Generalised Entropy and Asymptotic Complexities of Languages

Yuri Kalnishkan, Vladimir Vovk, and Michael V. Vyugin

Department of Computer Science, Royal Holloway, University of London, Egham, Surrey, TW20 0EX, UK {yura, vovk, misha}@cs.rhul.ac.uk

Abstract. In this paper the concept of asymptotic complexity of languages is introduced. This concept formalises the notion of learnability in a particular environment and generalises Lutz and Fortnow's concepts of predictability and dimension. Then asymptotic complexities in different prediction environments are compared by describing the set of all pairs of asymptotic complexities w.r.t. different environments. A geometric characterisation in terms of generalised entropies is obtained and thus the results of Lutz and Fortnow are generalised.

1 Introduction

We consider the following on-line learning problem: given a sequence of previous outcomes $x_1, x_2, \ldots, x_{n-1}$, a prediction strategy is required to output a prediction γ_n for the next outcome x_n.

We assume that outcomes belong to a finite set Ω; it may be thought of as an alphabet and sequences as words. We allow greater variation in predictions though. Predictions may be drawn from a compact set. A loss function $\lambda(\omega, \gamma)$ is used to measure the discrepancy between predictions and actual outcomes. The performance of the strategy is measured by the cumulative loss $\sum_{i=1}^{n} \lambda(x_i, \gamma_i)$. Different aspects of this prediction problem have been extensively studied; see [1] for an overview.

A loss function specifies a prediction environment. We study the notion of predictability in a particular environment. There are different approaches to formalising this intuitive notion. One is predictive complexity introduced in [2]. In this paper we introduce another formalisation, namely, asymptotic complexity.

Asymptotic complexity applies to languages, i.e., classes of sequences of outcomes. Roughly speaking, the asymptotic complexity of a language is the loss per element of the best prediction strategy. This definition can be made precise in several natural ways. We thus get several different variants. One of them, which we call lower non-uniform complexity, generalises the concepts of dimension and predictability from [3] (the conference version was presented at COLT 2002). In our framework dimension and predictability can be represented by means of complexities for two specific games.

In this paper we study the following problem. Let AC_1 be asymptotic complexity specified by a loss function λ_1 and let AC_2 be asymptotic complexity specified by a loss function λ_2. What relations exist between them? We give a complete answer to this question by describing the set $(\mathrm{AC}_1(L), \mathrm{AC}_2(L))$, where L ranges over all languages, on the Euclidean plane. The main theorem is formulated in Sect. 4. This set turns out to have a simple geometric description in terms of so called generalised entropy. Generalised entropy is the optimal expected loss per element. In the case of the logarithmic loss function, generalised entropy coincides with Shannon entropy. Generalised entropy is discussed in [4]. In [5] connections between generalised entropy and predictive complexity are studied. We thus generalise the result from [3], where only the case of predictability and dimension is considered.

Our main result holds for all convex games. We show that this requirement cannot be omitted.

The definitions and results in this paper are formulated without any reference to computability. However all constructions in the paper are effective. All the results from the paper can therefore be reformulated in either computable or polynomial-time computable fashion provided the loss functions are computable in a sufficiently efficient way. We discuss this in more detail in Sect. 6.

2 Preliminaries

The notation $\mathbb{N}$ refers to the set of all non-negative integers $\{0, 1, 2, \ldots\}$.

2.1 Games, Strategies, and Losses

A *game* $\mathfrak{G}$ is a triple $\langle \Omega, \Gamma, \lambda \rangle$, where Ω is an *outcome space*, Γ is a *prediction space*, and $\lambda : \Omega \times \Gamma \to [0, +\infty]$ is a *loss function*.

We assume that $\Omega = \{\omega^{(0)}, \omega^{(1)}, \ldots, \omega^{(M-1)}\}$ is a finite set of cardinality $M < +\infty$. If $M = 2$, then Ω may be identified with $\mathbb{B} = \{0, 1\}$; we will call this case *binary*. We denote the set of all finite sequences of elements of Ω by Ω^* and the set of all infinite sequences by Ω^∞; bold letters $\boldsymbol{x}$, $\boldsymbol{y}$ etc. are used to refer to both finite and infinite sequences. By $|\boldsymbol{x}|$ we denote the length of a finite sequence $\boldsymbol{x}$, i.e., the number of elements in it. The set of sequences of length n, $n = 0, 1, 2, \ldots$, is denoted by Ω^n. We will also be using the notation $\sharp_i \boldsymbol{x}$ for the number of $\omega^{(i)}$s among elements of $\boldsymbol{x}$. Clearly, $\sum_{i=0}^{M-1} \sharp_i \boldsymbol{x} = |\boldsymbol{x}|$ for any finite sequence $\boldsymbol{x}$. By $\boldsymbol{x}|_n$ we denote the prefix of length n of a (finite of infinite) sequence $\boldsymbol{x}$.

We also assume that Γ is a compact topological space and λ is continuous w.r.t. the topology of the extended real line $[-\infty, +\infty]$. We treat Ω as a discrete space and thus the continuity of λ in two arguments is the same as continuity in the second argument.

In order to take some important games into account we must allow λ to attain the value $+\infty$. However, we assume that for every $\gamma_0 \in \Gamma$ such that $\lambda(\omega^*, \gamma_0) = +\infty$ for some $\omega^* \in \Omega$, there is a sequence $\gamma_1, \gamma_2, \ldots \in \Gamma$ such that $\gamma_n \to \gamma_0$ and $\lambda(\omega, \gamma_n) < +\infty$ for all $n = 1, 2, \ldots$ and all $\omega \in \Omega$ (but, by continuity, $\lambda(\omega^*, \gamma_n) \to +\infty$ as $n \to +\infty$). In other terms, we assume that every prediction γ_0 leading to infinite loss can be approximated by predictions giving finite losses.

The following are examples of binary games with $\Omega = \mathbb{B}$ and $\Gamma = [0,1]$: the *square-loss* game with the loss function $\lambda(\omega,\gamma) = (\omega - \gamma)^2$, the *absolute-loss* game with the loss function $\lambda(\omega,\gamma) = |\omega - \gamma|$, and the *logarithmic* game with

$$\lambda(\omega,\gamma) = \begin{cases} -\log_2(1-\gamma) & \text{if } \omega = 0 \\ -\log_2 \gamma & \text{if } \omega = 1 \end{cases} .$$

A *prediction strategy* $\mathfrak{A} : \Omega^* \to \Gamma$ maps a finite sequence of outcomes to a prediction. We say that on a finite sequence $\boldsymbol{x} = x_1 x_2 \ldots x_n \in \Omega^n$ the strategy $\mathfrak{A}$ suffers loss $\mathrm{Loss}_{\mathfrak{A}}^{\mathfrak{G}}(\boldsymbol{x}) = \sum_{i=1}^{n} \lambda(x_i, \mathfrak{A}(x_1 x_2 \ldots x_{i-1}))$. By definition, we let $\mathrm{Loss}_{\mathfrak{A}}^{\mathfrak{G}}(\Lambda) = 0$, where Λ is the sequence of length 0.

We need to define one important class of games. The definition is in geometric terms. An M-tuple $(s_0, s_1, \ldots, s_{M-1}) \in [0, +\infty]^M$ is a *superprediction* w.r.t. $\mathfrak{G}$ if there is a prediction $\gamma \in \Gamma$ such that $\lambda(\omega^{(i)}, \gamma) \leq s_i$ for all $i = 0, 1, \ldots, M-1$. We say that the game $\mathfrak{G}$ is *convex* if the finite part of its set of superpredictions, $S \cap \mathbb{R}^M$, where S is the set of superpredictions, is convex.

It is shown in [6] that convexity is equivalent to another property called *weak mixability*. We will be using these terms as synonyms.

2.2 Generalised Entropies

Fix a game $\mathfrak{G} = \langle \Omega, \Gamma, \lambda \rangle$. Let $\mathbb{P}(\Omega)$ be the set of probability distributions on Ω. Since Ω is finite, we can identify $\mathbb{P}(\Omega)$ with the standard $(M-1)$-simplex $\mathbb{P}_M = \{(p_0, p_1, \ldots, p_{M-1}) \in [0,1]^M \mid \sum_{i=0}^{M-1} p_i = 1\}$.

Generalised entropy $H : \mathbb{P}(\Omega) \to \mathbb{R}$ is the infimum of expected loss over $\gamma \in \Gamma$, i.e., for $p^* = (p_0, p_1, \ldots, p_{M-1}) \in \mathbb{P}(\Omega)$

$$H(p^*) = \min_{\gamma \in \Gamma} \mathbf{E}_{p^*} \lambda(\omega, \gamma) = \min_{\gamma \in \Gamma} \sum_{i=0}^{M-1} p_i \lambda(\omega^{(i)}, \gamma) .$$

The minimum in the definition is achieved because λ is continuous and Γ compact.

Since p_i can accept the value 0 and $\lambda(\omega^{(i)}, \gamma)$ can be $+\infty$, we need to resolve the possible ambiguity. Let us assume that in this definition $0 \times (+\infty) = 0$. This is the same as replacing the minimum by the infimum over the values of $\gamma \in \Gamma$ such that $\lambda(\omega, \gamma) < +\infty$ for all $\omega \in \Omega$.

In the binary case $\Omega = \mathbb{B}$ the definition can be simplified. Let p be the probability of 1. Clearly, p fully specifies a distribution from $\mathbb{P}(\mathbb{B})$ and thus $\mathbb{P}(\mathbb{B})$ can be identified with the line segment $[0,1]$. We get $H(p) = \min_{\gamma \in \Gamma}[(1-p)\lambda(0,\gamma) + p\lambda(1,\gamma)]$.

If it is not clear from the context what game we are referring to, we will use subscripts for H. We will use the term $\mathfrak{G}$-*entropy* to refer to generalised entropy w.r.t. the game $\mathfrak{G}$. The notation ABS, SQ, and LOG will be used to refer to the absolute-loss, square-loss, and logarithmic games respectively, e.g., we will write 'ABS-entropy'.

3 Asymptotic Complexities

Fix a game $\mathfrak{G} = \langle \Omega, \Gamma, \lambda \rangle$. We are going to define measures of complexity for *languages*, i.e., sets of sequences. The finite and infinite sequences should be considered separately.

3.1 Finite Sequences

In this subsection we consider languages $L \subseteq \Omega^*$. We shall call the values

$$\overline{\mathrm{AC}}(L) = \inf_{\mathfrak{A}} \limsup_{n \to +\infty} \max_{\boldsymbol{x} \in L \cap \Omega^n} \frac{\mathrm{Loss}_{\mathfrak{A}}(\boldsymbol{x})}{n} \ , \tag{1}$$

$$\underline{\mathrm{AC}}(L) = \inf_{\mathfrak{A}} \liminf_{n \to +\infty} \max_{\boldsymbol{x} \in L \cap \Omega^n} \frac{\mathrm{Loss}_{\mathfrak{A}}(\boldsymbol{x})}{n} \tag{2}$$

upper and *lower asymptotic complexity* of L w.r.t. the game $\mathfrak{G}$. As with generalised entropies, we will use subscripts for AC to specify a particular game if it is not clear from the context.

In order to complete the definition, we must decide what to do if L contains no sequences of certain lengths at all. In this paper we are concerned only with infinite sets of finite sequences. One can say that asymptotic complexity of a finite language $L \subseteq \Omega^*$ is undefined. Let us also assume that the limits in (1) and (2) are taken over such n that $L \cap \Omega^n \neq \varnothing$. An alternative arrangement is to assume that in (1) $\max \varnothing = 0$, while in (2) $\max \varnothing = +\infty$.

3.2 Infinite Sequences

There are two natural ways to define complexities of languages $L \subseteq \Omega^\infty$.

First we can extend the notions we have just defined. Indeed, every nonempty set of infinite sequences can be identified with the set of all finite prefixes of all its sequences. The language thus obtained is infinite and has upper and lower complexities. For the resulting complexities we shall retain the notation $\overline{\mathrm{AC}}(L)$ and $\underline{\mathrm{AC}}(L)$. We shall refer to those complexities as *uniform*.

The second way is the following. Let

$$\overline{\overline{\mathrm{AC}}}(L) = \inf_{\mathfrak{A}} \sup_{\boldsymbol{x} \in L} \limsup_{n \to +\infty} \frac{\mathrm{Loss}_{\mathfrak{A}}(\boldsymbol{x}|_n)}{n} \text{ and } \underline{\underline{\mathrm{AC}}}(L) = \inf_{\mathfrak{A}} \sup_{\boldsymbol{x} \in L} \liminf_{n \to +\infty} \frac{\mathrm{Loss}_{\mathfrak{A}}(\boldsymbol{x}|_n)}{n} \ .$$

We shall refer to this complexity as *non-uniform*.

The concept of asymptotic complexity generalises certain complexity measures studied in the literature. The concepts of predictability and dimension studied in [3] can be easily reduced to asymptotic complexity: the dimension is the lower non-uniform complexity w.r.t. a multidimensional generalisation of the logarithmic game and predictability equals $1 - \underline{\underline{\mathrm{AC}}}$, where $\underline{\underline{\mathrm{AC}}}$ is the lower non-uniform complexity w.r.t. a multidimensional generalisation of the absolute-loss game.

3.3 Differences between Complexities

Let us show that the complexities we have introduced are different.

First let us show that upper and lower complexities differ. For example, consider the absolute-loss game. Let $0^{(n)}$ be the sequence of n zeros and let $\Xi_n = \{0^{(n)}\} \times \mathbb{B}^n$. Consider the language $L = \prod_{i=0}^{\infty} \Xi_{2^{2^i}} \subseteq \mathbb{B}^\infty$. In other terms, L consists of sequences that have alternating constant and random segments. It is easy to see that $\overline{\mathrm{AC}}(L) = \overline{\overline{\mathrm{AC}}}(L) = 1/2$, while $\underline{\mathrm{AC}}(L) = \underline{\underline{\mathrm{AC}}}(L) = 0$.

Secondly, let us show that uniform complexities differ from non-uniform. Once again, consider the absolute-loss game. Let $L \subseteq \mathbb{B}^\infty$ be the set of all sequences that have only zeros from some position on. In other terms, $L = \cup_{n=0}^{\infty} (\mathbb{B}^n \times \{0^{(\infty)}\})$, where $0^{(\infty)}$ is the infinite sequence of zeros. We have $\overline{\overline{\mathrm{AC}}}(L) = \underline{\underline{\mathrm{AC}}}(L) = 0$ while $\overline{\mathrm{AC}}(L) = \underline{\mathrm{AC}}(L) = 1/2$.

4 Main Result

Consider two games $\mathfrak{G}_1$ and $\mathfrak{G}_2$ with the same finite set of outcomes Ω. Let H_1 be $\mathfrak{G}_1$-entropy and H_2 be $\mathfrak{G}_2$-entropy. The $\mathfrak{G}_1/\mathfrak{G}_2$-*entropy set* is the set $\{(H_1(p), H_2(p)) \mid p \in \mathbb{P}(\Omega)\}$. The convex hull of the $\mathfrak{G}_1/\mathfrak{G}_2$-entropy set is called the $\mathfrak{G}_1/\mathfrak{G}_2$-*entropy hull*.

We say that a closed convex $\mathcal{S} \subseteq \mathbb{R}^2$ is a *spaceship* if for every pair of points $(x_1, y_1), (x_2, y_2) \in \mathcal{S}$ the point $(\max(x_1, x_2), \max(y_1, y_2))$ belongs to $\mathcal{S}$. The *spaceship closure* of a set $\mathcal{H} \subseteq \mathbb{R}^2$ is the smallest spaceship containing $\mathcal{H}$, i.e., the intersection of all spaceships containing $\mathcal{H}$.

We can now formulate the main result of this paper.

Theorem 1. *If games $\mathfrak{G}_1$ and $\mathfrak{G}_2$ have the same finite outcome space Ω and are convex, then the spaceship closure of the $\mathfrak{G}_1/\mathfrak{G}_2$-entropy hull coincides with the following sets, where* AC_1 *and* AC_2 *are asymptotic complexities w.r.t. $\mathfrak{G}_1$ and $\mathfrak{G}_2$:*

- $\{(\overline{\mathrm{AC}}_1(L),\ \overline{\mathrm{AC}}_2(L)) \mid L \subseteq \Omega^* \text{ and } L \text{ is infinite}\}$;
- $\{(\underline{\mathrm{AC}}_1(L),\ \underline{\mathrm{AC}}_2(L)) \mid L \subseteq \Omega^* \text{ and } L \text{ is infinite}\}$;
- $\{(\overline{\mathrm{AC}}_1(L),\ \overline{\mathrm{AC}}_2(L)) \mid L \subseteq \Omega^\infty \text{ and } L \neq \varnothing\}$;
- $\{(\underline{\mathrm{AC}}_1(L),\ \underline{\mathrm{AC}}_2(L)) \mid L \subseteq \Omega^\infty \text{ and } L \neq \varnothing\}$;
- $\{(\overline{\overline{\mathrm{AC}}}_1(L),\ \overline{\overline{\mathrm{AC}}}_2(L)) \mid L \subseteq \Omega^\infty \text{ and } L \neq \varnothing\}$;
- $\{(\underline{\underline{\mathrm{AC}}}_1(L),\ \underline{\underline{\mathrm{AC}}}_2(L)) \mid L \subseteq \Omega^\infty \text{ and } L \neq \varnothing\}$.

In other words, the spaceship closure $\mathcal{S}$ of the entropy hull contains all points $(\mathrm{AC}_1(L), \mathrm{AC}_2(L))$, where AC is one type of complexity, and these points fill the set $\mathcal{S}$ as L ranges over all languages that have complexity. The last item on the list covers Theorem 5.1 (Main Theorem) from [3].

Appendices A and B contain a discussion of shapes of the entropy hull and some examples. The theorem is proved in Sect. 5.

The requirement of convexity cannot be omitted. For example, consider the simple prediction game $\langle \mathbb{B}, \mathbb{B}, \lambda \rangle$, where $\lambda(\omega, \gamma)$ is 0 if $\omega = \gamma$ and 1 otherwise.

The convex hull of the set of superpredictions w.r.t. the simple prediction game coincides with the set of superpredictions w.r.t. the absolute-loss game. Geometric considerations imply that their generalised entropies coincide. Thus the maximum of the generalised entropy w.r.t. the simple prediction game is $1/2$ (see Appendix B). On the other hand, it is easy to check that $\mathrm{AC}(\mathbb{B}^*) = 1$, where AC is any of the asymptotic complexities w.r.t. the simple prediction game.

The statement of the theorem does not apply to pairs $(\overline{\mathrm{AC}}_1(L), \underline{\mathrm{AC}}_2(L))$ or pairs $(\overline{\overline{\mathrm{AC}}}_1(L), \underline{\underline{\mathrm{AC}}}_2(L))$. Indeed, let $\mathfrak{G}_1 = \mathfrak{G}_2$. Then $H_1 = H_2$ and the entropy hull with its spaceship closure are subsets of the bisector of the first quadrant. At the same time we know that upper and lower complexities differ and thus there will be pairs outside the bisector.

5 Proof of the Main Theorem

In this section we prove Theorem 1.

The following lemma proved in Appendix C allows us to 'optimise' the performance of a strategy w.r.t. two games. We shall call it *recalibration lemma*.

Lemma 1. *If $\mathfrak{G}_1$ and $\mathfrak{G}_2$ are convex games with the same finite set of outcomes Ω and $\mathcal{H}$ is the $\mathfrak{G}_1/\mathfrak{G}_2$-entropy hull, then for every prediction strategy $\mathfrak{A}$ and positive ε there are prediction strategies $\mathfrak{A}^1_\varepsilon$ and $\mathfrak{A}^2_\varepsilon$ and a function $f : \mathbb{N} \to \mathbb{R}$ such that $f(n) = o(n)$ as $n \to +\infty$ and for every sequence $\boldsymbol{x} \in \Omega^*$ there exists a point $(u_{\boldsymbol{x}}, v_{\boldsymbol{x}}) \in \mathcal{H}$ such that the following inequalities hold:*

$$u_{\boldsymbol{x}}|\boldsymbol{x}| \le \mathrm{Loss}^{\mathfrak{G}_1}_{\mathfrak{A}}(\boldsymbol{x}) + \varepsilon|\boldsymbol{x}| \ , \tag{3}$$

$$\mathrm{Loss}^{\mathfrak{G}_1}_{\mathfrak{A}^1_\varepsilon}(\boldsymbol{x}) \le |\boldsymbol{x}|(u_{\boldsymbol{x}} + \varepsilon) + f(|\boldsymbol{x}|) \ , \tag{4}$$

$$\mathrm{Loss}^{\mathfrak{G}_2}_{\mathfrak{A}^2_\varepsilon}(\boldsymbol{x}) \le |\boldsymbol{x}|(v_{\boldsymbol{x}} + \varepsilon) + f(|\boldsymbol{x}|) \ . \tag{5}$$

Below in Subsect. 5.1 we use this lemma to show that pairs of complexities belong to the spaceship closure. It remains to show that the pairs fill in the closure and it is done in Appendix D.

5.1 Every Pair of Complexities Belongs to the Spaceship Closure of the Hull

Let AC be one of the types of complexity we have introduced. Let us show that for every language L the pair $(\mathrm{AC}_1(L), \mathrm{AC}_2(L))$ belongs to the spaceship closure $\mathcal{S}$ of the entropy hull.

We start by showing that the pair $(\mathrm{AC}_1(L), \mathrm{AC}_2(L))$ belongs to the cucumber closure $\mathcal{C}$ of the $\mathfrak{G}_1/\mathfrak{G}_2$-entropy hull $\mathcal{H}$ (see Appendix A for a definition).

Let $\mathrm{AC}_1(L) = c$. Lemma 1 implies that $c_{\min} \le c \le c_{\max}$, where $c_{\min} = \min_{p\in\mathbb{P}(\Omega)} H_1(p)$ and $c_{\max} = \max_{p\in\mathbb{P}(\Omega)} H_1(p)$.

We need to show that $c_2 \le \mathrm{AC}_2(L) \le c_1$, where c_1 and c_2 correspond to intersections of the vertical line $x = c$ with the boundary of the cucumber as shown on Fig. 1.

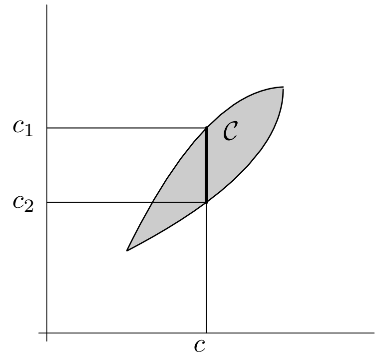

Fig. 1. A section of the cucumber hull

Let $f_1, f_2 : [c_{\min}, c_{\max}] \to \mathbb{R}$ be the non-decreasing functions that bound $\mathcal{C}$ from above and below, i.e., $\mathcal{C} = \{(x, y) \mid x \in [c_{\min}, c_{\max}] \text{ and } f_2(x) \le y \le f_1(x)\}$ (see Appendix A). We have $f_1(c) = c_1$ and $f_2(c) = c_2$. The function f_1 is concave and the function f_2 is convex; therefore they are continuous.

Since $\mathrm{AC}_1(L) = c$, for every $\varepsilon > 0$ there is a prediction strategy $\mathfrak{A}$ such that for certain infinite sets of finite sequences $\boldsymbol{x}$ there are functions $g : \mathbb{N} \to \mathbb{R}$ such that $g(n) = o(n)$ as $n \to +\infty$ and for all appropriate $\boldsymbol{x}$ we have

$$\mathrm{Loss}_{\mathfrak{A}}^{\mathfrak{G}_1}(\boldsymbol{x}) \le (c + \varepsilon)|\boldsymbol{x}| + g(|\boldsymbol{x}|) \ . \tag{6}$$

Let us apply Lemma 1 to $\mathfrak{A}$ and ε. We obtain the strategies $\mathfrak{A}_\varepsilon^1$ and $\mathfrak{A}_\varepsilon^2$ and a function $f : \mathbb{N} \to \mathbb{R}$ such that $f(n) = o(n)$ as $n \to +\infty$ and for every $\boldsymbol{x}$ there exists a point $(u_{\boldsymbol{x}}, v_{\boldsymbol{x}}) \in \mathcal{H}$ such that the inequalities

$$\begin{gathered}
\mathrm{Loss}_{\mathfrak{A}_\varepsilon^1}^{\mathfrak{G}_1}(\boldsymbol{x}) \le |\boldsymbol{x}|(u_{\boldsymbol{x}} + \varepsilon) + f(|\boldsymbol{x}|) \ , \\
\mathrm{Loss}_{\mathfrak{A}_\varepsilon^2}^{\mathfrak{G}_2}(\boldsymbol{x}) \le |\boldsymbol{x}|(v_{\boldsymbol{x}} + \varepsilon) + f(|\boldsymbol{x}|) \ , \\
|\boldsymbol{x}|u_{\boldsymbol{x}} \le \mathrm{Loss}_{\mathfrak{A}}^{\mathfrak{G}_1}(\boldsymbol{x}) + \varepsilon|\boldsymbol{x}| \le (c + 2\varepsilon)|\boldsymbol{x}| + g(|\boldsymbol{x}|)
\end{gathered}$$

hold. The last inequality implies that $u_{\boldsymbol{x}} \le c + 2\varepsilon + o(1)$ at $|\boldsymbol{x}| \to \infty$ and thus for all sufficiently long sequences $\boldsymbol{x}$ we have $u_{\boldsymbol{x}} \le c + 3\varepsilon$. Therefore the point $(u_{\boldsymbol{x}}, v_{\boldsymbol{x}})$ lies to the left of the line $x = c + 3\varepsilon$. This implies $v_{\boldsymbol{x}} \le f_1(c + 3\varepsilon)$ and $\mathrm{AC}_2(L) \le f_1(x + 3\varepsilon) + \varepsilon$. Since f_1 is continuous and $\varepsilon > 0$ is arbitrary, we get $\mathrm{AC}_2(L) \le f_1(c) = c_1$.

Now let us prove that $\mathrm{AC}_2(L) \ge c_2$. Assume the contrary. Let $\mathrm{AC}_2(L) = c_2 - \delta_2$, where $\delta_2 > 0$. There is $\delta_1 > 0$ such that $f_2(c - \delta_1) = c_2 - \delta_2$. By applying the same argument as above to the 'swapped' situation, one can show that $\mathrm{AC}_1(L) \le c - \delta_1$. This contradicts the assumption that $\mathrm{AC}_1(L) = c$.

In order to show that $(\mathrm{AC}_1(L), \mathrm{AC}_2(L)) \in \mathcal{S}$, it remains to prove that $(\mathrm{AC}_1(L), \mathrm{AC}_2(L)) \notin \mathcal{C} \setminus \mathcal{S}$. Let $U = \{(u, v) \in \mathbb{R}^2 \mid \exists (h_1, h_2) \in \mathcal{H} : h_1 \le u \text{ and } h_2 \le v\}$ be the set of points that lie 'above' the entropy set $\mathcal{H}$. Let e_i,

$i = 0, 1, \ldots, M-1$, be the vector with the i-th component equal to 1 and all other components equal to 0. Clearly, $e_i \in \mathbb{P}_M$; it represents a special degenerate distribution. We have $H_1(e_i) = \min_{\gamma\in\Gamma_1} \lambda_1(\omega^{(i)}, \gamma)$. For any prediction strategy $\mathfrak{A}$ we get

$$\mathrm{Loss}_{\mathfrak{A}}^{\mathfrak{G}_1}(\boldsymbol{x}) \geq \sum_{i=0}^{M-1} \sharp_i\boldsymbol{x} \min_{\gamma\in\Gamma_1} \lambda_1(\omega^{(i)}, \gamma) = |\boldsymbol{x}| \sum_{i=0}^{M-1} p_i H_1(e_i) \ ,$$

where $p_i = \sharp_i\boldsymbol{x}/|\boldsymbol{x}|$. The same holds for $\mathfrak{G}_2$. We thus get inequalities

$$\frac{\mathrm{Loss}_{\mathfrak{A}_1}^{\mathfrak{G}_1}(\boldsymbol{x})}{|\boldsymbol{x}|} \geq \sum_{i=0}^{M-1} p_i H_1(e_i) \quad \text{and} \quad \frac{\mathrm{Loss}_{\mathfrak{A}_2}^{\mathfrak{G}_2}(\boldsymbol{x})}{|\boldsymbol{x}|} \geq \sum_{i=0}^{M-1} p_i H_2(e_i) \ ,$$

where p_i depend only on $\boldsymbol{x}$, for all strategies $\mathfrak{A}_1$ and $\mathfrak{A}_2$. Therefore the pair $(\mathrm{Loss}_{\mathfrak{A}_1}^{\mathfrak{G}_1}(\boldsymbol{x})/|\boldsymbol{x}|, \mathrm{Loss}_{\mathfrak{A}_2}^{\mathfrak{G}_2}(\boldsymbol{x})/|\boldsymbol{x}|)$ belongs to U. Since U is closed, the same holds for every pair $(\mathrm{AC}_1(L), \mathrm{AC}_2(L))$.

6 Computability Aspects

The definition of the asymptotic complexity can be modified in the following way. The infima in definitions may be taken over a particular class of strategies. Examples of such classes are the computable strategies and polynomial-time computable strategies. This provides us with different definitions of asymptotic complexity. The theorems from this paper still hold for these modified complexities provided some straightforward adjustments are made.

If we want to take computability aspects into consideration, we need to impose computability restrictions on loss functions. If we are interested in computable strategies, it is natural to consider computable loss functions.

The definition of weak mixability needs modifying too. It is natural to require that the strategy $\mathfrak{A}$ obtained by aggregating $\mathfrak{A}_1$ and $\mathfrak{A}_2$ is computable by an algorithm that has access to oracles computing $\mathfrak{A}_1$ and $\mathfrak{A}_2$. Results from [6] still hold since strategies can be merged effectively provided λ is computable.

The recalibration procedure provides us with strategies $\mathfrak{A}_\varepsilon^1$ and $\mathfrak{A}_\varepsilon^2$ that are computable given an oracle computing $\mathfrak{A}$. The proof of the main theorem remains valid almost literally. Note that we do not require the languages L to be computable in any sense. We are only concerned with transforming some strategies into others. If the original strategies are computable, the resulting strategies will be computable too. All pairs $(\mathrm{AC}_1(L), \mathrm{AC}_2(L))$ still belong to the spaceship closure of the entropy hull and fill it.

Similar remarks can be made about polynomial-time computability.

Acknowledgements

The authors would like to thank Ivan Polikarov from the Moscow State University for his inspiring ideas on turnip and spaceships.

References

1. Cesa-Bianchi, N., Lugosi, G.: Prediction, Learning, and Games. Cambridge University Press (2006)
2. Vovk, V., Watkins, C.J.H.C.: Universal portfolio selection. In: Proceedings of the 11th Annual Conference on Computational Learning Theory, ACM Press (1998) 12–23
3. Fortnow, L., Lutz, J.H.: Prediction and dimension. Journal of Computer and System Sciences **70**(4) (2005) 570–589
4. Grünwald, P.D., Dawid, A.P.: Game theory, maximum entropy, minimum discrepancy and robust Bayesian decision theory. The Annals of Statistics **32**(4) (2004) 1367–1433
5. Kalnishkan, Y., Vovk, V., Vyugin, M.V.: Loss functions, complexities, and the Legendre transformation. Theoretical Computer Science **313**(2) (2004) 195–207
6. Kalnishkan, Y., Vyugin, M.V.: The weak aggregating algorithm and weak mixability. In: Learning Theory, Proceedings of the 18th Annual Conference (COLT 2005). Volume 3559 of Lecture Notes in Computer Science., Springer (2005) 188–203
7. Eggleston, H.G.: Convexity. Cambridge University Press, Cambridge (1958)
8. Hoeffding, W.: Probability inequalities for sums of bounded random variables. Journal of the American Statistical Association **58**(301) (March 1963) 13–30

Appendix A. Shapes of Entropy Hulls

In this section we discuss geometrical aspects of the statement of the main theorem in more detail.

We start with a fundamental property of entropy. The set $\mathbb{P}_M$ is convex. Therefore we can prove by direct calculation the following lemma.

Lemma 2. *If $H : \mathbb{P}_M \to \mathbb{R}$ is $\mathfrak{G}$-entropy, then H is concave.*

Note that concavity of H implies continuity of H. Therefore every entropy set is a closed set w.r.t. the standard Euclidean topology. It is also bounded. Thus the entropy hull is also bounded and closed (see, e.g., [7], Theorem 10).

We need to introduce a classification of planar convex sets. A closed convex $\mathcal{C} \subseteq \mathbb{R}^2$ is a *cucumber* if for every pair of points $(x_1, y_1), (x_2, y_2) \in \mathcal{C}$ the points $(\min(x_1, x_2), \min(y_1, y_2))$ and $(\max(x_1, x_2), \max(y_1, y_2))$ belong to $\mathcal{C}$. In other terms, a closed convex $\mathcal{C}$ is a cucumber if and only if there are nondecreasing functions $f_1, f_2 : I \to \mathbb{R}$, where I is an interval, perhaps infinite, such that $\mathcal{C} = \{(x, y) \mid x \in I \text{ and } f_1(x) \leq y \leq f_2(x)\}$.

If a closed convex set is not a cucumber, we call it a *turnip*.

We will formulate a criterion for $\mathcal{H}$ to be a cucumber. Let $\arg\min f$, where f is a function from I to $\mathbb{R}$, be the set of points of I where f achieves the value of its global minimum on I. If no global minimum exists, the set $\arg\min f$ is empty. The notation $\arg\max f$ is defined similarly.

Lemma 3. *If $H_1, H_2 : I \to \mathbb{R}$, where $I \subseteq \mathbb{R}^n$ is a closed bounded set, are two continuous functions, then the convex hull of the set $\{(H_1(p), H_2(p)) \mid p \in I\}$ is* *a cucumber if and only if the following pair of conditions hold:*

$$\arg\min H_1 \cap \arg\min H_2 \neq \varnothing$$
$$\arg\max H_1 \cap \arg\max H_2 \neq \varnothing \ .$$

In the binary case natural games, including the absolute-loss, square-loss and logarithmic games, are *symmetric*, i.e., their sets of superpredictions are symmetric w.r.t. the bisector of the positive quadrangle. For example, every game $\mathfrak{G} = \langle \mathbb{B}, [0,1], \lambda \rangle$ such that $\lambda(0,\gamma) = \lambda(1, 1-\gamma)$ for all $\gamma \in [0,1]$ is symmetric. Clearly, if H is $\mathfrak{G}$-entropy w.r.t. a symmetric game $\mathfrak{G}$, then $H(p) = H(1-p)$ for all $p \in [0,1]$. Thus H achieves its maximum at $p = 1/2$ and its minimum at $p = 0$ and $p = 1$. Therefore if H_1 and H_2 are entropies w.r.t. symmetric games $\mathfrak{G}_1$ and $\mathfrak{G}_2$, then their $\mathfrak{G}_1/\mathfrak{G}_2$-entropy hull is a cucumber.

The *cucumber closure* of a set $\mathcal{H} \subseteq \mathbb{R}^2$ is the smallest cucumber that contains $\mathcal{H}$, i.e., the intersection of all cucumbers that contain $\mathcal{H}$.

The definition of a spaceship given above uses only the upper point of the two that should belong to a cucumber. In terms of boundaries the definition is as follows. A closed convex $\mathcal{S} \subseteq \mathbb{R}^2$ is a spaceship if and only if there are functions $f_1, f_2 : I \to \mathbb{R}$, where I is an interval, perhaps infinite, such that f_2 is non-decreasing and $\mathcal{S} = \{(x,y) \mid x \in I \text{ and } f_1(x) \leq y \leq f_2(x)\}$.

Lemma 4. *If $H_1, H_2 : I \to \mathbb{R}$, where $I \subseteq \mathbb{R}$ is a closed bounded interval, are two continuous functions, then the convex hull of the set $\{(H_1(p), H_2(p)) \mid I\}$ is a spaceship if and only if the following condition holds:*

$$\arg\max H_1 \cap \arg\max H_2 \neq \varnothing \ . \tag{7}$$

Note that the definitions of turnips, cucumbers, and spaceships as well as of cucumber and spaceship closures are coordinate-dependent.

Appendix B. Examples of Entropy Hulls

This section contains some examples of entropy sets and hulls for the binary case.

It is easy to check by direct calculation that the ABS-entropy is given by $H^{\mathrm{ABS}}(p) = \min(p, 1-p)$, the SQ-entropy is given by $H^{\mathrm{SQ}}(p) = p(1-p)$, and the LOG-entropy is given by $H^{\mathrm{LOG}}(p) = -p\log_2 p - (1-p)\log_2(1-p)$, and thus it coincides with Shannon entropy. The entropy hulls are shown on Figs. 2, 3, and 4; the corresponding entropy sets are represented by bold lines. Since all the three games are symmetric, the entropy hulls are cucumbers.

Let us construct an entropy hull that is a turnip. It follows from the previous section, that the example must be rather artificial. Let $\mathfrak{G}_1 = \langle \mathbb{B}, [0,1], \lambda_1 \rangle$, where $\lambda_1(0,\gamma) = \gamma$ and $\lambda_1(1,\gamma) = 1 - \frac{\gamma}{2}$ for all $\gamma \in [0,1]$, and let $\mathfrak{G}_2 = \langle \mathbb{B}, [0,1], \lambda_2 \rangle$, where $\lambda_2(0,\gamma) = 1 + \frac{\gamma}{2}$ and $\lambda_2(1,\gamma) = \frac{3}{2} - \gamma$ for all $\gamma \in [0,1]$. It is easy to evaluate the corresponding entropies, $H_1(p) = \min(p, 1 - \frac{p}{2})$ and $H_2(p) = \min(1 + \frac{p}{2}, \frac{3}{2} - p)$.

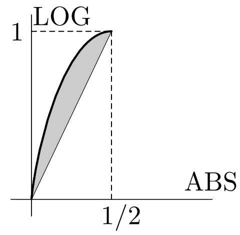

Fig. 2. The ABS/LOG-entropy set and hull

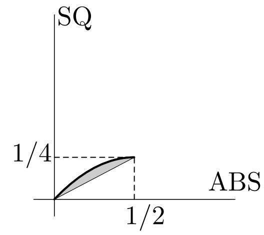

Fig. 3. The ABS/SQ-entropy set and hull

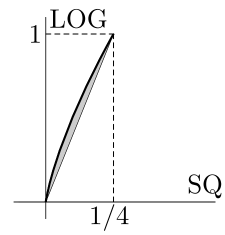

Fig. 4. The LOG/SQ-entropy set and hull

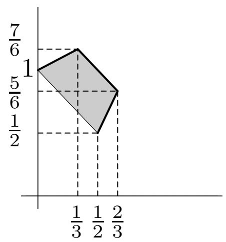

Fig. 5. The $\mathfrak{G}_1/\mathfrak{G}_2$-entropy hull (a turnip)

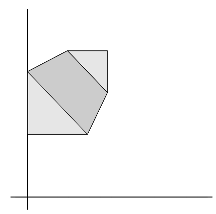

Fig. 6. The cucumber closure

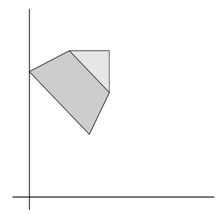

Fig. 7. The spaceship closure

Fig. 5 shows $\mathfrak{G}_1/\mathfrak{G}_2$-entropy hull, which is a turnip. Figure 6 shows its cucumber closure, while Fig. 7 shows its spaceship closure.

Appendix C. Proof of the Recalibration Lemma

Let $\mathfrak{G}_1 = \langle \Omega, \Gamma_1, \lambda_1 \rangle$ and $\mathfrak{G}_2 = \langle \Omega, \Gamma_2, \lambda_2 \rangle$. We shall describe the procedure transforming $\mathfrak{A}$ into $\mathfrak{A}^1_\varepsilon$ and $\mathfrak{A}^2_\varepsilon$. The construction is in four stages.

First let us perform an ε-*quantisation* of $\mathfrak{A}$.

Lemma 5. *For any $\mathfrak{G} = \langle \Omega, \Gamma, \lambda \rangle$ and $\varepsilon > 0$ there is a finite set Γ_ε such that for any $\gamma \in \Gamma$ there is $\gamma^* \in \Gamma_\varepsilon$ such that $\lambda(\omega, \gamma^*) \leq \lambda(\omega, \gamma) + \varepsilon$ for every $\omega \in \Omega$.*

Lemma 2 from [6] implies that it is sufficient to consider bounded loss functions λ. If λ is bounded, the lemma follows from continuity of λ and compactness of Γ.

Let us find such finite subsets $\Gamma_\varepsilon \subseteq \Gamma_1$ and $\Gamma''_\varepsilon \subseteq \Gamma_2$. Without restricting the generality, one can assume that they are of the same size $|\Gamma_\varepsilon| = |\Gamma''_\varepsilon| = N$; indeed, if this is not the case, we can add more elements to the smaller set. Let $\Gamma_\varepsilon = \{\gamma_1, \gamma_2, \ldots, \gamma_N\}$ and $\Gamma''_\varepsilon = \{\gamma''_1, \gamma''_2, \ldots, \gamma''_N\}$.

There is a strategy $\mathfrak{A}^{\mathrm{q}}_\varepsilon$ that outputs only predictions from Γ_ε and such that $\mathrm{Loss}^{\mathfrak{G}_1}_{\mathfrak{A}^{\mathrm{q}}_\varepsilon}(\boldsymbol{x}) \leq \mathrm{Loss}^{\mathfrak{G}_1}_{\mathfrak{A}}(\boldsymbol{x}) + \varepsilon|\boldsymbol{x}|$ for all $\boldsymbol{x} \in \Omega^*$.

Secondly let us construct the table of frequencies.

Given a sequence $\boldsymbol{x}$ of length n, let us count the number of times each of these predictions $\gamma_1, \gamma_2, \ldots, \gamma_N$ occurs as $\mathfrak{A}^{\mathrm{q}}_\varepsilon$ predicts elements of $\boldsymbol{x}$. For each $j = 1, 2, \ldots, N$ and $i = 0, 1, \ldots, M-1$ let $n^{(i)}_j$ be the number of occasions when

Table 1. Predictions and outcomes for a given sequence $\boldsymbol{x}$

Predictions	Number of $\omega^{(0)}$s	Number of $\omega^{(1)}$s	$\vdots$	Number of $\omega^{(M-1)}$s
γ_1	$n_1^{(0)}$	$n_1^{(1)}$		$n_1^{(M-1)}$
γ_2	$n_2^{(0)}$	$n_2^{(1)}$	$\vdots$	$n_2^{(M-1)}$
		$\dots$		
γ_j	$n_j^{(0)}$	$n_j^{(1)}$	$\vdots$	$n_j^{(M-1)}$
		$\dots$		
γ_N	$n_N^{(0)}$	$n_N^{(1)}$	$\vdots$	$n_N^{(M-1)}$

$\mathfrak{A}_\varepsilon^{\mathrm{q}}$ outputs the prediction γ_j just before the outcome $\omega^{(i)}$ occurs. We get Table 1, where $\sum_{j=1}^N n_j^{(i)} = \sharp_i \boldsymbol{x}$ for all $i = 0, 1, \dots, M-1$.

Thirdly we perform the recalibration and construct auxiliary 'strategies' $\widetilde{\mathfrak{A}_\varepsilon^{(1)}}$ and $\widetilde{\mathfrak{A}_\varepsilon^{(2)}}$. Formally they are not strategies because they have access to side information.

Suppose that we predict elements of $\boldsymbol{x}$ and have access to Table 1 right from the start. We can optimise the performance of $\mathfrak{A}_\varepsilon^{\mathrm{q}}$ as follows. If on some step $\mathfrak{A}_\varepsilon^{\mathrm{q}}$ outputs γ_j, we know that we are on the j-th line of the table. However γ_j is not necessarily the best prediction to output in this situation. Let γ_j^* be an element of Γ_ε where the minimum $\min_{\gamma \in \Gamma_\varepsilon} \sum_{i=0}^{M-1} n_j^{(i)} \lambda_1(\omega^{(i)}, \gamma)$ is attained. This minimum can be expressed though the generalised entropy H_1. Put $p_j^{(i)} = n_j^{(i)} / \sum_i n_j^{(i)}$; the M-tuple $p_j = (p_j^{(0)}, p_j^{(1)}, \dots, p_j^{(M-1)})$ is a distribution on Ω. We have $\sum_{i=0}^{M-1} n_j^{(i)} \lambda_1(\omega^{(i)}, \gamma) = \left(\sum_i n_j^{(i)}\right) \sum_{i=0}^{M-1} p_j^{(i)} \lambda_1(\omega^{(i)}, \gamma)$ and thus

$$\left(\sum_i n_j^{(i)}\right) H_1(p_j) \le \sum_{i=0}^{M-1} n_j^{(i)} \lambda_1(\omega^{(i)}, \gamma_j^*) \le \left(\sum_i n_j^{(i)}\right) (H_1(p_j) + \varepsilon) \ . \tag{8}$$

Let us output γ_j^* each time instead of γ_j.

This is how the 'strategy' $\widetilde{\mathfrak{A}_\varepsilon^1}$ works. The loss of $\widetilde{\mathfrak{A}_\varepsilon^1}$ on $\boldsymbol{x}$ is $\mathrm{Loss}_{\widetilde{\mathfrak{A}_\varepsilon^1}}^{\mathfrak{G}_1}(\boldsymbol{x}) = \sum_{j=1}^N \sum_{i=0}^{M-1} n_j^{(i)} \lambda_1(\omega^{(i)}, \gamma_j^*)$. Put $q_j = \left(\sum_{i=0}^{M-1} n_j^{(i)}\right)/n$; we have $\sum_j q_j = 1$. It follows from (8) that

$$\left| \mathrm{Loss}_{\widetilde{\mathfrak{A}_\varepsilon^1}}^{\mathfrak{G}_1}(\boldsymbol{x}) - n \sum_{j=1}^N q_j H_1(p_j) \right| \le \varepsilon n \ . \tag{9}$$

On the other hand, $\mathrm{Loss}_{\widetilde{\mathfrak{A}_\varepsilon^1}}^{\mathfrak{G}_1}(\boldsymbol{x}) \le \mathrm{Loss}_{\mathfrak{A}_\varepsilon^{\mathrm{q}}}^{\mathfrak{G}_1}(\boldsymbol{x}) \le \mathrm{Loss}_{\mathfrak{A}}^{\mathfrak{G}_1}(\boldsymbol{x}) + \varepsilon n$ because $\widetilde{\mathfrak{A}_\varepsilon^1}$ attempts to minimise losses. Note that each sequence $\boldsymbol{x} \in \Omega^*$ specifies its own sets of values p and q.

The strategy $\widetilde{\mathfrak{A}^2_\varepsilon}$ works as follows. It simulates $\widetilde{\mathfrak{A}^{\mathrm{q}}_\varepsilon}$ and when $\widetilde{\mathfrak{A}^{\mathrm{q}}_\varepsilon}$ outputs γ_j it finds itself on the j-th line of the table and outputs $\gamma_j^{**} \in \Gamma''_\varepsilon$ such that the minimum $\min_{\gamma\in\Gamma''_\varepsilon}\sum_{i=0}^{M-1} n_j^{(i)}\lambda_2(\omega^{(i)},\gamma)$ is attained on γ_j^{**}. We obtain

$$\left|\mathrm{Loss}^{\mathfrak{G}_2}_{\widetilde{\mathfrak{A}^2_\varepsilon}}(\boldsymbol{x}) - n\sum_{i=1}^{N} q_i H_2(p_i)\right| \leq \varepsilon n \ . \tag{10}$$

Finally, we should get rid of the side information; prediction strategies are not supposed to use any. There is a finite number (to be precise, N^N) of functions that map $\{1,2,\ldots,N\}$ into itself. Every $\sigma : \{1,2,\ldots,N\} \to \{1,2,\ldots,N\}$ defines a strategy that works as follows. The strategy runs $\mathfrak{A}^{\mathrm{q}}_\varepsilon$, and each time $\mathfrak{A}^{\mathrm{q}}_\varepsilon$ outputs γ_i our strategy outputs $\gamma_{\sigma(i)}$, $i = 1,2,\ldots,N$. For every finite sequence $\boldsymbol{x}$ there is a mapping σ such that the corresponding strategy works exactly like $\widetilde{\mathfrak{A}^1_\varepsilon}$.

Since the game $\mathfrak{G}_1$ is weakly mixable, we can obtain $\mathfrak{A}^1_\varepsilon$ that works nearly as good as each one from the final pool of strategies when the loss is measured w.r.t. the loss function λ_1. We get

$$\mathrm{Loss}^{\mathfrak{G}_1}_{\mathfrak{A}^1_\varepsilon}(\boldsymbol{x}) \leq n\sum_{i=1}^{N} q_i H_1(p_i) + \varepsilon n + f_1(n) \tag{11}$$

for every $\boldsymbol{x}$, where $f_1(n) = o(n)$ as $n \to +\infty$. Similarly, there is $\mathfrak{A}^2_\varepsilon$ such that

$$\mathrm{Loss}^{\mathfrak{G}_2}_{\mathfrak{A}^2_\varepsilon}(\boldsymbol{x}) \leq n\sum_{i=1}^{N} q_i H_2(p_i) + \varepsilon n + f_2(n) \tag{12}$$

for every $\boldsymbol{x}$, where $f_2(n) = o(n)$ as $n \to +\infty$. The lemma follows.

Appendix D. Filling in the Spaceship

We shall now construct languages $L \subseteq \Omega^\infty$ such that $\overline{\mathrm{AC}}_1(L) = \overline{\overline{\mathrm{AC}}}_1(L) = \underline{\mathrm{AC}}_1(L) = \underline{\underline{\mathrm{AC}}}_1(L)$ as well as $\overline{\mathrm{AC}}_2(L) = \overline{\overline{\mathrm{AC}}}_2(L) = \underline{\mathrm{AC}}_2(L) = \underline{\underline{\mathrm{AC}}}_2(L)$ and pairs $(\mathrm{AC}_1(L), \mathrm{AC}_2(L))$ fill the spaceship closure.

We start by constructing languages filling in the entropy set, then construct languages filling in the entropy hull and finally obtain languages filling in the spaceship closure. First let $u = H_1(p)$ and $v = H_2(p)$ for some $p \in \mathbb{P}(\Omega)$.

Lemma 6. *For every $p \in \mathbb{P}(\Omega)$ there is a set $L_p \subseteq \Omega^\infty$ such that for every game $\mathfrak{G} = \langle\Omega,\Gamma,\lambda\rangle$ we have $\overline{\mathrm{AC}}(L_p) = \overline{\overline{\mathrm{AC}}}(L_p) = \underline{\mathrm{AC}}(L_p) = \underline{\underline{\mathrm{AC}}}(L_p) = H(p)$.*

Proof (of the Lemma).

Let $p = (p_0, p_1, \ldots, p_{M-1}) \in \mathbb{P}(\Omega)$. If some of p_i are equal to 0, we can completely ignore those dimensions, or, in other words, consider the games with the sets of superpredictions that are the projection of the sets of superpredictions w.r.t. G_1 and $\mathfrak{G}_2$ to non-zero directions. So let us assume, without restricting the generality, that all $p_i \neq 0$.

Consider the set $\Xi_n^{(p)} \subseteq \Omega^n$ of sequences $\boldsymbol{x}$ of length n with the following property. For each $i = 0, 1, \ldots, M-1$ the number of $\omega^{(i)}$s among the elements of $\boldsymbol{x}$ is between the numbers $np_i - n^{3/4}$ and $np_i + n^{3/4}$, i.e., $np_i - n^{3/4} \le \sharp_i \boldsymbol{x} \le np_i + n^{3/4}$ for $i = 0, 1, \ldots, M-1$.

We need the Chernoff bound in Hoeffding's form (see Theorem 1 in [8]):

Proposition 1 (Chernoff bound). *If $\xi_1, \xi_2, \ldots, \xi_n$ are independent random variables with finite first and second moments and such that $0 \le \xi_i \le 1$ for all $i = 1, 2, \ldots, n$ then*

$$\Pr\{\overline{\xi} - \mu \ge t\} \le e^{-2nt^2} \quad ,$$

for all $t \in (0, 1-\mu)$, where $\overline{\xi} = (\xi_1 + \xi_2 + \ldots + \xi_n)/n$ and $\mu = \mathbf{E}\overline{\xi}$.

Let $\xi_1^{(p)}, \xi_2^{(p)}, \ldots, \xi_n^{(p)}$ be independent random variables that accept the values $\omega^{(i)}$ with probabilities p_i, $i = 0, 1, \ldots, M-1$. The Chernoff bound implies that

$$\Pr\left\{\left|\sharp_i(\xi_1^{(p)}\xi_2^{(p)}\ldots,\xi_n^{(p)}) - p_i n\right| \ge n^{3/4}\right\} \le 2e^{-2\sqrt{n}} \tag{13}$$

for all $n \ge N_0$ (the constant N_0 is necessary in order to ensure that $t \le 1-\mu$ in the bound) and all $i = 0, 1, \ldots, M-1$. If we denote by $\Pr_p(S)$ the probability that $\xi_1^{(p)}\xi_2^{(p)}\ldots,\xi_n^{(p)} \in S$, we get $\Pr_p(\Omega^n \setminus \Xi_n^{(p)}) \le 2Me^{-2\sqrt{n}}$.

Let $L_p = \Xi_{n_1}^{(p)} \times \Xi_{n_2}^{(p)} \times \Xi_{n_3}^{(p)} \ldots \times \Xi_{n_k}^{(p)} \times \ldots \subseteq \Omega^\infty$. In other terms, L_p consists of all infinite sequences $\boldsymbol{x}$ with the following property. For every non-negative integer k the elements of $\boldsymbol{x}$ from $\sum_{j=1}^{k-1} n_j$ to $\sum_{j=1}^{k} n_j$ form a sequence from $\Xi_{n_k}^{(p)}$. We will refer to these elements as the k-th segment of L_p. Take $n_j = N_0 + j$ so that $\sum_{j=1}^{k} n_j = kN_0 + k(k+1)/2$. We will show that L_p proves the lemma.

First let us prove that $\mathrm{AC}(L_p) \le H(p)$. Let $\mathfrak{A}^{(p)}$ be the strategy that always outputs the same prediction $\gamma^* \in \arg\min_{\gamma\in\Gamma} \sum_{i=0}^{M-1} p_i \lambda(\omega^{(i)}, \gamma)$. Let $n = \sum_{j=1}^{k} n_j$. There is a constant $C_{\gamma^*} > 0$ such that for every $\boldsymbol{x} \in L_p$ we get

$$\mathrm{Loss}_{\mathfrak{A}^{(p)}}(\boldsymbol{x}|_n) \le H(P)n + MC_{\gamma^*}\sum_{j=1}^{k} n_j^{3/4} \quad .$$

We have

$$\frac{\sum_{j=1}^{k} n_j^{3/4}}{n} = \frac{\sum_{j=1}^{k}(N_0+j)^{3/4}}{\sum_{j=1}^{k}(N_0+j)} \sim \frac{\frac{4}{7}k^{7/4}}{\frac{1}{2}k^2} \to 0$$

as $k \to +\infty$. Therefore $\underline{\mathrm{AC}}(L_p)$ and $\underline{\underline{\mathrm{AC}}}(L_p)$ do not exceed $H(p)$.

Now consider m such that $n - n_k = \sum_{j=1}^{k-1} n_j < m < \sum_{j=1}^{k} n_j = n$. It is easy to see that $|\,\mathrm{Loss}_{\mathfrak{A}^{(p)}}(\boldsymbol{x}|_n) - \mathrm{Loss}_{\mathfrak{A}^{(p)}}(\boldsymbol{x}|_m)| \le C_{\gamma^*} n_k$. We have

$$\frac{n_k}{m} \le \frac{(N_0+k)}{\sum_{j=1}^{k-1}(N_0+j)} = \frac{(N_0+k)}{N_0(k-1) + k(k-1)/2} \to 0 \tag{14}$$

and hence the upper complexities do not exceed $H(p)$ either.

Now let us prove that $\mathrm{AC}(L_p) \geq H(p)$. Consider a strategy $\mathfrak{A}$. First let us assume that λ is bounded and $D > 0$ is an upper bound on λ. We have

$$H(p)n \leq \mathbf{E}\,\mathrm{Loss}_{\mathfrak{A}}(\xi_1^{(p)}\xi_2^{(p)}\ldots,\xi_n^{(p)}) \leq \Pr_p(\Xi_n)\max_{\boldsymbol{x}\in\Xi_n}\mathrm{Loss}_{\mathfrak{A}}(\boldsymbol{x}) + \Pr_p(\Omega^n\setminus\Xi_n)Dn \ .$$

Therefore there is a sequence $\boldsymbol{x} \in \Xi_n$ such that

$$\mathrm{Loss}_{\mathfrak{A}}(\boldsymbol{x}) \geq H(p)n - \Pr_p(\Omega^n\setminus\Xi_n)Dn \geq H(p)n - 2nMDe^{-2\sqrt{n}}$$

provided $n \geq N_0$.

We construct $\boldsymbol{x} \in L_p$ from finite segments of length n_i. The series $\sum_{j=1}^{\infty}(N_0+j)e^{-2\sqrt{N_0+j}}$ converges and thus upper complexities are at least $H(p)$. We can extend this bound to lower complexity by using (14).

Now let λ be unbounded. Take $\lambda^{(D)} = \min(\lambda, D)$, where D is a constant. It is easy to see that for sufficiently large D we get $\min_{\gamma\in\Gamma}\sum_{i=0}^{M-1}p_i\lambda(\omega^{(i)},\gamma) = \min_{\gamma\in\Gamma}\sum_{i=0}^{M-1}p_i\lambda^{(D)}(\omega^{(i)},\gamma)$ (recall that $p \in \mathbb{P}(\Omega)$ is fixed).

Pick such D and let $\mathfrak{G}^{(D)}$ be the game with the loss function $\lambda^{(D)}$. It is easy to see that for every strategy $\mathfrak{A}$ and every sequence $\boldsymbol{x}$ we have $\mathrm{Loss}_{\mathfrak{A}}^{\mathfrak{G}}(\boldsymbol{x}) \geq \mathrm{Loss}_{\mathfrak{A}}^{\mathfrak{G}^{(D)}}(\boldsymbol{x})$. Since the loss function $\lambda^{(D)}$ is bounded, one can apply the above argument; for every $\mathfrak{A}$ and every n there is a sequence $\boldsymbol{x} \in L_p$ such that $\mathrm{Loss}_{\mathfrak{A}}^{\mathfrak{G}^{(D)}}(\boldsymbol{x}|_n) \geq H(p)n - o(n)$. The inequality implies $\mathrm{Loss}_{\mathfrak{A}}^{\mathfrak{G}}(\boldsymbol{x}|_n) \geq H(p)n - o(n)$. This proves the lemma. □

Secondly let (u,v) be some point from $\mathcal{H}$. The definition of convexity implies that there are probabilities $p_1,p_2,\ldots,p_N \in \mathbb{P}(\Omega)$ and weights $q_1,q_2,\ldots,q_N \in [0,1]$ such that $u = \sum_{j=1}^{N} q_jH_1(p_j)$ and $v = \sum_{j=1}^{N} q_jH_2(p_j)$.

Let us 'paint' all positive integers into N colours $1,2,\ldots,N$. Number 1 is painted colour 1. Suppose that all numbers from 1 to n have been painted and there are n_1 numbers among them painted colour 1, n_2 numbers painted colour 2 etc. The values $q_jn - n_j$ are *deficiencies*. Let j_0 corresponds to the largest deficiency (if there are several largest deficiencies, we take the one with the smallest j). We paint the number $n+1$ the colour j_0.

During the infinite construction process deficiencies never exceed N. Indeed, the value $-(q_jn - n_j)$ never exceeds 1 and the sum of all deficiencies is 0.

We now proceed to constructing $L \subseteq \Omega^\infty$ that corresponds to (u,v). The set L consists of all infinite sequences $\boldsymbol{x}$ with the following property. The subsequence of $\boldsymbol{x}$ formed by bits with numbers painted the colour j belongs to L_{p_j} from Lemma 6 for all $j = 1,2,\ldots,N$. One can easily check that L has all the required properties.

Thirdly let $(u,v) \in \mathcal{S}\setminus\mathcal{H}$. It is easy to check that if a game $\mathfrak{G}$ is weakly mixable, then for every pair of languages L_1, L_2 constructed above we have $\mathrm{AC}(L_1 \cup L_2) = \max(\mathrm{AC}(L_1), \mathrm{AC}(L_2))$.