4.4 Shallow Packings

"The highest activity a human being can attain is learning for understanding, because to understand is to be free."

Baruch Spinoza

As observed earlier, set systems $(X, \mathcal{F})$ of VC-dimension d behave, in many respects, like half-spaces or balls in $\mathbb{R}^d$. For example, the number of sets in $\mathcal{F}$ is $O(|X|)^d$. On the other hand, given a set P of points in $\mathbb{R}^d$, the number of subsets of P induced by half-spaces is $O\left(|P|^d\right)$. This section continues this correspondence through the lens of packing.

Given a set of P of n points in $[0, n]^d$, if the distance between every pair of points of P is at least δ, then $|P| = O\left(\left(\frac{n}{\delta}\right)^d\right)$. The analogous abstract case then becomes:

Theorem 4.7. *Given positive integers d and δ, let $\mathcal{F} = \{S_1, \ldots, S_m\}$ be a set system on a set X of n elements, with $\text{VC-dim}(\mathcal{F}) \leq d$. Further assume that for every $1 \leq i < j \leq m$, we have $\left|\Delta(S_i, S_j)\right| \geq \delta$. Then $|\mathcal{F}| = O\left((n/\delta)^d\right)$.*

Proof. From Lemma 4.3, we have

$$|\mathcal{F}| \leq 2 \cdot |\mathcal{F}|_A| , \quad \text{where } A \subseteq X \text{ is a uniform random sample of size at most } \frac{4dn}{\delta}.$$

The Primal Shatter Lemma 3.1 gives an upper-bound on $|\mathcal{F}|_A|$, in fact independent of the specific choice of A. Thus

$$|\mathcal{F}| \leq 2 \cdot |\mathcal{F}|_A| = O\left(\left(\frac{e|A|}{d}\right)^d\right) = O\left(\left(\frac{4en}{\delta}\right)^d\right).$$

□

⋆ ⋆ ⋆

Next, we present a packing statement for the case where each set of $\mathcal{F}$ further has size at most k, for some integer k. Unsurprisingly, the number of sets in the packing $\mathcal{F}$ can then be upper-bounded by the number of sets in the projection $\mathcal{F}|_A$ of a certain size.

Recall the definition of shallow-cell complexity of a set system:

Definition 3.1. *A set system $(X, \mathcal{F})$ has shallow-cell complexity $\varphi(\cdot, \cdot)$ if for any integer k and any subset $Y \subseteq X$, the number of sets in $\mathcal{F}|_Y$ of size at most k is at most $|Y| \cdot \varphi(|Y|, k)$.*

Theorem 4.8. *Given positive integers k, d and δ, let $\mathcal{F} = \{S_1, \ldots, S_m\}$ be a set system on a set X of n elements, and with shallow-cell complexity $\varphi(\cdot, \cdot)$ and $\text{VC-dim}(\mathcal{F}) \leq d$. Further assume that*

1. *$|S_i| \leq k$ for all $i = 1 \ldots m$, and*
2. *$|\Delta(S_i, S_j)| \geq \delta$ for all $1 \leq i < j \leq m$.*

Then $|\mathcal{F}| \leq \frac{24dn}{\delta} \cdot \varphi\left(\frac{4dn}{\delta}, \frac{12dk}{\delta}\right)$.

Intuition. Pick each element $p \in X$ into a sample R with probability $p = \frac{1}{\delta}$. Let

$$\mathcal{F}|_R = \{S_i' = S_i \cap R \colon S_i \in \mathcal{F}\}$$

be the projection of $\mathcal{F}$ onto R. We have $\mathrm{E}\left[|R|\right] = \frac{n}{\delta}$, and for each $S_i \in \mathcal{F}$,

$$\mathrm{E}\left[|S_i'|\right] = \mathrm{E}\left[|S_i \cap R|\right] = \frac{|S_i|}{\delta} \leq \frac{k}{\delta}.$$

Furthermore, for any two indices i and j:

$$\mathrm{E}\left[|\Delta(S_i', S_j')|\right] = \frac{|\Delta(S_i, S_j)|}{\delta} \geq 1.$$

So, in expectation, $\mathcal{F}|_R$ consists of sets of size at most $\frac{k}{\delta}$ over a set of $|R| = \frac{n}{\delta}$ elements. As the symmetric difference between any pair of sets in $\mathcal{F}|_R$ is at least one, the sets S_i' are distinct and thus

$$|\mathcal{F}| = |\mathcal{F}|_R| \leq \frac{n}{\delta} \cdot \varphi\left(\frac{n}{\delta}, \frac{k}{\delta}\right).$$

We next give the formal proof.

⋆ ⋆ ⋆

Proof. Let $A \subseteq X$ be a uniform random sample of size $\frac{4dn}{\delta} - 1$. Note that $\mathrm{E}\left[|S \cap A|\right] \leq \frac{4dk}{\delta}$ as $|S| \leq k$ for all $S \in \mathcal{F}$. Define

$$\mathcal{F}_1 = \left\{S \in \mathcal{F} \colon |S \cap A| > 3 \cdot \frac{4dk}{\delta}\right\}.$$

By Markov's inequality, for any $S \in \mathcal{F}$,

$$\Pr\left[S \in \mathcal{F}_1\right] = \Pr\left[|S \cap A| > 3 \cdot \frac{4dk}{\delta}\right] \leq \frac{1}{3}.$$

Thus

$$\begin{aligned}
\mathrm{E}\left[|\mathcal{F}|_A|\right] &\leq \mathrm{E}\left[|\mathcal{F}_1|\right] + \mathrm{E}\left[|(\mathcal{F} \setminus \mathcal{F}_1)|_A|\right] \\
&\leq \sum_{S \in \mathcal{F}} \Pr\left[S \in \mathcal{F}_1\right] + |A| \cdot \varphi\left(|A|, 3 \cdot \frac{4dk}{\delta}\right) \\
&\leq \frac{|\mathcal{F}|}{3} + \frac{4dn}{\delta} \cdot \varphi\left(\frac{4dn}{\delta}, \frac{12dk}{\delta}\right),
\end{aligned}$$

where the projection size of $\mathcal{F} \setminus \mathcal{F}_1$ to A is bounded by $\varphi(\cdot, \cdot)$. Now the bound follows from applying Lemma 4.3:

$$|\mathcal{F}| \leq 2 \cdot \mathrm{E}\left[|\mathcal{F}|_A|\right] \leq 2\left(\frac{|\mathcal{F}|}{3} + \frac{4dn}{\delta} \cdot \varphi\left(\frac{4dn}{\delta}, \frac{12dk}{\delta}\right)\right)$$

$$\implies |\mathcal{F}| \leq 6 \cdot \left(\frac{4dn}{\delta} \cdot \varphi\left(\frac{4dn}{\delta}, \frac{12dk}{\delta}\right)\right).$$

□

Bibliography and discussion. The shallow packing lemma for some geometric set systems was first shown in [1]. The statement was then generalized, and the proof simplified, in [2], whose presentation we have essentially followed here. The technical trick in the proof, at first somewhat counter-intuitive, is to upper-bound the desired quantity, $|\mathcal{F}|$, by a function that involves $|\mathcal{F}|$ itself. In the computer science literature, this is sometimes called boot-strapping.

[1] K. Dutta, E. Ezra, and A. Ghosh. Two proofs for shallow packings. *Discrete & Computational Geometry*, 56(4):910–939, 2016.

[2] N. H. Mustafa. A simple proof of the shallow packing lemma. *Discrete & Computational Geometry*, 55(3):739–743, 2016.

Chapter 7

Epsilon Approximations

Definition 7.1. *Given a set system $(X, \mathcal{F})$, an ϵ-approximation is a subset $A \subseteq X$ such that for any set $S \in \mathcal{F}$, we have*

$$\left| \frac{|S|}{|X|} - \frac{|S \cap A|}{|A|} \right| \leq \epsilon.$$

7.1 Epsilon-Approximations via Halving

The main theorem we will prove in this section is:

Theorem 7.1. *Let $(X, \mathcal{F})$ be a set system, $|X| = n$, with the property that*

$$\text{for any } Y \subseteq X, \qquad |\mathcal{F}|_Y| = O\left(|Y|^d\right).$$

Then there exists an ϵ-approximation for $\mathcal{F}$ of size $O\left(\frac{d}{\epsilon^2} \log \frac{d}{\epsilon}\right)$.

Basic idea. Imagine that we could pick a set $R_1 \subseteq X$ such that $|R_1| = \frac{n}{2}$ and R_1 is also 'equally representative' of all the other sets in $\mathcal{F}$. Namely, R_1 contains exactly half of the elements of each set of $\mathcal{F}$:

$$\text{for all } S \in \mathcal{F}\text{:} \qquad |R_1 \cap S| = \frac{|S|}{2}.$$

Repeating the above step, say that we could pick $R_2 \subseteq R_1$, with $\frac{|R_1|}{2} = \frac{n}{4}$ elements, such that $|R_2 \cap S| = \frac{|R_1 \cap S|}{2} = \frac{|S|}{4}$. See Figure 7.1. Continuing in this manner, let R_i be the set at the i-th iteration. Then we have

$$|R_i| = \frac{n}{2^i}, \qquad |R_i \cap S| = \frac{|S|}{2^i}, \quad \text{for all } S \in \mathcal{F}.$$

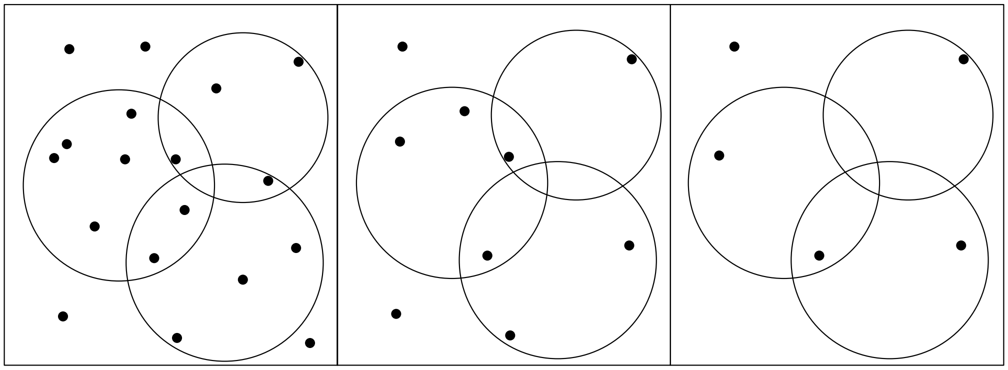

Figure 7.1: At each step, the total number of points halve; ideally, the number of points within *each* set—disks in this case—is also halved.

Then for any $S \in \mathcal{F}$, the proportion of points of S in any R_i will be equal to the proportion of points of S in X, namely

$$\left| \frac{|S|}{|X|} - \frac{|S \cap R_i|}{|R_i|} \right| = \left| \frac{|S|}{n} - \frac{|S|/2^i}{n/2^i} \right| = 0.$$

The problem is that, even at the first step, a set $R_1 \subseteq X$ that contains $\frac{|X|}{2}$ points, and also exactly halves *every* set is not always possible. The accuracy of R_1 will have to depend on the number of sets in $\mathcal{F}$. For example, for the complete set system $\mathcal{F} = 2^X$, for any choice R_1 of $\frac{n}{2}$ elements, there will be a set of $\frac{n}{2}$ elements in $\mathcal{F}$, namely the set $X \setminus R_1$, that will not share *any* points of R_1.

Thus one can only ask for the best R_1, $|R_1| = \frac{n}{2}$, such that for all $S \in \mathcal{F}$, $|R_1 \cap S|$ is as close as possible to $\frac{|S|}{2}$, although necessarily as a function of n and $|\mathcal{F}|$. Not surprisingly, we will pick R_1 randomly, by adding each element $p \in X$ to R_1 with probability $\frac{1}{2}$. First, note that the expected number of elements picked into R_1 is $\frac{n}{2}$. Not only that, but the expected number of elements picked from any *fixed* $S \in \mathcal{F}$ is also $\frac{|S|}{2}$. Of course that is unlikely to be true for all sets *simultaneously*[†].

$\star \quad \star \quad \star$

The key to proving our main theorem will be a halving lemma, which states the existence of a set R_1, $|R_1| = \frac{n}{2}$ such that:

$$\text{for all } S \in \mathcal{F}: \qquad |R_1 \cap S| = \frac{|S|}{2} \pm O\left(\sqrt{|X| \ln |\mathcal{F}|}\right).$$

[†]Take the complete set system example again.

Set $R_0 = X$, and using the halving lemma, we iteratively construct R_i from R_{i-1}, for $i \geq 1$. However, unlike earlier, errors accumulate with each iteration. As $R_i = \frac{n}{2^i}$, after i steps, a calculation shows that, for each $S \in \mathcal{F}$, we have

$$|R_i \cap S| = \frac{|S|}{2^i} \pm \sum_{j=0}^{i-1} \frac{O\left(\sqrt{|R_j| \ln |R_j|^d}\right)}{2^{(i-1)-j}} = \frac{|S|}{2^i} \pm \frac{1}{2^{(i-1)}} \sum_{j=0}^{i-1} 2^j \cdot O\left(\sqrt{|R_j| \ln |R_j|^d}\right).$$

The key here is in bounding the growth of the sum of errors. It turns out that the term 2^j dominates the error term, and so a calculation will show that the terms increase by a constant factor at each iteration. This gives an increasing geometric series, and so we have

$$|R_i \cap S| = \frac{|S|}{2^i} \pm O\left(\sqrt{|R_i| \ln |R_i|^d}\right).$$

Going to back ϵ-approximations, as $|R_i| = \frac{|X|}{2^i}$, we have

$$\left|\frac{|S|}{2^i} - |S \cap R_i|\right| = O\left(\sqrt{|R_i| \ln |R_i|^d}\right) \quad \underset{\text{dividing by } |R_i|}{\Longrightarrow} \quad \left|\frac{|S|}{|X|} - \frac{|S \cap R_i|}{|R_i|}\right| = O\left(\sqrt{\frac{\ln |R_i|^d}{|R_i|}}\right).$$

We now choose i to get this error to be at most ϵ. The reader can verify that if we set $i = \log\left(\frac{\epsilon^2 n}{d \log \frac{d}{\epsilon}}\right)$, R_i will be an ϵ-approximation. The size of R_i is then $\frac{|X|}{2^i} = O\left(\frac{d}{\epsilon^2} \log \frac{d}{\epsilon}\right)$.

⋆ ⋆ ⋆

Proof of Main Theorem

We now turn to the formal proof of the main theorem. The precise statement of the halving claim is the following.

Lemma 7.1 (Halving lemma). *Given a set system $(X, \mathcal{F})$, $X = \{p_1, \ldots, p_n\}$, $|\mathcal{F}| = m$, there exists a set $R_1 \subseteq X$, $|R_1| = \frac{n}{2}$, such that*

$$\textit{for all } S \in \mathcal{F}, \qquad |R_1 \cap S| = \frac{|S|}{2} \pm 2\sqrt{|X| \ln |\mathcal{F}|}.$$

Proof. We prove this claim by showing that a random R_1 suffices with non-zero probability. Independently with probability $\frac{1}{2}$, assign a value '+1' or '-1' to each $p_i \in X$. Let X_i be the value assigned to p_i.

Set $\Delta = 2\sqrt{|X| \ln |\mathcal{F}|}$. Then by the Chernoff bound, we have

$$\text{for each } S \in \mathcal{F}, \qquad \Pr\left[\sum_{p_i \in S} X_i > \Delta\right] < e^{-\frac{\Delta^2}{2|S|}} = e^{-\frac{\left(2\sqrt{|X| \ln |\mathcal{F}|}\right)^2}{2|S|}} \leq e^{-2 \ln |\mathcal{F}|} < \frac{1}{2|\mathcal{F}|}, \quad (7.1)$$

for $|\mathcal{F}| \geq 2$. By the union bound, the probability that there exists a set in $\mathcal{F}$ that fails to satisfy inequality (7.1) can be bounded as

$$\sum_{S \in \mathcal{F}} \Pr\left[\sum_{p_i \in S} X_i \geq \Delta\right] < \sum_{S \in \mathcal{F}} \frac{1}{2|\mathcal{F}|} = \frac{1}{2}.$$

Similarly, the probability that there exists a set S with $\sum_{p_i \in S} X_i \leq -\Delta$ is less than $\frac{1}{2}$.

Thus for each set S of $\mathcal{F}$, the number of elements assigned '+1' is equal, within an additive factor of Δ, to the elements assigned '-1'. Thus we can now take the elements of X assigned the value '+1' as our required set R_1. We are done, except a minor technical detail: we need to pick exactly $\frac{n}{2}$ elements†, which need not necessarily be true for either of '+1' or '-1' valued elements. We wrap up these details next.

Fix any set $S \in \mathcal{F}$. Let $S^+ \subseteq S$ be the elements with value $+1$, and $S^- = S \setminus S^+$ the elements with value -1. For each set $S \in \mathcal{F}$, we have

$$-\Delta \leq |S^+| - |S^-| \leq \Delta.$$

As $|S^+| + |S^-| = |S|$,

$$\begin{aligned} -\Delta \leq |S^+| - |S^-| \leq \Delta &\implies -\Delta \leq 2|S^+| - |S| \leq \Delta \\ &\implies \frac{|S|}{2} + \frac{\Delta}{2} \geq \left|S^+\right| \geq \frac{|S|}{2} - \frac{\Delta}{2}. \end{aligned}$$

Assuming w.l.o.g. that X is also a set in $\mathcal{F}$, and that there are at least $\frac{n}{2}$ elements assigned '+1'; then $\frac{n}{2} \leq |X^+| \leq \frac{n}{2} + \frac{\Delta}{2}$. By discarding at most $\frac{\Delta}{2}$ elements from X^+, we get the set R_1 with $\frac{n}{2}$ elements. For each set $S \in \mathcal{F}$,

$$|R_1 \cap S| \geq \left|S^+\right| - \frac{\Delta}{2} \geq \left(\frac{|S|}{2} - \frac{\Delta}{2}\right) - \frac{\Delta}{2} = \frac{|S|}{2} - 2\sqrt{|X| \log |\mathcal{F}|}.$$

$$|R_1 \cap S| \leq \left|S^+\right| \leq \frac{|S|}{2} + \frac{\Delta}{2} \leq \frac{|S|}{2} + 2\sqrt{|X| \log |\mathcal{F}|}.$$

□

We return to the main theorem:

Theorem 7.1. *Let $(X, \mathcal{F})$ be a set system, $|X| = n$, with the property that*

$$\textit{for any } Y \subseteq X, \qquad |\mathcal{F}|_Y| = O\left(|Y|^d\right).$$

Then there exists an ϵ-approximation for $\mathcal{F}$ of size $O\left(\frac{d}{\epsilon^2} \log \frac{d}{\epsilon}\right)$.

†This condition that $|R_1|$ be exactly equal to $\frac{n}{2}$ is not really necessary, but it simplifies the subsequent algebraic calculations.

Proof. Given $(X, \mathcal{F})$, let c be a constant such that for any $Y \subseteq X$, we have

$$|\mathcal{F}|_Y| \leq c \cdot |Y|^d.$$

We will iteratively compute R_i. Set the initial set $R_0 = X$. Apply the halving lemma to $(X, \mathcal{F})$ to get a set $R_1 \subset X$, $|R_1| = \frac{n}{2}$, such that

$$\text{for all } S \in \mathcal{F}: \quad |R_1 \cap S| = \frac{|S|}{2} \pm 2\sqrt{|R_0| \ln (c|R_0|^d)}. \tag{7.2}$$

Consider the set system $\mathcal{F}_1$ derived from projecting $\mathcal{F}$ to R_1, namely

$$\mathcal{F}_1 = \mathcal{F}|_{R_1} = \{R_1 \cap S \colon S \in \mathcal{F}\}.$$

Then we have

$$|\mathcal{F}_1| \leq c \cdot |R_1|^d = c \cdot \left(\frac{n}{2}\right)^d.$$

Now apply the halving lemma, this time on $(R_1, \mathcal{F}_1)$, to get a set $R_2 \subset R_1$, $|R_2| = \frac{|R_1|}{2} = \frac{n}{4}$, such that for all $S \in \mathcal{F}$:

$$\begin{aligned}
|R_2 \cap S| &= \frac{|R_1 \cap S|}{2} \pm 2\sqrt{|R_1| \ln (c\,|\mathcal{F}_1|)} \\
&= \frac{\frac{|S|}{2} \pm 2\sqrt{|R_0| \ln (c|R_0|^d)}}{2} \pm 2\sqrt{|R_1| \ln (c|R_1|^d)} \qquad \text{(here we used inequality (7.2))} \\
&= \frac{|S|}{4} \pm 2 \sum_{j=0}^{1} \frac{\sqrt{|R_j| \ln (c|R_j|^d)}}{2^{1-j}}.
\end{aligned}$$

Continuing on, we get the following.

Claim 7.2.

$$|R_i \cap S| = \frac{|S|}{2^i} \pm 2 \sum_{j=0}^{i-1} \frac{\sqrt{|R_j| \ln (c|R_j|^d)}}{2^{(i-1)-j}}.$$

Proof. Assume we have the set system $(R_{i-1}, \mathcal{F}_{i-1} = \{R_{i-1} \cap S \colon S \in \mathcal{F}\})$, where

$$|R_{i-1}| = \frac{n}{2^{i-1}}, \qquad |\mathcal{F}_{i-1}| \leq c \cdot |R_{i-1}|^d.$$

Furthermore, by inductive hypothesis, assume that

$$|R_{i-1} \cap S| = \frac{|S|}{2^{i-1}} \pm 2 \sum_{j=0}^{i-2} \frac{\sqrt{|R_j| \ln (c|R_j|^d)}}{2^{(i-2)-j}}.$$

Applying the halving lemma to $(R_{i-1}, \mathcal{F}_{i-1})$, we get a set R_i, $|R_i| = \frac{|R_{i-1}|}{2} = \frac{n}{2^i}$, with

$$|R_i \cap S| = \frac{|R_{i-1} \cap S|}{2} \pm 2\sqrt{|R_{i-1}| \ln (c|R_{i-1}|^d)}$$

$$= \frac{|S|}{2^i} \pm 2\sum_{j=0}^{i-2} \frac{\sqrt{|R_j| \ln (c|R_j|^d)}}{2^{(i-1)-j}} \pm 2\sqrt{|R_{i-1}| \ln (c|R_{i-1}|^d)}$$

$$= \frac{|S|}{2^i} \pm 2\sum_{j=0}^{i-1} \frac{\sqrt{|R_j| \ln (c|R_j|^d)}}{2^{(i-1)-j}}.$$

□

We bound the resulting error term:

$$2\sum_{j=0}^{i-1} \frac{\sqrt{|R_j| \ln (c|R_j|^d)}}{2^{(i-1)-j}} = \frac{2}{2^{i-1}} \sum_{j=0}^{i-1} 2^j \sqrt{|R_j| \ln (c|R_j|^d)}$$

$$\leq \frac{2}{2^{i-1}} \cdot c' \cdot 2^{i-1} \sqrt{|R_i| \ln (c|R_i|^d)}^{\dagger} = 2c' \cdot \sqrt{|R_i| \ln (c|R_i|^d)},$$

where c' is an absolute constant resulting from the geometric series.

Thus, at the i-th iteration, we have

$$|R_i \cap S| = \frac{|S|}{2^i} \pm 2c' \sqrt{|R_i| \ln \left(c\,|R_i|^d\right)},$$

$$\underset{\text{dividing by } |R_i|}{\Longrightarrow} \left| \frac{|R_i \cap S|}{|R_i|} - \frac{|S|}{n} \right| \leq 2c' \sqrt{\frac{\ln \left(c\,|R_i|^d\right)}{|R_i|}}.$$

We set the number of iterations t so that

$$2c' \sqrt{\frac{\ln \left(c\,|R_t|^d\right)}{|R_t|}} \leq \epsilon.$$

The reader can verify that indeed this is true when $t = \log\left(\frac{\epsilon^2 n}{dc_1 \log \frac{d}{\epsilon}}\right)$, for a large-enough constant $c_1 \geq 1$ (depending only on c and c').

Finally, the size of our approximation is

$$|R_t| = \frac{n}{2^t} = \frac{n}{\frac{\epsilon^2 n}{dc_1 \log \frac{d}{\epsilon}}} = \frac{dc_1 \log \frac{d}{\epsilon}}{\epsilon^2}.$$

Bibliography and discussion. This connection between balanced colorings (discrepancy) and ϵ-approximations was discovered in [1].

†Can be seen by a change of variables, from j to l. Set l such that $|R_j| = \frac{n}{2^j} = 2^l$. Then $\sum_{j=0}^{i-1} 2^j \sqrt{|R_j| \ln (c|R_j|^d)} = n \sum_{l=\log n-(i-1)}^{\log n} \sqrt{\frac{\ln(c2^{ld})}{2^l}} = O\left(n\sqrt{\frac{\ln(c(n/2^{i-1})^d)}{n/2^{i-1}}}\right) = O\left(2^{i-1}\sqrt{n/2^{i-1} \ln (c(n/2^{i-1})^d)}\right) = O\left(2^{i-1}\sqrt{|R_{i-1}| \ln (c|R_{i-1}|^d)}\right).$

[1] J. Matoušek, E. Welzl, and L. Wernisch. Discrepancy and approximations for bounded VC-dimension. *Combinatorica*, 13(4):455–466, 1993.