Safe Online Convex Optimization with First-order Feedback

Spencer Hutchinson and Mahnoosh Alizadeh

University of California, Santa Barbara
Department of Electrical and Computer Engineering

July 10, 2024

Safe Decision-making under Uncertainty

Many real-world applications require making decisions under uncertainty, while ensuring that constraints are *always* satisfied.

How do we ensure constraint satisfaction under incomplete information without sacrificing performance?

Online Convex Optimization (OCO)

Online convex optimization (OCO) is a powerful framework for handling decision-making under uncertainty.

Online Convex Optimization

In each round $t \in [T]$:

- Player chooses action $x_t \in \mathcal{X} \subseteq \mathbb{R}^d$.
- Adversary chooses $f_t : \mathbb{R}^d \to \mathbb{R}$ and player suffers cost $f_t(x_t)$.
- Player observes f_t.

Player aims to minimize *regret* with respect to single best action in hindsight, $R_T = \sum_{t=1}^T f_t(x_t) - \min_{x \in \mathcal{X}} \sum_{t=1}^T f_t(x)$.

However, most OCO approaches assume that either

- constraints on the player's actions are entirely known, or
- constraint violation is acceptable.

Safe OCO with First-order Feedback

Interaction Model

In each round $t \in [T]$:

- Player chooses action $x_t \in \mathcal{X} \subseteq \mathbb{R}^d$.
- Adversary chooses $f_t : \mathbb{R}^d \to \mathbb{R}$ and player suffers cost $f_t(x_t)$.
- Player observes $\nabla f_t(x_t), g(x_t), \nabla g(x_t)$.

Learning Goals

- Ensure constraint satisfaction: $g(x_t) \leqslant 0$ for all $t \in [T]$.
- Minimize regret w.r.t. to best feasible action

$$R_T = \sum_{t=1}^{T} f_t(x_t) - \min_{x \in \mathcal{X} : g(x) \leqslant 0} \sum_{t=1}^{T} f_t(x)$$

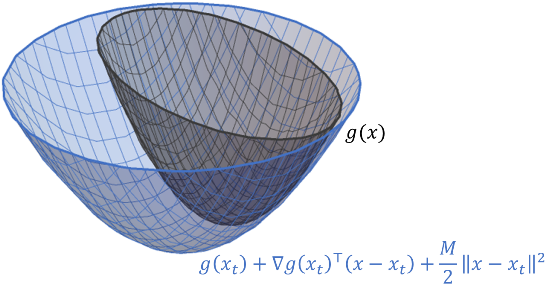
Handling Unknown Constraints: Optimistic Sets

Assumption

The constraint function g is M-strongly convex.

We construct an *optimistic set* that contains the feasible set,

$$\mathcal{Y}_t^o := \left\{ x \in \mathcal{X} : g(x_t) + \nabla g(x_t)^\top (x - x_t) + \frac{M}{2}\|x - x_t\|^2 \leqslant 0 \right\}$$
$$\supseteq \{x \in \mathcal{X} : g(x) \leqslant 0\} =: \mathcal{Y}.$$



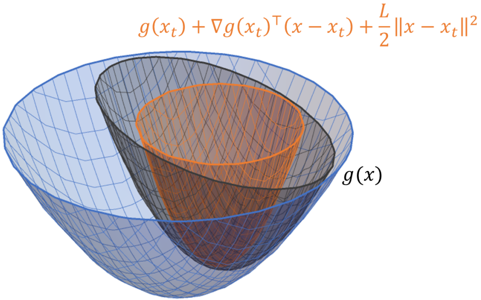
Maintaining Constraint Satisfaction: Pessimistic Sets

Assumption

The constraint function g is L-smooth.

We construct a *pessimistic set* that is contained by the feasible set,

$$\mathcal{Y}_t^p := \left\{ x \in \mathcal{X} : g(x_t) + \nabla g(x_t)^\top (x - x_t) + \frac{L}{2}\|x - x_t\|^2 \leqslant 0 \right\}$$
$$\subseteq \{x \in \mathcal{X} : g(x) \leqslant 0\} =: \mathcal{Y}.$$



Proposed Algorithm

Algorithm 1: Restrained Online Gradient Descent (ROGD)

Set $\tilde{x}_1 = \mathbf{0}$ and $x_1 = \mathbf{0}$.

for $t = 1$ **to** T **do**

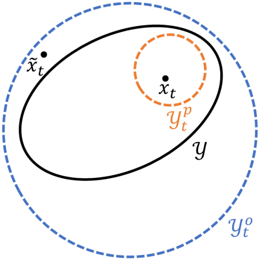
Play x_t and observe $\nabla f_t(x_t), g(x_t), \nabla g(x_t)$.

$\tilde{x}_{t+1} = \Pi_{\mathcal{Y}_t^o}(\tilde{x}_t - \eta \nabla f_t(x_t))$.

$\gamma_t = \max\{\mu \in [0,1] : x_t + \mu(\tilde{x}_{t+1} - x_t) \in \mathcal{Y}_t^p\}$.

$x_{t+1} = x_t + \gamma_t(\tilde{x}_{t+1} - x_t)$.

end



Regret Bound

Assumption

- *Gradients of f_t are bounded, i.e.* $\|\nabla f_t(x)\| \leqslant G$ *for all* $t \in [T], x \in \mathcal{X}$.
- $\mathcal{X}$ *is bounded, i.e.* $\|x - y\| \leqslant D$ *for all* $x, y \in \mathcal{X}$.

Theorem

Let $\kappa := L/M$ *be the condition number of the constraint. Then, the regret of ROGD satisfies,*

$$R_T \leqslant \underbrace{(\kappa - 1)G^2\eta T}_{\text{gap between } \tilde{x}_t \text{ and } x_t} + \underbrace{\frac{D^2}{2\eta} + \frac{1}{2}G^2\eta T}_{\text{regret of OGD}}.$$

Choosing $\eta = \frac{D}{G\sqrt{T}}$, *ensures that* $R_T \leqslant \kappa DG\sqrt{T}$.

Analysis Approach

$$
\begin{aligned}
R_T &= \sum_{t=1}^{T} (f_t(x_t) - f_t(x^*)) \qquad \left(x^* := \operatorname*{arg\,min}_{x \in \mathcal{X} : g(x) \leqslant 0} \sum_{t=1}^{T} f_t(x)\right) \\
&\leqslant \sum_{t=1}^{T} \nabla f_t(x_t)^\top (x_t - x^*) \qquad \text{(convexity)} \\
&= \sum_{t=1}^{T} \nabla f_t(x_t)^\top (x_t - \tilde{x}_t) + \sum_{t=1}^{T} \nabla f_t(x_t)^\top (\tilde{x}_t - x^*) \\
&= \underbrace{G \sum_{t=1}^{T} \|x_t - \tilde{x}_t\|}_{\text{Term I}} + \underbrace{\sum_{t=1}^{T} \nabla f_t(x_t)^\top (\tilde{x}_t - x^*)}_{\text{Term II}}
\end{aligned}
$$

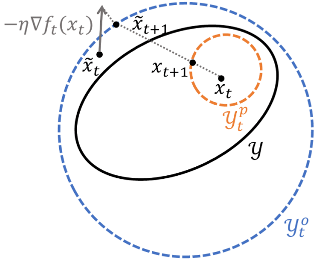
Term I = cumulative distance between x_t and $\tilde{x}_t$.
Term II = linearized regret of $\tilde{x}_t$.

Term I

Lemma

It holds that $\gamma_t \geqslant 1/\kappa$ *for all* $t \in [T]$.

$$\gamma_t = \max\{\mu \in [0,1] : x_t + \mu(\tilde{x}_{t+1} - x_t) \in \mathcal{Y}_t^p\}$$
$$x_{t+1} = x_t + \gamma_t(\tilde{x}_{t+1} - x_t)$$



Suppose that $\|x_t - \tilde{x}_t\| \leqslant G(\kappa - 1)\eta$. It follows that,

Term II

Since $x^* \in \mathcal{Y} \subseteq \mathcal{Y}_t^o$ and $\mathcal{Y}_t^o$ is convex, it holds that,

$$\begin{aligned}
\|\tilde{x}_{t+1} - x^*\|^2 &= \|\Pi_{\mathcal{Y}_t^o}(\tilde{x}_t - \eta \nabla f_t(x_t)) - x^*\|^2 \\
&\leqslant \|\tilde{x}_t - \eta \nabla f_t(x_t) - x^*\|^2 \\
&= \|\tilde{x}_t - x^*\|^2 - 2\eta \nabla f_t(x_t)^\top (\tilde{x}_t - x^*) + \eta^2 \|\nabla f_t(x_t)\|^2 \\
&\leqslant \|\tilde{x}_t - x^*\|^2 - 2\eta \nabla f_t(x_t)^\top (\tilde{x}_t - x^*) + \eta^2 G^2.
\end{aligned}$$

Then, rearranging,

$$\begin{aligned}
\text{Term II} &= \sum_t \nabla f_t(x_t)^\top (\tilde{x}_t - x^*) \\
&\leqslant \frac{1}{2\eta} \sum_t (\|\tilde{x}_t - x^*\|^2 - \|\tilde{x}_{t+1} - x^*\|^2) + \frac{1}{2} G^2 \eta T \\
&\leqslant \frac{1}{2\eta} (\|\tilde{x}_1 - x^*\|^2 - \|\tilde{x}_{T+1} - x^*\|^2) + \frac{1}{2} G^2 \eta T \\
&\leqslant \frac{1}{2\eta} D^2 + \frac{1}{2} G^2 \eta T = \mathcal{O}(\sqrt{T}). \qquad (\eta = \Theta(1/\sqrt{T}))
\end{aligned}$$

Extension to Time-varying Constraints

Our approach can handle time-varying constraints $g_1, g_2, ..., g_T$, assuming that:

- Constraint sets are monotone increasing, i.e. $\{x : g_1(x)\} \subseteq \{x : g_2(x)\} \subseteq ... \subseteq \{x : g_T(x)\}$.
- Player receives first-order feedback on next constraint, i.e. player observes $g_{t+1}(x_t), \nabla g_{t+1}(x_t)$.

In this case, we give a bound on the *dynamic* regret,

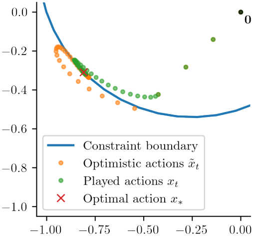
$$R_T^d = \sum_{t=1}^{T} f_t(x_t) - \sum_{t=1}^{T} \min_{x \in \mathcal{X} : g_t(x) \leqslant 0} f_t(x) = \mathcal{O}(\sqrt{T(P_T + 1)}),$$

where P_T is the path length of minimizer sequence, i.e.

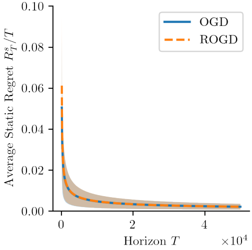
$$P_T = \sum_{t=1}^{T-1} \|x_t^* - x_{t+1}^*\|, \qquad x_t^* = \operatorname*{arg\,min}_{x \in \mathcal{X} : g_t(x) \leqslant 0} f_t(x).$$

Numerical Experiments

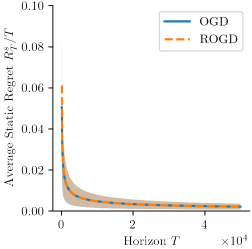
Empirical comparison of ROGD with first-order feedback to online gradient descent (OGD) with full constraint information.



Fixed cost function and fixed constraint.



Time-varying cost function and fixed constraint.



Time-varying cost function and time-varying constraint.

Future Directions

Some interesting directions for future work:

- Relax assumptions on the constraint function (smooth and strongly-convex).
- Use weaker feedback models, e.g. zero-order feedback. (See our L4DC paper)
- Apply our approach to related safe learning settings, e.g. distributed online optimization or online control.
- Apply our algorithm to real-world settings, such as demand response in the electric grid.

Thank you for listening!

shutchinson@ucsb.edu