Online Classification with Specificity Constraints Supplementary Material

October 26, 2010

1 Properties of the Relaxed Best-Response Envelope

Recall that the scalar-relaxed best-response envelope (SR-BE) is given by

$$r_{\alpha}^{\mathrm{SR}}(q) \triangleq q(1) \max\left\{0, \beta^{*}(q) - \alpha\right\}, \tag{1}$$

where

$$\beta^{*}(q) \triangleq \max_{a \in \{1,\ldots,m\}} \left\{\beta_{tp}(q;a)\right\}, \tag{2}$$

is the optimal (unconstrained) tp-rate in hindsight under distribution q. Also, the CBE is

$$r_{\gamma}^{*}(q) = q(1)\beta_{\gamma}^{*}(q), \tag{3}$$

where

$$\beta_{\gamma}^{*}(q) \triangleq \max_{p \in \Delta(\mathcal{A})} \left\{ \sum_{a \in \mathcal{A}} p(a)\beta_{tp}(q;a) : \quad \text{so that} \quad \sum_{a \in \mathcal{A}} p(a)\beta_{fp}(q;a) \leq \gamma \right\}, \tag{4}$$

is the optimal constrained tp-rate in hindsight under distribution q.

We prove the following lemmas, which state the properties of the SR-BE. The first one is trivial.

Lemma 1.1.

1. *For every $\alpha \geq 0$, r_{α}^{SR} is a convex continuous function.*
2. *For every $\alpha \geq 0$, we have that $0 \leq r_{\alpha}^{\mathrm{SR}}(q) \leq \mathrm{conv}\left(r_{\gamma}^{*}\right)(q)$ for all $q \in \Delta(\mathcal{Z})$.*

Proof. Part 1 of the Lemma follows since r_{α}^{SR} is a maximum of linear functions (which is convex and continuous). The lower bound of Part 2 is trivial by definition of r_{α}^{SR}. The upper bound follows by the fact that r_{α}^{SR} is a convex function which is nowhere larger than r_{γ}^{*}. □

In particular, this Lemma states that the SR-BE is always dominated from above by the convex hull of the CBE. Recall, however, that the motivation for using the former is computational.

The next lemma shows that the SR-BE is strictly above the value of the constrained game at some point, unless the game is in some sense trivial. Hence, it do provide a performance improvement over the constrained min-max solution. Recall that the value of the constrained game in our case is

$$\begin{aligned} v_\Gamma &\triangleq \min_{q\in\Delta(\mathcal{Z})} r^*_\gamma(q) \\ &= \min_{q\in\Delta(\mathcal{Z})} \left\{ q(1) \max_{p\in\Delta(\mathcal{A})} \left\{ \sum_{a\in\mathcal{A}} p(a)\beta_{tp}(q;a) : \quad \text{so that} \quad \sum_{a\in\mathcal{A}} p(a)\beta_{fp}(q;a) \le \gamma \right\} \right\} = 0. \end{aligned}$$

Lemma 1.2. *Suppose that*

$$\alpha^* = \max_{q\in\Delta(\mathcal{Z})} \left(\beta^*(q) - \beta^*_\gamma(q)\right) < 1. \tag{5}$$

Then, there exists $q \in \Delta(\mathcal{Z})$ such that $r^{\mathrm{SR}}_{\alpha^*}(q) > v_\gamma = 0$.

Proof. Note that under condition (5), we have that there exists $q \in \Delta(\mathcal{Z})$ with $q(1) > 0$ such that

$$\beta^*(q) = 1 > \max_{q\in\Delta(\mathcal{Z})} \left(\beta^*(q) - \beta^*_\gamma(q)\right) = \alpha^*,$$

implying that for this q

$$r^{\mathrm{SR}}_{\alpha^*}(q) = q(1)(\beta^*(q) - \alpha^*) > 0.$$

□

We note that the condition of Lemma 1.2 is satisfied, unless the constrained optimal tp-rate $\beta^*_\gamma(q)$ is zero at some q for which the unconstrained optimal tp-rate is 1.

2 Proof of Theorem 5.1

Fix $\alpha \ge \alpha^*$. We use approachability theory [2] (presented in detail in Section 5) to prove this theorem. In particular, we consider an extension of Hart and Mas-Colell's vector-valued game formulation [3] (see Section 5.3), with the following vector-valued payoff of the agent at time n:

$$m_n \triangleq \left(\left\{R^\alpha_n(a)\right\}_{a\in\mathcal{A}}, L_n\right) \in \mathbb{R}^{|\mathcal{A}|+1},$$

where

$$\begin{aligned} R^\alpha_k(a) &\triangleq [f_k(a) - f_k(a_k) - \alpha]\, \mathbb{I}\{b_k = 1\}, \quad a \in \mathcal{A}, \\ L_k &\triangleq c_\gamma(a_k, z_k), \end{aligned} \tag{6}$$

The average vector-valued payoff at time n is then

$$\overline{m}_n = \left(\left\{ \overline{R}_n^{\alpha}(a) \right\}_{a \in \mathcal{A}}, \overline{L}_n \right), \tag{7}$$

where

$$\overline{R}_n^{\alpha}(a) = \bar{q}_n(1) \left[\beta_{tp}(\bar{q}_n; a) - \bar{\beta}_{tp}(n) - \alpha \right], \quad a \in \mathcal{A}; \quad \overline{L}_n = \bar{q}_n(0)[\bar{\beta}_{fp}(n) - \gamma]. \tag{8}$$

Now, consider $S = \mathbb{R}_-^{|\mathcal{A}|+1} \triangleq \left\{ u \in \mathbb{R}^{|\mathcal{A}|+1} : \ u \leq 0 \right\}$, which is the non-positive orthant of $\mathbb{R}^{|\mathcal{A}|+1}$. The set S is convex, and we now show that it is approachable. For this, let

$$m(p,q) = \left(\left\{ q(1) \left[\beta_{tp}(q; a) - \alpha \right] - r(p,q) \right\}_{a \in \mathcal{A}}, c_{\gamma}(p,q) \right) \tag{9}$$

denote the expected vector-valued payoff under the pair of mixed actions (p,q). We need to show that for every $q \in \Delta(\mathcal{Z})$ there exists a $p \in \Delta(\mathcal{A})$ such that $m(p,q) \in S$, which is a sufficient condition for approachability of convex sets (see Theorem 5.1). Fix $q \in \Delta(\mathcal{Z})$. First note that, by our assumption that the constraints can always be satisfied, there exists a $p^* \in \Delta(\mathcal{A})$ such that $c_{\gamma}(p^*, q) \leq 0$ and

$$\begin{aligned} r(p^*, q) &= r_{\gamma}^*(q) \\ &= q(1)\beta_{\gamma}^*(q) \\ &\geq q(1)(\beta^*(q) - \max_{q \in \Delta(\mathcal{Z})} [\beta^*(q) - \beta_{\gamma}^*(q)]) \\ &\geq q(1)(\beta^*(q) - \alpha) \\ &\geq q(1)(\beta_{tp}(q; a) - \alpha), \quad \forall a \in \mathcal{A}, \end{aligned}$$

where the second inequality follows by the assumption that

$$\alpha \geq \alpha^* \triangleq \max_{q \in \Delta(\mathcal{Z})} \left(\beta^*(q) - \beta_{\gamma}^*(q) \right)$$

and the third inequality holds by (2). Thus $m(p^*, q) \in S$, and S is approachable. Also note that the corresponding approachability algorithm attains r_{α}^{SR}, since the approachability of $\overline{m}_n$ to S implies that (almost surely, for every strategy of the opponent)

$$\limsup_{n \to 0} \overline{R}_n^{\alpha}(a) \leq 0, \quad \forall a \in \mathcal{A}$$

and

$$\limsup_{n \to 0} \overline{L}_n \leq 0,$$

which in turn implies by (8) that

$$\limsup_{n \to 0} \left\{ \bar{q}_n(1) \left[\beta_{tp}(\bar{q}_n; a) - \bar{\beta}_{tp}(n) - \alpha \right] \right\} \leq 0, \quad \forall a \in \mathcal{A}$$

$$\limsup_{n \to 0} \left\{ \bar{q}_n(0)[\bar{\beta}_{fp}(n) - \gamma] \right\} \leq 0.$$

It remains to develop a more explicit version of this approachability algorithm and show that it coincides with the CRM algorithm. Since S is approachable, Blackwell's approachability theorem ensures that for every $u \in \mathbb{R}^{|\mathcal{A}|+1}$ there exists a $p \in \Delta(\mathcal{A})$ such that

$$u \cdot m(p,z) \leq \sup\{u \cdot s : \ s \in S\}, \ \forall z \in \mathcal{Z};$$

see [4] and Corollary 5.1 in Appendix 5. Moreover, an approaching strategy is to choose at stage n a mixed action p which corresponds to $u = \overline{m}_{n-1} - P_S(\overline{m}_{n-1})$, where $P_S(v)$ is the projection of v to S in the Euclidean distance. In our case, S is the non positive orthant and therefore $P_S(\overline{m}_{n-1}) = [\overline{m}_{n-1}]_-$, implying that $u = [\overline{m}_{n-1}]_+ \in \mathbb{R}_+^{|\mathcal{A}|+1}$. As a result, $\sup\{u \cdot s : \ s \in S\} = 0$. To summarize, an approaching strategy is to choose at time n a mixed action p which satisfies:

$$[\overline{m}_{n-1}]_+ \cdot m(p,z) \leq 0, \ \forall z \in \mathcal{Z}.$$

By using (7) and (9), this last condition is exactly the defining inequality for p in the CRM algorithm:

$$\begin{cases} \sum_{a\in\mathcal{A}} \left[\overline{R}^{\alpha}_{n-1}(a)\right]_+ \left(f(a) - \sum_{a'\in\mathcal{A}} p(a')f(a') - \alpha\right) \leq 0, & \forall z = (f,1) \in \mathcal{Z}, \\ \left[\overline{L}_{n-1}\right]_+ \left(\sum_{a'\in\mathcal{A}} p(a')f(a') - \gamma\right) \leq 0, & \forall z = (f,-1) \in \mathcal{Z}, \end{cases} \tag{10}$$

Finally we note that, whenever $\sum_{a\in\mathcal{A}} \left[\overline{R}^{\alpha}_{n-1}(a)\right]_+ > 0$, this p can be found by solving

$$\begin{aligned} &\min_{p\in B_n} \max_{f\in[0,1]^m} \sum_{a\in\mathcal{A}} f(a) \left(\left[\overline{R}^{\alpha}_{n-1}(a)\right]_+ - p(a)\right) / \sum_{a'\in\mathcal{A}} \left[\overline{R}^{\alpha}_{n-1}(a')\right]_+ \\ &= \min_{p\in B_n} \max_{f\in[0,1]^m} \sum_{a\in\mathcal{A}} f(a) \left(p^{\alpha}_n(a) - p(a)\right) \\ &= \min_{p\in B_n} \sum_{a\in\mathcal{A}: p^{\alpha}_n(a) > p(a)} \left(p^{\alpha}_n(a) - p(a)\right), \end{aligned}$$

where

$$B_n \triangleq \left\{ p \in \Delta(\mathcal{A}) : \left[\overline{L}_{n-1}\right]_+ \left(\sum_{a'\in\mathcal{A}} p(a')f(a') - \gamma\right) \leq 0, \ \ \forall z = (f,-1) \in \mathcal{Z} \right\}$$

and

$$p^{\alpha}_n(a) = \frac{\left[\overline{R}^{\alpha}_{n-1}(a)\right]_+}{\sum_{a'\in\mathcal{A}} \left[\overline{R}^{\alpha}_{n-1}(a')\right]_+}$$

is the α-regret matching strategy.

3 Adaptive Relaxation

Given a feasible $\alpha \geq \alpha^*$, the CRM algorithm attains the SR-BE r_α^{SR}. However, in practice, it may be possible to attain r_α^{SR} with $\alpha < \alpha^*$ if the opponent is not entirely adversarial. In order to capitalize on this possibility, we propose to use an adaptive algorithm that adjusts the value of α online. The idea is to start from some small initial value $\alpha_0 \geq 0$ (possibly $\alpha_0 = 0$). At each time step n, we would like to use a parameter $\alpha = \alpha_n$ for which inequality (10) can be satisfied. We remind that this inequality is always satisfied when $\alpha \geq \alpha^*$. If however $\alpha < \alpha^*$, the inequality may or may not be satisfied. In the latter case, we increase α so that the condition is satisfied. We propose two possibilities for the approximate adaptive search for parameter α, using an approximation parameter $\epsilon > 0$. Let α_n be the relaxation parameter that is used at stage n. (i) *Discretization.* Each time inequality (10) can not be satisfied, find the minimal integer $K \geq 1$ such that for $\alpha_n = \alpha_{n-1} + K\epsilon$ there exists a mixed action $p \in \Delta(\mathcal{A})$ required by (10). (ii) *Binary search.* Set α_b large enough, so that $\alpha_b > \alpha^*$, and use a binary search on the interval $[\alpha_{n-1}, \alpha_b]$ to obtain α_n such that (10) can be satisfied for it, while it can not satisfied for $\alpha_n - \epsilon$.

The next theorem ensures that the adaptive CRM algorithm attains r_α^{SR} with $\alpha \leq \alpha^* + \epsilon$, where $\epsilon > 0$ is the discretization parameter. Moreover, the theorem provides the convergence rate of this algorithm.

Theorem 3.1. *Suppose that the adaptive CRM algorithm uses an $\epsilon > 0$ to discretize the values of α. Let*[1] *$\alpha_\infty \triangleq \limsup_{n\to\infty} \alpha_n \leq \alpha^* + \epsilon$ denote the posterior relaxation parameter that the algorithm uses in the long term. Then, the algorithm attains $r_{\alpha_\infty}^{\text{SR}}$. In particular, for every $\delta > 0$ and $\eta > 0$, there exists $T > 0$ such that*

$$\mathbb{P}\left\{\bar{r}_n \geq r_{\alpha_\infty}^{\text{SR}}(\bar{q}_n) - \delta \quad \text{and} \quad d(\bar{c}_n, \Gamma) \leq \delta\right\} \geq 1 - \eta, \quad \forall n \geq T,$$

for any strategy of the opponent, where

$$T = \frac{8C^2}{\delta^4} \ln \frac{1}{\eta} + \frac{C}{\delta^2}$$

and $C = 16(|\mathcal{A}| + 1)$.

Note that T does not depend explicitly on the discretization parameter ϵ. However, this parameter controls the precision of the approximate adaptive search. Therefore, the smaller is ϵ, the better is the precision, and the longer it takes to perform the search. Finally, note that the adaptive CRM algorithm does not require the computation of the optimal α^*, as it discovers it online.

Proof. We follow the proof of Theorem 5.1 and modify the convergence proof of the approachability algorithm for our adaptive case. Let

$$m_k^{\alpha_n} \triangleq \left(\left\{R_k^{\alpha_n}(a)\right\}_{a\in\mathcal{A}}, L_k\right) \in \mathbb{R}^{|\mathcal{A}|+1}.$$

[1] We note that α_∞ is a random variable.

denote the vector payoff at time $k \leq n$, using a relaxation parameter α_n. The corresponding average vector payoff at time n is then

$$\overline{m}_n^{\alpha_n} = \left(\left\{ \overline{R}_n^{\alpha_n}(a) \right\}_{a \in \mathcal{A}}, \overline{L}_n \right).$$

As in the proof of Theorem 5.1, S denotes the non-positive orthant of $\mathbb{R}^{|\mathcal{A}|+1}$. Our goal is to show that

$$d\left(\overline{m}_n^{\alpha_\infty}, S\right) \to 0$$

as $n \to \infty$, almost surely. We proceed in the following stages.

Stage 1. Note that we surely have that

$$d\left(\overline{m}_n^{\alpha_n}, S\right) \geq d\left(\overline{m}_n^{\alpha_{n+1}}, S\right) \geq d\left(\overline{m}_n^{\alpha_\infty}, S\right), \ \forall n \geq 1, \tag{11}$$

since we only increase the relaxation parameter in the course of the algorithm. Thus, it is sufficient to show that

$$d\left(\overline{m}_n^{\alpha_n}, S\right) \to 0$$

almost surely. We prove this in the next stage.

Stage 2. Indeed, let

$$X_n \triangleq \left[d\left(\overline{m}_n^{\alpha_n}, S\right)\right]^2 = \left\| \overline{m}_n^{\alpha_n} - P_S(\overline{m}_n^{\alpha_n}) \right\|^2.$$

We show below that X_n almost surely converges to zero. Using the fact that

$$\overline{m}_n^{\alpha_n} = \frac{n-1}{n} \overline{m}_{n-1}^{\alpha_n} + \frac{1}{n} m_n^{\alpha_n},$$

we obtain

$$\begin{aligned}
X_n &= \left\| \overline{m}_n^{\alpha_n} - P_S(\overline{m}_n^{\alpha_n}) \right\|^2 \\
&\leq \left\| \overline{m}_n^{\alpha_n} - P_S(\overline{m}_{n-1}^{\alpha_n}) \right\|^2 \\
&= \left\| \frac{n-1}{n} \overline{m}_{n-1}^{\alpha_n} + \frac{1}{n} m_n^{\alpha_n} - P_S(\overline{m}_{n-1}^{\alpha_n}) \right\|^2 \\
&= \left\| \frac{n-1}{n} \left(\overline{m}_{n-1}^{\alpha_n} - P_S(\overline{m}_{n-1}^{\alpha_n}) \right) + \frac{1}{n} \left(m_n^{\alpha_n} - P_S(\overline{m}_{n-1}^{\alpha_n}) \right) \right\|^2 \\
&= \left(\frac{n-1}{n} \right)^2 \left\| \overline{m}_{n-1}^{\alpha_n} - P_S(\overline{m}_{n-1}^{\alpha_n}) \right\|^2 + \frac{1}{n^2} \left\| m_n^{\alpha_n} - P_S(\overline{m}_{n-1}^{\alpha_n}) \right\|^2 \\
&\quad + 2 \frac{n-1}{n^2} \left(\overline{m}_{n-1}^{\alpha_n} - P_S(\overline{m}_{n-1}^{\alpha_n}) \right) \cdot \left(m_n^{\alpha_n} - P_S(\overline{m}_{n-1}^{\alpha_n}) \right) \\
&\leq \left(\frac{n-1}{n} \right)^2 \left\| \overline{m}_{n-1}^{\alpha_{n-1}} - P_S(\overline{m}_{n-1}^{\alpha_{n-1}}) \right\|^2 + \frac{1}{n^2} \left\| m_n^{\alpha_n} - P_S(\overline{m}_{n-1}^{\alpha_n}) \right\|^2 \\
&\quad + 2 \frac{n-1}{n^2} \left(\overline{m}_{n-1}^{\alpha_n} - P_S(\overline{m}_{n-1}^{\alpha_n}) \right) \cdot \left(m_n^{\alpha_n} - P_S(\overline{m}_{n-1}^{\alpha_n}) \right),
\end{aligned}$$

where the last inequality follows by (11). Multiplying this inequality by n^2 and rearranging yields

$$\begin{aligned} n^2 X_n - (n-1)^2 X_{n-1} &\leq \left\| m_n^{\alpha_n} - P_S(\overline{m}_{n-1}^{\alpha_n}) \right\|^2 \\ &\quad +2(n-1)\left(\overline{m}_{n-1}^{\alpha_n} - P_S(\overline{m}_{n-1}^{\alpha_n})\right) \cdot \left(m_n^{\alpha_n} - P_S(\overline{m}_{n-1}^{\alpha_n})\right) \\ &\leq C + 2(n-1)\left(\overline{m}_{n-1}^{\alpha_n} - P_S(\overline{m}_{n-1}^{\alpha_n})\right) \cdot \left(m_n^{\alpha_n} - P_S(\overline{m}_{n-1}^{\alpha_n})\right), \quad (12) \end{aligned}$$

where the last inequality follows since the vector payoff is bounded, implying that

$$\left\| m_n^{\alpha_n} - P_S(\overline{m}_{n-1}^{\alpha_n}) \right\|^2 \leq C, \ \forall n \geq 1,$$

for some constant $C > 0$ (independent of n).

An explicit expression for C can be easily obtained in terms of the corresponding bounds on the reward and cost functions. In our case, we have that $r_{max} = 1$ and $c_{max} = 1$. Therefore, we have that

$$\begin{aligned} \left\| m_n^{\alpha_n} - P_S(\overline{m}_{n-1}^{\alpha_n}) \right\|^2 &\leq (|\mathcal{A}|+1) \max\left\{4\left(r_{max}+\alpha^*\right)^2, 4\left(c_{max}+\gamma\right)^2\right\} \\ &\leq 16(|\mathcal{A}|+1) \triangleq C, \quad (13) \end{aligned}$$

where the second inequality follows since we have that $\alpha^* \leq r_{max} = 1$ and $\gamma \leq c_{max} = 1$.

Summing both sides of inequality (12) for $k = 1, ..., n$, the left-hand side telescopes to $n^2 X_n$. Therefore, we have that

$$X_n \leq \frac{C}{n} + \frac{2}{n} \sum_{k=1}^{n} \frac{k-1}{n} \left(\overline{m}_{k-1}^{\alpha_k} - P_S(\overline{m}_{k-1}^{\alpha_k})\right) \cdot \left(m_k^{\alpha_k} - P_S(\overline{m}_{k-1}^{\alpha_k})\right)$$

Now, recall that the mixed action p_k used by the algorithm at time k satisfies the separation condition

$$\left(\overline{m}_{k-1}^{\alpha_k} - P_S(\overline{m}_{k-1}^{\alpha_k})\right) \cdot \left(m^{\alpha_k}(p_k, z) - P_S(\overline{m}_{k-1}^{\alpha_k})\right) \leq 0, \ \forall z \in \mathcal{Z},$$

where

$$m^{\alpha_k}(p, z) \triangleq \left(\left\{r(a, z) - r(p, z) - \alpha_k\right\}_{a \in \mathcal{A}}, \left\{c_i(p, q) - \gamma_i\right\}_{i=0}^{\ell}\right).$$

Hence, it follows that

$$\begin{aligned} X_n &\leq \frac{C}{n} + \frac{2}{n} \sum_{k=1}^{n} \left(\overline{m}_{k-1}^{\alpha_k} - P_S(\overline{m}_{k-1}^{\alpha_k})\right) \cdot \left(m_k^{\alpha_k} - m^{\alpha_k}(p_k, z_k)\right) \\ &\triangleq \frac{C}{n} + \frac{2}{n} \sum_{k=1}^{n} Y_k. \quad (14) \end{aligned}$$

To complete the proof, we show that $\{Y_n\}$ is a bounded martingale difference sequence, implying that its average almost surely converges to zero. Indeed, by the payoff boundedness assumption, we have that

$$|Y_n| \leq C, \ \forall n \geq 1.$$

Also, let $\mathcal{F}_n \triangleq \sigma(a_1, z_1, ..., a_n, z_n)$ denote the corresponding filtration. We then have that

$$\begin{aligned}\mathbb{E}[Y_n \mid \mathcal{F}_{n-1}] &= \mathbb{E}\left[\left(\overline{m}_{n-1}^{\alpha_n} - P_S(\overline{m}_{n-1}^{\alpha_n})\right) \cdot (m_n^{\alpha_n} - m^{\alpha_n}(p_n, z_n)) \;\middle|\; \mathcal{F}_{n-1}\right] \\ &= \left(\overline{m}_{n-1}^{\alpha_n} - P_S(\overline{m}_{n-1}^{\alpha_n})\right) \cdot \mathbb{E}\left(m^{\alpha_n}(a_n, z_n) - m^{\alpha_n}(p_n, z_n) \mid \mathcal{F}_{n-1}\right) \\ &= 0.\end{aligned}$$

Thus, $\{Y_n, \mathcal{F}_n\}$ is a bounded martingale difference sequence, and the claim follows.

Stage 3. Finally, we use the Hoeffding-Azuma inequality to obtain polynomial convergence rate. In particular, for any $\delta > 0$, it holds that

$$\begin{aligned}\mathbb{P}\left\{d\left(\overline{m}_n^{\alpha\infty}, S\right) \leq \delta\right\} \geq \mathbb{P}\left\{d\left(\overline{m}_n^{\alpha_n}, S\right) \leq \delta\right\} &= \mathbb{P}\left\{X_n \leq \delta^2\right\} \\ &\geq \mathbb{P}\left\{\frac{C}{n} + \frac{2}{n}\sum_{k=1}^{n} Y_k \leq \delta^2\right\} \\ &= \mathbb{P}\left\{\sum_{k=1}^{n} Y_k \leq \left(\delta^2 - \frac{C}{n}\right)\frac{n}{2}\right\} \\ &\geq 1 - \exp\left(-\frac{\left[\left(\delta^2 - \frac{C}{n}\right)\frac{n}{2}\right]^2}{2nC^2}\right) \\ &\geq 1 - \exp\left(-\frac{\delta^4 n}{8C^2} + \frac{\delta^2}{8C}\right),\end{aligned}$$

where the first inequality follows by (11), the second inequality follows by (14), and the third inequality holds by the Hoeffding-Azuma inequality. For a given $\eta > 0$, we thus require

$$\exp\left(-\frac{\delta^4 n}{8C^2} + \frac{\delta^2}{8C}\right) \leq \eta,$$

which yields

$$n \geq \frac{8C^2}{\delta^4}\ln\frac{1}{\eta} + \frac{C}{\delta^2} \triangleq T$$

as required. This completes the proof of the Theorem. □

4 Computational Aspects

In this section we discuss the computational issues related to the program

$$\alpha^* \triangleq \max_{q \in \Delta(\mathcal{Z})} \left(\beta^*(q) - \beta_\gamma^*(q)\right). \tag{15}$$

where β^* and β_γ^* are given in (2) and (4), respectively.

We propose below possible methods for the approximate (offline) computation of α^*. First note that we can write (15) as:

$$\alpha^* = \max_{a \in \mathcal{A}} \max_{q \in \Delta(\mathcal{Z})} \min_{p \in \Delta(\mathcal{A}):\ \beta_{fp}(q;p) \leq \gamma} \left(\beta_{tp}(q;a) - \beta_{tp}(q;p)\right),$$

where $\beta_{tp}(q;p) \triangleq \sum_{a\in\mathcal{A}} p(a)\beta_{tp}(q;a)$ and $\beta_{fp}(q;p) \triangleq \sum_{a\in\mathcal{A}} p(a)\beta_{fp}(q;a)$. Now, the inner minimization problem (which is in fact a linear program) can be expressed using Lagrange multipliers $\lambda \in \mathbb{R}^{\ell}_{+}$ as follows:

$$\begin{aligned}\min_{p\in\Delta(\mathcal{A}):\ \beta_{fp}(q;p)\le\gamma} \{\beta_{tp}(q;a)-\beta_{tp}(q;p)\} &= \min_{p\in\Delta(\mathcal{A})} \sup_{\lambda\ge 0} \{\beta_{tp}(q;a)-\beta_{tp}(q;p)+\lambda\left(\beta_{fp}(q;p)-\gamma\right)\} \\ &= \sup_{\lambda\ge 0} \min_{p\in\Delta(\mathcal{A})} \{\beta_{tp}(q;a)-\beta_{tp}(q;p)+\lambda\left(\beta_{fp}(q;p)-\gamma\right)\} \\ &= \sup_{\lambda\ge 0} \min_{a'\in\mathcal{A}} \{\beta_{tp}(q;a)-\beta_{tp}(q;a')+\lambda\left(\beta_{fp}(q;a')-\gamma\right)\},\end{aligned}$$

where the second equality holds due to the strong duality of the linear program, and the third equality follows since the argument of the minimization is a linear function of p. Thus, (15) takes the following form:

$$\alpha^* = \max_{a\in\mathcal{A}} \sup_{\lambda\ge 0} \max_{q\in\Delta(\mathcal{Z})} \min_{a'\in\mathcal{A}} \{\beta_{tp}(q;a)-\beta_{tp}(q;a')+\lambda\left(\beta_{fp}(q;a')-\gamma\right)\}. \tag{16}$$

We note that the resulting maximization problem over λ (or over q, if we exchange the max with sup) is not convex in general and therefore may lack efficient algorithms for its solution. Below we propose two solution approaches.

Observe that the inner term

$$\max_{q\in\Delta(\mathcal{Z})} \min_{a'\in\mathcal{A}} \{\beta_{tp}(q;a)-\beta_{tp}(q;a')+\lambda\left(\beta_{fp}(q;a')-\gamma\right)\}$$

can be easily transformed into a standard linear programming problem as is usually done when computing a value of a repeated game. We are thus left with the maximization problem on $\lambda \ge 0$, which can be solved numerically, using discretization. Another possibility is to solve directly (16) which can be easily transformed into the following *bilinear* program:

$$\begin{aligned} & \max t \\ \text{s.t.} \quad & \lambda \ge 0 \\ & q \in \Delta(\mathcal{Z}) \\ & \beta_{tp}(q;a)-\beta_{tp}(q;a')+\lambda\left(\beta_{fp}(q;a')-\gamma\right) \ge t,\ \forall a' \in A.\end{aligned}$$

This program should be solved for each $a \in \mathcal{A}$, and then the maximum should be taken. However, we are not aware of efficient algorithms for solving this program.

5 Approachability Theory

The approachability problem, introduced by Blackwell in [2], may be considered as a generalization of the basic online decision problem to *vector* rewards, taking values in $\mathbb{R}^I$. At each time step $n = 1, 2, ...$, the agent selects his action $a_n \in \mathcal{A}$, observes the action

$z_n \in \mathcal{Z}$ chosen by the opponent, and obtains a *vector* reward $m_n = m(a_n, z_n) \in \mathbb{R}^I$. We denote by

$$\bar{m}_n \triangleq \frac{1}{n} \sum_{k=1}^{n} m_n$$

the average reward obtained by the player up to time n, as before.

In the approachability problem, we consider a set $S \subseteq \mathbb{R}^I$, and ask if there exists a policy for the player that will bring the average reward $\bar{m}_n$ to S (asymptotically, almost surely) no matter what the opponent actions are. More formally, we have the following classical definition of approachable sets due to Blackwell [2].

Definition 5.1 (Approachable Set)**.** *Let $S \subseteq \mathbb{R}^I$ be a closed set. The set S is approachable by the player if there exist a policy π such that $d(\bar{m}_n, S) \to 0$ almost surely as $n \to \infty$, for any policy σ of the opponent. Here, $\bar{r}_n$ is the average reward obtained using π up to time n, and $d(\cdot, \cdot)$ is Euclidian distance.*

5.1 Approachability Theorem and Algorithms

Next, we present the original formulation of Blackwell's Theorem which provides us with necessary and sufficient conditions for approachability of *convex* sets. To this end, for any $p \in \Delta(\mathcal{A})$ denote by $C(p)$ the convex hull of the points $\{m(p, z)\}_{z \in \mathcal{Z}}$. Similarly, for any $q \in \Delta(\mathcal{Z})$ denote by $T(q)$ the convex hull of the points $\{m(a, q)\}_{a \in \mathcal{A}}$.

Theorem 5.1. *Let S be any closed convex set. Then, the following statements are equivalent:*

1. *S is approachable.*
2. *For every $q \in \Delta(\mathcal{Z})$ there exists $p \in \Delta(\mathcal{A})$ such that $m(p, q) \in S$.*
3. *For every $x \notin S$, there exists $p \in \Delta(\mathcal{A})$ such that*

$$(x - P_S(x)) \cdot (m(p, z) - P_S(x)) \leq 0, \ \forall z \in \mathcal{Z},$$

where $P_S(x)$ is the projection of x to S in the Euclidean distance.

Below we state a corollary to this theorem, which is a simple manipulation of condition (3) above.

Corollary 5.1. *A closed convex set S is approachable if and only if for every $u \in \mathbb{R}^I$ there exists $p \in \Delta(\mathcal{A})$ such that*

$$u \cdot m(p, z) \leq w(u) \triangleq \sup \{u \cdot y : y \in S\}, \ \forall z \in \mathcal{Z}.$$

In case of a general set S, Blackwell showed in [2] that in fact conditions 2 and 3 of Theorem 5.1 are, respectively, a necessary and a sufficient condition for approachability.

Theorem 5.2. *Let $S \subseteq \mathbb{R}^I$ be a given closed set.*

1. *If condition 3 of Theorem 5.1 holds, then S is approachable.*

2. *If S is approachable, then condition 2 of Theorem 5.1 holds.*

Recently, Spinat in [6] proved a necessary and sufficient condition for approachability of general sets.

Theorem 5.3. *A closed set S is approachable if and only if it contains a set which satisfies condition 3 of Theorem 5.1.*

In fact, this Theorem suggests that the study of approachability should focus on the sets which satisfy condition 3 of Theorem 5.1, the so-called *B-sets*. This is true since, by this Theorem, any approachability algorithm for B-sets is also the approachability algorithm for any approachable set.

Approachability algorithms (i.e. algorithms that actually approach a given approachable closed set S) require that condition 3 of Theorem 5.1 be satisfied for S. Thus, they are applicable to convex sets or, more generally, to B-sets. The first such algorithm is directly based on Blackwell's theorem, in particular on condition (3), and is presented as Algorithm 1. See Figure 1 for geometric interpretation of this algorithm.

Algorithm 1 Blackwell's Approachability Algorithm

1. If $\bar{m}_{n-1} \in S$, then choose arbitrary action a_n.

2. If $\bar{m}_{n-1} \notin S$, use a mixed action $p_n \in \Delta(\mathcal{A})$, such that

$$u \cdot m(p_n, z) \leq w(u), \ \forall z \in \mathcal{Z},$$

where $u = u(\bar{m}_{n-1}) \triangleq \bar{m}_{n-1} - P_S(\bar{m}_{n-1})$. Such a p_n exists by Corollary 5.1.

Note that Blackwell's algorithm uses a function $u : \mathbb{R}^I \setminus S \to \mathbb{R}^I$ to define the approachability direction, where $u(x)$ is the vector starting from the closest point $s \in S$ to x, and ending in x. Hart and Mas-Colell [4] proposed a more general class of approachability algorithms, by introducing a general class of directional mappings u, with the certain regularity properties.

5.2 Blackwell's Approachability Formulation of Regret Minimization

One of the first applications of approachability theory was proposed by Blackwell in [1]. In this work, the regret minimization problem was formulated as a special case of the approachability problem to a corresponding *convex* set. In particular, the following game with vector-valued rewards was defined in [1]. Recall that, at time n, the agent chooses

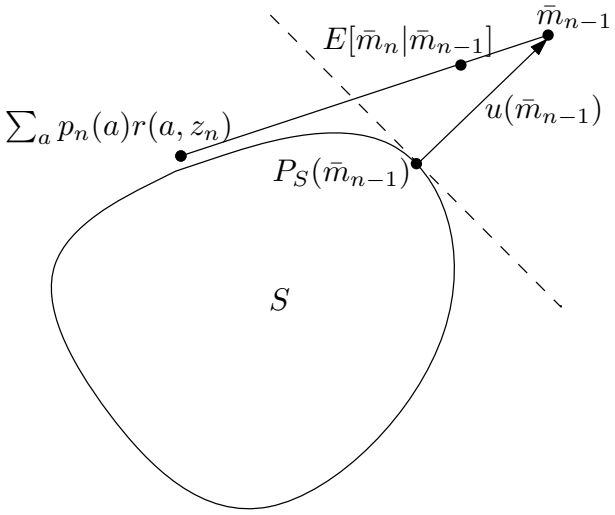

Figure 1: Geometric Interpretation of Approachability

a_n, the opponent chooses z_n, the agent obtains $r_n = r(a_n, z_n) \in \mathbb{R}$. Now, the vector-valued reward at time n, $m_n \in \mathbb{R} \times \Delta(\mathcal{Z})$, is defined as $m_n = (r_n, \mathbf{1}(z_n))$, where $\mathbf{1}(z)$ is the probability distribution concentrated on z. Note that the average reward for this game is

$$\bar{m}_n \triangleq \frac{1}{n}\sum_{k=1}^{n} m_k = (\bar{r}_n, \bar{q}_n)\,.$$

Finally, let

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

where r^* is the best-response envelope (BE). It is easy to see that S is convex approachable set, implying that Blackwell's algorithm can be used to approach it. Therefore, Blackwell's algorithm for this set is a no-regret algorithm for the original regret minimization problem. We note that this algorithm is implicit in the sense that it requires a computation of the projection P_S, and the complexity of this computation depends on the structure of the set S.

An extension of Blackwell's formulation to the constrained setting was proposed in [5]. In particular, instantaneous cost c_n was included in the vector-valued reward, that is

$$\bar{m}_n = (\bar{r}_n, \bar{c}_n, \bar{q}_n) \in \mathbb{R}^{1+\ell} \times \Delta(\mathcal{Z}).$$

Accordingly, the set S is defined as:

$$S = \left\{(r,c,q) \in \mathbb{R}^{1+\ell} \times \Delta(\mathcal{Z}) :\ r \geq r^*_\Gamma(q),\ c \in \Gamma\right\},$$

where r^*_Γ is the constrained best-response envelope CBE . In this case, S is *non-convex* since r^*_Γ is non-convex function. Moreover, it was shown in [5] that it is not approachable in general, implying that r^*_Γ need not be attainable. However, the *convex hull* $\mathrm{conv}\,(r^*_\Gamma)$ of r^*_Γ was shown to be attainable.

5.3 Regret Matching

An alternative vector-valued game formulation for the regret minimization problem was proposed by Hart and Mas-Colell in [3]. Let

$$R_k(a) \triangleq r(a, z_k) - r_k, \ a \in \mathcal{A},$$

denote the *instantaneous regret* at time k. The average regret at time n is then

$$\overline{R}_n(a) = r(a, \bar{q}_n) - \bar{r}_n,$$

Now consider the following vector-valued payoff of the agent at time n:

$$m_n \triangleq \{R_n(a)\}_{a \in \mathcal{A}} \in \mathbb{R}^{|\mathcal{A}|}.$$

The average vector-valued payoff at time n is then

$$\overline{m}_n = \left\{\overline{R}_n(a)\right\}_{a \in \mathcal{A}}.$$

Finally, define

$$S = \mathbb{R}_-^{|\mathcal{A}|} \triangleq \left\{u \in \mathbb{R}^{|\mathcal{A}|} : \ u \leq 0\right\},$$

which is the non positive orthant of $\mathbb{R}^{|\mathcal{A}|}$. S is a convex set, and it can be easily proved (using Theorem 5.1) that it is approachable by the agent. The advantage of this formulation is the fact that the corresponding no-regret algorithm is a simple *regret matching* strategy p_n, which is given by:

$$p_n(a) = \frac{\left[\overline{R}_{n-1}(a)\right]_+}{\sum_{a' \in \mathcal{A}} \left[\overline{R}_{n-1}(a')\right]_+}.$$

That is, p_n prescribes to play according to probabilities that are proportional to the (positive) regrets. In fact, a whole class of such no-regret algorithms was proposed in [4] based on a general class of approachability direction mappings.

References

[1] D. Blackwell. Controlled random walks. In *Proceedings of the International Congress of Mathematicians*, volume III, pages 335–338, 1954.

[2] D. Blackwell. An analog of the minimax theorem for vector payoffs. *Pacific Journal of Mathematics*, 6:1–8, 1956.

[3] S. Hart and A. Mas-Colell. A simple adaptive procedure leading to correlated equilibrium. *Econometrica*, 68:1127–1150, 2000.

[4] S. Hart and A. Mas-Colell. A general class of adaptive strategies. *Journal of Economic Theory*, 98:26–54, 2001.

[5] S. Mannor, J. N. Tsitsiklis, and J. Y. Yu. Online learning with sample path constraints. *Journal of Machine Learning Research*, 10:569–590, 2009.

[6] X. Spinat. A necessary and sufficient condition for approachability. *Mathematics of Operations Research*, 27(1):31–44, 2002.