Mathematics of Operations Research

Publication details, including instructions for authors and subscription information:
http://pubsonline.informs.org

Opportunistic Approachability and Generalized No-Regret Problems

Andrey Bernstein, Shie Mannor, Nahum Shimkin

To cite this article:
Andrey Bernstein, Shie Mannor, Nahum Shimkin (2014) Opportunistic Approachability and Generalized No-Regret Problems. Mathematics of Operations Research 39(4):1057-1083. https://doi.org/10.1287/moor.2014.0643

MATHEMATICS OF OPERATIONS RESEARCH
Vol. 39, No. 4, November 2014, pp. 1057–1083
ISSN 0364-765X (print) | ISSN 1526-5471 (online)

http://dx.doi.org/10.1287/moor.2014.0643

Opportunistic Approachability and Generalized No-Regret Problems

Andrey Bernstein, Shie Mannor, Nahum Shimkin
Technion–Israel Institute of Technology, Haifa 32000, Israel
{andreyb@tx.technion.ac.il, shie@ee.technion.ac.il, shimkin@ee.technion.ac.il}

Blackwell's theory of approachability, introduced in 1956, has since proved a useful tool in the study of a range of repeated multiagent decision problems. Given a repeated matrix game with vector payoffs, a target set S is approachable by a certain player if he can ensure that the average payoff vector converges to that set, for any strategy of the opponent. In this paper we consider the case where a set need not be approachable in general, but may be approached if the opponent played favorably in some sense. In particular, we consider nonconvex sets that satisfy Blackwell's dual condition, namely, can be approached when the opponent plays a stationary strategy. Whereas the convex hull of such a set is approachable, this is not generally the case for the original nonconvex set itself. We start by defining a sense of restricted play of the opponent (with stationary strategies being a special case), and then formulate appropriate goals for an *opportunistic* approachability algorithm that can take advantage of such restricted play as it unfolds during the game. We then consider a calibration-based approachability strategy that is opportunistic in that sense. A major motivation for this study comes from no-regret problems that lack a convex structure such as the problem of online learning with sample-path constraints, as formulated in Mannor et al. [Mannor S, Tsitsiklis JN, Yu JY (2009) Online learning with sample path constraints. *J. Machine Learn. Res.* 10:569–590]. Here the best-response-in-hindsight is not generally attainable, but only a convex relaxation thereof. Our proposed algorithm, while ensuring that relaxed goal, also comes closer to the nonrelaxed one when the opponent's play is restricted in a well-defined sense.

Keywords: Blackwell's approachability; calibrated play; no-regret algorithms
MSC2000 subject classification: Primary: 91A20; secondary: 90B99
OR/MS subject classification: Primary: games/group decisions, stochastic; secondary: decision analysis, sequential
History: Received August 9, 2012; revised April 2013, and October 2013. Published online in *Articles in Advance* April 16, 2014.

1. Introduction.

The concept of set approachability, as introduced in Blackwell [7], concerns a repeated matrix game with vector-valued payoffs that is played by two players, the agent and the opponent. Thus, for each pair of simultaneous actions a and z in the one-stage game, a payoff vector $r(a, z) \in \mathbb{R}^\ell$ is obtained. Given a target set S in $\mathbb{R}^\ell$, the agent's goal is to have the long-term average reward vector *approach* S, namely, converge to S almost surely in the point-to-set distance. If that convergence can be ensured irrespectively of the opponent's strategy, the set S is said to be *approachable*, and a strategy of the agent that satisfies this property is an approaching strategy (or algorithm) for S.

Approachability has close connections with online learning algorithms, and in particular with the notion of no-regret strategies. In fact, soon after no-regret strategies for repeated games were introduced in Hannan [15], it was shown by Blackwell [6] that the problem can be formulated and solved as a particular case of the general approachability problem, for a suitably defined set and payoff vector. An extensive overview of these concepts and their interrelations can be found in Fudenberg and Levine [14], Young [41], and Cesa-Bianchi and Lugosi [9].

By its very definition, the notion of an approachable set accommodates a worst-case scenario, as the target set must be approached for *any* strategy of the opponent. However, as the game unfolds, it may turn out that the temporal variability in the sequence of the opponent's actions is limited in some sense. For example, the opponent may choose to employ a stationary strategy, namely, repeat a single mixed action, or perhaps repeat a certain sequence of actions. If these restrictions were known in advance, the agent could possibly ensure convergence to a target set S that is not approachable in general, or perhaps converge to a subset of the target set S that is deemed more desirable. Our goal here is to formulate *opportunistic* approachability algorithms, in the sense that they may exploit such limitations on the opponent's action sequence in an online manner, *without knowing them beforehand*.

To illustrate the ideas involved, it will be useful to consider here some examples.

Example 1. Consider a scalar reward matrix given by $r(0, 0) = 2$, $r(1, 1) = -2$, and $r(0, 1) = r(1, 0) = 0$, where $\mathcal{A} = \mathcal{Z} = \{0, 1\}$ are the actions available to the agent and the opponent. Suppose the agent's goal is to have its long-term average reward larger or equal to 1 *in absolute value*, namely, $|\bar{R}_n| \geq 1 - o(1)$, where $\bar{R}_n = (1/n)\sum_{k=1}^n r(a_k, z_k)$. This clearly corresponds to an approachability problem, with the *nonconvex* target set $S = (-\infty, -1] \cup [1, \infty)$. Now, it is easily seen that for any mixed action $q = (q(0), q(1))$ of the opponent, the agent has a response $p = (p(0), p(1))$ so that $r(p, q) \triangleq \sum_{a,z} p(a)q(z)r(a, z) \in S$. Thus, if the opponent is restricted a priori to stationary strategies, the agent can easily devise a (possibly adaptive) strategy that approaches S. However, this is clearly not the case in general: for example, the opponent can ensure $\bar{R}_n \to 0$ by

playing $z_n = 0$ whenever $\bar{R}_{n-1} < 0$, and $z_n = 1$ otherwise. We see that the agent cannot approach the required target set S in general, but can hope to do so if the opponent happens to play a stationary strategy.

Example 1 satisfies the following property: For every mixed action q of the opponent, there exists a mixed action p of the agent so that $r(p, q) \in S$. This condition, referred to as Blackwell's dual condition, will play a central role in the following. As shown by Blackwell [7], for a *convex* set S this condition is necessary and sufficient for S to be approachable; however, for nonconvex sets this condition is only necessary but not sufficient, as seen above.

The next example demonstrates how nonconvex target sets that satisfy Blackwell's dual condition can arise naturally in the context of no-regret algorithms.

Example 2. Suppose the agent wishes to maximize a scalar average reward $\bar{R}_n$, defined as before, subject to a long-term average cost constraint of the form $\bar{C}_n \le \gamma + o(1)$, where $\bar{C}_n$ is the n-step average of a (scalar or vector-valued) cost function $c(a, z)$. Let $r^*_\gamma(q) = \max_{p \in \Delta(\mathcal{A})}\{r(p, q)\colon c(p, q) \le \gamma\}$ denote the maximal expected reward that the agent can secure against a mixed action $q \in \Delta(\mathcal{Z})$ of the opponent. We refer to $r^*_\gamma(q)$ as *the constrained best-reward-in-hindsight*. Now, consider the target set $S = \{(r, c, q) \in \mathbb{R}^2 \times \Delta(\mathcal{Z})\colon r \ge r^*_\gamma(q), c \le \gamma\}$.

Without the side constraint on the average cost, the above is exactly Blackwell's formulation of the no-regret problem (Blackwell [6]): The set S is convex and therefore approachable, and a strategy of the agent that approaches this sets obtains $\bar{R}_n \ge \max_p r(p, \bar{q}_n) - o(1)$. However, with a nontrivial side constraint, the target set S is generally nonconvex and nonapproachable, as shown in Mannor et al. [30]. As a consequence, the convex hull of S was suggested there as a feasible target set for a regret-minimizing algorithm. However, in the fortuitous event that the opponent plays a stationary strategy, or close to that, one should aim at a higher reward as presented by the original target set rather than its convex relaxation.

Several other online decision problems involve nonconvex target sets that satisfy Blackwell's dual condition, including regret minimization with global cost functions (Even-Dar et al. [11]), regret minimization in variable duration repeated games (Mannor and Shimkin [26]), and regret minimization in stochastic game models (Mannor and Shimkin [25]).

Our starting point, then, is a target set S that satisfies Blackwell's dual condition, but may be nonconvex. Such a target set can be approached when faced with a *statistically stationary* opponent (which is restricted to stationary strategies), by using a simple adaptive algorithm (e.g., estimate the opponent's mixed action online and choose an appropriate response to the current estimate at each stage). However, against an arbitrary opponent only the convex hull of S is approachable in general. Our goal is to devise *opportunistic* approachability algorithms that, in addition to approaching this convex hull, seek to approach strict subsets thereof when the opponent's play turns out to be restricted in an appropriate sense. In particular, in the extreme case that the opponent plays a stationary strategy, we require that the set S itself be approached.

In fact, the algorithms we devise are shown to converge to a single point in S when the opponent is stationary. Moreover, we establish this convergence (and generalizations thereof) under the broader property of *empirical stationarity*, which is defined only in terms of the *observed pure actions* of the opponent.

The central contributions of this paper are the following:

- We formulate an appropriate concept of opportunistic approachability, which relies on the accompanying notions of statistically and empirically restricted opponents.
- We propose a class of approachability algorithms that is based on a *calibrated forecast* of the opponent's actions (Dawid [10], Foster and Vohra [12]). In contrast to the standard approachability algorithms that are based on Blackwell's *primal* condition, our algorithms are based on Blackwell's *dual* condition. Hence, they do not require the computation of the projection to the target set S, or the computation of the convex hull of S whenever S is nonconvex. Instead, they require only a computation of a best response at a finite number of points. We note, however, that the computational complexity is transferred in some sense to that of the calibrated forecasts.
- We show that the calibration-based algorithms are opportunistic in the above mentioned sense when facing a statistically restricted opponent. Moreover, to establish the opportunistic properties against an empirically restricted opponent, we require the calibrated forecast to be *slowly time varying* in an appropriate sense, which we establish for a specific forecasting algorithm.
- We apply our opportunistic approachability framework to the constrained regret minimization problem introduced in Example 2. Our algorithms attain the convex relaxation of the constrained best-reward-in-hindsight, while satisfying the long-term constrains. In addition, in the fortuitous event that the opponent's play is empirically or statistically restricted, our algorithms attain the constrained best-reward-in-hindsight itself.

Since Blackwell's original construction, several approachability algorithms and related results have been proposed in the literature. Hart and Mas-Colell [16] proposed a class of approachability algorithms, by using a *general* directional mapping instead of the one based on Euclidean distance to the target set. Shimkin and Shwartz [37] and Milman [33] extended the basic approachability concept and results to *stochastic games*. Hou [18] and Spinat [39], independently, formulated a necessary and sufficient condition for approachability of *general* (not necessarily convex) sets. In Lehrer [20], approachability theory was extended to infinite dimensional spaces. Lehrer and Solan [22] give the characterization of the family of approachable sets when the player is restricted to use strategies with bounded memory. Abernethy et al. [1] established an equivalence between no-regret learning and the approachability problem. Mannor et al. [28] proposed a robust approachability algorithm for repeated games with partial monitoring and applied it to the corresponding regret minimization problem. Moreover, Perchet and Quincampoix [35] proposed a unified framework for approachability both in the full or partial monitoring case. The approachability policies discussed in all these papers are based on Blackwell's *primal* condition, which is a geometric separation condition with respect to the fixed target set. Therefore, the existing algorithms are *not opportunistic* in the sense we advocate in this paper.

The idea of approachability using a response function appears in Lehrer and Solan [21] in the context of internal no-regret strategies (see Lehrer and Solan [23] for the updated version of that paper), and in Bernstein and Shimkin [3] in the context of approachability without projection. It also resembles the ideas in recent papers such as Mannor et al. [28] and Perchet and Quincampoix [35]. However, opportunistic properties of the related algorithms are not analyzed in these works. A recent work by Bubeck and Slivkins [8] considered the multiarmed bandit problem, and proposed an algorithm that simultaneously achieves the optimal convergence rates against both arbitrary and stationary opponents. However, the opportunistic property there is with respect to the convergence rates rather than with respect to the achievable goal.

The idea of choosing a best response to calibrated forecasts was first introduced in Foster and Vohra [12] in the context of attaining correlated equilibrium, and was subsequently used in Mannor and Shimkin [26] and Mannor et al. [30] in the context of regret minimization. An approachability strategy that is based on calibrated forecasts was apparently proposed by Perchet [34]; however, the discussion there is limited only to *convex* sets, and hence the opportunistic properties of the algorithm are not analyzed.

We note that the calibration-based algorithm, while conceptually simple, is computationally challenging because of the computational complexity of obtaining calibrated forecasts. In particular, given the recent result of Hazan and Kakade [17], it is unlikely that there exists an efficient algorithm to compute an exact calibrated forecast when the number of actions available to the opponent is large. The only computationally efficient calibration algorithms known in the literature are for the case of *binary* sequences (Mannor et al. [29]). Thus, our calibration-based scheme is computationally efficient in this case. For opponents with nonbinary action sets, other methods need be considered. It should be emphasized that the main goal in this paper is in formulating the concept of opportunistic approachability and in showing that there exist algorithms that fit this concept. Hence, the computational issues are left for future work. Finally, we note that the convergence rates of our algorithms are that of the calibrated forecast used. E.g., for ϵ-calibration (and thus, ϵ approachability) using internal regret minimization, the rate is the standard rate of convergence of no-regret algorithms, that is of $O(1/\sqrt{n})$.

The paper is structured as follows. In §2, we review the approachability problem and standard approachability algorithms. In §3, we introduce the concept of opportunistic approachability along with the definitions of the response and goal functions, which will be used subsequently in our algorithms. Section 4 provides a background on calibrated forecasts, presents the calibration-based approachability algorithm, and analyzes its performance in the cases of a general and statistically restricted opponent. In §5, we introduce *slowly varying* calibrated forecasts, establish their existence, and analyze the performance of our algorithms in the case of *empirically* restricted opponent. Section 6 applies the proposed algorithms to the problem of constrained regret minimization. We conclude in §7 with some final remarks.

2. Review of the approachability problem. In this section, we present the approachability problem and review the basic conditions for a set to be approachable, as well as Blackwell's approachability algorithm.

Consider a repeated two-person game between an agent and an arbitrary opponent (that collectively represents that agent's environment, including the effect of Nature as well as that of other agents active there). The agent chooses its actions from a finite set $\mathcal{A}$, and the opponent chooses its actions from a finite set $\mathcal{Z}$. At each time instance $n = 1, 2, \ldots$, the agent selects its action $a_n \in \mathcal{A}$, observes the action $z_n \in \mathcal{Z}$ chosen by the opponent, and obtains a *vector* reward $r_n = r(a_n, z_n) \in \mathbb{R}^\ell$, $\ell \geq 1$, where $r\colon \mathcal{A} \times \mathcal{Z} \to \mathbb{R}^\ell$ is a given function. The average reward vector obtained by the agent up to time n is then $\bar{R}_n = n^{-1} \sum_{k=1}^n r_k$. A *mixed* action of the agent is the probability distribution $p \in \Delta(\mathcal{A})$, where $p(a)$ specifies the probability of choosing action $a \in \mathcal{A}$. Similarly,

$q \in \Delta(\mathcal{Z})$ denotes a mixed action of the opponent. Let $\bar{q}_n \in \Delta(\mathcal{Z})$ denote the empirical distribution of the opponent's actions at time n, with

$$\bar{q}_n(z) \triangleq \frac{1}{n}\sum_{k=1}^{n} \mathbb{I}\{z_k = z\}.$$

Also, define the span of the reward vector

$$\rho \triangleq \max_{a,z,a',z'} \|r(a,z) - r(a',z')\|, \tag{1}$$

where $\|\cdot\|$ is Euclidean norm.

In what follows, we will slightly abuse notation and let

$$r(p,q) \triangleq \sum_{a\in\mathcal{A},\, z\in\mathcal{Z}} p(a)q(z)r(a,z)$$

denote the expected reward under mixed actions $p \in \Delta(\mathcal{A})$ and $q \in \Delta(\mathcal{Z})$; the distinction between $r(a,z)$ and $r(p,q)$ should be clear by their arguments. Occasionally, we will use $r(p,z) = \sum_{a\in\mathcal{A}} p(a)r(a,z)$ for the expected reward under mixed action $p \in \Delta(\mathcal{A})$ and pure action $z \in \mathcal{Z}$. The notation $r(a,q)$ is interpreted similarly.

Let

$$h_{n-1} \triangleq \{a_1, z_1, \ldots, a_{n-1}, z_{n-1}\} \in (\mathcal{A}\times\mathcal{Z})^{n-1}$$

denote the history of the game up to time n. A *strategy* $\pi = (\pi_n)$ of the agent is a collection of the decision rules $\pi_n\colon (\mathcal{A}\times\mathcal{Z})^{n-1} \to \Delta(\mathcal{A})$, $n \geq 1$, where each mapping π_n specifies the mixed action for the agent at time n, based on the observed history:

$$p_n = \pi_n(h_{n-1}).$$

The pure action a_n taken by the agent is then selected randomly according to p_n. Similarly, the opponent's strategy is denoted by $\sigma = (\sigma_n)$, with $\sigma_n\colon (\mathcal{A}\times\mathcal{Z})^{n-1} \to \Delta(\mathcal{Z})$. Let $\mathbb{P}^{\pi,\sigma}$ denote the probability measure on $(\mathcal{A}\times\mathcal{Z})^{\infty}$ induced by the strategy pair (π,σ). In what follows, all the probabilistic statements are assumed to hold with respect to these measures.

In the approachability problem, we consider a set $S \subseteq \mathbb{R}^{\ell}$, and ask if there exists a strategy for the agent that will bring the average reward vector to S (asymptotically, almost surely) no matter what the opponent's strategy is. Below is the classical definition of an approachable set from Blackwell [7].

Definition 1 (Approachable Set). A closed set $S \subseteq \mathbb{R}^{\ell}$ is *approachable by the agent's strategy* π if the average reward $\bar{R}_n = n^{-1}\sum_{k=1}^{n} r_k$ converges to S almost surely for every strategy σ of the opponent.[1] The set S is *approachable* if there exists such a strategy for the agent.

In what follows, we find it convenient to state all our results in terms of the *expected* average reward, where the expected value is only with respect to the agent's mixed actions:

$$\bar{r}_n \triangleq \frac{1}{n}\sum_{k=1}^{n} r(p_k, z_k).$$

With this modified reward, the stated convergence results will be shown to hold *pathwise*, for any possible sequence of the opponent's actions. The corresponding almost sure results for the actual average reward can be easily deduced, using martingale convergence theory. Indeed, note that

$$d(\bar{R}_n, S) \leq \|\bar{R}_n - \bar{r}_n\| + d(\bar{r}_n, S).$$

Now, the first term is the norm of the mean of the martingale difference sequence $D_k = r(a_k,z_k) - r(p_k,z_k)$ and can readily be shown to converge to zero at a uniform rate of $O(1/\sqrt{n})$; see, e.g., Shiryaev [38] or Cesa-Bianchi and Lugosi [9].

Next, we present a formulation of Blackwell's Theorem (Blackwell [7]), which provides us with a *sufficient* condition for approachability of a general set S. To this end, for any $x \notin S$, let $c(x) \in S$ denote a closest point in S to x. Also, for any $p \in \Delta(\mathcal{A})$ let $T(p) \triangleq \{r(p,q)\colon q \in \Delta(\mathcal{Z})\}$, which equals the convex hull of the points $\{r(p,z)\}_{z\in\mathcal{Z}}$.

[1] Blackwell's original definition requires almost sure convergence at a uniform rate over the probability distributions induced by the strategies π and σ. Our algorithms satisfy this definition provided that the convergence of the employed calibrated forecasts is uniform, as for example in the case of the calibration forecaster discussed in §5.2 in this paper. However, we will not assume it here.

Definition 2 (Primal Condition: B-Sets). A closed set $S \subseteq \mathbb{R}^\ell$ is called a *B-set* (where B stands for *Blackwell*) if for every $x \notin S$ there exists a mixed action $p = p(x) \in \Delta(\mathcal{A})$ such that the hyperplane through $y = c(x)$ perpendicular to the line segment xy, separates x from $T(p)$.

Theorem 1 (Sufficient Condition and Algorithm). *Every B-set is approachable, by using at time n the mixed action $p(\bar{r}_{n-1})$ of Definition 2 whenever $\bar{r}_{n-1} \notin S$. (If $\bar{r}_{n-1} \in S$, an arbitrary action can be used.)*

Remark 1. Theorem 1 holds also if $\bar{r}_n$ is replaced with $\bar{R}_n$.

In addition, a dual *necessary* condition for approachability can be formulated as follows.

Definition 3 (Dual Condition: D-Sets). A closed set $S \subseteq \mathbb{R}^\ell$ is called a *D-set* (where D stands for *Dual*) if for every $q \in \Delta(\mathcal{Z})$ there exists a *response* $p \in \Delta(\mathcal{A})$ such that $r(p, q) \in S$.

Theorem 2 (Necessary Condition). *A closed set S is approachable only if it is a D-set.*

For *convex* target sets, Blackwell [7] showed that the primal and dual conditions coincide.

Theorem 3. *Let S be a closed convex set. Then, the following statements are equivalent*: (i) *S is approachable*, (ii) *S is a B-set, and* (iii) *S is a D-set.*

Theorem 3 has the following corollary.

Corollary 1. *The convex hull of a D-set is approachable* (*and is also a B-set*).

Proof. The convex hull of a D-set is a convex D-set. The claim then follows by Theorem 3. □

We note that a complete characterization of the family of approachable sets was provided independently by Hou [18] and Spinat [39]. They proved that a closed set S is approachable if and only if it contains a B-set. However, this result is *not* used in the present paper, as our main interest here is in the dual (rather than primal) condition. Hence, throughout the paper, we assume that the target set S satisfies the following condition.

Assumption 1. *The set S is a D-set.*

Observe that we do *not* assume that S is a convex set. Consequently, although $\mathrm{conv}(S)$ is approachable by Corollary 1, S itself need not be approachable.

Our focus in this paper is on a conceptually simple approachability strategy that is based on the dual condition, previously proposed by Perchet [34]: at each time n use the mixed action $p_n \in \Delta(\mathcal{A})$, which is a *response* (in the sense of Definition 3) to a *calibrated forecast* $y_n \in \Delta(\mathcal{Z})$ of the pure action $z_n \in \mathcal{Z}$. The definition of calibrated forecasts and the analysis of this strategy is presented in §4.

We note that in parts of this paper, the set S will only be implicitly defined through an appropriate *response function*, so that Assumption 1 is satisfied by definition of the latter; see Remark 2 in §3.

3. Opportunistic approachability.

In this section, we define the desiderata for an opportunistic approachability algorithm. To that end, we first define appropriate notions of a *statistically* and an *empirically* restricted play of the opponent, as well as the response and goal function for the given target set.

Before making formal definitions, we state the idea of our approach. We propose algorithms that simultaneously achieve the following goals, for any D-set S:

1. The *convex hull* of S is approached, for any strategy of the opponent.
2. If the *mixed actions* of the opponent or, more generally, the *empirical frequencies* of the opponent's actions are restricted to a subset of its mixed actions space (in the sense of Definitions 4 and 5), then the algorithm approaches a corresponding strict subset of $\mathrm{conv}(S)$. In particular, if the opponent is stationary, the set S itself is approached.

3.1. Restricted opponent play.

We start with a definition of restricted play of the opponent in terms of the sequence of its *mixed* actions $\{q_n\}$, which is intuitive and easy to state. In particular, we consider the notion of a *statistically restricted play*, in the sense that $\{q_n\}$ is asymptotically restricted to some set $Q \subseteq \Delta(\mathcal{Z})$. We note that this and other definitions below relate to a *given sample path* of the process (rather than to the overall policy of the opponent).

Definition 4 (Statistically Q-Restricted Play). We say that the play of the opponent is *statistically Q-restricted*, if there exists a convex subset $Q \subseteq \Delta(\mathcal{Z})$ so that the sequence $\{q_n\}$ of the mixed actions of the opponent satisfies, for the given sample path,

$$\lim_{n\to\infty} \frac{1}{n} \sum_{k=1}^{n} d(q_k, Q) = 0.$$

Here, $d(q, Q)$ is Euclidean point-to-set distance.

Observe that the Cesàro mean convergence of Definition 4 is a weaker assumption than the convergence of $d(q_n, Q)$ to zero.

A possible weakness of Definition 4 is that the mixed actions of the opponent are not generally revealed (when its strategy is not known), or may even lack an explicit meaning (e.g., when the opponent is Nature). We therefore proceed to define a notion of an *empirically restricted play* of the opponent, in terms of the *empirical frequencies* of the opponent's pure actions. To this end, we need to refer to a certain partition of the time axis into blocks on which these frequencies are computed. We let τ_m denote the length of block $m = 1, 2, \ldots$. Also, we let $n_M = \sum_{m=1}^{M} \tau_m$ denote the time at the end of block M. Finally, $\hat{q}_m \in \Delta(\mathcal{Z})$ denotes the empirical distribution of the opponent's actions of block m, namely,

$$\hat{q}_m(z) = \frac{1}{\tau_m} \sum_{k=n_{m-1}+1}^{n_m} \mathbb{I}\{z_k = z\}.$$

With this in hand, we introduce the following.

Definition 5 (Empirically Q-Restricted Play). We say that the play of the opponent is *empirically Q-restricted with respect to a partition* $\{\tau_m\}$, if there exists a convex subset $Q \subseteq \Delta(\mathcal{Z})$ so that, for the given sample path,

$$\lim_{M\to\infty} \frac{1}{n_M} \sum_{m=1}^{M} \tau_m d(\hat{q}_m, Q) = 0.$$

We note that the requirement of Definition 5 is much weaker than that of Definition 4 (as Lemma 2 shows), and hence is our focus. To see this, consider the following simple example.

Example 3. Consider binary sequences of actions, and let $Q = \{(0.5, 0.5)\}$ be a singleton. The deterministic sequence $0101\ldots$ is empirically Q-restricted with respect to *any* partition with fixed *even* block lengths, or with any strictly increasing blocks lengths. However, the opponent that generates this sequence using alternating mixed actions $(1, 0), (0, 1)$ is of course not statistically Q-restricted. □

Observe that our definition of empirically Q-restricted play involves a *general* partition $\{\tau_m\}$ rather than a partition with fixed lengths $\tau_m \equiv \tau$. The main reason behind this general definition is the fact that we would like to cover the case of *statistically* stationary sequences. The following example clarifies this point.

Example 4. Consider a stationary opponent that chooses its actions using a fixed strictly mixed action $q_0 \in \Delta(\mathcal{Z})$, and the corresponding restriction set $Q = \{q_0\}$. In this case, the sequence of pure actions will *not* satisfy Definition 5 with probability one if we choose a partition with fixed (or bounded) block lengths. Actually, we can satisfy Definition 5 with probability one only if we choose a partition with *superlogarithmically* increasing lengths (as is shown in general by Lemma 2). □

A given sequence of actions may satisfy Definition 5 under different partitions, as the following example demonstrates.

Example 5. Recall the setting of Example 3, and consider the sequence $01001100001111\ldots$. The empirical frequencies of this sequence do *not* converge to Q, but it is empirically Q-restricted with respect to a partition with *exponentially* increasing lengths $\tau_m = 2^m$. However, if we choose any partition with subexponentially increasing lengths, Definition 5 will not be satisfied. □

In general, we are interested in the *minimum* possible block lengths that will ensure that Definition 5 is satisfied. Moreover, we mostly focus on the sequences for which Definition 5 can be satisfied with a partition with *subexponentially* increasing block lengths. This is motivated by the following lemma, which shows that Definition 5 requires more than just convergence of $\bar{q}_n$ to Q.

LEMMA 1. *If Definition 5 is satisfied with respect to a partition with subexponentially increasing block lengths for some $Q \subseteq \Delta(\mathcal{Z})$, then $\bar{q}_n$ converges to Q. However, the converse is not true. Namely, there exist a sequence of actions so that $\bar{q}_n$ converges to Q, but there is no partition with subexponentially increasing block lengths so that Definition 5 is satisfied with respect to it.*

PROOF. For any $n_{M-1}+1 \le k \le n_M$, we have that

$$\begin{aligned} d(\bar{q}_k, Q) &\le d(\bar{q}_{n_M}, Q) + \|\bar{q}_{n_M} - \bar{q}_k\| \\ &= d\left(\frac{1}{n_M}\sum_{m=1}^{M} \tau_m \hat{q}_m, Q\right) + \|\bar{q}_{n_M} - \bar{q}_k\| \\ &\le \frac{1}{n_M}\sum_{m=1}^{M} \tau_m d(\hat{q}_m, Q) + \frac{\tau_M}{n_{M-1}}, \end{aligned}$$

where the second inequality holds by the convexity of the point-to-set Euclidean distance to a convex set, and by the fact that the changes in $\bar{q}_n$ are of the order of $1/n$. But, if τ_m increases subexponentially, we have that $\tau_M/n_{M-1} \to 0$, and the result follows.

To see why the converse need not be true, consider the sequence 01001100001111... from Example 5. It can be easily verified that $\bar{q}_n(1) \to Q = [1/3, 1/2]$ in this case. However, if we choose *any* partition with subexponentially increasing lengths, we have that, in the long-term, $\hat{q}_m(1)$ is either closed 0 or to 1, so that

$$\liminf_{M\to\infty} \frac{1}{n_M}\sum_{m=1}^{M} \tau_m d(\hat{q}_m, Q) > 0. \quad \square$$

It should be emphasized that the convergence of $\bar{q}_n$ to a restriction set Q seems to be not sufficient to guarantee opportunistic convergence of the average reward in terms of Q. In particular, the convergence of the empirical frequencies $\bar{q}_n$ does not say anything about the *rate* of this convergence. Indeed, consider again the sequence 01001100001111... from Example 5. As noted in the proof of Lemma 1, $\bar{q}_n$ converges to a strict subset of $[0, 1]$. However, if we choose any partition with subexponentially increasing lengths, we have that, in the long-term, the empirical frequency of 1 at any interval is either closed 0 or to 1, implying that opportunistic convergence is impossible.

Finally, we claim that, almost surely, the requirement in Definition 4 implies the requirement in Definition 5 (see Appendix A for a proof).

LEMMA 2. *Suppose that the play of the opponent is statistically Q-restricted in the sense of Definition 4, almost surely, with respect to the probability distribution induced by the strategies of the agent and the opponent. Then the requirement of Definition 5 is satisfied with respect to any partition $\{\tau_m\}$ with superlogarithmically increasing block lengths.*

3.2. Response and goal functions. By definition of a D-set, one can define a *response function* p^* that for any q returns p such that $r(p, q) \in S$. Below we demonstrate that p^* cannot be continuous in general. We therefore settle for the following piecewise continuity property.

DEFINITION 6 (REGULAR RESPONSE FUNCTION). A function $p^*\colon \Delta(\mathcal{Z}) \to \Delta(\mathcal{A})$ is a *regular response function* relative to the target set S if

(i) for each $q \in \Delta(\mathcal{Z})$, $r(p^*(q), q) \in S$; and
(ii) the function p^* is a *piecewise continuous* function.

That is, there exists a finite partition of $\Delta(\mathcal{Z})$ such that p^* is continuous on the interior of every element of that partition.

EXAMPLE 6 (EXAMPLE 1 CONTINUED). Recall the approachability problem with the scalar reward matrix

$$R = \begin{pmatrix} 2 & 0 \\ 0 & -2 \end{pmatrix}$$

and target D-set $S = (-\infty, -1] \cup [1, \infty)$, introduced in Example 1. For brevity, we identify any $p \in [0, 1]$ with a mixed action $(p, 1-p)$ of the agent (namely, p is the probability of action 0). Similarly, a $q \in [0, 1]$ is identified with a mixed action $(q, 1-q)$ of the opponent. Observe that for $q < 0.5$ and $p \le 0.5 - q$, we have

that $r(p,q) \le -1$ and therefore $r(p,q) \in S$. Similarly, for $q > 0.5$ and $p \ge 1.5 - q$, we have that $r(p,q) \ge 1$, implying that $r(p,q) \in S$. We thus can define a regular response function as follows:

$$p^*(q) = \begin{cases} 0, & \text{for } q \le 0.5 \\ 1, & \text{otherwise.} \end{cases} \tag{2}$$

Although other selections for $p^*(q)$ can be made, all of them will have a discontinuity at $q = 0.5$. □

Below we show that we can always choose a *regular* (that is, piecewise-continuous) response function.

Lemma 3. *Under Assumption 1, there exists a regular response function.*

Proof. To prove this lemma, we use standard results from *set-valued analysis* (see for instance Aubin and Frankowska [2]). We construct a set-valued function $f(q)$ as follows:

$$f(q) \triangleq \{p \in \Delta(\mathcal{A}) \colon r(p,q) \in S\}.$$

Since S is closed and $r(p,q)$ is continuous, it follows that for every $\{q_n\}$ and q such that $q_n \to q$, we have that $\lim_{n\to\infty} f(q_n) \subseteq f(q)$, in the sense that if $p_n \in f(q_n)$ is such that $p_n \to p$ then $p \in f(q)$. We conclude that f is an upper semi-continuous set-valued function. It follows from Fort's Theorem that f is also lower semi-continuous on a residual subset of S, namely, on a set whose complement is a meager set. In our case, this is an open set whose complement has measure 0.

Pick a finite partition of $\Delta(\mathcal{Z})$. Now, from Michael's selection theorem (Michael [32]), we know that on every element of the partition we have a continuous selection. Thus, we can choose $p^*(q) \in f(q)$ such that p^* is a piecewise continuous function. □

The actual choice of p^* is problem dependent. In §6 we will see an example where p^* is naturally defined as a best-response map. In general, we make the following assumption.

Assumption 2. *Let p^* be a regular response function relative to the given target set S, which we fix in the following. We assume that $p^*(q)$ can be efficiently computed for any given $q \in \Delta(\mathcal{Z})$.*

We note that Assumption 2 implies Assumption 1 by the definition of p^*. Hence, throughout, we suppose that Assumption 2 holds, and we usually do not refer to the target set S explicitly.

Remark 2. Observe that a given response function p^* induces the following set:

$$S(p^*) \triangleq \big\{r(p^*(q),q) \colon q \in \Delta(\mathcal{Z})\big\}.$$

This is the minimal target set for which p^* is a (regular) response function. Consequently, we can start from a given response function p^* that will define the target set $S(p^*)$. Moreover, the definition of the response function implies that any set S that contains $S(p^*)$ can be considered as a feasible target set.

The specified response function p^* leads naturally to our next definition.

Definition 7 (Goal Function). The *goal function* $r^*\colon \Delta(\mathcal{Z}) \to S$ is defined as $r^*(q) = r(p^*(q),q)$ for any $q \in \Delta(\mathcal{Z})$.

3.3. Opportunistic strategies. When the play of the opponent turns out to be statistically/empirically Q-restricted, we will essentially require the average reward to converge to $R(Q) = \operatorname{conv}\{r^*(q) \colon q \in Q\}$, the convex hull of the image of Q under the goal function r^* (see Figure 1). Because of possible discontinuities in r^*, we need to slightly expand that definition.

Definition 8 (Closed Convex Image). The *closed convex image* of a set $Q \subseteq \Delta(\mathcal{Z})$ under the goal function r^* is defined as

$$R^+(Q) \triangleq \bigcap_{\epsilon>0} \operatorname{conv}\{r^*(q) \colon d(q,Q) \le \epsilon\}.$$

In words, the set $R^+(Q)$ contains the convex hull of all the points of $r^*(q)$, $q \in Q$, *together with possible jumps in r^* on the boundary of Q*. Note that $R^+(Q) \subset \operatorname{conv}(S)$, as $r^*(q) \in S$ by its definition. We illustrate the inclusion of jumps using the following example.

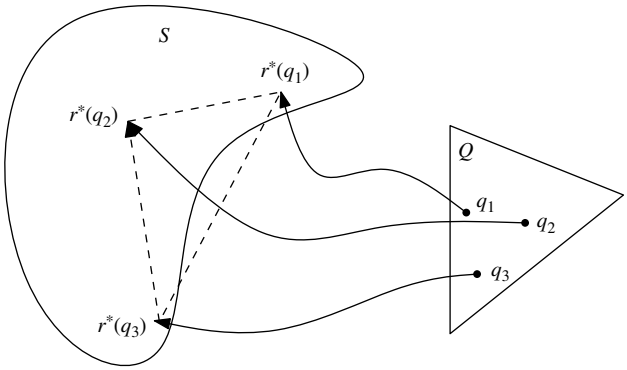

FIGURE 1. An illustration of the restriction set Q and the convex hull of the image of Q under the goal function r^*.

EXAMPLE 7 (EXAMPLE 6 CONTINUED). Consider the response function defined in (2). The corresponding goal function is given by

$$r^*(q) = \begin{cases} -2(1-q), & \text{for } q \le 0.5 \\ 2q, & \text{otherwise.} \end{cases}$$

We now compute the closed convex image of singeltons:

$$R^+(\{q\}) = \begin{cases} r^*(q), & \text{for } q \ne 0.5 \\ \text{conv}(\{-1,1\}) = [-1,1], & \text{for } q = 0.5. \end{cases}$$

Observe that the discontinuity of r^* at $q = 0.5$ is expressed by the fact that $R^+(\{q\})$ is the "jump interval" $[-1,1]$. □

With the above notions at hand, we can finally define opportunistic approachability strategies.

DEFINITION 9 (STATISTICALLY OPPORTUNISTIC APPROACHABILITY). A strategy π is *statistically opportunistic* for a given goal function r^* if it holds that

$$\lim_{n\to\infty} d(\bar{r}_n, R^+(Q)) = 0$$

whenever the play of the opponent is statistically Q-restricted (Definition 4) for some set $Q \subseteq \Delta(\mathcal{Z})$.

DEFINITION 10 (EMPIRICALLY OPPORTUNISTIC APPROACHABILITY). A strategy π is *empirically opportunistic* for a given goal function r^* w.r.t. a partition $\{\tau_m\}$ if

$$\lim_{n\to\infty} d(\bar{r}_n, R^+(Q)) = 0$$

whenever the play of the opponent is empirically Q-restricted w.r.t. $\{\tau_m\}$ (Definition 5) for some set $Q \subseteq \Delta(\mathcal{Z})$.

It should be emphasized that the definitions of opportunistic approachability strategies are based on the *sample path* properties of the opponent's play (either in pure or mixed actions). Also, observe that identifying whether the opponent is statistically restricted, or identifying the partition on which the opponent is empirically restricted are not straightforward tasks. However, no such tests are required in order to implement the suggested strategy, nor for its stated opportunistic properties to hold. Moreover, the related convergence results are required to hold *without knowing the restriction set* Q beforehand.

REMARK 3. Note that Definitions 9 and 10 imply the standard definition of approachability, by setting $Q = \Delta(\mathcal{Z})$. In this case, $R^+(Q) \subseteq \text{conv}(S)$, and

$$\lim_{n\to\infty} d(\bar{r}_n, \text{conv}(S)) = 0.$$

REMARK 4. Observe that if a strategy is empirically opportunistic with respect to some partition with super-logarithmically increasing lengths, it is also statistically opportunistic (as follows from Lemma 2). But the converse is not necessarily true.

We close this section with a simple example showing that a naive strategy that plays $p_n = p^*(\bar{q}_{n-1})$ fails to provide approachability guarantees, even in the case of a convex target set.

EXAMPLE 8. Consider the reward matrix

$$R = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$$

and target set $S = [0.5, \infty)$. Clearly, S is a convex set, and it is also a D-set: the response function can be taken the same as in (2) and the corresponding goal function is given by

$$r^*(q) = \begin{cases} 1 - q, & \text{for } q \le 0.5 \\ q, & \text{otherwise.} \end{cases}$$

Now, assume that the sequence of opponent's actions is a periodic sequence 010101..., and that the naive strategy $p_n = p^*(\bar{q}_{n-1})$ is employed by the agent. Since for all even n, $\bar{q}_{n-1} > 0.5$, we have that $a_n = 0$. Also, for odd n, $\bar{q}_{n-1} = 0.5$, and $a_n = 1$. Consequently, $\bar{R}_n \to 0 \notin S$. □

This example illustrates the well-known phenomenon of simple "best-response" strategies: since they choose actions deterministically, they can be "tricked" by the opponent all the time. This is also the case, for example, in the standard no-regret problem where best-response strategies that do not randomize fail to minimize the regret.

4. Calibration-based approachability. In this section, we present the basic calibration-based algorithm that is the subject of this paper. We first provide some background on calibrated forecasts, present and analyze our calibration-based approachability algorithm, and prove its properties in the case of a *statistically restricted opponent*.

4.1. Calibrated forecasts. A *forecaster* is an algorithm that specifies at each time instance n a probabilistic forecast $y_n \in \Delta(\mathcal{Z})$ of the opponent's action z_n, based on the history of observed actions and previous forecasts. The forecaster's policy may be *randomized*, i.e., at each time n it specifies a probability measure η_n over $\Delta(\mathcal{Z})$. In this case, the forecast $y_n \in \Delta(\mathcal{Z})$ is drawn at random according to η_n.

The following is a standard definition of a calibrated forecaster (see, e.g., Foster and Vohra [12]).

DEFINITION 11 (CALIBRATED FORECASTER). A forecaster is *calibrated* if for every Borel measurable set $Q \subseteq \Delta(\mathcal{Z})$ and every strategy of the opponent, it holds that

$$\lim_{n\to\infty} \frac{1}{n} \sum_{k=1}^{n} \mathbb{I}\{y_k \in Q\}(\mathbf{1}(z_k) - y_k) = 0, \quad \text{a.s.}, \tag{3}$$

where $\mathbf{1}(z)$ is the probability vector in $\Delta(\mathcal{Z})$ concentrated on z.

No deterministic forecaster can be calibrated for all possible sequences of outcomes (Dawid [10]). However, if the forecaster is allowed to randomize, calibration is possible. Several randomized calibrated forecasters were proposed in the literature (see the overview in Cesa-Bianchi and Lugosi [9], as well as Mannor et al. [29] and Foster et al. [13]). The common approach is to use a finite ϵ-grid over $\Delta(\mathcal{Z})$, which is gradually refined in order to fulfill the requirement of Definition 11. To achieve ϵ-calibration, the algorithms usually process the entire grid for each prediction. The only computationally efficient algorithms known in the literature are for the case of *binary* sequences (Mannor et al. [29]). Moreover, it was recently shown in Hazan and Kakade [17] that the existence of a general computationally efficient calibrated forecaster would imply the existence of an efficient algorithm for computing approximate Nash equilibria, thus implying the unlikely conclusion that every problem in PPAD (the class of problems that are polynomial time reducible to the problem of computing Nash equilibrium in a two player game) is solvable in polynomial time.

The calibration property in (3) can be interpreted as a *merging* or *averaging* property of the forecast relative to the pure actions of the opponent. The next lemma shows that a similar property holds with respect to the *mixed* (rather than pure) actions.

LEMMA 4. *Let $\{q_k\}_{k=1}^n$ denote the mixed actions of the opponent. The calibration property (3) is equivalent to*

$$\lim_{n\to\infty} \frac{1}{n} \sum_{k=1}^{n} \mathbb{I}\{y_k \in Q\}(q_k - y_k) = 0, \quad \textit{a.s.}$$

PROOF. The result follows by the strong law of large numbers, applied to the martingale difference sequence $D_k = \mathbb{I}\{y_k \in Q\}(\mathbf{1}(z_k) - q_k)$; see, e.g., Kalai et al. [19]. □

It will be convenient to reformulate the calibration property in Lemma 4 in terms of

$$\mu_n \triangleq \frac{1}{n}\sum_{k=1}^{n} \delta_{q_k, y_k} \in \Delta(\Delta(\mathcal{Z}) \times \Delta(\mathcal{Z})),$$

the joint empirical distribution of the opponent's mixed actions $\{q_k\}_{k=1}^n$ and forecasts $\{y_k\}_{k=1}^n$ by time n. Specifically, let

$$\mathcal{M} \triangleq \left\{\mu \in \Delta(\Delta(\mathcal{Z}) \times \Delta(\mathcal{Z})):\ \int \mathbb{I}\{y \in Q\}(q-y)\, d\mu(q,y) = 0, \forall\, Q \subseteq \Delta(\mathcal{Z})\right\} \tag{4}$$

denote the set of joint probability measures of the opponent's mixed actions (q) and forecasts (y) that satisfy the calibration property. We note that in this and subsequent definitions, we only refer to *Borel-measurable* sets Q. Then, the statement of Lemma 4 is equivalent to the statement that μ_n "converges" to $\mathcal{M}$ in the sense that

$$\lim_{n\to\infty} \int \mathbb{I}\{y \in Q\}(q-y)\, d\mu_n(q,y) = 0, \quad \forall\, Q \subseteq \Delta(\mathcal{Z}).$$

We next provide an important equivalent characterization of the set $\mathcal{M}$, and also prove *variability ordering* property (see, e.g., Whitt [40]) of the marginal distributions of $\mu \in \mathcal{M}$, that will be used in the sequel. For a given $\mu \in \mathcal{M}$, let

$$\mu_1(dq) = \int_y \mu(dq, dy) \qquad \text{and} \qquad \mu_2(dy) = \int_q \mu(dq, dy) \tag{5}$$

denote the marginal distributions of the mixed actions and forecasts, respectively. Also, let $(\mathbf{q}, \mathbf{y})$ denote a *random vector* distributed according to μ.

LEMMA 5. 1. *We have that $\mu \in \mathcal{M}$ if and only if*

$$\mathbb{E}(\mathbf{q} \mid \mathbf{y}) = \mathbf{y}, \quad \mu_2\text{-a.s.}$$

2. *For any $\mu \in \mathcal{M}$ and any convex function V on $\mathbb{R}^{|\mathcal{Z}|}$, we have that*

$$\int_y V(y)\mu_2(dy) \le \int_q V(q)\mu_1(dq).$$

PROOF. To prove part 1 we use the following standard definition of the conditional expectation (see, e.g., Shiryaev [38, p. 220]). The conditional expectation of the random variable $\mathbf{q}$ under the condition that $\mathbf{y} = y$ is any Borel-measurable function

$$\mathscr{E}(y) \triangleq \mathbb{E}(\mathbf{q} \mid \mathbf{y} = y)$$

for which

$$\int \mathbb{I}\{y \in Q\} q\, d\mu = \int_y \mathbb{I}\{y \in Q\}\mathscr{E}(y)\mu_2(dy), \quad \forall\, Q \subseteq \Delta(\mathcal{Z}). \tag{6}$$

However, by the calibration property,

$$\int \mathbb{I}\{y \in Q\} q\, d\mu = \int_y \mathbb{I}\{y \in Q\} y \mu_2(dy), \quad \forall\, Q \subseteq \Delta(\mathcal{Z}).$$

Therefore, $\mathscr{E}(y) = y$ satisfies (6), and the result follows by substituting y with the random variable $\mathbf{y}$.

Now, part 2 of the lemma easily follows by part 1 and Jensen's inequality. Indeed, for any convex function V on $\mathbb{R}^{|\mathcal{Z}|}$, it holds that

$$\mathbb{E}[V(\mathbf{y})] = \mathbb{E}[V(\mathbb{E}(\mathbf{q} \mid \mathbf{y}))] \le \mathbb{E}[\mathbb{E}(V(\mathbf{q}) \mid \mathbf{y})] = \mathbb{E}[V(\mathbf{q})]. \quad \square$$

We illustrate the relation between the two distributions μ_1 and μ_2 using the following example.

EXAMPLE 9. Suppose that the empirical distribution of the opponent's mixed actions converges to

$$\mu_1 = \alpha\delta_{q^{(1)}} + \bar{\alpha}\delta_{q^{(2)}}, \quad \alpha \in [0,1], \quad \alpha + \bar{\alpha} = 1.$$

That is, in the long term, the opponent chooses α % of the time the mixed action $q^{(1)}$, and $\bar{\alpha}$ % of the time he chooses the mixed action $q^{(2)}$. It is easy to see that in this case, the following two forecasters are calibrated: (i) with $\|y_k - q_k\| \to 0$, and (ii) with $y_k \to \alpha q^{(1)} + \bar{\alpha} q^{(2)} \triangleq q_0$. Indeed, the joint empirical distribution in the first case converges to

$$\mu = \alpha\delta_{q^{(1)}, q^{(1)}} + \bar{\alpha}\delta_{q^{(2)}, q^{(2)}}$$

and in the second case to

$$\mu = \alpha\delta_{q^{(1)}, q_0} + \bar{\alpha}\delta_{q^{(2)}, q_0}.$$

It can be easily verified that, in both cases, $\mu \in \mathcal{M}$. Also, in the first case, obviously

$$\int_y V(y)\mu_2(dy) = \alpha V(q^{(1)}) + \bar{\alpha} V(q^{(2)}) = \int_q V(q)\mu_1(dq)$$

since both marginal distributions are the same. However, in the second case

$$\int_y V(y)\mu_2(dy) = V(q_0) \le \alpha V(q^{(1)}) + \bar{\alpha} V(q^{(2)}) = \int_q V(q)\mu_1(dq),$$

where the inequality follows by convexity of V. Namely, there is less variability in μ_2 than in μ_1. □

4.2. The calibrated approachability algorithm. Recall that p^* denotes a regular response function relative to the given target set S (Definition 6). The algorithm that we analyze in the remainder of the paper is conceptually simple—at each time n use the mixed action p_n, which is specified by

$$p_n = p^*(y_n), \tag{7}$$

where y_n is the calibrated forecast at time n. This algorithm was previously proposed by Perchet [34].

4.3. Approachability results. We show that the proposed algorithm is an approachability algorithm for $\mathrm{conv}(S)$ in general, and establish its opportunistic properties in case of a *statistically* restricted play of the opponent. (The case of an *empirically* restricted opponent is analyzed in §5.)

In Theorem 4, we prove an abstract and general property of the calibrated approachability algorithm that relates the empirical distribution of the mixed actions of the opponent to the empirical distribution of the forecasts and the corresponding average reward. This result implies the general approachability result to $\mathrm{conv}(S)$, as well as the opportunistic property of the algorithm (see Corollary 2).

Recall the definitions of the set $\mathcal{M}$ in (4) and the corresponding marginal distributions $\mu_1(\cdot)$ and $\mu_2(\cdot)$ in (5). Also, let

$$f_n \triangleq \mu_{n,1} = \frac{1}{n}\sum_{k=1}^{n} \delta_{q_k} \in \Delta(\Delta(\mathcal{X}))$$

denote the empirical distribution of the mixed actions $\{q_k\}_{k=1}^n$. For a given f_n, we can define the following set of possible reward vectors:

$$R_n \triangleq \left\{\mathbb{E}_{Y\sim\mu_2}[r^*(Y)]\colon \mu \in \mathcal{M}, \mu_1 = f_n\right\}.$$

This is the set of all expected target rewards, where the expected value is with respect to a marginal distribution of the forecasts, which is "compatible" with the calibration property (i.e., belongs to the set $\mathcal{M}$ in (4)) and with the empirical distribution f_n of $\{q_k\}_{k=1}^n$. We note that R_n is a convex set by definition of $\mathcal{M}$. Still, in order to take into account the possible jumps in $r^*(y)$ on the boundary of $\mathcal{M}$, we need to augment R_n as follows (see also Definition 8):

$$R_n^+ \triangleq \left\{\mathbb{E}_{Y\sim\mu_2}[F(Y)]\colon F\colon \Delta(\mathcal{X}) \to \mathbb{R}^\ell, F(y) \in R^+(\{y\}), \mu \in \mathcal{M}, \mu_1 = f_n\right\}. \tag{8}$$

Observe that under Assumption 2, $R_n^+ \subseteq \mathrm{conv}(S)$. We note that R_n^+ can be interpreted as the *closed convex image* of a set $\mathcal{M}$ under the function $\mathbb{E}_{Y\sim\mu_2}[r^*(Y)]$, constrained that the marginal distribution of the opponent's actions equal to f_n. Also, observe that when r^* is continuous, we have that $R_n^+ = R_n$.

THEOREM 4. *Suppose that Assumption 2 holds. Then, if the agent uses the calibrated approachability algorithm specified by (7), we have that*

$$\lim_{n\to\infty} d(\bar{r}_n, R_n^+) = 0, \tag{9}$$

almost surely, for any strategy of the opponent.

We defer the proofs to §4.4. The following example clarifies the essence of the result of Theorem 4.

EXAMPLE 10 (EXAMPLE 9 CONTINUED). Recall the setting of Example 9, where two possible calibrated forecasts were considered: (i) with $\|y_k - q_k\| \to 0$, and (ii) with $y_k \to \alpha q^{(1)} + \bar{\alpha} q^{(2)} \triangleq q_0$. Observe that in the first case, Theorem 4 implies that $\bar{r}_n$ converges to $R^+(\{q^{(1)}, q^{(2)}\})$. In particular, if $r^*(q)$ is continuous at $q^{(1)}$ and $q^{(2)}$, $\bar{r}_n$ converges to $\mathrm{conv}(\{r^*(q^{(1)}), r^*(q^{(2)})\})$. In the second case, Theorem 4 implies that $\bar{r}_n$ converges to $R^+(\{q_0\})$. In particular, if $r^*(q)$ is continuous at q_0, $\bar{r}_n$ converges to $r^*(q_0) \in S$. In the special case, where the opponent chooses at each time instant the mixed action $q^{(1)}$ with probability α and $q^{(2)}$ with probability $\bar{\alpha}$, it is easy to see that only the calibrated forecast of case (ii) above is possible, and hence Theorem 4 implies that $\bar{r}_n$ converges to $R^+(\{q_0\})$. □

The following corollary establishes the opportunistic approachability property of the calibrated approachability algorithm illustrated by Example 10.

COROLLARY 2. *Consider the setting of Theorem 4. For any strategy of the opponent, the following implication holds true almost surely (i.e., on a set of probability 1): if the play of the opponent is statistically Q-restricted as per Definition 4, then*

$$\lim_{n\to\infty} d(\bar{r}_n, R^+(Q)) = 0,$$

where $R^+(Q)$ is the closed convex image of Q under r^* (Definition 8).

That is, the strategy specified by the calibrated approachability algorithm is statistically opportunistic in the sense of Definition 9. Specifically, if $Q = \{q_0\}$, where q_0 is a continuity point of the (piecewise continuous) response map $p^*(q)$, we have that $\lim_{n\to\infty} \bar{r}_n = r^*(q_0) \in S$.

4.4. Proofs.

PROOF OF THEOREM 4. We prove below that, for a general opponent,

$$\lim_{n\to\infty} \left\| \bar{r}_n - \frac{1}{n}\sum_{k=1}^{n} r(p_k, y_k) \right\| = 0, \quad \text{a.s.} \tag{10}$$

Now,

$$\frac{1}{n}\sum_{k=1}^{n} r(p_k, y_k) = \frac{1}{n}\sum_{k=1}^{n} r(p^*(y_k), y_k) = \frac{1}{n}\sum_{k=1}^{n} r^*(y_k) \in \mathrm{conv}(S).$$

But, by the definition of R_n^+ in (8),

$$\lim_{n\to\infty} d\left(\frac{1}{n}\sum_{k=1}^{n} r^*(y_k), R_n^+\right) = 0,$$

and the result of the theorem follows.

Fix $\epsilon > 0$. By compactness of $\Delta(\mathcal{A})$, there exists a partition of $\Delta(\mathcal{A})$ into a finite number l of measurable sets $P^1, P^2, \ldots, P^l$, with the property that if $p, p' \in P^i$ then $\|p - p'\| \le \epsilon$. That is, $\{P^i\}$ is an ϵ-partition of $\Delta(\mathcal{A})$. Also, let $Q^i = (p^*)^{-1}(P^i)$. By our definition of p^* (Definition 6), $\{Q^i\}$ are measurable sets that represent a partition of $\Delta(\mathcal{Z})$ (although not necessarily an ϵ partition), and since $p_k = p^*(y_k)$ we have $\mathbb{I}\{p_k \in P^i\} = \mathbb{I}\{y_k \in Q^i\}$. Finally, for every i, we fix a representative element $p^i \in P^i$ (e.g., central point of P^i). We have that

$$\begin{aligned}
\lim_{n\to\infty} \left\| \bar{r}_n - \frac{1}{n}\sum_{k=1}^{n} r(p_k, y_k) \right\| &= \lim_{n\to\infty} \left\| \bar{r}_n - \frac{1}{n}\sum_{k=1}^{n}\sum_{i=1}^{l} \mathbb{I}\{y_k \in Q^i\} r(p_k, y_k) \right\| \\
&\le \rho\epsilon + \lim_{n\to\infty} \left\| \bar{r}_n - \frac{1}{n}\sum_{k=1}^{n}\sum_{i=1}^{l} \mathbb{I}\{y_k \in Q^i\} r(p^i, y_k) \right\| \\
&= \rho\epsilon + \lim_{n\to\infty} \left\| \bar{r}_n - \sum_{i=1}^{l} r(p^i, \cdot) \frac{1}{n}\sum_{k=1}^{n} \mathbb{I}\{y_k \in Q^i\} y_k \right\|
\end{aligned}$$

$$
\begin{aligned}
&= \rho\epsilon + \lim_{n\to\infty} \left\| \bar{r}_n - \sum_{i=1}^{l} r(p^i,\cdot) \frac{1}{n} \sum_{k=1}^{n} \mathbb{I}\{y_k \in Q^i\} \mathbf{1}(z_k) \right\| \\
&= \rho\epsilon + \lim_{n\to\infty} \left\| \bar{r}_n - \frac{1}{n} \sum_{k=1}^{n} \sum_{i=1}^{l} \mathbb{I}\{p_k \in P^i\} r(p^i, z_k) \right\| \\
&\le 2\rho\epsilon + \lim_{n\to\infty} \left\| \bar{r}_n - \frac{1}{n} \sum_{k=1}^{n} \sum_{i=1}^{l} \mathbb{I}\{p_k \in P^i\} r(p_k, z_k) \right\| \\
&= 2\rho\epsilon.
\end{aligned}
$$

In this sequence, the first inequality holds since when $y_k \in Q^i$, we have $p_k \in P^i$ and

$$\|r(p_k, z_k) - r(p^i, z_k)\| \le \rho \|p_k - p^i\| \le \rho\epsilon, \tag{11}$$

where ρ is the span of the reward function (1). In the second equality, we use $r(p^i,\cdot)$ to denote the reward matrix that corresponds to a mixed strategy p^i, so that $r(p^i,\cdot)y_k = r(p^i, y_k)$. The third equality follows by the calibration property (Definition 11). Finally, the second inequality above holds again by (11). Since this inequality holds for any $\epsilon > 0$, the result follows. □

To prove Corollary 2, we need the following important result.

Lemma 6. *For any strategy of the opponent, the following implication holds true almost surely*: *if the sequence of the opponent's mixed actions is statistically Q-restricted as per Definition 4, then the sequence of calibrated forecasts* $\{y_k\}_{k=1}^{\infty}$ *is also statistically Q-restricted.*

Proof. Recall that by Lemma 4, the joint empirical distribution μ_n "converges" to the set $\mathcal{M}$ defined in (4). In addition, the Cesàro-convergence of q_k is equivalent to convergence of the marginal empirical distribution f_n of $\{q_k\}_{k=1}^{n}$ to the set

$$\left\{ \mu_1 \colon \int d(q, Q)\mu_1(dq) = 0 \right\}.$$

Using Lemma 5 with $V(\cdot)$ being the (convex) Euclidean point-to-set distance $d(\cdot, Q)$, we also have that the marginal empirical distribution g_n of $\{y_k\}_{k=1}^{n}$ satisfies

$$\lim_{n\to\infty} \frac{1}{n} \sum_{k=1}^{n} d(y_k, Q) = \lim_{n\to\infty} \int d(y, Q) g_n(dy) \le \lim_{n\to\infty} \int d(q, Q) f_n(dq) = 0. \quad \square$$

Proof of Corollary 2. The result follows by Lemma 6 since for $n \to \infty$, the support of the empirical distribution of the forecasts, $g_n \triangleq \mu_{n,2}$, is restricted to Q. Therefore

$$\lim_{n\to\infty} d\left(\frac{1}{n} \sum_{k=1}^{n} r^*(y_k), R^+(Q) \right) = \lim_{n\to\infty} d\big(\mathbb{E}_{Y\sim g_n}[r^*(Y)], R^+(Q) \big) = 0$$

by the definition of $R^+(Q)$. In particular, consider the case $Q = \{q_0\}$, where q_0 is a continuity point of the (piecewise continuous) response map $p^*(q)$. Then, q_0 is also a continuity point of $r^*(q)$, and $R^+(Q) = \{r(p^*(q_0), q_0)\} = \{r^*(q_0)\}$ is a singleton by its definition. □

4.5. Additional remarks. The rate of convergence of (9) is that of the calibrated forecast used. E.g., for ϵ-calibration (and thus, ϵ approachability) using internal regret minimization, the rate is the standard rate of convergence of no-regret algorithms, that is of $O(1/\sqrt{n})$ (Cesa-Bianchi and Lugosi [9]).

Our algorithm assumes that an *exact* calibration algorithms is used. If, instead, an ϵ-calibration forecaster is employed, our results carry over with minor modifications as follows. First, the set $\mathcal{M}$ should be replaced with

$$\mathcal{M}_\epsilon \triangleq \left\{ \mu \in \Delta(\Delta(\mathcal{Z}) \times \Delta(\mathcal{Z})) \colon \left\| \int \mathbb{I}\{y \in Q\}(q - y) d\mu(q, y) \right\|_2 \le \epsilon, \forall\, Q \subseteq \Delta(\mathcal{Z}) \right\}.$$

Also, it is easy to see that the convergence results of Theorem 4 and Corollary 2 hold with $\lim_{n\to\infty}(\cdot) = 0$ replaced by $\limsup_{n\to\infty}(\cdot) \le \epsilon$. Finally, instead of using the exact closed convex image of a set, $R^+(Q)$, one should use

$$R^+_\epsilon(Q) \triangleq \operatorname{conv}\{r^*(q) \colon d(q, Q) \le \epsilon\}.$$

This is required since in the case of ϵ-calibration, Cesàro-convergence condition of Corollary 2 only implies that

$$\limsup_{n\to\infty} \frac{1}{n} \sum_{k=1}^{n} d(y_k, Q) \le \epsilon, \quad \text{a.s.}$$

5. Approachability using slowly varying calibration. In §4, we analyzed the behaviour of a basic calibrated approachability algorithm and proved its opportunistic properties against a statistically restricted play of the opponent. In this section, we turn to the analysis of our algorithm in the case when the opponent's play is *empirically* restricted. In particular, we pose the following question: does the fact that the empirical frequencies of the pure (observed) actions are restricted to a set Q in the sense of Definition 5 implies that the calibrated approachability algorithm (7) converges to $R^+(Q)$ (similarly to the result of Corollary 2 for the case of statistically Q-restricted opponent)? The following example shows that this is not necessarily the case.

Example 11 (Example 6 continued). Recall the setting of Example 6, where the goal is to approach the nonconvex set $S = (-\infty, -1] \cup [1, \infty)$. Suppose that the opponent's actions are $0, 0, 1, 0, 0, 1, \ldots$, implying that $\bar{q}_n \to q_0 = 2/3$. An opportunistic approachability algorithm should ideally converge in this case to $R^+(\{q_0\})$ (see Definition 10). Indeed, the fixed forecaster $y_n = 2/3$ is calibrated, and the calibrated approachability algorithm that uses this forecaster will approach

$$R^+(\{q_0\}) = r(p^*(q_0), q_0) = r\big((1,0), (\tfrac{2}{3}, \tfrac{1}{3})\big) = \tfrac{4}{3},$$

where the first equality follows since q_0 is a continuity point of the response function p^* defined in (2). Now since $4/3 \in S$, the algorithm will approach S. However, consider a *perfect* forecaster that predicts $y_n = \mathbf{1}(z_n)$. If the calibrated approachability algorithm uses this forecaster, it approaches

$$\tfrac{2}{3} r^*((1,0)) + \tfrac{1}{3} r^*((0,1)) = \tfrac{2}{3} r((1,0),(1,0)) + \tfrac{1}{3} r((0,1),(0,1)) = \tfrac{2}{3},$$

which is *not* in S. Hence, in this case, only convergence to $\mathrm{conv}(S)$ is guaranteed. □

This example illustrates the fact that a perfect forecaster is bad for the purpose of empirically opportunistic approachability. In fact, we would prefer a fixed forecaster, or more generally, a *slowly time-varying* forecaster. This motivates us to introduce the following assumption in terms of the *probability distributions* of the forecasts $\{\eta_n\}$. To this end, for any probability measures $\eta_1, \eta_2 \in \Delta(\Delta(\mathcal{Z}))$, let

$$\|\eta_1 - \eta_2\|_{\mathrm{TV}} \triangleq \sup_{A \subseteq \Delta(\mathcal{Z})} |\eta_1(A) - \eta_2(A)|$$

denote the *total variation distance*, where the suprimum is taken over Borel-measurable sets $A \subseteq \Delta(\mathcal{Z})$.

Assumption 3 (Slowly Varying Calibration Algorithm). *The probability distribution η_n is changing slowly. Namely, there exists $n_0 < \infty$ such that for all $n \geq n_0$,*

$$\|\eta_n - \eta_{n-1}\|_{TV} \leq \frac{C}{n^\xi},$$

for some $\xi > 0$ and $C < \infty$.

We note that Assumption 3 is *not probabilistic* since it is stated in terms of the randomizing probabilities of the calibrated forecaster (and not in terms of the actual forecasts, which are random). In §5.2, we will show that there exists a specific calibration algorithm that satisfies this property (see Corollary 3). We leave open the interesting question of whether some slow variation property, in the spirit of the above, is intrinsically related to the calibration requirement, or is a property of the specific algorithm used.

5.1. Approachability result. The following theorem shows that if the calibrated approachability algorithm uses a slowly varying calibrated forecaster, it is empirically opportunistic[2] in the sense of Definition 10.

Theorem 5. *Suppose that Assumption 2 holds, and a calibration algorithm satisfies Assumption 3 with a parameter $\xi > 0$. For any strategy of the opponent, the following implication then holds true almost surely: if the play of the opponent is empirically Q-restricted (as per Definition 5) with respect to a partition $\{\tau_m\}$ with either*

(1) *bounded blocks lengths $\tau_m \leq \bar{\tau} < \infty$, or*
(2) *growing blocks lengths $\tau_m = O(m^\nu)$ with $\nu > 0$, under the condition that $\xi > \nu/(\nu+1)$,*

then,

$$\lim_{n \to \infty} d(\bar{r}_n, R^+(Q)) = 0.$$

[2] Note that it is statistically opportunistic as well, as follows by the result of Lemma 2.

Below is an outline of the proof. We first claim in Lemma 7 that the calibration property (3) can be stated in terms of the *distributions* of the forecasts η_n (rather than the forecasts y_n themselves). We then prove in Lemma 8 that Assumption 3 actually implies a calibration property in terms of the *empirical frequencies* of the actions (rather than pure or mixed actions themselves). Finally, we show that this last property implies opportunistic approachability in terms of the empirical frequencies.

Lemma 7. *The calibration property* (3) *is equivalent to the following lifted calibration property in terms of the forecast distributions* η_n:

$$\lim_{n\to\infty}\frac{1}{n}\sum_{k=1}^{n}\big(\mathbb{E}_{y_k\sim\eta_k}[\mathbb{I}\{y_k\in Q\}]\mathbf{1}(z_k)-\mathbb{E}_{y_k\sim\eta_k}[\mathbb{I}\{y_k\in Q\}y_k]\big)=0. \tag{12}$$

The proof of Lemma 7 is based on standard analysis of a suitably defined martingale difference sequence. See Appendix B for details.

Recall that $n_M \triangleq \sum_{m=1}^{M}\tau_m$, and $\hat{q}_m$ denotes the empirical distribution of the opponent's actions at block m. Also, let

$$\tilde{\eta}_m=\eta_{k_m},$$

be *any* sequence of forecast distributions with the corresponding (deterministic) index subsequence $n_{m-1}+1\le k_m\le n_m$. We show below that Assumption 3 implies the following calibration property in terms of the empirical distributions $\{\hat{q}_m\}$.

Definition 12 (Calibration for Empirical Frequencies). A calibrated forecaster is said to be *calibrated for empirical frequencies* on a given partition $\{\tau_m\}$, if the forecast distributions can be *fixed* during each block, without violating the (lifted) calibration property (12). That is,

$$\begin{aligned}&\lim_{M\to\infty}\frac{1}{n_M}\sum_{m=1}^{M}\sum_{k=n_{m-1}+1}^{n_m}\big(\mathbb{E}_{y_k\sim\tilde{\eta}_m}[\mathbb{I}\{y_k\in Q\}]\mathbf{1}(z_k)-\mathbb{E}_{y_k\sim\tilde{\eta}_m}[\mathbb{I}\{y_k\in Q\}y_k]\big)\\&\quad=\lim_{M\to\infty}\frac{1}{n_M}\sum_{m=1}^{M}\tau_m\big(\mathbb{E}_{\tilde{y}_m\sim\tilde{\eta}_m}[\mathbb{I}\{\tilde{y}_m\in Q\}]\hat{q}_m-\mathbb{E}_{\tilde{y}_m\sim\tilde{\eta}_m}[\mathbb{I}\{\tilde{y}_m\in Q\}\tilde{y}_m]\big)=0,\end{aligned} \tag{13}$$

for all $\{\tilde{\eta}_m\}$ with $\tilde{\eta}_m=\eta_{k_m}$ and $n_{m-1}+1\le k_m\le n_m$, and all Borel-measurable $Q\subseteq\Delta(\mathcal{Z})$.

The following lemma shows that a calibration algorithm is calibrated for empirical frequencies if it is slowly varying.

Lemma 8. *Suppose that a calibration algorithm satisfies Assumption* 3 *with a parameter* $\xi>0$. *Then, it is calibrated for empirical frequencies* (*as per Definition* 12) *for any partition with either*

(1) *bounded blocks lengths* $\tau_m\le\bar{\tau}<\infty$, *or*
(2) *growing blocks lengths* $\tau_m=O(m^\nu)$ *with* $\nu>0$, *under the condition that* $\xi>\nu/(\nu+1)$.

The proof of Lemma 8 is based on fixing the forecast distribution during each block and employing the slow varying calibration property to bound the difference of the corresponding expected values. The detailed proof can be found in Appendix B.

Finally, we prove Theorem 5.

Proof of Theorem 5. Let $\tilde{\mu}_M$ be the empirical joint distribution of $\{\hat{q}_m\}$ and $\{\tilde{\eta}_m\}$ using the partition $\{\tau_m\}$, that is

$$\tilde{\mu}_M=\frac{1}{n_M}\sum_{m=1}^{M}\tau_m\delta_{\hat{q}_m,\,\tilde{\eta}_m}\in\Delta\big(\Delta(\mathcal{Z})\times\Delta(\Delta(\mathcal{Z}))\big).$$

(I.e., $\tilde{\mu}_M$ is a distribution over pairs (q,η), where q is a probability vector in $\Delta(\mathcal{Z})$ and η is a distribution over probability vectors in $\Delta(\mathcal{Z})$.) Note that Lemma 8 implies that $\tilde{\mu}_M$ "converges" to

$$\bar{\mathcal{M}}\triangleq\left\{\mu\in\Delta\big(\Delta(\mathcal{Z})\times\Delta(\Delta(\mathcal{Z}))\big):\ \int\big(\mathbb{E}_{y\sim\eta}[\mathbb{I}\{y\in Q\}]q-\mathbb{E}_{y\sim\eta}[\mathbb{I}\{y\in Q\}y]\big)\,d\mu(q,\eta)=0,\forall\,Q\subseteq\Delta(\mathcal{Z})\right\}$$

for *any* choice of $\{\tilde{\eta}_m\}$, in the sense that

$$\lim_{M\to\infty}\int\big(\mathbb{E}_{y\sim\eta}[\mathbb{I}\{y\in Q\}]q-\mathbb{E}_{y\sim\eta}[\mathbb{I}\{y\in Q\}y]\big)\,d\tilde{\mu}_M(q,\eta)=0,\quad\forall\,Q\subseteq\Delta(\mathcal{Z}).$$

Now, since the opponent's play satisfies Definition 5, it holds that

$$\lim_{M\to\infty}\int d(q,Q)\tilde{\mu}_{M,1}(dq)=\lim_{M\to\infty}\frac{1}{n_M}\sum_{m=1}^{M}\tau_m d(\hat{q}_m,Q)=0. \tag{14}$$

Also, it can be easily verified that an analogue of Lemma 5 holds for the set $\bar{\mathscr{M}}$. In particular, we have that $\mu\in\bar{\mathscr{M}}$ if and only if

$$\mathbb{E}(\mathbf{q}\mid\mathbf{y})=\mathbf{y},\quad \mu\text{-a.s.} \tag{15}$$

where $(\mathbf{q},\mathbf{y})$ denote a pair of *random variables* distributed according to μ. Indeed, by the definition of the conditional expectation $\mathbb{E}(\mathbf{q}\mid\mathbf{y}=y)$, we have that

$$\mathbb{E}[\mathbb{I}\{\mathbf{y}\in Q\}\mathbf{q}]=\mathbb{E}[\mathbb{I}\{\mathbf{y}\in Q\}\,\mathbb{E}(\mathbf{q}\mid\mathbf{y}=y)],\quad \forall\, Q\subseteq\Delta(\mathcal{X}). \tag{16}$$

Now, $\mu\in\bar{\mathscr{M}}$ if and only if

$$\mathbb{E}[\mathbb{I}\{\mathbf{y}\in Q\}\mathbf{q}]=\mathbb{E}[\mathbb{I}\{\mathbf{y}\in Q\}\mathbf{y}],\quad \forall\, Q\subseteq\Delta(\mathcal{X}),$$

implying that (15) holds. Consequently, for any convex function V, it holds that

$$\mathbb{E}[V(\mathbf{y})]=\mathbb{E}[V(\mathbb{E}(\mathbf{q}\mid\mathbf{y}))]\le\mathbb{E}[\mathbb{E}(V(\mathbf{q})\mid\mathbf{y})]=\mathbb{E}[V(\mathbf{q})],$$

or, equivalently

$$\int_\eta\mathbb{E}_{y\sim\eta}[V(y)]\mu_2(d\eta)\le\int_q V(q)\mu_1(dq).$$

We use this last result to prove the restriction property of $\{y_n\}$ as in the proof of Lemma 6. In particular, using $V(q)=d(q,Q)$ and (14), we have that

$$\begin{aligned}\lim_{M\to\infty}\frac{1}{n_M}\sum_{m=1}^{M}\tau_m\,\mathbb{E}_{\tilde{y}_m\sim\tilde{\eta}_m}d(\tilde{y}_m,Q)&=\lim_{M\to\infty}\int\mathbb{E}_{y\sim\eta}[d(y,Q)]\tilde{\mu}_{M,2}(d\eta)\\&\le\lim_{M\to\infty}\int d(q,Q)\tilde{\mu}_{M,1}(dq)=0.\end{aligned}$$

Therefore,

$$\lim_{M\to\infty}\frac{1}{n_M}\sum_{k=1}^{n_M}\mathbb{E}_{y_k\sim\eta_k}d(y_k,Q)\le\lim_{M\to\infty}\frac{1}{n_M}\sum_{m=1}^{M}\tau_m\,\mathbb{E}_{y_m^*\sim\eta_m^*}d(y_m^*,Q)=0,$$

where

$$\eta_m^*\in\operatorname*{arg\,max}_{n_{m-1}+1\le k\le n_m}\mathbb{E}_{y_k\sim\eta_k}d(y_k,Q).$$

Consequently,

$$\lim_{M\to\infty}\frac{1}{n_M}\sum_{k=1}^{n_M}d(y_k,Q)=0,\quad \text{a.s.}$$

by the strong law of large numbers applied to the martingale difference sequence $\mathbb{E}_{y_k\sim\eta_k}d(y_k,Q)-d(y_k,Q)$, and the result of the theorem follows similarly to the proof of Corollary 4. □

Remark 5. Observe that if the forecast distribution is *fixed* during each block, then Definition 12 is equivalent to the definition of calibration (see (12)). In general, this is of course not true. Lemma 8 shows that Definition 12 is satisfied for slowly varying calibrated forecasters. In fact, it is sufficient that the requirement (13) of Definition 12 holds for a given sample path of the opponent's play in order to obtain the result of Theorem 5.

5.2. The existence of slowly varying calibration algorithms. In this section, we show that there exists a calibration algorithm that satisfies Assumption 3. The algorithm that we analyze is based on a specific method for *internal regret minimization*. We present a general connection between calibration and internal no-regret in §5.2.2. But first, we prove a result of independent interest that establishes a slowly varying property of a specific internal no-regret algorithm.

5.2.1. Slowly varying internal regret minimization. Consider, as before, a repeated two-person game between an agent and an arbitrary opponent. At time instant n, the agent chooses its action I_n from a finite set $\{1, 2, \ldots, N\}$ randomly according to a probability distribution $\eta_n \in \Delta(\{1, 2, \ldots, N\})$, while the opponent chooses its action $z_n \in \mathcal{Z}$. The average expected loss incurred by the agent up to time n is

$$\bar{\ell}_n = \frac{1}{n} \sum_{k=1}^{n} \ell(\eta_k, z_k),$$

where $\ell(i, z)$ is a given one-stage loss function, and $\ell(\eta, z)$ is its expected value with respect to η.

For each $i \neq j \in \{0, 1, \ldots, N\}$, let $\eta_n^{i \to j} \in \Delta(\{0, 1, \ldots, N\})$ be defined by

$$\eta_n^{i \to j}(l) = \begin{cases} 0, & l = i, \\ \eta_n(i) + \eta_n(j), & l = j, \\ \eta_n(l), & \text{otherwise.} \end{cases} \tag{17}$$

That is, this is a probability distribution that transfers the weight of i to j. We say that the agent *minimizes internal no-regret* with respect to the loss function ℓ if

$$\limsup_{n \to \infty} \left\{ \frac{1}{n} \sum_{k=1}^{n} \ell(\eta_k, z_k) - \min_{i \neq j} \frac{1}{n} \sum_{k=1}^{n} \ell(\eta_k^{i \to j}, z_k) \right\} \leq 0, \tag{18}$$

for any strategy of the opponent. See Cesa-Bianchi and Lugosi [9] for an overview of internal no-regret algorithms.

Below, we analyze a specific internal no-regret algorithm from Cesa-Bianchi and Lugosi [9] that uses *exponentially weighted average strategy* to define η_n. In particular, let

$$\Delta_{(i,j),n} \triangleq \frac{\exp(-\alpha_n \sum_{k=1}^{n-1} \ell(\eta_k^{i \to j}, z_k))}{\sum_{l \neq l'} \exp(-\alpha_n \sum_{k=1}^{n-1} \ell(\eta_k^{l \to l'}, z_k))} \tag{19}$$

with $\alpha_n = 1/\sqrt{n}$. Also, define the stochastic matrix P_n with the following elements:

$$(P_n)_{i,j} \triangleq \begin{cases} \Delta_{(i,j),n}, & i \neq j \\ \sum_{l \neq i, l' \neq l} \Delta_{(l,l'),n}, & i = j. \end{cases} \tag{20}$$

Note that $(P_n)_{i,j}$ can be interpreted as a *transition probability* from i to j in the sense of transferring the weight from i to j in the previously used strategies η_k, $k = 1, \ldots, n-1$. The internal regret minimizing strategy at time n, η_n, is then defined as a solution of the fixed point equation

$$\begin{aligned} \eta^\top &= \eta^\top P_n, \\ \sum_i \eta(i) &= 1. \end{aligned} \tag{21}$$

Therefore, η_n is a stationary distribution of the Markov chain that corresponds to P_n. Note that this solution is unique.

Lemma 9. *The Markov chains that correspond to P_n are irreducible and aperiodic, and therefore there exists a unique solution to (21) for all n.*

Proof. Note that the state space is finite, and for all n

$$\alpha_n \sum_{k=1}^{n-1} \ell(\eta_k^{i \to j}, z_k) < \infty.$$

Therefore, $(P_n)_{i,j} > 0$ for all i, j, and in particular $(P_n)_{i,i} > 0$. Hence the chains are irreducible and aperiodic. □

The next theorem establishes that the distributions of the agent's actions change slowly with time.

Theorem 6. *The distributions η_n satisfy Assumption 3 for all $n \geq 3$.*

Proof Outline. This theorem follows by the smoothness property of the transition matrices P_n, which in turn implies smoothness of the corresponding stationary distributions η_n. The detailed proof is rather involved and can be found in Appendix C. □

5.2.2. Calibration and internal no-regret. Next, we outline a general connection between calibration and internal no-regret. For simplicity, we focus on the binary case $\mathcal{Z} = \{0, 1\}$ and ϵ-calibration algorithms. The results for exact calibrated forecasters and for general action set $\mathcal{Z}$ follow by the arguments similar to those in Cesa-Bianchi and Lugosi [9].

For brevity, we let $y_n \in [0, 1]$ denote the forecast of $z_n = 1$. As was mentioned, the construction of the calibration algorithm starts from discretization of $[0, 1]$ into N intervals. For a fixed N, an ϵ-calibration algorithm is then constructed, where $\epsilon \to 0$ as $N \to \infty$. A calibrated forecast can then be obtained by using a simple application of the doubling trick, letting $N \to \infty$.

We present the connection between *any* ϵ-calibration algorithm and internal regret minimization of the *squared* loss function. Assume that the ϵ-calibrated forecaster is given by

$$y_n = \frac{I_n}{N},$$

where $I_n \in \{0, 1, \ldots, N\}$ is randomly selected according to a probability distribution

$$\eta_n \in \Delta(\{0, 1, \ldots, N\}).$$

The following result was shown in Cesa-Bianchi and Lugosi [9], using the so-called *Brier score*.

Lemma 10. *A forecaster is ϵ-calibrated if and only if it minimizes the internal regret with respect to the loss function*

$$\ell(i, z) \triangleq \left(\frac{i}{N} - z\right)^2. \tag{22}$$

Hence, we have the following corollary that establishes the existence of slowly varying calibrated forecasters.

Corollary 3. *The calibrated forecaster that is based on internal regret minimization technique presented in §5.2.1 satisfies Assumption 3 for all $n \geq 3$.*

Proof. The result follows by Theorem 6 and Lemma 10. □

6. Constrained regret minimization. We next apply our opportunistic approachability framework to the problem of regret minimization subject to average cost constraints (Mannor et al. [30]).

Consider first the standard (unconstrained) regret minimization problem, where as before, the agent faces an arbitrarily varying environment (the opponent). The repeated game model is the same as above, except that the vector reward function r is replaced by a scalar reward (or utility) function u: $\mathcal{A} \times \mathcal{Z} \to \mathbb{R}$. Let $\bar{u}_n \triangleq n^{-1} \sum_{k=1}^{n} u_k$ denote the average reward by time n. The goal of the agent is to maximize $\bar{u}_n$. Suppose that the agent knew in advance that the empirical distribution $\bar{q}_n$ of the opponent's actions is say $\bar{q}_n = q$. He could then maximize its average reward by repeatedly choosing the action that solved

$$u^*(q) = \max_{p \in \Delta(\mathcal{Z})} u(p, q) = \max_{a \in \mathcal{Z}} u(a, q).$$

However, in the online setting, when the actions of the opponent can be arbitrary and are not known in advance, a suitable goal introduced in Hannan [15] is to minimize the *regret*, namely, to ensure that

$$\limsup_{n \to \infty} (u^*(\bar{q}_n) - \bar{u}_n) \leq 0, \tag{23}$$

almost surely, for every strategy of the opponent. Right after Hannan's seminal paper, Blackwell [6] used approachability theory in order to elegantly show the existence of regret minimizing algorithms. Define the vector-valued rewards $r_n \triangleq (u_n, \mathbf{1}(z_n)) \in \mathbb{R} \times \Delta(\mathcal{Z})$. The corresponding average reward is then $\bar{r}_n \triangleq n^{-1} \sum_{k=1}^{n} r_k = (\bar{u}_n, \bar{q}_n)$. Finally, define the target set

$$S = \{(u, q) \in \mathbb{R} \times \Delta(\mathcal{Z}) \colon u \geq u^*(q)\}.$$

It can be easily verified that this set is a D-set, and it is convex by the convexity of $u^*(q)$. Hence, S is approachable, and by the continuity of $u^*(q)$, an algorithm that approaches S also minimizes the regret in the sense of (23).

In the *constrained* regret minimization problem, in addition to the scalar reward function u, we are given a vector-valued cost function c: $\mathcal{A} \times \mathcal{Z} \to \mathbb{R}^s$. We are also given a closed and convex set $\Gamma \subseteq \mathbb{R}^s$, the constraint

set, defining the allowed values for the long-term average cost (see below). The typical case is that of linear constraints, that is $\Gamma = \{c \in \mathbb{R}^s\colon\ c_i \le \gamma_i, i = 1, \ldots, s\}$ for some vector $\gamma \in \mathbb{R}^s$. The constraint set is assumed to be *feasible*, in the sense that for every $q \in \Delta(\mathcal{Z})$, there exists $p \in \Delta(\mathcal{A})$, such that $c(p, q) \in \Gamma$.

Let $\bar{c}_n \triangleq n^{-1}\sum_{k=1}^n c_k$ denote the average cost by time n. The agent is required to satisfy the cost constraints, in the sense that $\lim_{n\to\infty} d(\bar{c}_n, \Gamma) = 0$ must hold, irrespectively of the opponent's play. Subject to these constraints, the agent wishes to maximize his average reward $\bar{u}_n$.

Suppose the agent knew in advance that the empirical distribution $\bar{q}_n$ equals to q. He could then maximize its expected average reward subject to the constraints by always choosing the mixed action p that solved the following program:

$$u^*_\Gamma(q) \triangleq \max_{p\in\Delta(\mathcal{A})} \{u(p, q)\colon\ c(p, q) \in \Gamma\}. \tag{24}$$

We consider $u^*_\Gamma(q)$ as the *best-reward-in-hindsight* for the constrained problem. The goal of the agent would be then to attain u^*_Γ in the following sense.

Definition 13 (Constrained No-Regret). A strategy of the agent π is a *constrained no-regret strategy with respect to a function* u^*_Γ if (i) $\limsup_{n\to\infty}(u^*_\Gamma(\bar{q}_n) - \bar{u}_n) \le 0$; and (ii) $\lim_{n\to\infty} d(\bar{c}_n, \Gamma) = 0$ both hold almost surely, for every strategy of the opponent. If such a strategy exists, we say that $u^*_\Gamma(\cdot)$ is *attainable*.[3]

The problem of attaining $u^*_\Gamma(\cdot)$ can be formulated as an approachability problem, which extends Blackwell's original formulation for the unconstrained case presented above. Define the vector-valued rewards $r_n \triangleq (u_n, c_n, \mathbf{1}(z_n)) \in \mathbb{R}^{s+1} \times \Delta(\mathcal{Z})$. The corresponding average reward becomes $\bar{r}_n \triangleq n^{-1}\sum_{k=1}^n r_k = (\bar{u}_n, \bar{c}_n, \bar{q}_n)$. Finally, define the target set

$$S = \{(u, c, q) \in \mathbb{R}^{s+1} \times \Delta(\mathcal{Z})\colon\ u \ge u^*_\Gamma(q), c \in \Gamma\}. \tag{25}$$

It is easily verified that an algorithm that approaches S also attains $u^*_\Gamma(q)$ if $u^*_\Gamma(q)$ is continuous. Furthermore, the set S is a D-set by construction (see below). However, the function $u^*_\Gamma(q)$ is *not* convex in general, which implies that the set S is not convex. Therefore, one cannot invoke the dual condition to infer approachability of S, but only of its convex hull. Indeed, it was shown in Mannor et al. [30] that S is not approachable in general.

A feasible (approachable) target set is then the convex hull of S. This may be written as

$$\mathrm{conv}(S) = \{(u, c, q) \in \mathbb{R}^{s+1} \times \Delta(\mathcal{Z})\colon\ u \ge \mathrm{conv}(u^*_\Gamma)(q), c \in \Gamma\}, \tag{26}$$

where the function $\mathrm{conv}(u^*_\Gamma)$ is the lower convex hull of $u^*_\Gamma(\cdot)$ (i.e., the largest convex function over $\Delta(\mathcal{Z})$ that is smaller than u^*_Γ). Now, since $\mathrm{conv}(S)$ is approachable, it follows that $\mathrm{conv}(u^*_\Gamma)(q)$ is attainable, in the sense of Definition 13.

Two algorithms were proposed in Mannor et al. [30] for attaining the relaxed goal function $\mathrm{conv}(u^*_\Gamma)$. The first is a standard approachability algorithm for $\mathrm{conv}(S)$, which requires the demanding calculation of projection directions to the convex hull of S. Further, this algorithm is not opportunistic, in the sense described below. The second algorithm relies on computing calibrated forecasts of the opponent's actions, and as we show below is actually equivalent to our calibration-based approachability algorithm when used for this special case. Our opportunistic convergence results thus apply to this algorithm.

To apply our algorithm, a regular response function p^* (Definition 6) is required. It is easily seen that any choice of

$$p^*(q) \in \arg\max_{p\in\Delta(\mathcal{A})} \{u(p, q)\colon\ c(p, q) \in \Gamma\}$$

yields a response function, in the sense that $r(p^*(q), q) \in S$. A regular (piecewise continuous) selection can be induced, for example, by imposing a lexicographic precedence over the coordinates of p in case the maximizing set is not a singleton. The goal function in this case is then

$$r^*(q) = (u^*_\Gamma(q), c(p^*(q), q), q). \tag{27}$$

Thus, our calibrated approachability algorithm can be applied, and the results of §§4 and 5 imply that the algorithm approaches $\mathrm{conv}(S)$, hence attains the relaxed goal function $\mathrm{conv}(u^*_\Gamma)$. In particular, Corollary 2 and Theorem 5 show that S itself is approached whenever the opponent is either statistically or empirically restricted to a singleton $Q = \{q_0\}$ that is a continuity point of $p^*(q)$. Interestingly, in the present case the last continuity requirement can be removed.

[3] The term "attainability" was recently used by Lehrer et al. [24] in a different context, to describe a certain kind of approachability. In this paper, we, however, use this term as in Mannor and Shimkin [25] and Mannor et al. [30] to describe the goals of a generalized no-regret algorithm.

LEMMA 11. *For the model of the present section,* $R^+(\{q\}) \subseteq S$ *(rather than* $\mathrm{conv}(S)$*) for every* $q \in \Delta(\mathcal{Z})$.

PROOF. Observe that the first component of r^* (defined in (27)) is continuous in q (see Mannor et al. [30]). Also, note that the jumps of $c(p^*(q), q)$, the second component of r^*, lie entirely in S. This is true since, for the fixed first component, the induced set is convex because of convexity of Γ. Consequently, the jumps of $r^*(q)$ around a given $q \in \Delta(\mathcal{Z})$ lie in S, which implies that $R^+(\{q\}) \subseteq S$ by its definition. □

This brings us back to our requirement for a constrained no-regret algorithm, in Definition 13. Although this requirement cannot be attained for any strategy of the opponent, we have just seen that it is attained whenever the opponent is asymptotically stationary, in the sense that its actions are (statistically or empirically) converging to a singleton. In that case, the algorithm attains $u^*_\Gamma(q)$, the best-reward-in-hindsight, rather than a relaxed goal, while satisfying the average cost constraints.

EXAMPLE (CLASSIFICATION WITH SPECIFICITY CONSTRAINTS). In Bernstein et al. [4], we considered the online problem of merging the output of multiple binary classifiers, with the goal of maximizing the true-positive rate, while keeping the false-positive rate under a given threshold $0 < \gamma < 1$. As shown there, this problem may be formulated as a constrained regret minimization problem, and provides a concrete learning application for the theory developed here.

7. Conclusion and future work. In this paper, our central goal was to formulate the concept of opportunistic approachability. We have also devised a class of calibration-based approachability algorithms and shown that they are opportunistic in the sense advocated here. The presented algorithms are computationally challenging in that they require the computation of calibrated forecasts. In addition, a procedure for the computation of the response function p^* is required, the complexity of which is problem dependent. However, given these two components, the computational burden is much lighter than standard approachability that requires computing the projection to the target set and solving a zero-sum game in every stage. Specifically, it is sometimes difficult to compute the projection to the convex hull of a nonconvex set; a step that our approach avoids.

We have applied our opportunistic approachability framework to the problem of regret minimization subject to average cost constraints, and shown that the best-reward-in-hindsight (rather than its convex relaxation) is attained when the opponent turns out to be stationary in our sense.

Below we present some topics of future interest. Our calibration-based algorithm is conceptually simple and of general applicability. However, Although considerable progress has been made recently toward the efficient computation of calibrated forecasts (Mannor and Stoltz [27], Rakhlin et al. [36], Abernethy et al. [1]), this remains a demanding task. Therefore, it should be of interest to devise alternative algorithms that are computationally efficient and have optimal convergence rates. Initial results on a new class of opportunistic approachability algorithms that is based on online convex optimization methods appear in Bernstein et al. [5]. In addition, the work in Bernstein and Shimkin [3] focuses on designing simple and efficient algorithms that are based on the concept of a response function (rather than projection), which are however not opportunistic in the sense advocated in this paper.

We note that there are several other regret minimization problems where our framework can be applied, such as online learning with global cost functions (Even-Dar et al. [11]), regret minimization in variable duration repeated games (Mannor and Shimkin [26]), and regret minimization in stochastic game models (Mannor and Shimkin [25]). In particular, as in the problem of constrained regret minimization, the best-reward-in-hindsight is not attainable in these models in general, but only its convex relaxation. Our approach only requires that we can compute the response function, and this can be done efficiently in these cases. Of course some technical adaptation is needed to account for the models' dynamics in the two latter cases, which is a subject of future work.

Acknowledgments. This research has been supported by the Israel Science Foundation [Grants 1319/11 and 920/12]. The authors wish to thank the reviewers and the editorial team for many useful comments on this paper.

Appendix A. Proof of Lemma 2. Let $\{\tau_m\}$ be a given partition and set

$$\bar{\eta}_m \triangleq \tau_m^{-1} \sum_{k=n_{m-1}+1}^{n_m} q_k.$$

First, observe that, by the convexity of the point-to-set Euclidean distance to a convex set, Definition 4 implies that

$$\lim_{M\to\infty} \frac{1}{n_M} \sum_{m=1}^{M} \tau_m d(\bar{\eta}_m, Q) = 0.$$

This is equivalent to

$$\lim_{M\to\infty}\frac{1}{n_M}\sum_{m=1}^{M}\tau_m\mathbb{I}\{d(\bar{\eta}_m,Q)>\epsilon\}=0,\quad\forall\epsilon>0. \tag{A1}$$

Fix $\epsilon>0$. We have that,

$$\begin{aligned}\frac{1}{n_M}\sum_{m=1}^{M}\tau_m\mathbb{P}\{d(\hat{q}_m,Q)>\epsilon\}&=\frac{1}{n_M}\sum_{m=1}^{M}\tau_m\mathbb{P}\{d(\hat{q}_m,Q)>\epsilon\}\mathbb{I}\{d(\bar{\eta}_m,Q)\le\epsilon/2\}\\&\quad+\frac{1}{n_M}\sum_{m=1}^{M}\tau_m\mathbb{P}\{d(\hat{q}_m,Q)>\epsilon\}\mathbb{I}\{d(\bar{\eta}_m,Q)>\epsilon/2\}.\end{aligned}$$

Now, the second term above converges to zero by (A1). The first term above can be bounded as follows:

$$\begin{aligned}&\frac{1}{n_M}\sum_{m=1}^{M}\tau_m\mathbb{P}\{d(\hat{q}_m,Q)>\epsilon\}\mathbb{I}\{d(\bar{\eta}_m,Q)\le\epsilon/2\}\\&\quad\le\frac{1}{n_M}\sum_{m=1}^{M}\tau_m\mathbb{P}\{\|\hat{q}_m-\bar{\eta}_m\|>\epsilon/2\}\mathbb{I}\{d(\bar{\eta}_m,Q)\le\epsilon/2\}\\&\quad\le\frac{1}{n_M}\sum_{m=1}^{M}\tau_m\sum_{z\in\mathcal{Z}}\mathbb{P}\left\{|\hat{q}_m(z)-\bar{\eta}_m(z)|>\frac{\epsilon}{2\sqrt{|\mathcal{Z}|}}\right\}\\&\quad\le 2|\mathcal{Z}|\frac{1}{n_M}\sum_{m=1}^{M}\tau_m\exp\left(-\frac{\tau_m\epsilon^2}{8|\mathcal{Z}|}\right),\end{aligned}$$

where the second inequality holds by union bound for Euclidean distance, and the last inequality follows by Hoeffding's inequality for the average of the martingale difference sequence

$$D_n(z)=\mathbb{I}\{z_n=z\}-q_n(z).$$

Thus, for all $\epsilon>0$ and all *superlogarithmically growing* blocks lengths $\{\tau_m\}$,

$$\lim_{M\to\infty}\frac{1}{n_M}\sum_{m=1}^{M}\tau_m\mathbb{P}\{d(\hat{q}_m,Q)>\epsilon\}=0$$

implying that

$$\lim_{M\to\infty}\frac{1}{n_M}\sum_{m=1}^{M}\tau_m\mathbb{I}\{d(\hat{q}_m,Q)>\epsilon\}=0,\quad\text{a.s.}$$

by the almost sure convergence to zero of the mean of the martingale difference sequence $D_m=\tau_m\mathbb{I}\{d(\hat{q}_m,Q)>\epsilon\}-\tau_m\mathbb{P}\{d(\hat{q}_m,Q)>\epsilon\}$. This completes the proof. □

Appendix B. Proofs of Lemmas 7 and 8.

Proof of Lemma 7. We first write

$$\begin{aligned}&\frac{1}{n}\sum_{k=1}^{n}\big(\mathbb{E}_{y_k\sim\eta_k}[\mathbb{I}\{y_k\in Q\}]\mathbf{1}(z_k)-\mathbb{E}_{y_k\sim\eta_k}[\mathbb{I}\{y_k\in Q\}y_k]\big)\\&\quad=\frac{1}{n}\sum_{k=1}^{n}\mathbb{I}\{y_k\in Q\}(\mathbf{1}(z_k)-y_k)+\frac{1}{n}\sum_{k=1}^{n}\big(\mathbb{E}_{y_k\sim\eta_k}[\mathbb{I}\{y_k\in Q\}]-\mathbb{I}\{y_k\in Q\}\big)\mathbf{1}(z_k)\\&\qquad+\frac{1}{n}\sum_{k=1}^{n}\big(\mathbb{I}\{y_k\in Q\}y_k-\mathbb{E}_{y_k\sim\eta_k}[\mathbb{I}\{y_k\in Q\}y_k]\big).\end{aligned}$$

The first term in this equation converges to zero (almost surely) by (3), and the two other terms converge to zero (almost surely) by the strong law of large numbers, applied to the martingale difference sequences

$$D_k^{(1)}=\big(\mathbb{E}_{y_k\sim\eta_k}[\mathbb{I}\{y_k\in Q\}]-\mathbb{I}\{y_k\in Q\}\big)\mathbf{1}(z_k)$$

and

$$D_k^{(2)}=\mathbb{I}\{y_k\in Q\}y_k-\mathbb{E}_{y_k\sim\eta_k}[\mathbb{I}\{y_k\in Q\}y_k],$$

respectively. □

PROOF OF LEMMA 8. On a given partition $\{\tau_m\}$, fix a sequence $\{\tilde{\eta}_m\}$ with the corresponding (deterministic) index subsequence $n_{m-1}+1 \le k_m \le n_m$. We have that

$$\begin{aligned}
&\frac{1}{n_M}\sum_{m=1}^{M}\sum_{k=n_{m-1}+1}^{n_m}\left(\mathbb{E}_{y_k\sim\tilde{\eta}_m}[\mathbb{I}\{y_k\in Q\}]\mathbf{1}(z_k)-\mathbb{E}_{y_k\sim\tilde{\eta}_m}[\mathbb{I}\{y_k\in Q\}y_k]\right)\\
&\quad=\frac{1}{n_M}\sum_{k=1}^{n_M}\left(\mathbb{E}_{y_k\sim\eta_k}[\mathbb{I}\{y_k\in Q\}]\mathbf{1}(z_k)-\mathbb{E}_{y_k\sim\eta_k}[\mathbb{I}\{y_k\in Q\}y_k]\right)\\
&\qquad+\frac{1}{n_M}\sum_{m=1}^{M}\sum_{k=n_{m-1}+1}^{n_m}\mathbf{1}(z_k)\left(\mathbb{E}_{\tilde{y}_m\sim\tilde{\eta}_m}[\mathbb{I}\{\tilde{y}_m\in Q\}]-\mathbb{E}_{y_k\sim\eta_k}[\mathbb{I}\{y_k\in Q\}]\right)\\
&\qquad+\frac{1}{n_M}\sum_{m=1}^{M}\sum_{k=n_{m-1}+1}^{n_m}\left(\mathbb{E}_{y_k\sim\eta_k}[\mathbb{I}\{y_k\in Q\}y_k]-\mathbb{E}_{\tilde{y}_m\sim\tilde{\eta}_m}[\mathbb{I}\{\tilde{y}_m\in Q\}\tilde{y}_m]\right). \qquad \text{(B1)}
\end{aligned}$$

Now, the first term above converges to zero by (12). We prove that the other two terms also converge to zero under Assumption 3. Let us assume, without loss of generality, that n_0 of Assumption 3 equals to 1. The result for any finite n_0 follows similarly. We start by observing that, for any bounded function V, we have

$$\|\mathbb{E}_{\eta_k}[V(y_k)]-\mathbb{E}_{\eta_{k-1}}[V(y_{k-1})]\|\le V_{\max}\|\eta_k-\eta_{k-1}\|_{\mathrm{TV}}. \qquad \text{(B2)}$$

Fix $m\ge 2$ and $k_1,k_2\in[n_{m-1}+1,n_m]$. It holds that

$$\|\eta_{k_1}-\eta_{k_2}\|_{\mathrm{TV}}\le\sum_{k=n_{m-1}+2}^{n_m}\|\eta_k-\eta_{k-1}\|_{\mathrm{TV}}\le C\sum_{k=n_{m-1}+1}^{n_m}\frac{1}{k^\xi}\le\frac{C\tau_m}{n_{m-1}^\xi},$$

where the second inequality follows by Assumption 3. Using this inequality and (B2) in (B1), and taking the limit, we obtain that

$$\left\|\lim_{M\to\infty}\frac{1}{n_M}\sum_{m=1}^{M}\sum_{k=n_{m-1}+1}^{n_m}\left(\mathbb{E}_{y_k\sim\tilde{\eta}_m}[\mathbb{I}\{y_k\in Q\}]\mathbf{1}(z_k)-\mathbb{E}_{y_k\sim\tilde{\eta}_m}[\mathbb{I}\{y_k\in Q\}y_k]\right)\right\|\le 2\lim_{M\to\infty}\left(\frac{\tau_1}{n_M}+\frac{C}{n_M}\sum_{m=2}^{M}\frac{\tau_m^2}{n_{m-1}^\xi}\right). \qquad \text{(B3)}$$

Now, to prove that the bound in (B3) converges to zero, consider the two cases of the partition mentioned in the hypothesis of the lemma. First, if $\tau_m\le\bar{\tau}<\infty$ for all m, we have that

$$\frac{C}{n_M}\sum_{m=2}^{M}\frac{\tau_m^2}{n_{m-1}^\xi}\le\frac{C(\bar{\tau})^2}{M}\sum_{m=2}^{M}\frac{1}{(m-1)^\xi}\le\begin{cases}\dfrac{C(\bar{\tau})^2}{M}\left[\dfrac{1}{1-\xi}(M^{1-\xi}-1)+1\right]=\dfrac{C(\bar{\tau})^2}{(1-\xi)}\left(\dfrac{1}{M^\xi}-\dfrac{\xi}{M}\right), & \xi\ne 1\\ \dfrac{C(\bar{\tau})^2\log M}{M}, & \xi=1,\end{cases}$$

where the first inequality holds since $n_m\ge m$ and in the second inequality integral upper bound of a sum is used. Thus, the bound in (B3) goes to zero in this case.

For the second case, suppose $\tau_m=m^\nu$. Using integral approximation of a sum, we have that

$$\frac{1}{\nu+1}M^{\nu+1}\le n_M\triangleq\sum_{m=1}^{M}m^\nu\le M^{\nu+1}.$$

Therefore, the right-hand side of (B3) can be bounded by

$$\begin{aligned}
\frac{\tau_1}{n_M}+\frac{C}{n_M}\sum_{m=2}^{M}\frac{\tau_m^2}{n_{m-1}^\xi}&\le\frac{\nu+1}{M^{\nu+1}}\left[1+C(\nu+1)^\xi\sum_{m=2}^{M}\frac{m^{2\nu}}{(m-1)^{(\nu+1)\xi}}\right]\\
&\le\frac{\nu+1}{M^{\nu+1}}\left[1+C(\nu+1)^\xi 2^{\nu+1}\sum_{m=2}^{M}\frac{1}{m^{(\nu+1)\xi-2\nu}}\right]\\
&\le\begin{cases}O\left(\dfrac{\log M}{M^{\nu+1}}\right), & (\nu+1)\xi-2\nu=1\\ O\left(\dfrac{1}{M^{(\nu+1)\xi-\nu}}\right), & \text{otherwise,}\end{cases}
\end{aligned}$$

where the second inequality holds since $m-1\ge m/2$ for $m\ge 2$, and the third inequality follows by integral approximation of a sum. Hence, the bound in (B3) goes to 0 for $M\to\infty$ whenever $(\nu+1)\xi-\nu>0$, or $\xi>\nu/(\nu+1)$. □

Appendix C. Proof of existence of slowly varying internal no-regret algorithms. Below we prove Theorem 6. We first express the sensitivity of change in the stationary distribution η_n due to the changes in the transition matrices P_n defined in (20). As a second step, we provide a bound on $\|P_n - P_{n-1}\|_\infty$, where $\|A\|_\infty$ is the ℓ_∞ induced norm of matrix A, namely,

$$\|A\|_\infty \triangleq \max_i \sum_j |a_{ij}|.$$

Since the measures η_n are discrete, we identify them with the corresponding probability vectors, and provide bounds for the ℓ_∞ distance $\|\eta_n - \eta_{n-1}\|_\infty$. The total variation distance can be then bounded by

$$\|\eta_n - \eta_{n-1}\|_{\mathrm{TV}} \le N\|\eta_n - \eta_{n-1}\|_\infty.$$

For the first step, we follow Meyer [31] and introduce the following definitions.

Definition 14 (Group Inverse). Let A be a given squared matrix with $\mathrm{rank}(A^2) = \mathrm{rank}(A)$. The group inverse of A is the unique matrix $A^\#$ such that $A^2A^\# = A$, $A^\#AA^\# = A^\#$, and $AA^\# = A^\#A$.

Definition 15 (Limiting Condition Number). The limiting condition number of the sequence of irreducible and aperiodic Markov chains $\{P_n\}$ is given by

$$\kappa \triangleq \sup_{n\ge 1}\left\{\max_{i,j}[(I - P_n)^\#]_{i,j}\right\} < \infty.$$

Using these definitions, the bound on the variation of η_n can be expressed as follows.

Proposition 1. *We have that,*

$$\|\eta_n - \eta_{n-1}\|_\infty \le \kappa\|P_n - P_{n-1}\|_\infty$$

for all n.

Proof. Given the result of Lemma 9, this proposition follows by applying the results for two perturbed Markov chains, presented, e.g., in Meyer [31], to a sequence of Markov chains. □

As a second step, we provide a bound on $\|P_n - P_{n-1}\|_\infty$, where P_n is given in (20). The potential difficulty in providing such a bound is the fact that the denominator in the exponentially weighted forecaster (19) can in principle go to zero as n goes to infinity. To overcome this difficulty, we introduce a modified definition of this forecaster. Let $\{\beta_n\}$ be a given sequence, and set

$$\phi_{(i,j),n} \triangleq \exp\left(\alpha_n\left[\sum_{k=1}^{n-1}\beta_k - \sum_{k=1}^{n-1}\ell(\eta_k^{i\to j}, z_k)\right]\right).$$

It is easy to see that the exponentially weighted forecaster (19) is equivalent to

$$\Delta_{(i,j),n} \triangleq \frac{\phi_{(i,j),n}}{\sum_{l\ne l'}\phi_{(l,l'),n}}. \tag{C1}$$

We first argue that we can construct the sequence $\{\beta_n\}$ so that the $\max_{i\ne j}\phi_{(i,j),n}$ is bounded from below and above in the limit. Indeed, consider the following *history-dependent* sequence:

(i) Starting from time $n = 0$, set $\beta_n = 0$ as long as

$$\exists\, i \ne j\colon \sum_{k=1}^{n-1}\ell(\eta_k^{i\to j}, z_k) \le \sqrt{n}.$$

(ii) If at time instance n_0 the above condition fails to hold true, set

$$\beta_{n_0} = \max_{i\ne j}\ell(\eta_{n_0}^{i\to j}, z_{n_0}).$$

Set $\beta_n = \beta_{n_0}$ for all $n > n_0$, until

$$\exists\, i \ne j\colon \sum_{k=1}^{n-1}\beta_k - \sum_{k=1}^{n-1}\ell(\eta_k^{i\to j}, z_k) \ge 0.$$

(iii) Suppose that the above condition is first satisfied at time instance n_1. Then, set $\beta_n = 0$ for $n \ge n_1$ as long as

$$\exists\, i \ne j\colon \sum_{k=1}^{n-1}\beta_k - \sum_{k=1}^{n-1}\ell(\eta_k^{i\to j}, z_k) \ge -\sqrt{n}.$$

When the above condition fails to hold true, go to (ii).

Hence, for all n

$$\sum_{k=1}^{n-1}\beta_k - \sum_{k=1}^{n-1}\ell(\eta_k^{i\to j}, z_k) \le 0, \quad \forall\, i \ne j,$$

and

$$\exists\, i \ne j\colon \sum_{k=1}^{n-1}\beta_k - \sum_{k=1}^{n-1}\ell(\eta_k^{i\to j}, z_k) \ge -\sqrt{n},$$

for *any* sample path of the play. This implies that

$$\frac{1}{e} \le \max_{i\ne j}\phi_{(i,j),n} \le 1, \quad \forall\, n, \tag{C2}$$

for any sample path of the play.

With this at hand, we prove the following lemma.

Lemma 12. *We have that*

$$\|P_n - P_{n-1}\|_\infty \le 16eN(N^2+1)\frac{1}{\sqrt{n}}$$

for $n \ge 3$.

Proof. First observe that by (20), we have that

$$\|P_n - P_{n-1}\|_\infty \le 2N \max_{i\ne j}|\Delta_{(i,j),n} - \Delta_{(i,j),n-1}|. \tag{C3}$$

Also, by Equation (C1), it holds that

$$\begin{aligned}|\Delta_{(i,j),n} - \Delta_{(i,j),n-1}| &\le \frac{1}{\sum_{l\ne l'}\phi_{(l,l'),n-1}}|\phi_{(i,j),n} - \phi_{(i,j),n-1}| \\ &\quad + \frac{\phi_{(i,j),n}}{(\sum_{l\ne l'}\phi_{(l,l'),n})(\sum_{l\ne l'}\phi_{(l,l'),n-1})}\sum_{l\ne l'}|\phi_{(l,l'),n} - \phi_{(l,l'),n-1}| \\ &\le (e + eN^2)\max_{l,l'}|\phi_{(l,l'),n} - \phi_{(l,l'),n-1}|, \end{aligned}\tag{C4}$$

where the second inequality follows by (C2). However,

$$\begin{aligned}\phi_{(i,j),n} &= \exp\left(\alpha_{n-1}L_{(i,j),n-1}\frac{\alpha_n}{\alpha_{n-1}}\right)\exp\big(\alpha_n[\beta_{n-1} - \ell(\eta_{n-1}^{i\to j}, z_{n-1})]\big) \\ &= (\phi_{(i,j),n-1})^{\alpha_n/\alpha_{n-1}}\exp\big(\alpha_n[\beta_{n-1} - \ell(\eta_{n-1}^{i\to j}, z_{n-1})]\big),\end{aligned}$$

where

$$L_{(i,j),n-1} \triangleq \sum_{k=1}^{n-2}\beta_k - \sum_{k=1}^{n-2}\ell(\eta_k^{i\to j}, z_k).$$

The rest of the proof is technical: it provides an explicit upper bound for the difference

$$\phi_{(i,j),n-1} - \phi_{(i,j),n} = \phi_{(i,j),n-1} - (\phi_{(i,j),n-1})^{\alpha_n/\alpha_{n-1}}\exp\big(\alpha_n[\beta_{n-1} - \ell(\eta_{n-1}^{i\to j}, z_{n-1})]\big) \tag{C5}$$

in terms of n.

Now, since

$$\lim_{n\to\infty}\frac{\alpha_n}{\alpha_{n-1}} = \lim_{n\to\infty}\frac{\sqrt{n-1}}{\sqrt{n}} = 1$$

and

$$\lim_{n\to\infty}\exp\big(\alpha_n[\beta_{n-1} - \ell(\eta_{n-1}^{i\to j}, z_{n-1})]\big) = 1,$$

we use Taylor series of the first order to approximate the function

$$f(x,y) = \phi - y\phi^x$$

around $(x_0, y_0) = (1,1)$. We have that

$$\begin{aligned}f(x,y) &= 0 + (x-1)(-\phi\log\phi) + (y-1)(-\phi) + R_1(\tilde{x},\tilde{y}) \\ &= \phi\log\phi(1-x) + \phi(1-y) + R_1(\tilde{x},\tilde{y}),\end{aligned}$$

where

$$R_1(\tilde{x},\tilde{y}) = \frac{1}{2}\left(\frac{\partial^2 f}{\partial x^2}(x-\tilde{x})^2 + 2\frac{\partial^2 f}{\partial x \partial y}(x-\tilde{x})(y-\tilde{y}) + \frac{\partial^2 f}{\partial y^2}(y-\tilde{y})^2\right)$$
$$= \frac{1}{2}(-(x-\tilde{x})^2\tilde{y}\phi^{\tilde{x}}\log^2\phi - 2(x-\tilde{x})(y-\tilde{y})\phi^{\tilde{x}}\log\phi)$$

is the Lagrange remainder, and $(\tilde{x},\tilde{y})$ is a point with $\tilde{x}$ between x and 1 and $\tilde{y}$ between y and 1.

By (C2), $0<\phi<1$. Since $x=\sqrt{n-1}/\sqrt{n}\in[1/\sqrt{2},1)$, it is easily verified (by taking derivatives) that $\phi\log(1/\phi)\le 1/e$, $\phi^x\log^2\phi\le 8/e^2$, and $\phi^x\log(1/\phi)\le\sqrt{2}/e$. Therefore,

$$\begin{aligned}|f(x,y)| &\le |\phi\log\phi||1-x|+\phi|1-y|+0.5(x-\tilde{x})^2\tilde{y}\phi^{\tilde{x}}\log^2\phi+|x-\tilde{x}||y-\tilde{y}|\phi^x\log(1/\phi)\\ &\le (1/e)(1-x)+|1-y|+0.5(x-1)^2e(8/e^2)+(1-x)|1-y|(\sqrt{2}/e)\\ &\le \frac{1}{e}(1-x)+|1-y|+\frac{4}{e}(x-1)^2+\frac{\sqrt{2}}{e}(1-x)|1-y|. \qquad \text{(C6)}\end{aligned}$$

Now, since $x<1$, $(x-1)^2\le 1-x$. Also,

$$|1-y| = \left|1-\exp\left(\alpha_n[\beta_{n-1}-\ell(\eta_{n-1}^{i\to j},z_{n-1})]\right)\right|$$
$$\le \exp\left(\frac{1}{\sqrt{n}}\right)-1$$

since $-1\le\beta_{n-1}-\ell(\eta_{n-1}^{i\to j},z_{n-1})\le 1$. Using these results and (C6) for (C5) yields

$$|\phi_{(i,j),n-1}-\phi_{(i,j),n}|\le\frac{5}{e}\left(1-\sqrt{\frac{n-1}{n}}\right)+\left(\exp\left(\frac{1}{\sqrt{n}}\right)-1\right)+\frac{\sqrt{2}}{e}\left(1-\sqrt{\frac{n-1}{n}}\right)\left(\exp\left(\frac{1}{\sqrt{n}}\right)-1\right).$$

Finally, it can be verified that

$$1-\sqrt{\frac{n-1}{n}}\le\exp\left(\frac{1}{\sqrt{n}}\right)-1,\quad\forall n,$$

and, trivially,

$$\exp\left(\frac{1}{\sqrt{n}}\right)-1<1,\quad\forall n\ge 3.$$

This results in the following upper bound:

$$|\phi_{(i,j),n-1}-\phi_{(i,j),n}|\le\left(\frac{5+\sqrt{2}}{e}+1\right)\left(\exp\left(\frac{1}{\sqrt{n}}\right)-1\right),\quad\forall n\ge 3. \qquad \text{(C7)}$$

To conclude this proof, we use again Taylor series of the first order to approximate the function $\exp(x)-1$ around $x=0$. Namely,

$$\exp(x)-1=0+x+\tfrac{1}{2}\exp(\tilde{x})(x-\tilde{x})^2,$$

where $\tilde{x}$ is between x and 0. Since $x=1/\sqrt{n}\le 1$, this yields the upper bound

$$\exp\left(\frac{1}{\sqrt{n}}\right)-1\le\frac{1}{\sqrt{n}}+\frac{e}{2}\frac{1}{n}\le\frac{1+e/2}{\sqrt{n}}.$$

By plugging this inequality in (C7), we obtain

$$|\phi_{(i,j),n-1}-\phi_{(i,j),n}|\le\left(\frac{5+\sqrt{2}}{e}+1\right)\left(1+\frac{e}{2}\right)\frac{1}{\sqrt{n}}\le\frac{8}{\sqrt{n}},\quad\forall n\ge 3.$$

Combining this inequality with (C4) and (C3) completes the proof. □

Proof of Theorem 6. Combining the results of Proposition 1 and Lemma 12, we have that

$$\|\eta_n-\eta_{n-1}\|_{\mathrm{TV}}\le N\|\eta_n-\eta_{n-1}\|_\infty\le 16e\kappa N^2(N^2+1)\frac{1}{\sqrt{n}}$$

for $n\ge 3$. □

References

[1] Abernethy J, Bartlett PL, Hazan E (2011) Blackwell approachability and low-regret learning are equivalent. Kakade SM, von Luxburg U, eds., *Proc. 24th Annual Conf. Learn. Theory* (*COLT'11*), 27–46.
[2] Aubin J-P, Frankowska H (1990) *Set-Valued Analysis* (Birkhauser, Berlin).
[3] Bernstein A, Shimkin N (2013) Approachability without projection and generalized no-regret algorithms. Manuscript, submitted.
[4] Bernstein A, Mannor S, Shimkin N (2010) Online classification with specificity constraints. Lafferty J, Williams CKI, Shawe-Taylor J, Zemel RS, Culotta A, eds., *Proc. Adv. Neural Inform. Processing Systems*, Vol. 23 (Neural Information Processing Foundation, Vancouver), 190–198.
[5] Bernstein A, Mannor S, Shimkin N (2012) Opportunistic approachability: Algorithms based on online convex optimization. Manuscript.
[6] Blackwell D (1954) Controlled random walks. Gerretsen JCH, deGroot J, eds., *Proc. Internat. Congress of Mathematicians*, Vol. III (E.P. Noordhoff, Amsterdam), 335–338.
[7] Blackwell D (1956) An analog of the minimax theorem for vector payoffs. *Pacific J. Math.* 6:1–8.
[8] Bubeck S, Slivkins A (2012) The best of both worlds: Stochastic and adversarial bandits. Mannor S, Srebro N, Williamson RC, eds., *Proc. 25th Annual Conf. Learn. Theory* (*COLT'12*).
[9] Cesa-Bianchi N, Lugosi G (2006) *Prediction, Learning, and Games* (Cambridge University Press, New York).
[10] Dawid AP (1985) The impossibility of inductive inference. *J. Amer. Statist. Assoc.* 80:340–341.
[11] Even-Dar E, Kleinberg R, Mannor S, Mansour Y (2009) Online learning with global cost functions. Dasgupta S, Klivans A, eds., *Proc. 22nd Annual Conf. Learn. Theory* (*COLT'09*).
[12] Foster DP, Vohra RV (1997) Calibrated learning and correlated equilibrium. *Games Econom. Behav.* 21:40–55.
[13] Foster DP, Rakhlin A, Sridharan K, Tewari A (2011) Complexity-based approach to calibration with checking rules. Kakade SM, von Luxburg U, eds., *Proc. 24th Annual Conf. Learn. Theory* (*COLT'11*), 293–314.
[14] Fudenberg D, Levine DK (1998) *The Theory of Learning in Games* (MIT Press, Cambridge, MA).
[15] Hannan J (1957) Approximation to Bayes risk in repeated play. *Contributions to the Theory of Games* 3:97–139.
[16] Hart S, Mas-Colell A (2001) A general class of adaptive strategies. *J. Econom. Theory* 98:26–54.
[17] Hazan E, Kakade S (2012) (Weak) Calibration is computationally hard. Mannor S, Srebro N, Williamson RC, eds., *Proc. 25th Annual Conf. Learn. Theory* (*COLT'12*).
[18] Hou T-F (1971) Approachability in a two-person game. *Ann. Math. Statist.* 42(2):735–744.
[19] Kalai E, Lehrer E, Smorodinsky R (1999) Calibrated forecasting and merging. *Games Econom. Behav.* 29(1–2):151–169.
[20] Lehrer E (2002) Approachability in infinite dimensional spaces. *Internat. J. Game Theory* 31:253–268.
[21] Lehrer E, Solan E (2007) Learning to play partially-specified equilibrium. Manuscript.
[22] Lehrer E, Solan E (2009) Approachability with bounded memory. *Games Econom. Behav.* 66(2):995–1004.
[23] Lehrer E, Solan E (2013) A general internal regret-free strategy. Manuscript.
[24] Lehrer E, Solan E, Bauso D (2011) Repeated games over networks with vector payoffs: The notion of attainability. Cominetti R, Sorin S, Tuffin B, eds., *Proc. 5th Internat. Conf. Network Games, Control Optim.* (*NetGCooP*), 1–5.
[25] Mannor S, Shimkin N (2003) The empirical Bayes envelope and regret minimization in competitive Markov decision processes. *Math. Oper. Res.* 28(2):327–345.
[26] Mannor S, Shimkin N (2008) Regret minimization in repeated matrix games with variable stage duration. *Games Econom. Behav.* 63(1):227–258.
[27] Mannor S, Stoltz G (2010) A geometric proof of calibration. *Math. Oper. Res.* 35(4):721–727.
[28] Mannor S, Perchet V, Stoltz G (2011) Robust approachability and regret minimization in games with partial monitoring. Kakade SM, von Luxburg U, eds., *Proc. 24th Annual Conf. Learn. Theory* (*COLT'11*).
[29] Mannor S, Shamma JS, Arslan G (2007) Online calibrated forecasts: Memory efficiency versus universality for learning in games. *Machine Learn.* 67:77–115.
[30] Mannor S, Tsitsiklis JN, Yu JY (2009) Online learning with sample path constraints. *J. Machine Learn. Res.* 10:569–590.
[31] Meyer CD (1994) Sensitivity of the stationary distribution of a Markov chain. *SIAM J. Matrix Anal. Appl.* 15:715–728.
[32] Michael E (1956) Continuous selections. I. *Ann. Math.* 63(2):361–382.
[33] Milman E (2006) Approachable sets of vector payoffs in stochastic games. *Games Econom. Behav.* 56(1):135–147.
[34] Perchet V (2009) Calibration and internal no-regret with partial monitoring. Gavaldà R, Lugosi G, Zeugmann T, Zilles S, eds., *Proc. 20th Internat. Conf. Algorithmic Learn. Theory* (*ALT'09*), 68–82.
[35] Perchet V, Quincampoix M (2013) On an unified framework for approachability in games with or without signals. CoRR abs/1301.3609.
[36] Rakhlin A, Sridharan K, Tewari A (2011) Online learning: Beyond regret. *Proc. 24th Annual Conf. Learn. Theory* (*COLT'11*).
[37] Shimkin N, Shwartz A (1993) Guaranteed performance regions in Markovian systems with competing decision makers. *IEEE Trans. Automatic Control* 38(1):84–95.
[38] Shiryaev AN (1995) *Probability* (Springer, Berlin).
[39] Spinat X (2002) A necessary and sufficient condition for approachability. *Math. Oper. Res.* 27(1):31–44.
[40] Whitt W (1985) Uniform conditional variability ordering of probability distributions. *J. Appl. Probab.* 22(3):619–633.
[41] Young HP (2004) *Strategic Learning and Its Limits* (Oxford University Press, New York).