ONLINE CONVEX OPTIMISATION WITH PERTURBED CONSTRAINTS

Anubhav Mittal

November 26, 2019

Dynamic Regret with dynamic step-size

1 Introduction

The Online convex optimization (OCO) problem consists of a sequence of games where an agent selects an action x_t from a feasible convex set C and suffers a cost $f_t(x_t)$, whose functional form is not known at the time of making a decision. The goal is to mimimize the cost incurred, which can be obtained by mimicking the respective offline version by minimizing the regret:

$$R_s(T) = \sum_{t=1}^{T} f_t(x_t) - min_{x \in C} \sum_{t=1}^{T} f_t(x)$$

We call the above as 'static' regret as the benchmark is minimum of the sum $\sum_{t=1}^{T} f_t(x)$ over a static variable $x \in C$. Another popular performance metric is the 'dynamic' regret. Let $z_t = \arg\min_{xinC} f_t(x)$, then:

$$R_d(T) = \sum_{t=1}^{T} f_t(x_t) - \sum_{t=1}^{T} f_t(z_t)$$

.

Extending the above problem, we also have a collection of m convex constraints $g^j(x) : \mathbb{R}^n \to \mathbb{R}$, each with an associated perturbation b_t^j, that needs to be satisfied on average. The optimal decision y_t for timestep t should satisfy $g(z_t) + b_t \leq 0$.

We try to find a sublinear dynamic regret for the primal-dual proximal algorithm described in [1]. As we are not aware of the perturbation b_t at the time of selecting x_t, each action also contributes to the constraint violation:

$$V(T) = \left\| \left[\sum_{t=1}^{T} g(x_t) + b_t \right]^+ \right\|$$

2 Assumptions

We take the same assumptions as given in [1] with some extensions:

Bounded set Let C be the convex set that contains the admissible actions x_t. Set C is bounded i.e. $\|u - v\| \leq D, u, v \in C$

Bounded perturbation $\|b_t\| < \infty$ for all $t \in \mathbb{N}$.

Bounded Subgradients There exists constants F_*, G_*, G such that $\left\|f'(x)\right\|_* \leq F_*, \left\|g(x) + b_t\right\|_* \leq G_*, *, \left\|g(x) + b_t\right\| \leq G$ for all $x \in C, t \in N$.

Slater's condition There exists $\eta > 0$ s.t. $g(x) + b_t + \eta \leq 0$ for some $x \in C$ for all $t \in N$.

Bregman functions The functions ψ, φ are $\sigma_\psi, \sigma_\varphi$ Strongly convex and L_ψ, L_φ smooth. Additionally, ψ is K Lipschitz continous.

3 Lemmas and Proofs

3.1 Lemma 1

We produce the Lemma 1 from [1] for convenience. The proof is straightforward and can be found there.

Consider the following update:

$$x_{t+1} = \arg\min_{u \in C} \phi(u) + \frac{1}{\rho_t} B_\psi(u, x_t) \tag{1}$$

For $z_t \in C$, the following bound holds:

$$\phi(x_{t+1}) - \phi(z_t) \leq \frac{1}{\rho_t}(B_\psi(z_t, x_t) - B_\psi(z_t, x_{t+1}) - B_\psi(x_{t+1}, x_t)) \tag{2}$$

3.2 Lemma 2

Again, we produce the Lemma 2 from [1].

Let $\phi(u) = \langle f'(x), u\rangle + \theta(u)$. Then, for $z_t \in C$, the following bound holds:

$$f(x_t) - f(z_t) + \theta(x_{t+1}) - \theta(z_t) \leq \frac{1}{\rho_t}(B_\psi(z_t, x_t) - B_\psi(z_t, x_{t+1})) + 2\frac{F_*^2}{\sigma_\psi}\rho_t \tag{3}$$

3.3 Lemma 3

Consider the update:

$$x_{t+1} = \arg\min_{u \in C} \left\{ \langle f_t'(x_t), u\rangle + \langle y_t, g(u) + b_t\rangle + \frac{1}{\rho_t} B_\psi(u, x_t) \right\} \tag{4}$$

Then, for arbitrary sequence $\{y_t\}$ and optimal decision variables $\{z_t\}$, the following holds:

$$R(T) \leq \frac{L_\psi D^2}{2\rho_1} + \sum_{t=1}^{T} \frac{K}{\rho_{t+1}} \|z_{t+1} - z_t\| + 2\frac{F_*^2}{\sigma_\psi} \sum_{t=1}^{T} \rho_t - \sum_{t=1}^{T} \langle y_t, g(x_{t+1}) + b_t\rangle \tag{5}$$

Proof:

Let $\phi(u) = \langle f_t'(x_t), u\rangle + \langle y_t, g(u) + b_t\rangle$.

We have

$$f(x_t) - f(z_t) + \langle y_t, g(x_{t+1}) + b_t\rangle - \langle y_t, g(z_t) + b_t\rangle \leq \frac{1}{\rho_t}(B_\psi(z_t, x_t) - B_\psi(z_t, x_{t+1})) + 2\frac{F_*^2}{\sigma_\psi}\rho_t$$

Since $\langle y_t, g(z_t) + b_t\rangle \leq 0$,

$$f(x_t) - f(z_t) + \langle y_t, g(x_{t+1}) + b_t\rangle \leq \frac{1}{\rho_t}(B_\psi(z_t, x_t) - B_\psi(z_t, x_{t+1})) + 2\frac{F_*^2}{\sigma_\psi}\rho_t$$

Summing over $t = 1, 2, ..., T$ and rearranging,

$$\sum_{t=1}^{T}[f(x_t) - f(z_t)] \leq \sum_{t=1}^{T} \frac{1}{\rho_t}(B_\psi(z_t, x_t) - B_\psi(z_t, x_{t+1})) + 2\frac{F_*^2}{\sigma_\psi}\sum_{t=1}^{T}\rho_t - \sum_{t=1}^{T}\langle y_t, g(x_{t+1}) + b_t\rangle$$

Now, let's consider the first term of the right of the inequality.

$$\begin{aligned}
\frac{1}{\rho_t}(B_\psi(z_t, x_t) - B_\psi(z_t, x_{t+1})) &= \frac{1}{\rho_t}B_\psi(z_t, x_t) - \frac{1}{\rho_{t+1}}B_\psi(z_{t+1}, x_{t+1}) && \text{(a)}\\
&+ \frac{1}{\rho_{t+1}}B_\psi(z_{t+1}, x_{t+1}) - \frac{1}{\rho_{t+1}}B_\psi(z_t, x_{t+1}) && \text{(b)}\\
&+ \frac{1}{\rho_{t+1}}B_\psi(z_t, x_{t+1}) - \frac{1}{\rho_t}B_\psi(z_t, x_{t+1}) && \text{(c)}
\end{aligned}$$

We bound each of the three terms above.

(a) The first term telescopes when summed from $t = 1$ to T to give:

$$\begin{aligned}
\sum_{t=1}^{T}(\frac{1}{\rho_t}B_\psi(z_t, x_t) - \frac{1}{\rho_{t+1}}B_\psi(z_{t+1}, x_{t+1})) &= \frac{1}{\rho_1}B_\psi(z_1, x_1) - \frac{1}{\rho_{T+1}}B_\psi(z_{T+1}, x_{T+1})\\
&\leq \frac{1}{\rho_1}B_\psi(z_1, x_1) \leq \frac{L_\psi D^2}{2\rho_1}
\end{aligned}$$

(b) As ψ is K-Lipschitz continous,

$$\frac{1}{\rho_{t+1}}B_\psi(z_{t+1}, x_{t+1}) - \frac{1}{\rho_{t+1}}B_\psi(z_t, x_{t+1}) \leq \frac{K}{\rho_{t+1}}\|z_{t+1} - z_t\|$$

$$\implies \sum_{t=1}^{T}(\frac{1}{\rho_{t+1}}B_\psi(z_{t+1}, x_{t+1}) - \frac{1}{\rho_{t+1}}B_\psi(z_t, x_{t+1})) \leq \sum_{t=1}^{T}\frac{K}{\rho_{t+1}}\|z_{t+1} - z_t\|$$

(c) Since $\rho_{t+1} \leq \rho_t$, and Bregman distance is always postive, the third term

$$B_\psi(z_t, x_{t+1})(\frac{1}{\rho_{t+1}} - \frac{1}{\rho_t}) \leq 0$$

Combining, we get:

$$\sum_{t=1}^{T}\frac{1}{\rho_t}(B_\psi(z_t, x_t) - B_\psi(z_t, x_{t+1})) \leq \frac{L_\psi D^2}{2\rho_1} + \sum_{t=1}^{T}\frac{K}{\rho_{t+1}}\|z_{t+1} - z_t\|$$

Finally, using the inequality above, we get:

$$\sum_{t=1}^{T}[f(x_t) - f(z_t)] \leq \frac{L_\psi D^2}{2\rho_1} + \sum_{t=1}^{T}\frac{K}{\rho_{t+1}}\|z_{t+1} - z_t\| + 2\frac{F_*^2}{\sigma_\psi}\sum_{t=1}^{T}\rho_t - \sum_{t=1}^{T}\langle y_t, g(x_{t+1}) + b_t\rangle$$

3.4 Lemma 4

To make the $-\sum_{t=1}^{T}\langle y_t, g(x_{t+1}) + b_t\rangle$ term vanish, we apply the proximal gradient update:

$$y_{t+1} \longleftarrow \arg\max_{v \in \mathbb{R}_+^m}\left\{\langle v, g(x_{t+1} + b_t) - \frac{1}{\rho_t}B_\varphi(v, y_t)\right\} \tag{6}$$

For the above update, the following bound holds:

$$\sum_{t=1}^{T}\langle y_t, -(g(x_{t+1}) + b_t)\rangle \leq \frac{L_\varphi E^2}{2\rho_T} + 2\frac{G_*^2}{\sigma_\varphi}\sum_{t=1}^{T}\rho_t \tag{7}$$

Proof:

Let

$$\phi(y) = \langle y, -(g(x_{t+1}) + b_t)\rangle \tag{8}$$

in Lemma 2. Summing from $t = 1, 2, ..., T$,

$$\sum_{t=1}^{T}\langle y_t, -(g(x_{t+1}) + b_t)\rangle - \langle z, -(g(x_{t+1}) + b_t)\rangle \leq \frac{1}{\rho_t}\sum_{t=1}^{T}(B_\varphi(z, y_t) - B_\varphi(z, y_{t+1})) + 2\frac{G_*^2}{\sigma_\varphi}\sum_{t=1}^{T}\rho_t$$

Choosing $z = 0$, we get:

$$\begin{aligned}\sum_{t=1}^{T}\langle y_t, -(g(x_{t+1}) + b_t)\rangle &\leq \frac{1}{\rho_t}\sum_{t=1}^{T}(B_\varphi(0, y_t) - B_\varphi(0, y_{t+1})) + 2\frac{G_*^2}{\sigma_\varphi}\sum_{t=1}^{T}\rho_t \\ &= \frac{1}{\rho_1}B_\varphi(0, y_1) - \frac{1}{\rho_T}B_\varphi(0, y_{T+1}) + \sum_{t=2}^{T}B_\varphi(0, y_t)\left(\frac{1}{\rho_t} - \frac{1}{\rho_{t-1}}\right) + 2\frac{G_*^2}{\sigma_\varphi}\sum_{t=1}^{T}\rho_t\end{aligned}$$

Now, by smoothness of φ, $B_\varphi(0, y_t) \leq \frac{L_\varphi\|y_t\|^2}{2} \leq \frac{L_\varphi E^2}{2}$ (Lemma 7 of [1] gives $\|y_t\| \leq E$). We now get:

$$\begin{aligned}\sum_{t=1}^{T}\langle y_t, -(g(x_{t+1}) + b_t)\rangle &\leq \frac{L_\varphi E^2}{2}\left(\frac{1}{\rho_1} + \sum_{t=2}^{T}\left(\frac{1}{\rho_t} - \frac{1}{\rho_{t-1}}\right)\right) + 2\frac{G_*^2}{\sigma_\varphi}\sum_{t=1}^{T}\rho_t \\ &= \frac{L_\varphi E^2}{2\rho_T} + 2\frac{G_*^2}{\sigma_\varphi}\sum_{t=1}^{T}\rho_t\end{aligned}$$

Now, from the above inequality and 24, we get,

$$\boxed{R(T) \leq \frac{L_\psi D^2}{2\rho_1} + \sum_{t=1}^{T}\frac{K}{\rho_{t+1}}\|z_{t+1} - z_t\| + 2\left(\frac{F_*^2}{\sigma_\psi} + \frac{G_*^2}{\sigma_\varphi}\right)\sum_{t=1}^{T}\rho_t + \frac{L_\varphi E^2}{2\rho_T}} \tag{9}$$

3.5 Lemma 5

Starting from $y_1 = 0, x_1 \in C$, the updates in 4 and 6, and for $\|y_t\| \leq E$, the following bound on constrain violation holds:

$$\boxed{\left\|\left[\sum_{t=1}^{T} g(x_t) + b_t\right]^+\right\| \leq G + \frac{1}{2\rho_T}L_\varphi E} \tag{10}$$

Proof:

The analysis followed is almost identical to that of Proposition 1 in [1], so we omit most of it here and get:

$$\left\|\left[\sum_{t=1}^{T} g(x_t) + b_t\right]^+\right\| \leq \left\|g(x_1) + b_T + \frac{\nabla\varphi(y_T)}{\rho_T}\right\| \leq \|g(x_1) + b_T\| + \frac{1}{\rho_T}\|\nabla\varphi(y_T)\|$$

Now, by smoothness of φ, $\|\nabla\varphi(y_T)\| \leq \frac{L_\varphi\|y_T\|}{2} \leq \frac{L_\varphi E}{2}$ (Proved in Lemma 7 of [1] that $\|y_t\| \leq E$). Also, by assumption, $\|g(x_1) + b_T\| \leq G$.

$$\left\|\left[\sum_{t=1}^{T} g(x_t) + b_t\right]^+\right\| \leq G + \frac{1}{2\rho_T}L_\varphi E$$

3.6 Remark

For simplicity, we can choose $\rho_t = \frac{1}{t^\epsilon}$. Then,

$$\boxed{R_d(T) = O(\sum_{t=1}^{T-1} t^{\epsilon+1} K \|z_{t+1} - z_t\| \bigwedge T^\epsilon \bigwedge T^{1-\epsilon})} \tag{11}$$

$$\boxed{\left\| \left[\sum_{t=1}^{T} g(x_t) + b_t \right]^+ \right\| = O(T^\epsilon)} \tag{12}$$

No Constraint Violation Case

1 Introduction

We try to use the idea suggested in [2] to remove the violation of constraints, although possibly trading it with restrictions in the choice of optimal values for regret calculation. Change the Langrangain to the form

$$\mathcal{L}_t(x, y) = \langle f_t(x_t), x\rangle + \langle y, g(x) + b_t + \delta\rangle \tag{13}$$

where δ is a variable that we choose later. We introduce it to the Dynamic step-size-static regret approach in [1] (changing the analysis to fit our Constant Step-size-Dynamic regret case should be straightforward). This leads to the following changes in the updates:

$$x_{t+1} \leftarrow \arg\min_{u\in C}\{\mathcal{L}_t(u, y_t) + \frac{1}{\rho}B_\psi(u, x_t)\} \tag{14}$$

$$y_{t+1} \leftarrow \arg\max_{v\in R^m_+}\{\langle v, g(x_{t+1}) + b_{t+1} + \delta\rangle - \frac{1}{\rho}B_\varphi(v, y_t)\} \tag{15}$$

2 Lemmas and Proofs

Lemma 1 and 2 from [1] remain unchanged.

2.1 Lemma 3

Consider the updates as described in 14 and 15. Then the following bound holds:

$$R_s(T) \leq \frac{1}{\rho_T}\left(\frac{L_\psi}{2}D^2 + \frac{L_\varphi}{2}E^2\right) + \left(\frac{2F_*^2}{\sigma_\psi} + \frac{2G_*^2}{\sigma_\varphi}\right)\sum_{t=1}^{T}\rho_t \tag{16}$$

Proof:

Let $\phi(u) = \langle f_t'(x_t), u\rangle + \langle y_t, g(u) + b_{t+1} + \delta\rangle$.

We have

$$f(x_t) - f(z) + \langle y_t, g(x_{t+1}) + b_t + \delta\rangle - \langle y_t, g(z) + b_t + \delta\rangle \leq \frac{1}{\rho_t}(B_\psi(z, x_t) - B_\psi(z, x_{t+1})) + 2\frac{F_*^2}{\sigma_\psi}\rho_t$$

Now, the term $-\langle y_t, g(z) + b_t + \delta\rangle$ is similar to that mentioned in [1] as $-\langle y_t, g(z) + b_t\rangle$ with the added δ term. We modify the definition of the 'optimal' z selected from the set:

$$X_T := \{x \in C | g(x) + \bar{b}_T + \delta \leq 0\} \tag{17}$$

We later show that the value of δ approaches 0 as $T \to \infty$, where the two inequalities become the same. Assuming that z takes values from the set in 17, we get:

$$f(x_t) - f(z) \leq \frac{1}{\rho_t}(B_\psi(z, x_t) - B_\psi(z, x_{t+1})) + 2\frac{F_*^2}{\sigma_\psi}\rho_t - \langle y_t, g(x_{t+1}) + b_t + \delta\rangle$$

Following the updates for the primal and dual variables as described in 14and 15,the rest of the proof for regret follows exactly as in [1], to get

$$R_s(T) \leq \frac{1}{\rho_T}\left(\frac{L_\psi}{2}D^2 + \frac{L_\varphi}{2}E^2\right) + \left(\frac{2F_*^2}{\sigma_\psi} + \frac{2G_*^2}{\sigma_\varphi}\right)\sum_{t=1}^{T}\rho_t$$

2.2 Lemma 4

(Constraint Violation) If we choose $\delta = \frac{L_\varphi E}{2\rho_T T}$*, we ensure that there will be **no** long-term constraint violation.*

Proof:

For constraint violation, following similar steps as in Proposition 1 of [1], we get,

$$\sum_{t=1}^{T}(g(x)+b_t)+\delta T \leq \frac{\nabla\varphi(y_T)}{\rho_T}$$
$$\implies \sum_{t=1}^{T}(g(x)+b_t) \leq \frac{\nabla\varphi(y_T)}{\rho_T} - \delta T$$

Now, $\left\|\nabla\varphi(y_T)\right\| \leq \frac{L_\varphi\|y_T\|}{2} \leq \frac{L_\varphi E}{2}$.

Choose $\delta = \frac{L_\varphi E}{2\rho_T T}$. This ensures that the long-term constraint is not violated. If $\rho_t = t^{-\epsilon}$ as given in [1], we get

$$\delta = \frac{L_\varphi E}{2T^{1-\epsilon}} = O(1/T^{1-\epsilon}) \tag{18}$$

which goes to 0 as $T \to \infty$, reducing feasible set defined in 17 to its standard form.

Dsitributed Online Convex Constrained Optimization with primal consensus step

1 Introduction

We try to extend the centralized algorithm to a distributed form, taking ideas from [3] and [4]. In addition to the convex optimisation problem defined before, we now want to solve the constrained optimisation problem in a decentralised fashion. In particular, the global function at each time t can be written as sum of n local functions as

$$f_t(x) = \frac{1}{n}\sum_{i=1}^{n} f_{i,t}(x) \quad g_t(x) = \frac{1}{n}\sum_{i=1}^{n} g_{i,t}(x)$$

We have a network of n agents, and the agent j receives information only about $f_{i,t}(.), g_{i,t}(.)$. The agents can exchange information with each other via an undirected graph W, and agent i assigns a positive weight $[W]_{ij}$ to the information received from $j \neq i$. The dynamic regret is defined as:

$$R_d(T) = \frac{1}{n}\sum_{i=1}^{n}\sum_{t=1}^{T} f_t(x_{i,t}) - \sum_{t=1}^{T} f_t(z_t)$$

The optimal decision for timestep t must satisfy $z_t = \arg\min_{xinC} f_t(x)$ and $g_t(z_t) \leq 0$ for all i.

Also, assume that a dynamics A is common knowledge in the network,

$$z_{t+1} = Az_t + v_t$$

and we get regret in terms of $\sum_{t=1}^{T}\|z_{t+1} - Az_t\| = \sum_{t=1}^{T}\|v_t\|$. If there is no such A, then we simply put $A = I$ and calculate regret in terms of $\sum_{t=1}^{T}\|z_{t+1} - z_t\|$.

2 Analysis

The assumptions are the same as defined before, except that they now hold for each individual $f_{i,t}(.)$ and $g_{i,t}(.)$ respectively. Adding to these are the following,

Assumption 1: Connectedness and Double-stochasticity The network is connected i.e. there exists a path from i to j for all i, j. Also, the matrix W is doubly stochastic with a positive diagonal,

$$\sum_{i=1}^{n}[W]_{ij} = \sum_{j=1}^{n}[W]_{ij} = 1$$

Assumption 2 : Non-expansiveness The mapping A is assumed to be non-expansive w.r.t. ψ, that is $B_\psi(Ax, Ay) \leq B_\psi(x, y)$ for all $x, y \in C$ and $\|A\| \leq 1$

Assumption 3: Bounded set Let C be the convex set that contains the admissible actions x_t. Set C is bounded i.e. $\|u - v\| \leq D, u, v \in C$

Assumption 4: Bounded perturbation $\left\|b_{i,t}\right\| < \infty$ for all $t \in \mathbb{N}$.

Assumption 5: Bounded Subgradients There exists constants F_*, G_*, G such that $\left\|f'(x)\right\|_* \leq F_*, \left\|g_i(x) + b_{i,t}\right\|_* \leq G_*, \left\|g_i(x) + b_{i,t}\right\| \leq G$ for all $x \in C, t \in N$ and for all $i \in 1, 2, ..., n$.

Algorithm 1: Distributed Online convex constrained optimization

Input: Bregman functions ψ and φ, constraint function $g(.)$, set C
Initialize: $x_{i,1} \in C, \nabla f_{i,1} = 0, b_{i,1} = 0, \ \forall i \in [n]$
for $t = 1, ..., T$ **do**
  $\rho_t \leftarrow \frac{1}{t^\epsilon}$
  for $i = 1, ..., n$ **do**
    $\hat{x}_{i,t+1} \leftarrow \arg\min_{u \in C} \{ \langle \nabla f_{i,t}(x_{i,t}), u \rangle + \langle y_{i,t}, g_i(u) + b_{i,t} \rangle + \frac{1}{\rho_t} B_\psi(u, x_{i,t}) \ \}$
    $\hat{y}_{i,t+1} \leftarrow \arg\max_{v \in \mathbb{R}^m_+} \{ \langle v, g_i(x_{i,t}) + b_{i,t} \rangle - \frac{1}{\rho_t} B_\varphi(v, y_{i,t}) \}$
    $x_{i,t+1} \leftarrow \sum_{i=1}^n [W]_{ij} A \hat{x}_{i,t+1}$
    $y_{i,t+1} \leftarrow \sum_{i=1}^n [W]_{ij} \hat{y}_{i,t+1}$
    Play $x_{i,t+1}$ and receive $\nabla f_{i,t+1}(x_{i,t+1}), b_{i,t+1}$

Assumption 6: Slater's condition There exists $\eta > 0$ s.t. $g_i(x) + b_t + \eta \leq 0$ for some $x \in C$ for all $t \in N$.
Assumption 7: Bregman functions The functions ψ, φ are $\sigma_\psi, \sigma_\varphi$ Strongly convex and L_ψ, L_φ smooth. Additionally, ψ is K Lipschitz continous.

2.1 Lemma 1

We produce the Lemma 1 from [1] for convenience. The proof is straightforward and can be found there.

Consider the following update:

$$\hat{x}_{i,t+1} = \arg\min_{u \in C} \phi(u) + \frac{1}{\rho_t} B_\psi(u, x_{i,t}) \tag{19}$$

For $z_t \in C$, the following bound holds:

$$\phi(\hat{x}_{i,t+1}) - \phi(z_t) \leq \frac{1}{\rho_t} (B_\psi(z_t, x_{i,t}) - B_\psi(z_t, \hat{x}_{i,t+1}) - B_\psi(\hat{x}_{i,t+1}, x_{i,t})) \tag{20}$$

2.2 Lemma 2

Let $\phi(u) = \langle \nabla f(x_{i,t}), u \rangle + \theta(u)$. Then, for $z_t \in C$, the following bound holds:

$$f_{i,t}(x_{i,t}) - f_{i,t}(z_t) + \theta(\hat{x}_{i,t+1}) - \theta(z_t) \leq \frac{1}{\rho_t} (B_\psi(z_t, x_{i,t}) - B_\psi(z_t, \hat{x}_{i,t+1})) + 2 \frac{F_*^2}{\sigma_\psi} \rho_t \tag{21}$$

2.3 Lemma 3

Consider the update:

$$\hat{x}_{i,t+1} \leftarrow \arg\min_{u \in C} \left\{ \langle \nabla f_{i,t}(x_{i,t}), u \rangle + \langle y_{i,t}, g_i(u) + b_{i,t} \rangle + \frac{1}{\rho_t} B_\psi(u, x_{i,t}) \right\} \tag{22}$$

$$x_{i,t+1} = \sum_{j=1}^n [W]_{ij} A \hat{x}_{i,t+1} \tag{23}$$

Then, for arbitrary sequence $\{y_t\}$ and optimal decision variables $\{z_t\}$, the following holds:

$$R(T) \leq 2 \frac{F_*^2}{\sigma_\psi} \sum_{t=1}^T \rho_t + \sum_{t=1}^T \frac{K}{\rho_{t+1}} \|z_{t+1} - z_t\| + \frac{L_\psi D^2}{2\rho_{T+1}} - \frac{1}{n} \sum_{t=1}^T \sum_{i=1}^n \langle y_{i,t}, g_i(\hat{x}_{i,t+1}) + b_t \rangle \tag{24}$$

Proof:

Let $\phi(u) = \langle \nabla f_{i,t}(x_{i,t}), u \rangle + \langle y_{i,t}, g_i(u) + b_{i,t} \rangle$.

We have

$$f_{i,t}(x_{i,t}) - f_{i,t}(z_t) + \langle y_{i,t}, g_i(\hat{x}_{i,t+1}) + b_{i,t}\rangle - \langle y_{i,t}, g_i(z_t) + b_{i,t}\rangle \leq \frac{1}{\rho_t}(B_\psi(z_t, x_{i,t}) - B_\psi(z_t, \hat{x}_{i,t+1})) + 2\frac{F_*^2}{\sigma_\psi}\rho_t$$

$$\implies f_{i,t}(x_{i,t}) - f_{i,t}(z_t) + \langle y_{i,t}, g_i(\hat{x}_{i,t+1}) + b_{i,t}\rangle - \langle \tilde{y}_t, g_i(z_t) + b_{i,t}\rangle \leq \frac{1}{\rho_t}(B_\psi(z_t, x_{i,t}) - B_\psi(z_t, \hat{x}_{i,t+1})) + 2\frac{F_*^2}{\sigma_\psi}\rho_t + \langle y_{i,t} - \tilde{y}_t, g_i(z_t) + b_{i,t}\rangle$$

Summing over $t = 1, 2, ..., T$, $i = 1, 2, ..., n$ and rearranging,

$$\sum_{t=1}^{T}\sum_{i=1}^{n}[f_{i,t}(x_{i,t}) - f_{i,t}(z_t)] + \sum_{t=1}^{T}\sum_{i=1}^{n}\langle y_{i,t}, g_i(\hat{x}_{i,t+1}) + b_{i,t}\rangle - \sum_{t=1}^{T}\langle \tilde{y}_t, \sum_{i=1}^{n} g_i(z_t) + b_{i,t}\rangle$$
$$\leq \sum_{t=1}^{T}\sum_{i=1}^{n}\frac{1}{\rho_t}(B_\psi(z_t, x_{i,t}) - B_\psi(z_t, \hat{x}_{i,t+1})) + 2\frac{nF_*^2}{\sigma_\psi}\sum_{t=1}^{T}\rho_t + \sum_{t=1}^{T}\sum_{i=1}^{n}\langle y_{i,t} - \tilde{y}_t, g_i(z_t) + b_{i,t}\rangle$$

Now, $y_{i,t} \geq 0 \implies \tilde{y}_t \geq 0$, and $\sum_{i=1}^{n} g_i(z_t) + b_{i,t} \leq 0$, from which, $\langle \tilde{y}_t, \sum_{i=1}^{n} g_i(z_t) + b_{i,t}\rangle \leq 0$ follows.
Also, $\langle y_{i,t} - \tilde{y}_t, g_i(z_t) + b_{i,t}\rangle \leq G_* \|y_{i,t} - \tilde{y}_t\|$.

$$\sum_{t=1}^{T}\sum_{i=1}^{n}[f_{i,t}(x_{i,t}) - f_{i,t}(z_t)] + \sum_{t=1}^{T}\sum_{i=1}^{n}\langle y_{i,t}, g_i(\hat{x}_{i,t+1}) + b_{i,t}\rangle$$
$$\leq \sum_{t=1}^{T}\sum_{i=1}^{n}\frac{1}{\rho_t}(B_\psi(z_t, x_{i,t}) - B_\psi(z_t, \hat{x}_{i,t+1})) + 2\frac{nF_*^2}{\sigma_\psi}\sum_{t=1}^{T}\rho_t + G_*\sum_{t=1}^{T}\sum_{i=1}^{n}\|y_{i,t} - \tilde{y}_t\|$$

Now, let's consider the first term of the right of the inequality.

$$\begin{aligned}
\frac{1}{\rho_t}(B_\psi(z_t, x_{i,t}) - B_\psi(z_t, \hat{x}_{i,t+1})) &= \frac{1}{\rho_t}B_\psi(z_t, x_{i,t}) - \frac{1}{\rho_{t+1}}B_\psi(z_{t+1}, x_{i,t+1}) && \text{(a)}\\
&+ \frac{1}{\rho_{t+1}}B_\psi(z_{t+1}, x_{i,t+1}) - \frac{1}{\rho_{t+1}}B_\psi(Az_t, x_{i,t+1}) && \text{(b)}\\
&+ \frac{1}{\rho_{t+1}}B_\psi(Az_t, x_{i,t+1}) - \frac{1}{\rho_{t+1}}B_\psi(z_t, \hat{x}_{i,t+1}) && \text{(c)}\\
&+ \frac{1}{\rho_{t+1}}B_\psi(z_t, \hat{x}_{i,t+1}) - \frac{1}{\rho_t}B_\psi(z_t, \hat{x}_{i,t+1}) && \text{(d)}
\end{aligned}$$

We bound each of the three terms above.

(a) The first term telescopes when summed from $t = 1$ to T to give:

$$\begin{aligned}
\sum_{t=1}^{T}(\frac{1}{\rho_t}B_\psi(z_t, x_{i,t}) - \frac{1}{\rho_{t+1}}B_\psi(z_{t+1}, x_{i,t+1})) &= \frac{1}{\rho_1}B_\psi(z_{i,1}, x_{i,1}) - \frac{1}{\rho_{T+1}}B_\psi(z_{t+1}, x_{i,T+1})\\
&\leq \frac{1}{\rho_1}B_\psi(z_{i,1}, x_{i,1}) \leq \frac{L_\psi D^2}{2\rho_1}
\end{aligned}$$

(b) As ψ is K-Lipschitz continous,

$$\frac{1}{\rho_{t+1}}B_\psi(z_{t+1}, x_{i,t+1}) - \frac{1}{\rho_{t+1}}B_\psi(Az_t, x_{i,t+1}) \leq \frac{K}{\rho_{t+1}}\|z_{t+1} - Az_t\|$$
$$\implies \sum_{t=1}^{T}\sum_{i=1}^{n}(\frac{1}{\rho_{t+1}}B_\psi(z_{t+1}, x_{i,t+1}) - \frac{1}{\rho_{t+1}}B_\psi(z_t, x_{i,t+1})) \leq \sum_{t=1}^{T}\frac{nK}{\rho_{t+1}}\|z_{t+1} - Az_t\|$$

(c) For the third term, summing over i, we get

$$\begin{aligned}\sum_{i=1}^{n} B_\psi(Az_t, x_{i,t+1}) - \sum_{i=1}^{n} B_\psi(z_t, \hat{x}_{i,t+1}) &= \sum_{i=1}^{n} B_\psi(Az_t, \sum_{j=1}^{n}[W]_{ij} A\hat{x}_{j,t+1}) - \sum_{i=1}^{n} B_\psi(z_t, \hat{x}_{i,t+1}) \\ &\leq \sum_{i=1}^{n}\sum_{j=1}^{n}[W]_{ij} B_\psi(Az_t, A\hat{x}_{j,t+1}) - \sum_{i=1}^{n} B_\psi(z_t, \hat{x}_{i,t+1}) \\ &= \sum_{j=1}^{n} B_\psi(Az_t, A\hat{x}_{j,t+1}) - \sum_{i=1}^{n} B_\psi(z_t, \hat{x}_{i,t+1}) \leq 0\end{aligned}$$

where the first inequality follows from the assumption that Bregman divergence is separately convex, and the second inequality follows as A is non-expansive.

(d) For the fourth term, summing over t gives

$$\begin{aligned}\sum_{t=1}^{T} B_\psi(z_t, \hat{x}_{i,t+1})(\frac{1}{\rho_{t+1}} - \frac{1}{\rho_t}) &\leq \frac{L_\psi D^2}{2} \sum_{t=1}^{T}(\frac{1}{\rho_{t+1}} - \frac{1}{\rho_t}) \\ &= \frac{L_\psi D^2}{2}(\frac{1}{\rho_{T+1}} - \frac{1}{\rho_1})\end{aligned}$$

Combining, we get:

$$\sum_{t=1}^{T}\sum_{i=1}^{n} \frac{1}{\rho_t}(B_\psi(z_t, x_t) - B_\psi(z_t, x_{t+1})) \leq \frac{L_\psi D^2}{2\rho_{T+1}} + \sum_{t=1}^{T} \frac{nK}{\rho_{t+1}} \|z_{t+1} - Az_t\|$$

Finally, using the inequality above, we get:

$$\begin{aligned}\sum_{t=1}^{T}\sum_{i=1}^{n}[f(x_{i,t}) - f(z_t)] &\leq 2\frac{nF_*^2}{\sigma_\psi}\sum_{t=1}^{T}\rho_t + \sum_{t=1}^{T}\frac{nK}{\rho_{t+1}}\|z_{t+1} - Az_t\| + \frac{nL_\psi D^2}{2\rho_{T+1}} - \sum_{t=1}^{T}\sum_{i=1}^{n}\langle y_{i,t}, g_i(\hat{x}_{i,t+1}) + b_t\rangle \\ &\quad + G_*\sum_{t=1}^{T}\sum_{i=1}^{n}\|y_{i,t} - \tilde{y}_t\| \\ \implies \frac{1}{n}\sum_{t=1}^{T}\sum_{i=1}^{n}[f(x_{i,t}) - f(z_t)] &\leq 2\frac{F_*^2}{\sigma_\psi}\sum_{t=1}^{T}\rho_t + \sum_{t=1}^{T}\frac{K}{\rho_{t+1}}\|z_{t+1} - Az_t\| + \frac{L_\psi D^2}{2\rho_{T+1}} - \frac{1}{n}\sum_{t=1}^{T}\sum_{i=1}^{n}\langle y_{i,t}, g_i(\hat{x}_{i,t+1}) + b_t\rangle \\ &\quad + G_*\sum_{t=1}^{T}\sum_{i=1}^{n}\|y_{i,t} - \tilde{y}_t\|\end{aligned}$$

2.4 Lemma 4

To make the $-\sum_{t=1}^{T}\langle y_{i,t}, g(\hat{x}_{i,t+1}) + b_{i,t}\rangle$ term vanish, we apply the proximal gradient update:

$$\hat{y}_{i,t+1} \longleftarrow \arg\max_{v\in\mathbb{R}_+^m}\left\{\langle v, g(\hat{x}_{i,t+1} + b_{i,t}) - \frac{1}{\rho_t}B_\varphi(v, y_{i,t})\right\} \tag{25}$$

$$y_{i,t+1} = \sum_{j=1}^{n}[W]_{ij}\hat{y}_{i,t+1} \tag{26}$$

For the above update, the following bound holds:

$$\frac{1}{n}\sum_{t=1}^{T}\sum_{i=1}^{n}\langle y_{i,t}, -(g(x_{t+1}) + b_t)\rangle \leq \frac{L_\varphi E^2}{2\rho_T} + 2\frac{G_*^2}{\sigma_\varphi}\sum_{t=1}^{T}\rho_t \tag{27}$$

Proof:

Let

$$\phi(y) = \langle y, -(g_i(\hat{x}_{i,t+1}) + b_{i,t})\rangle \tag{28}$$

in Lemma 2. Summing from $t = 1, 2, ..., T$,

$$\sum_{t=1}^{T}\langle y_{i,t}, -(g(\hat{x}_{i,t+1}) + b_{i,t})\rangle - \langle z, -(g(\hat{x}_{i,t+1}) + b_{i,t})\rangle \leq \frac{1}{\rho_t}\sum_{t=1}^{T}(B_\varphi(z, y_{i,t}) - B_\varphi(z, \hat{y}_{i,t+1})) + 2\frac{G_*^2}{\sigma_\varphi}\sum_{t=1}^{T}\rho_t$$

Choosing $z = 0$, we get:

$$\sum_{t=1}^{T}\langle y_{i,t}, -(g(x_{t+1}) + b_t)\rangle \leq \frac{1}{\rho_t}\sum_{t=1}^{T}(B_\varphi(0, y_{i,t}) - B_\varphi(0, \hat{y}_{i,t+1})) + 2\frac{G_*^2}{\sigma_\varphi}\sum_{t=1}^{T}\rho_t$$

Now, by smoothness of φ, $B_\varphi(0, y_t) \leq \frac{L_\varphi \|y_{i,t}\|^2}{2} \leq \frac{L_\varphi E^2}{2}$ (Lemma 7 of [1] gives $\|y_{i,t}\| \leq E$). Following the same steps as in Lemma 4, we now get:

$$\begin{aligned}
\sum_{t=1}^{T}\langle y_{i,t}, -(g(x_{t+1}) + b_t)\rangle &\leq \frac{L_\varphi E^2}{2\rho_{T+1}} + 2\frac{G_*^2}{\sigma_\varphi}\sum_{t=1}^{T}\rho_t \\
\implies \frac{1}{n}\sum_{t=1}^{T}\sum_{i=1}^{n}\langle y_{i,t}, -(g(x_{t+1}) + b_t)\rangle &\leq \frac{L_\varphi E^2}{2\rho_{T+1}} + 2\frac{G_*^2}{\sigma_\varphi}\sum_{t=1}^{T}\rho_t
\end{aligned}$$

Now, from the above inequality and 24, we get,

$$\boxed{R(T) \leq 2\left(\frac{F_*^2}{\sigma_\psi} + \frac{G_*^2}{\sigma_\varphi}\right)\sum_{t=1}^{T}\rho_t + \sum_{t=1}^{T}\frac{K}{\rho_{t+1}}\|z_{t+1} - z_t\| + \frac{L_\psi D^2}{2\rho_{T+1}} + \frac{L_\varphi E^2}{2\rho_T}} \tag{29}$$

2.5 Lemma 5

Starting from $y_1 = 0, x_1 \in C$, the updates in 4 and 6, and for $\|y_t\| \leq E$, the following bound on constrain violation holds:

$$\boxed{\left\|\left[\frac{1}{n}\sum_{t=1}^{T}\sum_{i=1}^{n} g_i(x_{i,t}) + b_{i,t}\right]^+\right\| \leq G + \frac{1}{2\rho_T}L_\varphi E} \tag{30}$$

Proof:

The analysis followed is almost identical to that of Proposition 1 in [1], so we omit most of it here and get:

$$\left\|\left[\frac{1}{n}\sum_{t=1}^{T}\sum_{i=1}^{n} g_i(x_{i,t}) + b_{i,t}\right]^+\right\| \leq \left\|\frac{1}{n}\sum_{i=1}^{n}\left(g_i(x_{i,1}) + b_{i,T} + \frac{\nabla\varphi(y_{i,T})}{\rho_T}\right)\right\| \leq \frac{1}{n}\sum_{i=1}^{n}(\|g_i(x_{i,1}) + b_{i,T}\| + \frac{1}{\rho_T}\|\nabla\varphi(y_T)\|)$$

Now, by smoothness of φ, $\|\nabla\varphi(y_{i,T})\| \leq \frac{L_\varphi \|y_{i,T}\|}{2} \leq \frac{L_\varphi E}{2}$ (Proved in Lemma 7 of [1] that $\|y_t\| \leq E$). Also, by assumption, $\|g_i(x_{i,1}) + b_{i,T}\| \leq G$.

$$\left\|\left[\frac{1}{n}\sum_{t=1}^{T}\sum_{i=1}^{n} g_i(x_{i,t}) + b_{i,t}\right]^+\right\| \leq G + \frac{1}{2\rho_T}L_\varphi E$$

2.6 Remark 1

For simplicity, we can choose $\rho_t = \frac{1}{t^\epsilon}$. Then,

$$\boxed{R_d(T) = O(\sum_{t=1}^{T-1}(t+1)^\epsilon K\|z_{t+1} - z_t\| \bigwedge T^\epsilon \bigwedge T^{1-\epsilon})} \tag{31}$$

$$\boxed{\left\|\left[\frac{1}{n}\sum_{t=1}^{T}\sum_{i=1}^{n} g_i(x_{i,t}) + b_{i,t}\right]^+\right\| = O(T^\epsilon)} \tag{32}$$

3 Removing Constraint Violation

Assumption 8: Extended Slater's condition There exists $\eta > 0$ s.t. $g_i(x) + b_t + \delta + \eta \leq 0$ for some $x \in C$ for all $t \in N$, and $\delta = \frac{L_\varphi E}{2T^{1-\epsilon}}$. Since $\delta = O(1/T^{1-\epsilon})$, as $T \to \infty, \delta \to 0$ and we recover our original Slater's condition. Note that the feasible set remains the same as before i.e. $z_{i,t} \in \{x | g_i(z_{i,t}) + b_{i,t} \leq 0\}$

Algorithm 2: Distributed Online convex constrained optimization with no contraint violation

Input: Bregman functions ψ and φ, constraint function $g(.)$, set C, time duration T
Initialize: $x_{i,1} \in C, \nabla f_{i,1} = 0, b_{i,1} = 0, \ \forall i \in [n], \delta = \frac{L_\varphi E}{2T^{1-\epsilon}}\mathbf{1}^m$
for $t = 1, ..., T$ **do**
  $\rho_t \leftarrow \frac{1}{t^\epsilon}$
  for $i = 1, ..., n$ **do**
    $\hat{x}_{i,t+1} \leftarrow \arg\min_{u \in C}\{\langle \nabla f_{i,t}(x_{i,t}), u\rangle + \langle y_{i,t}, g_i(u) + b_{i,t} + \delta\rangle + \frac{1}{\rho_t}B_\psi(u, x_{i,t})\ \}$
    $y_{i,t+1} \leftarrow \arg\max_{v \in \mathbb{R}^m_+}\{\langle v, g_i(x_{i,t}) + b_{i,t} + \delta\rangle - \frac{1}{\rho_t}B_\varphi(v, y_{i,t})\}$
    $x_{i,t+1} \leftarrow \sum_{i=1}^n [W]_{ij} A\hat{x}_{i,t+1}$
    Play $x_{i,t+1}$ and receive $\nabla f_{i,t+1}(x_{i,t+1}), b_{i,t+1}$

3.1 Lemma 6

For the updates of primal and dual variables as given in Algorithm 2, the following holds:

$$\boxed{R(T) \leq 2\bigg(\frac{F_*^2}{\sigma_\psi} + \frac{G_*^2}{\sigma_\varphi}\bigg)\sum_{t=1}^T \rho_t + \sum_{t=1}^T \frac{K}{\rho_{t+1}}\|z_{t+1} - z_t\| + \frac{L_\psi D^2}{2\rho_{T+1}} + \frac{L_\varphi E^2}{2}\bigg(\frac{1}{\rho_T} + T^\epsilon\bigg)}$$

Proof:

We borrow most of the proof from that used in Lemma 3, as there are minor differences between the two.

Let $\phi(u) = \langle \nabla f_{i,t}(x_{i,t}), u\rangle + \langle y_{i,t}, g_i(u) + b_{i,t} + \delta\rangle$.

We have

$$f(x_{i,t}) - f(z_t) + \langle y_{i,t}, g(\hat{x}_{i,t+1}) + b_{i,t} + \delta\rangle - \langle y_{i,t}, g(z_t) + b_{i,t}\rangle - \langle y_{i,t}, \delta\rangle \leq \frac{1}{\rho_t}(B_\psi(z_t, x_{i,t}) - B_\psi(z_t, \hat{x}_{i,t+1})) + 2\frac{F_*^2}{\sigma_\psi}\rho_t$$

Since $\langle y_{i,t}, g(z_t) + b_{i,t}\rangle \leq 0$,

$$\begin{aligned} f(x_{i,t}) - f(z_t) + \langle y_{i,t}, g(\hat{x}_{i,t+1}) + b_{i,t} + \delta\rangle &\leq \frac{1}{\rho_t}(B_\psi(z_t, x_{i,t}) - B_\psi(z_t, \hat{x}_{i,t+1})) + 2\frac{F_*^2}{\sigma_\psi}\rho_t + \langle y_{i,t}, \delta\rangle \\ &= \frac{1}{\rho_t}(B_\psi(z_t, x_{i,t}) - B_\psi(z_t, \hat{x}_{i,t+1})) + 2\frac{F_*^2}{\sigma_\psi}\rho_t + \frac{L_\varphi E^2}{2T^{1-\epsilon}} \end{aligned}$$

Summing over $t = 1, 2, ..., T$, $i = 1, 2, ..., n$ and rearranging,

$$\begin{aligned} \sum_{t=1}^T\sum_{i=1}^n [f(x_{i,t}) - f(z_t)] \leq &\sum_{t=1}^T\sum_{i=1}^n \frac{1}{\rho_t}(B_\psi(z_t, x_{i,t}) - B_\psi(z_t, \hat{x}_{i,t+1})) + 2\frac{nF_*^2}{\sigma_\psi}\sum_{t=1}^T \rho_t \\ &+ \frac{nL_\varphi E^2}{2}T^\epsilon - \sum_{t=1}^T\sum_{i=1}^n \langle y_{i,t}, g(\hat{x}_{i,t+1}) + b_{i,t} + \delta\rangle \end{aligned}$$

The rest of the steps for bound of the non-simplified terms follows parallel from that in Lemma 3 and 4. The only difference is the term $-\sum_{t=1}^T\sum_{i=1}^n \langle y_{i,t}, g(\hat{x}_{i,t+1}) + b_{i,t} + \delta\rangle$ instead of $-\sum_{t=1}^T\sum_{i=1}^n \langle y_{i,t}, g(\hat{x}_{i,t+1}) + b_{i,t}\rangle$, which is compensated by the change in the dual variable update.

3.2 Remark 2

For simplicity, we can choose $\rho_t = \frac{1}{t^\epsilon}$. Then,

$$\boxed{R_d(T) = O(\sum_{t=1}^{T-1}(t+1)^\epsilon K\|z_{t+1} - z_t\| \bigwedge T^\epsilon \bigwedge T^{1-\epsilon})} \tag{33}$$

which is the same as in the previous case of Algorithm 1 (which had $O(T^\epsilon)$ constraint violation but did not require the extended slater's condition).

3.3 Lemma 7

(Constraint Violation) If we choose $\delta = \frac{L_\varphi E}{2\rho_T T}$*, we ensure that there will be* ***no*** *long-term constraint violation.*

Proof:

For constraint violation, following similar steps as in Proposition 1 of [1], we get,

$$\begin{aligned}
&\frac{1}{n}\sum_{t=1}^{T}\sum_{i=1}^{n}(g_i(x_{i,t}) + b_{i,t}) + \delta T \leq \frac{1}{n}\sum_{i=1}^{n}\frac{\nabla\varphi(y_{i,T})}{\rho_T} \\
&\implies \frac{1}{n}\sum_{t=1}^{T}\sum_{i=1}^{n}(g_i(x_{i,t}) + b_{i,t})\frac{1}{n}\sum_{i=1}^{n}\left(\frac{\nabla\varphi(y_{i,T})}{\rho_T} - \delta T\right)
\end{aligned}$$

Now, $\left\|\nabla\varphi(y_{i,T})\right\| \leq \frac{L_\varphi \left\|y_{i,T}\right\|}{2} \leq \frac{L_\varphi E}{2}$.

Choose $\delta = \frac{L_\varphi E}{2\rho_T T}$. This ensures that the long-term constraint is not violated. If $\rho_t = t^{-\epsilon}$ as given in [1], we get

$$\delta = \frac{L_\varphi E}{2T^{1-\epsilon}} = O(1/T^{1-\epsilon}) \tag{34}$$

which as stated before, goes to 0 as $T \to \infty$ and our extended slater's condition goes back to the original one.

References

[1] Valls, Víctor and Iosifidis, George and Leith, Douglas J and Tassiulas, Leandros. Online Convex Optimization with Perturbed Constraints. *arXiv preprint arXiv:1906.00049*, 2019.

[2] Mahdavi, Mehrdad, Rong Jin, and Tianbao Yang. "Trading regret for efficiency: online convex optimization with long term constraints." *Journal of Machine Learning Research* 13.Sep (2012): 2503-2528.

[3] Yi, X., Li, X., Xie, L. and Johansson, K.H., 2019. Distributed Online Convex Optimization with Time-Varying Coupled Inequality Constraints. *arXiv preprint arXiv:1903.04277*.

[4] Shahrampour, Shahin, and Ali Jadbabaie. "Distributed online optimization in dynamic environments using mirror descent." *IEEE Transactions on Automatic Control* 63, no. 3 (2017): 714-725.

[5] Lee, Soomin, and Michael M. Zavlanos. "On the sublinear regret of distributed primal-dual algorithms for online constrained optimization." arXiv preprint arXiv:1705.11128 (2017)