \sqrt{\frac{L}{2}}\|u\|\Big)^2. \tag{13}$$

Then, with the concavity of the square-root, we get

$$f(z+u) \leq \Big(\sqrt{w} - \frac{\mu}{2^r\sqrt{w}}\gamma(1-\gamma)\|x-y\|^r + \sqrt{\frac{L}{2}}\|u\|\Big)^2. \tag{14}$$

Hence for any u such that

$$\|u\| = \frac{\mu}{2^{r-1}\sqrt{2wL}}\gamma(1-\gamma)\|x-y\|^r,$$

we have $z + u \in \mathcal{L}_w$. Hence $\mathcal{L}_w$ is a $(\frac{\mu}{2^{r-1}\sqrt{2wL}}, r)$-uniformly convex set. ■

Lemma C.2 restrictively requires smoothness of the uniformly convex function f. Hence we provide the analogous of [Garber and Hazan, 2015, Lemma 3].

Lemma C.3. *Consider a finite dimensional normed vector space $(\mathbb{X}, \|\cdot\|)$. Assume $f(x) = \|x\|^2$ is (μ, s)-uniformly convex function (with $r \geq 2$) with respect to $\|\cdot\|$. Then the norm balls $B_{\|\cdot\|}(r) = \Big\{x \in \mathbb{X} \mid \|x\| \leq r\Big\}$ are $(\frac{\mu}{2^{r-1}r}, s)$-uniformly convex.*

Proof. The proof follows exactly that of [Garber and Hazan, 2015, Lemma 3] which itself follows that of [Journée et al., 2010, Theorem 12], where operations involving L-smoothness are replaced by an application of the triangular inequality.

Let's consider $s \geq 2$, $(x, y) \in B_{\|\cdot\|}(r)$ and $\gamma \in [0, 1]$. We denote $z = \gamma x + (1-\gamma)y$. For $u \in \mathbb{X}$, applying successively the triangular inequality and with the (μ, s)-uniform convexity of $f(x) = \|x\|^2$, we get

$$\begin{aligned} f(z+u) = \|z+u\|^2 &\leq \Big(\|z\| + \|u\|\Big)^2 = \Big(\sqrt{f(z)} + \|u\|\Big)^2 \\ &\leq \Big(\sqrt{\gamma f(x) + (1-\gamma)f(y) - \frac{\mu}{2^{r-1}}\gamma(1-\gamma)\|x-y\|^s} + \|u\|\Big)^2 \\ &\leq \Big(\sqrt{r^2 - \frac{\mu}{2^{r-1}}\gamma(1-\gamma)\|x-y\|^s} + \|u\|\Big)^2. \end{aligned}$$

We then use concavity of the square root as before to get

$$\|z+u\|^2 \leq \Big(r - \frac{\mu}{2^{r-1}r}\gamma(1-\gamma)\|x-y\|^s + \|u\|\Big)^2.$$

In particular, for $u \in \mathbb{X}$ such that $\|u\| = \frac{\mu}{2^{r-1}r}\gamma(1-\gamma)\|x-y\|^s$, we have $z + u \in B_{\|\cdot\|}(r)$. Hence $B_{\|\cdot\|}(r)$ is $(\frac{\mu}{2^{r-1}r}, s)$- uniformly convex with respect to $\|\cdot\|$. ∎

These previous lemmas hence allow to translate functional uniformly convex results into results for classic balls norms. For instance, [Shalev-Shwartz, 2007, Lemma 17] showed that for $p \in]1, 2]$ $f(x) = 1/2\|x\|_p^2$ was $(p-1)$-uniformly convex with respect to $\|\cdot\|_p$.