Supplementary Material for Online Learning under Adversarial Nonlinear Constraints

A Polyhedral Intersection

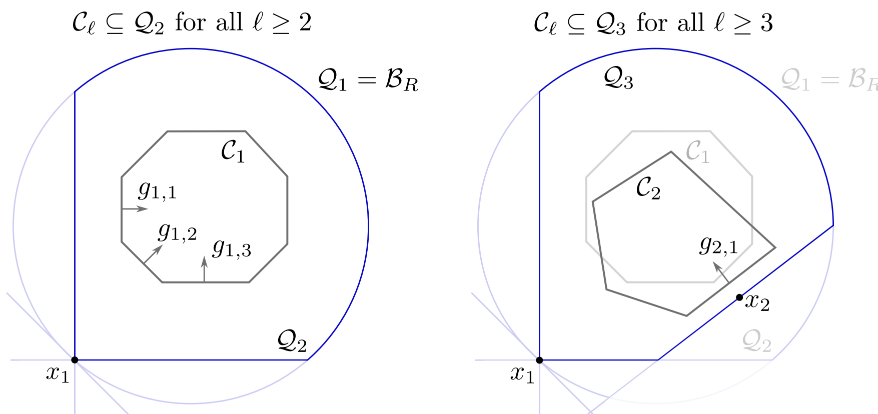

Figure S1: Illustration of polyhedral intersection.

We present here a simplified geometric setting of the Assumption 1.2 Part 2, where the time-varying constraints are linear $g_{t,i}(x) = g_{t,i}^\top x$.

In the interaction protocol of Assumption 1.2, the learner first commits a decision $x_1 \in \mathcal{B}_R$. Then the environment selects a feasible set $\mathcal{C}_1 \subseteq \mathcal{Q}_1 = \mathcal{B}_R$ (by Assumption 1.1 Part 3) and reveals to the learner a cost value $f_1(x_1)$, a gradient $\nabla f_1(x_1)$, and information for all violated constraints $(g_{1,i}(x_1), \nabla g_{1,i}(x_1))_{i=1}^3$. This constraint violation information restricts the adversary to selecting successive feasible sets $\mathcal{C}_\ell$, for all $\ell \geq 2$, from a polyhedral intersection $\mathcal{Q}_2 = \mathcal{Q}_1 \cap S_1$, where the cone intersection $S_1 = \{z \in \mathbb{R}^n : G_1(x_1)^\top (z - x_1) \geq 0\}$ and the gradient matrix $G_1(x_1) = [\nabla g_{1,i}(x_1)]_{i=1}^3$.

In the next iteration, the learner makes an update and commits a decision $x_2 = x_1 + \eta_1 v_1$. Then, the process is repeated: the environment selects $\mathcal{C}_2 \subseteq \mathcal{Q}_2$, reveals a cost value $f_2(x_2)$, a gradient $\nabla f_2(x_2)$ and constraint violation information $g_{2,1}(x_2), \nabla g_{2,1}(x_2)$. All successive feasible sets $\mathcal{C}_\ell$, for all $\ell \geq 3$, are restricted to belong to a polyhedral intersection $\mathcal{Q}_3 = \mathcal{Q}_2 \cap S_2$, where the cone intersection $S_2 = \{z \in \mathbb{R}^n : G_2(x_2)^\top (z - x_2) \geq 0\}$ and the gradient matrix $G_2(x_2) = [\nabla g_{2,1}(x_1)]$.

B Further Applications

We consider here a system identification and optimal control application where an agent must predict a sequence of actions to minimize costs and satisfy constraints. Many real-world systems are subject to wear, tear and drift (e.g., sensors), which naturally leads to non-stationary costs and constraints, corresponding to slowly time-varying functions f_t and $\{g_{t,i}\}_{i=1}^m$, respectively. Furthermore, it is common in optimal control to know analytically both the dynamics model and the cost and constraint functions, so the gradients are naturally available. Assuming access to a constraint violation oracle, the above scenario can be cast into our online problem formulation. More specifically, in each episode t, an agent ϕ parameterized by weights $\theta_t \in \mathbb{R}^n$ generates a sequence of actions $\{x_\ell\}_{\ell=1}^H$ and upon their deployment in the environment, receives a cost value $f_t(\theta_t)$, gradient $\nabla f_t(\theta_t)$ and information for all violated constraints $\{(g_{t,i}(\theta_t), \nabla g_{t,i}(\theta_t)\}_{i \in I(\theta_t)}$.

C Contrasting CVV-Pro and OGD: A Comparative Study

In this section, we compare the runtime performance and regret guarantee of the standard Online Gradient Descent (OGD) algorithm and our (CVV-Pro) algorithm in the two-player game setting (defined in Section 4). More concretely, we consider shared constraints of the form $C_x x + C_y y \leq b$. We report results from numerical simulations with decision dimension $n = 1000$, $m = 100$ shared resource constraints, capacity $b = 1.3$, $T = 2000$ iterations, and 5 independently sampled instances of the two-player game. We report below the results:

Regret: The 25th percentile of OGD has a higher regret around iteration 1400 than the function $5\sqrt{t}$ and stays above it. In contrast, CVV-Pro achieves better regret, with the 75th percentile being strictly bounded by the function $5\sqrt{t}$, see Figure S2a.

$\%$ **Constraints Violation** In each iteration, CVV-Pro requires an oracle access only to the currently violated constraints. The percentage of violated constraints first increases from 0.01% to 57% in the first four iterations, and then decreases rapidly to plateau at 20%, see Figure S2b.

Runtime: In Figure S2c, we report the average runtime per iteration for computing a projection. Since CVV-Pro solves the velocity projection problem with a decreasing number of constraints, it achieves a faster average runtime of 0.11 ± 0.01s compared to OGD, which requires solving the full projection problem each time and runs in 0.18 ± 0.01s. Thus, for the two-player game with shared constraints, our algorithm CVV-Pro achieves a runtime improvement of around 60% over OGD. Further, we report in Figure S2d the total cumulative runtime of CVV-Pro and OGD for computing the projection.

The amount of improvement in execution time is likely to be greater for higher-dimensional problems, where fewer constraints tend to be active at each iteration. Moreover, there are important situations, for example if constraints are non-convex, where projections are very difficult to compute (and/or might not even be well defined). In contrast, the velocity projection step in CCV-Pro is always a convex problem, regardless of whether the underlying feasible set is convex or not.

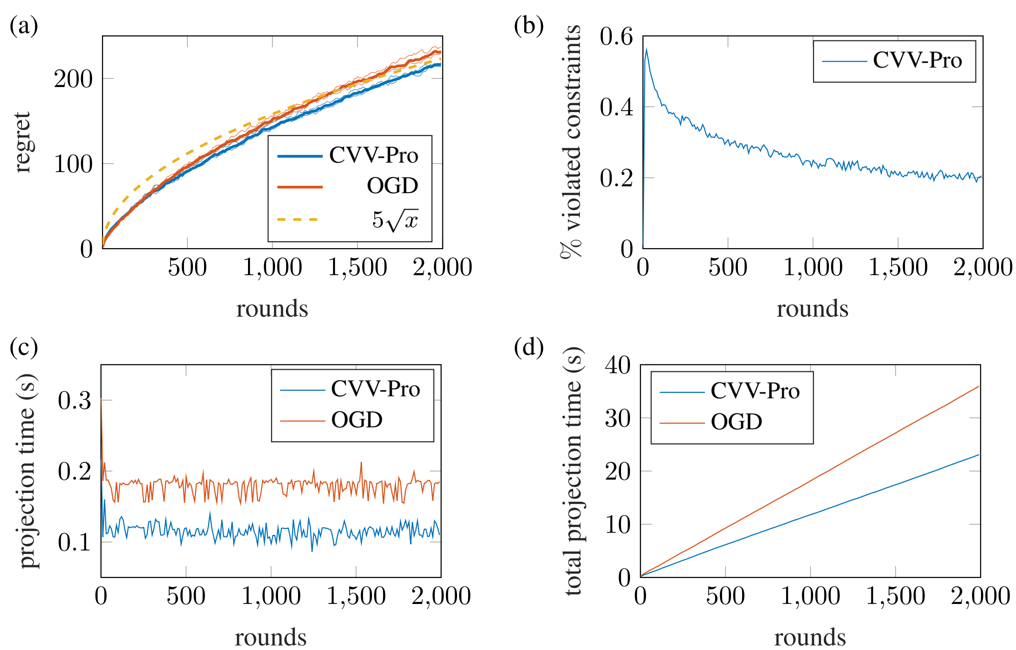

Figure S2: The figure contrasts CVV-Pro and OGD by comparing the resulting regret (a) and execution time (c,d). Panel (b) shows how the number of violated constraints evolves over time.

D Proof of Theorem 2.1

In this section, we consider an online optimization problem with time-invariant constraints and a bounded iterate assumption. The bounded iterate assumption will be removed subsequently in Section E, which however, will require a more complex analysis.

We restate Theorem 2.1 below for the convenience of the reader.

Theorem D.1 (Structural)**.** *Suppose Assumption 1.1 holds and in addition $x_t \in \mathcal{B}_R$ for all $t \in \{1, \ldots, T\}$. Then, on input $\alpha = L_{\mathcal{F}}/R$, Algorithm 1 with step sizes $\eta_t = \frac{1}{\alpha\sqrt{t}}$ guarantees the following for all $T \geq 1$:*

(regret) $\sum_{t=1}^{T} f_t(x_t) - \min_{x \in \mathcal{C}} \sum_{t=1}^{T} f_t(x) \leq 18 L_{\mathcal{F}} R \sqrt{T}$;

(feasibility) $g_i(x_t) \geq -8 \left[\frac{L_{\mathcal{G}}}{R} + 2\beta_{\mathcal{G}}\right] \frac{R^2}{\sqrt{t}}$, *for all $t \in \{1, \ldots, T\}$ and $i \in \{1, \ldots, m\}$.*

The rest of this section is devoted to proving the preceding statement.

D.1 Structural Properties

Lemma D.2. *Suppose g_i is concave for every $i \in \{1, \ldots, m\}$. Then, for any $\alpha > 0$ and all $x \in \mathcal{C}$ the following holds*

$$\max_{t \geq 0} \|v_t\| \leq \alpha \|x - x_t\| + 2\|\nabla f_t(x_t)\|.$$

In particular, when f_t satisfies $\|\nabla f_t(z)\| \leq L_{\mathcal{F}}$ for all $z \in \mathcal{B}_{cR}$, it follows that $\|v_t\| \leq (c+1)\alpha R + 2L_{\mathcal{F}}$ for any $x \in \mathcal{B}_R$, $x_t \in \mathcal{B}_{cR}$, and $c > 0$.

Proof. By Claim 2.2, we have $\alpha(x - x_t) \in V_\alpha(x_t)$ for every $x \in \mathcal{C}$. Combining the triangle inequality with the fact that v_t is an optimal solution of the velocity projection problem in Step 8, yields

$$\begin{aligned} \|v_t\| - \|\nabla f_t(x_t)\| &\leq \|v_t + \nabla f_t(x_t)\| \\ &\leq \|\alpha(x - x_t) + \nabla f_t(x_t)\| \\ &\leq \alpha\|x - x_t\| + \|\nabla f_t(x_t)\|. \end{aligned}$$

Using $x \in \mathcal{B}_R$, $x_t \in \mathcal{B}_{cR}$ and $\|\nabla f_t(x_t)\| \leq L_{\mathcal{F}}$, we conclude

$$\|v_t\| \leq \alpha\|x - x_t\| + 2\|\nabla f_t(x_t)\| \leq (c+1)\alpha R + 2L_{\mathcal{F}}.$$

□

D.2 Cost Regret

Lemma D.3. *Suppose Assumption 1.1 holds and $x_t \in \mathcal{B}_{cR}$ for all $t \in \{1, \ldots, T\}$ with $c \in (0, 4]$. Let $d \geq 0$ be a constant. Then, Algorithm 1 applied with $\alpha = L_{\mathcal{F}}/R$ and step sizes $\eta_t = \frac{1}{\alpha\sqrt{t+d}}$, guarantees the following for all $T \geq 1$:*

$$R_T = \sum_{t=1}^{T} f_t(x_t) - \min_{x^\star \in \mathcal{C}} \sum_{t=1}^{T} f_t(x^\star) \leq \sqrt{d+1}\left[(c+3)^2 + \frac{1}{2}(c+1)^2\right] L_{\mathcal{F}} R \sqrt{T}.$$

In particular, for $c = 1$ and $d = 0$ we have $R_T \leq 18 L_{\mathcal{F}} R \sqrt{T}$.

Proof. We denote an optimal decision in hindsight by $x^\star \in \arg\min_{x \in \mathcal{C}} \sum_{t=1}^{T} f_t(x)$. For any points $x^\star$, x_t we have $f_t(x_t) - f_t(x^\star) \leq [\nabla f_t(x_t)]^\top (x_t - x^\star)$, since f_t is convex. Summing over the number of rounds t results in

$$\sum_{t=1}^{T} f_t(x_t) - f_t(x^\star) \leq \sum_{t=1}^{T} [\nabla f_t(x_t)]^\top (x_t - x^\star).$$

We proceed by upper bounding the expression $[\nabla f_t(x_t)]^\top (x_t - x^\star)$. Using $x_{t+1} = x_t + \eta_t v_t$ and $v_t = r_t - \nabla f_t(x_t)$, we have

$$\begin{aligned} \|x_{t+1} - x^\star\|^2 &= \|x_t + \eta_t (r_t - \nabla f_t(x_t)) - x^\star\|^2 \\ &= \|x_t - x^\star\|^2 + \eta_t^2 \|r_t - \nabla f_t(x_t)\|^2 + 2\eta_t [r_t - \nabla f_t(x_t)]^\top (x_t - x^\star). \end{aligned}$$

Then, Lemma 2.3 gives $r_t^\top(x_t - x^\star) \le 0$ and thus

$$
\begin{aligned}
[\nabla f_t(x_t)]^\top(x_t - x^\star) &= r_t^\top(x_t - x^\star) + \frac{\|x_t - x^\star\|^2 - \|x_{t+1} - x^\star\|^2}{2\eta_t} + \frac{\eta_t}{2}\|v_t\|^2 \\
&\le \frac{\|x_t - x^\star\|^2 - \|x_{t+1} - x^\star\|^2}{2\eta_t} + \frac{\eta_t}{2}\|v_t\|^2.
\end{aligned}
$$

Since $x^\star \in \mathcal{B}_R$ and $x_t \in \mathcal{B}_{cR}$ for all $t \in \{1, \dots, T\}$, by Lemma D.2 it follows for all $t \in \{1, \dots, T\}$ that

$$
\|v_t\| \le (c+1)\alpha R + 2L_{\mathcal{F}} = (c+3)L_{\mathcal{F}} =: \mathcal{V}_\alpha. \tag{S1}
$$

Summing over the whole sequence, using the fact that $\eta_t = \frac{1}{\alpha\sqrt{t+d}}$ is a decreasing positive sequence and applying Claim D.4, $x^\star \in \mathcal{B}_R$, $x_t \in \mathcal{B}_{cR}$, and (S1), yields

$$
\begin{aligned}
2\sum_{t=1}^{T}[\nabla f_t(x_t)]^\top(x_t - x^\star) &\le \sum_{t=1}^{T} \frac{\|x_t - x^\star\|^2 - \|x_{t+1} - x^\star\|^2}{\eta_t} + \eta_t\|v_t\|^2 \\
&\le \mathcal{V}_\alpha^2\left(\sum_{t=1}^{T}\eta_t\right) + \frac{(c+1)^2R^2}{\eta_T} \\
&\le (c+3)^2L_{\mathcal{F}}^2\frac{2}{\alpha}\sqrt{T+d} + (c+1)^2L_{\mathcal{F}}R\sqrt{T+d} \\
&= \left[2(c+3)^2 + (c+1)^2\right]L_{\mathcal{F}}R\sqrt{T+d},
\end{aligned}
$$

where last inequality uses

$$
\sum_{t=1}^{T}\eta_t = \frac{1}{\alpha}\sum_{t=1}^{T}\frac{1}{\sqrt{t+d}} < \frac{1}{\alpha}\sum_{t=1}^{T+d}\frac{1}{\sqrt{t}} \le \frac{2}{\alpha}\sqrt{T+d}.
$$

The statement follows by combining the fact that $\sqrt{T+d} \le \sqrt{d+1}\sqrt{T}$ for any $d \ge 0$ and all $T \ge 1$, and

$$
\sum_{t=1}^{T} f_t(x_t) - f_t(x^\star) \le \sum_{t=1}^{T}[\nabla f_t(x_t)]^\top(x_t - x^\star) \le \sqrt{d+1}\left[(c+3)^2 + \frac{1}{2}(c+1)^2\right]L_{\mathcal{F}}R\sqrt{T}.
$$

□

Claim D.4 (Series)**.** *For any positive sequence $\{a_t\}_{t=1}^{T+1}$ and any decreasing positive sequence $\{\eta_t\}_{t=1}^{T}$, it holds that*

$$
\sum_{t=1}^{T}\frac{a_t - a_{t+1}}{\eta_t} \le \frac{A}{\eta_T}, \quad \textit{where} \quad A := \max_{t=\{1,\dots,T\}} a_t.
$$

Proof. Observe that

$$
\begin{aligned}
\sum_{t=1}^{T}\frac{a_t - a_{t+1}}{\eta_t} &= \frac{a_1 - a_2}{\eta_1} + \frac{a_2 - a_3}{\eta_2} + \frac{a_3 - a_4}{\eta_3} + \dots + \frac{a_T - a_{T+1}}{\eta_T} \\
&= \frac{a_1}{\eta_1} - \frac{a_{T+1}}{\eta_T} + \sum_{i=2}^{\top} a_i\left(\frac{1}{\eta_i} - \frac{1}{\eta_{i-1}}\right) \\
&\le \frac{A}{\eta_T},
\end{aligned}
$$

where the last inequality follows by

$$
\sum_{i=2}^{\top} a_i\left(\frac{1}{\eta_i} - \frac{1}{\eta_{i-1}}\right) \le A\sum_{i=2}^{\top}\left(\frac{1}{\eta_i} - \frac{1}{\eta_{i-1}}\right) = A\left(\frac{1}{\eta_T} - \frac{1}{\eta_1}\right) \le \frac{A}{\eta_T} - \frac{a_1}{\eta_1}.
$$

□

D.3 Convergence Rate of Constraint Violations

Lemma D.5 (Convergence Rate of Constraint Violations)**.** *Suppose Assumption 1.1 holds and* $\{x_t\}_{t\geq 1} \in \mathcal{B}_{cR}$ *with* $x_1 \in \mathcal{B}_R$ *and* $c \in (0, 4]$*. Then, for any* $\alpha > 0$ *and* $d \geq 0$*, step sizes* $\eta_t = 1/(\alpha\sqrt{t+d})$ *and* $\mathcal{V}_\alpha > 0$ *such that* $\|v_t\| \leq \mathcal{V}_\alpha$ *for all* $t \geq 1$*, it follows for every* $i \in \{1, \ldots, m\}$ *and* $t \geq 1$ *that*

$$g_i(x_t) \geq -c_1\eta_t,$$

where

$$c_1 = \mathcal{V}_\alpha \left[2L_\mathcal{G} + \frac{\beta_\mathcal{G}\mathcal{V}_\alpha}{\alpha}\right] + \mathcal{Z}_d \quad \text{and} \quad \mathcal{Z}_d = \left(1 - \frac{1}{\sqrt{d+1}}\right)\sqrt{d+2}\left[\frac{L_\mathcal{G}}{R} + \beta_\mathcal{G}\right] 2\alpha R^2.$$

In particular, when Assumption 1.1 holds, $\{x_t\}_{t\geq 1} \in \mathcal{B}_R$*,* $\alpha = L_\mathcal{F}/R$ *and* $d = 0$*, it follows that*

$$g_i(x_t) \geq -8\left[\frac{L_\mathcal{G}}{R} + 2\beta_\mathcal{G}\right]\frac{R^2}{\sqrt{t}} \quad \text{for all } t \geq 1.$$

Proof. The proof is by induction on t. We start with the base case $t = 1$. The proof proceeds by case distinction.

Case 1. Suppose $i \in \{1, \ldots, m\}\backslash I(x_1)$, i.e., $g_i(x_1) > 0$. Then, by Claim D.6 Part ii) we have

$$g_i(x_2) \geq -\eta_2\mathcal{V}_\alpha\left[2L_\mathcal{G} + \frac{\mathcal{V}_\alpha\beta_\mathcal{G}}{\alpha\sqrt{1+d}}\right] \geq -c_1\eta_2.$$

Case 2. Suppose $i \in I(x_1)$, i.e., $g_i(x_1) \leq 0$. By combining $x_1 \in \mathcal{B}_R$ and g_i is concave $\beta_\mathcal{G}$-smooth, it follows for every $x \in \mathcal{C} \subseteq \mathcal{B}_R$ that

$$\begin{aligned}
g_i(x_1) &\geq g_i(x) + \nabla g_i(x)^T(x_1 - x) - \frac{\beta_\mathcal{G}}{2}\|x_1 - x\|^2 \\
&\geq -2L_\mathcal{G}R - 2\beta_\mathcal{G}R^2 \\
&= -\eta_1\sqrt{d+1}\left[\frac{L_\mathcal{G}}{R} + \beta_\mathcal{G}\right] 2\alpha R^2 \geq -c_1\eta_2.
\end{aligned}$$

Using $\eta_t = 1/(\alpha\sqrt{t+d})$, $\eta_1/\eta_2 \leq \sqrt{2}$ and $\eta_1^2\frac{\mathcal{V}_\alpha^2\beta_\mathcal{G}}{2} \leq \eta_2^2\mathcal{V}_\alpha^2\beta_\mathcal{G} = \eta_2\frac{\mathcal{V}_\alpha^2\beta_\mathcal{G}}{\alpha\sqrt{d+2}}$, it follows by Claim D.6 Part i) that

$$\begin{aligned}
g_i(x_2) &\geq (1 - \alpha\eta_1)g_i(x_1) - \eta_1^2\frac{\mathcal{V}_\alpha^2\beta_\mathcal{G}}{2} \\
&\geq -\eta_2\left[\left(1 - \frac{1}{\sqrt{d+1}}\right)\sqrt{d+2}\left[\frac{L_\mathcal{G}}{R} + \beta_\mathcal{G}\right]2\alpha R^2 + \frac{\mathcal{V}_\alpha^2\beta_\mathcal{G}}{\alpha\sqrt{d+2}}\right] \geq -c_1\eta_2.
\end{aligned}$$

Our inductive hypothesis is $g_i(x_t) \geq -c_1\eta_t$ for all i. We now show that it holds for $t + 1$.

Case 1. Suppose $i \in \{1, \ldots, m\}\backslash I(x_1)$, i.e., $g_i(x_t) > 0$. Then by Claim D.6 Part ii)

$$g_i(x_{t+1}) \geq -\eta_{t+1}\mathcal{V}_\alpha\left[2L_\mathcal{G} + \frac{\beta_\mathcal{G}\mathcal{V}_\alpha}{\alpha\sqrt{d+1}}\right] \geq -c_1\eta_{t+1}.$$

Case 2. Suppose $i \in I(x_t)$, i.e., $g_i(x_t) \leq 0$. Combining Claim D.6 Part ii) and the inductive hypothesis we have

$$\begin{aligned}
g_i(x_{t+1}) &\geq (1 - \alpha\eta_t)g_i(x_t) - \eta_t^2\frac{\mathcal{V}_\alpha^2\beta_\mathcal{G}}{2} \\
&\geq -c_1\eta_t + c_1\alpha\eta_t^2 - \eta_t^2\frac{\mathcal{V}_\alpha^2\beta_\mathcal{G}}{2} \\
&= -c_1\eta_{t+1} + c_1\eta_{t+1} - c_1\eta_t + c_1\alpha\eta_t^2 - \eta_t^2\frac{\mathcal{V}_\alpha^2\beta_\mathcal{G}}{2} \\
&= -c_1\eta_{t+1} + c_1\eta_t\left[\frac{\eta_{t+1}}{\eta_t} - 1 + \alpha\eta_t - \eta_t\frac{\mathcal{V}_\alpha^2\beta_\mathcal{G}}{2c_1}\right].
\end{aligned}$$

Since $c_1\eta_t > 0$, it suffices to show that

$$\alpha - \frac{\eta_t - \eta_{t+1}}{\eta_t^2} \geq \frac{\mathcal{V}_\alpha^2 \beta_\mathcal{G}}{c_1} \tag{S2}$$

or equivalently (using $\eta_t = \frac{1}{\alpha\sqrt{t+d}}$ for $t \geq 1$)

$$\alpha - \alpha\sqrt{\frac{t+d}{t+d+1}}\left(\sqrt{t+d+1} - \sqrt{t+d}\right) \geq \frac{\mathcal{V}_\alpha^2 \beta_\mathcal{G}}{2c_1}.$$

Straightforward checking shows that $\max_{t\geq 1}\sqrt{\frac{t}{t+1}}\left(\sqrt{t+1} - \sqrt{t}\right) < \frac{1}{3}$. Hence, inequality (S2) is implied for $c_1 \geq \beta_\mathcal{G}\mathcal{V}_\alpha^2/\alpha$ and thus $g_i(x_{t+1}) \geq -c_1\eta_{t+1}$.

Furthermore, for $c = 1$ and $\alpha = L_\mathcal{F}/R$, by Lemma D.2, we can set $\mathcal{V}_\alpha = 4L_\mathcal{F}$. Then, for $d = 0$ we have $g_i(x_t) \geq -8\left[\frac{L_\mathcal{G}}{R} + 2\beta_\mathcal{G}\right]\frac{R^2}{\sqrt{t}}$ for all $t \geq 1$. □

Claim D.6 (Constraint Violation)**.** *Suppose g_i is concave, $\beta_\mathcal{G}$-smooth and satisfies $\|\nabla g_i(x)\| \leq L_\mathcal{G}$ for all $x \in \mathcal{B}_{cR}$ and $i \in \{1, \ldots, m\}$, where $c > 0$ is a constant. Suppose further that there exists a constant $\mathcal{V}_\alpha > 0$ such that $x_t \in \mathcal{B}_{cR}$ and $\|v_t\| \leq \mathcal{V}_\alpha$, for all $t \geq 1$. Then, for all $t \geq 1$ we have*

i) $g_i(x_{t+1}) \geq (1 - \alpha\eta_t)g_i(x_t) - \eta_t^2\mathcal{V}_\alpha^2\beta_\mathcal{G}/2$ for every $i \in I(x_t)$;

ii) $g_i(x_{t+1}) \geq -\eta_{t+1}\mathcal{V}_\alpha\left[2L_\mathcal{G} + \mathcal{V}_\alpha\beta_\mathcal{G}/(\alpha\sqrt{1+d})\right]$ for every $i \in \{1, \ldots, m\}\backslash I(x_t)$.

Proof. The proof proceeds by case distinction.

Case 1. Suppose $i \in I(x_t)$, i.e., $g_i(x_t) \leq 0$. By combining the facts that g_i is concave and $\beta_\mathcal{G}$-smooth, $x_{t+1} = x_t + \eta_t v_t$ and $[\nabla g_i(x_t)]^\top v_t \geq -\alpha g_i(x_t)$, it follows that

$$\begin{aligned} g_i(x_{t+1}) &\geq g_i(x_t) + [\nabla g_i(x_t)]^\top[x_{t+1} - x_t] - \frac{\beta_\mathcal{G}}{2}\|x_{t+1} - x_t\|_2^2 \\ &\geq (1 - \alpha\eta_t)g_i(x_t) - \eta_t^2\mathcal{V}_\alpha^2\frac{\beta_\mathcal{G}}{2}. \end{aligned} \tag{S3}$$

Case 2. Suppose $i \notin I(x_t)$, i.e., $g_i(x_t) > 0$. Using $\|\nabla g_i(x)\| \leq L_\mathcal{G}$ for $x_t \in \mathcal{B}_{cR}$, we have

$$[\nabla g_i(x_t)]^\top[x_{t+1} - x_t] \leq \|\nabla g_i(x_t)\|\|x_{t+1} - x_t\| \leq \eta_t L_\mathcal{G}\mathcal{V}_\alpha.$$

Hence,

$$\begin{aligned} g_i(x_{t+1}) &\geq g_i(x_t) + [\nabla g_i(x_t)]^\top[x_{t+1} - x_t] - \frac{\beta_\mathcal{G}}{2}\|x_{t+1} - x_t\|_2^2 \\ &\geq -\eta_t L_\mathcal{G}\mathcal{V}_\alpha - \eta_t^2\mathcal{V}_\alpha^2\frac{\beta_\mathcal{G}}{2} \\ &= -\eta_{t+1}\frac{\eta_t}{\eta_{t+1}}\mathcal{V}_\alpha\left[L_\mathcal{G} + \frac{\eta_t}{2}\mathcal{V}_\alpha\beta_\mathcal{G}\right] \\ &> -\eta_{t+1}\mathcal{V}_\alpha\left[2L_\mathcal{G} + \frac{\mathcal{V}_\alpha\beta_\mathcal{G}}{\alpha\sqrt{1+d}}\right], \end{aligned}$$

where the last inequality follows by $\eta_t \leq \eta_1 = 1/(\alpha\sqrt{1+d})$ and

$$\max_{\ell\geq 1}\frac{\eta_\ell}{\eta_{\ell+1}} \leq \max_{\ell\geq 1}\sqrt{\frac{\ell+1}{\ell}} = \sqrt{2}.$$

□

E Guaranteeing a Bounded Decision Sequence

We now show that the second assumption in Theorem 2.1, namely, "$x_t \in \mathcal{B}_R$ for all $t \in \{1, \ldots, T\}$" can be enforced algorithmically. We achieve this by introducing an additional hypersphere constraint $g_{m+1}(x_t) = \frac{1}{2}[R^2 - \|x_t\|^2]$ that attracts the decision sequence $\{x_t\}_{t \geq 1}$ to a hypersphere $\mathcal{B}_R$ and guarantees that it always stays inside a hypersphere $\mathcal{B}_{4R}$ with a slightly larger radius. Technically, we modify the velocity polyhedron in Step 3 of Algorithm 1 as follows: $V'_\alpha(x_t) = V_\alpha(x_t)$ if $\|x\| \leq R$, and otherwise

$$V'_\alpha(x_t) = \{v \in V_\alpha(x_t) \mid [\nabla g_{m+1}(x_t)]^\top v \geq -\alpha g_{m+1}(x_t)\}.$$

We are now ready to state our main algorithmic result for the setting of time-invariant constraints.

Theorem E.1 (Algorithm)**.** *Suppose Assumption 1.1 holds. Then, on input $R, L_\mathcal{F} > 0$, $\alpha = L_\mathcal{F}/R$ and $x_1 \in \mathcal{B}_R$, Algorithm 1 with augmented velocity polyhedron $V'_\alpha(\cdot)$ and step sizes $\eta_t = \frac{1}{\alpha\sqrt{t+15}}$ guarantees the following for all $T \geq 1$:*

- ***(regret)*** $\sum_{t=1}^T f_t(x_t) - \min_{x^\star \in \mathcal{C}} \sum_{t=1}^T f_t(x^\star) \leq 246 L_\mathcal{F} R\sqrt{T}$;
- ***(feasibility)*** $g_i(x_t) \geq -21\left[\frac{L_\mathcal{G}}{R} + 3\beta_G\right]\frac{R^2}{\sqrt{t+15}}$, *for all $t \in \{1, \ldots, T\}$ and $i \in \{1, \ldots, m\}$;*
- ***(attraction)*** $g_{m+1}(x_t) \geq -27\frac{R^2}{\sqrt{t+15}}$ *for all $t \in \{1, \ldots, T\}$.*

In addition, $\|x_t\| \leq 4R$ and $\|v_t\| \leq 7L_\mathcal{F}$, for all $t \geq 1$.

To ensure convergence of the hypersphere constraint $-\min\{g_{m+1}(x_t), 0\}$ at a rate of $\mathcal{O}(1/\sqrt{t})$, we use an inductive argument similar to Lemma D.5. We note that compared to the simplified setting of Appendix D, our analysis requires an additional refined inductive argument, which is summarized in Lemma E.5.

E.1 Hypersphere constraint

Definition E.2. We consider the following hypersphere constraint, parameterized by $R > 0$,

$$g_{m+1}(x) = \frac{1}{2}[R^2 - \|x_t\|^2].$$

By construction, g_{m+1} is concave and 1-smooth.

Claim E.3. *Suppose g_i is concave for every $i \in \{1, \ldots, m\}$ such that $\mathcal{C} \subseteq \mathcal{B}_R$ and f_t is convex such that $\|\nabla f_t(x)\| \leq L_\mathcal{F}$ for all $x \in \mathcal{B}_{cR}$, where $c > 0$ is a constant. Then for any decision $x_t \in \mathcal{B}_{cR}$, it holds that*

$$\|v_t\| \leq \alpha\|x_t\| + (\alpha R + 2L_\mathcal{F}) \qquad \text{and} \qquad \frac{1}{2}\|v_t\|^2 < -2\alpha^2 g_{m+1}(x) + \left[\alpha^2 R^2 + (\alpha R + 2L_\mathcal{F})^2\right].$$

Proof. Due to the fact that g_{m+1} and g_i are concave for every $i \in \{1, \ldots, m\}$, it follows by Lemma D.2 that

$$\begin{aligned} \|v_t\| &\leq 2\|\nabla f_t(x_t)\| + \alpha\|x^\star - x_t\| \\ &\leq \alpha\|x_t\| + \alpha R + 2L_\mathcal{F}. \end{aligned}$$

Further, by definition of $g_{m+1}(x)$ we have

$$\begin{aligned} \frac{1}{2}\|v_t\|^2 &\leq \frac{1}{2}\left[\alpha\|x_t\| + (\alpha R + 2L_\mathcal{F})\right]^2 \\ &\leq \alpha^2\|x_t\|^2 + (\alpha R + 2L_\mathcal{F})^2 \\ &= -2\alpha^2 g_{m+1}(x) + \left[\alpha^2 R^2 + (\alpha R + 2L_\mathcal{F})^2\right]. \end{aligned}$$

□

Claim E.4. *Suppose the assertions in Claim E.3 hold. Let the step sizes be $\{\eta_t = \frac{1}{\alpha\sqrt{t+15}}\}_{t \geq 1}$ and $\alpha = L_\mathcal{F}/R$. Then, we have*

i) If $g_{m+1}(x_t) > 0$ then $g_{m+1}(x_{t+1}) \geq -\eta_t \cdot 6L_\mathcal{F}R$; and

ii) If $g_{m+1}(x_t) \leq 0$ then $g_{m+1}(x_{t+1}) \geq \left(1 - \frac{\alpha}{2}\eta_t\right)g_{m+1}(x_t) - \eta_t^2 10L_\mathcal{F}^2$.

Proof. The proof is by case distinction.

Case 1. Suppose $g_{m+1}(x_t) > 0$. Using $\|x_t\| < R$ it follows by Claim E.3 that

$$\|v_t\| \leq 2\left(\alpha R + L_{\mathcal{F}}\right) = 4L_{\mathcal{F}}.$$

Using g_{m+1} is concave and 1-smooth, $g_{m+1}(x_t) > 0$, $\nabla g_{m+1}(x_t) = -x_t$ and $\|x_t\| < R$, we have

$$\begin{aligned} g_{m+1}(x_{t+1}) &\geq g_{m+1}(x_t) + \nabla g_{m+1}(x_t)^\top (x_{t+1} - x_t) - \frac{1}{2}\|x_{t+1} - x_t\|^2 \\ &\geq -\eta_t R\|v_t\| - \frac{1}{2}\eta_t^2\|v_t\|^2 \\ &\geq -\eta_t \cdot 6L_{\mathcal{F}}R \\ &\geq -\eta_{t+1} \cdot 7L_{\mathcal{F}}R, \end{aligned}$$

where we used

$$\frac{1}{2}\eta_t 16L_{\mathcal{F}}^2 = \frac{8}{\sqrt{t+15}}L_{\mathcal{F}}R \leq 2L_{\mathcal{F}}R.$$

Case 2. Suppose $g_{m+1}(x_t) \leq 0$, i.e., $\|x_t\| \geq R$. Using $\alpha^2 R^2 + (\alpha R + 2L_{\mathcal{F}})^2 = 10L_{\mathcal{F}}^2$, it follows by Claim E.3 that

$$\frac{1}{2}\|v_t\|^2 < -2\alpha^2 g_{m+1}(x) + 10L_{\mathcal{F}}^2.$$

Combining g_{m+1} is concave and 1-smooth, and $\nabla g_{m+1}(x_t)^\top v_t \geq -\alpha g_{m+1}(x_t)$ yields

$$\begin{aligned} g_{m+1}(x_{t+1}) &\geq g_{m+1}(x_t) + \nabla g_{m+1}(x_t)^\top (x_{t+1} - x_t) - \frac{1}{2}\|x_{t+1} - x_t\|^2 \\ &\geq (1 - \alpha\eta_t) g_{m+1}(x_t) - \frac{1}{2}\eta_t^2\|v_t\|^2 \\ &> \left(1 - \alpha\eta_t + 2\alpha^2\eta_t^2\right) g_{m+1}(x_t) - \eta_t^2 10L_{\mathcal{F}}^2 \\ &\geq \left(1 - \frac{\alpha}{2}\eta_t\right) g_{m+1}(x_t) - \eta_t^2 10L_{\mathcal{F}}^2, \end{aligned}$$

where the last inequality follows by: $-\eta_t\alpha + 2\eta_t^2\alpha^2 \leq -\eta_t\frac{\alpha}{2}$, which is implied by $\eta_t = \frac{1}{\alpha\sqrt{t+15}}$. □

Lemma E.5 (Main). *Suppose the assertions in Claim E.3 hold for $c = 4$. Given $\alpha = L_{\mathcal{F}}/R$, step sizes $\{\eta_t = \frac{1}{\alpha\sqrt{t+15}}\}_{t\geq 1}$ and an arbitrary initial decision x_1 with $\|x_1\| < R$, then it holds that*

$$g_{m+1}(x_t) \geq -27\frac{R^2}{\sqrt{t}}, \qquad \|x_t\| \leq 4R, \qquad \|v_t\| \leq 7L_{\mathcal{F}}, \qquad \textit{for all } t \geq 1. \tag{S4}$$

Proof. The proof is by induction on $t \geq 1$.

Part I) We show first that $g_{m+1}(x_{t+1}) \geq -c_0\eta_{t+1}$, for some $c_0 > 0$. The proof proceeds by case distinction.

Case 1. Suppose $g_{m+1}(x_t) > 0$, then by Claim E.4 we have

$$g_{m+1}(x_{t+1}) \geq -\eta_t \cdot 6L_{\mathcal{F}}R, \qquad (\text{implying } c_0 \geq 6L_{\mathcal{F}}R).$$

Case 2. Suppose $g_{m+1}(x_t) \leq 0$. Let $A := 10L_{\mathcal{F}}^2$, then by combining Claim E.4 and the inductive hypothesis, we have

$$\begin{aligned} g_{m+1}(x_{t+1}) &\geq \left(1 - \eta_t\frac{\alpha}{2}\right) g_{m+1}(x_t) - \eta_t^2 A \\ &\geq -\left(1 - \eta_t\frac{\alpha}{2}\right) c_0\eta_t - \eta_t^2 A \\ &= -c_0\eta_t - \left(A - \frac{\alpha}{2}c_0\right)\eta_t^2 \\ &= -c_0\eta_{t+1} - c_0\eta_t + c_0\eta_{t+1} - \left(A - \frac{\alpha}{2}c_0\right)\eta_t^2 \\ &= -c_0\eta_{t+1} + c_0\eta_t\left[-1 + \frac{\eta_{t+1}}{\eta_t} - \eta_t\left(\frac{A}{c_0} - \frac{\alpha}{2}\right)\right]. \end{aligned}$$

Since $c_0\eta_t > 0$, it suffices to show that

$$-1 + \frac{\eta_{t+1}}{\eta_t} - \eta_t\left(\frac{A}{c_0} - \frac{\alpha}{2}\right) \geq 0 \iff \frac{\alpha}{2} - \frac{\eta_t - \eta_{t+1}}{\eta_t^2} \geq \frac{A}{c_0}.$$

The previous condition is equivalent to (using $\eta_t = \frac{1}{\alpha\sqrt{t+15}}$ for $t \geq 1$)

$$\alpha\left[\frac{1}{2} - \frac{\sqrt{t+15}}{\sqrt{t+16}}\left[\sqrt{t+16} - \sqrt{t+15}\right]\right] \geq \frac{A}{c_0}.$$

Straightforward checking shows that $\max_{t\geq 16}\sqrt{\frac{t}{t+1}}\left(\sqrt{t+1} - \sqrt{t}\right) < 0.12$ and thus

$$c_0 \geq 2.7\frac{A}{\alpha} = 27L_{\mathcal{F}}R.$$

Hence, for $c_0 = 27L_{\mathcal{F}}R$ it holds that $g_{m+1}(x_{t+1}) \geq -c_0\eta_{t+1}$. We set c_0 to the maximum over the preceding two case, i.e.,

$$c_0 := \max\left\{7L_{\mathcal{F}}R,\ 27L_{\mathcal{F}}R\right\},$$

and obtain

$$g_{m+1}(x_t) \geq -c_0\eta_t = -\frac{27R^2}{\sqrt{t+15}}.$$

Part II) We now show that $\|x_{t+1}\| \leq 4R$. Combining **Part I)** and the definition of step size $\eta_t = \frac{1}{\alpha\sqrt{t+15}}$, we have

$$\frac{1}{2}\left[R^2 - \|x_{t+1}\|^2\right] = g_{m+1}(x_{t+1}) \geq -c_0\eta_{t+1} \geq -c_0\eta_1 = -\frac{c_0}{4\alpha}$$

and thus

$$\|x_{t+1}\|^2 \leq R^2 + \frac{c_0}{2\alpha} < 15R^2 < (4R)^2.$$

Part III) By Claim E.3, it follows that

$$\|v_{t+1}\| \leq \frac{L_{\mathcal{F}}}{R}\|x_{t+1}\| + 3L_{\mathcal{F}} < 7L_{\mathcal{F}}.$$

$\square$

E.2 Concluding Remarks

By Lemma E.5, the decision sequence $\{x_t\}_{t\geq 1}$ is attracted to the hypersphere $\mathcal{B}_R$ and always stays inside a slightly larger hypersphere $\mathcal{B}_{4R}$.

Then, by Lemma D.3 applied with $c = 4$, $d = 15$, $\alpha = L_{\mathcal{F}}/R$ and step size $\eta_t = 1/(\alpha\sqrt{t+d})$ we obtain

$$\begin{aligned}\text{Regret}_T &\leq \sqrt{15+1}\left[(4+3)^2 + \frac{1}{2}(4+1)^2\right]L_{\mathcal{F}}R\sqrt{T} \\ &= 246L_{\mathcal{F}}R\sqrt{T}.\end{aligned}$$

Moreover, by Lemma D.5, we have $\mathcal{Z}_d = \frac{3}{2}\sqrt{17}\left[\frac{L_{\mathcal{G}}}{R} + \beta_{\mathcal{G}}\right]L_{\mathcal{F}}R$ and

$$c_1 = \mathcal{V}_\alpha\left[2L_{\mathcal{G}} + \frac{\beta_G\mathcal{V}_\alpha}{\alpha}\right] + \mathcal{Z}_d \leq 21\left[\frac{L_{\mathcal{G}}}{R} + 3\beta_G\right]L_{\mathcal{F}}R.$$

Hence, the convergence rate to the feasible $\mathcal{C}$ satisfies for every $t \geq 1$ and $i \in \{1,\ldots,m\}$

$$g_i(x_t) \geq -c_1\eta_t \geq -21\left[\frac{L_{\mathcal{G}}}{R} + 3\beta_G\right]\frac{R^2}{\sqrt{t+15}}.$$

F Proof of Theorem 3.3

In this section, we consider an online optimization problem with adversarially generated time-varying constraints. More precisely, at each time step t, the learner receives partial information on the current cost f_t and feasible set $\mathcal{C}_t$, and seeks to minimize (1). To make this problem well posed, we restrict the environment such that each feasible set $\mathcal{C}_t$ is contained in $\mathcal{Q}_t$ (see Section 1) and the rate of change between consecutive time-varying constraints *decreases* over time. We quantify a sufficient rate of decay in Assumption 3.1, which we restate below for the convenience of the reader.

Assumption F.1 (TVC Decay Rate). We assume that the adversarially generated sequence $\{g_t\}_{t\geq 1}$ of time-varying constraints are such that for every $x \in \mathcal{B}_{4R}$ and all $t \geq 1$, the following holds $\|g_{t+1}(x) - g_t(x)\|_\infty \leq \frac{98}{t+16}\left[\frac{L_\mathcal{G}}{R} + 3\beta_\mathcal{G}\right]R^2$.

We note that Assumption F.1 essentially only requires $\|g_{t+1}(x) - g_t(x)\|_\infty \leq \mathcal{O}(1/t)$, as R can be chosen large enough such that the bound is satisfied. Of course, R will appear in our regret and feasibility bounds, but it will not affect the dependence on t or T (up to constant factors).

We restate Theorem 3.3 below for the convenience of the reader.

Theorem F.2 (Time-Varying Constraints). *Suppose the functions $\{f_t, g_t\}_{t\geq 1}$ satisfy Assumptions 1.1, 1.2 and F.1. Then, on input $R, L_\mathcal{F} > 0$ and $x_1 \in \mathcal{B}_R$, Algorithm 1 applied with $\alpha = L_\mathcal{F}/R$, augmented velocity polyhedron $V'_\alpha(\cdot)$ and step sizes $\eta_t = \frac{1}{\alpha\sqrt{t+15}}$ guarantees the following for all $T \geq 1$:*

(regret) $\sum_{t=1}^{T} f_t(x_t) - \min_{x^\star \in \mathcal{C}} \sum_{t=1}^{T} f_t(x^\star) \leq 246 L_\mathcal{F} R\sqrt{T}$;

(feasibility) $g_{t,i}(x_t) \geq -265\left[\frac{L_\mathcal{G}}{R} + 4\beta_\mathcal{G}\right]\frac{R^2}{\sqrt{t+15}}$, *for all $t \in \{1,\ldots,T\}$ and $i \in \{1,\ldots,m\}$;*

(attraction) $g_{m+1}(x_t) \geq -27\frac{R^2}{\sqrt{t+15}}$ *for all $t \in \{1,\ldots,T\}$.*

Outline This section is organized as follows. In Subsection F.1, we introduce a key geometric property that allows us to generalize the standard online gradient descent analysis to the setting of time-varying constraints. In Subsection F.2, we give an overview of our proof approach for Theorem 3.3. In Subsection F.3, we present the analysis that quantifies the convergence rate to the feasible set for the setting of slowly time-varying constraints. Finally, in Subsection F.4, we give an important special case, slightly generalizing Lemma 3.2, for which Assumption F.1 is satisfied.

F.1 Key Geometric Property

Our regret analysis builds upon the following key geometric property that generalizes Lemma 2.3 to time-varying constraints. We show that for any subset $\mathcal{C}_T$ of the polyhedral intersection $\mathcal{Q}_T$, every decision $x \in \mathcal{C}_T$ satisfies the normal cone constraint $-r_t^\top(x - x_t) \leq 0$, for every pair (x_t, r_t) in the decision sequence $\{(x_t, r_t)\}_{t=1}^{T}$ up to step T. As a result, a similar argument as in (2) yields $\mathcal{O}(\sqrt{T})$ regret in the time-varying constraint setting.

Lemma F.3 (Polyhedral Intersection). *Let $\mathcal{C}_T$ be any subset of the polyhedral intersection $\mathcal{Q}_T$. Then, every decision $x \in \mathcal{C}_T$ satisfies the normal cone constraint $-r_t^\top(x - x_t) \leq 0$, $\forall t \in \{1,\ldots,T\}$.*

Proof. Using $\mathcal{S}_0 = \mathbb{R}^n$, $\mathcal{C}_T$ is contained in $\cap_{t=1}^{T-1}\{x \in \mathbb{R}^n \mid G(x_t)^\top(x - x_t) \geq 0\}$. Since $x \in \mathcal{C}_T$, it follows by Lemma 2.3 that $r_T^\top(x - x_T) \leq 0$. The proof proceeds by case distinction. Let $t \in \{1,\ldots,T-1\}$ be arbitrary. Suppose $x_t \in \mathcal{C}_t$, then by Part 1 in the proof of Lemma 2.3 we have $r_t = 0$. Suppose $x_t \notin \mathcal{C}_t$, then $x \in \mathcal{C}_T$ implies $G(x_t)^\top(x - x_t) \geq 0$ or equivalently $\nabla g_{t,i}(x_t)^\top(x - x_t) \geq 0$ for all $i \in I(x_t)$. Since $v_t \in V_\alpha(x_t)$, it follows that $v(x) = v_t + x - x_t \in V_\alpha(x_t)$. Moreover, the vector $-r_t$ belongs to the normal cone $N_{V_\alpha(x_t)}(v_t)$, which implies $-r_t^\top(v - v_t) \leq 0$ for all $v \in V_\alpha(x_t)$. In particular, for $v(x)$ we have $-r_t^\top(x - x_t) \leq 0$. $\square$

F.2 Proof Overview of Theorem 3.3

By Assumption 1.1, the slowly time-varying constraints $g_{t,i}(x)$ are concave and $\beta_{\mathcal{G}}$-smooth such that $\|\nabla g_{t,i}(x)\| \leq L_{\mathcal{G}}$ for all $x \in \mathcal{B}_{4R}$, $t \geq 1$ and $i \in \{1, \ldots, m\}$. By construction, see Lemma E.5, $\eta_t = 1/(\alpha\sqrt{t+15})$, $\alpha = L_{\mathcal{F}}/R$ and $\mathcal{V}_\alpha = 7L_{\mathcal{F}}$ implies that $\eta_{t+1}\mathcal{V}_\alpha = 7R/\sqrt{t+16}$. We note that Lemma E.5 still holds for time-varying constraints, which implies $\|x_t\| \leq 4R$ and $\|v_t\| \leq 7L_{\mathcal{F}}$.

Further, by Assumption F.1 we have for every $x \in \mathcal{B}_{4R}$, $t \geq 1$ and $i \in \{1, \ldots, m\}$ that

$$|g_{t+1,i}(x) - g_{t,i}(x)| \leq \frac{98}{t+16}\left[\frac{L_{\mathcal{G}}}{R} + 3\beta_{\mathcal{G}}\right] R^2 = 2\eta_{t+1}^2 \left[\frac{L_{\mathcal{G}}}{R} + 3\beta_{\mathcal{G}}\right] \mathcal{V}_\alpha^2. \tag{S5}$$

Then, applying the preceding inequality and using similar arguments as in Part 2) of Section 2.5, we give in Corollary F.5 bounds on the slowly time-varying constraints $g_{t,i}(x)$ from below. In particular, we show that

$$g_{t+1,i}(x_{t+1}) \geq (1 - \alpha\eta_t) g_{t,i}(x_t) - \eta_t^2 \left[2\frac{L_{\mathcal{G}}}{R} + 7\beta_{\mathcal{G}}\right] \mathcal{V}_\alpha^2 \qquad \text{for all } i \in I(x_t),$$

and

$$g_{t+1,i}(x_{t+1}) \geq -\eta_{t+1} 7\mathcal{V}_\alpha \left[L_{\mathcal{G}} + \frac{\beta_{\mathcal{G}}\mathcal{V}_\alpha}{4\alpha}\right] \qquad \text{for all } i \in \{1, \ldots, m\} \backslash I(x_t).$$

Using a similar inductive argument as in Lemma D.5, we show in Lemma F.4 that in the setting of slowly time-varying constraints, the following feasibility convergence rate holds

$$g_{t,i}(x_t) \geq -\left[265\frac{L_{\mathcal{G}}}{R} + 927\beta_{\mathcal{G}}\right] \frac{R^2}{\sqrt{t+15}}, \qquad \text{for all } t \in \{1, \ldots, T\} \text{ and } i \in \{1, \ldots, m\}.$$

Then, the regret and the attraction to the feasible sets follow as in Theorem E.1.

F.3 Slowly Time-Varying Constraints

Lemma F.4 (Slowly TVC). *Suppose Assumption 1.1 holds, $x_1 \in \mathcal{B}_R$, $\alpha = L_{\mathcal{F}}/R$ and step sizes $\eta_t = 1/(\alpha\sqrt{t+15})$. Then, for every $i \in \{1, \ldots, m\}$ and $T \geq 1$ we have*

$$g_{t,i}(x_t) \geq -\left[265\frac{L_{\mathcal{G}}}{R} + 927\beta_{\mathcal{G}}\right] \frac{R^2}{\sqrt{t+15}}, \qquad \textit{for all } t \in \{1, \ldots, T\} \textit{ and } i \in \{1, \ldots, m\}.$$

Proof. The proof is by induction on t. We start with the base case $t = 1$. The proof proceeds by case distinction.

Case 1. Suppose $i \in \{1, \ldots, m\} \backslash I(x_1)$, i.e., $g_{1,i}(x_1) > 0$. Then, by Corollary F.5 Part ii) we have

$$g_{2,i}(x_2) \geq -\eta_2 7\mathcal{V}_\alpha \left[L_{\mathcal{G}} + \frac{\beta_{\mathcal{G}}\mathcal{V}_\alpha}{4\alpha}\right] \geq -\eta_2 \left[49\frac{L_{\mathcal{G}}}{R} + 86\beta_{\mathcal{G}}\right] L_{\mathcal{F}} R.$$

Case 2. Suppose $i \in I(x_1)$, i.e., $g_{1,i}(x_1) \leq 0$. By combining $x_1 \in \mathcal{B}_R$ and $g_{1,i}$ is concave $\beta_{\mathcal{G}}$-smooth, it follows for every $x \in \mathcal{C}_1 \subseteq \mathcal{B}_R$ that

$$\begin{aligned} g_{1,i}(x_1) &\geq g_{1,i}(x) + \nabla g_{1,i}(x)^T (x_1 - x) - \frac{\beta_{\mathcal{G}}}{2}\|x_1 - x\|^2 \\ &\geq -2L_{\mathcal{G}}R - 2\beta_{\mathcal{G}}R^2 \\ &= -\eta_1 \left[8\frac{L_{\mathcal{G}}}{R} + 8\beta_{\mathcal{G}}\right] L_{\mathcal{F}} R. \end{aligned}$$

Using $\eta_t = 1/(\alpha\sqrt{t+15})$ and $\eta_1/\eta_2 \leq \sqrt{2}$, it follows that

$$(1 - \alpha\eta_1) g_{1,i}(x_1) \geq -\eta_1 \left[\frac{L_{\mathcal{G}}}{R} + \beta_{\mathcal{G}}\right] 6L_{\mathcal{F}}R \geq -\eta_2 \left[9\frac{L_{\mathcal{G}}}{R} + 9\beta_{\mathcal{G}}\right] L_{\mathcal{F}}R$$

and

$$\eta_1^2 \left[2\frac{L_{\mathcal{G}}}{R} + 7\beta_{\mathcal{G}}\right] \mathcal{V}_\alpha^2 \leq \eta_2^2 \left[4\frac{L_{\mathcal{G}}}{R} + 14\beta_{\mathcal{G}}\right] \mathcal{V}_\alpha^2 \leq \eta_2 \left[49\frac{L_{\mathcal{G}}}{R} + 172\beta_{\mathcal{G}}\right] L_{\mathcal{F}}R.$$

Then, by Corollary F.5 Part i) we have

$$\begin{aligned} g_{2,i}(x_2) &\geq (1-\alpha\eta_1)g_{1,i}(x_1) - \eta_1^2\left[2\frac{L_{\mathcal{G}}}{R} + 7\beta_{\mathcal{G}}\right]\mathcal{V}_\alpha^2 \\ &\geq -\eta_2\left[58\frac{L_{\mathcal{G}}}{R} + 181\beta_{\mathcal{G}}\right]L_F R. \end{aligned}$$

Our inductive hypothesis is $g_{t,i}(x_t) \geq -c_2\eta_t$ for all i. We now show that it holds for $t+1$.

Case 1. Suppose $i \in \{1,\dots,m\}\backslash I(x_1)$, i.e., $g_i(x_t) > 0$. Then by Corollary F.5 ii)

$$g_{t+1,i}(x_{t+1}) \geq -\eta_{t+1}7\mathcal{V}_\alpha\left[L_{\mathcal{G}} + \frac{\beta_{\mathcal{G}}\mathcal{V}_\alpha}{4\alpha}\right] \geq -\eta_{t+1}\left[49\frac{L_{\mathcal{G}}}{R} + 86\beta_{\mathcal{G}}\right]L_{\mathcal{F}}R.$$

Case 2. Suppose $i \in I(x_t)$, i.e., $g_i(x_t) \leq 0$. Let $A = \left[2\frac{L_{\mathcal{G}}}{R} + 7\beta_{\mathcal{G}}\right]\mathcal{V}_\alpha^2$. By combining Corollary F.5 Part i), the inductive hypothesis and using similar arguments as in the proof of Lemma E.5 Case 2, yields

$$g_{t+1,i}(x_{t+1}) \geq -c_2\eta_{t+1}, \quad \text{where} \quad c_2 = 2.7\frac{A}{\alpha} = \left[265\frac{L_{\mathcal{G}}}{R} + 927\beta_{\mathcal{G}}\right]L_{\mathcal{F}}R.$$

The feasibility convergence rate is then given by

$$g_{t,i}(x_t) \geq -\left[265\frac{L_{\mathcal{G}}}{R} + 927\beta_{\mathcal{G}}\right]\frac{R^2}{\sqrt{t+15}}.$$

$\square$

Corollary F.5. *Suppose Assumptions 1.1 and Assumption F.1 hold. Let* $\alpha = L_{\mathcal{F}}/R$, $\mathcal{V}_\alpha = 7L_{\mathcal{F}}$ *and step sizes* $\eta_t = 1/(\alpha\sqrt{t+15})$. *Then, for every* $t \geq 1$ *we have*

i) $g_{t+1,i}(x_{t+1}) \geq (1-\alpha\eta_t)g_{t,i}(x_t) - \eta_t^2\left[2\frac{L_{\mathcal{G}}}{R} + 7\beta_{\mathcal{G}}\right]\mathcal{V}_\alpha^2$ *for all* $i \in I(x_t)$*; and*

ii) $g_{t+1,i}(x_{t+1}) \geq -\eta_{t+1}7\mathcal{V}_\alpha\left[L_{\mathcal{G}} + \frac{\beta_{\mathcal{G}}\mathcal{V}_\alpha}{4\alpha}\right]$ *for all* $i \in \{1,\dots,m\}\backslash I(x_t)$.

Proof. Combining Assumption F.1 and (S5) gives

$$g_{t+1,i}(x_{t+1}) \geq g_{t,i}(x_{t+1}) - 2\eta_{t+1}^2\left[\frac{L_{\mathcal{G}}}{R} + 3\beta_{\mathcal{G}}\right]\mathcal{V}_\alpha^2.$$

Then, by Claim D.6, it follows for every $i \in I(x_t)$ that

$$\begin{aligned} g_{t+1,i}(x_{t+1}) &\geq g_{t,i}(x_{t+1}) - 2\eta_{t+1}^2\left[\frac{L_{\mathcal{G}}}{R} + 3\beta_{\mathcal{G}}\right]\mathcal{V}_\alpha^2. \\ &\geq (1-\alpha\eta_t)g_{t,i}(x_t) - \eta_t^2\frac{\mathcal{V}_\alpha^2\beta_{\mathcal{G}}}{2} - \eta_t^2\left[2\frac{L_{\mathcal{G}}}{R} + 6\beta_{\mathcal{G}}\right]\mathcal{V}_\alpha^2 \\ &> (1-\alpha\eta_t)g_{t,i}(x_t) - \eta_t^2\left[2\frac{L_{\mathcal{G}}}{R} + 7\beta_{\mathcal{G}}\right]\mathcal{V}_\alpha^2, \end{aligned}$$

and for every $i \in \{1,\dots,m\}\backslash I(x_t)$ that

$$\begin{aligned} g_{t+1,i}(x_{t+1}) &\geq g_{t,i}(x_{t+1}) - 2\eta_{t+1}^2\left[\frac{L_{\mathcal{G}}}{R} + 3\beta_{\mathcal{G}}\right]\mathcal{V}_\alpha^2 \\ &\geq -\eta_{t+1}\mathcal{V}_\alpha\left[2L_{\mathcal{G}} + \frac{\beta_{\mathcal{G}}\mathcal{V}_\alpha}{4\alpha}\right] - \eta_{t+1}\mathcal{V}_\alpha\left[\frac{L_{\mathcal{G}}\mathcal{V}_\alpha}{2\alpha R} + \frac{3\beta_{\mathcal{G}}\mathcal{V}_\alpha}{2\alpha}\right] \\ &\geq -\eta_{t+1}7\mathcal{V}_\alpha\left[L_{\mathcal{G}} + \frac{\beta_{\mathcal{G}}\mathcal{V}_\alpha}{4\alpha}\right]. \end{aligned}$$

where we used that $\alpha = L_{\mathcal{F}}/R$ and $\mathcal{V}_\alpha = 7L_{\mathcal{F}}$ implies $\frac{L_{\mathcal{G}}\mathcal{V}_\alpha}{R\alpha} = 7L_{\mathcal{G}}$. $\square$

F.4 Average Time-Varying Constraints

An important special case where Assumption F.1 is satisfied, is summarized in the following slightly more general version of Lemma 3.2.

Lemma F.6. *Suppose the functions $\widetilde{g}_{t,i}$ satisfy Assumption 1.1 and in addition there is a decision $x_{t,i} \in \mathcal{B}_R$ such that $|\widetilde{g}_{t,i}(x_{t,i})| \leq \frac{1}{2}\left[\frac{L_\mathcal{G}}{R} + 3\beta_\mathcal{G}\right]R^2$, for every $t \geq 1$ and $i \in \{1, \ldots, m\}$. Then the following average time-varying constraints, satisfy Assumption 1.1 and Assumption F.1:*

$$g_{t,i}(x) := \frac{1}{t}\sum_{\ell=1}^{t} \widetilde{g}_{\ell,i}(x) \in \mathbb{R}^m. \tag{S6}$$

The rest of this subsection is devoted to proving Lemma F.6. We achieve this in two steps. We start by showing in Lemma F.7 that the average time-varying constraints satisfy Assumption 1.1, and then in Lemma F.8 we demonstrate that they also satisfy Assumption F.1.

Lemma F.7. *Suppose $\widetilde{g}_{t,i}$ is concave $\beta_\mathcal{G}$-smooth such that $\|\nabla \widetilde{g}_{t,i}(x)\| \leq L_\mathcal{G}$ for all $x \in \mathcal{B}_{4R}$, $t \geq 1$ and $i \in \{1, \ldots, m\}$. Then, the average function*

$$g_{t,i}(x) := \frac{1}{t}\sum_{\ell=1}^{t} \widetilde{g}_{\ell,i}(x)$$

is concave and $\beta_\mathcal{G}$-smooth and $\|\nabla g_{t,i}(x)\| \leq L_\mathcal{G}$ holds for all $x \in \mathcal{B}_{4R}$, $t \geq 1$ and $i \in \{1, \ldots, m\}$.

Proof. By assumption, each $\widetilde{g}_{\ell,i}$ is concave and $\beta_\mathcal{G}$-smooth, which implies

$$\widetilde{g}_{\ell,i}(x_{t+1}) \geq \widetilde{g}_{\ell,i}(x_t) + [\nabla \widetilde{g}_{\ell,i}(x_t)]^\top [x_{t+1} - x_t] - \frac{\beta_\mathcal{G}}{2}\|x_{t+1} - x_t\|^2.$$

Summing over all $\ell \in \{1, ..., t\}$ yields

$$\frac{1}{t}\sum_{\ell=1}^{t} \widetilde{g}_{\ell,i}(x_{t+1}) \geq \frac{1}{t}\sum_{\ell=1}^{t} \widetilde{g}_{\ell,i}(x_t) + \left[\frac{1}{t}\sum_{\ell=1}^{t} \nabla \widetilde{g}_{\ell,i}(x_t)\right]^\top [x_{t+1} - x_t] - \frac{1}{t}\sum_{\ell=1}^{t} \frac{\beta_\mathcal{G}}{2}\|x_{t+1} - x_t\|^2,$$

since $\frac{1}{t}\sum_{\ell=1}^{t} \nabla \widetilde{g}_{\ell,i}(x) = \nabla g_{t,i}(x)$, which is equivalent to

$$g_{t,i}(x_{t+1}) \geq g_{t,i}(x_t) + [\nabla g_{t,i}(x)]^\top [x_{t+1} - x_t] - \frac{\beta_\mathcal{G}}{2}\|x_{t+1} - x_t\|_2^2.$$

Hence, $g_{t,i}$ is concave and $\beta_\mathcal{G}$-smooth.

Moreover, since $\|\nabla \widetilde{g}_{t,i}(x)\| \leq L_\mathcal{G}$ for all $x \in \mathcal{B}_{4R}$, we have

$$\|\nabla g_{t,i}(x)\| = \left\|\frac{1}{t}\sum_{\ell=1}^{t} \nabla \widetilde{g}_{\ell,i}(x)\right\| \leq \frac{1}{t}\sum_{\ell=1}^{t} \|\nabla \widetilde{g}_{\ell,i}(x)\| \leq L_\mathcal{G}.$$

□

We show next that the average time-varying constraints satisfy Assumption F.1.

Lemma F.8 (Average TVC)**.** *Suppose $\widetilde{g}_{t,i}$ is concave $\beta_\mathcal{G}$-smooth such that $\|\nabla \widetilde{g}_{t,i}(x)\| \leq L_\mathcal{G}$ for all $x \in \mathcal{B}_{4R}$, $t \geq 1$ and $i \in \{1, \ldots, m\}$. Further, suppose for every $t \geq 1$ and $i \in \{1, \ldots, m\}$, there exists a decision $x_{t,i} \in \mathcal{B}_R$ such that*

$$|\widetilde{g}_{t,i}(x_{t,i})| \leq \frac{1}{2}\left[\frac{L_\mathcal{G}}{R} + 3\beta_\mathcal{G}\right]R^2. \tag{S7}$$

Then, for $\alpha = L_\mathcal{F}/R$, step sizes $\eta_t = 1/(\alpha\sqrt{t+15})$ and $\mathcal{V}_\alpha = 7L_\mathcal{F}$, it holds for every $x \in \mathcal{B}_{4R}$ that

$$|g_{t+1,i}(x) - g_{t,i}(x)| \leq 2\eta_{t+1}^2\left[\frac{L_\mathcal{G}}{R} + 3\beta_\mathcal{G}\right]\mathcal{V}_\alpha^2.$$

Proof. Using the inequality $\frac{1}{t+1} \leq \frac{17}{2}\frac{1}{t+16}$ for every $t \geq 1$ and $\eta_{t+1}^2 = 1/(\alpha^2(t+16))$, it follows by construction that

$$\begin{aligned}
|g_{t+1,i}(x) - g_{t,i}(x)| &= \left| \frac{1}{t+1}\widetilde{g}_{t+1,i}(x) + \frac{t}{t+1}g_{t,i}(x) - g_{t,i}(x) \right| \\
&= \frac{1}{t+1}\left|\widetilde{g}_{t+1,i}(x) - g_{t,i}(x)\right| \\
&= \frac{1}{t+1}\frac{1}{t}\left|\sum_{\ell=1}^{t}\widetilde{g}_{t+1,i}(x) - \widetilde{g}_{\ell,i}(x)\right| \\
&\leq \eta_{t+1}^2\frac{17}{2}\alpha^2 \cdot \frac{1}{t}\sum_{\ell=1}^{t}\left|\widetilde{g}_{t+1,i}(x) - \widetilde{g}_{\ell,i}(x)\right|. \qquad \text{(S8)}
\end{aligned}$$

By triangle inequality $|\widetilde{g}_{t+1,i}(x) - \widetilde{g}_{\ell,i}(x)| \leq |\widetilde{g}_{t+1,i}(x)| + |\widetilde{g}_{\ell,i}(x)|$ and thus it suffices to bound the term $|\widetilde{g}_{t,i}(x)|$ for every $t \geq 1$, $i \in \{1, \ldots, m\}$ and $x \in \mathcal{B}_{4R}$.

By assumption, $x \in \mathcal{B}_{4R}$ and there is $x_{t,i} \in \mathcal{B}_R$ satisfying inequality (S7). Further, $\widetilde{g}_{t,i}$ is concave, which implies

$$\widetilde{g}_{t,i}(x) - \widetilde{g}_{t,i}(x_{t,i}) \leq [\nabla\widetilde{g}_{t,i}(x_{t,i})]^\top[x - x_{t,i}] \leq 5L_{\mathcal{G}}R$$

and the fact that $\widetilde{g}_{t,i}$ is concave $\beta_{\mathcal{G}}$-smooth yields

$$\begin{aligned}
\widetilde{g}_{t,i}(x) - \widetilde{g}_{t,i}(x_{t,i}) &\geq [\nabla\widetilde{g}_{t,i}(x_{t,i})]^\top[x - x_{t,i}] - \frac{\beta_{\mathcal{G}}}{2}\|x_{t,i} - x\|^2 \\
&\geq -5\left[\frac{L_{\mathcal{G}}}{R} + 3\beta_{\mathcal{G}}\right]R^2.
\end{aligned}$$

Further, by combining $\left|\widetilde{g}_{t,i}(x) - \widetilde{g}_{t,i}(x_{t,i})\right| \leq 5\left[\frac{L_{\mathcal{G}}}{R} + 3\beta_{\mathcal{G}}\right]R^2$, triangle inequality and assumption (S7), we obtain for every $x \in \mathcal{B}_{4R}$ that

$$\begin{aligned}
|\widetilde{g}_{t,i}(x)| &= |\widetilde{g}_{t,i}(x) - \widetilde{g}_{t,i}(x_{t,i}) + \widetilde{g}_{t,i}(x_{t,i})| \\
&\leq |\widetilde{g}_{t,i}(x) - \widetilde{g}_{t,i}(x_{t,i})| + |\widetilde{g}_{t,i}(x_{t,i})| \\
&\leq \frac{11}{2}\left[\frac{L_{\mathcal{G}}}{R} + 3\beta_{\mathcal{G}}\right]R^2.
\end{aligned}$$

The statement follows by combining $\alpha = L_{\mathcal{F}}/R$, $\mathcal{V}_\alpha = 7L_{\mathcal{F}}$, (S8) and

$$\begin{aligned}
|g_{t+1,i}(x) - g_{t,i}(x)| &\leq \eta_{t+1}^2\frac{17}{2}\alpha^2 \cdot \frac{1}{t}\sum_{\ell=1}^{t}|\widetilde{g}_{t+1,i}(x) - \widetilde{g}_{\ell,i}(x)| \\
&\leq \eta_{t+1}^2\left[\frac{L_{\mathcal{G}}}{R} + 3\beta_{\mathcal{G}}\right]\frac{11}{2} \cdot 17\alpha^2R^2 \\
&< 2\eta_{t+1}^2\left[\frac{L_{\mathcal{G}}}{R} + 3\beta_{\mathcal{G}}\right]\mathcal{V}_\alpha^2.
\end{aligned}$$

□