Online Optimisation in Convexity Spaces

Thomas Gärtner[1,2] **Olana Missura**[1]

[1]University of Bonn and [2]Fraunhofer IAIS, Sankt Augustin, Germany
{thomas.gaertner, olana.missura}@uni-bonn.de

Abstract

We consider optimisation and online convex programming in convexity spaces and derive efficient algorithms for these problems. Convexity spaces [Kay and Womble, 1971] generalise the Euclidean notion of convexity to a much broader setting in which neither optimisation nor online convex programming have been considered so far. We show that, given a particular sampling oracle, a generalisation of the iterated Radon points algorithm [Clarkson et al., 1996] is able to approximate the maximum of any quasiconcave function in a convexity space with high probability. Applying this algorithm to find points that ensure large belief updates leads to an efficient online convex programming algorithm.

1 Introduction

We propose an algorithm for maximising quasiconcave functions in convexity spaces and derive an efficient online convex programming algorithm from it. Convexity spaces [Kay and Womble, 1971] generalise Euclidean convexity and include many discrete and combinatorial domains. While offline optimisation aims at finding a point with a large value of a fixed function, online convex programming can be seen as an iterative game between an algorithm and an adversary: In each iteration of this game, the adversary arbitrarily chooses a 'feasible' convex set, the algorithm queries a point, and if this point is infeasible then the adversary reveals a convex 'cutting' set, i.e. a convex region fully containing the feasible set but not the query point. Each infeasible query is considered a mistake and the goal is to make as few as possible mistakes.

Convexity spaces [Kay and Womble, 1971] consist of a domain and a family of its subsets that is closed under intersection. They can be formed in various ways and apart from Euclidean convexity, they include hyperrectangles, linear extensions of partially ordered sets, as well as shortest path or metric convexities. Quasiconcave functions in convexity spaces are defined as functions with convex superlevel sets. We introduce convexity spaces in more detail in Section 2 where we also characterise the size of the intersection of large convex sets by their Vapnik and Chervonenkis [1971] dimension.

Radon [1921] Theorem shows that any set of $d + 2$ points in $\mathbb{R}^d$ can be partitioned such that the convex hulls of the two partitions overlap. Any point in this overlap is called a Radon point. Considering such partitions in non-Euclidean convexity spaces leads to the definition of the Radon number as the smallest number for which such a partitioning exists for any set of that size in the convexity space. Clarkson et al. [1996] proposed an algorithm that iteratively replaces sets of points in $\mathbb{R}^d$ with their Radon point, the iterated Radon points algorithm . They showed that, given a finite set of points as an input, the output of this algorithm is a point of large Tukey [1975] depth of the original set. In Section 3 we show that a generalisation of this algorithm can also be used to approximate the maximum of a quasiconcave function in a convexity space with high probability, assuming only a particular sampling oracle.

To derive an online convex programming algorithm for non-Euclidean spaces, we follow the principle of maximising the information gained. This means, conceptually, to maintain a belief distribution over the convexity space and multiplicatively updating the belief outside the convex cutting sets. The amount of belief that can be updated, in the case that the adversary chooses the least informative cutting set, is a generalisation of Tukey [1975] depth to convexity spaces. In Section 4 we describe our approach to online convex programming in more detail and show that for any learning algorithm that picks sufficiently deep points in the belief distribution, the number of infeasible queries can be bounded by a function linear in the number of mistakes of the best static ϵ-environment.

2 Measurable Convexity Spaces

General interest in convex geometry has been raised by Danzer et al. [1963] who also gave an overview over different possible generalizations of the geometric notion of convexity. The generalisation which we use in this paper and which seems to be the most popular one to date, follows along the lines of Kay and Womble [1971] and goes back to Levi [1951, cf Eckhoff [1968]]. It considers any intersection closed family of sets as a convexity structure.

In the following, we denote by $\mathcal{P}(\cdot)$ the power set and for any family of sets $\mathcal{F} \subseteq \mathcal{P}(\mathcal{X})$ from a domain $\mathcal{X}$, we denote $\bigcup \mathcal{F} = \bigcup_{F\in\mathcal{F}} F$, $\bigcap \mathcal{F} = \bigcap_{F\in\mathcal{F}} F$, and $\bigcap \emptyset = \mathcal{X}$.

Definition 1. *A* set system *over a domain $\mathcal{X}$ is any collection of subsets of $\mathcal{X}$. A* convexity structure *$\mathcal{C}$ over a set $\mathcal{X}$ is a collection of subsets of $\mathcal{X}$ such that $\emptyset \in \mathcal{C}$, $\mathcal{X} \in \mathcal{C}$, and $\forall \mathcal{F} \subseteq \mathcal{C} : \bigcap \mathcal{F} \in \mathcal{C}$. A* measurable convexity space *is a triple $(\mathcal{X}, \mathcal{C}, \mu)$ where $\mathcal{C}$ is a convexity structure and $\mu : \Sigma \to \mathbb{R}$ is a measure defined on a suitable σ-algebra $\Sigma \supseteq \mathcal{C}$. The* convex hull *operator corresponding to $\mathcal{C}$ is denoted by $\langle\cdot\rangle : \mathcal{P}(\mathcal{X}) \to \mathcal{C}$ and defined as $\langle S\rangle = \bigcap\{C \in \mathcal{C} \mid S \subseteq C\}$.*

In literature, set systems are often also called *hypergraphs* or *range spaces*, convexity structures are often also called *closure systems*. We prefer the term convexity structure to emphasise the relation to Euclidean convexity and the importance for our analyses of some concepts borrowed from Euclidean convexity. We will often refer to a measurable convexity space just as a *convexity space*.

It is well known that the above convex hull operator satisfies all usual properties of a *closure operator* and that vice versa the closed sets of every closure operator form a convexity structure.

Definition 2. *A* closure operator *is a function $\langle\cdot\rangle : \mathcal{P}(\mathcal{X}) \to \mathcal{P}(\mathcal{X})$ such that for all $A, B \subseteq \mathcal{X}$ it holds that $A \subseteq \langle A\rangle$, $A \subseteq B \Rightarrow \langle A\rangle \subseteq \langle B\rangle$, and $\langle A\rangle = \langle\langle A\rangle\rangle$.*

In the remainder of this paper, unless explicitly stated, $(\mathcal{X}, \mathcal{C}, \mu)$ will refer to one and the same fixed (but otherwise arbitrary) measurable convexity space and $\langle\cdot\rangle$ will always be the associated closure/convex hull operator. Furthermore, for any set $X \subseteq \mathcal{X}$ we use the notation $\bar{X}$ to denote its complement in $\mathcal{X}$, i.e., $\bar{X} = \{x \in \mathcal{X} \mid x \notin X\}$.

Our first result concerns the measure of the intersection of convex sets in convexity spaces with finite Vapnik and Chervonenkis [1971] dimension.

Definition 3. *The* VC-dimension *of a set system $(\mathcal{X}, \mathcal{C})$ is the largest integer $\mathbf{v}$ such that there is a set $X \subseteq \mathcal{X}$ with $|X| = \mathbf{v}$ which can be* shattered*, i.e.,*

$$\forall Z \subset X\ \exists C \in \mathcal{C} :\ X \cap C = Z\ .$$

For convexity structures $\mathcal{C}$ the above condition can be written as $\forall Z \subseteq X : \langle Z\rangle \cap X = Z$.

The following theorem gives a lower bound on the measure of the intersection of convex sets that contain a large fraction of the total measure:

Theorem 1. *In any measurable convexity space $(\mathcal{X}, \mathcal{C}, \mu)$ with VC-dimension less than or equal to $\mathbf{v}$, it holds for all $\theta \in \mathbb{R}$ that*

$$\mu\left[\bigcap\{C \in \mathcal{C} \mid \mu(C) > \theta\}\right] > \mu(\mathcal{X}) - \mathbf{v}(\mu(\mathcal{X}) - \theta)\ .$$

This theorem can be shown by making use of a characterisation of the VC dimension as the smallest integer $\mathbf{v}$ for which $\forall \mathcal{F} \subseteq \mathcal{C}\ \exists \mathcal{G} \subseteq \mathcal{F} : \bigcap \mathcal{G} = \bigcap \mathcal{F}\ \wedge\ |\mathcal{G}| \leq \mathbf{v}$.

3 Radon Iterations in Convexity Spaces

In this section we investigate *quasiconcave functions* on measurable convexity spaces with finite *Radon [1921] number*. The main result of this section concerns the iterated Radon points algorithm

[Clarkson et al., 1996]. After giving necessary definitions and introducing the algorithm, we will show that with high probability the output of the algorithm is approximating the maximiser of any quasiconcave function under some assumptions on the sampling distribution.

Definition 4. *A* Radon partition *of a set S in a convexity space is a pair $A, B \subset S$ such that $A \cap B = \emptyset$ but $\langle A \rangle \cap \langle B \rangle \neq \emptyset$. A* Radon point *of a set S is any $r \in \mathcal{X}$ such that there is a Radon partition A, B of S and $r \in \langle A \rangle \cap \langle B \rangle$. The* Radon number *of a convexity structure is the smallest $\mathbf{r} \in \mathbb{N} \cup \{\infty\}$ such that for all $S \subseteq \mathcal{X}$ it holds that $|S| \geq \mathbf{r}$ implies the existence of a Radon partition.*

It holds that the Radon number of a convexity space is never larger than the VC-dimension plus one.

From now on we will denote the Radon number of the convexity structure $\mathcal{C}$ by $\mathbf{r}$ and consider only convexity structures with finite Radon number. The **iterated Radon points algorithm on measurable convexity spaces with finite Radon number** proceeds as follows: Let P be a probability distribution on $\mathcal{X}$ and let $\mathcal{T}$ be a perfect $\mathbf{r}$-ary tree, i.e., all leaves of $\mathcal{T}$ have the same distance from the root and all nodes other than the leaves have $\mathbf{r}$ children. For any vertex v of the tree, let $\delta(v)$ denote its children. The iterated Radon points algorithm traverses the tree in a bottom-up manner and selects a point from $\mathcal{X}$ at each node v which we will denote by $I(v)$. In particular, for all leaves $I(v)$ is sampled iid from P and for all internal nodes $I(v)$ is a Radon point of its children $\delta(v)$. By symmetry, the distribution over $I(v)$ for all vertices v of the same height is identical and we will denote the random variable representing vertices of height h by $\mathcal{T}_h$. The output of the algorithm is the point that was assigned to the root of $\mathcal{T}$.

Next we show that the iterated Radon points algorithm approximates a maximiser of a quasiconcave function with high probability.

Definition 5. *A function $f : \mathcal{X} \to \mathbb{R}$ is called* quasiconcave *with respect to a convexity structure $\mathcal{C}$ if and only if all its superlevel sets are convex, i.e.,*

$$\forall \theta \in \mathbb{R} : \ \{x \in \mathcal{X} \mid f(x) > \theta\} \in \mathcal{C} \ .$$

Note that 1) a minimum of a quasiconcave function over the convex hull of a set of points in a convexity space can be found in the set itself; 2) for any Radon point r of a set and any quasiconcave function f, there are two points a, b in the set such that $f(a) \leq f(r)$ and $f(b) \leq f(r)$. A simple application of the union bound is then sufficient to show the basic inequality $P\left[f(Y) \leq \theta\right] \leq \left(\mathbf{r}\ P[f(X) \leq \theta]\right)^2$ where X is a sample from a fixed distribution over the convexity space and Y is the Radon point of a set of size $\mathbf{r}$ sampled iid from that same distribution. The iterative application of the basic inequality leads to the following theorem:

Theorem 2. *Let $(\mathcal{X}, \mathcal{C}, P)$ be a measurable convexity space with probability measure P and finite Radon number $\mathbf{r}$ and denote by X the random variable representing any point drawn according to P. Let furthermore $\mathcal{T}_h$ denote the random variable representing the output of the iterated Radon points algorithm at height h with leaves drawn according to P. Then for any quasiconcave function $f : \mathcal{X} \to \mathbb{R}$ and all $\theta \in \mathbb{R}$ it holds that*

$$P\left[f(\mathcal{T}_h) \leq \theta\right] \leq \left(\mathbf{r}\ P[f(X) \leq \theta]\right)^{2^h} \ .$$

Assuming there is a sampling oracle such that the probability of sampling a point x with $f(x) \leq \theta$ is strictly upper bounded by $1/2\mathbf{r}$, the theorem implies that the iterated Radon points algorithm returns a point which—with high probability—has a constant approximation rate to the maximiser of a quasiconcave function.

Having generalised the iterated Radon points algorithm to measurable convexity spaces, we will show in the next section how it can be used as a part of an online learning algorithm with bounded number of mistakes.

4 Maximising Information Gained

The online convex programming setting we consider takes place in a *measurable convexity space* $(\mathcal{X}, \mathcal{C}, \mu)$. **We consider the following iterative game**: In each round t the environment or adversary fixes a convex set $C_t^* \in \mathcal{C}$ of 'feasible' points and the learner queries a point $x_t \in \mathcal{X}$. The aim of the learner is to 'hit' $x_t \in C_t^*$. For each infeasible query point, the learner receives as a feedback a convex 'cutting' set $C_t \supseteq C_t^*$ such that $x_t \notin C_t$.

We aim at bounding the number of mistakes of the learner against the number of mistakes of the best static ϵ-environment: The number of mistakes of the learner until time T is the number of infeasible queries $m_T = |\{1 \leq t \leq T \mid x_t \notin C_t^*\}|$. For any $\epsilon \in \mathbb{R}^+$, we define the set of ϵ-environments as $\mathcal{B}^\epsilon = \{C \in \mathcal{C} \mid \mu(C) \geq \epsilon\}$. We denote the number of mistakes of any point $x \in \mathcal{X}$ by $M_T(x) = |\{1 \leq t \leq T \mid x \notin C_t^*\}|$, the mistakes of a convex set $C \in \mathcal{C}$ by $M_T^C = \sup_{x \in C} M_T(x)$, and the ϵ-BSIH (best static in hindsight) mistake for any $\epsilon \in \mathbb{R}^+$ by $M_T^\epsilon = \inf_{C \in \mathcal{B}^\epsilon} M_T^C$.

The general online learning strategy which we consider maintains a belief over the measurable convexity space and chooses each query such that the information that is gained in the worst case over all possible replies of the adversary in this round can be bounded from below by a constant fraction of the current belief.

In particular, we consider the following multiplicative belief sequence:

Definition 6. *For any $\beta \in (0,1)$, the* β-belief sequence *corresponding to a sequence of convex sets $C_1, C_2, \ldots \in \mathcal{C}$ in a measurable convexity space $(\mathcal{X}, \mathcal{C}, \mu)$ is the sequence of measures $\mu_0, \mu_1, \ldots$ with $\mu_0 = \mu$ and*

$$\mu_t(F) = \mu_{t-1}(F \cap C_t) + \beta\mu_{t-1}(F \setminus C_t)$$

where F is any measurable set and $t \in \{1, 2, \ldots\}$.

The least amount of belief that can be updated after a particular query is a generalisation of the *halfspace depth*, centrality, or Tukey [1975] depth in a distribution.

Definition 7. *The* depth *of a point x with respect to measure μ in a measurable convexity space is*

$$d_\mu(x) = \inf\{\mu(\bar{C}) \mid C \in \mathcal{C} \wedge x \notin C\} .$$

We note that depth is quasiconcave, as its superlevel sets are intersections of convex sets.

To bound the number of mistakes of an online learning algorithm in this setting, we rely on the sequence of queries to be 'trustworthy' with respect to the β-belief sequence:

Definition 8. *For any $\beta, \gamma \in (0,1)$, a (β, γ)*-trustworthy online learning algorithm *in a measurable convexity space $(\mathcal{X}, \mathcal{C}, \mu)$, is an online learning algorithm for which the sequence of queries $x_0, x_1, \ldots$ is such that each query has depth at least $d_{\mu_t}(x_t) \geq \gamma\mu_t(\mathcal{X})$ in the corresponding element of the β-belief sequence. Such points x_t are called γ-centers.*

We are now ready to bound the number of mistakes of trustworthy online learners as a function of the best static ϵ-environment (chosen in hindsight):

Theorem 3. *The number of mistakes m_T until time T of any (β, γ)*-trustworthy online learner *in a measurable convexity space $(\mathcal{X}, \mathcal{C}, \mu)$ is bounded by a linear function of the mistakes M_T^ϵ of the best static ϵ-environment:*

$$m_T \leq \frac{M_T^\epsilon \ln 1/\beta + \ln 1/\epsilon + \ln \mu(\mathcal{X})}{\ln 1/(1 - \gamma + \beta\gamma)} .$$

An obvious strategy for online learning in measurable convexity spaces is now the *maximising information gained* (MIG) algorithm: On each iteration query a γ-center in the corresponding element of the β-belief sequence. Whenever such a point can be found efficiently, the above bound on the number of mistakes naturally holds. Since the depth is a quasiconcave function, the generalised iterated Radon points algorithm from the previous section can be applied to find points of sufficient depths with high probability, provided there is a suitable sampling oracle.

5 Related Work

Most previous work on online learning has been concentrating on the case where the problems' domain is finite, i.e. there are K actions, arms, experts, etc, and on each round the learner chooses one (or several) of them. The traditional measure of the learner's performance is the so-called regret: the difference between the loss of the best static choice in hindsight and of the learner (the expected regret in case of randomized algorithms). The overall goal is to create a learner that achieves a regret sublinear in t. Various bounds on the regret were proven for various assumptions on the adversary's (environment's) behaviour, for more information we refer the reader to the excellent book by Cesa-Bianchi and Lugosi [2006].

The stochastic continuum-armed bandit problem was introduced by Agrawal [1995], who considered a one-dimensional interval of arms and a continuous loss function. Subsequently, the setting has been generalized to Euclidean spaces and has since been termed 'online convex programming' or 'online linear programming' due to its relationship to minimising a fixed convex or linear function, respectively. Most approaches for both stochastic and adversarial settings are in this case based on using gradient information from the loss function (in the full-information setting) or estimating the gradient (in the case of bandit feedback). Most of the research in this area not concerned with linear loss functions is based on assuming either global or local smoothness of the payoff function, two examples of such assumptions are Lipschitz smoothness and Hölder continuity. For a detailed overview, the reader is referred to the outstanding survey of Shalev-Shwartz [2012].

A generalization of the Euclidean space setting can be found in Bubeck et al. [2011]. The authors present their results for a stochastic bandit setting in a generic measurable space, while making the assumption that the loss function is locally Lipschitz continuous with respect to a dissimilarity function known to the learner. A different generalization of online learning concerns a case where the learning is happening on a graph. One particular variant of it restricts the adversary to assign a labeling to the vertices of a graph in advance. Then the learner is presented with one vertex at a time (in the order decided by the adversary as well) and is requested to predict its label. The learner's goal is to minimize the amount of mistakes it makes while labelling the graph. Various algorithms and bounds exploiting the structure of the graph were designed and proven for this setting (for details see Herbster et al. [2008b,a, 2005], Vitale et al. [2013], Fakcharoenphol and Kijsirikul [2008]).

Missura and Gärtner [2011] present a case of online learning on a poset, where no constraints are placed on the labelling (it can change in an adversarial way), but it has to respect the partial order. In order to leverage the information encoded in the partial order of arms, in addition to the loss value incurred by the played arm, the order relation of the played arm to the minimum-loss arms is revealed. This allows the agent to infer the direction of the minimum-loss arms. The authors introduce an algorithm that in each round plays the arm that allows to gain a maximal amount of information about the set of arms of minimum loss.

Cesa-Bianchi and Lugosi [2012] consider combinatorial bandits, which are a special case of bandit linear optimization where the problem domain S is restricted to be a subset of the binary hypercube $\{0,1\}^d$. Casting different online learning problems, such as perfect matchings [Helmbold and Warmuth, 2009] or the online version of the travelling salesman problem, into this setting and exploiting its combinatorial structure allows for better regret bounds in some particular cases.

To the best of our knowledge, we are the first to consider online learning in general, non-Euclidean convexity spaces.

6 Conclusion

We considered the problem of optimising a quasiconcave function in a measurable convexity space. In this setting we generalised the iterated Radon points algorithm and showed that, given a sampling oracle, the iterated Radon points algorithm approximates a maximiser of a quasiconcave function with high probability. We proposed the generalisation of adversarial online learning to measurable convexity spaces. Since the depth of a distribution is a quasiconcave function, the iterated Radon points algorithm can be used to produce points with high probability guarantees on their depth. Based on this and using the principle of maximising the information gained upon each mistake, we proposed an algorithm that, in each iteration, queries a point of sufficiently large depth.

By applying Theorems 1 and 2 to the quasiconcave depth function, we can characterise the depth of the output of the iterated Radon points algorithm as:

Corollary 1. *Let $(\mathcal{X}, \mathcal{C}, P)$ be a measurable convexity space with probability measure P, finite Radon number $\mathbf{r}$, and finite VC-dimension $\mathbf{v}$. It holds that the iterated Radon points algorithm with height h and thus $L = 2^h$ leaves returns a point of depth at least $1/(2\mathbf{r}\mathbf{v})$ with respect to P with probability at least $1 - 2^{-L}$.*

Note, that the runtime of the iterated Radon points algorithm is linear in the time needed to sample a point from P, the time needed to find a Radon point of a set of size $\mathbf{r}$, and the number of leaves L. We also note that for a convexity structure $\mathcal{C}$ with constant, finite VC-dimension $\mathbf{v}$ and assuming access to an oracle for sampling according to μ from any convex set $C \in \mathcal{C}$, samples according to any belief μ_T from a β-belief sequence can be produced in deterministic time polynomial in T without explicitly updating the belief.

Combining Corollary 1 and Theorem 3 leads to the following mistake bound:

Corollary 2. *Consider online learning in a measurable convexity space $(\mathcal{X}, \mathcal{C}, \mu)$ with Radon number $\mathbf{r}$ and VC dimension $\mathbf{v}$. For any constant l, the number of mistakes m_T until time T of the maximising information gained by iterated Radon points algorithm with learning rate β and $L = 2^h \geq l + \log_2 T$ leaves is bounded with probability at least $1 - 2^{-l}$ by a linear function of the mistakes M_T^ϵ of the best static ϵ-environment:*

$$m_T \leq \frac{M_T^\epsilon \ln 1/\beta + \ln 1/\epsilon + \ln \mu(\mathcal{X})}{\ln 2\mathbf{rv}/(2\mathbf{rv} - 1 + \beta))} .$$

Acknowledgements: Part of this work was supported by the German Science Foundation (DFG) under the reference number 'GA 1615/1-1'.

References

R. Agrawal. The continuum-armed bandit problem. *SIAM Journal on Control and Optimization*, 33, 1995.

S. Bubeck, R. Munos, G. Stoltz, and C. Szepesvári. X-Armed Bandits. *JMLR*, 12, 2011.

N. Cesa-Bianchi and G. Lugosi. Combinatorial bandits. *Journal of Computer and System Sciences*, 78, 2012.

Nicolò Cesa-Bianchi and Gábor Lugosi. *Prediction, learning, and games*. 2006.

K. L. Clarkson, D. Eppstein, G. L. Miller, C. Sturtivant, and S.-H. Teng. Approximating center points with iterative Radon points. *International Journal of Computational Geometry & Applications*, 6, 1996.

Ludwig Danzer, Branko Grünbaum, and Victor Klee. *Helly's theorem and its relatives*. 1963.

Jürgen Eckhoff. Der Satz von Radon in konvexen Produktstrukturen I. *Monatshefte für Mathematik*, 72, 1968.

Jittat Fakcharoenphol and Boonserm Kijsirikul. Low congestion online routing and an improved mistake bound for online prediction of graph labeling. *CoRR*, abs/0809.2075, 2008.

D. P. Helmbold and M. K. Warmuth. Learning Permutations with Exponential Weights. *JMLR*, 10, 2009.

M. Herbster, M. Pontil, and L. Wainer. Online learning over graphs. In *ICML*, 2005.

M. Herbster, G. Lever, and M. Pontil. Online prediction on large diameter graphs. In *NIPS*, 2008a.

M. Herbster, M. Pontil, and S. R. Galeano. Fast prediction on a tree. In *NIPS*, 2008b.

David C. Kay and Eugene W. Womble. Axiomatic convexity theory and relationships between the Carathéodory, Helly, and Radon numbers. *Pacific Journal of Mathematics*, 38, 1971.

Friedrich W Levi. On Helly's theorem and the axioms of convexity. *J. Indian Math. Soc*, 15, 1951.

Olana Missura and Thomas Gärtner. Predicting Dynamic Difficulty. In *NIPS 24*, 2011.

J. Radon. Mengen konvexer Körper, die einen gemeinsamen Punkt enthalten. *Mathematische Annalen*, 1921.

Shai Shalev-Shwartz. Online learning and online convex optimization. *Found. Trends Mach. Learn.*, 4, 2012.

John W Tukey. Mathematics and the picturing of data. In *Proceedings of the international congress of mathematicians*, volume 2, 1975.

Vladimir N. Vapnik and Alexey Chervonenkis. On the uniform convergence of relative frequencies of events to their probabilities. *Theory of Probability & Its Applications*, 16, 1971.

F. Vitale, N. Cesa-Bianchi, C. Gentile, and G. Zappella. Random Spanning Trees and the Prediction of Weighted Graphs. *JMLR*, 14, 2013.