# Written Examination Online Convex Optimization
# 7 January, 2021
# Lecturer: Olivier Wintenberger

- This is a closed book examination. You may not use any lecture note or other support.
- This examination paper consists of 2 pages, duration 1h30.
- Solutions can be written in English or French.
- The distribution of the points over the problem will be approximatively as follows.

| Question | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | sum |
|---|---|---|---|---|---|---|---|---|---|---|
| # points | 1 | 2 | 1 | 3 | 1 | 2 | 3 | 4 | 3 | 20 |

**Problem:** The aim of the problem is to get a regret bound for the agile version of the OMD algorithm and to apply it to get a $\log d$ regret bound over any $\ell^1$-ball. We recall that $R$ is any regularization function (strongly convex and twice differentiable), $(f_t)$ is a sequence of convex loss function and the regret is

$$Regret_T = \sum_{t=1}^{T} f_t(x_t) - \min_{x\in\mathcal{K}} \sum_{t=1}^{T} f_t(x) = \sum_{t=1}^{T} f_t(x_t) - \sum_{t=1}^{T} f_t(x^*)$$

where $\mathcal{K}$ is a bounded close set. We recall also the notions of convex conjugate and Bergman divergence

$$R^*(y^*) = \max_{y\in\mathcal{K}}\{y^T y^* - R(y)\} \text{ and } B_R(y||x) = R(y) - R(x) - \nabla R(x)^T(y-x)\,,$$

where $y^*$ is any point in $\mathcal{K}^* = \{\nabla R(x), x \in \mathcal{K}\}$. We want to show that

**Theorem 0.1** *The agile OMD Algorithm 1 satisfies the regret bound*

$$Regret_T \le \frac{\eta}{2}\sum_{t=1}^{T} \|\nabla f_t(x_t)\|_t^{*2} + \frac{R(x^*) - R(x_1)}{\eta}\,,$$

*where* $\|\cdot\|_t^{*2} = \|\cdot\|^2_{\nabla^2 R^*(z_t^*)}$ *for* $R^*$ *the convex conjugate of* $R$ *and* $z_t^*$ *some point in* $\mathcal{K}^*$.

**Algorithm 1:** Online Mirror Descent (agile version)

**Parameters:** Regularization function $R$, step-size $\eta > 0$.
**Initialization:** Initial prediction $x_1 = \arg\min_{x\in\mathcal{K}} B_R(x||y_1)$ with $y_1 \in \mathbf{R}^d$ such that $\nabla R(y_1) = 0$.
For each recursion $t \geq 1$:
**Predict:** $x_t$
**Incur:** $f_t(x_t)$
**Observe:** $\nabla f_t(x_t)$
**Recursion:** Update
$$\nabla R(y_{t+1}) = \nabla R(x_t) - \eta \nabla f_t(x_t)\,,$$
$$x_{t+1} = \arg\min_{x\in\mathcal{K}} B_R(x||y_{t+1})\,.$$

1. Explain what is the difference with the lazy version of OMD?

2. Apply the gradient trick on the regret in order to get a bound depending on $\nabla f_t(x_t)$, $x_t$, $1 \leq t \leq T$ and $x^*$ only.

3. Show that
$$Regret_T \leq \frac{1}{\eta} \sum_{t=1}^{T} (\nabla R(x_t) - \nabla R(y_{t+1}))^T (x_t - x^*)\,.$$

4. Express each summand in terms of $B_R(x^*||x_t)$, $B_R(x^*||y_{t+1})$ and $B_R(x_t||y_{t+1})$.

5. Using that $B_R(x^*, y_{t+1}) \geq B_R(x^*, x_{t+1})$ and a telescoping sum argument, derive that
$$Regret_T \leq \frac{1}{\eta}(B_R(x^*||x_1) - B_R(x^*||x_{T+1})) + \frac{1}{\eta}\sum_{t=1}^{T} B_R(x_t||y_{t+1})\,.$$

6. Check that by definition of $x_1$ we have $R(x_1) = \min_{x\in\mathcal{K}} R(x)$ and $B_R(x^*||x_1) \leq R(x^*) - R(x_1)$.

7. Assuming the identities
$$B_R(x||y) = B_{R^*}(\nabla R(y)||\nabla R(x)) = \frac{1}{2}\|x-y\|^2_{\nabla^2 R(z)} \tag{1}$$
for some $z \in \mathcal{K}$ and any $R$, derive the desired bound.

8. *Application.* We consider $R(x) = \frac{1}{2}\|x\|_p^2$, $1 < p \leq 2$, admitting that $R$ is a $p-1$ strongly convex function such that $R^*(y^*) = \frac{1}{2(p-1)}\|y^*\|_q^2$ with $\frac{1}{p} + \frac{1}{q} = 1$. Design an algorithm satisfying the regret
$$Regret_T \leq G_\infty z \sqrt{Tq} d^{1/q}\,,$$
over $\mathcal{K} = B_1(z)$, $z > 0$ with $G_\infty$ satisfying $\|\nabla f_t(x_t)\|_\infty \leq G_\infty$ for all $t \geq 1$.

9. Optimizing the above regret bound in $q$, derive an algorithm with regret $O(\sqrt{T\log d})$. The performances are comparable with EG+/-, why EG+/- is preferable?

Bonus Prove the identities (1).

# End of Examination Paper