An Optimistic Approach to Online Optimization with Unknown Linear Constraints

Spencer Hutchinson
Advisor: Mahnoosh Alizadeh

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

November 1, 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 [Zinkevich, 2003]

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

- Play action $x_t \in \mathcal{X} \subseteq \mathbb{R}^d$.
- Suffer cost $f_t(x_t)$ according to (adversarily-chosen) cost function $f_t : \mathbb{R}^d \to \mathbb{R}$.
- Observe f_t.

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

Online Convex Optimization

The OCO problem has garnered significant attention as both:

- *a fundamental building block in modern learning and control*
    - stochastic optimization (online-to-batch)
    - non-convex optimization [Cutkosky et al., 2023]
    - adaptive algorithms [Duchi et al., 2011]
    - online control [Cohen et al., 2018]
- *a model for real-world problems*
    - online advertising [McMahan et al., 2013]
    - internet of things [Chen and Giannakis, 2019]
    - smart grid [Lesage-Landry and Callaway, 2020]
    - healthcare [Tewari and Murphy, 2017]

However, classical OCO approaches either (1) assume that any constraints on the actions are known a priori or (2) allow constraints to be violated.

OCO with Unknown Linear Constraints

Interaction Model

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

- Play action $x_t \in \mathcal{X} \subseteq \mathbb{R}^d$.
- Suffer cost $f_t(x_t)$ according to cost function $f_t : \mathbb{R}^d \to \mathbb{R}$.
- Observe f_t and noisy constraint feedback $y_t = Ax_t + \epsilon_t$.[1]

Learning Goals

- Ensure constraint satisfaction: $Ax_t \leqslant b$ for all $t \in [T]$.
- Minimize regret w.r.t. to best feasible action:

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

[1] The random noise ϵ_t is assumed to be conditionally-subgaussian.

Prior Work

Prior work has studied the same problem:[2]

- *Regret:* $\tilde{\mathcal{O}}(T^{2/3})$
- *Constraint guarantee:* $Ax_t \leqslant b \; \forall t$ w.h.p.
- *Approach:* explore-exploit

$$t \in [1, T^{2/3}] : \; x_{t+1} \sim \mathcal{U}$$
$$t \in [T^{2/3}, T] : \; x_{t+1} = \Pi_{\mathcal{Y}_t^p}(x_t - \eta \nabla f_t(x_t))$$

Our results:

- *Regret:* $\tilde{\mathcal{O}}(\sqrt{T})$
- *Constraint guarantee:* $Ax_t \leqslant b \; \forall t$ w.h.p.
- *Approach:* optimistic

[2] Chaudhary and Kalathil, 2022; Chang et al., 2023

Least-squares Confidence Bounds

In each round t, we know (1) action and feedback history $x_1, ..., x_{t-1}, y_1, ..., y_{t-1}$, (2) that expected feedback is linear $\mathbb{E}[y_t] = Ax_t$.

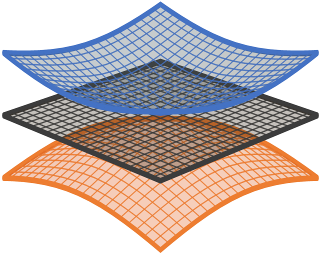
Estimate A with (regularized) least-squares:

$$\hat{A}_t = \left(\sum_{s=1}^{t-1} y_t x_t^\top \right) V_t^{-1}, \text{ where, } V_t = \sum_{s=1}^{t-1} x_s x_s^\top + \lambda I$$

Then use $\hat{A}_t$ for high probability confidence bounds on the true constraint,[3]

$$\hat{A}_t x - \beta_t \|x\|_{V_t^{-1}} \mathbf{1} \leqslant Ax \leqslant \hat{A}_t x + \beta_t \|x\|_{V_t^{-1}} \mathbf{1}$$

where $\|x\|_{V_t^{-1}} = \sqrt{x^\top V_t^{-1} x}$.

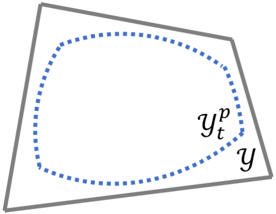


[3] Abbasi-yadkori et al., 2011

Pessimistic Sets

A "pessimistic set" can be constructed with the confidence bounds:[4]

$$\mathcal{Y}_t^p := \{x \in \mathcal{X} : \hat{A}_t x + \beta_t \|x\|_{V_t^{-1}} \mathbf{1} \leqslant b\}$$
$$\subseteq \{x \in \mathcal{X} : Ax \leqslant b\} =: \mathcal{Y}$$



Playing in the pessimistic set ensures constraint satisfaction:

$$x_t \in \mathcal{Y}_t^p \quad \forall t \in [T] \underset{\text{w.h.p.}}{\implies} Ax_t \leqslant b \quad \forall t \in [T]$$

Prior work performs regret minimization (e.g. gradient descent) on the pessimistic set.

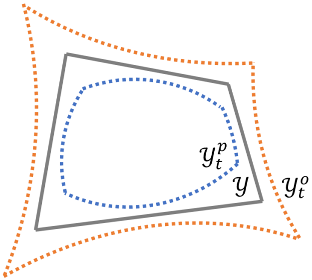
However, to sufficiently "grow" the pessimistic set, this approach requires $T^{2/3}$ rounds of pure exploration $\implies \tilde{\mathcal{O}}(T^{2/3})$ regret.

[4] Amani et al., 2019; Chaudhary and Kalathil, 2022

Optimistic Sets

Our approach: We additionally use "optimistic" sets that contain the feasible set.

$$\mathcal{Y}_t^o := \{x \in \mathcal{X} : \hat{A}_t x - \beta_t \|x\|_{V_t^{-1}} \mathbf{1} \leqslant b\}$$
$$\supseteq \{x \in \mathcal{X} : Ax \leqslant b\} =: \mathcal{Y}$$



Regret bound on optimistic set $\implies$ regret bound on true feasible set.

We ensure low regret with regret minimization on the *optimistic sets*.

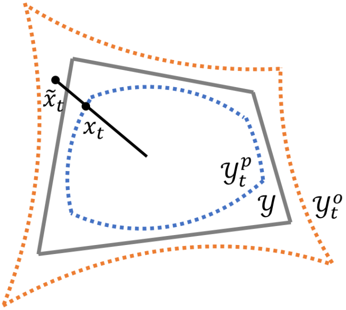
However, we can't directly play actions chosen in the optimistic set because they violate the constraints.

Safety Through Scaling

How do we ensure constraint satisfaction when choosing actions from the optimistic set?

Given actions $\tilde{x}_t$ in $\mathcal{Y}_t^o$, we scale in to the pessimistic set:

$$\begin{aligned} x_t &= \gamma_t \tilde{x}_t, \\ \gamma_t &= \max\{\gamma \in [0,1] : \gamma \tilde{x}_t \in \mathcal{Y}_t^p\}. \end{aligned} \tag{1}$$



Lemma

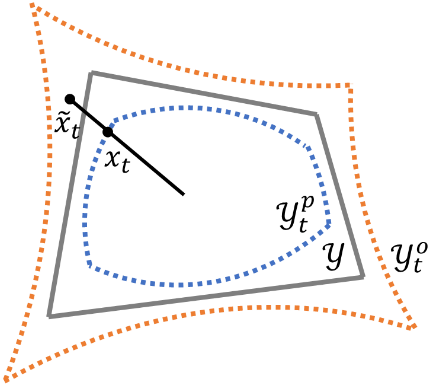
If x_t are chosen as in (1), then $\sum_{t=1}^{T} \|\tilde{x}_t - x_t\| = \tilde{\mathcal{O}}(\sqrt{T})$.

There is only a small cost for scaling in to the pessimistic set.

Overall Approach

For $t \in [T]$:

1. Construct $\mathcal{Y}_t^o$ and $\mathcal{Y}_t^p$.
2. Regret minimizer chooses $\tilde{x}_t \in \mathcal{Y}_t^o$.
3. $\gamma_t = \max\left\{\mu \in [0,1] : \mu \tilde{x}_t \in \mathcal{Y}_t^p\right\}$.
4. Play $x_t = \gamma_t \tilde{x}_t$



Regret Minimization on Optimistic Sets

We need an (efficient) OCO algorithm that chooses actions $(\tilde{x}_t)_t$,

$$\sum_{t=1}^{T} f_t(\tilde{x}_t) - \min_{x \in \bigcap_t \mathcal{Y}_t^o} \sum_{t=1}^{T} f_t(x) \leqslant C\sqrt{T}.$$

Challenges: the optimistic set $\mathcal{Y}_t^o$ is (1) time-varying, and (2) non-convex.

$\implies$ Standard OCO algorithms are not suitable.

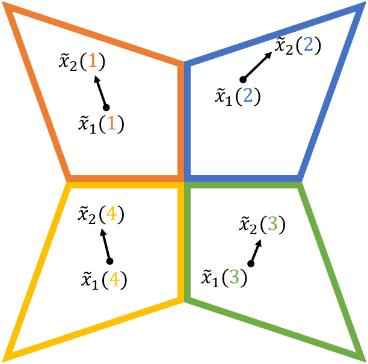
Key ideas:

- **Rare updating:** Only update $\mathcal{Y}_t^o$ when $\det(V_t)$ doubles.
- **Relaxed optimistic set:** Use ∞-norm relaxation of optimistic set. $\tilde{\mathcal{Y}}_t^o := \{x \in \mathcal{X} : \hat{A}_t x - \sqrt{d}\beta_t \| V_t^{-1/2} x\|_\infty \mathbf{1} \leqslant b\} \supseteq \mathcal{Y}_t^o$
- **Finite union of convex sets:** Represent as finite union. $\tilde{\mathcal{Y}}_t^o = \bigcup_{j \in [k], \xi \in \{-1,1\}} \{x \in \mathcal{X} : \hat{A}_t x - \xi\sqrt{d}\beta_t [V_t^{-1/2}]_j x \mathbf{1} \leqslant b\}$
- **Hedge over finite sets:** Use Hedge to select set and gradient descent to minimize regret in each set.

HedgeDescent

Initialize: $p_1(m) = 1/(M), \tilde{x}_1(m) = \mathbf{0}$ for all $m \in [M]$.
For $\tau = 1, 2, \ldots$:

- Sample Hedge: $m_\tau \sim p_\tau$.
- Play $\tilde{x}_\tau(m_\tau)$ and observe f_τ.
- Update Gradient Descent: $\tilde{x}_{\tau+1}(m) = \Pi_{\tilde{\mathcal{Y}}^o(m_\tau)}(\tilde{x}_\tau(m_\tau) - \eta \nabla_{\tau,m})$.
- Update Hedge: $p_{\tau+1}(m) \propto p_{\tau+1}(m) \exp(-\zeta f_\tau(\tilde{x}_\tau(m)))$.



Extension to OCO with Stochastic Constraints

Our approach can be extended to OCO with stochastic constraints g_t, provided that:

- $g_t(x) := A_t x - b_t$ where $(A_t, b_t)_t$ is an i.i.d. sequence.
- Receive "bandit" feedback $g_t(x_t)$.
- A_t, b_t are bounded.
- $\mathbb{E}[b_t]$ is known.

In this case, regret is defined as,

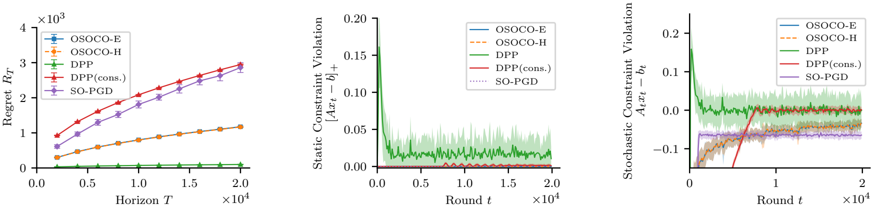
$$R_T := \sum_{t=1}^{T} f_t(x_t) - \min_{x \in \mathcal{X} : \mathbb{E}[A_t] x \leqslant \mathbb{E}[b_t]} \sum_{t=1}^{T} f_t(x).$$

By tightening the constraint in our algorithm, we guarantee:

$$\mathbb{E}[R_T] = \tilde{\mathcal{O}}(\sqrt{T}) \quad \text{and} \quad \mathbb{E}[g_t(x_t)] \leqslant 0$$

Numerical Experiments

Empirical comparison of our algorithm (OSOCO) with existing algorithms in settings with both static and stochastic constraints.



Existing algorithms are SO-PGD [Chaudhary and Kalathil, 2021] and DPP [Yu et al., 2017].

Conclusion

- Efficient algorithm with $\tilde{\mathcal{O}}(\sqrt{T})$ regret and no constraint violation (w.h.p.) for OCO with unknown linear constraints.
- Improves on prior work that has shown $\tilde{\mathcal{O}}(T^{2/3})$ regret and same constraint violation guarantees.
- Extension to stochastic linear constraints shows $\tilde{\mathcal{O}}(\sqrt{T})$ regret and no constraint violation in expectation.
- Possible future work includes extensions to safe online control or distributed online optimization, and applications to smart grid.

shutchinson@ucsb.edu