A General Framework for Safe Decision Making: A Convex Duality Approach

Martino Bernasconi* **Federico Cacciamani*** **Nicola Gatti*** **Francesco Trovò***

Abstract

We study the problem of online interaction in general decision making problems, where the objective is not only to find optimal strategies, but also to satisfy some *safety* guarantees, expressed in terms of costs accrued. We propose a theoretical framework to address such problems and present BAN-SOLO, a UCB-like algorithm that, in an online interaction with an unknown environment, attains sublinear regret of order $\mathcal{O}(\sqrt{T})$ and plays safely with high probability at each iteration. At its core, BAN-SOLO relies on tools from *convex duality* to manage environment exploration while satisfying the safety constraints imposed by the problem.

1 Introduction

In the recent years, improvements in the field of Artificial Intelligence, more in particular in the subjects of Algorithmic Game Theory (AGT) and Reinforcement Learning (RL), made it possible to achieve outstanding results in various applications. These range from recreative applications, e.g., Chess [Silver et al., 2018], Texas hold'em poker [Brown and Sandholm, 2018], Go [Silver et al., 2017], to security applications such as anti-poaching patrolling [Tambe, 2011]. The increasing number of successful applications of decision making algorithms to real-world tasks raises fundamental concerns in terms of *safety*. Indeed, when considering critical tasks with humans in the loop, it becomes of upmost importance to avoid the occurrence of undesirable and potentially dangerous behaviour of the algorithms, especially during the learning process. In this work, we propose a theoretical framework to address the aforementioned problem. The decision-making model that we adopt is general and can capture many popular strategic scenarios, (*e.g., sequential decision making, multi-armed bandits, partially observable markov decision processes*). The concept of safety is modeled by constraints on a cost function that can be suitably specified to represent any particular, domain-dependent safety constraint. The algorithm proposed is called BANdit Safe Online Linear Optimization (BAN-SOLO). During the online interaction with the environment, BAN-SOLO pursues a twofold objective: (i) minimize the regret against the best possible safe strategy and (ii) always select safe strategies. In order to achieve such a goal, BAN-SOLO uses the feedback received from the environment to estimate an high confidence region for the environment model and exploits the concept of convex dual set of such an high confidence region to achieve $\mathcal{O}(\sqrt{T})$ regret and high-probability safety at each iteration.

2 Preliminaries

In this section we will define the model of interaction that is adopted throughout the rest of the paper, and then we will review some concepts from variational analysis.

*DEIB, Politecnico di Milano, {martino.bernasconideluca, federico.cacciamani, nicola.gatti, francesco1.trovo.gatti}@polimi.it.
Politecnico di Milano

ML Safety Workshop, 36th Conference on Neural Information Processing Systems (NeurIPS 2022).

2.1 Decision Making with Bandit Feedback and Costs

We consider a general decision making scenario in which the set of strategies available to the agent is a bounded and convex set $\mathcal{X}$. After selecting a strategy $\boldsymbol{x} \in \mathcal{X}$, the agent receives a utility $u(\boldsymbol{x})$ and pays a cost $c(\boldsymbol{x})$. We focus on the case in which the utility and cost functions are linear functions with unknown parameters $\boldsymbol{y}^\star$ and $\boldsymbol{\omega}^\star$, respectively, *i.e.,* such that $u(\boldsymbol{x}) = \langle \boldsymbol{x}, \boldsymbol{y}^\star \rangle$ and $c(\boldsymbol{x}) = \langle \boldsymbol{x}, \boldsymbol{\omega}^\star \rangle$.

At each time instant $t \in [T]$[2], the agent selects a strategy $\boldsymbol{x}^t \in \mathcal{X}$ and receives a partial feedback ℓ^t from the environment. For each t, the objective of the agent is to select strategies that maximize her utility, while being *safe*. The concept of safety that we adopt in this work is expressed in terms of costs. In particular, a strategy $\boldsymbol{x}$ is said to be safe if it guarantees that the expected cost is within an interval $\mathcal{C} := [\alpha, \beta] \subset \mathbb{R}$. Thus, the set of *safe strategies* is defined as $\mathcal{X}^\star := \{\boldsymbol{x} \in \mathcal{X} \mid c(\boldsymbol{x}) \in \mathcal{C}\}$.

The performances of the agent are evaluated in terms of *cumulative regret* R^T which is defined as $R^T = \sup_{\boldsymbol{x} \in \mathcal{X}^\star} \sum_{t=1}^T \langle \boldsymbol{x}, \boldsymbol{y}^\star \rangle - \sum_{t=1}^T \langle \boldsymbol{x}^t, \boldsymbol{y}^\star \rangle$. We demand that the cumulative regret is sub-linear in T, while $\langle \boldsymbol{x}^t, \boldsymbol{\omega}^\star \rangle \in \mathcal{C}$, for all t, with high probability. Additionally, we assume that $\mathcal{X}^\star$ is not empty, otherwise the problem is trivially impossible to solve, and that we have a starting set $\mathcal{X}^0 \subset \mathcal{X}^\star$ for the safe exploration, such that for all $\boldsymbol{x} \in \mathcal{X}^0$ it holds $\langle \boldsymbol{x}, \boldsymbol{\omega}^\star \rangle \in \mathcal{C}$.

2.2 Variational Analysis

First we are going to recall the definition of Hausdorff distance which defines a metric between compact sets. We will use extensively this notion throughout the paper.

Definition 1. *Let A and B, two non-empty subsets of a metric space $(\mathcal{M}, d)$. The Hausdorff distance between A and B is defined as $d_H(A, B) = \inf_{\epsilon > 0} \{A \subseteq B_\epsilon \text{ and } B \subseteq A_\epsilon\}$, where A_ϵ is the ϵ-flattening of the set A defined as $A_\epsilon := \bigcup_{z \in A} \{z' \in \mathcal{M} \mid d(z, z') \leq \epsilon\}$.*

We exploit this definition to define the rate of convergence of a sequence of sets.

Definition 2. *We write $A^t \xrightarrow{K(t)} B$ when a sequence of sets $\{A^t\}_{t \in \mathbb{N}}$ converges to B with rate $K(t) = o(t)$, i.e., $d_H(A^t, B) \leq K(t)$, for all t.*

We will also use the notion of *polar sets* of convex polytopes. Namely we will use the fact that any bounded convex polytope can be defined in two equivalent ways: as a *convex hull* of a finite set of points ($\mathcal{V}$-Polytope) and as the intersection of a finite number of closed *half-spaces* ($\mathcal{H}$-Polytope). Indeed, starting from the definition a $\mathcal{V}$-Polytope (respectively $\mathcal{H}$-Polytope) it is always possible to obtain the polar set (indicated with the superscript $\circ$) in terms of $\mathcal{H}$-Polytope (respectively $\mathcal{V}$-Polytope).

Lemma 1 (Polar Set Ziegler [2012])**.** *Given any set of p points, $\{a_1, ..., a_p\}$ with $a_i \in \mathbb{R}^n \; \forall i$, if A is the $n \times p$ matrix whose i^{th} column is a_i, then hull$(\{a_1, ..., a_p\})^\circ = \{\boldsymbol{x} \in \mathbb{R}^n | A^\top \boldsymbol{x} \leq \mathbf{1}\}$. Conversely we have that $(\{\boldsymbol{x} \in \mathbb{R}^n | A^\top \boldsymbol{x} \leq \mathbf{1}\})^\circ =$ hull$(\{a_1, ..., a_p\})$, assuming the left hand side is bounded.*

Additional Notation We indicate the convergence of a sequence of real numbers $\{\alpha^t\}_{t \geq 0}$ to $\alpha^\star$ with $|\alpha^t - \alpha^\star| \leq K(t)$ as $\alpha^t \xrightarrow{K(t)} \alpha^\star$. We denote with $epi(f) \subset \mathbb{R}^{n+1}$ the epigraph of a function $f : \mathbb{R}^n \to \bar{\mathbb{R}}$.[3] Finally, we use $\delta_{\mathcal{X}}(\boldsymbol{x})$ to denote a function that is 0 if $\boldsymbol{x} \in \mathcal{X}$ and $+\infty$ if $\boldsymbol{x} \notin \mathcal{X}$.

2.3 Structure of the paper

This subsection outlines the main idea behind our approach and the structure of the paper. We start the discussion by imagining to have an oracle returning at each time t a feasibility set $\mathcal{X}^t \subset \mathcal{X}^\star$ s.t. for all $\boldsymbol{x} \in \mathcal{X}^t$ we have $\langle \boldsymbol{x}, \boldsymbol{\omega}^\star \rangle \in \mathcal{C}$ with high probability. As a starting point of our discussion, we assume the oracle to return better and better feasibility sets $\mathcal{X}^t$ as the time t progresses, and more experience is collected. Formally, we use Definition 2 of set-convergence and ask that $\mathcal{X}^t \xrightarrow{K(t)} \mathcal{X}^\star$. In Section 3 we will show that when $\mathcal{X}^t \xrightarrow{K(t)} \mathcal{X}^\star$ and we know $\boldsymbol{y}^\star$, we obtain a convergence rate

[2] In this work we let $[N]$ be the first N natural numbers.

[3] $\bar{\mathbb{R}}$ is the *extended* real line that contains $\pm\infty$.

of $\mathcal{O}(K(t))$ to the optimal value $\sup_{\boldsymbol{x}\in\mathcal{X}^\star}\langle \boldsymbol{x}^\star, \boldsymbol{y}^\star\rangle$, by playing $\boldsymbol{x}^t = \arg\sup_{\boldsymbol{x}\in\mathcal{X}^t}\langle \boldsymbol{x}, \boldsymbol{y}^\star\rangle$[4]. Next, in Section 4, we will continue the discussion by proposing a procedure that returns safe feasibility sets $\mathcal{X}^t$, by exploiting a suitable high probability region $\mathcal{W}^t$ over the vector $\boldsymbol{\omega}^\star$. This construction of $\mathcal{X}^t$ exploits the *polar set* of the high probability region $\mathcal{W}^t$. We will prove that such construction generates a set $\mathcal{X}^t$ that satisfies the assumptions made in Section 3. Finally, in Section 5, we will describe the BAN-SOLO algorithm, based on Lin-UCB, that achieves sub-linear regret in the case in which $\boldsymbol{y}^\star$ and $\boldsymbol{\omega}^\star$ are not known.

3 Rate of Convergence for Linear Programs

In this section we prove the following theorem which relates the Hausdorff distance of the approximate feasibility set $\mathcal{X}^t$ to $\mathcal{X}^\star$ and the difference in the objective of the respective optimization problems.[5]

Theorem 1. *Let $\mathcal{X}^t \subset \mathcal{X}^\star$ and $\mathcal{X}^t \xrightarrow{K(t)} \mathcal{X}^\star$ for some function $K(t) = o(t)$ and bounded set $\mathcal{X}^\star \subset \mathbb{R}^n$. Then $\sup_{x\in\mathcal{X}^t}\langle \boldsymbol{x}, \boldsymbol{y}^\star\rangle \xrightarrow{K'(t)} \sup_{x\in\mathcal{X}^\star}\langle \boldsymbol{x}, \boldsymbol{y}^\star\rangle$, where $K'(t) = 3K(t)$.*

Moreover, for ease of notation, we introduce the following functions: $g^t(\boldsymbol{x}) = -\langle \boldsymbol{x}, \boldsymbol{y}^\star\rangle + \delta_{\mathcal{X}^t}(\boldsymbol{x})$, $g^\star(\boldsymbol{x}) = -\langle \boldsymbol{x}, \boldsymbol{y}^\star\rangle + \delta_{\mathcal{X}^\star}(\boldsymbol{x})$, and their epigraphs $E^t = epi(g^t)$, $E^\star = epi(g^\star)$, while H^+ is the epigraph of the linear function $-\langle \cdot, \boldsymbol{y}^\star\rangle$.[6] Figure 1 is a schematic representation of the epigraphs of the function involved.

3.1 Sketch of the Proof of Theorem 1

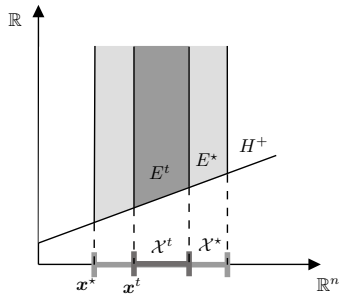

Figure 1: Schematic representation of the epigraphs E^t and $E^\star$ of the functions g^t and $g^\star$.

In this section we report a sketch of the proof that provides the intuition of the techniques used. The idea of the proof can be divided in two main points. **(i)** Relate the Hausdorff distance of $\mathcal{X}^t$ and $\mathcal{X}^\star$ to the Hausdorff distance of the epigraphs of g^t and $g^\star$. **(ii)** Relate the Hausdorff distance of the two epigraphs to the difference in values of their optima. Intuitively we can understand why this point of view is convenient. The epigraph of g^t and $g^\star$ differ only in terms of the epigraph of the indicator functions of the feasibility sets $\delta_{\mathcal{X}^t}$ and $\delta_{\mathcal{X}^\star}$. Indeed, the epigraph E^t and $E^\star$ can be expressed as intersection of the half-space H^+ and the epigraph of the indicator functions $\delta_{\mathcal{X}^t}$ and $\delta_{\mathcal{X}^\star}$, respectively.

This observation suggest to solve point **(i)** by proving that $d_H(E^t, E^\star) \approx d_H(\mathcal{X}^t, \mathcal{X}^\star)$. On the other hand we can see that the optimal value of functions g^t and $g^\star$ can be written as the distance of their epigraphs E^t and $E^\star$ to the optimal value $\boldsymbol{x}^t \in \arg\min g^t$ and $\boldsymbol{x}^\star \in \arg\inf g^\star$, when $\boldsymbol{x}^t$ and $\boldsymbol{x}^\star$ are tough as embedded in $\mathbb{R}^{n+1}$. This helps solving point **(ii)** of the proof, that informally states that $\inf g^t - \inf g^\star \approx d_H(E^t, E^\star)$. Finally, by combining **(i)** and **(ii)**, we can conclude that $\inf g^t - \inf g^\star \approx d_H(\mathcal{X}^t, \mathcal{X}^\star)$.

4 Exploiting Convex Duality to Deal with Uncertainty on $\boldsymbol{\omega}^\star$

In this section we drop the assumption made in the previous section about the existence of an oracle that provides the feasibility sets and propose an explicit construction of such region. More precisely, we loosen the assumptions and we assume only to have a high confidence region $\mathcal{W}^t$ for the vector $\boldsymbol{\omega}^\star$ and define the sets $\mathcal{X}^t$ and $\mathcal{X}^\star$ as polarization of the sets $\mathcal{W}_t$ and $\mathcal{W}^\star$, respectively:

$$\mathcal{X}^t = \left\{\boldsymbol{x} \in \mathcal{X} | \langle \boldsymbol{x}, \boldsymbol{\omega}\rangle \in \mathcal{C}\ \forall \boldsymbol{\omega} \in \mathcal{W}^t\right\}, \quad \mathcal{X}^\star = \left\{\boldsymbol{x} \in \mathcal{X} | \langle \boldsymbol{x}, \boldsymbol{\omega}^\star\rangle \in \mathcal{C}\ \right\}. \tag{1}$$

We use the partial feedback ℓ^t to estimate $\boldsymbol{\omega}^\star$ by building a region in which such parameter lies with high probability. In practice this consists in having an high confidence region $\mathcal{W}^t$ that converges to

[4] Note that it is customary in sensitivity analysis to investigate convergence of the $\inf$ of an objective function f. It is possible to trivially obtain our case simply by considering a different objective function $g = -f$

[5] Proofs missing from Section 3 Section 4 and Section 5 can be found in Appendix B

[6] Note that H^+ is also an half-space.

the singleton $\mathcal{W}^\star := \{\boldsymbol{\omega}^\star\}$ with some convergence rate $K(t) = o(t)$, namely $\mathcal{W}^t \xrightarrow{K(t)} \mathcal{W}^\star$. Under these assumptions, and by defining $\mathcal{C} = [\alpha, \beta]$ we can state the following result.

Theorem 2. *Assume that $\mathcal{W}^t$ is a bounded linear set so that, with probability at least $1-\delta$ we have $\mathcal{W}^\star \subset \mathcal{W}^t$, and that $\mathcal{W}^t \xrightarrow{K(t)} \mathcal{W}^\star$ for some function $K(t) = o(t)$. Define $\mathcal{C} := [\alpha, \beta]$ and $\mathcal{X}^t$ and $\mathcal{X}^\star$ as in Equation 1. Then $\sup_{x \in \mathcal{X}^t} \langle \boldsymbol{x}, \boldsymbol{y}^\star \rangle \xrightarrow{K'(t)} \sup_{x \in \mathcal{X}^\star} \langle \boldsymbol{x}, \boldsymbol{y}^\star \rangle$, where $K'(t) = 3R^2 K(t)/\psi$ with $R = \sup_{\boldsymbol{x}, \boldsymbol{x}' \in \mathcal{X}^\star \times \mathcal{X}^\star} ||\boldsymbol{x} - \boldsymbol{x}'||_2$ and $\psi = \min\{|\alpha|, |\beta|\}$.*

5 Algorithm

Algorithm 1 BAN-SOLO

1: **for** $t \in \{1, \dots, T\}$ **do**
2: Build confidence regions $\mathcal{Y}^t$ and $\mathcal{W}^t$ from past feedback $\{\ell^1, \dots, \ell^{t-1}\}$
3: Build linear set $\mathcal{X}^t$ by exploiting $\mathcal{W}^t$ ▷ Subsection 4
4: $\boldsymbol{x}^t \leftarrow \arg \sup_{\boldsymbol{x} \in \mathcal{X}^t} \sup_{\boldsymbol{y} \in \mathcal{Y}^t} \langle \boldsymbol{x}, \boldsymbol{y} \rangle$
5: Play the game according to strategy $\boldsymbol{x}^t$
6: Observe feedback ℓ^t from the environment

In the previous section we showed how the assumption of Section 3 of having a oracle returning the feasibility set $\mathcal{X}^t$ can be substituted by a high probability region $\mathcal{W}^t$ for $\boldsymbol{\omega}^\star$. Moreover, while up to now the results were given for a known $\boldsymbol{y}^\star$, in this section we will present a practical algorithm that by having a high probability region $\mathcal{Y}^t$ achieves sub-linear regret, given that $\mathcal{Y}^t \xrightarrow{\mathcal{O}(1/\sqrt{t})} \mathcal{Y}^\star$. In order to guarantee that BAN-SOLO attains sub-linear regret $R^T = o(T)$ after T iterations (with high probability), we adopt an approach inspired from *linear* multi-armed bandit problems [Abbasi-Yadkori et al., 2011]. The pseudo code for the BAN-SOLO algorithm is provided in Algorithm 1. It is based on the *optimism in face of uncertainty* principle, and, thus, it selects a strategy $\boldsymbol{x}^t \in \mathcal{X}^t$ that maximizes the expected payoff $\langle \boldsymbol{x}, \boldsymbol{y} \rangle$ under the assumption that, for every $\boldsymbol{x} \in \mathcal{X}^t$, strategy $\boldsymbol{y}$ is an optimistic estimate of $\boldsymbol{y}^\star$ taken from the confidence region $\mathcal{Y}^t$, that is, $\boldsymbol{y} \in \mathcal{Y}^t$ maximizes the same player i's expected payoff $\langle \boldsymbol{x}, \boldsymbol{y} \rangle$ (line 4). The only assumption is that one can build the high confidence region $\mathcal{Y}^t$ (line 2). The following theorem provides an high-probability sub-linear regret guarantee for BAN-SOLO (the complete proof can be found in Appendix D).

Theorem 3. *If $\mathcal{W}^t \xrightarrow{K(\delta)/\sqrt{t}} \mathcal{W}^\star$ and $\mathcal{Y}^t \xrightarrow{K(\delta)/\sqrt{t}} \mathcal{Y}^\star$, and both $\mathcal{W}^\star \subset \mathcal{W}^t$ and $\mathcal{Y}^\star \subset \mathcal{Y}^t$ holds with probability at least $1-\delta$, then we have that $R^T \leq \left(3R^2/\psi + D\right) K(\delta)\sqrt{T}$, where ψ and R are defined as in Theorem 2 and $D = \sup_{\mathcal{X}^\star} ||\boldsymbol{x}||$.*

Note that in Theorem3 the parameter K is written as a function of the parameter δ, because, in general, to guarantee $\boldsymbol{\omega}^\star \in \mathcal{W}^t$ with higher probability (smaller δ) one needs a larger $\mathcal{W}^t$ (larger K). Moreover, notice that BAN-SOLO needs to solve a linearly-constrained *bilinear* optimization problem at each iteration, which can be done efficiently by cutting-hedge solvers. Theorem 3 shows that one can achieve sub-linear regret just by having good (e.g., of order $1/\sqrt{t}$) estimators for the parameters $\boldsymbol{y}^\star$ and $\boldsymbol{\omega}^\star$. Due to space constraints, we defer to the appendix the application of BAN-SOLO to common strategic scenarios, like normal-form and sequential decision making.

6 Conclusions

In this work we study the problem of playing safely decision making problems with bandit feedback and linear cost and payoff functions. We first derived a general result on the the sensitivity of a linear optimization problem based on their feasibility set and their Hausdorff distance. Then we showed how such result can be used in our case, by exploiting the polar set formulation that follows from the safety constraints. Finally we proposed BAN-SOLO, a Lin-UCB algorithm, which guarantees sub-linear regret and, at the same time, the satisfaction with high probability of the safety constraints during the entire learning process.

References

Y. Abbasi-Yadkori, D. Pál, and C. Szepesvári. Improved algorithms for linear stochastic bandits. *Advances in neural information processing systems*, 24:2312–2320, 2011.

M. Bernasconi, F. Cacciamani, M. Castiglioni, A. Marchesi, N. Gatti, and F. Trovò. Safe learning in tree-form sequential decision making: Handling hard and soft constraints. In *International Conference on Machine Learning*, pages 1854–1873. PMLR, 2022.

M. Bernasconi, M. Castiglioni, A. Marchesi, N. Gatti, and F. Trovò. Sequential information design: Learning to persuade in the dark. *Advances in Neural Information Processing Systems*, 35, 2022.

M. Bernasconi-de-Luca, F. Cacciamani, S. Fioravanti, N. Gatti, A. Marchesi, and F. Trovò. Exploiting opponents under utility constraints in sequential games. *Advances in Neural Information Processing Systems*, 34, 2021.

N. Brown and T. Sandholm. Superhuman ai for heads-up no-limit poker: Libratus beats top professionals. *Science*, 359(6374):418–424, 2018.

L. Devroye. The equivalence of weak, strong and complete convergence in l1 for kernel density estimates. *The Annals of Statistics*, pages 896–904, 1983.

L. Dümbgen and G. Walther. Rates of convergence for random approximations of convex sets. *Advances in applied probability*, 28(2):384–393, 1996.

G. Farina, R. Schmucker, and T. Sandholm. Bandit linear optimization for sequential decision making and extensive-form games. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 35, pages 5372–5380, 2021.

J. L. Higle and S. W. Wallace. Sensitivity analysis and uncertainty in linear programming. *Interfaces*, 33(4):53–60, 2003.

S. Hoda, A. Gilpin, J. Pena, and T. Sandholm. Smoothing techniques for computing nash equilibria of sequential games. *Mathematics of Operations Research*, 35(2):494–512, 2010.

D. Silver, J. Schrittwieser, K. Simonyan, I. Antonoglou, A. Huang, A. Guez, T. Hubert, L. Baker, M. Lai, A. Bolton, et al. Mastering the game of go without human knowledge. *nature*, 550(7676): 354–359, 2017.

D. Silver, T. Hubert, J. Schrittwieser, I. Antonoglou, M. Lai, A. Guez, M. Lanctot, L. Sifre, D. Kumaran, T. Graepel, et al. A general reinforcement learning algorithm that masters chess, shogi, and go through self-play. *Science*, 362(6419):1140–1144, 2018.

M. Tambe. *Security and game theory: algorithms, deployed systems, lessons learned*. Cambridge university press, 2011.

G. M. Ziegler. *Lectures on polytopes*, volume 152. Springer Science & Business Media, 2012.

A Structure of the Appendix

The appendix is structured as follows:

- Appendix B provides the proof of the main theorem of Section 3.
- Appendix C provides the proofs omitted from Section 4.
- Appendix D provides the proofs omitted from Section 5.
- Appendix E provides applications of our methodology to the context of games.

B Proofs Omitted from Section 3

In this section we are going to formally write the intermediate results needed to complete the informal discussion on the proof of Theorem 1. The relationship between the distance of the feasibility sets $\mathcal{X}^t$ and $\mathcal{X}^\star$ and the epigraphs of g^t and $g^\star$ is formalized in the following lemma.

Lemma 2. *Let A, B and C some empty sets of a metric space $(\mathcal{M}, d)$. Then we have the following $d_H(A \cap B, A \cap C) \leq 3d_H(B, C)$.*

Proof. First note that $(A \cap B)_\epsilon \subset (A_\epsilon \cap B_\epsilon)$. By definition we have that for $x \in (A_\epsilon \cap B_\epsilon)$ then $d(x, A) < \epsilon$ and $d(x, B) < \epsilon$. Let's now consider the following inequalities for $x \in (A_\epsilon \cap B_\epsilon)$:

$$\begin{aligned} d(x, (A \cap B)_\epsilon) &:= \inf_{y \in (A \cap B)_\epsilon} d(x, y) \\ &\leq d(x, P_A(x)) + \inf_{y \in (A \cap B)_\epsilon} d(P_A(x), y), \end{aligned} \tag{2}$$

where $P_A(x) := \arg\inf_{z \in A} d(x, z)$.

Since $x \in (A_\epsilon \cap B_\epsilon)$ the first term of Equation (2) can be bounded by ϵ since $d(x, z) < \epsilon$ for all $z \in A$. For the second term of Equation (2) we can consider the following chain of inequalities. Consider $\bar{x} \in (A \cap B)_\epsilon$ and

$$||\bar{x} - x||^2 = ||\bar{x} - P_A(x) + P_A(x) - x||^2 \tag{3}$$

$$\geq ||\bar{x} - P_A(x)||^2 + ||P_A(x) - x||^2, \tag{4}$$

which in turn implies that $||\bar{x} - P_A(x)||^2 \leq ||\bar{x} - x||^2$. From this we can conclude that $d(P_A(x), (A \cap B)_\epsilon) < \epsilon$. We can thus conclude that $d(x, (A \cap B)_\epsilon) < 2\epsilon$ and since $(A \cap B)_\epsilon \subset (A_\epsilon \cap B_\epsilon)$ we can infer that:

$$d_H((A \cap B)_\epsilon, A_\epsilon \cap B_\epsilon) < 2\epsilon.$$

Hence we have the following set inclusions:

$$(A \cap B)_\epsilon \subset A_\epsilon \cap B_\epsilon \subset ((A \cap B)_\epsilon)_{2\epsilon} \subset (A \cap B)_{3\epsilon}.$$

Finally consider the following inequalities:

$$d_H(A \cap B, A \cap C) := \inf_{\epsilon > 0}((A \cap B) \subset (A \cap C)_\epsilon \text{ and } (A \cap C) \subset (A \cap B)_\epsilon) \tag{5}$$

$$\leq \inf_{\epsilon > 0}((A \cap B) \subset A_{\epsilon/3} \cap C_{\epsilon/3} \text{ and } (A \cap C) \subset A_{\epsilon/3} \cap B_{\epsilon/3}) \tag{6}$$

$$= \inf_{\epsilon > 0}(B \subset C_{\epsilon/3} \text{ and } C \subset B_{\epsilon/3}) = 3d_H(C, B), \tag{7}$$

where the inequality comes from the fact that $(A_{\epsilon/3} \cap C_{\epsilon/3}) \subset (A \cap C)\epsilon$ and similarly with the set B instead of the set C. □

This result states that factoring out a common set A, does only change the Hausdorff distance up to a constant multiplicative factor. On the other hand the following lemma relates the distance in epigraph and the distance in optimal values of the functions.

Lemma 3. *Let $E^{(1)}$ and $E^{(2)}$ be the epigraphs of two linear programs described by the functions $f^{(1)}$ and $f^{(2)}$. Then we have that $\left|\inf f^{(1)} - \inf f^{(2)}\right| \leq d_H\left(E^{(1)}, E^{(2)}\right)$.*

Proof. Define $E^{(1)} := epi(f^{(1)})$, $E^{(2)} := epi(f^{(2)})$ and $\boldsymbol{x}^{(1)} \in \arg\inf f^{(1)}$. Moreover one has that $\inf f^{(1)} = d(\boldsymbol{x}^{(1)}, E^{(1)})$ where, with abuse of notation we denoted with $\boldsymbol{x}^{(1)}$ the vector $(\boldsymbol{x}^{(1)}, 0) \in \mathbb{R}^{n+1}$. Similarly we have $\inf f^{(2)} = d(\boldsymbol{x}^{(2)}, E^{(2)})$. Consider now the following inequalities:

$$\inf f^{(1)} - \inf f^{(2)} = d(\boldsymbol{x}^{(1)}, E^{(1)}) - d(\boldsymbol{x}^{(2)}, E^{(2)}) \tag{8}$$

$$\leq d(\boldsymbol{x}^{(2)}, E^{(1)}) - d(\boldsymbol{x}^{(2)}, E^{(2)}) \tag{9}$$

$$\leq d_H(E^{(1)}, E^{(2)}). \tag{10}$$

This concludes the proof of the statement. □

This result can be seen as a case of sensitivity analysis of linear programs Higle and Wallace [2003] in which we exploited the Hausdorff distance between their epigraph. To the best of our knowledge, such general sensitivity analysis is not found elsewhere in the literature, but due to its simplicity, we do not claim it to be novel. Finally, by using Lemma 3 and Lemma 2 one can easily prove Theorem 1.

Theorem 1. *Let $\mathcal{X}^t \subset \mathcal{X}^\star$ and $\mathcal{X}^t \xrightarrow{K(t)} \mathcal{X}^\star$ for some function $K(t) = o(t)$ and bounded set $\mathcal{X}^\star \subset \mathbb{R}^n$. Then $\sup_{x\in\mathcal{X}^t}\langle \boldsymbol{x}, \boldsymbol{y}^\star\rangle \xrightarrow{K'(t)} \sup_{x\in\mathcal{X}^\star}\langle \boldsymbol{x}, \boldsymbol{y}^\star\rangle$, where $K'(t) = 3K(t)$.*

Proof. Note that one can express $E^t = H^+ \cap A^t$ where $A^t = epi(\delta_{\mathcal{X}^t})$. Similarly $E^\star = H^+ \cap A^\star$, where $A^\star = epi(\delta_{\mathcal{X}^\star})$. Consider now the following inequality:

$$\begin{aligned} d_H(E^t, E^\star) &= d_H(H^+ \cap A^t, H^+ \cap A^\star) \\ &\leq 3d_H(A^t, A^\star) = 3d_H(\mathcal{X}^t, \mathcal{X}^\star), \end{aligned}$$

where we used Lemma 2 for the first inequality, while the last equality comes from the fact that A^t and $A^\star$ are epigraphs of indicator functions over $\mathcal{X}^t$ and $\mathcal{X}^\star$, respectively. Hence the Hausdorff distance between $A^t, A^\star \subset \mathbb{R}^{n+1}$ is equivalent to the distance on the plane $\mathbb{R}^n \times 0 \subset \mathbb{R}^{n+1}$, which is in turn equivalent to the Hausdorff distance between $\mathcal{X}^t, \mathcal{X}^\star \subset \mathbb{R}^n$. We can then conclude that $d_H(E^t, E^\star) \leq 3K(t)$, since by assumption, we have that $d_H(\mathcal{X}^t, \mathcal{X}^\star) \leq K(t)$. Then, from Lemma 3 we have that:

$$\left|\inf g^t(\boldsymbol{x}) - \inf g^\star(\boldsymbol{x})\right| \leq d_H\left(E^t, E^\star\right) \leq 3K(t). \tag{11}$$

The theorem follows from the fact that $\inf g^t = \sup_{\boldsymbol{x}\in\mathcal{X}^t}\langle \boldsymbol{x}, \boldsymbol{y}^\star\rangle$ and $\inf g^\star = \sup_{\boldsymbol{x}\in\mathcal{X}^\star}\langle \boldsymbol{x}, \boldsymbol{y}^\star\rangle$ □

C Proofs omitted from Section 4

Theorem 2. *Assume that $\mathcal{W}^t$ is a bounded linear set so that, with probability at least $1-\delta$ we have $\mathcal{W}^\star \subset \mathcal{W}^t$, and that $\mathcal{W}^t \xrightarrow{K(t)} \mathcal{W}^\star$ for some function $K(t) = o(t)$. Define $\mathcal{C} := [\alpha, \beta]$ and $\mathcal{X}^t$ and $\mathcal{X}^\star$ as in Equation 1. Then $\sup_{x\in\mathcal{X}^t}\langle \boldsymbol{x}, \boldsymbol{y}^\star\rangle \xrightarrow{K'(t)} \sup_{x\in\mathcal{X}^\star}\langle \boldsymbol{x}, \boldsymbol{y}^\star\rangle$, where $K'(t) = 3R^2K(t)/\psi$ with $R = \sup_{\boldsymbol{x},\boldsymbol{x}'\in\mathcal{X}^\star\times\mathcal{X}^\star} ||\boldsymbol{x} - \boldsymbol{x}'||_2$ and $\psi = \min\{|\alpha|, |\beta|\}$.*

Proof. In order to show the convergence of the LP solution we need to prove that $\mathcal{X}^t \xrightarrow{R^2K(t)/\psi} \mathcal{X}^\star$ and then apply Theorem 1 to obtain the statement. Thus we need to study the relation between the convergence rates of $\mathcal{W}^t$ and of $\mathcal{X}^t$. Define $\mathcal{C} = [\alpha, \beta]$ with $-\infty < \alpha < 0 < \beta < +\infty$.[7] Since the set $\mathcal{Y}^t$ is assumed to be a bounded polytope it can be expressed as the convex hull of the set of vertices $\mathcal{V}^t$ that define the high probability region for $\boldsymbol{y}^\star$, formally: $\mathcal{Y}^t = \text{hull}(\mathcal{V}^t)$, where $\mathcal{V}^t = \{\boldsymbol{y}_1, ..., \boldsymbol{y}_d\}$. We will exploit Lemma 3, in Dümbgen and Walther [1996], which gives bounds on the Hausdorff distance between sets and their corresponding polar sets. Formally it states that:

$$d_H(\mathcal{X}^\star, \mathcal{X}^t) \leq R^2 d_H(\mathcal{X}^{\star,\circ}, \mathcal{X}^{t,\circ}), \tag{12}$$

where R is the diameter of $\mathcal{X}^\star$, defined as the radius of the smallest ball centered in zero that contains $\mathcal{X}^\star$. Thus, in order to bound the Hausdorff distance between $\mathcal{X}^t$ and $\mathcal{X}^\star$ we have to study the distance

[7] The proof can be trivially adapted in the case of α and β have the same sign or in the case of $\alpha = -\infty$ or $\beta = +\infty$.

between their polar sets. Given Lemma 1 it is possible to derive an expression for the polar sets $\mathcal{X}^{t,\circ}$ and $\mathcal{X}^{\star,\circ}$:

$$\mathcal{X}^{\star,\circ} = \text{hull}\left(\{\boldsymbol{x}\}_{\boldsymbol{x}\in\text{vert}(\mathcal{X}^\circ)} \cup \left\{\frac{\boldsymbol{y}^\star}{\alpha}, \frac{\boldsymbol{y}^\star}{\beta}\right\}\right),$$

$$\mathcal{X}^{t,\circ} = \text{hull}\left(\{\boldsymbol{x}\}_{\boldsymbol{x}\in\text{vert}(\mathcal{X}^\circ)} \cup \left\{\frac{\boldsymbol{y}^t}{\alpha}\right\}_{\boldsymbol{y}^t\in\mathcal{V}^t} \cup \left\{\frac{\boldsymbol{y}^t}{\beta}\right\}_{\boldsymbol{y}^t\in\mathcal{V}^t}\right),$$

where $\text{vert}(\mathcal{X}^\circ)$ is the set of vertices of $\mathcal{X}^\circ$.

An equivalent definition of the Hausdorff distance between $\mathcal{X}^{\star,\circ}$ and $\mathcal{X}^{t,\circ}$ is:

$$d_H(\mathcal{X}^{\star,\circ}, \mathcal{X}^{t,\circ}) = \max\left\{\sup_{\boldsymbol{x}\in\mathcal{X}^{\star,\circ}} d(\boldsymbol{x}, \mathcal{X}^{t,\circ}), \sup_{\boldsymbol{x}\in\mathcal{X}^{t,\circ}} d(\boldsymbol{x}, \mathcal{X}^{\star,\circ})\right\}.$$

Let us consider $\sup_{\boldsymbol{x}\in\mathcal{X}^{\star,\circ}} d(\boldsymbol{x}, \mathcal{X}^{t,\circ})$. The sup can be tackled by considering exclusively the vertices of $\mathcal{X}^{\star,\circ}$, thus:

$$\sup_{\boldsymbol{x}\in\mathcal{X}^{\star,\circ}} d(\boldsymbol{x}, \mathcal{X}^{t,\circ}) = \sup_{\boldsymbol{x}\in\text{vert}(\mathcal{X}^{\star,\circ})} d(\boldsymbol{x}, \mathcal{X}^{t,\circ}),$$

where $\text{vert}\,(\mathcal{X}^{\star,\circ}) = \{\boldsymbol{x}\}_{\boldsymbol{x}\in\text{vert}(\mathcal{X}^\circ)} \cup \left\{\frac{\boldsymbol{y}^\star}{\alpha}, \frac{\boldsymbol{y}^\star}{\beta}\right\}$ is the set of vertices of the convex set $\mathcal{X}^{\star,\circ}$.

Considering that the two sets $\mathcal{X}^{\star,\circ}$ and $\mathcal{X}^{t,\circ}$ have the vertices of $\mathcal{X}^\circ$ in common, we can ignore such vertices as they are characterized by 0 distance. Hence:

$$\sup_{\boldsymbol{x}\in\mathcal{X}^{\star,\circ}} d(\boldsymbol{x}, \mathcal{X}^{t,\circ}) = \sup_{\boldsymbol{x}\in\left\{\frac{\boldsymbol{y}^\star}{\alpha}, \frac{\boldsymbol{y}^\star}{\beta}\right\}} d(\boldsymbol{x}, \mathcal{X}^{t,\circ}).$$

Now let us consider $d(\frac{\boldsymbol{y}^\star}{\alpha}, \mathcal{X}^{t,\circ}) := \inf_{\boldsymbol{x}\in\mathcal{X}^{t,\circ}} d(\frac{\boldsymbol{y}^\star}{\alpha}, \boldsymbol{x})$. Since the set $\mathcal{X}^{t,\circ}$ is linear we can upper bound the distance with the distance computed with respect to the vertices. Hence we have that:

$$\begin{aligned} d\left(\frac{\boldsymbol{y}^\star}{\alpha}, \mathcal{X}^{t,\circ}\right) &\leq \min\left\{\inf_{\boldsymbol{y}^t\in\mathcal{V}^t} d\left(\frac{\boldsymbol{y}^\star}{\alpha}, \frac{\boldsymbol{y}^t}{\alpha}\right), \inf_{\boldsymbol{y}^t\in\mathcal{V}^t} d\left(\frac{\boldsymbol{y}^\star}{\alpha}, \frac{\boldsymbol{y}^t}{\beta}\right)\right\} \\ &\leq \inf_{\boldsymbol{y}^t\in\mathcal{V}^t} d\left(\frac{\boldsymbol{y}^\star}{\alpha}, \frac{\boldsymbol{y}^t}{\alpha}\right). \end{aligned}$$

Where the inequality comes from the fact that $min(a, b) \leq a$. Now we can bound the last term as follows:

$$d\left(\frac{\boldsymbol{y}^\star}{\alpha}, \frac{\boldsymbol{y}^t_i}{\alpha}\right) = \frac{1}{|\alpha|} d(\boldsymbol{y}^\star, \boldsymbol{y}^t_i) \leq \frac{K(t)}{|\alpha|},$$

since by assumption we had that $d_H(\mathcal{Y}^\star, \mathcal{Y}^t) \leq K(t)$ and that $\mathcal{Y}^\star$ is a singleton. With a similar reasoning we can conclude that:

$$d\left(\frac{\boldsymbol{y}^\star}{\beta}, \mathcal{X}^{t,\circ}\right) \leq \frac{K(t)}{|\beta|},$$

and it follows that:

$$\sup_{\boldsymbol{x}\in\mathcal{X}^{\star,\circ}} d(\boldsymbol{x}, \mathcal{X}^{t,\circ}) \leq \max\left\{\frac{K(t)}{|\alpha|}, \frac{K(t)}{|\beta|}\right\}.$$

Similarly it is possible to derive a bound for $\sup_{\boldsymbol{x}\in\mathcal{X}^{t,\circ}} d(\boldsymbol{x}, \mathcal{X}^{\star,\circ})$:

$$\sup_{\boldsymbol{x}\in\mathcal{X}^{t,\circ}} d(\boldsymbol{x}, \mathcal{X}^{\star,\circ}) \leq \max\left\{\frac{K(t)}{|\alpha|}, \frac{K(t)}{|\beta|}\right\},$$

and thus:

$$d_H(\mathcal{X}^{\star,\circ}, \mathcal{X}^{t,\circ}) \leq \frac{K(t)}{\psi}, \tag{13}$$

where $\psi = \min\{|\alpha|, |\beta|\}$. Now we can plug Equation (13) into Equation (12) and prove that $d_H(\mathcal{X}^\star, \mathcal{X}^t) \leq R^2 \frac{K(t)}{\psi}$. By using Theorem 1 we can conclude the proof. □

D Proofs omitted from Section 5

Theorem 3. *If $\mathcal{W}^t \xrightarrow{K(\delta)/\sqrt{t}} \mathcal{W}^\star$ and $\mathcal{Y}^t \xrightarrow{K(\delta)/\sqrt{t}} \mathcal{Y}^\star$, and both $\mathcal{W}^\star \subset \mathcal{W}^t$ and $\mathcal{Y}^\star \subset \mathcal{Y}^t$ holds with probability at least $1-\delta$, then we have that $R^T \leq \left(3R^2/\psi + D\right) K(\delta)\sqrt{T}$, where ψ and R are defined as in Theorem 2 and $D = \sup_{\mathcal{X}^\star} ||\boldsymbol{x}||$.*

Proof. We will use the following instantaneous regret decomposition:

$$\begin{aligned} r_t &= \sup_{\boldsymbol{x}\in\mathcal{X}^\star} \langle \boldsymbol{x}, \boldsymbol{y}^\star\rangle - \langle \boldsymbol{x}^t, \boldsymbol{y}^\star\rangle \\ &= \underbrace{\sup_{\boldsymbol{x}\in\mathcal{X}^\star} \langle \boldsymbol{x}, \boldsymbol{y}^\star\rangle - \sup_{\boldsymbol{x}\in\mathcal{X}^t} \langle \boldsymbol{x}, \boldsymbol{y}^\star\rangle}_{:=r_t^A} + \underbrace{\sup_{\boldsymbol{x}\in\mathcal{X}^t} \langle \boldsymbol{x}, \boldsymbol{y}^\star\rangle - \langle \boldsymbol{x}^t, \boldsymbol{y}^\star\rangle}_{:=r_t^B}, \end{aligned}$$

which decompose the regret in the regret due to the fact that we have to guarantee safety (r_t^A) and $\boldsymbol{y}^\star$ is unknown (r_t^B). Now we are going to consider the two terms of the instantaneous regret separately.

Bounding r_t^A: From Theorem 2 we have that $r_t^A \leq \frac{3R^2K}{\phi\sqrt{t}}$, where $\psi = \min\{|\alpha|, |\beta|\}$.

Bounding r_t^B: Define $(\boldsymbol{x}^t, \boldsymbol{y}^t) := \arg\left(\sup_{\boldsymbol{x}\in\mathcal{X}^t}\sup_{\boldsymbol{y}\in\mathcal{Y}^t} \langle \boldsymbol{x}, \boldsymbol{y}\rangle\right)$. Then with probability at least $1-\delta$ we have that:

$$\begin{aligned} r_t^B &:= \sup_{\boldsymbol{x}\in\mathcal{X}^t} \langle \boldsymbol{x}, \boldsymbol{y}^\star\rangle - \langle \boldsymbol{x}^t, \boldsymbol{y}^\star\rangle && (14)\\ &\leq \langle \boldsymbol{x}^t, \boldsymbol{y}^t - \boldsymbol{y}^\star\rangle && (15)\\ &\leq ||\boldsymbol{x}^t||_2 ||\boldsymbol{y}^t - \boldsymbol{y}^\star||_2 && (16)\\ &\leq D\frac{K}{\sqrt{t}}, && (17) \end{aligned}$$

where the first inequality follows from the fact that $(\boldsymbol{x}^t, \boldsymbol{y}^t)$ is a pessimistic estimate of the payoff, the second inequality comes from the Cauchy-Schwartz inequality while the last inequality come from the fact that $\mathcal{Y}^t \xrightarrow{K/\sqrt{t}} \mathcal{Y}^\star$. D is defined as $D := \sup_{\boldsymbol{x}\in\mathcal{X}^\star} ||\boldsymbol{x}||_2$, which upper bounds $||\boldsymbol{x}||_2$ for all $\boldsymbol{x} \in \mathcal{X}^t$ as $\mathcal{X}^t \subset \mathcal{X}^\star$.

Hence to conclude we have that:

$$r_t \leq \left(\frac{3R^2}{\psi} + D\right)\frac{K}{\sqrt{t}},$$

which implies that $R^T \leq \left(\frac{3R^2}{\psi} + D\right) K\sqrt{T}$, hence concluding the proof. □

E Applications

Normal Form Games A normal form game against a stochastic player is the simplest case of decision making, in which we have a single initial state, N^2 terminal states, where N is the dimension of the action space. Indeed, we have a stochastic adversary that picks an action $\boldsymbol{y}^t \in \{\boldsymbol{e}_1, \dots, \boldsymbol{e}_N\}$, with a categorical distribution defined by $\boldsymbol{y}^\star \in \Delta^N$, where Δ^N is the $N-1$ dimensional simplex. In turn the agent picks an action $\boldsymbol{x}^t \in \{\boldsymbol{e}_1, \dots, \boldsymbol{e}_N\}$ and gets a utility of $\langle \boldsymbol{x}^t, U\boldsymbol{y}^t\rangle$ where $U \in \mathbb{R}^{N\times N}$ is the payoff matrix and a cost of $\langle \boldsymbol{x}^t, B\boldsymbol{y}^t\rangle$ where $B \in \mathbb{R}^{N\times N}$ is the costs matrix. In this setting, the feedback ℓ^t correspond to the actions $\boldsymbol{y}^t$ played by the adversary. Thus by observing $\boldsymbol{y}^1, \dots, \boldsymbol{y}^t$ one we can build a confidence polytope for $\boldsymbol{y}^\star$, trough the concentration inequality of [Devroye, 1983], that states that for any $\delta \leq 3e^{-4N/5}$ we have that with probability at least $1-\delta$, $\boldsymbol{y}^\star$ lies in the set: $\mathcal{Y}^t := \left\{\boldsymbol{y} \,\middle|\, ||\boldsymbol{y} - U\hat{\boldsymbol{y}}^t|| \leq \sqrt{2\log(1/\delta)/t}\right\}$, where $\hat{\boldsymbol{y}}^t$ is the empirical mean after t turns. Finally $\mathcal{W}^t$ can be defined similarly.

Sequential Decision Making Not in all application we have that ℓ^t coincide with $\boldsymbol{y}^t$. An example of when this does not happen are sequential decision making problems (SDM) Farina et al. [2021]. A SDM against a stochastic adversary, is a special case instance of our model that allows to model imperfect information game such as Poker Brown and Sandholm [2018], in which the strategy space $\mathcal{X}$ and $\mathcal{Y}$ can be described by *treeplexes* [Hoda et al., 2010], which allows to have a linear structure for the payoffs, as the case for NFG described above. The main difference between NFG and SDM is the construction of the confidence set $\mathcal{Y}^t$. In the case of SDM, the feedback $\tilde{\boldsymbol{y}}^t$, only traverses a subset of states at each turn, and the set $\mathcal{Y}^t$ can only be updated at such states. This problem is called Opponent Modelling and it can be shown [Bernasconi-de-Luca et al., 2021, Bernasconi et al., 2022, Bernasconi et al., 2022] that both $\mathcal{Y}^t$ and $\mathcal{W}^t$ can be obtained from past feedback $\{\ell^t\}_t$.