



Efficient Approximations of High-Dimensional Data

Mónika Csikós

► To cite this version:

Mónika Csikós. Efficient Approximations of High-Dimensional Data. Logic [math.LO]. Université Gustave Eiffel, 2022. English. NNT : 2022UEFL2004 . tel-03783594

HAL Id: tel-03783594

<https://theses.hal.science/tel-03783594>

Submitted on 22 Sep 2022

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Efficient Approximations of High-Dimensional Data

Thèse de doctorat de l'Université Gustave Eiffel

École doctorale n°532, Mathématiques et STIC (MSTIC),
Université Paris-Est,
77455 MARNE LA VALLEE CEDEX 2
Spécialité de doctorat: Informatique
Unité de recherche : LIGM

Thèse présentée et soutenue à l'Université Gustave Eiffel, le 31/01/2022, par

Mónika CSIKÓS

Composition du Jury

Venceslas BIRI
Professeur, Université Gustave Eiffel

Directeur de thèse

Jérémie CHALOPIN
Directeur de recherche CNRS, Aix-Marseille Université

Rapporteur

Claire MATHIEU
Directrice de recherche CNRS, Université de Paris

Président du jury

Wolfgang MULZER
Professeur, Freie Universität Berlin

Rapporteur

Nabil MUSTAFA
Professeur, Université Sorbonne Paris Nord

Directeur de thèse

Alantha NEWMAN
Chargée de recherche CNRS, Université Grenoble Alpes

Examinatrice

Lionel POURNIN
Professeur, Université Sorbonne Paris Nord

Examineur

Acknowledgements

I would like to express my sincere gratitude to all my supervisors. Thank you Venceslas for all our enlightening discussions on illumination problems! Nabil, I am really grateful for the last 3 years: for your time, patience, and continuous support; for the excellent pedagogical introduction to randomized algorithms, computational geometry, and real research; and for *treating me as if I was what I ought to be to help me to become what I am capable of being*.

I would like to thank Jérémie Chalopin and Wolfgang Mulzer for taking their time to review this manuscript and all of my jury Claire Mathieu, Alantha Newman, and Lionel Pournin to have accepted to be on my defense.

I would not be here without the continuous support of my family. My father, who sparked my interest 20 years ago by asking ‘*Is it easier to guess 5 numbers or 85 numbers out of 90 on a lottery?*’ and has been guiding me through more and more advanced topics in mathematics since then. My mother, who always looked after me and inspired me with her hard working. My brother, who showed that ‘anything is possible’ with a great deal of mindfulness and dedication; Dóri and Kinga who brought much joy to our family.

I am thankful for the people that I am happy to call my friends. For Lena, Barbi and Füli who were always there for me. For Ildi, Lilla, Dorka, Kori, Móni, Attila, Ágoston, István, and Tomi, who made my years at ELTE more fun. For Dasha, Deise, Lama, Tina, Daniel, Duc, Jonathan, Nicolas, Misha, Mohamed, Torsten, and Uwe who helped me to feel at home abroad.

I would like to thank all the professors at ELTE and UPEM, especially the EGRES group and András Frank – attending your inspiring courses and our discussions were the first steps towards my PhD. I am also grateful for Labex Bézout and the École Doctorale MSTIC for giving me an opportunity to live and work in the beautiful country of France.

Last, but (definitely) not least Женя, would all of this be even possible without you? Спасибо, szeretlek!

Summary

In the last decades, a lot of scientific and engineering efforts have been dedicated to processing and storing large data-sets. This thesis focuses on the development of theoretical tools and algorithmic solutions which allow to ease the burden of the “big data era”. In particular, we study fundamental tools of combinatorial data reduction: VC-dimension, ε -nets, ε -approximations, low-discrepancy colorings, and matchings with low crossing number.

The main algorithmic contribution of the thesis is an improved algorithm for constructing matchings with low crossing numbers. Matchings and spanning paths with low crossing numbers were introduced by [Welzl \(1988\)](#) for geometric range searching. Since then, they became a key structure in computational geometry and have found many other applications in various fields such as discrepancy theory, algorithmic graph theory, or learning theory. We propose a simple, randomized algorithm which given a set system with dual VC-dimension D , constructs a matching of X with crossing number $\tilde{O}(|X|^{1-1/D})$ in time $O(|X|^{1/D}|\mathcal{S}|)$, improving upon the previous-best construction of time $\tilde{O}(|X|^2|\mathcal{S}|)$. This contribution allows us to obtain improved algorithms for constructing low-discrepancy colorings and ε -approximations of sub-quadratic size. Our method does not use any input-specific tools, which makes it capable to handle abstract set systems and geometric set systems in high dimensions, without additional complications.

We resolve a long standing central open problem of VC theory: bounding the VC-dimension of k -fold unions and intersections of half-spaces in high dimensions. We show that the set system induced by k -fold union of half-spaces in \mathbb{R}^d have VC dimension $\Omega(dk \log k)$ if $d \geq 4$. This settles the open problem of [Blumer et al. \(1989\)](#), studied by [Eisenstat and Angluin \(2007\)](#); [Eisenstat \(2009\)](#). Furthermore, we show that asymptotic lower bounds on the VC dimension of k -fold unions of set systems imply asymptotic lower bounds on ε -net sizes.

Keywords: data approximation, sampling, VC-dimension

Résumé

Au cours des dernières décennies, de nombreux efforts scientifiques et techniques ont été consacrés au traitement et au stockage de grands ensembles de données. Cette thèse se concentre sur le développement d’outils théoriques et de solutions algorithmiques qui permettent d’alléger le fardeau de “l’ère des grandes données”. En particulier, nous étudions les outils fondamentaux de la réduction combinatoire des données : La dimension VC, les ε -réseaux, les ε -approximations, les colorations à faible discrédance, et les couplages à faible nombre de croisements.

La principale contribution algorithmique de la thèse est un algorithme amélioré pour construire des couplages avec un faible nombre de croisements. Les couplages et les chemins couvrant à faible nombre de croisements ont été introduits par [Welzl \(1988\)](#) pour le problème de la recherche par plage géométriques. Depuis, ils sont devenus une structure clé en géométrie computationnelle et ont trouvé de nombreuses autres applications dans des domaines variés tels que la théorie de la discrédance, la théorie algorithmique des graphes ou la théorie de l’apprentissage. Nous proposons un algorithme simple et aléatoire qui, étant donné un système d’ensembles avec une dimension VC duale D , construit un couplage de X avec un nombre de croisement $\tilde{O}(|X|^{1-1/D})$ en temps $O(|X|^{1/D}|\mathcal{S}|)$, améliorant la meilleure construction précédente en temps $\tilde{O}(|X|^2|\mathcal{S}|)$. Cette contribution nous permet d’obtenir des algorithmes améliorés pour construire des colorations à faible discrédance et des ε -approximations de taille sub-quadratique. Notre méthode n’utilise pas d’outils spécifiques à l’entrée, ce qui la rend capable de traiter des systèmes d’ensembles abstraits et des systèmes d’ensembles géométriques en haute dimension, sans complications supplémentaires.

Nous résolvons un problème ouvert central de la théorie VC: la borne de la dimension VC des k -unions et k -intersections de demi-espaces en haute dimension. Nous montrons que le système d’ensembles induit par les k -unions de demi-espaces dans \mathbb{R}^d ont une dimension VC $\Omega(dk \log k)$ si $d \geq 4$. Ceci résout le problème ouvert de [Blumer et al. \(1989\)](#), étudié par [Eisenstat and Angluin \(2007\)](#); [Eisenstat \(2009\)](#). De plus, nous montrons que les bornes inférieures asymptotiques sur la dimension VC des k -unions de systèmes d’ensembles impliquent des bornes inférieures asymptotiques sur les tailles des ε -réseaux.

Mots clés: approximation des données, échantillonnage, dimension VC

Organization of the thesis

Chapter 1 highlights my main contributions and provides a high-level overview on the essential background. Chapter 2 further elaborates on the studied notions and the fundamental links between them, and presents a detailed historic overview. The rest of the manuscript is partitioned into four parts: Chapter 4 presents an improved algorithm for constructing low-discrepancy colorings; Chapter 5 provides tight lower bounds for the VC-dimension of k -fold unions of half-spaces; Chapter 6 studies the upper and lower bounds on ε -net sizes by uniform sample; Chapter 7 present a simple and self-contained proof of optimal ε -approximations by uniform sampling. The following works are at the core of this thesis.

- Chapter 4
 - M Csikós, N. H. Mustafa. “Escaping the Curse of Spatial Partitioning: Matchings With Low Crossing Numbers and Their Applications” *37th International Symposium on Computational Geometry (SoCG 2021)*.
[Csikós and Mustafa \(2021\)](#)
 - M. Csikós, N. H. Mustafa. “Practical Algorithm for Colorings with Optimal Discrepancy” *Submitted*. [Csikós and Mustafa \(2022\)](#)
- Chapter 5
 - M. Csikós, A. Kupavskii, N. H. Mustafa. “Tight Lower Bounds on the VC-dimension of Geometric Set Systems” *Journal of Machine Learning Research*.
[Csikós et al. \(2019\)](#)
- Chapter 7
 - M. Csikós, N. H. Mustafa. “Optimal Approximations Made Easy” *Information Processing Letters*. [Csikós and Mustafa \(2022\)](#)
- Not included in the manuscript
 - M. Axenovich, M. Csikós. “Induced Saturation of Graphs”. *Discrete Mathematics*. [Axenovich and Csikós \(2019\)](#)

Contents

1	Introduction	1
1.1	Epsilon-approximations	1
1.2	Matchings with low crossing number	3
1.2.1	Previous constructions	4
1.2.2	Our method	5
1.3	VC-dimension	7
1.4	Epsilon-nets.	9
2	Preliminaries on set systems and approximations	13
2.1	Set systems and epsilon-approximations	13
2.1.1	VC-dimension in learning theory	18
2.1.2	Bounding the VC-dimension	20
2.2	Epsilon approximations and discrepancy	21
2.2.1	Halving method	21
2.2.2	Combinatorial discrepancy	23
3	Introduction française	25
3.1	Approximations de systèmes d'ensembles	25
3.2	Dimension VC	26
3.2.1	k -union et k -intersection	27
3.3	Au-delà de l'échantillonnage uniforme	28
3.4	Couplages avec un nombre de croisement faible	30
3.4.1	Constructions précédentes	30
3.4.2	Notre méthode	31
3.5	Epsilon-réseaux.	32
4	Improved algorithm for low-discrepancy colorings	35
4.1	Previous results	36

4.2	Our results	40
4.3	General framework	42
4.3.1	Matchings with low crossing number	44
4.3.2	Epsilon-approximations	46
4.4	Proof of our main discrepancy result	48
4.5	Proof of our main result on approximations	51
4.6	Proof of our main result on low-crossing matchings	52
4.7	Our results for geometric set systems	61
4.7.1	Semialgebraic set systems	62
4.7.2	Half-spaces and balls	63
4.8	Empirical aspects of the algorithm LOWDISCCOLOR	66
4.8.1	Experimental setup	66
4.8.2	Pre-sampling	66
4.8.3	Implementation details	68
4.8.4	Evaluation	72
4.8.5	Proof of pre-sampling guarantees	74
5	VC dimension of unions	83
5.1	Lower bound for unions of half-spaces	86
5.2	Lower bound for simplices	89
6	Epsilon-nets	93
6.1	Previous results	93
6.2	Our results	94
7	Simple proof of optimal approximations	101
7.1	Chaining	105
7.2	Iterative sampling	108
7.3	Sensitive approximations	109
8	Conclusion	111
8.1	Non-uniform sampling	111
8.1.1	Applicability - hereditary condition for different graphs	112
8.1.2	Relation to Primal-Dual reweighing algorithms	113
8.1.3	Matchings with low relative crossing number	114
8.1.4	Algorithmic complexity	114
8.2	VC-dimension	115

Chapter 1

Introduction

With recent technical developments, processing and storing large data-sets became increasingly important. As a consequence, the representation of massive data with compact summaries became a key theme in many modern research areas. This thesis contributes to fundamental tools of combinatorial data reduction: VC-dimension, ε -nets, ε -approximations, low-discrepancy colorings, and matchings with low crossing number. In this introductory chapter, we first give a brief, non-technical introduction to the topics of the thesis and then state our main contributions. To keep the current chapter focused, the presentation of technical details, examples, and extended historical overview is postponed to Chapter 2. Readers who are unfamiliar with the main notions appearing in the context of set systems, might find it useful to first read Chapter 2 and then return to Chapter 1.

1.1 Epsilon-approximations

We work with approximations of a general, abstract class of objects—set systems. A *set system* is a pair (X, \mathcal{S}) , where X is a set of elements and \mathcal{S} is a set of subsets of X . Throughout this manuscript, we use the notation $n = |X|$ and $m = |\mathcal{S}|$. Given a set system (X, \mathcal{S}) and a parameter $\varepsilon \in (0, 1)$, we say that $A \subseteq X$ is an ε -approximation of (X, \mathcal{S}) if for *any* $S \in \mathcal{S}$,

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

The notion of ε -approximation was initially introduced in one of the foundational papers of learning theory (Vapnik and Chervonenkis, 1971) and since then it became a fundamental tool in many areas. For instance, ε -approximations are used in geometric computing to design efficient approximate combinatorial queries on geometric data (see

e.g. [Har-Peled, 2011](#)) and play the key role in the core-set construction of [Feldman and Langberg \(2011\)](#).

The pioneering work of [Vapnik and Chervonenkis \(1971\)](#) showed that if (X, \mathcal{S}) has bounded complexity, then it admits ε -approximations of size independent of $|X|$ and $|\mathcal{S}|$, moreover, such an approximation can be constructed via uniform sampling. They measured the complexity of a set system via its *VC-dimension*, which is defined as the size of the largest $Y \subseteq X$ such that

$$|\{Y \cap S : S \in \mathcal{S}\}| = 2^{|Y|}.$$

The fundamental limits of approximations via uniform sampling was described by [Li et al. \(2001\)](#). They showed that if (X, \mathcal{S}) has VC-dimension at most d , then a uniform random sample $A \subseteq X$ of size $O(d/\varepsilon^2)$ is an ε -approximation with positive probability¹, moreover this bound is tight for uniform sampling guarantees.

The first problem we investigate in this thesis is developing efficient algorithms for constructing ε -approximations of sub-quadratic size. A possible way to obtain $o(d/\varepsilon^2)$ -sized approximations is by applying the iterative halving technique of [Matoušek et al. \(1991\)](#). They observed that finding a set of $|X|/2$ elements with low approximation error is essentially equivalent to the problem of finding a low-discrepancy coloring of (X, \mathcal{S}) . Given a two-coloring $\chi: X \rightarrow \{-1, 1\}$, the *discrepancy* of χ with respect to \mathcal{S} is defined as

$$\text{disc}_{\mathcal{S}}(\chi) = \max_{S \in \mathcal{S}} \left| \sum_{x \in S} \chi(x) \right|.$$

The halving method iteratively constructs a coloring with low discrepancy and proceeds with only half of the elements which belong to the same color class (for details see [Section 2.2.1](#)). Let δ and T be functions such that for any $Y \subseteq X$, a coloring of Y with discrepancy at most $\delta(|Y|)$ can be found in time $T(|Y|)$. The halving method guarantees that for any integer $t \geq 1$, one can construct an ε_t -approximation of size $n/2^t$ in time $T(n) + T(n/2) + \dots + T(n/2^{t-1})$, where

$$\varepsilon_t \leq \frac{2}{n} \left(\delta(n) + 2\delta\left(\frac{n}{2}\right) + \dots + 2^{t-1}\delta\left(\frac{n}{2^{t-1}}\right) \right). \quad (1.1)$$

It is known that for any set system, a random two-coloring χ of X satisfies $\text{disc}_{\mathcal{S}}(\chi) = O(\sqrt{n \ln m})$ with probability at least $1/2$. Furthermore, if no additional properties are known for \mathcal{S} and $m = \Omega(n^2)$, then this bound is essentially optimal. Not so surprisingly, a random coloring with discrepancy $\tilde{O}(\sqrt{n})$ is not enough to improve upon the uniform

¹In [Chapter 7](#), we give a simple proof of this result of [Li et al. \(2001\)](#).

sampling guarantee for approximations: by Equation (1.1), if we want to use the halving method to construct approximations of size $o(d/\varepsilon^2)$, we need to find colorings with discrepancy is $o(\sqrt{n})$. A series of research has demonstrated that if one assumes additional structure on (X, \mathcal{S}) , then X admits colorings of lower-order discrepancy.

Bounded VC-dimension. Matoušek (1995) showed that if the VC-dimension of (X, \mathcal{S}) is at most d , then X has a coloring with discrepancy $O\left(\sqrt{n^{1-1/d}}\right)$. This implies, by the halving method (Equation (1.1)), that set systems with VC-dimension at most d admit ε -approximations of size

$$O\left(\left(\frac{1}{\varepsilon^2}\right)^{\frac{d}{d+1}}\right).$$

While this coloring result implies improved guarantees, its proof is solely existential—it does not lead to an algorithm to construct an ε -approximation of the sub-quadratic size.

Bounded dual VC-dimension. The practical situation is better if one assumes the finiteness of the dual VC-dimension. The *dual VC-dimension* of (X, \mathcal{S}) is defined the VC-dimension of its dual set system (\mathcal{S}, X^*) , where $X^* = \{R_x : x \in X\}$ and $R_x = \{S \in \mathcal{S} : x \in S\}$. If the dual VC-dimension of (X, \mathcal{S}) is at most D , then one can find a two-coloring of X with discrepancy $O\left(\sqrt{n^{1-1/D} \ln m}\right)$ in time $O(n^3 m)$ (Matoušek et al., 1991). Thus, by the halving method, if (X, \mathcal{S}) has dual VC-dimension at most D , then one can find ε -approximations of size

$$\left(\frac{c_D \ln \frac{1}{\varepsilon}}{\varepsilon^2}\right)^{\frac{D}{D+1}} \tag{1.2}$$

in time $O(n^3 m)$. The key structure used to construct a coloring with $\tilde{O}\left(\sqrt{n^{1-1/D}}\right)$ discrepancy is a matching of X with low crossing number, which is also the algorithmic bottleneck requiring $O(n^3 m)$ time to find.

The first contribution of this thesis is an improved algorithm to construct matchings with low crossing number. In the next section, we shall focus on this problem: we give a formal definition, present previous algorithmic results, and state the guarantees of our method.

1.2 Matchings with low crossing number

A matching of a set X is a partition of X into pairs, which we refer to as edges. We say that a set $S \in \mathcal{S}$ crosses an edge $\{x, y\} \in X$ if $|S \cap \{x, y\}| = 1$. The crossing number of a

matching M with respect to \mathcal{S} is the maximum number of pairs in M that are crossed by a set in \mathcal{S} . Matchings and spanning paths² with low crossing numbers were introduced by Welzl (1988) for geometric range searching (see also Chazelle and Welzl, 1989; Welzl, 1992). Since then, they became key structures in computational geometry (see e.g. Pach and Agarwal, 2011) and has found many applications in other fields such as algorithmic graph theory (Ducoffe et al., 2020) or learning theory (Alon et al., 2016).

1.2.1 Previous constructions

The classical algorithm to construct a matching with low crossing number is based on the multiplicative weight update (MWU) method (Welzl, 1988; Chazelle and Welzl, 1989). Briefly, the algorithm maintains a weight function π on \mathcal{S} , with initial weights set to 1. It selects edges iteratively, always choosing an edge that is guaranteed to be crossed by sets of low total weight in π ; it then updates π based on the chosen edge. The algorithmic bottleneck is in finding such an edge: for an abstract set system without additional structure, this takes $O(n^2m)$ time for each of the $n/2$ iterations.

Another way of creating matchings in abstract set systems can be deduced using Linear Programming. The starting point is the observation that a spanning tree of crossing number $O(n^{1-1/D})$ can be found by rounding the solution of an LP on $\binom{n}{2}$ variables and $m + n$ constraints (Har-Peled, 2009). The resulting spanning tree then can be used to construct a matching with crossing number $O(n^{1-1/D})$ in linear time (see e.g. Pach and Agarwal, 2011, Chapter 15). Combining this with an efficient approximate LP solver (e.g., Chekuri and Quanrud (2018)) leads to a randomized $\tilde{O}(mn^2)$ time algorithm. While this method has an improved running time, it is rather involved and requires an explicit representation of a matrix storing the crossings between all the $\binom{n}{2}$ edges on X and the ranges in \mathcal{S} .

In case of geometric set systems, improved bounds are made possible using spatial partitioning. The current-best algorithms for set systems induced by half-spaces recursively construct simplicial partitions, stored in a hierarchical structure called the partition tree, which then at its base level gives a matching with low crossing number. This approach is used in the breakthrough result of Chan (2012) who gave an $O(n \log n)$ time algorithm to build partition trees with respect to half-spaces in \mathbb{R}^d , which then implies the same for computing matchings with crossing number $O(n^{1-1/d})$. However, optimal bounds for constructing spatial partitions are only known for the case of half-spaces. Even for half-space

²Spanning path is an ordering x_1, \dots, x_n of X , its edges are $\{x_i, x_{i+1}\}_{i=1}^{n-1}$, with analogous definition of crossing number.

ranges, practical implementation of spatial partitioning in \mathbb{R}^d , $d > 2$ remains an open problem in geometric computing. In particular, for $d > 2$, there were no previous implementations of low-crossing matchings; nor of constructing $o\left(\frac{d}{\varepsilon^2}\right)$ -sized ε -approximations even for $d = 3$.

1.2.2 Our method

The first contribution of this thesis is an improved algorithm to construct a matching (or spanning path) with low crossing number.

CONTRIBUTION

Let (X, \mathcal{S}) be a set system with dual VC-dimension at most D . Then there is a randomized algorithm which returns a matching of X with expected crossing number at most

$$\frac{36}{D} \cdot n^{1-1/D} + 30 \ln m \ln n + 26 \ln^2 n$$

with an expected $O(mn^{1/D} \ln m \cdot \min\{D, \ln n\} + n^{2+1/D} \ln n)$ calls to the membership Oracle of (X, \mathcal{S}) .

This result implies that one can construct colorings with discrepancy $\tilde{O}(\sqrt{n^{1-1/D}})$ and ε -approximations of size $O\left((\varepsilon^{-2} \ln \frac{1}{\varepsilon})^{\frac{D}{D+1}}\right)$ in time $\tilde{O}(n^{1/D}m + n^{2+1/D})$. Our method is outlined in Algorithm 1.

Importantly, we do not use spatial partitioning, which makes it possible to handle abstract set systems and geometric set systems in \mathbb{R}^d (not only in \mathbb{R}^2) without additional complications. The only black-box needed is the membership Oracle that returns for a given $x \in X$ and $S \in \mathcal{S}$, if $x \in S$.

A preliminary multi-threaded implementation of our algorithm in C++ for set systems induced on points by half-spaces in \mathbb{R}^d is available on [GitHub](#). It is approximately 200 lines of basic code without any non-trivial data-structures.

To further speed-up our algorithm, we implemented an accelerated version, where instead of maintaining the weights on all the $O(n^2)$ edges, we work with an initial uniform random sample of $O(n^{1+\alpha} \log n)$ edges, where $0 < \alpha \leq 1$ is a parameter to be set. We theoretically describe the trade-off between α and the expected crossing number guarantees.

Algorithm 1: SIMPLEMATCHING $((X, \mathcal{S}), D)$

```

 $M \leftarrow \emptyset$ 
while  $|X| \geq 4$  do
     $n \leftarrow |X|, m \leftarrow |\mathcal{S}|$ 
    Set the weight of every set in  $\mathcal{S}$  and every edge in  $E$  to 1 //  $E$  denotes  $\binom{X}{2}$ 
    for  $i = 1, \dots, n/4$  do
        sample  $e_i$  according to current edge weights
        sample  $S_i$  according to current set weights
        set the weight of  $e_i$  and of each edge adjacent to  $e_i$  to zero

         $E_i \leftarrow$  sample of  $\tilde{O}(n^{1+1/D})$  edges from  $E$ 
         $\mathcal{S}_i \leftarrow$  sample of  $\tilde{O}(mn^{-(1-1/D)})$  sets from  $\mathcal{S}$ 
        foreach  $e \in E_i$  do
            | halve weight of  $e$  if  $S_i$  crosses  $e$ 
        foreach  $S \in \mathcal{S}_i$  do
            | double weight of  $S$  if  $S$  crosses  $e_i$ 
         $M \leftarrow M \cup \{e_1, \dots, e_{n/4}\}$ 
         $X \leftarrow X \setminus \text{endpoints}(e_1, \dots, e_{n/4})$ 
return  $M$ 

```

CONTRIBUTION

Let (X, \mathcal{S}) be a set system with dual VC-dimension at most D , $0 < \alpha \leq 1$, and $0 < \delta < 1$. Let E be a uniform sample of $\tilde{O}(n^{1+\alpha} + n^\alpha \ln(1/\delta))$ edges from $\binom{X}{2}$. Then with probability at least $1 - \delta$, E contains a matching of size $n/4$ with crossing number $O(n^{1-\alpha/D} + \ln |\mathcal{S}|)$.

Figure 1.1 shows the matchings constructed by our algorithm on different input and query types. For reference, the last column contains randomly constructed matchings. We emphasize that each of these figures contain exactly 2500 edges. Our method, that is based only on sampling, gives a matching which adapts well to each specific instance. It is also important to observe that, our method explicitly takes into account the information about ranges, which leads to different outcomes for half-spaces as ranges (**Left column**) and disks as ranges (**Middle column**). Meanwhile, random sampling is oblivious to ranges.

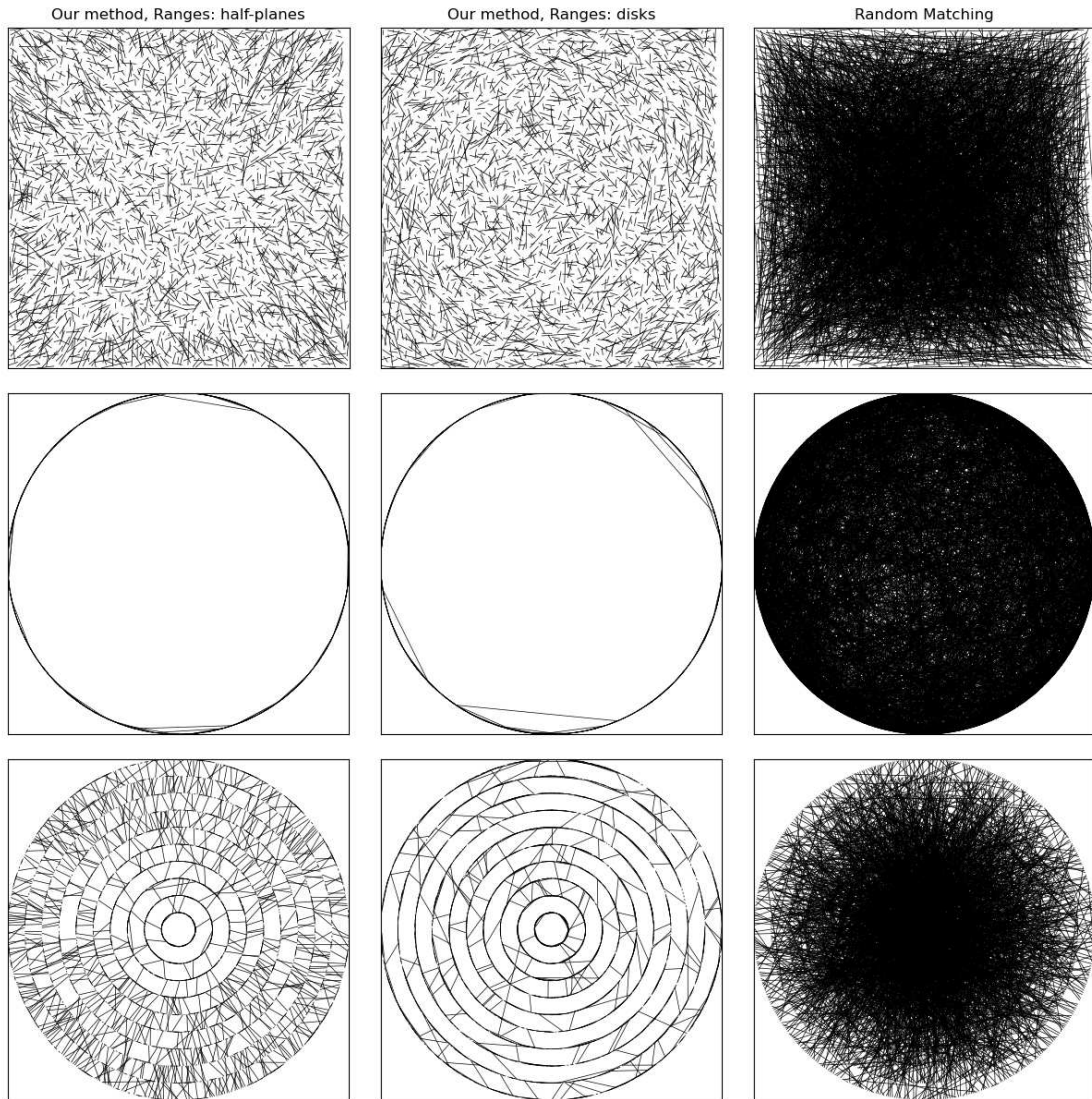


Figure 1.1: Matchings on 5000 points with different input and query types.

This makes progress towards the goals expressed in the survey on range searching and its applications ([Agarwal, 2017](#)): “...an interesting open question is to develop simple data structures that work well under some assumptions on input points and query ranges”.

1.3 VC-dimension

Since the seminal work of [Vapnik and Chervonenkis \(1971\)](#), one of the key quantities describing the complexity of a set system is the VC-dimension. As such, VC-dimension is the foundation of many theoretical results in computational geometry ([Matoušek, 2013](#)) and learning theory ([Shalev-Shwartz and Ben-David, 2014](#)). What is more important,

it plays a central role in uniform sampling guarantees and is often required as an input parameter of algorithms.

While for simple geometric set systems (e.g. ones induced by half-spaces), one can bound the VC-dimension using elementary arguments, many applications require more complex set systems. A long standing central open problem since 1989 was bounding the VC-dimension of k -fold unions and intersections of half-spaces in high dimensions. The k -fold union of a range set \mathcal{S} is defined as $\mathcal{S}^{k\cup} = \{S_1 \cup \dots \cup S_k : S_1, \dots, S_k \in \mathcal{S}\}$. Similarly, one can define the k -fold intersection of \mathcal{S} , denoted by $\mathcal{S}^{k\cap}$, as the range set consisting of all subsets derived from the common intersection of at most k sets of \mathcal{S} .

The foundational paper of [Blumer et al. \(1989\)](#) showed that for any set system (X, \mathcal{S}) both VC-dimensions $d_{\text{VC}}(\mathcal{S}^{k\cup})$ and $d_{\text{VC}}(\mathcal{S}^{k\cap})$ are at most $O(d_{\text{VC}}(\mathcal{S}) \cdot k \log k)$. They also studied the question whether this general upper bound is tight for the basic geometric case, where \mathcal{S} consists of sets induced by half-spaces in \mathbb{R}^d . It turned out that for dimensions 2 and 3, an improved bound of $O(k)$ holds ([Blumer et al., 1989](#); [Dobkin and Gunopulos, 1995](#)). The prevailing conjecture was that the $\log k$ factor can also be removed for dimensions $d \geq 4$.

Despite several attempts to settle this conjecture, the resolution of the VC-dimension of k -fold unions and intersections of half-spaces was left as one of the main open problems in the PhD thesis of [Johnson \(2008\)](#), as well as in the papers [Eisenstat and Angluin \(2007\)](#), and [Eisenstat \(2009\)](#). Yet, as it was pointed out by [Bachem \(2018\)](#), several papers in learning theory and computational geometry literature assume that the VC-dimension of k -fold unions and intersections of half-spaces in \mathbb{R}^d is $O(dk)$. For example, the coresset size bounds in the constructions of [Feldman and Langberg \(2011\)](#), [Balcan et al. \(2013\)](#), and [Lucic et al. \(2016\)](#) would require an additional $\log k$ factor in the coresset size if the upper-bound of $O(dk \log k)$ was tight for the k -fold intersection of half-spaces.

We prove that in dimensions $d \geq 4$, the $\log k$ factor is in fact necessary. In particular, we describe a deterministic lower-bound construction, which shows that the $O(dk \log k)$ bound is optimal for k -fold unions or intersections of half-spaces in dimensions at least 4.

CONTRIBUTION

There exists a universal constant c such that for any positive integers k and $d \geq 4$, there exists a set P of at least $c \cdot dk \log k$ points in \mathbb{R}^d such that the set system \mathcal{H} induced on P by half-spaces satisfies $d_{\text{VC}}(P, \mathcal{H}^{k\cup}) = d_{\text{VC}}(P, \mathcal{H}^{k\cap}) = |P|$.

Apart from settling affirmatively the 30 year old problem posed by [Blumer et al. \(1989\)](#), this statement provides the first non-probabilistic construction of a geometric

set system for which $d_{\text{VC}}(\mathcal{S}^{k\cup}) = \Omega(d \cdot k \log k)$ —previous lower bound constructions were randomized and only proved the existence of an abstract set system for which the $d_{\text{VC}}(\mathcal{S}^{k\cup}) = O(d \cdot k \log k)$ bound is tight (Eisenstat and Angluin, 2007; Eisenstat, 2009).

Set systems induced by Boolean combinations of geometric objects arise in other fields as well. For instance, one of the basic set systems in computational geometry is defined on a set of hyperplanes \mathcal{H} in \mathbb{R}^d with ranges

$$\Delta(\mathcal{H}) = \left\{ \mathcal{H}' \subseteq \mathcal{H} : \exists \text{ an open } d\text{-dimensional simplex } S \text{ in } \mathbb{R}^d \text{ such that } \right. \\ \left. H \in \mathcal{H}' \text{ if and only if } H \text{ intersects } S \right\}.$$

The importance of $(\mathcal{H}, \Delta(\mathcal{H}))$ derives from the fact that it is the set system underlying the construction of *cuttings* via random sampling (Chazelle and Friedman, 1990). Cuttings are the key tool for fast point-location algorithms and were studied in detail recently by Ezra et al. (2020), who provided the best bounds so far for the VC-dimension of $\Delta(\mathcal{H})$.

Lemma 1 (Ezra et al. 2020). *For $d \geq 9$, we have*

$$d(d+1) \leq d_{\text{VC}}(\Delta(\mathcal{H})) \leq 5 \cdot d^2 \log d.$$

We close the asymptotic gap between the upper and lower bounds of Lemma 1 by establishing an $\Omega(d^2 \log d)$ bound on the VC-dimension of $\Delta(\mathcal{H})$.

CONTRIBUTION

There is a universal constant c such that for any $d \geq 4$ there exists a set \mathcal{H} of $c \cdot d^2 \log d$ hyperplanes in \mathbb{R}^d such that $d_{\text{VC}}(\mathcal{H}, \Delta(\mathcal{H})) = |\mathcal{H}|$.

1.4 Epsilon-nets.

Given a parameter $0 < \varepsilon < 1$, a set $N \subseteq X$ is an ε -net of (X, \mathcal{S}) if any range $S \in \mathcal{S}$ with size $|S| > \varepsilon|X|$ contains at least one element of N . In contrast to ε -approximations that uniformly approximate the proportion of points in each range, ε -nets guarantee only 1 representative element in each large range (no matter if the range contains $2\varepsilon \cdot |X|$ or $|X|$ elements). For many application, this weaker property is sufficient: for example, ε -nets (of the set system $(\mathcal{H}, \Delta(\mathcal{H}))$ studied in the previous section) can be used to de-randomize sampling-based linear programming solvers (Chan, 2018) or to create cuttings (Clarkson, 1987; Chazelle and Friedman, 1990)—in the latter example, one uses an ε -net of the set

system $(\mathcal{H}, \Delta(\mathcal{H}))$ studied in the previous section. Similar to the case of ε -approximations, VC-dimension turned out to be the key parameter in the uniform sampling guarantees for ε -nets. [Haussler and Welzl \(1987\)](#) showed if (X, \mathcal{S}) has VC-dimension at most d , then a random sample of

$$\max \left\{ \frac{8d}{\varepsilon} \log \frac{8d}{\varepsilon}, \frac{4}{\varepsilon} \log \frac{2}{\delta} \right\}$$

points from X forms an ε -net of (X, \mathcal{S}) with probability at least $1 - \delta$. Later, [Komlós et al. \(1992\)](#) removed the logarithmic dependence on d by proving that for any $0 < \varepsilon < \varepsilon_0$, a uniform sample of

$$\frac{d}{\varepsilon} \log \frac{1}{\varepsilon} + o \left(\frac{d}{\varepsilon} \log \frac{1}{\varepsilon} \right).$$

elements from X is an ε -net with large probability. Moreover, they showed that this bound is asymptotically tight: for any $d \geq 2$, that there exists of a set system with VC-dimension d , in which any ε -net ($\varepsilon < \varepsilon_0$) has size at least $(1 - 2/d)(d/\varepsilon) \log(1/\varepsilon)$. Thus, in contrast to ε -approximation, uniform sampling yields optimal ε -nets.

In Chapter 6, we prove the following bound that has the current-best leading term up to constant and guarantees arbitrarily large success probability.

CONTRIBUTION

Let (X, \mathcal{S}) be a set system with VC-dimension at most d , $0 < \delta < 1$, and $0 < \varepsilon < \varepsilon_0(d, \delta)$. Then a uniform random sample $N \subseteq X$ of size

$$\frac{d}{\varepsilon} \left(\log \frac{1}{\varepsilon} + \frac{1}{d} \log \frac{1}{\delta} + 2 \log \left(\log \frac{1}{\varepsilon} + \frac{1}{d} \log \frac{1}{\delta} \right) + 5 \right)$$

is an ε -net for (X, \mathcal{S}) with probability at least $1 - \delta$.

The aforementioned lower bound construction of [Komlós et al. \(1992\)](#) was of probabilistic nature, proving the existence of an abstract set system of VC-dimension at most d that does not have small ε -nets. The investigation of the tightness of $O(d/\varepsilon \log(1/\varepsilon))$ upper-bound for set systems induced by geometric objects has been an active area and is still ongoing. For instance, the current best lower and upper bounds for the ε -net sizes in set systems induced by balls in \mathbb{R}^3 are $\Omega(1/\varepsilon)$ and $O(1/\varepsilon \log(1/\varepsilon))$ respectively.

Our last contribution is a lemma that can potentially help tackling the problem of lower-bounding ε -net sizes as it gives a new tool to establish these bounds. In a way, it can be seen as a complement to the above upper bounds: we show that asymptotic lower bounds of the VC-dimension of $(X, \mathcal{S}^{k\cup})$ imply *lower bounds* for the sizes of ε -nets in a sub-system of (X, \mathcal{S}) .

CONTRIBUTION

Let (X, \mathcal{S}) be a set system, $\varepsilon \in (0, 1)$ be a parameter, and $f : \mathbb{N} \rightarrow \mathbb{R}$ be a function that satisfies $\text{d}_{\text{VC}}(X, \mathcal{S}^{k \cup}) \geq \text{d}_{\text{VC}}(X, \mathcal{S}) \cdot k \cdot f(k)$. Then there is a subset X' of X so that any ε -net for $(X', \mathcal{S}|_{X'})$ must have size at least

$$\frac{\text{d}_{\text{VC}}(X', \mathcal{S}|_{X'})}{4\varepsilon} \cdot f\left(\frac{1}{2\varepsilon}\right).$$

For example, the combination of the above result with the lower bound for VC-dimension of k -fold unions of half-spaces presented in Section 1.3, immediately yields a lower bound of $\Omega(d/\varepsilon \log(1/\varepsilon))$ on the ε -net sizes in set systems induced by half-spaces, recovering the result of [Pach and Tardos \(2013\)](#).

Chapter 2

Preliminaries on set systems and approximations

In this part we further elaborate on the main concepts appearing in the previous chapter. In particular, since the main contributions of this work have been already stated in Chapter 1, we only give a (nearly) self-contained, pedagogical introduction to the necessary background, giving some examples from the computational geometry and learning theory, and provide a more detailed historic overview. The content of this chapter is relatively standard and can generally be skipped by readers familiar with the subject at hand.

2.1 Set systems and epsilon-approximations

To motivate the notion of approximations, consider the following fundamental problem from computational geometry.

Half-space counting. Given a set P of n points X in \mathbb{R}^d and a query half-space H , count the number of points in P , which lie in H .

A straightforward solution has linear complexity: we can simply test for each $p \in P$ whether it is contained in H . This trivial query time of $O(n)$ can be significantly improved to $O(\log n)$ by pre-processing P into a suitable (but rather involved) data-structure. There has been a long line of research studying different query-time/space trade-offs for exact half-space counting, see the survey of [Agarwal \(2017\)](#).

If we allow ourselves to make an error of 0.1%, there is a much simpler solution: we can replace P with a smaller set A , which can be used to approximately count the number of

points in any half-space H . In particular, a uniform random sample $A \subseteq X$ of *constant* size satisfies

$$|A \cap H| \cdot \frac{|P|}{|A|} \in \left[|P \cap H| - 0.001 \cdot |P|, |P \cap H| + 0.001 \cdot |P| \right]$$

with high probability. This implies that for any half-space H , it is sufficient to count $|A \cap H|$ (which we can do in a constant time with the naive linear-time algorithm) to get an approximation of $|P \cap H|$.

In the approximate half-space counting example, the key property that the sample A satisfies is that for *any* query half-space H , the proportion of points in A which lie in H is close to the proportion of points in P that are in H . Formally, A is a subset of P such that for *any* half-space H

$$\left| \frac{|P \cap H|}{|P|} - \frac{|A \cap H|}{|A|} \right| \leq 0.001. \quad (2.1)$$

We formalize this property for a more abstract class of objects—set systems. A *set system* is a pair (X, \mathcal{S}) , where X is a set of elements and \mathcal{S} is a set of subsets of X . We refer to the elements of \mathcal{S} as *ranges*. For instance, the half-space counting problem can be modeled with a set system (P, \mathcal{H}) , where \mathcal{H} consists of those subsets of P that can be obtained as an intersection with a half-space.

The key property formulated in Equation (2.1) can be generalized to abstract set systems as follows. Given a set system (X, \mathcal{S}) and a parameter $0 < \varepsilon < 1$, we say that $A \subseteq X$ is an ε -*approximation* of (X, \mathcal{S}) if for each $S \in \mathcal{S}$,

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

We can obtain ε -approximations via uniform sampling. A first bound can be deduced from the following two basic probabilistic ingredients:

Chernoff's bound. Let X be a set of n elements and A be a uniform random sample of X of size t . Then for any $S \subseteq X$ and $\eta > 0$, we have

$$\mathbb{P} \left[|A \cap S| \notin \left(\frac{|S|t}{n} - \eta, \frac{|S|t}{n} + \eta \right) \right] \leq 2 \exp \left(-\frac{\eta^2 n}{2|S|t + \eta n} \right). \quad (2.2)$$

Union bound. For a countable set of events $\{E_i\}_{i=1}^{\infty}$

$$\mathbb{P} \left[\bigcup_{i=1}^{\infty} E_i \right] \leq \sum_{i=1}^{\infty} \mathbb{P} [E_i]. \quad (2.3)$$

Indeed, applying Chernoff's bound with $\eta = \varepsilon t$, we get that a uniform random sample A of size t fails to be an ε -approximation for a *fixed* $S \in \mathcal{S}$ with probability at most

$$2 \exp \left(-\frac{\varepsilon^2 t^2 n}{2|S|t + \varepsilon t n} \right) \leq 2 \exp \left(-\frac{\varepsilon^2 t}{\frac{2|S|}{n} + \varepsilon} \right) \leq 2 \exp \left(-\frac{\varepsilon^2 t}{3} \right).$$

The quantity above bounds the probability of failure for an arbitrary, but fixed range in \mathcal{S} . In order to establish a uniform guarantee over all ranges $S \in \mathcal{S}$, we use the union bound. Hence, we obtain that a uniform random sample A of size t fails to be an ε -approximation for (X, \mathcal{S}) with probability at most

$$2|\mathcal{S}| \exp \left(-\frac{\varepsilon^2 t}{3} \right).$$

In other words, a uniform random sample from X of size

$$\frac{3}{\varepsilon^2} \ln(4|\mathcal{S}|) \tag{2.4}$$

is an ε -approximation of (X, \mathcal{S}) with probability at least $1/2$.

A breakthrough in the study of approximations dates back to 1971 when [Vapnik and Chervonenkis](#) described a key property which guarantees the existence of ε -approximations of size independent of $|X|$ or $|\mathcal{S}|$. Before we present this property, let us introduce additional notions. For a set $Y \subseteq X$, define the *projection* of \mathcal{S} onto Y as

$$\mathcal{S}|_Y := \{S \cap Y : S \in \mathcal{S}\}.$$

We say that the sets in $\mathcal{S}|_Y$ are the subsets of Y *induced by* \mathcal{S} , and that \mathcal{S} *shatters* Y if each of the $2^{|Y|}$ subsets of Y are induced by \mathcal{S} , that is,

$$|\mathcal{S}|_Y| = 2^{|Y|}.$$

[Vapnik and Chervonenkis \(1971\)](#) characterized the complexity of a set system via its *shatter function*. The *shatter function* $\pi_{\mathcal{S}} : [1, |X|] \rightarrow \mathbb{N}$ is defined as

$$\pi_{\mathcal{S}}(k) := \max_{Y \subseteq X, |Y|=k} |\mathcal{S}|_Y|,$$

that is, $\pi_{\mathcal{S}}(k)$ is the maximum number of ranges that \mathcal{S} can induce on a k -element subsets of X . It is easy to see that if \mathcal{S} contains every subset of X , then $\pi_{\mathcal{S}}(k) = 2^k$ for all k .

Why is this a good measurement of complexity? To get an intuition, observe that in Equation (2.4), the dependence of the sample size on the input parameters arises

from the application of the union bound over all ranges of \mathcal{S} . However, if for instance, the shatter function of (X, \mathcal{S}) has polynomial growth, that is, $\pi_{\mathcal{S}}(k) = O(k^d)$ for some constant d , then a uniform sample from X of size

$$O\left(\frac{d}{\varepsilon^2} \log |X|\right) \quad (2.5)$$

is an ε -approximation of (X, \mathcal{S}) with positive probability. This is a more refined description of failure probability, which can be used to remove the dependence on $|X|$ completely with more advanced probabilistic techniques.

Vapnik and Chervonenkis (1971) proved that the growth of the shatter function exhibits a dichotomy.

Theorem 2. *The shatter function $\pi_{\mathcal{S}}(k)$ is either identically equals to 2^k or else $\pi_{\mathcal{S}}(k) \leq k^{d+1} + 1$, where d is the smallest positive number for which the equation $\pi_{\mathcal{S}}(d) = 2^d$ holds and $\pi_{\mathcal{S}}(d+1) < 2^{d+1}$.*

For honoring their contribution, the ‘threshold’ d of Theorem 2 entered the community as the *Vapnik-Chervonenkis dimension*, or shortly, the VC-dimension. This definition can be rephrased as: the VC-dimension of (X, \mathcal{S}) is the smallest positive d such that X has a d -element subset which is shattered, but no $d+1$ -element subset of X can be shattered. Observe that the property of ‘being shattered by \mathcal{S} ’ is monotone in the sense that if Y is shattered, then each subset of Y is shattered. This implies that if the $d_{\text{VC}}(X, \mathcal{S}) = d$, then no subset of X of size larger than d can be shattered. Therefore, the VC-dimension of (X, \mathcal{S}) can be defined as the largest d such that X has a subset of size d which is shattered by \mathcal{S} .

The notion of VC-dimension became a central tool in many fields including learning theory, computational and combinatorial geometry. One of the most fundamental set systems arising in these fields is the one induced by half-spaces (in learning theory this is the underlying set systems of linear classifiers often deployed in practice). When bounding the VC-dimension of set systems induced by geometric objects on points, we are usually interested in the ‘worst-case scenario’, that is, what is the largest obtainable VC-dimension induced by these objects on any set of points. Hence, in these cases, we often write $d_{\text{VC}}(\mathcal{S})$, which denotes

$$\max_X \{d_{\text{VC}}(X, \mathcal{S})\}.$$

For instance, the VC-dimension of half-spaces in \mathbb{R}^d is $d+1$ —this means that for any point-set P in \mathbb{R}^d , the VC-dimension of the set system induced by half-spaces on P is at most $d+1$.

Vapnik and Chervonenkis (1971) also showed that if (X, \mathcal{S}) has finite VC-dimension, then it admits an ε -approximation of size independent of $|X|$ or $|\mathcal{S}|$. Moreover, such approximations can be constructed simply by uniform sampling.

Theorem 3. *Let (X, \mathcal{S}) be a set system with VC-dimension d and $0 < \varepsilon < 1$ be a parameter. Then a uniform random sample $A \subset X$ of size*

$$O\left(\frac{d}{\varepsilon^2} \log \frac{d}{\varepsilon}\right)$$

is an ε -approximation of (X, \mathcal{S}) with positive probability.

30 years after the pioneering work of Vapnik and Chervonenkis (1971), Li et al. (2001) proved optimal bounds on approximations via uniform sampling. In particular, they showed that with a more careful analysis of the failure probability, the $\log(d/\varepsilon)$ factor in Theorem 3 can be removed and that this is the best possible ε -approximation size obtainable by uniform sampling.

Theorem 4 (Li et al. (2001)). *There exists an absolute constant $c \geq 1$ such that the following holds. Let (X, \mathcal{S}) be a set system with VC-dimension at most d and let $0 < \varepsilon, \gamma < 1$ be given parameters. Then for any integer t at least*

$$\frac{c}{\varepsilon^2} \left(d + \ln \frac{1}{\gamma} \right),$$

a uniform random sample $A \subseteq X$ of size t is an ε -approximation for (X, \mathcal{S}) with probability at least $1 - \gamma$. Moreover, this bound is tight.

Remark. Li et al. (2001) work with a more general notion of approximations: (ε, δ) -samples, for its definition and further details, see Chapter 7, where we also provide a simpler proof of their result for finite set systems.

To appreciate the power of these guarantees, let us again have a look at a simple example¹.

Public opinion poll. Imagine we are planning the election campaign of a party and would like to understand its popularity within certain social groups. We can model this problem as a set system, where X is the set of all citizens of our country and each set in \mathcal{S} represents voters of the party that belong to one of the social groups of our interest. The ranges in \mathcal{S} are unknown to us (yet, we can reasonably assume that \mathcal{S} is sufficiently simple), all we can do is to ask the preference of a random

¹Special thanks to János Pach, who inspired this example with a similar one on ε -nets.

citizen². By Theorem 3, it is possible to create a 0.01-approximation A of (X, \mathcal{S}) by uniform sampling. This way, if our observations on A tell us that 70% of women would vote for our party, then we know that the real proportion of our supporters among women is between 69% and 71%. Moreover, Theorem 3 also implies that the number of people we need to ask to create such a representative set only depends on the error parameter 0.01 and the VC-dimension of (X, \mathcal{S}) , which captures the demographic structure of our voters. Therefore, on a high level, we would need to make the same number of polls to represent the citizens of France, Germany, or the USA, even though the population of these countries largely differ.

2.1.1 VC-dimension in learning theory

In this subsection, we give a brief description of basic problems in learning theory, where the notion of ε -approximations was originally introduced. While in computational geometry we usually want to reduce a large, but finite input data-set, learning theory works in a different setup, where we do not have access to the whole data. Instead, we only have observations (samples) from an underlying unknown probability distribution. A basic problem is to determine, how many samples do we need to understand this unknown distribution.

Realizable binary classification. Let (X, \mathcal{S}) be a set system³ with each element of the base set X labeled by 0 or 1. For each range $S \in \mathcal{S}$, define a classifier

$$\ell_S(x) = \mathbf{I}(x \in S) = \begin{cases} 1 & \text{if } x \in S \\ 0 & \text{if } x \notin S \end{cases}.$$

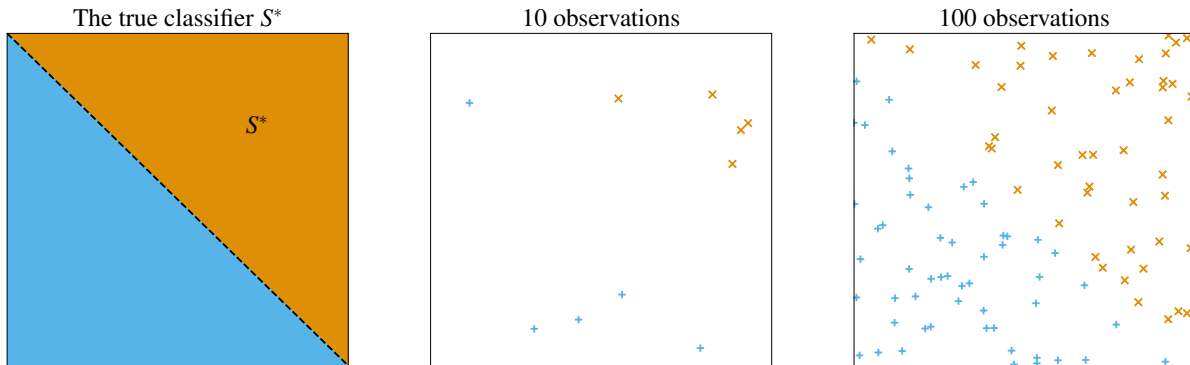
In the realizable case of binary classification, we assume that there exists $S^* \in \mathcal{S}$ such that the label of each $x \in X$ is given by $\ell_{S^*}(x)$. Naturally, we would like to find or approximate the S^* , which assumed to be unknown. To do so, we can sample a point x according to some unknown probability distribution \mathbb{P} on X and observe its associated label $\ell_{S^*}(x)$. In this setup, the goodness of a candidate range S is given by its error—the average disagreement with S^* :

$$\text{err}(S) := \mathbb{P}[x : \ell_{S^*}(x) \neq \ell_S(x)].$$

²While in practice, it is challenging to create a uniform random sample of people without any bias, assume now for simplicity that it is possible.

³In learning theory, the elements of \mathcal{S} are also called *concepts*, and \mathcal{S} is called a *concept class* on X .

The above framework is often called PAC (Probably Approximately Correct) learning. A fundamental question in this theory is: how many points does one need to sample in order to guarantee the existence of an algorithm which finds a candidate range \hat{S} with $\text{err}(\hat{S}) \leq \varepsilon$ (no matter \mathbb{P} and S^*)?



The answer to this question is known as the sample complexity of PAC learning. It was shown in the foundational paper of [Blumer et al. \(1989\)](#) that in case of realizable binary classification, the existence of such an algorithm is equivalent to the finiteness of the VC-dimension of (X, \mathcal{S}) . Moreover, they showed that if $d_{\text{VC}}(X, \mathcal{S}) = d$, then for any $\varepsilon \in (0, 1)$, the sample complexity is

$$O\left(\frac{d}{\varepsilon} \log \frac{d}{\varepsilon}\right).$$

The main drawback of PAC learning lies in the realizability assumption—we explicitly assumed that the label of each element in X is determined by a range $S^* \in \mathcal{S}$, which is very restrictive in practice.

A more realistic setup is known as the *agnostic* PAC learning. In agnostic PAC learning we no longer assume that there is a range that perfectly separates zeros and ones. While, from the perspective of sample complexity, the situation changes drastically, the learnability in this generalized setup is still equivalent to the finiteness of the VC-dimension of (X, \mathcal{S}) . Let us also mention, without going into further details, that agnostic PAC learnability is in some sense equivalent to a notion of ε -approximations. In particular, one of the main results of [Vapnik and Chervonenkis \(1971\)](#) states that if $d_{\text{VC}}(X, \mathcal{S}) = d$, then for any $\varepsilon \in (0, 1)$, the sample complexity⁴ is

$$O\left(\frac{d}{\varepsilon^2} \log \frac{d}{\varepsilon}\right).$$

For further bibliographic and methodological details related to learning theory, we refer to ([Shalev-Shwartz and Ben-David, 2014](#), Section 6).

⁴One needs to appropriately modify the notion of sample complexity for the agnostic PAC learning.

2.1.2 Bounding the VC-dimension

In learning theory, using only half-space classifiers can be too restrictive. An interesting and much more complex range space is the one induced by all convex polytopes. However, the VC-dimension of the set system induced by these objects is infinity as one can shatter the vertices of *any* convex polytope. Hence, recalling the results from Section 2.1.1, learning these ranges with samples is impossible.

A very natural question is: what type of geometric range spaces are richer than half-spaces but still have finite VC-dimension? This was one of the main motivations of [Blumer et al. \(1989\)](#), who studied k -fold unions and intersections of set systems. Formally, the k -fold union of a range set \mathcal{S} is defined as

$$\mathcal{S}^{k\cup} = \{S_1 \cup \dots \cup S_k : S_1, \dots, S_k \in \mathcal{S}\}.$$

Similarly, one can define the k -fold intersection of \mathcal{S} , denoted by $\mathcal{S}^{k\cap}$, as the set system consisting of all subsets derived from the common intersection of at most k sets of \mathcal{S} . Note that as the subsets S_1, \dots, S_k need not necessarily be distinct, we have $\mathcal{S} \subseteq \mathcal{S}^{k\cup}$ and $\mathcal{S} \subseteq \mathcal{S}^{k\cap}$. Among other contributions related to learning theory, [Blumer et al. \(1989\)](#) established bounds on the asymptotic growth of $d_{VC}(\mathcal{S}^{k\cup})$ and $d_{VC}(\mathcal{S}^{k\cap})$ as a function of k and $d_{VC}(X, \mathcal{S})$.

Theorem 5 ([Blumer et al. 1989](#)). *For any set system (X, \mathcal{S}) with finite VC-dimension and integer k*

$$d_{VC}(\mathcal{S}^{k\cup}) = O(d_{VC}(\mathcal{S}) \cdot k \log k),$$

$$d_{VC}(\mathcal{S}^{k\cap}) = O(d_{VC}(\mathcal{S}) \cdot k \log k).$$

Moreover, there are set systems such that $d_{VC}(\mathcal{S}^{k\cup}) = \Omega(d_{VC}(\mathcal{S}) \cdot k)$ and $d_{VC}(\mathcal{S}^{k\cap}) = \Omega(d_{VC}(\mathcal{S}) \cdot k)$.

The upper bounds of Theorem 5 can be deduced from the following refinement of Theorem 2 which is implicit in the work of [Vapnik and Chervonenkis \(1971\)](#) and was independently discovered by [Sauer \(1972\)](#); [Shelah \(1972\)](#).

Lemma 6 (Sauer-Shelah lemma). *Let (X, \mathcal{S}) be a set system with VC-dimension at most d . Then for any $Y \subseteq X$,*

$$|\mathcal{S}|_Y \leq (e|Y|/d)^d.$$

In the subsequent works of [Eisenstat and Angluin \(2007\)](#) and [Eisenstat \(2009\)](#), the authors proved that the upper bound of Theorem 5 is asymptotically optimal for k -fold unions if $d_{\text{VC}}(\mathcal{S}) \geq 2$ and that for $d_{\text{VC}}(\mathcal{S}) = 1$, a tight upper bound of $O(k)$ holds. Their lower bound constructions were of probabilistic nature, proving the existence of an abstract set system (X, \mathcal{S}) for which $d_{\text{VC}}(\mathcal{S}^{k\cup}) = \Omega(d_{\text{VC}}(\mathcal{S}) \cdot k \log k)$.

2.2 Epsilon approximations and discrepancy

As it was shown by [Li et al. \(2001\)](#), the $O(d/\varepsilon^2)$ bound is optimal for uniformly sampled approximations of set systems with VC-dimension at most d . In the last part of this chapter, we present the halving method of [Matoušek et al. \(1991\)](#) which can be used to obtain smaller-sized approximations.

2.2.1 Halving method

To better understand the method, let us first sketch a proof of Theorem 3 by halving.

Let (X, \mathcal{S}) be a set system with VC-dimension at most d . For the sake of simplicity, assume that $n := |X|$ is a power of 2. Let A_1 be a uniform random sample from X of size $n/2$. Define the approximation error $\varepsilon(A_1, X, \mathcal{S})$ of A_1 with respect to (X, \mathcal{S}) as

$$\varepsilon(A_1, X, \mathcal{S}) := \max_{S \in \mathcal{S}} \left| \frac{|S \cap A_1|}{|A_1|} - \frac{|S|}{n} \right|.$$

By Chernoff's bound (Equation (2.2)), for any fixed set $S \in \mathcal{S}$

$$\mathbb{P} \left[\left| |S \cap A_1| - \frac{|S|}{2} \right| \leq \sqrt{2n \log \frac{2}{\gamma}} \right] \geq 1 - \gamma.$$

Combining this with the union bound (Equation (2.3)), we get that

$$\mathbb{P} \left[\max_{S \in \mathcal{S}} \left| |S \cap A_1| - \frac{|S|}{2} \right| \leq \sqrt{2n \log \frac{2|\mathcal{S}|}{\gamma}} \right] \geq 1 - \gamma.$$

Therefore, with probability at least $1/2$,

$$\varepsilon(A_1, X, \mathcal{S}) = \frac{2}{n} \cdot \max_{S \in \mathcal{S}} \left| |S \cap A_1| - \frac{|S|}{2} \right| \leq \frac{2}{n} \cdot \sqrt{2n \log (4|\mathcal{S}|)} \leq \sqrt{\frac{8}{n} \ln \left(4 \left(\frac{en}{d} \right)^d \right)},$$

where the last bound follows from the Sauer-Shelah lemma (Lemma 6). While the expected approximation error is very small $O\left(\sqrt{\frac{d}{n} \cdot \ln n}\right)$, the size of A_1 is rather large—we have only halved the initial input size.

We can reduce the approximation size by halving A_1 again: let A_2 be a random subset from A_1 of size $|A_1|/2$. By our previous argument, with probability at least $3/4$, A_2 is an $\varepsilon(A_2, A_1, \mathcal{S}|_{A_1})$ -approximation of $(A_1, \mathcal{S}|_{A_1})$ with

$$\varepsilon(A_2, A_1, \mathcal{S}|_{A_1}) \leq \sqrt{\frac{16}{n} \ln \left(8 \left(\frac{en}{2d} \right)^d \right)}.$$

We can easily bound the approximation error of A_2 with respect to (X, \mathcal{S}) using the triangle inequality:

$$\begin{aligned} \max_{S \in \mathcal{S}} \left| \frac{|A_2 \cap S|}{|A_2|} - \frac{|S|}{n} \right| &= \max_{S \in \mathcal{S}} \left| \frac{|A_2 \cap S|}{|A_2|} - \frac{|A_1 \cap S|}{|A_1|} + \frac{|A_1 \cap S|}{|A_1|} - \frac{|S|}{n} \right| \\ &\leq \max_{S \in \mathcal{S}} \left(\left| \frac{|A_2 \cap S|}{|A_2|} - \frac{|A_1 \cap S|}{|A_1|} \right| + \left| \frac{|A_1 \cap S|}{|A_1|} - \frac{|S|}{n} \right| \right) \\ &\leq \max_{S \in \mathcal{S}} \left| \frac{|A_2 \cap S|}{|A_2|} - \frac{|A_1 \cap S|}{|A_1|} \right| + \max_{S \in \mathcal{S}} \left| \frac{|A_1 \cap S|}{|A_1|} - \frac{|S|}{n} \right| \\ &\leq \varepsilon(A_2, A_1, \mathcal{S}|_{A_1}) + \varepsilon(A_1, X, \mathcal{S}). \end{aligned}$$

This calculation is summarized in the following well-known property of ε -approximations.

Property 1. *Let (X, \mathcal{S}) be a set system, A_1 be an ε_1 -approximation of (X, \mathcal{S}) and A_2 be an ε_2 -approximation of $(A_1, \mathcal{S}|_{A_1})$. Then A_2 is an $(\varepsilon_1 + \varepsilon_2)$ -approximation of (X, \mathcal{S}) .*

Therefore, A_2 is an $(\varepsilon(A_2, A_1, \mathcal{S}|_{A_1}) + \varepsilon(A_1, X, \mathcal{S}))$ -approximation of (X, \mathcal{S}) with probability at least $1/4$.

We see that each time we halve the size of our approximation, the error increases with an additive term. Given an allowed error ε , we can continue halving until the accumulated approximation error remains at most ε . One can verify that we can continue the process until

$$t(\varepsilon, n) = \Omega \left(\log \frac{\varepsilon^2 n}{d \ln \frac{1}{\varepsilon}} \right)$$

halving steps. This implies that a random sample of size

$$\frac{n}{2^{t(\varepsilon, n)}} = O \left(\frac{d}{\varepsilon^2} \ln \frac{1}{\varepsilon} \right)$$

is an ε -approximation of (X, \mathcal{S}) with positive probability.

The first idea leading to improved constructions is to follow the above halving method, but at each step, halve along more carefully selected sets. Recall, that the approximation error of a set $A \subset X$ of size $n/2$ is defined as

$$\varepsilon(A, X, \mathcal{S}) = \frac{1}{n} \cdot \max_{S \in \mathcal{S}} |2|S \cap A| - |S||. \quad (2.6)$$

The next topic that we study is how small $\varepsilon(A, X, \mathcal{S})$ can be for a set $A \subseteq X$ of size $n/2$. As we will see, this question can be reduced to the classical *combinatorial discrepancy problem*.

2.2.2 Combinatorial discrepancy

Given a set system (X, \mathcal{S}) , the combinatorial discrepancy problem asks for a two-coloring $\chi: X \rightarrow \{-1, 1\}$, that minimizes the *discrepancy*

$$\text{disc}_{\mathcal{S}}(\chi) = \max_{S \in \mathcal{S}} \left| \sum_{x \in S} \chi(x) \right|.$$

In words, the discrepancy of χ measures how evenly it colors the ranges in \mathcal{S} . For instance, if $\text{disc}_{\mathcal{S}}(\chi) = 0$ then in every set $S \in \mathcal{S}$ precisely half of the elements have color 1, while the other half has color -1 assigned. We can express the discrepancy of χ as

$$\begin{aligned} \text{disc}_{\mathcal{S}}(\chi) &= \max_{S \in \mathcal{S}} \left| |\chi^{-1}(1) \cap S| - |\chi^{-1}(-1) \cap S| \right| \\ &= \max_{S \in \mathcal{S}} \left| |\chi^{-1}(1) \cap S| - |S| + \underbrace{|S| - |\chi^{-1}(-1) \cap S|}_{|\chi^{-1}(1) \cap S|} \right| \\ &= \max_{S \in \mathcal{S}} |2|\chi^{-1}(1) \cap S| - |S||. \end{aligned}$$

Compare this formulation with Equation (2.6). We see that the approximation error of A is precisely $1/n$ -times the discrepancy of the coloring

$$\chi_A(x) = \begin{cases} 1 & \text{if } x \in A \\ -1 & \text{if } x \in X \setminus A \end{cases}.$$

Thus, a set $A \subset X$, $|A| = n/2$ with low approximation error can be used to construct a coloring with low discrepancy.

On the other hand, let χ be a 2-coloring of X and let A_{χ} be a set of $n/2$ elements from the larger color class of χ say, without loss of generality, $A_{\chi} \subseteq \chi^{-1}(1)$. Assume that

\mathcal{S} contains X as a range, which implies

$$||A_\chi| - |\chi^{-1}(1)|| = \frac{1}{2}|n - 2|\chi^{-1}(1)|| \leq \frac{\text{disc}_{\mathcal{S}}(\chi)}{2}.$$

In particular, we get that

$$\begin{aligned} \max_{S \in \mathcal{S}} |2|S \cap A_\chi| - |S|| &\leq \max_{S \in \mathcal{S}} \left| 2|S \cap \chi^{-1}(1)| - \frac{|S|}{2} \right| + 2 \max_{S \in \mathcal{S}} ||S \cap \chi^{-1}(1)| - |S \cap A_\chi|| \\ &\leq \text{disc}_{\mathcal{S}}(\chi) + \text{disc}_{\mathcal{S}}(\chi), \end{aligned}$$

and thus, the approximation error of A_χ with respect to (X, \mathcal{S}) satisfies

$$\varepsilon(A_\chi, X, \mathcal{S}) \leq \frac{2}{n} \cdot \text{disc}_{\mathcal{S}}(\chi).$$

In conclusion, the problem of finding a set of $|X|/2$ elements with low approximation error is essentially equivalent to the problem of finding a low-discrepancy coloring of (X, \mathcal{S}) . This link between approximations and low-discrepancy colorings was first established by [Matoušek et al. \(1991\)](#) and can be summarized as follows.

Lemma 7. *Let (X, \mathcal{S}) be a set system such that \mathcal{S} contains the entire X as a range.*

- a) *If χ is a coloring with discrepancy $\text{disc}_{\mathcal{S}}(\chi) = \delta$, then there is a set A_χ of size $|X|/2$, which is a $(2\delta/n)$ -approximation of (X, \mathcal{S}) .*
- b) *If $A \subset X$ is a δ -approximation of (X, \mathcal{S}) of size $n/2$, then there is a coloring $\chi_A : X \rightarrow \{-1, 1\}$ with discrepancy $n\delta$.*

Let us return to the halving technique. Let δ be a function such that any $Y \subseteq X$ has a coloring with discrepancy at most $\delta(|Y|)$ in time $T(|Y|)$. Then, if at the i^{th} halving step, we use a coloring with discrepancy $\delta(n/2^{i-1})$ as described above, we get that for any integer $t \geq 1$, an ε_t -approximation of size $n/2^t$ can be constructed in time $T(n) + T(n/2) + \dots + T(n/2^{t-1})$ with

$$\varepsilon_t \leq \frac{2}{n} \left(\delta(n) + 2\delta\left(\frac{n}{2}\right) + \dots + 2^{t-1}\delta\left(\frac{n}{2^{t-1}}\right) \right).$$

As we have already mention in Chapter 1, if the dual VC-dimension of (X, \mathcal{S}) is at most D , then one can find a two-coloring of X with discrepancy $O\left(\sqrt{n^{1-1/D} \ln m}\right)$ in time $O(n^3 m)$ ([Matoušek et al., 1991](#)). Thus, if (X, \mathcal{S}) has dual VC-dimension at most D , then one can find ε -approximations of size

$$\left(\frac{c_D \ln \frac{1}{\varepsilon}}{\varepsilon^2} \right)^{\frac{D}{D+1}} \tag{2.7}$$

in time $O(n^3 m)$. In Chapter 4, we present an improved $\tilde{O}(n^{1/D} m)$ time algorithm for constructing colorings and approximation with the same guarantees.

Chapter 3

Introduction française

Avec les récents développements techniques, le traitement et le stockage de grands ensembles de données sont devenus de plus en plus importants. En conséquence, la représentation de données massives par des résumés compacts est devenue un thème central dans de nombreux domaines de recherche modernes. Cette thèse contribue aux outils fondamentaux de la réduction combinatoire des données : la dimension VC, les ε -réseaux (ε -nets), les ε -approximations, les colorations à faible discrédance et les couplages à faible nombre de croisements (par rapport aux ensembles).

3.1 Approximations de systèmes d'ensembles

Pour motiver la notion d'approximations, considérons le problème fondamental suivant issu du traitement des données géométriques.

Comptage du demi-espace. Étant donné un ensemble P de n points X dans \mathbb{R}^d et un demi-espace H , comptez le nombre de points dans $P \cap H$.

Une solution simple a une complexité linéaire: nous pouvons simplement tester pour chaque $p \in P$ s'il est contenu dans H . Ce temps d'exécution trivial de $O(n)$ peut être considérablement amélioré à $O(\log n)$ en prétraitant P dans une structure de données appropriée (mais plutôt complexe). De nombreuses recherches ont étudié les différents trade-offs temps/espace pour le comptage du le demi-espace, voir l'étude de [Agarwal \(2017\)](#).

Si nous nous autorisons une erreur de 0,1%, il existe une solution beaucoup plus simple: nous pouvons remplacer P par un ensemble plus petit A , qui peut être utilisé

pour compter approximativement le nombre de points dans tout demi-espace H . En particulier, un échantillon aléatoire uniforme $A \subset X$ de taille *constante* satisfait à la condition suivante

$$|A \cap H| \cdot \frac{|P|}{|A|} \in \left[|P \cap H| - 0.001 \cdot |P|, |P \cap H| + 0.001 \cdot |P| \right]$$

avec une forte probabilité. Ceci implique que pour tout demi-espace H , il suffit de compter $|A \cap H|$ (ce que nous pouvons faire en un temps constant avec l'algorithme naïf de temps linéaire) pour obtenir une approximation de $|P \cap H|$.

Dans l'exemple du comptage approximatif dans le demi-espace, la propriété clé que l'échantillon A satisfait est que pour *n'importe quel* demi-espace H , la proportion de points dans A qui se trouvent dans H est proche de la proportion de points dans P qui sont dans H . Formellement, A est un sous-ensemble de P tel que pour *n'importe quel* demi-espace H

$$\left| \frac{|P \cap H|}{|P|} - \frac{|A \cap H|}{|A|} \right| \leq 0.001. \quad (3.1)$$

Nous formalisons cette propriété pour une classe d'objets plus abstraite: les systèmes d'ensembles. Un *système d'ensembles* est une paire (X, \mathcal{S}) , où X est un ensemble d'éléments et \mathcal{S} est un ensemble de sous-ensembles de X . Par exemple, le problème du comptage des demi-espaces peut être modélisé par un système d'ensembles (P, \mathcal{H}) , où \mathcal{H} est constitué des sous-ensembles de P qui peuvent être obtenus par intersection avec un demi-espace, nous disons que ce sont les sous-ensembles *induits* par \mathcal{H} .

La propriété formulée dans Equation (3.1) peut être généralisée aux systèmes d'ensembles abstraits comme suit. Étant donné un système d'ensembles (X, \mathcal{S}) et un paramètre $0 < \varepsilon < 1$, on dit que $A \subseteq X$ est une ε -*approximation* de (X, \mathcal{S}) si pour chaque $S \in \mathcal{S}$,

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

3.2 Dimension VC

Les travaux pionniers de [Vapnik and Chervonenkis \(1971\)](#) ont montré que si (X, \mathcal{S}) a une complexité bornée, alors il admet des ε -approximations de taille indépendante de $|X|$ et de $|\mathcal{S}|$, de plus, une telle approximation peut être construite par échantillonnage

uniforme. Ils ont mesuré la complexité d'un système d'ensembles par sa *dimension VC*, qui est définie comme la taille du plus grand $Y \subseteq X$ tel que

$$|\{Y \cap S : S \in \mathcal{S}\}| = 2^{|Y|}.$$

La dimension VC est le fondement de nombreux résultats théoriques en géométrie computationnelle (Matoušek, 2013) et en théorie de l'apprentissage (Shalev-Shwartz and Ben-David, 2014). Plus important encore, elle joue un rôle central dans les garanties d'échantillonnage uniforme et est souvent requise comme paramètre d'entrée des algorithmes. Les limites fondamentales des approximations par échantillonnage uniforme ont été décrites par Li et al. (2001). Ils ont montré que si (X, \mathcal{S}) a une dimension VC d'au plus d , alors un échantillon aléatoire uniforme $A \subseteq X$ de taille $O(d/\varepsilon^2)$ est une ε -approximation avec une probabilité positive¹, de plus cette borne est optimale pour les garanties d'échantillonnage uniforme.

3.2.1 k -union et k -intersection

Alors que pour les systèmes d'ensembles géométriques simples (par exemple ceux induits par les demi-espaces), on peut déterminer la dimension VC à l'aide d'arguments élémentaires, de nombreuses applications nécessitent des systèmes d'ensembles plus complexes. Un problème ouvert depuis 1989 était la borne de la dimension VC des unions et des intersections de k demi-espaces en haute dimension. La k -union d'une collection d'ensembles \mathcal{S} est définie comme $\mathcal{S}^{k\cup} = \{S_1 \cup \dots \cup S_k : S_1, \dots, S_k \in \mathcal{S}\}$. De façon similaire, on peut définir la k -intersection de \mathcal{S} , dénotée par $\mathcal{S}^{k\cap}$, comme la collection d'ensembles constituées de tous les sous-ensembles dérivés de l'intersection commune d'au plus k ensembles de \mathcal{S} . L'article fondateur de Blumer et al. (1989) a montré que pour tout système d'ensembles (X, \mathcal{S}) , $d_{VC}(\mathcal{S}^{k\cup})$ et $d_{VC}(\mathcal{S}^{k\cap})$ sont $O(d_{VC}(\mathcal{S}) \cdot k \log k)$. Ils ont également étudié la question de savoir si cette borne supérieure générale est optimale pour le cas géométrique de base, où \mathcal{S} est constitué d'ensembles induits par des demi-espaces dans \mathbb{R}^d . Il s'est avéré que pour les dimensions 2 et 3, une borne améliorée de $O(k)$ tient (Blumer et al., 1989; Dobkin and Gunopulos, 1995). La conjecture dominante était que le facteur $\log k$ peut également être supprimé pour des dimensions $d \geq 4$.

Nous prouvons qu'en dimension $d \geq 4$, le facteur $\log k$ est en fait nécessaire. En particulier, nous décrivons une construction déterministe de la borne inférieure, qui montre que la borne $O(dk \log k)$ est optimale pour les k -unions ou k -intersections de demi-espaces en dimensions au moins 4.

¹En Chapitre 7, nous donnons une preuve simple de ce résultat de Li et al. (2001).

CONTRIBUTION

Il existe une constante universelle c telle que pour tout nombre entier positif k et $d \geq 4$, il existe un ensemble P d'au moins $c \cdot dk \log k$ points dans \mathbb{R}^d tel que le système d'ensembles \mathcal{H} induit sur P par les demi-espaces satisfasse

$$d_{\text{VC}}(P, \mathcal{H}^{k \cup}) = d_{\text{VC}}(P, \mathcal{H}^{k \cap}) = |P|.$$

Les systèmes d'ensembles induits par des combinaisons booléennes d'objets géométriques sont également utilisés dans d'autres domaines. Par exemple, l'un des systèmes d'ensembles fondamentaux en géométrie algorithmique est défini sur un ensemble d'hyperplans \mathcal{H} dans \mathbb{R}^d avec des ensembles

$$\Delta(\mathcal{H}) = \left\{ \mathcal{H}' \subseteq \mathcal{H} : \exists \text{ un simplex ouvert } S \text{ de dimension d'ordre } d \text{ dans } \mathbb{R}^d \text{ tel que } H \in \mathcal{H}' \text{ si et seulement si } H \text{ intersecte } S \right\}.$$

L'importance du système d'ensembles $(\mathcal{H}, \Delta(\mathcal{H}))$ provient du fait qu'il s'agit du système d'ensemble utilisé dans la construction de *cuttings* par l'échantillonnage aléatoire (Chazelle and Friedman, 1990). Les cuttings sont l'outil central des algorithmes de localisation rapide de points et ont été étudiés en détail récemment par Ezra et al. (2020), qui a fourni les meilleures bornes à ce jour pour la dimension VC de $\Delta(\mathcal{H})$.

Lemma 8 (Ezra et al. 2020). *Pour $d \geq 9$, nous avons*

$$d(d+1) \leq d_{\text{VC}}(\Delta(\mathcal{H})) \leq 5 \cdot d^2 \log d.$$

Nous montrons que la borne supérieure de Lemma 8 est asymptotiquement optimale en établissant une borne de $\Omega(d^2 \log d)$ sur la dimension VC de $\Delta(\mathcal{H})$ pour $d \geq 4$.

CONTRIBUTION

Il existe une constante universelle c telle que pour tout nombre entier positif k et $d \geq 4$, il existe un ensemble \mathcal{H} de $c \cdot d^2 \log d$ hyperplans dans \mathbb{R}^d tel que

$$d_{\text{VC}}(\mathcal{H}, \Delta(\mathcal{H})) = |\mathcal{H}|.$$

3.3 Au-delà de l'échantillonnage uniforme

Rappelons que si (X, \mathcal{S}) a une dimension VC d'au plus d , alors un échantillon aléatoire uniforme $A \subseteq X$ de taille $O(d/\varepsilon^2)$ est une ε -approximation avec une probabilité positive, et que cette borne est optimale pour des garanties d'échantillonnage uniforme.

Le principal problème algorithmique que nous étudions dans cette thèse est le développement d'algorithmes efficaces pour construire des ε -approximations de taille sub-quadratique. Une façon possible d'obtenir des ε -approximations de taille $o(d/\varepsilon^2)$ est d'appliquer la technique itérative de division en deux de Matoušek et al. (1991) : la méthode construit itérativement une coloration avec une faible discrédance et procède avec seulement la moitié des éléments qui appartiennent à la même classe de couleur (pour plus de détails voir Section 2.2.1). Étant donné une bicoloration $\chi: X \rightarrow \{-1, 1\}$, la *discrédance* de χ par rapport à \mathcal{S} est définie comme suit

$$\text{disc}_{\mathcal{S}}(\chi) = \max_{S \in \mathcal{S}} \left| \sum_{x \in S} \chi(x) \right|.$$

Dans ce qui suit, nous utiliserons la notation $n = |X|$ et $m = |\mathcal{S}|$. Il est connu que pour tout système d'ensembles, une bicoloration aléatoire χ de X satisfait à $\text{disc}_{\mathcal{S}}(\chi) = O(\sqrt{n \ln m})$ avec une probabilité d'au moins $1/2$. De plus, si aucune autre propriété n'est connue pour \mathcal{S} et $m = \Omega(n^2)$, alors cette borne est essentiellement optimale. Il n'est pas surprenant qu'une coloration aléatoire avec une discrédance de $\tilde{O}(\sqrt{n})$ n'est pas suffisante pour améliorer la garantie d'échantillonnage uniforme des approximations: si nous voulons utiliser la méthode itérative de division en deux pour construire des approximations de taille $o(d/\varepsilon^2)$, nous devons trouver des colorations dont la discrédance est $o(\sqrt{n})$. Une série de recherches a démontré que si l'on suppose une structure supplémentaire sur (X, \mathcal{S}) , alors X admet des colorations de discrédance d'ordre inférieur.

Dimension VC limitée. Matoušek (1995) a montré que si $d_{\text{VC}}(X, \mathcal{S}) \leq d$, alors X a une coloration avec une discrédance $O(\sqrt{n^{1-1/d}})$. Cela implique, par la méthode de division en deux, que les systèmes d'ensembles de dimension VC au plus égale à d admettent des ε -approximations de taille

$$O\left(\left(\frac{1}{\varepsilon^2}\right)^{\frac{d}{d+1}}\right).$$

Bien que ce résultat de coloration implique des garanties améliorées, sa preuve est uniquement existentielle—elle ne conduit pas à un algorithme permettant de construire une ε -approximation de taille subquadratique.

Dimension VC duale limitée. La situation pratique est meilleure si l'on suppose la finitude de la dimension VC duale. La *dimension VC duale* de (X, \mathcal{S}) est définie comme la dimension VC de son système d'ensembles dual (\mathcal{S}, X^*) . Le système d'ensembles dual de (X, \mathcal{S}) est le système d'ensembles sur \mathcal{S} , où chaque point de $x \in X$ induit un ensemble

constitué des éléments de \mathcal{S} qui contiennent x . Formellement, c'est une paire (\mathcal{S}, X^*) , où $X^* = \{R_x : x \in X\}$ et $R_x = \{S \in \mathcal{S} : x \in S\}$. Si la dimension VC duale de (X, \mathcal{S}) est au plus égale à D , alors on peut trouver un bicolore de X avec une discrédance $O\left(\sqrt{n^{1-1/D} \ln m}\right)$ en temps $O(n^3 m)$ (Matoušek et al., 1991). Ainsi, par la méthode de division en deux, si (X, \mathcal{S}) a une dimension VC duale d'au plus D , alors on peut trouver des approximations de ε de taille

$$\left(\frac{c_D \ln \frac{1}{\varepsilon}}{\varepsilon^2}\right)^{\frac{D}{D+1}}$$

en temps $O(n^3 m)$. La structure clé utilisée pour construire une coloration avec une discrédance de $\tilde{O}\left(\sqrt{n^{1-1/D}}\right)$ est un couplage de X avec un nombre de croisement faible, qui est aussi le goulot d'étranglement algorithmique nécessitant $O(n^3 m)$ de temps pour être trouvé.

La principale contribution algorithmique de cette thèse est un algorithme amélioré pour construire des couplages avec un nombre de croisement faible. Dans la section suivante, nous nous concentrons sur ce problème : nous donnons une définition formelle, nous présentons les résultats algorithmiques précédents, et nous énonçons les garanties de notre méthode.

3.4 Couplages avec un nombre de croisement faible

Un *couplage* d'un ensemble X est une partition de X en paires, que nous appelons des arêtes. Nous disons qu'un ensemble $S \in \mathcal{S}$ croise une arête $\{x, y\}$ si $|S \cap \{x, y\}| = 1$. Le nombre de croisements d'un couplage M par rapport à \mathcal{S} est le nombre maximal de paires dans M qui sont croisées par un ensemble dans \mathcal{S} . Les graphes (couplages, chemins couvrants, arbres couvrants) à faible nombre de croisement ont été introduites par Welzl (1988) pour les algorithmes de recherche géométrique (voir aussi Chazelle and Welzl, 1989; Welzl, 1992). Depuis, elles sont devenues des structures clés en géométrie computationnelle (voir par exemple Pach and Agarwal, 2011) et ont trouvé de nombreuses applications dans d'autres domaines tels que la théorie algorithmique des graphes (Ducoffe et al., 2020) ou la théorie de l'apprentissage (Alon et al., 2016).

3.4.1 Constructions précédentes

L'algorithme classique pour construire un couplage à faible nombre de croisements est basé sur la méthode des poids multiplicatifs (Welzl, 1988; Chazelle and Welzl, 1989)

et a une complexité temporelle $O(mn^3)$. Les arbres couvrants à faible croisement des systèmes d'ensembles abstraits peuvent également être trouvés en arrondissant (rounding) la solution d'un LP sur $\binom{n}{2}$ variables et $m+n$ contraintes (Har-Peled, 2009). En combinant cette méthode avec un solveur de LP approximatif efficace (par exemple, Chekuri and Quanrud (2018)), on obtient un algorithme aléatoire en temps $\tilde{O}(mn^2)$. Bien que cette méthode ait un temps d'exécution amélioré, elle est plutôt complexe et nécessite une représentation explicite d'une matrice stockant les croisements entre toutes les $\binom{n}{2}$ arêtes sur X et les ensembles dans \mathcal{S} .

Dans le cas des systèmes d'ensembles géométriques, des bornes améliorées sont possibles grâce au partitionnement spatial. Les meilleurs algorithmes actuels pour les systèmes d'ensembles induits par les demi-espaces construisent récursivement des partitions simpliciales, stockées dans une structure hiérarchique appelée l'arbre de partition, qui donne ensuite à son niveau de base un couplage avec un nombre de croisement faible. Cette approche est utilisée dans le résultat révolutionnaire de Chan (2012) qui a donné un algorithme en temps $O(n \log n)$ pour construire des arbres de partition par rapport aux demi-espaces dans \mathbb{R}^d , ce qui implique ensuite la même chose pour construire des couplages avec un nombre de croisement $O(n^{1-1/d})$. Cependant, l'implémentation pratique du partitionnement spatial dans \mathbb{R}^d , $d > 2$ reste un problème ouvert majeur en calcul géométrique. En particulier, pour $d > 2$, il n'y a pas eu d'implémentations précédentes de couplages à faible croisement, ni de construction de ε -approximations de la taille de $o\left(\frac{d}{\varepsilon^2}\right)$, même pour $d = 3$.

3.4.2 Notre méthode

Nous proposons un algorithme amélioré pour construire un couplage (ou chemin couvrant) avec un faible nombre de croisements.

CONTRIBUTION

Soit (X, \mathcal{S}) un système d'ensembles avec une dimension VC duale d'au plus D . Il existe alors un algorithme aléatoire qui produit un couplage de X avec un nombre de croisements attendu d'au plus

$$\frac{36}{D} \cdot n^{1-1/D} + 30 \ln m \ln n + 26 \ln^2 n$$

avec un nombre attendu de $O\left(mn^{1/D} \ln m \cdot \min\{D, \ln n\} + n^{2+1/D} \ln n\right)$ appels à l'Oracle d'appartenance^a de (X, \mathcal{S}) .

^aUn Oracle d'appartenance de (X, \mathcal{S}) décide si $x \in S$ est vrai pour $x \in X$ et $S \in \mathcal{S}$.

Ce résultat implique que l'on peut construire des colorations avec discrédance $\tilde{O}(\sqrt{n^{1-1/D}})$ et des ε -approximations de taille $O\left((\varepsilon^{-2} \ln \frac{1}{\varepsilon})^{\frac{D}{D+1}}\right)$ en temps $\tilde{O}(n^{1/D}m + n^{2+1/D})$.

Il est important de noter que nous n'utilisons pas le partitionnement spatial, ce qui permet de traiter les systèmes d'ensembles abstraits et les systèmes d'ensembles géométriques dans \mathbb{R}^d (et pas seulement dans \mathbb{R}^2) sans complications supplémentaires. Une implémentation préliminaire de notre algorithme en C++ pour les systèmes d'ensembles géométriques induits sur les points par les demi-espaces dans \mathbb{R}^d est disponible sur [GitHub](#). Il représente environ 200 lignes de code sans aucune structure de données non triviale.

Figure 3.1 montre les couplages construits par notre algorithme sur différents types d'entrées et d'ensembles. Pour référence, la dernière colonne contient des couplages construits de manière aléatoire. Nous soulignons que chacune de ces figures contient exactement 2500 arêtes. Notre méthode, qui est basée uniquement sur l'échantillonnage, donne un couplage qui s'adapte bien à chaque instance spécifique, en particulier, elle conduit à des résultats différents pour les demi-espaces en tant qu'ensembles (colonne de gauche) et les disques en tant qu'ensembles (colonne du milieu).

3.5 Epsilon-réseaux.

Étant donné un système d'ensembles (X, \mathcal{S}) et un paramètre $\varepsilon \in (0, 1)$, un ensemble $N \subseteq X$ est un ε -réseau de (X, \mathcal{S}) si tout ensemble $S \in \mathcal{S}$ de taille $|S| > \varepsilon|X|$ contient au moins un élément de N . Contrairement aux ε -approximations qui approximent uniformément la proportion de points dans chaque ensemble, les ε -réseaux garantissent seulement 1 élément représentatif dans chaque grand ensemble (peu importe s'il contient $2\varepsilon \cdot |X|$ ou $|X|$ points). Pour de nombreuses applications, cette propriété plus modeste est suffisante: par exemple, les ε -réseaux peuvent être utilisés pour dé-randomiser les solveurs de programmation linéaire basés sur l'échantillonnage (Chan, 2018) ou pour créer des cuttings (Clarkson, 1987; Chazelle and Friedman, 1990)—dans ce dernier exemple, on utilise un ε -réseau du système d'ensembles $(\mathcal{H}, \Delta(\mathcal{H}))$ étudié dans Section 3.2.1. Haussler and Welzl (1987) ont montré que si (X, \mathcal{S}) a une dimension VC d'au plus d , alors un échantillon aléatoire de

$$\max \left\{ \frac{8d}{\varepsilon} \log \frac{8d}{\varepsilon}, \frac{4}{\varepsilon} \log \frac{2}{\delta} \right\}$$

points de X forme un ε -réseau de (X, \mathcal{S}) avec une probabilité d'au moins $1 - \delta$. Plus tard, Komlós et al. (1992) ont supprimé la dépendance logarithmique sur d en prouvant que pour tout $0 < \varepsilon < \varepsilon_0$, un échantillon uniforme de

$$\frac{d}{\varepsilon} \log \frac{1}{\varepsilon} + o\left(\frac{d}{\varepsilon} \log \frac{1}{\varepsilon}\right).$$

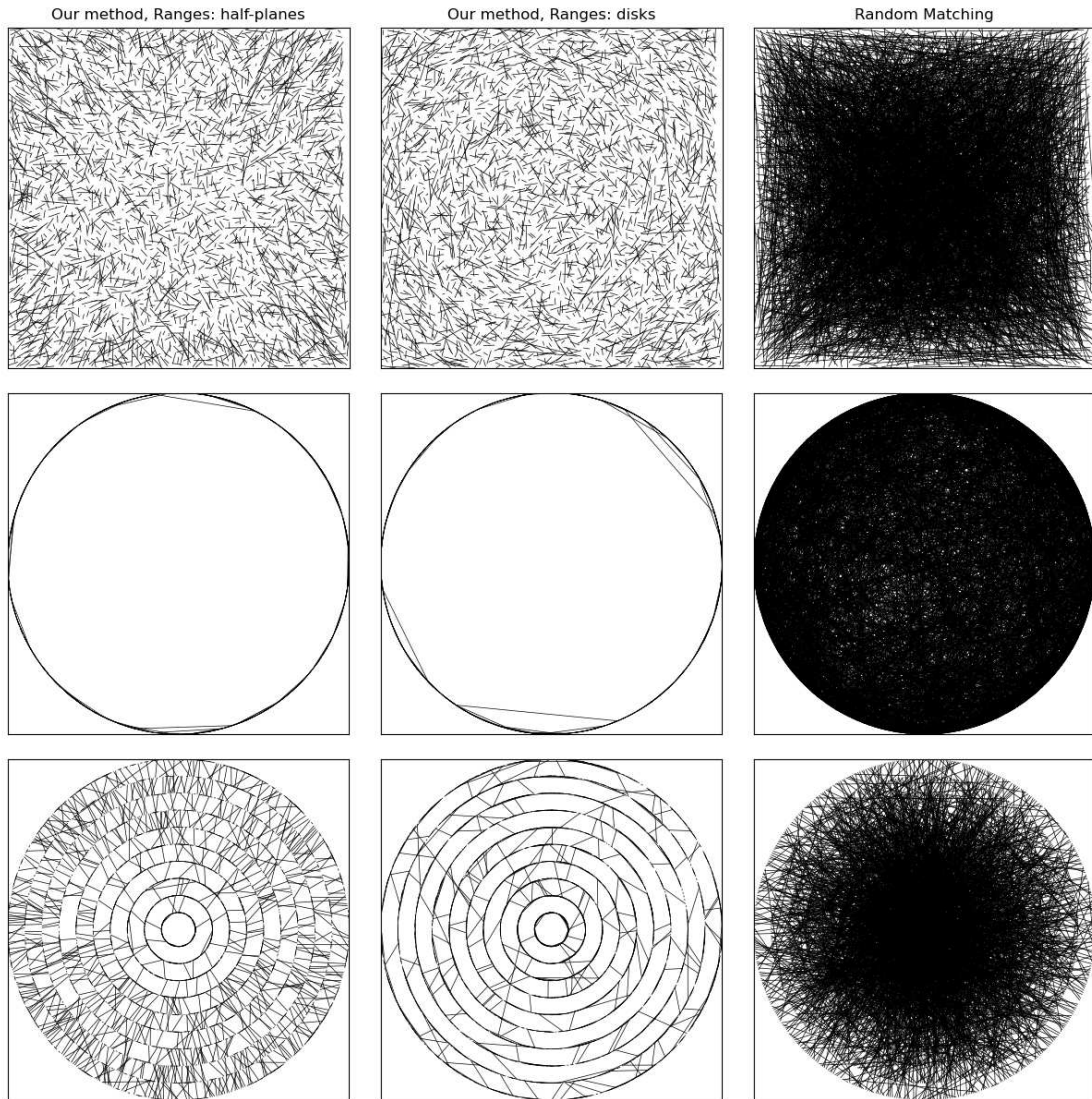


Figure 3.1: Couplages de 5000 points avec différents types d'entrée et d'ensemble. **Colonne de gauche:** notre méthode avec des ensembles de demi-plans. **Colonne du milieu:** notre méthode avec des ensembles de disques. **Colonne de droite:** échantillonnage aléatoire.

éléments de X est un ε -réseau avec une grande probabilité. De plus, ils ont montré que cette borne est asymptotiquement optimale : pour tout $d \geq 2$, il existe un système d'ensembles de dimension VC égale à d , dans lequel tout ε -réseau ($\varepsilon < \varepsilon_0$) a une taille d'au moins $(1 - 2/d)(d/\varepsilon) \log(1/\varepsilon)$. Ainsi, contrairement aux ε -approximations, l'échantillonnage uniforme produit des ε -réseaux optimaux. Dans Chapitre 6, nous prouvons la borne suivante.

CONTRIBUTION

Soit (X, \mathcal{S}) un système d'ensembles avec une dimension VC d'au plus d , $0 < \delta < 1$, et $0 < \varepsilon < \varepsilon_0(d, \delta)$. Alors un échantillon aléatoire uniforme $N \subseteq X$ de taille

$$\frac{d}{\varepsilon} \left(\log \frac{1}{\varepsilon} + \frac{1}{d} \log \frac{1}{\delta} + 2 \log \left(\log \frac{1}{\varepsilon} + \frac{1}{d} \log \frac{1}{\delta} \right) + 5 \right)$$

est un ε -réseau de (X, \mathcal{S}) avec une probabilité d'au moins $1 - \delta$.

Notre dernière contribution est un lemme qui peut potentiellement aider à résoudre le problème de la borne inférieure de la taille des ε -réseaux car il donne un nouvel outil pour établir ces bornes. D'une certaine manière, il peut être considéré comme un complément aux bornes supérieures ci-dessus : nous montrons que les bornes inférieures asymptotiques de la dimension VC de $(X, \mathcal{S}^{k\cup})$ impliquent des bornes inférieures pour les tailles des ε -réseaux dans un sous-système de (X, \mathcal{S}) .

CONTRIBUTION

Soit (X, \mathcal{S}) un système d'ensembles, $\varepsilon \in (0, 1)$ un paramètre, et $f : \mathbb{N} \rightarrow \mathbb{R}$ une fonction qui satisfait $d_{VC}(X, \mathcal{S}^{k\cup}) \geq d_{VC}(X, \mathcal{S}) \cdot k \cdot f(k)$. Alors il existe un sous-ensemble X' de X tel que n'importe quel ε -réseau de $(X', \mathcal{S}|_{X'})$ doit avoir une taille d'au moins

$$\frac{d_{VC}(X', \mathcal{S}|_{X'})}{4\varepsilon} \cdot f\left(\frac{1}{2\varepsilon}\right).$$

Par exemple, la combinaison du résultat ci-dessus avec la borne inférieure pour la dimension VC des k -unions de demi-espaces présentée dans Section 3.2.1, donne immédiatement une borne inférieure de $\Omega(d/\varepsilon \log(1/\varepsilon))$ sur les tailles des réseaux ε dans les systèmes d'ensembles induits par les demi-espaces, récupérant le résultat de [Pach and Tardos \(2013\)](#).

Chapter 4

Improved algorithm for low-discrepancy colorings

Part of the results presented in this chapter were published in the conference paper

Escaping the Curse of Spatial Partitioning:

Matchings With Low Crossing Numbers and Their Applications

37th International Symposium on Computational Geometry (SoCG 2021).

It is a joint work with Nabil Mustafa.

In this chapter, we study the *combinatorial discrepancy problem*, which asks for a two-coloring $\chi: X \rightarrow \{-1, 1\}$, that minimizes the *discrepancy*

$$\text{disc}_S(\chi) = \max_{S \in \mathcal{S}} \left| \sum_{x \in S} \chi(x) \right|.$$

Starting from 1980s, the study of low-discrepancy colorings have been an active area of research which has found many applications in various branches of mathematics and computer science. As it is often termed, the ‘discrepancy method’ can be used to tackle problems arising in combinatorial optimisation, data approximation, computational geometry, or communicational complexity just to name a few. Low-discrepancy colorings are also closely connected to the sample complexity of learning. For instance, in [Bartlett et al. \(2002\)](#), discrepancy of a random balanced coloring is used to construct penalized empirical risk minimization algorithms, leading to improved statistical guarantees. Furthermore, the study of Rademacher complexity—one of the fundamental quantities in learning literature—can be seen as a study of discrepancy of a random coloring. For further details and other examples of applications, we refer the interested reader to dedicated books [Chazelle \(2000\)](#); [Chen et al. \(2014\)](#); [Matoušek \(1999\)](#).

4.1 Previous results

A first bound on the combinatorial discrepancy of (X, \mathcal{S}) follows immediately from Chernoff's bound, which implies that with probability at least $\frac{1}{2}$, a random two-coloring χ of X satisfies $\text{disc}_{\mathcal{S}}(\chi) = O(\sqrt{n \ln m})$. Furthermore, if no additional properties are known for \mathcal{S} and $m = \Omega(n^2)$, then this bound is essentially optimal. We immediately get a randomized algorithm to obtain such a coloring and it is possible to de-randomize the method yielding a deterministic algorithm with running time $O(nm)$ (see Chazelle (2000)).

Spencer (1985) showed that there exists a coloring of X with discrepancy of order $O(\sqrt{n \ln(m/n)})$, which improves the general bound for $m = O(n)$. His original proof only demonstrated the existence of such a coloring and did not give a polynomial-time algorithm for it. This remained a major open problem for 25 years, until a recent breakthrough of Bansal (2010), who gave a randomized polynomial-time SDP rounding algorithm to compute a coloring with discrepancy $O(\sqrt{n \ln(m/n)})$ coloring, which matches the bound of Spencer when $m = O(n)$. Later Lovett and Meka (2015) gave a different combinatorial randomized algorithm for constructing colorings with discrepancy $O(\sqrt{n \ln(m/n)})$ and improved the expected running time to $\tilde{O}(n^3 + m^3)$; see also Rothvoss (2017) for a different proof. More recently, Levy et al. (2017) used the multiplicative weights technique to give a deterministic $O(n^4 m)$ -time algorithm to compute a two-coloring with discrepancy $O(\sqrt{n \ln(m/n)})$ for an arbitrary set system. See also Bansal et al. (2018) for a random-walk algorithm for Banaszczyk's discrepancy bound, with running time $O(n^{3.37\dots} + m^{2.37\dots})$ (the exponent depends on the running time for matrix multiplication).

These guarantees are computationally tight for general set systems where $m = O(n)$, in particular, it was shown by Charikar et al. (2011) that there exists a set system with $m = O(n)$ for which it is NP-hard to decide whether discrepancy zero or $\Omega(\sqrt{n})$.

Set systems with dual shatter function of polynomial growth. A series of research has demonstrated that if one assumes additional structure on \mathcal{S} , then X admits colorings of lower-order discrepancy. A common way to bound the complexity of (X, \mathcal{S}) is through the shatter function of its dual set system, shortly, the dual-shatter function of (X, \mathcal{S}) .

Definition 9 (Dual-shatter function). *For any $\mathcal{R} \subseteq \mathcal{S}$, we say that the elements $x, y \in X$ are equivalent with respect to \mathcal{R} if x belongs to the same sets of \mathcal{R} as y . The dual-shatter function $\pi_{\mathcal{S}}^*$ of (X, \mathcal{S}) is a function, whose value at any $k \in [1, m]$ is defined as the maximum number of equivalence classes on X defined by a k -element subfamily $\mathcal{R} \subseteq \mathcal{S}$.*

In this work, we focus on set systems whose dual-shatter function satisfies $\pi_{\mathcal{S}}^*(k) = O(k^d)$ for some constant d . This class contains several well-studied examples, for instance:

- set system with dual VC-dimension d : this property implies $\pi_{\mathcal{S}}^*(k) \leq \left(\frac{ek}{d}\right)^d$ by the Sauer-Shelah lemma (Lemma 6);
- geometric set systems, where X is a set of n points and each range in \mathcal{S} can be obtained as an intersection of X with a semialgebraic set of bounded complexity. This includes set systems induced by (unions or intersections of) half-spaces, balls, etc. For details see Section 4.7;
- set systems such that any d range has intersection at most c for some constant c (Matoušek, 1997).

The following theorem gives an upper-bound on the combinatorial discrepancy of set systems with $\pi_{\mathcal{S}}^*(k) = O(k^d)$, complemented with a lower-bound from Alon et al. (1999).

Theorem 10 (Alon et al. (1999); Matoušek et al. (1991); Matoušek (1997)). *Let (X, \mathcal{S}) be a finite set system, $n = |X|$, $m = |\mathcal{S}|$, and let d be a constant such that $\pi_{\mathcal{S}}^*(k) = O(k^d)$. Then there exists a polynomial-time algorithm to compute a two-coloring of X with discrepancy $O\left(\sqrt{n^{1-1/d} \ln m}\right)$. Furthermore, for any d , there exists a set system with dual shatter function $\pi_{\mathcal{S}}^*(k) = O(k^d)$ such that any coloring has discrepancy $\Omega\left(\sqrt{n^{1-1/d} \ln n}\right)$.*

If d is considered as a constant, the upper and lower bounds of Theorem 10 match. This follows from the observation that if the dual-shatter function of (X, \mathcal{S}) is $\pi_{\mathcal{S}}^*(k) = O(k^d)$, then $d_{\text{VC}}(X, \mathcal{S}) \leq 2^d$ and thus by the Sauer-Shelah lemma (Lemma 6), $m = |\mathcal{S}| \leq \left(\frac{en}{2^d}\right)^{2^d}$; see (Matoušek, 1999, Chapter 5) for further details.

Practical aspects

The classical proof of the upper-bound in Theorem 10 is in fact constructive. It uses the multiplicative weights update (MWU) technique (see Arora et al. (2012)) as follows. The algorithm maintains a weight function π on \mathcal{S} , with initial weights set to 1. For any pair $\{x, y\} \in X$, let $\Delta_{\mathcal{S}}(x, y)$ denote the set of those sets $S \in \mathcal{S}$ which satisfy $|S \cap \{x, y\}| = 1$ and let

$$\tilde{\pi}(x, y) = \sum_{S \in \Delta_{\mathcal{S}}(x, y)} \pi(S).$$

The algorithm colors two elements of X at a time (for simplicity, we assume that X is even) as follows

for $i = 1, \dots, n/2$

1. Find a pair $\{x_i, y_i\} \in X$ that minimizes $\tilde{\pi}(x, y)$.
2. Set $\chi(x_i) = \begin{cases} 1 & \text{with probability } 1/2 \\ -1 & \text{with probability } 1/2 \end{cases}$, and define $\chi(y_i) = -\chi(x_i)$
3. Remove x_i, y_i from X .
4. Update π by doubling the weight of each set in $\Delta_S(x_i, y_i)$.

The reweighing scheme ensures the key property that for each $S \in \mathcal{S}$,

$$|\{i : S \in \Delta_S(x_i, y_i)\}| = O(n^{1-1/d}). \quad (4.1)$$

This implies, using Chernoff's bound (Equation (2.2)) and the union bound (Equation (2.3)), that

$$\text{disc}_S(\chi) = O\left(\sqrt{n^{1-1/d} \ln |\mathcal{S}|}\right)$$

with probability at least $1/2$ (Matoušek et al., 1991). The algorithmic bottleneck is finding the pair $\{x_i, y_i\}$ that minimizes $\tilde{\pi}$. Using the incidence matrix for \mathcal{S} , this can be done in $O(n^2m)$ steps, and thus the algorithm has overall running time $O(n^3m)$.

Matchings with low crossing numbers. Given a set X , a *matching* M of X is a set of $\lfloor n/2 \rfloor$ disjoint pairs (edges) from X . We say that a range $S \in \mathcal{S}$ *crosses* a pair $\{x, y\} \in M$ if and only if $|S \cap \{x, y\}| = 1$ and define the *crossing number* of M with respect to \mathcal{S} as maximum the number of pairs of M crossed by a single range $S \in \mathcal{S}$. Finding matchings with optimal crossing number is NP-hard, even in the special case of set systems induced by half-planes (Fekete et al., 2008).

Notice that the pairs $\{x_i, y_i\}$ selected by the MWU algorithm form a matching of X . Furthermore, the key property stated in Equation (4.1) means simply that the matching $M = \{\{x_i, y_i\}\}_{i=1}^{n/2}$ has crossing number $O(n^{1-1/d})$ with respect to \mathcal{S} . The study of matchings, along with spanning paths and spanning trees with low crossing number was originally introduced for geometric range searching (Welzl, 1988; Chazelle and Welzl, 1989). Since then, matchings and spanning paths have found applications in various fields including learning theory (Alon et al., 2016) and algorithmic graph theory (Ducoffe et al., 2020).

Another way of creating matchings in abstract set systems can be deduced using Linear Programming. The starting point is the observation that a spanning tree of crossing

number $O(n^{1-1/d})$ can be found by rounding the solution of an LP on $\binom{n}{2}$ variables and $m + n$ constraints (Har-Peled, 2009; Fekete et al., 2008; Giannopoulos et al., 2014). The resulting spanning tree then can be used to construct a matching with crossing number $O(n^{1-1/d})$ in linear time. Combining this with an efficient approximate LP solver (e.g., Chekuri and Quanrud (2018)) leads to a randomized $\tilde{O}(mn^2)$ time algorithm. While this method has an improved running time, it is rather involved and requires an explicit representation of a matrix storing the crossings between all the $\binom{n}{2}$ edges on X and the ranges in \mathcal{S} .

Geometric set systems. For set systems, where X is a set of n points in \mathbb{R}^d and \mathcal{S} consists of subsets of X that are induced by certain geometric objects, improved bounds are made possible using spatial partitioning. For instance, if \mathcal{S} consists of subsets of X that are induced by half-spaces, one can apply the algorithm of Chan (2012) to construct a matching with crossing number $O(n^{1-1/d})$ in time $O(n \ln n)$, which then implies the same running-time for computing coloring with discrepancy $O\left(\sqrt{n^{1-1/d} \ln |\mathcal{S}|}\right)$.

Remark. *The method of Chan recursively constructs simplicial partitions, stored in a hierarchical structure called the partition tree, which at its base level gives a matching with low crossing number.*

While the use of spatial partitioning gives $o(mn^3)$ running times, progress remains blocked in several ways:

- a) Simplicial partitions only exist in certain geometric settings. Indeed, as shown by Alon et al. (1987), they do not always exist in settings satisfying the requirements of Theorem 10 (e.g., the projective plane). Furthermore, spatial partitioning is not possible when dealing with abstract set systems such as those arising in graph theory or learning theory.
- b) Optimal bounds for constructing simplicial partitions are only known for the case of half-spaces; this is one of the main problems left open by Chan (2012). Despite a series of research for semi-algebraic set systems (using linearization, cuttings, and more recently, polynomial partitioning Agarwal et al. (2013)), the bounds are still sub-optimal for polynomials of degree larger than 2, with exponential dependence on the dimension.
- c) There are large constants in the asymptotic notation depending on the dimension d both in the running time as well as the crossing number bounds, due to the use

of cuttings (see Ezra et al. (2020)). For instance, in Chan’s algorithm the constants are quite large—Theorem 3.2 Chan (2012) requires $\delta \leq \frac{1}{d^2}$, $b = 22$ (see Matheny and Phillips (2018)), which then implies that it constructs a spanning tree with a guaranteed crossing number no better than $12 \cdot 22 \cdot d^4 n^{1-1/d}$; this is at least $20000 \cdot n^{1-1/d}$ even for $d = 3$. Furthermore, the actual construction running time is at least $264 \cdot d^2 n \log n$, not counting the typically large constants in the several complex data structures that the algorithm needs (simplex range searching in \mathbb{R}^d with dynamic insertion; see Matheny and Phillips (2018) for a discussion of its practical aspects in \mathbb{R}^2).

- d) Practical implementation of spatial partitioning in \mathbb{R}^d , $d > 2$, even cuttings for hyperplanes, remains an open problem in geometric computing. Cuttings have been implemented in the planar case by Har-Peled (2000), which have then been used recently for computing ε -approximations w.r.t. half-spaces in \mathbb{R}^2 by Matheny and Phillips (2018). In particular, for $d > 2$, we know of no implementations for low-crossing matchings; nor for constructing $o\left(\frac{d}{\varepsilon^2}\right)$ -sized ε -approximations even for half-spaces in \mathbb{R}^3 .

4.2 Our results

The main result of this chapter is an improved construction of low-discrepancy colorings of set systems with polynomially bounded dual-shatter function.

Theorem 11 (Main Theorem). *Let (X, \mathcal{S}) be a finite set system and c, d be constants such that $\pi_{\mathcal{S}}^*(k) \leq c \cdot k^d$. Then there is a randomized algorithm that constructs a coloring χ of X with expected discrepancy at most*

$$3\sqrt{\frac{9c^{1/d}}{2} \cdot n^{1-1/d} \ln m + 19 \ln^2 m \ln n}$$

with at most

$$\min \left\{ \frac{34n^{2+1/d} \ln n}{c^{1/d}} + \frac{25mn^{1/d} \ln(mn)}{c^{1/d}} \cdot \min \{2d, \log n\}, \frac{n^3}{7} + \frac{mn}{2} \right\}$$

calls in expectation to the membership Oracle of (X, \mathcal{S}) .

Remark. *Note that our algorithm always terminates in $O(n^3 + mn)$ Oracle calls (the second term in min above). However, thanks to the use of random sample within each iteration, the expected running time is significantly better for large d (the first term in min above).*

The above result implies improved constructions to many set systems, including geometric ones, see Table 4.1.

LOW-DISCREPANCY COLORINGS				
Set system	Our method		Previous-best	
	Discrepancy	time	Discrepancy	time
arbitrary with $\pi_S^*(k) \leq ck^d$	$(7c^{1/2d} + o(1)) \sqrt{n^{1-1/d} \ln m}$	$\tilde{O}(mn^{1/d} + n^{2+1/d})$ (Main Theorem)	$O(\sqrt{n^{1-1/d} \ln m})$	$\tilde{O}(mn^2)$ Har-Peled; Chekuri and Quanrud
geometric induced by \mathcal{B}_d	$(12d + o(d)) \sqrt{n^{1-1/d} \ln m}$	$\tilde{O}(dn^{2+1/d})$ (Corollary 36)	$O(\sqrt{n^{1-1/d} \ln m})$	$O(n^{3+1/d})$ Har-Peled; Chekuri and Quanrud
geometric induced by $\Gamma_{d,\Delta,s}$	$(24\sqrt{\Delta s} + o(1)) \sqrt{n^{1-1/d} \ln m}$	$\tilde{O}(s\Delta^d(mn^{1/d} + n^{2+1/d}))$ (Corollary 30)	$O(\sqrt{10^d s \Delta n^{1-1/d} \ln m})$	$O(n^{O(d^3)})$ Agarwal et al.
			$O(\sqrt{\Delta s n^{1-1/d} \ln m})$	$\tilde{O}(s\Delta^d mn^2)$ Har-Peled; Chekuri and Quanrud
geometric induced by \mathcal{H}_d	$(8d + o(d)) \sqrt{n^{1-1/d} \ln m}$	$\tilde{O}(dn^{2+1/d})$ (Corollary 35)	$\geq d^2 \sqrt{n^{1-1/d} \ln m}$	$\tilde{O}(d^2 n)$ Chan

Table 4.1: Summary of our results for set systems (X, \mathcal{S}) with $n = |X|$, $m = |\mathcal{S}|$, $n \leq m$, and $d \geq 2$. We use the notation $\pi_S^*(\cdot)$ for the dual shatter function of (X, \mathcal{S}) , \mathcal{H}_d for half-spaces in \mathbb{R}^d , \mathcal{B}_d for balls in \mathbb{R}^d , and $\Gamma_{d,\Delta,s}$ for semialgebraic ranges in \mathbb{R}^d described by at most s equations of degree at most Δ (see Section 4.7).

Our method rests on the following three key ideas:

1. We replace the bottleneck algorithmic step of finding a light edge in the multiplicative weights update technique by simply sampling an edge according to a carefully maintained distribution. In particular, we maintain weights not only on the sets in \mathcal{S} , but also on $\binom{X}{2}$. At each iteration we sample an edge e and a set S according to the current weights. Then we color the endpoints of e and update the weights by *doubling* the weight of each set that crosses e and *halving* the weight of each edge that is crossed by S . The idea of maintaining ‘primal-dual’ weights has been used earlier to approximately solve matrix games (Grigoriadis and Khachiyan, 1995) and in geometric optimization (Agarwal and Pan, 2014).
2. In our case, the process is more elaborate as we are constructing a coloring at the same time as reweighing. Therefore, at the end of each iteration, as we color the endpoints e , we are forced to set the weights of e and all edges adjacent to e to 0. This breaks down the reweighing scheme, as the removal of the edges amplifies the error introduced in later iterations and thus our maintained weights degrade over time. However, we prove that restarting the algorithm by ‘resetting’ all the weights a logarithmic number of times suffices to ensure the required low crossing numbers.

3. This still does not get us to our goal as updating the weights of *all* edges and sets crossing the randomly picked set and edge would be too expensive. Instead, we show that updating the weights of a *uniform* sample of $\tilde{O}(n^{1+1/d})$ edges and $\tilde{O}(m/n^{1-1/d})$ sets at each iteration is sufficient for our purposes. The key observation here is that the standard multiplicative weights proof has an additive smaller-order term; we take advantage of this gap to improve the running time at the cost of amplifying this term, just enough so that it is still within a constant factor of the optimal solution.

Importantly, our algorithm does not use spatial partitioning, which makes it possible to handle abstract set-systems, and geometric set systems in \mathbb{R}^d (not only in \mathbb{R}^2) without additional complications. The only black-box needed is the membership Oracle that returns for a given $x \in X$ and $S \in \mathcal{S}$, if $x \in S$. The time complexity of this operation depends on the precise way (X, \mathcal{S}) is given; typically this is independent of $|X|$ and $|\mathcal{S}|$ (using indexing, hashing). A preliminary multi-threaded implementation of our algorithm in C++ for set systems induced on points by half-spaces in \mathbb{R}^d is available on [GitHub](#). It is approximately 200 lines of basic code without any non-trivial data-structures. Our experimental setup and a preliminary empirical evaluation of our algorithm is presented in Section 4.8.

4.3 General framework

We prove our results under a more general assumption, which uses matchings with low crossing number.

Assumption ($\text{MAINASSUMPTION}(a, b, \gamma)$). (X, \mathcal{S}) is a finite set system with $m \geq n$, $m \geq 34$, and any $Y \subseteq X$ has a matching with crossing number at most $a|Y|^\gamma + b$ with respect to \mathcal{S} .

This class indeed includes set systems with polynomially bounded dual-shatter function:

Lemma 12. *If c, d are such that (X, \mathcal{S}) has dual shatter function $\pi_{\mathcal{S}}^*(k) \leq ck^d$ for all $k \in [1, n]$, then (X, \mathcal{S}) satisfies $\text{MAINASSUMPTION}(a, b, \gamma)$ with parameters $a = \frac{(2c)^{1/d}}{2 \ln 2(1-1/d)}$, $b = \frac{\ln m}{\ln 2}$, and $\gamma = 1 - 1/d$.*

Proof. To see this, one can use the following classical theorem which states that set systems with polynomially bounded dual shatter function possess matchings with sublinear crossing number ([Matoušek, 1999](#), Chap. 5.4):

Claim 1. *Let (X, \mathcal{S}) be a set system and c, d be constants such that $\pi_{\mathcal{S}}^*(k) \leq ck^d$ for all $k \in [1, n]$. Then there is a matching of X such that any set $S \in \mathcal{S}$ crosses at most*

$$\frac{(2c)^{1/d}}{2 \ln 2(1 - 1/d)} n^{1-1/d} + \frac{\ln m}{\ln 2}$$

edges of the matching.

Observe that by definition, for any $Y \subseteq X$, the dual shatter function of $(Y, \mathcal{S}|_Y)$ is upper-bounded by the dual shatter function of (X, \mathcal{S}) . Thus Claim 1 implies that any $Y \subseteq X$ has a matching with crossing number at most $\frac{(2c)^{1/d}}{2 \ln 2(1-1/d)} |Y|^{1-1/d} + \frac{\ln m}{\ln 2}$ with respect to \mathcal{S} . \square

We prove the following result on low-discrepancy colorings of set systems satisfying $\text{MAINASSUMPTION}(a, b, \gamma)$, which together with Lemma 12 immediately imply **Main Theorem**.

Theorem 13. *Let (X, \mathcal{S}) be a set system, $n = |X|$, $m = |\mathcal{S}|$ with $m \geq \max\{n, 34\}$ that satisfies $\text{MAINASSUMPTION}(a, b, \gamma)$. The algorithm **LOWDISCCOLOR** $((X, \mathcal{S}), a, b, \gamma)$ constructs a coloring χ of X of with expected discrepancy at most*

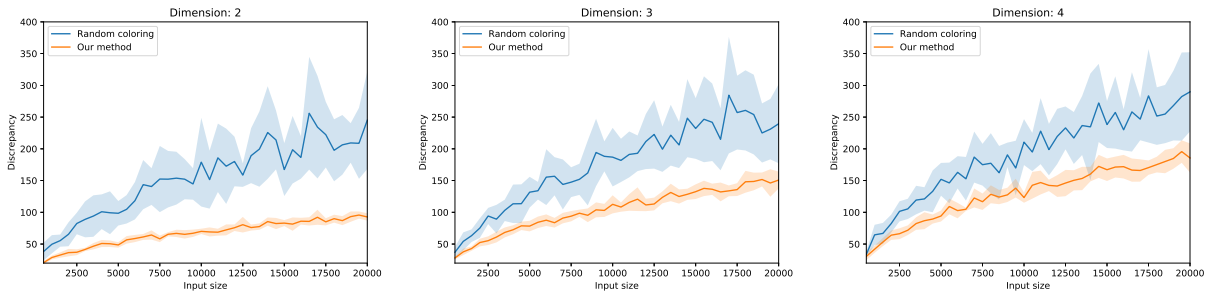
$$3\sqrt{\frac{an^\gamma \ln m}{\gamma} + \frac{b \ln m \log n}{2} + 12 \ln^2 m \log n},$$

with an expected number of Oracle calls at most

$$\min \left\{ \frac{24n^{3-\gamma} \ln n}{a} + \frac{18mn^{1-\gamma} \ln mn}{a} \cdot \min \left\{ \frac{2}{1-\gamma}, \log n \right\}, \frac{n^3}{7} + \frac{mn}{2} \right\}.$$

The algorithm **LOWDISCCOLOR** is presented in Algorithm 2. Our method can be used directly to construct a matching with low crossing number, thus to avoid repetition, we present the Algorithm 2 using a subroutine to build matchings. This subroutine is presented in Algorithm 3.

The figures below show the average discrepancies of the coloring returned by **LOWDISCCOLOR** with respect to half-spaces in dimensions 2, 3, and 4. We performed 10 repetitions of our method and random coloring. The shaded areas denote ± 1 standard deviation from the mean.



Algorithm 2: LOWDISCCOLOR $((X, \mathcal{S}), a, b, \gamma)$

```

 $n \leftarrow |X|$ 
 $(e_1, \dots, e_{\lceil n/2 \rceil}) \leftarrow \text{BUILDMATCHING}((X, \mathcal{S}), a, b, \gamma)$ 
for  $i = 1, \dots, \lfloor n/2 \rfloor$  do
     $\{x_i, y_i\} \leftarrow \text{endpoints}(e_i)$ 
     $\chi(x_i) = \begin{cases} 1 & \text{with probability } 1/2 \\ -1 & \text{with probability } 1/2 \end{cases}$ 
     $\chi(y_i) = -\chi(x_i)$ 
if  $n$  is odd then
     $\chi(x_{\lceil n/2 \rceil}) = 1$  // by construction,  $e_{\lceil n/2 \rceil}$  is a loop-edge
return  $\chi$ 

```

4.3.1 Matchings with low crossing number

Now we present the subroutine **BUILDMATCHING** and state the running time and crossing number guarantees of its output.

Theorem 14. *Let (X, \mathcal{S}) be a set system, $n = |X|$, $m = |\mathcal{S}|$ with $m \geq n$, which satisfies **MAINASSUMPTION** (a, b, γ) . Then **BUILDMATCHING** $((X, \mathcal{S}), a, b, \gamma)$ returns a matching with expected crossing number at most*

$$\frac{3a}{\gamma} n^\gamma + \frac{3b \log n}{2} + 18 \ln(mn) \log n$$

with an expected number of Oracle calls at most

$$\min \left\{ \frac{24n^{3-\gamma} \ln n}{a} + \frac{18mn^{1-\gamma} \ln mn}{a} \cdot \min \left\{ \frac{2}{1-\gamma}, \log n \right\}, \frac{n^3}{7} + \frac{mn}{2} \right\}.$$

Corollary 15. *Let (X, \mathcal{S}) be a set system and c, d be constants such that $\pi_{\mathcal{S}}^*(k) \leq ck^d$ for all $k \in [1, n]$. Then **BUILDMATCHING** $((X, \mathcal{S}), \frac{(2c)^{1/d}}{2 \ln 2(1-1/d)}, \frac{\ln m}{\ln 2}, 1 - \frac{1}{d})$ returns a matching $\{e_1, \dots, e_{n/2}\}$ with expected crossing number at most*

$$13c^{1/d} \cdot n^{1-1/d} + 30 \ln m \ln n + 26 \ln^2 n$$

with an expected $O(mn^{1/d} \ln m \cdot \min\{d, \ln n\} + n^{2+1/d} \ln n)$ calls to the membership Oracle of (X, \mathcal{S}) .

We note that the algorithm **BUILDMATCHING** can easily be modified to construct a spanning tree or a spanning path with the same guarantees up to a constant factor.

Algorithm 3: BUILDMATCHING $((X, \mathcal{S}), a, b, \gamma)$

```

 $M \leftarrow \emptyset$ 
while  $|X| \geq 4$  do
     $n \leftarrow |X|, m \leftarrow |\mathcal{S}|$ 
     $\omega_1(e) \leftarrow 1, \pi_1(S) \leftarrow 1 \quad \forall e \in E, S \in \mathcal{S} \quad // E \text{ denotes } \binom{X}{2}$ 
     $\mathbf{p} \leftarrow \min \left\{ \frac{48 \ln(|E| \cdot n/4)}{an^\gamma + b}, 1 \right\}$ 
     $\mathbf{q} \leftarrow \min \left\{ \frac{72 \ln(m \cdot n/4)}{an^\gamma + b}, 1 \right\}$ 
    for  $i = 1, \dots, n/4$  do
         $\omega_i(E) \leftarrow \sum_{e \in E} \omega_i(e)$ 
         $\pi_i(\mathcal{S}) \leftarrow \sum_{S \in \mathcal{S}} \pi_i(S)$ 
        choose  $e_i \sim \omega_i$   $// \mathbb{P}[e_i = e] = \frac{\omega_i(e)}{\omega_i(E)} \quad \forall e \in E$ 
        set the weight in  $\omega_i$  of  $e_i$  and of each edge adjacent to  $e_i$  to zero
        choose  $S_i \sim \pi_i$   $// \mathbb{P}[S_i = S] = \frac{\pi_i(S)}{\pi_i(\mathcal{S})} \quad \forall S \in \mathcal{S}$ 
         $E_i \leftarrow \text{sample from } E \text{ with probability } \mathbf{p} \quad // \mathbb{P}[e \in E_i] = \mathbf{p} \quad \forall e \in E$ 
         $\mathcal{S}_i \leftarrow \text{sample from } \mathcal{S} \text{ with probability } \mathbf{q} \quad // \mathbb{P}[S \in \mathcal{S}_i] = \mathbf{q} \quad \forall S \in \mathcal{S}$ 
         $// \mathbf{I}(e, S) = 1 \text{ if } e \text{ crosses } S, \mathbf{I}(e, S) = 0 \text{ otherwise}$ 
        for  $e \in E_i$  do
             $\omega_{i+1}(e) \leftarrow \omega_i(e) \left(1 - \frac{1}{2} \mathbf{I}(e, S_i)\right) \quad // \text{halve weight if } S_i \text{ crosses } e$ 
        for  $S \in \mathcal{S}_i$  do
             $\pi_{i+1}(S) \leftarrow \pi_i(S) (1 + \mathbf{I}(e_i, S)) \quad // \text{double weight if } S \text{ crosses } e_i$ 
     $M \leftarrow M \cup \{e_1, \dots, e_{n/4}\}$ 
     $X \leftarrow X \setminus \text{endpoints}(e_1, \dots, e_{n/4}) \quad // \text{remove elements covered by } e_i$ 
return  $M$ 

```

4.3.2 Epsilon-approximations

Our main result on ε -approximations is the following.

Theorem 16. *Let (X, \mathcal{S}) be a set system that satisfies **MAINASSUMPTION** (a, b, γ) and let $\varepsilon \in (0, 1)$. Then **APPROXIMATE** $((X, \mathcal{S}), a, b, \gamma, \varepsilon)$ returns a set $A \subset X$ of size at most*

$$2 \max \left\{ \left(30 \sqrt{\frac{a \ln m}{\gamma}} \cdot \frac{1}{\varepsilon} \right)^{\frac{2}{2-\gamma}}, \frac{12 \sqrt{\left(\frac{b}{2} + 12 \ln m\right) \ln m \log n}}{\varepsilon} \right\} + 1,$$

with expected approximation guarantee

$$\mathbb{E}[\varepsilon(A, X, \mathcal{S})] \leq \varepsilon,$$

and with an expected

$$\min \left\{ \frac{8n^{3-\gamma} \ln n}{a} + \frac{18mn^{1-\gamma} \ln(mn)}{a} \min \left\{ \frac{4}{(1-\gamma)^2}, \log^2 n \right\}, \frac{n^3}{49} + \frac{mn}{2} \right\}$$

calls to the membership Oracle of (X, \mathcal{S}) .

Algorithm 4: **APPROXIMATE** $((X, \mathcal{S}), a, b, \gamma, \varepsilon)$

$A_0 \leftarrow X$

$j = \left\lfloor \log |X| + \min \left\{ \frac{2}{2-\gamma} \log \frac{\varepsilon \sqrt{\gamma}}{30 \sqrt{a \ln(|\mathcal{S}|)}}, \log \frac{\varepsilon}{12 \sqrt{\left(\frac{b}{2} + 12 \ln(|\mathcal{S}|)\right) \ln(|\mathcal{S}|) \log |X|}} \right\} \right\rfloor$

for $i = 1, \dots, j$ **do**

$\chi \leftarrow \text{LOWDISCCOLOR}((A_{i-1}, \mathcal{S}|_{A_{i-1}}), a, b, \gamma)$
 $A_i \leftarrow \chi^{-1}(1)$

return A_j

Corollary 17. *Let (X, \mathcal{S}) be a set system and c, d be constants such that $\pi_{\mathcal{S}}^*(k) \leq ck^d$ for all $k \in [1, n]$. Then for any $\varepsilon \in (0, 1)$, **APPROXIMATE** $((X, \mathcal{S}), \frac{(2c)^{1/d}}{2 \ln 2(1-1/d)}, \frac{\ln m}{\ln 2}, 1 - \frac{1}{d}, \varepsilon)$ returns a set $A \subset X$ of size*

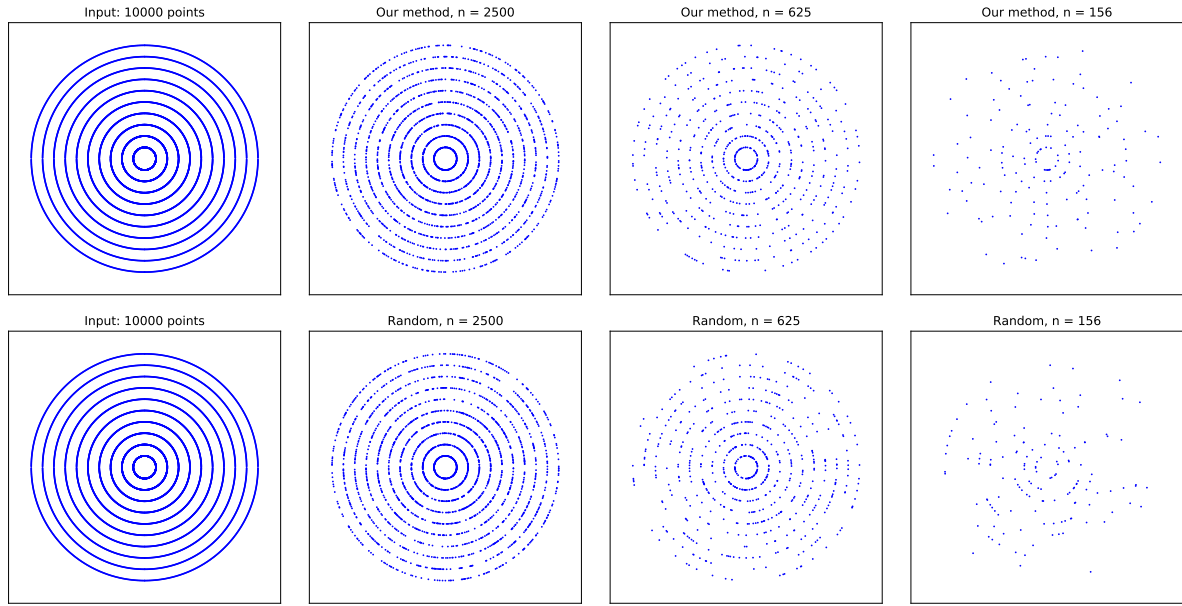
$$O \left(\max \left\{ \left(\frac{\ln m}{\varepsilon^2} \right)^{\frac{d}{d+1}}, \frac{\sqrt{\ln n \ln m}}{\varepsilon} \right\} \right)$$

with expected approximation guarantee satisfying $\mathbb{E}[\varepsilon(A, X, \mathcal{S})] \leq \varepsilon$, and with an expected

$$O \left(mn^{1/d} \ln(mn) \min \{d^2, \ln^2 n\} + n^{2+1/d} \ln n \right)$$

calls to the membership Oracle of (X, \mathcal{S}) .

The figures below show how the algorithm **APPROXIMATE** creates iteratively a data approximation with respect to disks (Top row) compared to random sampling (Bottom row). The input consists of 10000 points, placed evenly on 10 concentric circles. We note that, from visual perspective, the random sampling is spread less evenly on the concentric circles. Meanwhile, our method preserves better (again visually) the structure of the initial input with 10000 points.



For set systems, where uniform sampling yields small-sized ε -approximations, the construction time can be improved using the observation that $\varepsilon/2$ -approximation of an $\varepsilon/2$ -approximation is an ε -approximation of the original system (see Property 1). To this end, we use the following formulation of the uniform sampling guarantee of Li et al. (2001) (Theorem 4 in Chapter 2):

Theorem 18 ((Vershynin, 2018, Theorem 8.3.23)). *There is a universal constant C_{apx} such that if (X, \mathcal{S}) is a set system with VC-dimension at most V and A is a uniform random sample of X , then*

$$\mathbb{E}[\varepsilon(A, X, \mathcal{S})] \leq \sqrt{\frac{C_{\text{apx}} V}{|A|}}.$$

Thus, we can pre-process our point-set using Theorem 18 and get the following.

Corollary 19. *Let (X, \mathcal{S}) be a set system that satisfies **MAINASSUMPTION**(a, b, γ), and let V be such that $\text{d}_{\text{VC}}(X, \mathcal{S}) \leq V$, and let $\varepsilon \in (0, 1)$. Let A_0 be a uniform random sample of $\frac{C_{\text{apx}} V}{(\varepsilon/2)^2}$ elements from X . Then **APPROXIMATE**(($A_0, \mathcal{S}|_{A_0}$), $a, b, \gamma, \varepsilon/2$) returns a*

set $A \subset X$ of size at most

$$2 \max \left\{ \left(30 \sqrt{\frac{a \ln |\mathcal{S}|_{A_0}}{\gamma}} \cdot \frac{2}{\varepsilon} \right)^{\frac{2}{2-\gamma}}, \frac{24 \sqrt{\left(\frac{b}{2} + 12 \ln |\mathcal{S}|_{A_0}\right) \ln |\mathcal{S}|_{A_0} \log |A_0|}}{\varepsilon} \right\} + 1,$$

with expected approximation guarantee

$$\mathbb{E}[\varepsilon(A_j, X, \mathcal{S})] \leq \varepsilon,$$

and with an expected

$$\min \left\{ \frac{8|A_0|^{3-\gamma} \ln |A_0|}{a} + \frac{18|\mathcal{S}|_{A_0}||A_0|^{1-\gamma} \ln (|\mathcal{S}|_{A_0}||A_0|)}{a} \min \left\{ \frac{4}{(1-\gamma)^2}, \log^2 |A_0| \right\}, \right. \\ \left. \frac{|A_0|^3}{49} + \frac{|\mathcal{S}|_{A_0}||A_0|}{2} \right\}$$

calls to the membership Oracle of (X, \mathcal{S}) .

Outline of the remaining part of this chapter

We present the proofs of the stated results in the following order. In Sections 4.4 and 4.5 we prove Theorem 13 and Theorem 16 respectively. Section 4.6 contains the main technical part—the proof of Theorem 14. In Section 4.7, we state our guarantees for geometric set systems, and finally, in Section 4.8 we present an experimental evaluation of our method.

4.4 Proof of our main discrepancy result

In this section, we present the proof Theorem 13. One can observe that the algorithm **LOWDISCCOLOR** can be written in a compact form, using **BUILDMATCHING** as a subroutine, see Algorithm 2.

This presentation allows us to see that the algorithm **LOWDISCCOLOR** defines a coloring of X using the matching returned by **BUILDMATCHING**. In particular, for a matching M , **LOWDISCCOLOR** constructs a random coloring χ_M such that for each edge $\{x, y\} \in M$,

$$\chi_M(x) = \begin{cases} -1 & \text{with probability } 1/2, \\ 1 & \text{with probability } 1/2 \end{cases},$$

and set $\chi_M(y) = -\chi_M(x)$. The key lemma that connects the notions of low-discrepancy colorings and low-crossing matchings is the following.

Lemma 20. *Let (X, \mathcal{S}) be a set system, $n = |X|$, $m = |\mathcal{S}| \geq 34$, and let M be a matching of X with crossing number κ with respect to \mathcal{S} . Then the expected discrepancy of χ_M is at most $\sqrt{3\kappa \ln m}$.*

A ‘high probability version’ of this statement is well-known (Matoušek et al., 1991, Lemma 2.5) and implies the above bound via basic probabilistic calculations. For completeness, we present the proof of Lemma 20.

Proof of Lemma 20 Let $S \in \mathcal{S}$ be a fixed set. We express the sum $\chi_M(S)$ of colors over elements of S as

$$\chi_M(S) \stackrel{\text{def}}{=} \sum_{x \in S} \chi_M(x) = \sum_{\{x,y\} \in M; x,y \in S} (\chi_M(x) + \chi_M(y)) + \sum_{x \in \text{cr}(S,M)} \chi_M(x),$$

where $\text{cr}(S, M) = \{x \in S : \{x, y\} \in M, y \notin S\}$. By the definition of χ_M , each term in the first summation is zero, thus we obtain

$$\chi_M(S) = \sum_{x \in \text{cr}(S,M)} \chi_M(x).$$

Since $\text{cr}(S, M) \leq \kappa$ for any $S \in \mathcal{S}$, $\text{disc}(S, \chi_M)$ is a sum of at most κ independent random variables. We use the following concentration bound from Alon and Spencer (2016)

Claim 2 (Theorem A.1.1 from Alon and Spencer (2016)). *Let X_1, \dots, X_k be independent $\{-1, 1\}$ -valued random variables with $\mathbb{P}[X_i = -1] = \mathbb{P}[X_i = 1] = 1/2$. Then for any $\alpha \geq 0$*

$$\mathbb{P} \left[\left| \sum_{i=1}^k X_i \right| > \alpha \right] \leq 2e^{-\alpha^2/2k}.$$

Applying Claim 2, we get that for any fixed $S \in \mathcal{S}$ and $\alpha > 0$,

$$\mathbb{P} [|\chi_M(S)| > \alpha] \leq 2e^{-\alpha^2/2\kappa}.$$

By the union bound, we get

$$\mathbb{P} [\text{disc}_{\mathcal{S}}(\chi_M) > \alpha] \stackrel{\text{def}}{=} \mathbb{P} \left[\max_{S \in \mathcal{S}} |\chi_M(S)| > \alpha \right] \leq m \cdot 2e^{-\alpha^2/2\kappa}.$$

Finally, we bound the expected discrepancy by applying Fubini’s theorem

$$\mathbb{E} [\text{disc}_{\mathcal{S}}(\chi_M)] \stackrel{\text{def}}{=} \int_0^\infty \mathbb{P} [\text{disc}_{\mathcal{S}}(\chi_M) > \alpha] d\alpha \leq \int_0^\infty \min \left\{ 2m \cdot e^{-\alpha^2/2\kappa}, 1 \right\} d\alpha$$

$$\begin{aligned}
&= \int_0^{\sqrt{2\kappa \ln(2m)}} 1 d\alpha + \int_{\sqrt{2\kappa \ln(2m)}}^{\infty} 2m \cdot e^{-\alpha^2/2\kappa} d\alpha \\
&= \sqrt{2\kappa \ln(2m)} + 2m\sqrt{2\kappa} \int_{\sqrt{\ln(2m)}}^{\infty} e^{-t^2} dt \\
&= \sqrt{2\kappa \ln(2m)} + 2m\sqrt{2\kappa} \int_{\sqrt{\ln(2m)}}^{\infty} \frac{t}{t} \cdot e^{-t^2} dt \\
&\leq \sqrt{2\kappa \ln(2m)} + 2m\sqrt{\frac{2\kappa}{\ln(2m)}} \int_{\sqrt{\ln(2m)}}^{\infty} te^{-t^2} dt \\
&= \sqrt{2\kappa \ln(2m)} + 2m\sqrt{\frac{2\kappa}{\ln(2m)}} \left[-\frac{e^{-t^2}}{2} \right]_{\sqrt{\ln(2m)}}^{\infty} \\
&= \sqrt{2\kappa \ln(2m)} + \sqrt{\frac{\kappa}{2\ln(2m)}} \leq \sqrt{3\kappa \ln m},
\end{aligned}$$

if $m \geq 34$. This concludes the proof of Lemma 20. \square

In our case, the matching returned by **BUILDMATCHING** is also random. Using Lemma 20, taking total expectation over the matchings returned by **BUILDMATCHING**, and applying Jensen's inequality, we get

$$\mathbb{E}[\text{disc}_{\mathcal{S}}(\chi_M)] \leq \mathbb{E}\left[\sqrt{3\kappa(M) \ln m}\right] \leq \sqrt{3\mathbb{E}[\kappa(M)] \ln m},$$

where $\kappa(M)$ denotes the crossing number of M with respect to \mathcal{S} . Now we are ready to prove Theorem 13.

Proof of Theorem 13.

Let M be the matching returned by **BUILDMATCHING** $((X, \mathcal{S}), a, b, \gamma)$. By Theorem 14,

$$\mathbb{E}[\kappa(M)] \leq \frac{3a}{\gamma} n^\gamma + \frac{3b \log n}{2} + 18 \ln(nm) \log n.$$

Therefore, **LOWDISCCOLOR** $((X, \mathcal{S}), a, b, \gamma)$ returns a coloring with expected discrepancy at most

$$\sqrt{3 \left(\frac{3a}{\gamma} n^\gamma + \frac{3b \log n}{2} + 18 \ln(nm) \log n \right) \ln m}.$$

Each call of the membership Oracle is performed during the call of **BUILDMATCHING**, thus the bound on the expected number of membership Oracle calls follows immediately from Theorem 14. \square

4.5 Proof of our main result on approximations

Now, we prove Theorem 16. The algorithm **APPROXIMATE** constructs a sequence of sets $A_0, A_1, A_2, \dots, A_j \subseteq X$ iteratively. In particular, it sets $A_0 = X$ and for $i = 1, \dots, j$, $A_i \subseteq A_{i-1}$ is defined as $\chi_i^{-1}(1)$, where $\chi_i : A_i \rightarrow \{-1, 1\}$ is the coloring provided by **LOWDISCCOLOR** $((A_i, \mathcal{S}|_{A_i}), a, b, \gamma)$. Note that $|A_{i+1}| = \lceil |A_i|/2 \rceil = \lceil n/2^{i+1} \rceil$. Recall the following lemma from the introduction

Lemma 21. *Let (X, \mathcal{S}) be a set system with $|X| = n$, $X \in \mathcal{S}$ and let χ be a coloring with discrepancy $\text{disc}_{\mathcal{S}}(\chi) = \delta$ and let $A \subset X$ be a set of $\lceil n/2 \rceil$ elements from the larger color class of χ . Then A is a $(2\delta/n)$ -approximation of (X, \mathcal{S}) .*

By Theorem 13,

$$\mathbb{E} \left[\text{disc}_{\mathcal{S}|_{A_i}}(\chi_i) \right] \leq 3 \sqrt{\frac{a}{\gamma} |A_i|^\gamma \ln |\mathcal{S}|_{A_i}| + \left(\frac{b}{2} + 12 \ln |\mathcal{S}|_{A_i}| \right) \log |A_i| \cdot \ln |\mathcal{S}|_{A_i}|}.$$

Thus for $i = 0, \dots, j-1$, by Lemma 21,

$$\mathbb{E} \left[\varepsilon(A_{i+1}, A_i, \mathcal{S}|_{A_i}) \right] \leq \frac{6}{\lceil n/2^i \rceil} \cdot \sqrt{\frac{a}{\gamma} \cdot \lceil n/2^i \rceil^\gamma \ln m + \left(\frac{b}{2} + 12 \ln m \right) \log \lceil n/2^i \rceil \ln m}. \quad (4.2)$$

Moreover, by Property 1 (on approximations of approximations),

$$\varepsilon(A_j, X, \mathcal{S}) \leq \varepsilon(A_j, A_{j-1}, \mathcal{S}|_{A_{j-1}}) + \varepsilon(A_{j-1}, A_{j-2}, \mathcal{S}|_{A_{j-2}}) + \dots + \varepsilon(A_2, A_1, \mathcal{S}|_{A_1}) + \varepsilon(A_1, X, \mathcal{S}),$$

which by linearity of expectation and Equation (4.2) yield

$$\begin{aligned} \mathbb{E} \left[\varepsilon(A_j, X, \mathcal{S}) \right] &\leq \sum_{i=0}^{j-1} \frac{6}{\lceil n/2^i \rceil} \cdot \sqrt{\frac{a}{\gamma} \cdot \lceil n/2^i \rceil^\gamma \ln m + \left(\frac{b}{2} + 12 \ln m \right) \log \lceil n/2^i \rceil \ln m} \\ &\leq \frac{6}{n^{1-\gamma/2}} \sqrt{\frac{a \ln m}{\gamma}} \cdot \sum_{i=0}^{j-1} (2^{1-\gamma/2})^i + \frac{6}{n} \sqrt{\frac{b \log n \ln m}{2} + 12 \log n \ln^2 m} \cdot \sum_{i=0}^{j-1} 2^i \\ &\leq \frac{15}{n^{1-\gamma/2}} \sqrt{\frac{a \ln m}{\gamma}} \cdot (2^{1-\gamma/2})^j + \frac{6 \cdot (2^j - 1)}{n} \sqrt{\frac{b \log n \ln m}{2} + 12 \log n \ln^2 m} \\ &\leq 15 \sqrt{\frac{a \ln m}{\gamma}} \cdot \left(\frac{2^j}{n} \right)^{1-\gamma/2} + \frac{6 \cdot 2^j}{n} \sqrt{\frac{b \log n \ln m}{2} + 12 \log n \ln^2 m}. \end{aligned}$$

Substituting

$$j = \left\lceil \log n + \min \left\{ \frac{2}{2-\gamma} \log \frac{\varepsilon \sqrt{\gamma}}{30 \sqrt{a \ln m}}, \log \frac{\varepsilon}{12 \sqrt{(\frac{b}{2} + 12 \ln m) \ln(m) \log n}} \right\} \right\rceil,$$

we get that $\mathbb{E}[\varepsilon(A_j, X, \mathcal{S})] \leq \varepsilon$ and

$$|A_j| = \frac{n}{2^j} \leq 2 \max \left\{ \left(\frac{30\sqrt{a \ln m}}{\varepsilon\sqrt{\gamma}} \right)^{\frac{2}{2-\gamma}}, \frac{12\sqrt{(\frac{b}{2} + 12 \ln m) \ln(m) \log n}}{\varepsilon} \right\}.$$

By Theorem 13, constructing a the coloring χ_i requires at most

$$\min \left\{ \frac{24|A_i|^{3-\gamma} \ln |A_i|}{a} + \frac{18m|A_i|^{1-\gamma} \ln(m|A_i|)}{a} \min \left\{ \frac{2}{1-\gamma}, \log |A_i| \right\}, \frac{1}{7}|A_i|^3 + \frac{m|A_i|}{2} \right\}$$

calls to the membership Oracle, in expectation. Since $|A_i| = \lceil n/2^i \rceil$, the expected number of membership Oracle calls that **APPROXIMATE** $((X, \mathcal{S}), a, b, \gamma, j)$ performs is at most

$$\begin{aligned} & \sum_{i=0}^j \min \left\{ \frac{24 \left(\frac{n}{2^i}\right)^{3-\gamma} \ln \frac{n}{2^i}}{a} + \frac{18m \left(\frac{n}{2^i}\right)^{1-\gamma} \ln \frac{mn}{2^i}}{a} \min \left\{ \frac{2}{1-\gamma}, \log \frac{n}{2^i} \right\}, \frac{1}{7} \left(\frac{n}{2^i}\right)^3 + \frac{mn}{2^{i+1}} \right\} \\ & \leq \min \left\{ \sum_{i=0}^j \left(\frac{24 \left(\frac{n}{2^i}\right)^{3-\gamma} \ln \frac{n}{2^i}}{a} + \frac{18m \left(\frac{n}{2^i}\right)^{1-\gamma} \ln \frac{mn}{2^i}}{a} \min \left\{ \frac{2}{1-\gamma}, \log \frac{n}{2^i} \right\} \right), \right. \\ & \quad \left. \sum_{i=0}^j \left(\frac{n^3}{7 \cdot 2^{3i}} + \frac{mn}{2^{i+1}} \right) \right\} \\ & \leq \min \left\{ \frac{32n^{3-\gamma} \ln n}{a} + \frac{18mn^{1-\gamma} \ln(mn)}{a} \left(\min \left\{ \frac{2}{1-\gamma}, \log n \right\} \right)^2, \frac{8n^3}{49} + mn \right\}. \end{aligned}$$

This concludes the proof of Theorem 16 □

4.6 Proof of our main result on low-crossing matchings

In this section, we give a proof of Theorem 14. Later in the section, we will prove the following statement for **MATCHHALF** (see Algorithm 5), which captures one phase of **BUILDMATCHING**.

Theorem 22. *Let (X, \mathcal{S}) be a set system, $n = |X|$, $m = |\mathcal{S}|$ with $m \geq n$, which satisfies **MAINASSUMPTION** (a, b, γ) and let E denote the set of all pairs (edges) from X . Then **MATCHHALF** $((X, \mathcal{S}), E, a, b, \gamma, n/4)$ returns a matching of size $n/4$ with expected crossing number at most*

$$a \left(\frac{n}{2}\right)^\gamma + b + \max \left\{ \frac{a \left(\frac{n}{2}\right)^\gamma + b}{2}, 18 \ln(mn/4) \right\},$$

with an expected number of Oracle calls at most

$$\min \left\{ \frac{6}{a} \left(n^{3-\gamma} \ln \frac{n^3}{4} + 3mn^{1-\gamma} \ln \frac{mn}{4} \right), \frac{n^3 + 2mn}{8} \right\}.$$

Algorithm 5: MATCHHALF $((X, \mathcal{S}), E, a, b, \gamma, t)$

 $\omega_1(e) \leftarrow 1, \quad \pi_1(S) \leftarrow 1 \quad \forall e \in E, S \in \mathcal{S}$

for $i = 1, \dots, t$ **do** $\omega_i(E) \leftarrow \sum_{e \in E} \omega_i(e)$ $\pi_i(\mathcal{S}) \leftarrow \sum_{S \in \mathcal{S}} \pi_i(S)$ choose $e_i \sim \omega_i$ // $\mathbb{P}[e_i = e] = \frac{\omega_i(e)}{\omega_i(E)} \quad \forall e \in E$ choose $S_i \sim \pi_i$ // $\mathbb{P}[S_i = S] = \frac{\pi_i(S)}{\pi_i(\mathcal{S})} \quad \forall S \in \mathcal{S}$ $\mathbf{p} \leftarrow \min \left\{ \frac{48 \ln(|E| \cdot t)}{a|X|^{\gamma+b}}, 1 \right\}$ $\mathbf{q} \leftarrow \min \left\{ \frac{72 \ln(|\mathcal{S}| \cdot t)}{a|X|^{\gamma+b}}, 1 \right\}$ $E_i \leftarrow$ sample from E with probability \mathbf{p} // $\mathbb{P}[e \in E_i] = \mathbf{p} \quad \forall e \in E$ $\mathcal{S}_i \leftarrow$ sample from \mathcal{S} with probability \mathbf{q} // $\mathbb{P}[S \in \mathcal{S}_i] = \mathbf{q} \quad \forall S \in \mathcal{S}$ // $\mathbf{I}(e, S) = 1$ if e crosses S , $\mathbf{I}(e, S) = 0$ otherwise**for** $e \in E_i$ **do** $\omega_{i+1}(e) \leftarrow \omega_i(e) \left(1 - \frac{1}{2} \mathbf{I}(e, S_i)\right)$ // halve weight if S_i crosses e **for** $S \in \mathcal{S}_i$ **do** $\pi_{i+1}(S) \leftarrow \pi_i(S) \left(1 + \mathbf{I}(e_i, S)\right)$ // double weight if S crosses e_i set the weight in ω_{i+1} of e_i and of each edge adjacent to e_i to zero**return** $\{e_1, \dots, e_t\}$

First we show how Theorem 22 implies Theorem 14.

The algorithm BUILDMATCHING makes $\log n$ calls to MATCHHALF with exponentially decreasing input sizes. In particular, the overall expected number of membership Oracle calls of BUILDMATCHING can be bounded as

$$\begin{aligned}
& \sum_{i=0}^{\log n} \min \left\{ \frac{6}{a} \left(\left(\frac{n}{2^i} \right)^{3-\gamma} \ln \frac{n^3}{2^{3i+2}} + 3m \left(\frac{n}{2^i} \right)^{1-\gamma} \ln \frac{mn}{2^{i+2}} \right), \frac{\left(\frac{n}{2^i} \right)^3 + \frac{mn}{2^{i-1}}}{8} \right\} \\
& \leq \min \left\{ \sum_{i=0}^{\log n} \frac{6}{a} \left(\left(\frac{n^{3-\gamma}}{4^i} \right) \ln \frac{n^3}{2^{3i+2}} + 3m \left(\frac{n}{2^i} \right)^{1-\gamma} \ln \frac{mn}{2^{i+2}} \right), \sum_{i=0}^{\log n} \frac{\frac{n^3}{8^i} + \frac{2mn}{2^i}}{8} \right\} \\
& \leq \min \left\{ \frac{6}{a} \left(\frac{4}{3} n^{3-\gamma} \ln n^3 + 3m \cdot \min \left\{ \frac{2n^{1-\gamma}}{1-\gamma}, n^{1-\gamma} \log n \right\} \ln mn \right), \frac{n^3}{7} + \frac{mn}{2} \right\}.
\end{aligned}$$

As for the crossing number, Theorem 22 implies that the algorithm BUILDMATCHING $((X, \mathcal{S}), a, b, \gamma)$ returns a matching with expected crossing number at most

$$\sum_{i=1}^{\log n} \left[\frac{3a}{2} \left(\frac{n}{2^i} \right)^{\gamma} + \frac{3b}{2} + 18 \ln \frac{mn}{2^{i+1}} \right] < \frac{3an^{\gamma}}{2} \left[\sum_{i=1}^{\infty} \left(\frac{1}{2^{\gamma}} \right)^i \right] + \left(\frac{3b}{2} + 18 \ln(mn) \right) \log n$$

$$< \frac{3a}{\gamma} n^\gamma + \left(\frac{3b}{2} + 18 \ln(mn) \right) \log n.$$

This concludes the proof of Theorem 14. \square

Proof of Theorem 22. The proof relies on the following technical lemma, whose proof is presented later in this section. For an edge e and a set S , we define

$$I(e, S) = \begin{cases} 1 & \text{if } S \text{ crosses } e \\ 0 & \text{otherwise.} \end{cases}$$

Lemma 23 (Main Lemma). *Let $t \in [1, |X|/4]$ be an integer and $\tilde{E}_t \subseteq E$ denote the set of edges that have non-zero weight when $\text{MATCHHALF}((X, \mathcal{S}), E, a, b, \gamma, t)$ terminates. Then*

$$\mathbb{E} \left[\max_{S \in \mathcal{S}} \sum_{i=1}^t I(e_i, S) \right] \leq \frac{1}{2} \mathbb{E} \left[\min_{e \in \tilde{E}_t} \sum_{i=1}^t I(e, S_i) \right] + \frac{\left(\frac{128}{13} + \frac{8}{\ln 2}\right) \ln(|E|t)}{3\mathbf{p}} + \frac{\left(16 + \frac{4}{\ln 2}\right) \ln(|\mathcal{S}|t)}{3\mathbf{q}}. \quad (4.3)$$

The left-hand side of Equation (4.3) is precisely the expected crossing number of the edges $\{e_1, \dots, e_t\}$ picked by MATCHHALF . To bound the expectation in the right-hand side of Equation (4.3), we use the following lemma.

Lemma 24. *Let (Y, \mathcal{R}) be a set system, $w : \mathcal{R} \rightarrow \mathbb{R}_{\geq 0}$, and κ be such that Y has a matching with crossing number at most κ with respect to \mathcal{R} . Then there is an edge xy spanned by the points of Y such that*

$$\sum_{R \text{ crosses } \{x,y\}} w(R) \leq \frac{2w(\mathcal{R}) \cdot \kappa}{|Y|}.$$

Proof. Let M be a matching of Y such that any set of \mathcal{R} crosses at most κ edges of M . Then if we consider the weighted sum there are at most $w(\mathcal{R}) \cdot \kappa$ crossings between the edges of M and sets in \mathcal{R} counted with weights. By the pigeonhole principle, there is an edge in M that is crossed by sets of total weight at most

$$\frac{w(\mathcal{R}) \cdot \kappa}{|M|} = \frac{w(\mathcal{R}) \cdot \kappa}{|Y|/2} = \frac{2w(\mathcal{R})\kappa}{|Y|}$$

sets of \mathcal{R} . \square

Let $\tilde{X}_t \subset X$ denote the set of points that are not covered by the edges $\{e_1, \dots, e_t\}$ picked by $\text{MATCHHALF}((X, \mathcal{S}), E, a, b, \gamma)$. Note that $\tilde{E} = \binom{\tilde{X}}{2}$. Applying Lemma 24 to

$Y = \tilde{X}_t$ and $\mathcal{R} = \{S_1, \dots, S_t\}$ and weights $w(S_i) = 1$, we get that there is an edge $e \in \tilde{E}_t$ that satisfies

$$\sum_{i=1}^t \mathbf{I}(e, S_i) \leq \frac{2 \cdot \sum_{i=1}^t (a|\tilde{X}|^\gamma + b)}{|\tilde{X}|} = \frac{2a(|X| - 2t)^\gamma + 2b}{|X| - 2t} \cdot t \leq \left(\frac{2a}{(2t)^{1-\gamma}} + \frac{2b}{2t} \right) \cdot t. \quad (4.4)$$

Equations (4.3) and (4.4) imply

$$\begin{aligned} \mathbb{E} \left[\max_{S \in \mathcal{S}} \sum_{i=1}^t \mathbf{I}(e_i, S) \right] &\leq \left(\frac{2a}{(2t)^{1-\gamma}} + \frac{2b}{2t} \right) \cdot t + \frac{\left(\frac{128}{13} + \frac{8}{\ln 2} \right) \ln(|E|t)}{3\mathbf{p}} + \frac{\left(16 + \frac{4}{\ln 2} \right) \ln(|\mathcal{S}|t)}{3\mathbf{q}} \\ &\leq a(2t)^\gamma + b + \frac{\left(\frac{128}{13} + \frac{8}{\ln 2} \right) \ln(|E|t)}{3 \min \left\{ \frac{48}{a|X|^{\gamma+b}} \cdot \ln(|E| \cdot |X|/4), 1 \right\}} + \frac{\left(16 + \frac{4}{\ln 2} \right) \ln(|\mathcal{S}|t)}{3 \min \left\{ \frac{72}{a|X|^{\gamma+b}} \cdot \ln(|\mathcal{S}| \cdot |X|/4), 1 \right\}} \\ &\leq a \left(\frac{|X|}{2} \right)^\gamma + b + \frac{128 + \frac{104}{\ln 2}}{39} \max \left\{ \frac{a|X|^\gamma + b}{48}, \ln(|E|t) \right\} + \\ &\quad + \frac{16 + \frac{4}{\ln 2}}{3} \max \left\{ \frac{a|X|^\gamma + b}{72}, \ln(|\mathcal{S}|t) \right\} \\ &\leq a \left(\frac{|X|}{2} \right)^\gamma + b + \frac{128 + \frac{104}{\ln 2}}{39} \max \left\{ \frac{a|X|^\gamma + b}{48}, \frac{3}{2} \ln(|\mathcal{S}|t) \right\} + \\ &\quad + \frac{16 + \frac{4}{\ln 2}}{3} \max \left\{ \frac{a|X|^\gamma + b}{72}, \ln(|\mathcal{S}|t) \right\} \\ &= a \left(\frac{|X|}{2} \right)^\gamma + b + \left(\frac{128 + \frac{104}{\ln 2}}{26} + \frac{16 + \frac{4}{\ln 2}}{3} \right) \max \left\{ \frac{a|X|^\gamma + b}{72}, \ln(|\mathcal{S}|t) \right\} \\ &\leq a \left(\frac{|X|}{2} \right)^\gamma + b + 18 \cdot \max \left\{ \frac{a|X|^\gamma + b}{72}, \ln(|\mathcal{S}|t) \right\} \\ &= a \left(\frac{|X|}{2} \right)^\gamma + b + \max \left\{ \frac{a|X|^\gamma + b}{4}, 18 \ln(|\mathcal{S}|t) \right\}. \end{aligned}$$

Finally, we bound the number of membership Oracle calls. At each iteration $i = 1, \dots, n/4$, we update the weights of at most $\frac{n^2}{2}\mathbf{p} + m\mathbf{q}$ elements in expectation, each requiring one call to the membership Oracle. Thus in expectation, the total number of membership Oracle calls is at most

$$\begin{aligned} &\frac{n}{4} \left(\frac{n^2}{2} \min \left\{ \frac{48 \ln \frac{n^3}{4}}{an^\gamma + b}, 1 \right\} + m \min \left\{ \frac{72 \ln \frac{mn}{4}}{an^\gamma + b}, 1 \right\} \right) \\ &\leq \min \left\{ \frac{6}{a} \left(n^{3-\gamma} \ln \frac{n^3}{4} + 3mn^{1-\gamma} \ln \frac{mn}{4} \right), \frac{n^3 + 2mn}{8} \right\}. \end{aligned}$$

□

Proof of Main Lemma. The proof is subdivided into three lemmas. The first lemma is proved by examining the total weight of the sets of \mathcal{S} in π_{t+1} .

Lemma 25.

$$\mathbb{E} \left[\max_{S \in \mathcal{S}} \sum_{i=1}^t \mathbf{I}(e_i, S) \right] \leq \frac{4}{3 \ln 2} \cdot \sum_{i=1}^t \mathbb{E} \left[\sum_{S \in \mathcal{S}} \frac{\pi_i(S)}{\pi_i(\mathcal{S})} \mathbf{I}(e_i, S) \right] + \frac{(16 + \frac{4}{\ln 2}) \ln(|\mathcal{S}|t)}{3q}$$

Proof. Let $\pi_{t+1}(\mathcal{S})$ denote the total weight of the sets of \mathcal{S} in π_{t+1} . We bound $\pi_{t+1}(\mathcal{S})$ in two different ways. On the one hand, $\pi_{t+1}(\mathcal{S})$ is clearly lower-bounded by the weight of the set of maximum weight in π_{t+1} . Recall that the weight of a set S is doubled in iteration i if and only if $S \in \mathcal{S}_i$ and S crosses e_i , therefore

$$\pi_{t+1}(\mathcal{S}) \geq \max_{S \in \mathcal{S}} \pi_{t+1}(S) = 2^{\max_{S \in \mathcal{S}} \sum_{i=1}^t \mathbf{I}(e_i, S) \cdot \mathbf{1}_{\{S \in \mathcal{S}_i\}}},$$

where $\mathbf{1}_{\mathcal{A}}$ denotes the indicator random variable of the event \mathcal{A} . On the other hand, we can express $\pi_{t+1}(\mathcal{S})$ using the update rule of the algorithm

$$\begin{aligned} \pi_{t+1}(\mathcal{S}) &= \sum_{S \in \mathcal{S}} \pi_{t+1}(S) = \sum_{S \in \mathcal{S}} \pi_t(S) (1 + \mathbf{I}(e_t, S) \cdot \mathbf{1}_{\{S \in \mathcal{S}_t\}}) \\ &= \sum_{S \in \mathcal{S}} \pi_t(S) + \sum_{S \in \mathcal{S}} \pi_t(S) \mathbf{I}(e_t, S) \cdot \mathbf{1}_{\{S \in \mathcal{S}_t\}} \\ &= \pi_t(\mathcal{S}) + \pi_t(\mathcal{S}) \sum_{S \in \mathcal{S}} \frac{\pi_t(S)}{\pi_t(\mathcal{S})} \mathbf{I}(e_t, S) \cdot \mathbf{1}_{\{S \in \mathcal{S}_t\}} \\ &= \pi_t(\mathcal{S}) \left(1 + \sum_{S \in \mathcal{S}} \frac{\pi_t(S)}{\pi_t(\mathcal{S})} \mathbf{I}(e_t, S) \cdot \mathbf{1}_{\{S \in \mathcal{S}_t\}} \right). \end{aligned}$$

Unfolding this recursion and using the fact that $1 + a \leq \exp(a)$, we get

$$\begin{aligned} \pi_{t+1}(\mathcal{S}) &= \pi_1(\mathcal{S}) \prod_{i=1}^t \left(1 + \sum_{S \in \mathcal{S}} \frac{\pi_i(S)}{\pi_i(\mathcal{S})} \mathbf{I}(e_i, S) \cdot \mathbf{1}_{\{S \in \mathcal{S}_i\}} \right) \\ &\leq |\mathcal{S}| \cdot \exp \left(\sum_{i=1}^t \sum_{S \in \mathcal{S}} \frac{\pi_i(S)}{\pi_i(\mathcal{S})} \mathbf{I}(e_i, S) \cdot \mathbf{1}_{\{S \in \mathcal{S}_i\}} \right). \end{aligned}$$

Putting together the obtained upper and lower bounds on $\pi_{t+1}(\mathcal{S})$, we get

$$2^{\max_{S \in \mathcal{S}} \sum_{i=1}^t \mathbf{I}(e_i, S) \cdot \mathbf{1}_{\{S \in \mathcal{S}_i\}}} \leq |\mathcal{S}| \cdot \exp \left(\sum_{i=1}^t \sum_{S \in \mathcal{S}} \frac{\pi_i(S)}{\pi_i(\mathcal{S})} \mathbf{I}(e_i, S) \cdot \mathbf{1}_{\{S \in \mathcal{S}_i\}} \right).$$

Taking the logarithm of each side yields

$$\ln(2) \cdot \max_{S \in \mathcal{S}} \sum_{i=1}^t \mathbf{I}(e_i, S) \cdot \mathbf{1}_{\{S \in \mathcal{S}_i\}} \leq \sum_{i=1}^t \sum_{S \in \mathcal{S}} \frac{\pi_i(S)}{\pi_i(\mathcal{S})} \mathbf{I}(e_i, S) \cdot \mathbf{1}_{\{S \in \mathcal{S}_i\}} + \ln |\mathcal{S}|. \quad (4.5)$$

If $\mathbf{q} = 1$, then $\mathbf{1}_{\{S \in \mathcal{S}_i\}} = 1$ for all i and $S \in \mathcal{S}$, thus taking total expectation we conclude

$$\mathbb{E} \left[\max_{S \in \mathcal{S}} \sum_{i=1}^t \mathbf{I}(e_i, S) \right] \leq \frac{1}{\ln 2} \sum_{i=1}^t \mathbb{E} \left[\sum_{S \in \mathcal{S}} \frac{\pi_i(S)}{\pi_i(\mathcal{S})} \mathbf{I}(e_i, S) \right] + \frac{\ln |\mathcal{S}|}{\ln 2}.$$

Assume that $\mathbf{q} < 1$. Since $\max f(x) - \max g(x) \leq \max(f(x) - g(x))$, Equation (4.5) implies

$$\begin{aligned} \ln(2) \cdot \frac{3}{4} \cdot \max_{S \in \mathcal{S}} \sum_{i=1}^t \mathbf{I}(e_i, S) \cdot \mathbf{q} &\leq \ln(2) \cdot \max_{S \in \mathcal{S}} \sum_{i=1}^t \mathbf{I}(e_i, S) \cdot \left(\frac{3\mathbf{q}}{4} - \mathbf{1}_{\{S \in \mathcal{S}_i\}} \right) \\ &\quad + \sum_{i=1}^t \sum_{S \in \mathcal{S}} \frac{\pi_i(S)}{\pi_i(\mathcal{S})} \mathbf{I}(e_i, S) \cdot \mathbf{1}_{\{S \in \mathcal{S}_i\}} + \ln |\mathcal{S}|. \end{aligned}$$

Taking total expectation of each side, we obtain

$$\begin{aligned} \frac{3}{4} \ln(2) \cdot \mathbb{E} \left[\max_{S \in \mathcal{S}} \sum_{i=1}^t \mathbf{I}(e_i, S) \cdot \mathbf{q} \right] &\leq \ln(2) \cdot \mathbb{E} \left[\max_{S \in \mathcal{S}} \sum_{i=1}^t \mathbf{I}(e_i, S) \cdot \left(\frac{3\mathbf{q}}{4} - \mathbf{1}_{\{S \in \mathcal{S}_i\}} \right) \right] \\ &\quad + \sum_{i=1}^t \sum_{S \in \mathcal{S}} \mathbb{E} \left[\frac{\pi_i(S)}{\pi_i(\mathcal{S})} \mathbf{I}(e_i, S) \cdot \mathbf{1}_{\{S \in \mathcal{S}_i\}} \right] + \ln |\mathcal{S}|. \end{aligned} \tag{4.6}$$

Observe that for each fixed i , the random variables $\{\pi_i, e_i\}$ and \mathcal{S}_i are independent, thus

$$\sum_{i=1}^t \sum_{S \in \mathcal{S}} \mathbb{E} \left[\frac{\pi_i(S)}{\pi_i(\mathcal{S})} \mathbf{I}(e_i, S) \cdot \mathbf{1}_{\{S \in \mathcal{S}_i\}} \right] = \mathbf{q} \cdot \sum_{i=1}^t \sum_{S \in \mathcal{S}} \mathbb{E} \left[\frac{\pi_i(S)}{\pi_i(\mathcal{S})} \mathbf{I}(e_i, S) \right]. \tag{4.7}$$

To bound the expectation of $\max_{S \in \mathcal{S}} \sum_{i=1}^t \mathbf{I}(e_i, S) \cdot \left(\frac{3\mathbf{q}}{4} - \mathbf{1}_{\{S \in \mathcal{S}_i\}} \right)$, we will need the following Azuma-type inequality for martingales.

Lemma 26 ((Koufogiannakis and Young, 2014, Lemma 10)). *Let $X = \sum_{i=1}^T x_i$ and $Y = \sum_{i=1}^T y_i$ be sums of non-negative random variables, where T is a random stopping time with finite expectation, and, for all i , $|x_i - y_i| < 1$ and*

$$\mathbb{E} \left[x_i - y_i \mid \sum_{s < i} x_s, \sum_{s < i} y_s \right] \leq 0.$$

Let $\varepsilon \in [0, 1]$ and $A \in \mathbb{R}$, then

$$\mathbb{P} [(1 - \varepsilon)X \geq Y + A] \leq \exp(-\varepsilon A).$$

Claim 3.

$$\mathbb{P} \left[\max_{S \in \mathcal{S}} \sum_{i=1}^t \mathbf{I}(e_i, S) \cdot \left(\frac{3\mathbf{q}}{4} - \mathbf{1}_{\{S \in \mathcal{S}_i\}} \right) \geq 3 \ln(|\mathcal{S}|t) \right] \leq \frac{1}{t}.$$

Proof. For each $i \in [1, t]$ and $S \in \mathcal{S}$, consider the random variables $x_i(S) = \mathbf{I}(e_i, S) \cdot \mathbf{q}$ and $y_i(S) = \mathbf{I}(e_i, S) \cdot \mathbf{1}_{\{S \in \mathcal{S}_i\}}$, which are measurable with respect to e_i and \mathcal{S}_i . For any i and $S \in \mathcal{S}$, we have $|x_i(S) - y_i(S)| \leq 1$. Since \mathcal{S}_i is independent of e_i , $\sum_{k < i} x_k(S)$, and $\sum_{k < i} y_k(S)$, we have

$$\mathbb{E} \left[x_i(S) - y_i(S) \mid \sum_{k < i} x_k(S), \sum_{k < i} y_k(S) \right] = 0$$

as $\mathbb{E} [\mathbf{q} - \mathbf{1}_{\{S \in \mathcal{S}_i\}}] = 0$ for all $i \in [1, t]$ and $S \in \mathcal{S}$.

Therefore, Lemma 26 with $\varepsilon = 1/4$, combined with the union bound implies for any $A \in \mathbb{R}$,

$$\mathbb{P} \left(\max_{S \in \mathcal{S}} \sum_{i=1}^t \mathbf{I}(e_i, S) \cdot \left(\frac{3\mathbf{q}}{4} - \mathbf{1}_{\{S \in \mathcal{S}_i\}} \right) \geq A \right) \leq |\mathcal{S}| \exp \left(-\frac{A}{4} \right).$$

Setting $A = 4 \ln(|\mathcal{S}|t)$, we conclude the proof of Claim 3. \square

Applying Claim 3 and using that $\sum_{i=1}^t \mathbf{I}(e_i, S) \cdot \left(\frac{3\mathbf{q}}{4} - \mathbf{1}_{\{S \in \mathcal{S}_i\}} \right) \leq t$ always holds, we get

$$\mathbb{E} \left[\max_{S \in \mathcal{S}} \sum_{i=1}^t \mathbf{I}(e_i, S) \cdot \left(\frac{3\mathbf{q}}{4} - \mathbf{1}_{\{S \in \mathcal{S}_i\}} \right) \right] \leq 4 \ln(|\mathcal{S}|t) + t \cdot \frac{1}{t} \leq 4 \ln(|\mathcal{S}|t) + 1. \quad (4.8)$$

Hence Equations (4.6), (4.7), and (4.8) imply

$$\begin{aligned} & \frac{3 \ln 2}{4} \cdot \mathbf{q} \cdot \mathbb{E} \left[\max_{S \in \mathcal{S}} \sum_{i=1}^t \mathbf{I}(e_i, S) \right] \\ & \leq \sum_{i=1}^t \mathbf{q} \cdot \mathbb{E} \left[\sum_{S \in \mathcal{S}} \frac{\pi_i(S)}{\pi_i(\mathcal{S})} \mathbf{I}(e_i, S) \right] + \ln(2) \cdot (4 \ln(|\mathcal{S}|t) + 1) + \ln |\mathcal{S}|. \end{aligned}$$

Dividing both sides by $\frac{\mathbf{q} \cdot 3 \ln 2}{4}$, we obtain

$$\begin{aligned} \mathbb{E} \left[\max_{S \in \mathcal{S}} \sum_{i=1}^t \mathbf{I}(e_i, S) \right] & \leq \frac{4}{3 \ln 2} \cdot \sum_{i=1}^t \mathbb{E} \left[\sum_{S \in \mathcal{S}} \frac{\pi_i(S)}{\pi_i(\mathcal{S})} \mathbf{I}(e_i, S) \right] + \frac{16 \ln(|\mathcal{S}|t) + 4 + 4 \frac{\ln |\mathcal{S}|}{\ln 2}}{3\mathbf{q}} \\ & \leq \frac{4}{3 \ln 2} \cdot \sum_{i=1}^t \mathbb{E} \left[\sum_{S \in \mathcal{S}} \frac{\pi_i(S)}{\pi_i(\mathcal{S})} \mathbf{I}(e_i, S) \right] + \frac{(16 + \frac{4}{\ln 2}) \ln(|\mathcal{S}|t)}{3\mathbf{q}}. \end{aligned}$$

This concludes the proof of Lemma 25. \square

The next lemma is proven by applying analogous arguments for the total weight of edges in ω_{t+1} with a small adjustment as in each iteration we set some edge weights to zero. Recall that \tilde{E}_t denotes the set of edges that have non-zero weight in ω_{t+1} .

Lemma 27.

$$\sum_{i=1}^t \sum_{e \in E} \mathbb{E} \left[\frac{\omega_i(e)}{\omega_i(E)} \mathbf{I}(e, S_i) \right] < \frac{3 \ln 2}{8} \cdot \mathbb{E} \left[\min_{e \in \tilde{E}_t} \sum_{i=1}^t \mathbf{I}(e, S_i) \right] + \frac{\left(\frac{32 \ln 2}{13} + 2\right) \ln(|E|t)}{\mathbf{p}}.$$

Proof. Let $\omega_{t+1}(E)$ denote the total weight of edges in ω_{t+1} . Again, we lower-bound $\omega_{t+1}(E)$ by the largest edge-weight in ω_{t+1} , which is now attained at some edge of \tilde{E}_t :

$$\omega_{t+1}(E) \geq \max_{e \in E} \omega_{t+1}(e) = \max_{e \in \tilde{E}_t} \omega_{t+1}(e) = \left(\frac{1}{2}\right)^{\min_{e \in \tilde{E}_t} \sum_{i=1}^t \mathbf{I}(e, S_i) \cdot \mathbf{1}_{\{e \in E_i\}}}}.$$

The upper bound is obtained by using the algorithm's weight update rule. Since e_t has positive weight in ω_t , but its weight in ω_{t+1} is set to 0, we have a strict inequality

$$\begin{aligned} \omega_{t+1}(E) &= \sum_{e \in E} \omega_{t+1}(e) < \sum_{e \in E} \omega_t(e) \left(1 - \frac{1}{2} \mathbf{I}(e, S_t) \cdot \mathbf{1}_{\{e \in E_t\}}\right) \\ &= \sum_{e \in E} \omega_t(e) - \frac{1}{2} \sum_{e \in E} \omega_t(e) \mathbf{I}(e, S_t) \cdot \mathbf{1}_{\{e \in E_t\}} \\ &= \omega_t(E) \left(1 - \frac{1}{2} \sum_{e \in E} \frac{\omega_t(e)}{\omega_t(E)} \mathbf{I}(e, S_t) \cdot \mathbf{1}_{\{e \in E_t\}}\right). \end{aligned}$$

Unfolding this recursion and using the fact that $1 + a \leq \exp(a)$, we get

$$\omega_{t+1}(E) < |E| \cdot \exp \left(-\frac{1}{2} \sum_{i=1}^t \sum_{e \in E} \frac{\omega_i(e)}{\omega_i(E)} \mathbf{I}(e, S_i) \cdot \mathbf{1}_{\{e \in E_i\}} \right).$$

Combining the obtained upper and the lower bounds on $\omega_{t+1}(E)$ and taking the logarithm of each side, we get

$$\ln \left(\frac{1}{2} \right) \cdot \min_{e \in \tilde{E}_t} \sum_{i=1}^t \mathbf{I}(e, S_i) \cdot \mathbf{1}_{\{e \in E_i\}} < -\frac{1}{2} \sum_{i=1}^t \sum_{e \in E} \frac{\omega_i(e)}{\omega_i(E)} \mathbf{I}(e, S_i) \cdot \mathbf{1}_{\{e \in E_i\}} + \ln |E|,$$

which is equivalent to

$$\sum_{i=1}^t \sum_{e \in E} \frac{\omega_i(e)}{\omega_i(E)} \mathbf{I}(e, S_i) \cdot \mathbf{1}_{\{e \in E_i\}} < 2 \ln(2) \cdot \min_{e \in \tilde{E}_t} \sum_{i=1}^t \mathbf{I}(e, S_i) \cdot \mathbf{1}_{\{e \in E_i\}} + 2 \ln |E|. \quad (4.9)$$

If $\mathbf{p} = 1$, then $\mathbf{1}_{\{e \in E_i\}} = 1$ for all i and $e \in E$, thus taking total expectation we conclude

$$\sum_{i=1}^t \mathbb{E} \left[\sum_{e \in E} \frac{\omega_i(e)}{\omega_i(E)} \mathbf{I}(e, S_i) \right] < 2 \ln(2) \cdot \mathbb{E} \left[\min_{e \in \tilde{E}_t} \sum_{i=1}^t \mathbf{I}(e, S_i) \right] + 2 \ln |E|.$$

Assume that $\mathbf{p} < 1$. Since $\min f(x) - \min g(x) \leq \max(f(x) - g(x))$, Equation (4.9) implies

$$\begin{aligned} \sum_{i=1}^t \sum_{e \in E} \frac{\omega_i(e)}{\omega_i(E)} \mathbf{I}(e, S_i) \cdot \mathbf{1}_{\{e \in E_i\}} &< 2 \ln(2) \cdot \max_{e \in \tilde{E}_t} \sum_{i=1}^t \mathbf{I}(e, S_i) \cdot \left(\mathbf{1}_{\{e \in E_i\}} - \frac{3\mathbf{p}}{16} \right) \\ &+ 2 \ln(2) \cdot \min_{e \in \tilde{E}_t} \sum_{i=1}^t \mathbf{I}(e, S_i) \cdot \frac{3\mathbf{p}}{16} + 2 \ln |E|. \end{aligned}$$

Taking total expectation of each side, and using that for each fixed i , the random variables $\{\omega_i, S_i\}$ and E_i are independent, we get

$$\begin{aligned} \mathbf{p} \cdot \sum_{i=1}^t \sum_{e \in E} \mathbb{E} \left[\frac{\omega_i(e)}{\omega_i(E)} \mathbf{I}(e, S_i) \right] &< 2 \ln(2) \cdot \mathbb{E} \left[\max_{e \in \tilde{E}_t} \sum_{i=1}^t \mathbf{I}(e, S_i) \cdot \left(\mathbf{1}_{\{e \in E_i\}} - \frac{3\mathbf{p}}{16} \right) \right] \\ &+ 2 \ln(2) \cdot \mathbb{E} \left[\min_{e \in \tilde{E}_t} \sum_{i=1}^t \mathbf{I}(e, S_i) \cdot \frac{3\mathbf{p}}{16} \right] + 2 \ln |E|. \end{aligned} \quad (4.10)$$

We need the following claim whose proof uses Lemma 26 and is similar to Claim 3.

Claim 4.

$$\mathbb{P} \left[\max_{e \in \tilde{E}_t} \sum_{i=1}^t \mathbf{I}(e, S_i) \cdot \left(\mathbf{1}_{\{e \in E_i\}} - \frac{3\mathbf{p}}{16} \right) \geq \frac{16}{13} \ln(|E|t) \right] \leq \frac{1}{t}.$$

This, together with the fact that $\sum_{i=1}^t \mathbf{I}(e, S_i) \cdot \left(\mathbf{1}_{\{e \in E_i\}} - \frac{3\mathbf{p}}{16} \right) \leq t$ always holds imply

$$\mathbb{E} \left[\max_{e \in \tilde{E}_t} \sum_{i=1}^t \mathbf{I}(e, S_i) \cdot \left(\mathbf{1}_{\{e \in E_i\}} - \frac{3\mathbf{p}}{16} \right) \right] \leq \frac{16}{13} \ln(|E|t) + t \cdot \frac{1}{t} \leq \frac{16}{13} \ln(|E|t) + 1.$$

Hence Equation (4.10) yields

$$\begin{aligned} \sum_{i=1}^t \mathbf{p} \cdot \sum_{e \in E} \mathbb{E} \left[\frac{\omega_i(e)}{\omega_i(E)} \mathbf{I}(e, S_i) \right] &< \frac{6 \ln 2}{16} \cdot \mathbb{E} \left[\min_{e \in \tilde{E}_t} \sum_{i=1}^t \mathbf{I}(e, S_i) \cdot \mathbf{p} \right] + 2 \ln(2) \cdot \left(\frac{16}{13} \ln(|E|t) + 1 \right) + 2 \ln |E|. \end{aligned}$$

Dividing both sides by \mathbf{p} , we get

$$\sum_{i=1}^t \sum_{e \in E} \mathbb{E} \left[\frac{\omega_i(e)}{\omega_i(E)} \mathbf{I}(e, S_i) \right] < \frac{3 \ln 2}{8} \cdot \mathbb{E} \left[\min_{e \in \tilde{E}_t} \sum_{i=1}^t \mathbf{I}(e, S_i) \right] + \frac{\left(\frac{32 \ln 2}{13} + 2 \right) \ln(|E|t)}{\mathbf{p}}.$$

□

We need one more lemma to tie the previous two together.

Lemma 28. *For any $i \in [1, t]$, we have*

$$\mathbb{E} \left[\sum_{S \in \mathcal{S}} \frac{\pi_i(S)}{\pi_i(\mathcal{S})} \mathbf{I}(e_i, S) \right] = \mathbb{E} \left[\sum_{e \in E} \frac{\omega_i(e)}{\omega_i(E)} \mathbf{I}(e, S_i) \right].$$

Proof. Let $F_i = \sigma(e_1, \dots, e_i, S_1, \dots, S_i, E_1, \dots, E_i, \mathcal{S}_1, \dots, \mathcal{S}_i)$. We have

$$\begin{aligned} \mathbb{E} \left[\sum_{S \in \mathcal{S}} \frac{\pi_i(S)}{\pi_i(\mathcal{S})} \mathbf{I}(e_i, S) \right] &= \mathbb{E} \left[\mathbb{E} \left[\sum_{S \in \mathcal{S}} \frac{\pi_i(S)}{\pi_i(\mathcal{S})} \mathbf{I}(e_i, S) \mid F_{i-1} \right] \right] \quad \text{and} \\ \mathbb{E} \left[\sum_{e \in E} \frac{\omega_i(e)}{\omega_i(E)} \mathbf{I}(e, S_i) \right] &= \mathbb{E} \left[\mathbb{E} \left[\sum_{e \in E} \frac{\omega_i(e)}{\omega_i(E)} \mathbf{I}(e, S_i) \mid F_{i-1} \right] \right]. \end{aligned}$$

Observe that ω_i and π_i are measurable with respect to F_{i-1} , thus

$$\begin{aligned} \mathbb{E} \left[\sum_{S \in \mathcal{S}} \frac{\pi_i(S)}{\pi_i(\mathcal{S})} \mathbf{I}(e_i, S) \mid F_{i-1} \right] &= \sum_{e \in E} \frac{\omega_i(e)}{\omega_i(E)} \cdot \left(\sum_{S \in \mathcal{S}} \frac{\pi_i(S)}{\pi_i(\mathcal{S})} \mathbf{I}(e, S) \right) \\ &= \sum_{e \in E} \sum_{S \in \mathcal{S}} \frac{\omega_i(e)}{\omega_i(E)} \cdot \frac{\pi_i(S)}{\pi_i(\mathcal{S})} \mathbf{I}(e, S) \\ &= \sum_{S \in \mathcal{S}} \frac{\pi_i(S)}{\pi_i(\mathcal{S})} \cdot \left(\sum_{e \in E} \frac{\omega_i(e)}{\omega_i(E)} \mathbf{I}(e, S) \right) = \mathbb{E} \left[\sum_{e \in E} \frac{\omega_i(e)}{\omega_i(E)} \mathbf{I}(e, S_i) \mid F_{i-1} \right]. \end{aligned}$$

□

Finally, we combine Lemmas 25, 27, and 28 in the following way

$$\begin{aligned} \mathbb{E} \left[\max_{S \in \mathcal{S}} \sum_{i=1}^t \mathbf{I}(e_i, S) \right] &\leq \frac{4}{3 \ln 2} \cdot \sum_{i=1}^t \mathbb{E} \left[\sum_{S \in \mathcal{S}} \frac{\pi_i(S)}{\pi_i(\mathcal{S})} \mathbf{I}(e_i, S) \right] + \frac{(16 + \frac{4}{\ln 2}) \ln(|\mathcal{S}|t)}{3\mathbf{q}} \\ &= \frac{4}{3 \ln 2} \cdot \sum_{i=1}^t \mathbb{E} \left[\sum_{e \in E} \frac{\omega_i(e)}{\omega_i(E)} \mathbf{I}(e, S_i) \right] + \frac{(16 + \frac{4}{\ln 2}) \ln(|\mathcal{S}|t)}{3\mathbf{q}} \\ &< \frac{1}{2} \cdot \mathbb{E} \left[\min_{e \in \tilde{E}} \sum_{i=1}^t \mathbf{I}(e, S_i) \right] + \frac{(\frac{128}{13} + \frac{8}{\ln 2}) \ln(|E|t)}{3\mathbf{p}} + \frac{(16 + \frac{4}{\ln 2}) \ln(|\mathcal{S}|t)}{3\mathbf{q}} \end{aligned}$$

This completes the proof of the Main Lemma and thus of Theorem 22. □

4.7 Our results for geometric set systems

In this section, we state the implications of our results in Section 4.2 for set systems induced by geometric objects. As before, let (X, \mathcal{S}) be a finite set system, $n = |X|$ and

$m = |\mathcal{S}|$ and assume that $m \geq n$. For a dimension $d \geq 2$, let X be a finite set of points in \mathbb{R}^d and let \mathcal{C} be a collection of geometric objects in \mathbb{R}^d . We say that a set $Y \subset X$ is induced by \mathcal{C} if Y can be written as $\{x \in X : x \in C\}$ for some $C \in \mathcal{C}$. We say that a set system (X, \mathcal{S}) is induced by \mathcal{C} if each range in \mathcal{S} is induced by \mathcal{C} .

4.7.1 Semialgebraic set systems

Let $\Gamma_{d,\Delta,s}$ denote the collection of semialgebraic sets in \mathbb{R}^d that can be defined as the solution set of a Boolean combination of at most s polynomial inequalities of degree at most Δ . First, we give a bound on its dual shatter function.

Lemma 29. *Let X be a set of points in \mathbb{R}^d and (X, \mathcal{S}) be a set system induced by $\Gamma_{d,\Delta,s}$. Then $d_{\text{VC}}(X, \mathcal{S}) \leq 2s \log(es) \binom{\Delta+d}{d}$ and the dual shatter function of (X, \mathcal{S}) can be upper-bounded as $\pi_{\mathcal{S}}^*(k) \leq (4e\Delta s)^d \cdot k^d$.*

Proof. The bound on the VC-dimension can be deduced from Propositions 10.3.2 and Proposition 10.3.3 in Matoušek (2002). Let $\mathcal{R} \subseteq \Gamma_{d,\Delta,s}$ be a set of k ranges, defined by $\mathcal{P} = \{p_{ij} : 1 \leq i \leq k, 1 \leq j \leq s\}$, where each element is a d -variate polynomial of degree at most Δ . Observe that if $\text{sign}[p(x)] = \text{sign}[p(y)]$ for all $p \in \mathcal{P}$, then x, y are equivalent with respect to \mathcal{R} . Therefore, $\pi_{\Gamma_{d,\Delta,s}}^*(k)$ can be upper-bounded by the number of different sign patterns in $\{-1, 1\}^{ks}$ induced by ks d -variate polynomials of degree at most Δ . This quantity is bounded by $(4e\Delta s)^d \cdot k^d$, see (Warren, 1968, Theorem 3). \square

Now we can apply Theorems 13, 14, and Theorem 16 and obtain the following.

Corollary 30. *Let X be a set of points n in \mathbb{R}^d and (X, \mathcal{S}) be a set system with m ranges, each induced by $\Gamma_{d,\Delta,s}$. Then*

- i) **LOWDISCCOLOR** $((X, \mathcal{S}), \frac{4e\Delta s}{\ln 2(1-1/d)}, \frac{\ln m}{\ln 2}, 1 - \frac{1}{d})$ constructs a coloring χ of X of with expected discrepancy at most

$$3 \sqrt{\frac{4e\Delta s \ln m}{\ln 2(1-1/d)^2} \cdot n^{1-1/d} + 19 \ln^2 m \ln n}$$

with expected running time $O(s\Delta^d (mn^{1/d} \ln(mn) \cdot \min\{d, \ln n\} + n^{2+1/d} \ln n))$.

- ii) **BUILDMATCHING** $((X, \mathcal{S}), \frac{4e\Delta s}{\ln 2(1-1/d)}, \frac{\ln m}{\ln 2}, 1 - \frac{1}{d})$ returns a matching $\{e_1, \dots, e_{n/2}\}$ with expected crossing number at most

$$\frac{12es\Delta}{(1-1/d)^2 \ln 2} \cdot n^{1-1/d} + O(\ln m \ln n)$$

with expected running time $O(s\Delta^d (mn^{1/d} \ln(mn) \cdot \min\{d, \ln n\} + n^{2+1/d} \ln n))$.

iii) if $\varepsilon \in (0, 1)$, $V := d_{VC}(X, \mathcal{S})$, and A_0 is a uniform random sample of X of size $\frac{4C_{\text{apx}}V}{\varepsilon^2}$, then **APPROXIMATE** $((A_0, \mathcal{S}|_{A_0}), \frac{4e\Delta s}{\ln 2(1-1/d)}, \frac{\ln |\mathcal{S}|_{A_0}|}{\ln 2}, 1 - \frac{1}{d}, \varepsilon)$ returns a set $A \subset X$ of size

$$O\left(\max\left\{\left(\Delta s \cdot \frac{V}{\varepsilon^2} \ln \frac{1}{\varepsilon}\right)^{\frac{d}{d+1}}, \frac{V}{\varepsilon} \ln^{3/2}\left(\frac{V}{\varepsilon}\right)\right\}\right)$$

with expected approximation guarantee satisfying $\mathbb{E}[\varepsilon(A, X, \mathcal{S})] \leq \varepsilon$, and with an expected running time $O\left(n + \left(\frac{V}{\varepsilon^2}\right)^{2+1/d} \ln \frac{V}{\varepsilon^2} + \left(\frac{V}{\varepsilon^2}\right)^{V+1/d} \ln \left(\frac{V}{\varepsilon^2}\right)^{V+1} \min\{d^2, \ln^2 \frac{V}{\varepsilon^2}\}\right)$.

Remark. The previous best algorithm for constructing matchings with low crossing numbers with respect to $\Gamma_{d,\Delta,s}$ relies on the polynomial partitioning technique (Agarwal et al., 2013). It computes a matching of n points in general position with crossing number $O(10^d s \Delta n^{1-1/d})$ with respect to any set in $\Gamma_{d,\Delta,s}$ in time $O(n^{O(d^3)})$, notably the running time is independent of m . Our algorithm provides improved running time bounds for specific instances with $m = n^{o(d^3)}$.

4.7.2 Half-spaces and balls

Let \mathcal{H}_d and \mathcal{B}_d denote the set of all half-spaces and balls in \mathbb{R}^d respectively. Half-spaces and balls are semialgebraic sets, in particular, $\mathcal{H}_d = \Gamma_{d,1,1}$ and $\mathcal{B}_d \subset \Gamma_{d,2,1}$. What distinguishes their case from the general one is the existence of *test-sets*. Test-sets are small-sized subfamilies (of half-spaces and balls) such that if a matching has low crossing number with respect to this subfamily, then it is guaranteed to have low crossing number with respect to any member of the family (of half-spaces and balls resp.). In particular, we can use a result of Matoušek (1992) on test-sets for half-spaces.

Lemma 31 (Test set lemma). *Let X be a set of n points in \mathbb{R}^d and t be a parameter. There exists a set $\mathcal{T}(t)$ of at most $(d+1)t^d$ hyperplanes such that if a matching of X has crossing number κ with respect to $\mathcal{T}(t)$, then its crossing number with respect to any half-space in \mathbb{R}^d is at most $(d+1)\kappa + \frac{6d^2n}{t}$.*

We will use Lemma 31 as black-box to obtain a test-set lemma for balls. It is well known that there are mappings $\alpha : \mathbb{R}^d \rightarrow \mathbb{R}^{d+1}$ and $\beta : \mathcal{B}_d \rightarrow \mathcal{H}_{d+1}$ such that for any $p \in \mathbb{R}^d$ and $B \in \mathcal{B}_d$, we have $p \in B$ if and only if $\alpha(p) \in \beta(B)$, see e.g. (Matoušek, 2002, Chap. 10). This mapping and Lemma 31 applied in \mathbb{R}^{d+1} with $t = n^{1/d}$ give the following lemma.

Lemma 32. *Let X be a set of n points in \mathbb{R}^d . There exists a set \mathcal{Q} of at most $(d+2)n^{1+1/d}$ balls such that if a matching of X has crossing number κ with respect to \mathcal{Q} , then its crossing number with respect to \mathcal{B}_d is at most $(d+2)\kappa + 6(d+1)^2n^{1-1/d}$.*

These lemmas allow us to efficiently construct matchings with low crossing number with respect to *any* half-space or ball in \mathbb{R}^d . (This is in contrast to previous setups, where we required to have a finite set of m ranges as an input.) Notice however that we cannot use directly Lemmas 31 and 32 to obtain test-sets for ordered matchings.

Lemma 31 and Corollary 30 implies the following corollary unordered matchings.

Corollary 33. *Let X be a set of n points in \mathbb{R}^d and $\mathcal{T} = \mathcal{T}(n^{1/d})$ be the set of half-spaces provided by Lemma 31. Then $\text{BUILDMATCHING}((X, \mathcal{T}), \frac{4e}{(1-1/d)\ln 2}, \frac{\ln((d+1)n)}{\ln 2}, 1 - \frac{1}{d})$ returns a matching of X with expected crossing number at most*

$$\left(6d^2 + \frac{12e(d+1)}{(1-1/d)^2 \ln 2}\right) n^{1-1/d} + O(\ln(dn) \ln n)$$

with respect to half-spaces in \mathbb{R}^d . The expected running time is $O(dn^{2+1/d} \ln n)$.

Similarly, we can apply Corollary 30 to set systems induced by balls. Note that in case of balls, the Oracle complexity can be improved to $O(d)$ from the $O(2^d)$ bound used in Corollary 30 for $\Gamma_{d,2,1}$.

Corollary 34. *Let X be a set of n points in \mathbb{R}^d and let \mathcal{Q} be the set of balls provided by Lemma 32. Then $\text{BUILDMATCHING}((X, \mathcal{Q}), \frac{8e}{(1-1/d)\ln 2}, \frac{\ln((d+2)n^{1+1/d})}{\ln 2}, 1 - \frac{1}{d})$ returns a matching $\{e_1, \dots, e_{n/2}\}$ with expected crossing number at most*

$$\left(6(d+1)^2 + \frac{24e(d+2)}{(1-1/d)^2 \ln 2}\right) n^{1-1/d} + O(\ln(dn) \ln n)$$

with respect to balls in \mathbb{R}^d . The expected running time is $O(dn^{2+1/d} \ln n)$.

Remark. *An algorithm to construct spanning trees with crossing number $O(n^{1-1/d})$ with respect to \mathcal{B}_d can be obtained by randomized LP rounding and has time complexity $\tilde{O}(mn^2)$ (Har-Peled, 2009; Chekuri and Quanrud, 2018), which combined with Lemma 32 yields an $\tilde{O}(n^{3+1/d})$ time algorithm. Alternatively, one can obtain a matching with sub-optimal crossing number $O(n^{1-1/(d+1)})$ by lifting X into \mathbb{R}^{d+1} , where the image of each range in \mathcal{B}_d can be represented by a range in \mathcal{H}_{d+1} and applying the algorithm of Chan (2012) with time complexity $\tilde{O}(n)$.*

Test-sets can also be used within the algorithms LOWDISCCOLOR and APPROXIMATE : since a matching M returned by BUILDMATCHING on a test-set has low crossing number with respect to \mathcal{H}_d , M can be used to construct a low-discrepancy coloring χ with respect to \mathcal{H}_d (recall Lemma 20). In turn, χ can be used to construct a small-sized ε -approximation (recall Lemma 21). These observations lead to the last two corollaries of this section.

Corollary 35. *Let X be a set of n points in \mathbb{R}^d and $\mathcal{T} = \mathcal{T}(n^{1/d})$ be the set of $(d+1)n$ half-spaces provided by Lemma 31. Then*

- i) **LOWDISCCOLOR** $((X, \mathcal{T}), \frac{4e}{\ln 2(1-1/d)}, \frac{\ln((d+1)n)}{\ln 2}, 1 - \frac{1}{d})$ constructs a coloring χ of X of with expected discrepancy at most

$$3\sqrt{\left(6d^2 + \frac{12e(d+1)}{(1-1/d)^2 \ln 2}\right) n^{1-1/d} \ln m + O(\ln(dn) \ln n \ln m)}$$

with respect to half-spaces in \mathbb{R}^d . The expected running time is $O(dn^{2+1/d} \ln n)$.

- ii) if $\varepsilon \in (0, 1)$ and A_0 is a uniform random sample of X of size $\frac{4C_{\text{apx}}(d+1)}{\varepsilon^2}$, then **APPROXIMATE** $((A_0, \mathcal{T}|_{A_0}), \frac{4e}{\ln 2(1-1/d)}, \frac{\ln |\mathcal{T}|_{A_0}|}{\ln 2}, 1 - \frac{1}{d}, \varepsilon)$ returns a set $A \subset X$ of size

$$O\left(\max\left\{\left(\frac{d}{\varepsilon^2} \ln \frac{1}{\varepsilon}\right)^{\frac{d}{d+1}}, \frac{d}{\varepsilon} \ln^{3/2}\left(\frac{d}{\varepsilon}\right)\right\}\right)$$

with expected approximation guarantee satisfying $\mathbb{E}[\varepsilon(A, X, \mathcal{H}_d)] \leq \varepsilon$, and with an expected running time $O\left(n + d\left(\frac{d}{\varepsilon^2}\right)^{2+1/d} \ln \frac{d}{\varepsilon}\right)$.

Corollary 36. *Let X be a set of n points in \mathbb{R}^d and let \mathcal{Q} be the set of balls provided by Lemma 32. Then*

- i) **LOWDISCCOLOR** $((X, \mathcal{Q}), \frac{8e}{\ln 2(1-1/d)}, \frac{\ln((d+2)n^{1+1/d})}{\ln 2}, 1 - \frac{1}{d})$ constructs a coloring χ of X of with expected discrepancy at most

$$3\sqrt{\left(6(d+1)^2 + \frac{24e(d+2)}{(1-1/d)^2 \ln 2}\right) n^{1-1/d} \ln m + O(\ln(dn) \ln n \ln m)}$$

with respect to balls in \mathbb{R}^d . The expected running time is $O(dn^{2+1/d} \ln n)$.

- ii) if $\varepsilon \in (0, 1)$, A_0 is a uniform random sample of X of size $\frac{4C_{\text{apx}}(d+2)}{\varepsilon^2}$, then the algorithm **APPROXIMATE** $((A_0, \mathcal{Q}|_{A_0}), \frac{8e}{\ln 2(1-1/d)}, \frac{\ln |\mathcal{Q}|_{A_0}|}{\ln 2}, 1 - \frac{1}{d}, \varepsilon)$ returns a set $A \subset X$ of size at most

$$O\left(\max\left\{\left(\frac{d}{\varepsilon^2} \ln \frac{1}{\varepsilon}\right)^{\frac{d}{d+1}}, \frac{d}{\varepsilon} \ln^{3/2}\left(\frac{d}{\varepsilon}\right)\right\}\right)$$

with expected approximation guarantee satisfying $\mathbb{E}[\varepsilon(A, X, \mathcal{B}_d)] \leq \varepsilon$, and with an expected running time $O\left(n + d\left(\frac{d}{\varepsilon^2}\right)^{2+1/d} \ln \frac{d}{\varepsilon}\right)$.

4.8 Empirical aspects of the algorithm LOWDISCCOLOR

In this section, present an experimental validation of our algorithm for set systems induced by half-spaces.

4.8.1 Experimental setup

We apply our algorithm for geometric set systems induced by half-spaces in dimensions 2, 4, 6, 8, and 10. We consider two types of inputs:

Grid: each point is picked randomly in a cell of the uniform grid;

Moment: each point is a slightly perturbed element of the moment curve;

These two examples capture two extreme cases: in the case of the **Grid** input, the optimal discrepancy is $\Theta\left(\sqrt{n^{1-1/d}}\right)$ (Alexander, 1990; Matoušek, 1995), while it is $O\left(\sqrt{d}\right)$ in the case of the **Moment** input. All the experiments are performed with dual Xeon E5-2643 v3 processors, each with 6 cores, 12 threads, at 3.4 GHz.

4.8.2 Pre-sampling

Half-spaces. Recall that our algorithm maintains weights on the set of all pairs (edges) in X and on \mathcal{S} , and construct a coloring iteratively, each time coloring a pair of X . In general, half-spaces induce $O(n^d)$ ranges on n points. However, as we have shown in Section 4.7, for our problem at hand it is sufficient to work with a set of $O(dn)$ half-spaces (see Lemma 31, Corollary 35). A test set that achieves the guarantee of Lemma 31 can be constructed via cuttings, which are impractical in higher dimensions. Alternatively, it can be shown that a uniform sample of $\Theta(n \log^d n)$ half-spaces also have the test-set properties stated in Lemma 31 with high probability. Since the study of test sets is not the main focus of this work, our implementation builds the input ranges by $\Theta(n \log n)$ random d -tuples of the input points.

We also measure the quality of our method with respect to a random sample of $O(n \log n)$ half-spaces, which sample is independent of our initial input range set. To verify that our measurements are still reasonably reliable, we compare this method with the theoretically grounded one: in Table 4.2 we present a comparison of the crossing numbers of matchings created by BUILDMATCHING with respect to $O(n \log n)$ random half-spaces, and with respect to $O(n \log^d n)$ random half-spaces.

Input size	Grid									
	$d = 2$		$d = 3$		$d = 4$		$d = 5$		$d = 6$	
	$n \log n$	$n \log^2 n$	$n \log n$	$n \log^3 n$	$n \log n$	$n \log^4 n$	$n \log n$	$n \log^5 n$	$n \log n$	$n \log^6 n$
100	21	20	30	31	29	36	35	39	30	37
250	35	38	55	62	65	73	71	78	66	82
500	51	58	92	100	110	119	118	134	131	149
1000	85	89	161	168	218	232	243	249	237	258
2500	141	144	285	296	411	425	491	509	543	575
5000	215	220	473	482	690	718	888	920	995	1044
10000	347	360	880	818	1241	1258	1595	1637	1962	1910
25000	701	702	1584	1602	2698	2734	3600	3683	4293	4390
50000	1191	1197	2808	2865	4765	4814	6685	6763	7946	8070

Table 4.2: Testing crossing number of matchings with $n \log n$ and $n \log^d n$ random planes.

Edges. To further speed-up our algorithm, we implemented an accelerated version, where instead of maintaining the weights on all the $O(n^2)$ edges, we work with an initial uniform random sample of $O(n^{1+\alpha} \log n)$ edges, where $0 < \alpha \leq 1$ is a parameter to be set. The resulting algorithm **LOWDISCCOLORPRESAMPLED** is presented in Algorithm 6.

The following theorem describes the trade-off between α and the expected discrepancy guarantees.

Theorem 37. *Let (X, \mathcal{S}) be a set system, $n = |X|$, $m = |\mathcal{S}|$, and let c, d be constants such that $\pi_{\mathcal{S}}^*(k) \leq c \cdot k^d$. For any $0 < \alpha \leq 1$, the algorithm **LOWDISCCOLORPRESAMPLED** $((X, \mathcal{S}), d, \alpha)$ constructs a coloring χ of X with expected discrepancy*

$$O\left(\sqrt{n^{1-\alpha/d} \ln m + \ln^2 m \log n}\right),$$

with an expected number of Oracle calls at most

$$O\left(\min\left\{n^{1+\alpha+\alpha/d} \ln^2 n + mn^{\alpha/d} \ln(mn) \cdot \min\{d, \log n\}, n^3 + mn\right\}\right).$$

The core of its proof is the following lemma on matchings in random edge-sets, which might be of independent interest.

Lemma 38. *Let (X, \mathcal{S}) be a set system with dual shatter function $\pi^*(k) = O(k^d)$, $0 < \alpha \leq 1$, and $\delta \in (0, 1)$. Let E be a uniform sample from $\binom{X}{2}$, where each edge is picked i.i.d. with probability*

$$p = \frac{2 \ln n}{n^{1-\alpha}} + \frac{4 \ln(2/\delta)}{n^{2-\alpha}}.$$

Algorithm 6: LOWDISCCOLORPRESAMPLED $((X, \mathcal{S}), d, \alpha)$

```

while  $|X| \geq 16$  do
     $n \leftarrow |X|$ 
     $E \leftarrow$  sample of  $O(n^{1+\alpha} \ln n)$  edges from  $\binom{X}{2}$ 
     $(e_1, \dots, e_{\lceil n/16 \rceil}) \leftarrow \text{MATCHHALF}((X, \mathcal{S}), E, (2c)^{1/d}, \ln m, 1 - \alpha/d, n/16)$ 
    for  $i = 1, \dots, \lceil n/16 \rceil$  do
         $\{x_i, y_i\} \leftarrow$  endpoints  $(e_i)$ 
         $\chi(x_i) = \begin{cases} 1 & \text{with probability } 1/2 \\ -1 & \text{with probability } 1/2 \end{cases}$ 
         $\chi(y_i) = -\chi(x_i)$ 
         $X \setminus \{x_i, y_i\}$ 
    Color the remaining vertices randomly
return  $\chi$ 

```

Then with probability at least $1 - \delta$, E contains a matching of size¹ $n/4$ with crossing number

$$O(n^{1-\alpha/d} + \ln |\mathcal{S}|).$$

Moreover, for any $d \geq 2$, and $n_0 \in \mathbb{N}$ there is a set system (X, \mathcal{S}) with $|X| = n \geq n_0$ and dual shatter function $\pi_{\mathcal{S}}^*(k) = O(k^d)$ such that for any $0 < \alpha \leq 1$ and $p(n) = o(n^{\alpha-1})$ if E is a random edge-set obtained by selecting each edge in $\binom{X}{2}$ i.i.d. with probability $p(n)$, then with probability at least $1/2$, every matching in E of size $n/4$ has crossing number $\omega(n^{1-\alpha/d})$ with respect to \mathcal{S} .

The proofs of Theorem 37 and Lemma 38 are presented later, in Section 4.8.5.

4.8.3 Implementation details

Recall that our algorithm maintains weights on each pre-sampled edge, and these weights can be halved at each iteration. Instead of storing these potentially exponentially small weights explicitly, we simply keep track of how many times the weight of an edge has been updated. In particular, we maintain a partition of the edges into groups such that each group consists of elements that have been updated the same number of times, and thus have the same weight. We store the (exponentially increasing) weights of the test set half-spaces in the same way. To sample an edge or a half-space with respect to the

¹The size of a matching is the number of its edges. A matching of size $n/4$ is a matching of $n/2$ elements of X .

current weights, it suffices to sample from the heaviest $\Theta(\log n)$ groups. The remaining groups have $o\left(\frac{1}{n}\right)$ -th fraction of the total weight, which can be shown to not effect the analysis.

We perform an initial $\frac{n}{4}$ iterations to set more accurate edge weights and start constructing the final coloring only afterwards.

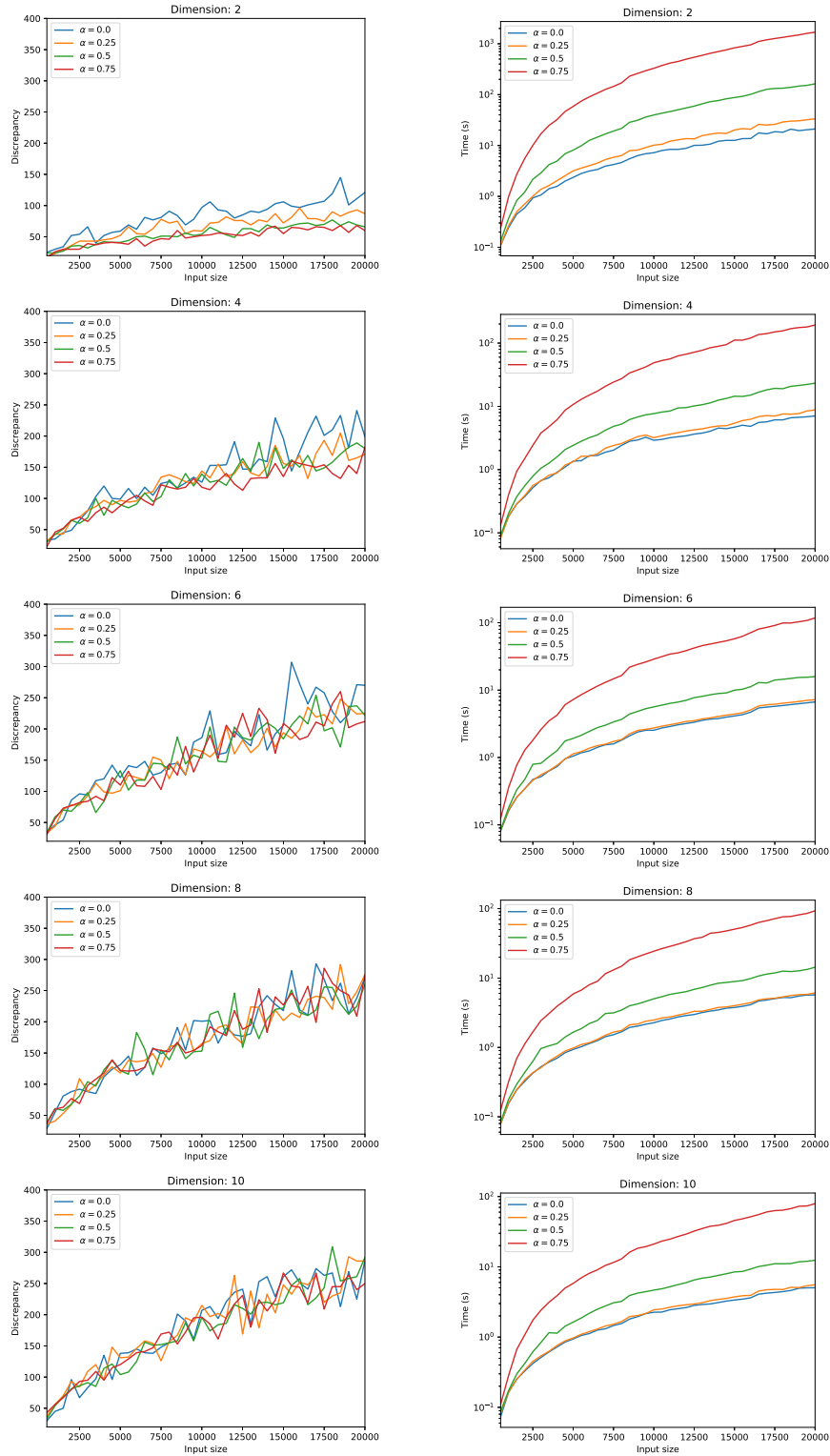


Figure 4.1: Comparing the quality and time of our algorithm with different settings of α on Grid input type.

Theorem 37 states that the discrepancy of the coloring produced by our algorithm depends on the pre-sampling parameter $\alpha \in [0, 1]$. In Figure 4.1, we highlight the impact of α on the discrepancy errors and running times using the `Grid` input. The discrepancy errors are plotted on a linear scale and the running times are in seconds, plotted on a log-scale. As derived in Theorem 37, from the perspective of the discrepancy error, the α parameter becomes less significant with the growth of the dimension. At the same time, α has a significant impact on the running time, even in large dimensions. For the remaining of our experiments, we set $\alpha = 0.25$, which gives a reasonable trade-off between the running time and the improvement on the discrepancy error.

Figure 4.2 demonstrates how the choice of α impacts the underlying matchings of `LOWDISCCOLORPRESAMPLED`.

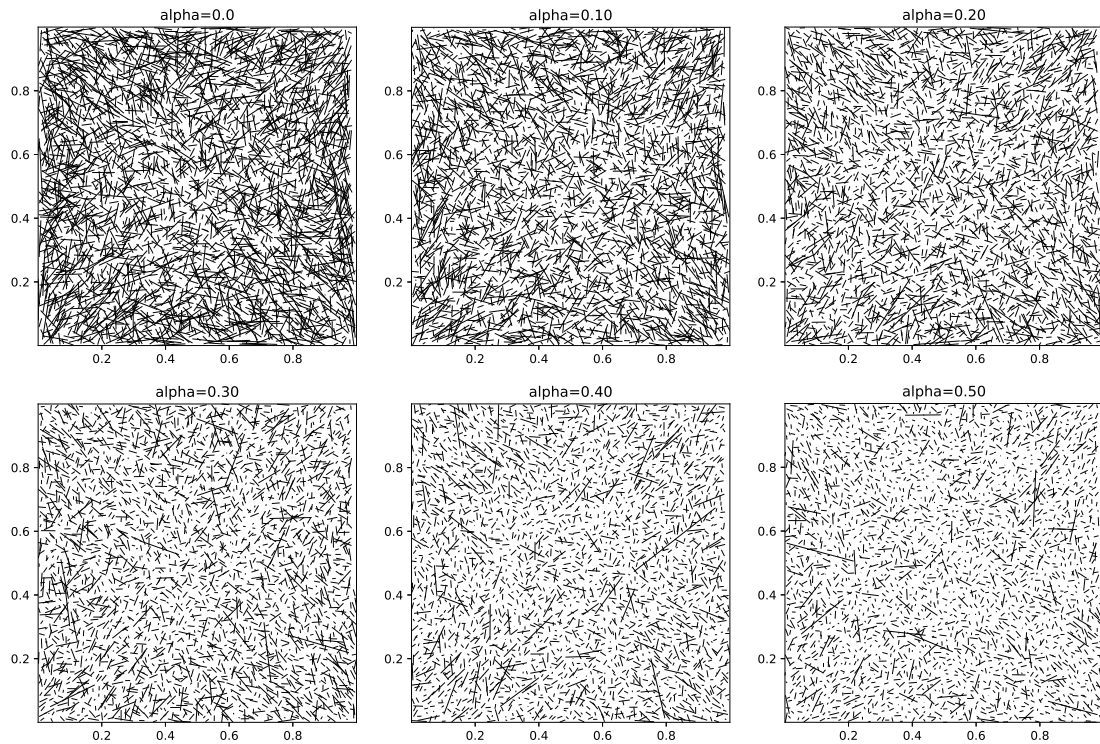


Figure 4.2: Matchings created by our method on 5000 points (`Grid` input type) with different pre-sampling parameters α .

4.8.4 Evaluation

Table 4.3 shows the observed discrepancies and running times on the two types of input: **Grid** and **Moment**. For comparison, we added a column **Rdisc** (in grey), which shows the discrepancy of a random coloring on the same input, with respect to the same quality testing set.

Input size	$d = 2$			$d = 4$			$d = 6$			$d = 8$			$d = 10$		
	disc	time	Rdisc	disc	time	Rdisc	disc	time	Rdisc	disc	time	Rdisc	disc	time	Rdisc
Grid															
100	9	0.02	12	13	0.02	23	13	0.02	24	15	0.02	20	14	0.02	14
250	15	0.05	31	22	0.05	41	26	0.04	24	25	0.05	33	25	0.04	26
500	22	0.11	48	35	0.09	66	37	0.08	47	39	0.09	42	40	0.11	32
1000	22	0.27	31	40	0.18	73	45	0.17	54	57	0.16	48	66	0.17	49
2500	39	1.03	84	67	0.54	94	85	0.45	88	89	0.44	118	76	0.44	94
5000	49	3.12	109	95	1.32	129	97	1.02	143	145	0.94	153	123	0.98	211
10000	66	9.75	206	142	3.16	234	140	2.78	235	191	2.42	181	186	2.31	227
25000	95	48.48	144	222	11.58	294	283	9.66	343	290	8.14	323	313	7.18	422
50000	126	161.40	253	287	31.68	621	338	26.49	454	388	19.08	524	425	16.94	522
100000	184	554.27	407	436	75.98	769	536	54.17	723	630	56.01	694	562	47.79	694
Moment															
100	11	0.02	10	14	0.02	19	13	0.02	31	15	0.02	16	17	0.01	16
250	17	0.06	19	15	0.05	33	25	0.05	26	27	0.04	46	29	0.04	29
500	17	0.12	37	32	0.09	46	31	0.09	69	31	0.09	42	35	0.10	48
1000	27	0.27	47	35	0.18	52	54	0.17	52	52	0.17	60	51	0.16	52
2500	32	1.02	52	54	0.55	91	72	0.46	154	81	0.45	84	86	0.44	83
5000	41	3.20	132	75	1.28	104	112	1.05	129	95	0.96	174	124	0.99	176
10000	54	10.17	170	105	3.14	274	122	2.66	251	161	2.46	218	182	2.34	220
25000	61	47.47	403	164	11.43	310	202	9.33	297	260	7.81	383	265	6.92	426
50000	81	164.04	266	180	30.12	552	382	25.13	649	323	19.02	522	560	16.57	580
100000	106	550.13	789	265	75.07	555	393	53.98	585	511	55.45	936	512	47.58	839

Table 4.3: Experimental results for set systems induced by half-spaces.

The results of Table 4.3 exhibit a consistent improvement of our method over the random coloring, which becomes less pronounced with the growth of the dimension. Note that this phenomenon agrees with our theoretical results and with the known lower bound on the discrepancy. Furthermore, as our theory suggests, our algorithm becomes faster as the dimension increases.

We illustrate these behaviors on Figure 4.3, where we plot the average outcome of our method over 10 repetitions on the **Grid** input type. Interestingly, our experiments suggest that the variance of the discrepancy error increase with the dimension, while the variance of the runtime slightly decreases with dimension. At last, even though we ran

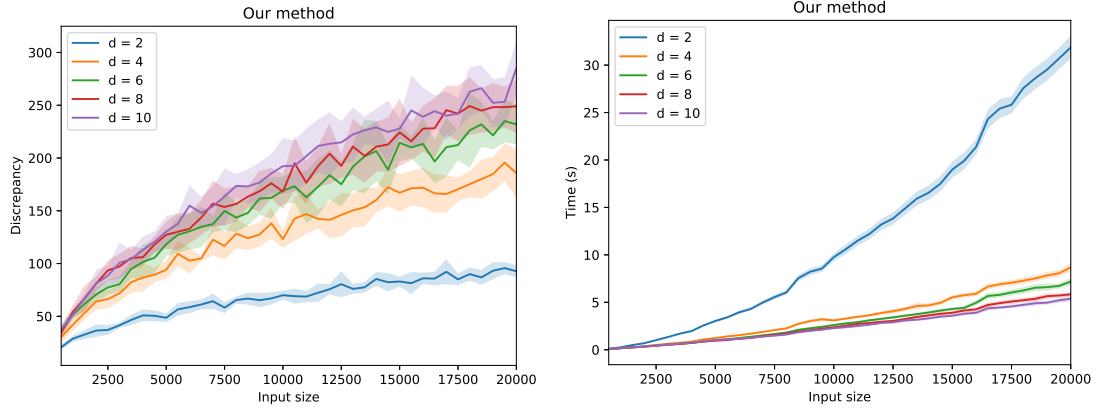


Figure 4.3: Mean and variance of our discrepancy error and runtime on input type **Grid**. The shaded area denote ± 1 standard deviation.

the algorithm with the same parameter setting (we are not adapting to the input) for both input types, the discrepancy errors are consistently smaller for **Moment**.

We also compared the mean and the variance of our discrepancy errors to those of random sampling. Figure 4.4 presents this comparison over 10 repetitions on the **Grid** input type. Again, the average performances of the two methods get closer with the increase of dimension. However, remarkably, the variance of our method is much smaller even on such small number of repetitions.

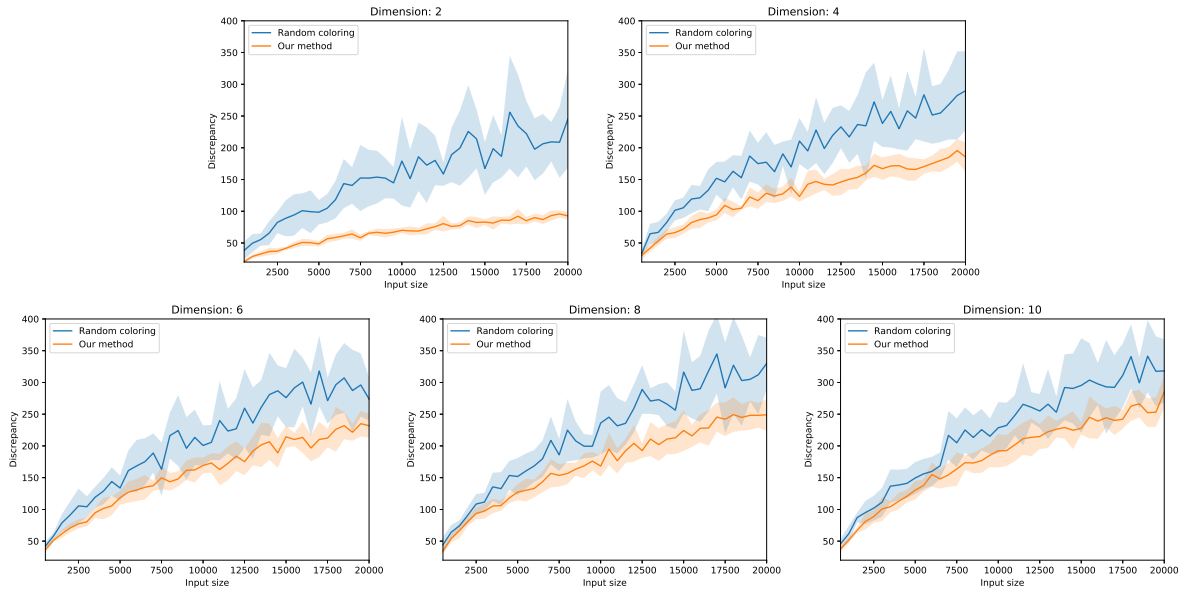


Figure 4.4: Mean and variance of our method and random sampling on input type **Grid**. The shaded areas denote ± 1 standard deviation.

Overall, the empirical evidences presented in this section are coherent with the derived theory and they highlight the often significant (depending on the dimension) improvement over the basic random coloring. Previous experimental results on matchings with low crossing number—the core structure of our algorithm—only presented results for inputs of size at most 159 in \mathbb{R}^2 , see [Giannopoulos et al. \(2014\)](#)².

4.8.5 Proof of pre-sampling guarantees

In this section we prove the discrepancy guarantee of **LOWDISCCOLORPRESAMPLED** stated in Theorem 37. First, we show how Lemma 38 implies Theorem 37, then we prove Lemma 38.

Proof of Theorem 37

By Lemma 20, it is enough to show that the algorithm **MATCHINGPRESAMPLED** (Algorithm 7) constructs a matching with expected crossing number $O(n^{1-\alpha/d} + \ln |\mathcal{S}| \log n)$:

Algorithm 7: **MATCHINGPRESAMPLED** $((X, \mathcal{S}), d, \alpha)$

$M \leftarrow \emptyset$

while $|X| \geq 16$ **do**

$n \leftarrow |X|$

$E \leftarrow$ sample of $O(n^{1+\alpha} \ln n)$ edges from $\binom{X}{2}$

$\{e_1, \dots, e_{\lceil n/16 \rceil}\} \leftarrow \text{MATCHHALF}((X, \mathcal{S}), E, (2c)^{1/d}, \ln m, 1 - \alpha/d, n/16)$

$M \leftarrow M \cup \{e_1, \dots, e_{\lceil n/16 \rceil}\}$

return M

Recall the following statement on **MATCHHALF**.

Lemma 23 (Main Lemma). *Let $t \in [1, |X|/4]$ be an integer and $\tilde{E}_t \subseteq E$ denote the set of edges that have non-zero weight when **MATCHHALF** $((X, \mathcal{S}), E, a, b, \gamma, t)$ terminates. Then*

$$\mathbb{E} \left[\max_{S \in \mathcal{S}} \sum_{i=1}^t \mathbf{I}(e_i, S) \right] \leq \frac{1}{2} \mathbb{E} \left[\min_{e \in \tilde{E}_t} \sum_{i=1}^t \mathbf{I}(e, S_i) \right] + \frac{\left(\frac{128}{13} + \frac{8}{\ln 2}\right) \ln(|E|t)}{3\mathbf{p}} + \frac{\left(16 + \frac{4}{\ln 2}\right) \ln(|\mathcal{S}|t)}{3\mathbf{q}}. \quad (4.3)$$

²The paper studies the related problem of finding spanning trees with low crossing number, which can be used to construct matchings with low crossing number.

Substituting \mathbf{p} and \mathbf{q} , we get the following bound on the expected crossing number of $\{e_1, \dots, e_{\lceil n/8 \rceil}\}$:

$$\mathbb{E} \left[\max_{S \in \mathcal{S}} \sum_{i=1}^{\lceil n/16 \rceil} I(e_i, S) \right] \leq \frac{1}{2} \mathbb{E} \left[\min_{e \in \tilde{E}_t} \sum_{i=1}^{\lceil n/16 \rceil} I(e, S_i) \right] + O(n^{1-\alpha/d}). \quad (4.11)$$

It remains to show that

$$\mathbb{E} \left[\min_{e \in \tilde{E}_t} \sum_{i=1}^{\lceil n/16 \rceil} I(e, S_i) \right] = O(n^{1-\alpha/d}).$$

By Lemma 38, with probability at least $1 - \frac{1}{n}$, E contains a matching of size $n/4$ with crossing number

$$O(n^{1-\alpha/d} + \ln |\mathcal{S}|).$$

Assume that E contains a matching M_0 of size $n/4$ with the above crossing number. Then $M_0 \cap \tilde{E}_{\lceil n/16 \rceil}$ has crossing number $C_0 \cdot (n^{1-\alpha/d} + \ln |\mathcal{S}|)$ with respect to \mathcal{S} and $|M_0 \cap \tilde{E}_{\lceil n/16 \rceil}| \geq n/16$. By the pigeonhole principle, there is an edge in $M_0 \cap \tilde{E}_{\lceil n/16 \rceil}$ which is crossed by at most

$$\frac{C_0 \cdot (n^{1-\alpha/d} + \ln |\mathcal{S}|) \cdot \lceil n/16 \rceil}{n/16} = O(n^{1-\alpha/d} + \ln |\mathcal{S}|)$$

sets from $S_1, \dots, S_{\lceil n/16 \rceil}$.

We conclude that with probability at least $1 - \frac{1}{n}$, the matching returned by **MATCHING-PRESAMPLED** has crossing number $O(n^{1-\alpha/d} + \ln |\mathcal{S}| \log n)$. Since the crossing number of any matching is $O(n)$, the expected crossing number of the matching returned by **MATCHINGPRESAMPLED** is $O(n^{1-\alpha/d} + \ln |\mathcal{S}| \log n)$.

This concludes the proof of Theorem 37. \square

Proof of Lemma 38

In the remaining of this section, we discuss the proof of Lemma 38. Our starting point is the algorithm **RELAXEDMWU** (Algorithm 8), which is a variant of the classical MWU method.

Analysis of the **RELAXEDMWU algorithm.** Assume that for any X and ω , we can find an edge which is crossed by sets of total weight at most $\tau(|X|, \omega(\mathcal{S}))$. Then at each iteration, the total weight of set changes as

$$\omega_{i+1}(\mathcal{S}) \leq \omega_i(\mathcal{S}) + \tau(n - 2(i - 1), \omega_i(\mathcal{S}))$$

Algorithm 8: RELAXEDMWU $((X, \mathcal{S}), \alpha, E)$

 $\omega_1(S) \leftarrow 1$ for all $S \in \mathcal{S}$
for $i = 1, \dots, n/2$ **do**
 $S_i \leftarrow$ the $|X|^{2-\alpha}$ lightest edges in $\binom{X}{2}$ w.r.t. ω_i
if $E \cap S_i = \emptyset$ **then**
 $\quad \lfloor$ set $T = i - 1$ and **return** $\{e_1, \dots, e_{i-1}\}$
else
 $\quad \lfloor$ Pick an edge e_i from $E \cap S_i$
 $\quad \lfloor$ Define ω_{i+1} from ω_i by doubling the weights of each set crossing e_i
 $\quad \lfloor$ $X \leftarrow X \setminus \text{vertices}(e_i)$
 $\quad \lfloor$ **return** $\{e_1, \dots, e_{n/2}\}$

$$\begin{aligned}
&= \omega_i(\mathcal{S}) \left(1 + \frac{\tau(|X| - 2(i-1), \omega_i(\mathcal{S}))}{\omega_i(\mathcal{S})} \right) \\
&\leq \omega_1(\mathcal{S}) \prod_{j=1}^i \left(1 + \frac{\tau(|X| - 2j + 2, \omega_j(\mathcal{S}))}{\omega_j(\mathcal{S})} \right) \\
&= |\mathcal{S}| \cdot \prod_{j=1}^i \left(1 + \frac{\tau(|X| - 2j + 2, \omega_j(\mathcal{S}))}{\omega_j(\mathcal{S})} \right)
\end{aligned}$$

Let $n = |X|$, $t \in [1, n/2]$ be a stopping time and let κ_t denote the maximum number of edges in $\{e_1, \dots, e_t\}$ that are crossed by any set in \mathcal{S} , then by the update rule,

$$\omega_{t+1}(\mathcal{S}) \geq \max_{S \in \mathcal{S}} \omega_{t+1}(S) = 2^{\kappa_t}.$$

We get that

$$\begin{aligned}
2^{\kappa_t} \leq \omega_{t+1}(\mathcal{S}) &\leq |\mathcal{S}| \cdot \prod_{j=1}^t \left(1 + \frac{\tau(n - 2j + 2, \omega_j(\mathcal{S}))}{\omega_j(\mathcal{S})} \right) \\
&\leq |\mathcal{S}| \cdot \exp \left(\sum_{j=1}^t \frac{\tau(n - 2j + 2, \omega_j(\mathcal{S}))}{\omega_j(\mathcal{S})} \right)
\end{aligned}$$

which implies

$$\kappa_t \leq \frac{1}{\ln 2} \left(\ln |\mathcal{S}| + \sum_{j=1}^t \frac{\tau(n - 2j + 2, \omega_j(\mathcal{S}))}{\omega_j(\mathcal{S})} \right). \quad (4.12)$$

We use the following lemma to bound $\tau(\cdot, \cdot)$.

Lemma 39. *Let (X, \mathcal{S}) be a set system with dual shatter function $\pi_{\mathcal{S}}^*(k) \leq c_1 \cdot k^d$. Then for any $Y \subset X$, $w : \mathcal{S} \rightarrow \mathbb{N}$, and parameter $|Y| \leq k \leq \binom{|Y|}{2}$ there are at least k distinct*

edges in $\binom{Y}{2}$ such that any of these edges are crossed by sets of total weight at most $\tau_k(|Y|, w(\mathcal{S})) = (10c_1)^{1/d} \cdot \frac{w(\mathcal{S}) \cdot k^{1/d}}{|Y|^{2/d}}$.

Proof. The proof uses the packing lemma of [Haussler \(1995\)](#).

Lemma 40 (Packing lemma). *Let (X, \mathcal{S}) be a set system with shatter function $\pi_{\mathcal{S}}(k) \leq c_1 \cdot k^d$ and $1 < \delta < |X|$ be a parameter. Furthermore, let $\mathcal{P} \subset \mathcal{S}$ be a δ -separated set, that is, $|S_1 \Delta S_2| \geq \delta$ for any $S_1, S_2 \in \mathcal{P}$. Then*

$$|\mathcal{P}| \leq 2c_1 \left(\frac{|X|}{\delta} \right)^d.$$

Let $(\mathcal{S}_w, \mathcal{R}_Y)$ denote the set system where \mathcal{S}_w contains $w(S)$ copies of each $S \in \mathcal{S}$, $\mathcal{R}_Y = \{R_y : y \in Y\}$, and $R_y = \{S \in \mathcal{S}_w : y \in S\}$. Note that $|\mathcal{S}_w| = w(\mathcal{S})$ and the shatter function of $(\mathcal{S}_w, \mathcal{R}_X)$ is the dual shatter function of (Y, \mathcal{S}) . Let

$$\delta_k = \left(10c_1 \cdot \frac{w(\mathcal{S})^d k}{|Y|^2} \right)^{1/d}$$

By the Packing lemma, any δ_k -separated subset of ranges in \mathcal{R}_Y has cardinality at most

$$C_k = 2c_1 \left(\frac{w(\mathcal{S})}{\delta_k} \right)^d = \frac{|Y|^2}{5k}$$

Observe that for any pair $x, y \in Y$, the set $R_x \Delta R_y$ contains precisely the sets in \mathcal{S}_w that cross the edge xy . Therefore, the Packing Lemma implies that the graph on Y spanned by the edges which are crossed by at least δ_k sets in \mathcal{S}_w does not contain a clique on $C_k + 1$ vertices.

Thus by the classical theorem of [Turán \(1941\)](#), the number of δ_k -short edges is at least

$$C_k \binom{\lfloor |Y|/C_k \rfloor}{2} \geq C_k \cdot \frac{(|Y|/C_k - 1)(|Y|/C_k - 2)}{2} \geq \frac{|Y|^2}{2C_k} - \frac{3|Y|}{2} = \frac{5k}{2} - \frac{3|Y|}{2} = k,$$

where we used that $|Y| \leq k$.

□

Lemma 39 and Equation (4.12) imply that the matching $\{e_1, \dots, e_t\}$ returned by **RELAXEDMWU** has crossing number at most

$$\frac{\ln |\mathcal{S}|}{\ln 2} + \frac{(10c_1)^{1/d}}{\ln 2} \sum_{j=1}^t \frac{1}{(n - 2j + 2)^{\alpha/d}} \leq \frac{\ln |\mathcal{S}|}{\ln 2} + \frac{(10c_1)^{1/d}}{\ln 2} \cdot \frac{t^{1-\alpha/d}}{1 - \alpha/d}. \quad (4.13)$$

By the above analysis, it is enough to show that if $E \subseteq \binom{X}{2}$, where each edge is picked i.i.d with probability

$$p = \frac{2 \ln n}{n^{1-\alpha}} + \frac{4 \ln(2/\delta)}{n^{2-\alpha}},$$

then with probability at least $1 - \delta$, the algorithm **RELAXEDMWU** $((X, \mathcal{S}), \alpha, E)$ returns at least $n/4$ edges. Briefly, we need to show that

$$\mathbb{P}[T \leq n/4] \leq \delta,$$

where T denotes the number of edges returned by **RELAXEDMWU** on the set E . We can express the above probability as

$$\mathbb{P}[T \leq n/4] = \sum_{i=0}^{n/4} \mathbb{P}[T = i],$$

therefore it is enough to bound the probabilities $\mathbb{P}[T = i]$ for each $i = 1, \dots, n/4$. Since E is an i.i.d. uniform random sample of $\binom{X}{2}$,

$$\mathbb{P}[T = 1] = \mathbb{P}[E \cap S_1 = \emptyset] = (1 - p)^{|S_1|}.$$

Observe that in iteration $i \geq 2$ of the algorithm **RELAXEDMWU**, the short-edge set S_i depends on the previously picked edges. To signify this, for any set of edges e_1, \dots, e_{i-1} , we denote the set of $(n - 2i)^\alpha$ shortest edges of $X \setminus \text{vertices}(e_1, \dots, e_{i-1})$ as $S_i(e_1, \dots, e_{i-1})$. We say that a vector of edges (e^1, \dots, e^i) is *feasible* if $e^1 \in S_1, e^2 \in S_2(e^1), \dots, e^i \in S_i(e^1, \dots, e^{i-1})$. Observe that

$$\begin{aligned} \mathbb{P}[T = i + 1] &= \mathbb{P}[E \cap S_1 \neq \emptyset, E \cap S_2 \neq \emptyset, \dots, E \cap S_{i+1} = \emptyset] \\ &= \sum_{(e^1, \dots, e^i) \text{ feasible}} \mathbb{P}[E \cap S_{i+1} = \emptyset, E \cap S_j \neq \emptyset \forall j \in [1, i] \mid e_j = e^j, \forall j \in [1, i]] \\ &\quad \cdot \mathbb{P}[e_j = e^j, \forall j \in [1, i]] \\ &= \sum_{(e^1, \dots, e^i) \text{ feasible}} \mathbb{P}[E \cap S_{i+1}(e^1, \dots, e^i) = \emptyset \mid e_j = e^j, \forall j \in [1, i]] \mathbb{P}[e_j = e^j, \forall j \in [1, i]]. \end{aligned} \tag{4.14}$$

Note that $S_{i+1}(e^1, \dots, e^i)$ is a fixed, non-random set. We use Bayes' theorem which states that for events A, B

$$\mathbb{P}[A \mid B] = \frac{\mathbb{P}[B \mid A] \cdot \mathbb{P}[A]}{\mathbb{P}[B]}.$$

Using this rule, we can express the conditional probabilities in the right hand side of Equation (4.14) as

$$\mathbb{P}[E \cap S_{i+1}(e^1, \dots, e^i) = \emptyset \mid e_j = e^j, \forall j \in [1, i]]$$

$$= \frac{\mathbb{P} [e_j = e^j, \forall j \in [1, i] \mid E \cap S_{i+1}(e^1, \dots, e^i) = \emptyset] \cdot \mathbb{P} [E \cap S_{i+1}(e^1, \dots, e^i) = \emptyset]}{\mathbb{P} [e_j = e^j, \forall j \in [1, i]]}.$$

Substituting this back to Equation (4.14), we get

$$\begin{aligned} & \mathbb{P} [T = i + 1] \\ &= \sum_{(e^1, \dots, e^i) \text{ feasible}} \mathbb{P} [e_j = e^j, \forall j \in [1, i] \mid E \cap S_{i+1}(e^1, \dots, e^i) = \emptyset] \\ & \quad \cdot \mathbb{P} [E \cap S_{i+1}(e^1, \dots, e^i) = \emptyset] \\ &= \sum_{(e^1, \dots, e^i) \text{ feasible}} \mathbb{P} [e_j = e^j, \forall j \in [1, i] \mid E \cap S_{i+1}(e^1, \dots, e^i) = \emptyset] \cdot (1 - p)^{|S_{i+1}(e^1, \dots, e^i)|} \end{aligned} \quad (4.15)$$

Recall that $|S_{i+1}(e^1, \dots, e^i)| = (n - 2i)^\alpha$ for any choice of e^1, \dots, e^i . Therefore, the only (so far) unknown factor in the right hand side of Equation (4.15) is the probability

$$\mathbb{P} [e_j = e^j, \forall j \in [1, i] \mid E \cap S_{i+1}(e^1, \dots, e^i) = \emptyset].$$

We continue by bounding this probability. For brevity, let $\mathbf{e} = (e^1, \dots, e^i)$ be a fixed feasible vector of edges and $S_{i+1}(\mathbf{e}) = S_{i+1}(e^1, \dots, e^i)$. Observe that

$$\begin{aligned} & \mathbb{P} [e_j = e^j, \forall j \in [1, i] \mid E \cap S_{i+1}(\mathbf{e}) = \emptyset] \\ &= \mathbb{P} [e_i = e^i \mid E \cap S_{i+1}(\mathbf{e}) = \emptyset, e_j = e^j, \forall j \in [1, i - 1]] \\ & \quad \cdot \mathbb{P} [e_j = e^j, \forall j \in [1, i - 1] \mid E \cap S_{i+1}(\mathbf{e}) = \emptyset] \\ &= \prod_{j=2}^i \mathbb{P} [e_j = e^j \mid E \cap S_{i+1}(\mathbf{e}) = \emptyset, e_\ell = e^\ell, \forall \ell \in [1, j - 1]] \\ & \quad \cdot \mathbb{P} [e_1 = e^1 \mid E \cap S_{i+1}(\mathbf{e}) = \emptyset] \end{aligned}$$

Recall that e_1 was picked uniformly at random from $S_1 \cap E$, where S_1 is a fixed set such that $e^1 \in S_1$, and E is random. This implies the following for $\mathbb{P} [e_1 = e^1 \mid E \cap S_{i+1}(\mathbf{e}) = \emptyset]$

$$\begin{aligned} & \mathbb{P} [e_1 = e^1 \mid E \cap S_{i+1}(\mathbf{e}) = \emptyset] \\ &= \sum_{S' \subset S_1} \mathbb{P} [e_1 = e^1 \mid E \cap S_{i+1}(\mathbf{e}) = \emptyset, E \cap S_1 = S'] \cdot \mathbb{P} [E \cap S_1 = S' \mid E \cap S_{i+1}(\mathbf{e}) = \emptyset] \\ &= \sum_{e^1 \in S' \subset S_1 \setminus S_{i+1}(\mathbf{e})} \frac{1}{|S'|} \cdot p^{|S'|} \cdot (1 - p)^{|S_1 \setminus S_{i+1}(\mathbf{e})| - |S'|} \\ &= \sum_{\ell=1}^{|S_1 \setminus S_{i+1}(\mathbf{e})|} \binom{|S_1 \setminus S_{i+1}(\mathbf{e})| - 1}{\ell - 1} \cdot \frac{1}{\ell} \cdot p^\ell \cdot (1 - p)^{|S_1 \setminus S_{i+1}(\mathbf{e})| - \ell} \end{aligned}$$

$$\begin{aligned}
&= \sum_{\ell=1}^{|S_1 \setminus S_{i+1}(\mathbf{e})|} \frac{1}{|S_1 \setminus S_{i+1}(\mathbf{e})|} \binom{|S_1 \setminus S_{i+1}(\mathbf{e})|}{\ell} \cdot p^\ell \cdot (1-p)^{|S_1 \setminus S_{i+1}(\mathbf{e})|-\ell} \\
&= \frac{1}{|S_1 \setminus S_{i+1}(\mathbf{e})|} (p + (1-p))^{|S_1 \setminus S_{i+1}(\mathbf{e})|} - (1-p)^{|S_1 \setminus S_{i+1}(\mathbf{e})|} \\
&= \frac{1}{|S_1 \setminus S_{i+1}(\mathbf{e})|} (1 - (1-p)^{|S_1 \setminus S_{i+1}(\mathbf{e})|}).
\end{aligned}$$

The last piece is to bound the probabilities

$$\mathbb{P} [e_j = e^j \mid E \cap S_{i+1}(\mathbf{e}) = \emptyset, e_\ell = e^\ell, \forall \ell \in [1, j-1]]$$

for $j \geq 2$. Note that, given the choices of the edges $e_\ell = e^\ell$ for $\ell = 1, \dots, j-1$, the set

$$S_j = S_j(e^1, \dots, e^{j-1})$$

is not random, and thus we have a similar relation as before

$$\begin{aligned}
&\mathbb{P} [e_j = e^j \mid E \cap S_{i+1}(\mathbf{e}) = \emptyset, e_\ell = e^\ell, \forall \ell \in [1, j-1]] \\
&= \mathbb{P} [e_j = e^j \mid E \cap S_{i+1}(\mathbf{e}) = \emptyset] \\
&= \sum_{e^j \in S' \subset S_j \setminus S_{i+1}(\mathbf{e})} \frac{1}{|S'|} \cdot p^{|S'|} \cdot (1-p)^{|S_j \setminus S_{i+1}(\mathbf{e})| - |S'|} \\
&= \sum_{\ell=1}^{|S_j \setminus S_{i+1}(\mathbf{e})|} \binom{|S_j \setminus S_{i+1}(\mathbf{e})| - 1}{\ell - 1} \cdot \frac{1}{\ell} \cdot p^\ell \cdot (1-p)^{|S_j \setminus S_{i+1}(\mathbf{e})| - \ell} \\
&= \frac{1}{|S_j \setminus S_{i+1}(\mathbf{e})|} (1 - (1-p)^{|S_j \setminus S_{i+1}(\mathbf{e})|}).
\end{aligned}$$

Observe that for each $1 \leq j \leq i$ we have

$$(n - 2(j-1))^\alpha - (n - 2i)^\alpha \leq |S_j \setminus S_{i+1}(\mathbf{e})| \leq (n - 2(j-1))^\alpha$$

and that the above probability is maximized if

$$|S_j \setminus S_{i+1}(\mathbf{e})| = (n - 2(j-1))^\alpha - (n - 2i)^\alpha. \quad .37$$

Putting all this together and using the notation $k_i = (n - 2(i-1))^\alpha$, we get

$$\begin{aligned}
\mathbb{P} [T = i + 1] &= (1-p)^{k_{i+1}} \cdot \sum_{\mathbf{e} \text{ feasible}} \prod_{j=1}^i \frac{1}{|S_j \setminus S_{i+1}(\mathbf{e})|} (1 - (1-p)^{|S_j \setminus S_{i+1}(\mathbf{e})|}) \\
&\leq (1-p)^{k_{i+1}} \cdot \sum_{\mathbf{e} \text{ feasible}} \prod_{j=1}^i \frac{1}{k_j - k_{i+1}} (1 - (1-p)^{k_j - k_{i+1}})
\end{aligned}$$

$$\begin{aligned}
&= (1-p)^{k_{i+1}} \cdot |S_1| \cdot |S_2| \cdots |S_i| \cdot \prod_{j=1}^i \frac{1}{k_j - k_{i+1}} (1 - (1-p)^{k_j - k_{i+1}}) \\
&= (1-p)^{k_{i+1}} \cdot \prod_{j=1}^i \frac{k_j}{k_j - k_{i+1}} (1 - (1-p)^{k_j - k_{i+1}})
\end{aligned}$$

For any $i \geq 2$, we get that the probability of $T \leq i+1$ can be bounded as

$$\begin{aligned}
\mathbb{P}[T \leq i+1] &= \sum_{j=1}^{i+1} \mathbb{P}[T = j] \\
&\leq (1-p)^{k_1} + \sum_{\ell=1}^{i+1} (1-p)^{k_{\ell+1}} \cdot \prod_{j=1}^{\ell} \frac{k_j}{k_j - k_{\ell+1}} (1 - (1-p)^{k_j - k_{\ell+1}})
\end{aligned}$$

Using the bounds $k_1 \geq k_2 \dots k_{n/4-1} \geq k_{n/4}$ and $k_j - k_{i+1} \geq 1$ for all $1 \leq j \leq i$, we can bound the probability of $T \leq n/4$ as

$$\begin{aligned}
\mathbb{P}[T \leq n/4] &\leq (1-p)^{k_{n/4}} + (1-p)^{k_{n/4}} \sum_{\ell=1}^{n/4} \prod_{j=1}^{\ell} k_j \cdot p \leq (1-p)^{k_{n/4}} \sum_{\ell=0}^{n/4} (k_1 \cdot p)^{\ell} \\
&= (1-p)^{k_{n/4}} \frac{1 - (pk_1)^{n/4+1}}{1 - pk_1} \leq 2(1-p)^{k_{n/4}} \cdot (pk_1)^{n/4} \leq 2 \exp(-pk_{n/4}) \cdot k_1^{n/4}
\end{aligned}$$

Substituting $k_1 = n^{2-\alpha}$, $k_{n/4} \geq (n/2)^{2-\alpha} \geq n^{2-\alpha}/4$ and $p = \frac{2 \ln n}{n^{1-\alpha}} + \frac{4 \ln(2/\delta)}{n^{2-\alpha}}$, we conclude

$$\mathbb{P}[T \leq n/4] \leq 2 \exp(-(n/2) \ln n - \ln(2/\delta)) \cdot (n^{2-\alpha})^{n/4} = 2 \cdot \frac{n^{n/2 - \alpha n/4}}{n^{n/2}} \cdot \frac{\delta}{2} \leq \delta$$

Therefore, with probability at least $1 - \delta$, **RELAXEDMWU** returns a matching of size $n/4$. This, together with Equation (4.13) implies the upper bound of Lemma 38.

Lower bound construction. We define X to be a subset of the integer grid, in particular $X = \times_{i=1}^d \left[1, \left\lceil n_0^{1/d} \right\rceil \right] \subset \mathbb{Z}^d$. Let \mathcal{S} consist of all subsets of X induced by half-spaces of the form

$$\left\{ x_i \leq j + 1/2 \mid i = 1, \dots, d, j = 1, \dots, \left\lfloor n_0^{1/d} \right\rfloor \right\}.$$

Observe that for any edge $\{\mathbf{x}, \mathbf{y}\} \in \binom{X}{2}$, the number of sets in \mathcal{S} that crosses $\{\mathbf{x}, \mathbf{y}\}$ is precisely the ℓ_1 -distance of \mathbf{x} and \mathbf{y} , which is defined as

$$\ell_1(\mathbf{x}, \mathbf{y}) = \sum_{i=1}^d |x_i - y_i|.$$

Using this observation, it is easy to see that for any fixed k , the number of edges crossed by at most k sets from \mathcal{S} is at most nk^d . We refer to these edges as k -good and denote their set with G_k .

Now we set $k = \left(\frac{1}{16p(n)}\right)^{1/d}$ and examine the expected number of k -good edges in E .

$$\mathbb{E}[|E \cap G_k|] \leq nk^d \cdot p(n) = \frac{n}{16}$$

By Markov's inequality,

$$\mathbb{P}\left[|E \cap G_k| \leq \frac{n}{8}\right] \geq \frac{1}{2}.$$

Thus, with probability at least $1/2$, we have $|E \cap G_k| \leq \frac{n}{8}$. Assume that this event holds and let $M \subset E$ be any subset of size $n/4$. Then M contains at least $n/8$ edges which are not k -good. Therefore, the number of crossings between the edges of M and the sets of \mathcal{S} is at least

$$\frac{n}{8} \cdot \left(\frac{1}{16p(n)}\right)^{1/d}$$

Recall that $|\mathcal{S}| = d \cdot n^{1/d}$ and so by the pigeonhole principle, we get that there is a set in \mathcal{S} that crosses at least

$$\frac{\frac{n}{8} \cdot \left(\frac{1}{16p(n)}\right)^{1/d}}{dn^{1/d}} = \frac{n^{1-1/d}}{8d \cdot (16)^{1/d}} \cdot \underbrace{\left(\frac{1}{p(n)}\right)^{1/d}}_{\omega(n^{(1-\alpha)/d})} = \omega(n^{1-\alpha/d})$$

edges of M .

□

Chapter 5

VC dimension of unions

*The results presented in this chapter were published in the article
Tight Lower Bounds on the VC dimension of Geometric Set Systems
Journal of Machine Learning Research 20(81):1-8, 2019.
It is a joint work with Andrey Kupavskii and Nabil Mustafa.*

In this chapter, we provide tight lower bounds for the VC dimension of two fundamental set systems: k -fold unions/intersections of half-spaces and the simplices set system. Among other implications, our result settles an open question in machine learning that was first studied in the foundational paper of [Blumer et al. \(1989\)](#) as well as by [Eisenstat and Angluin \(2007\)](#) and [Johnson \(2008\)](#).

As before, let (X, \mathcal{S}) denote a *set system*, where X is a set of elements and \mathcal{S} is a set of subsets of X . For any integer $k \geq 2$, define the k -fold union of \mathcal{S} as the set system induced on X by the ranges $\mathcal{S}^{k\cup} = \{S_1 \cup \dots \cup S_k : S_1, \dots, S_k \in \mathcal{S}\}$. Similarly, one can define the k -fold intersection of \mathcal{S} , denoted by $\mathcal{S}^{k\cap}$, as the set system consisting of all subsets derived from the common intersection of at most k sets of \mathcal{S} . Note that as the ranges S_1, \dots, S_k need not necessarily be distinct, we have $\mathcal{S} \subseteq \mathcal{S}^{k\cup}$ and $\mathcal{S} \subseteq \mathcal{S}^{k\cap}$. Analogously, the k -fold symmetric difference of \mathcal{S} is defined as $\mathcal{S}^{k\oplus} = \{S_1 \oplus \dots \oplus S_k : S_1, \dots, S_k \in \mathcal{S}\}$, where $S_1 \oplus \dots \oplus S_k$ is the set of those elements that are contained in an odd number of S_1, \dots, S_k .

Theorem 41 ([Blumer et al. 1989](#)). *Let (X, \mathcal{S}) be a set system and k be any positive integer. Then*

$$d_{\text{VC}}(\mathcal{S}^{k\cup}) = O(d_{\text{VC}}(\mathcal{S}) \cdot k \log k),$$

$$d_{\text{VC}}(\mathcal{S}^{k\cap}) = O(d_{\text{VC}}(\mathcal{S}) \cdot k \log k).$$

Moreover, there are set systems such that $d_{VC}(\mathcal{S}^{k\cup}) = \Omega(d_{VC}(\mathcal{S}) \cdot k)$ and $d_{VC}(\mathcal{S}^{k\cap}) = \Omega(d_{VC}(\mathcal{S}) \cdot k)$.

Remark. The upper bound of Theorem 41 holds in a more general setting: for any fixed set-theoretic expression $F(S_1, S_2, \dots, S_k)$ (consisting of operations of set union, intersection, and difference) and range set $\mathcal{S}^{k*} := \{F(S_1, \dots, S_k) : S_1, \dots, S_k \in \mathcal{S}\}$, we have $d_{VC}(\mathcal{S}^{k*}) = O(d_{VC}(\mathcal{S}) \cdot k \log k)$ (see eg. Matoušek, 2002, chap. 10), in particular, $d_{VC}(\mathcal{S}^{k\oplus}) = O(d_{VC}(\mathcal{S}) \cdot k \log k)$.

Half-spaces. Blumer et al. (1989) also considered the question of whether the upper bounds of Theorem 41 are tight in the most basic geometric case when $X \subseteq \mathbb{R}^d$ is a set of points and \mathcal{S} is the projection of the family of all half-spaces of \mathbb{R}^d onto X . They proved that the VC dimension of the k -fold union of half-spaces in two dimensions is exactly $2k + 1$. For general dimensions $d \geq 3$, they upper-bound the VC dimension of the k -fold union of half-spaces by $O(d \cdot k \log k)$. This follows from Theorem 41 together with the fact that the VC dimension of the set system induced by half-spaces in \mathbb{R}^d is $d + 1$. The same upper bound holds for the VC dimension of the k -fold intersection of half-spaces in \mathbb{R}^d . Later Dobkin and Gunopulos (1995) showed that the VC dimension of the k -fold union of half-spaces in \mathbb{R}^3 is upper-bounded by $4k$.

Eisenstat and Angluin (2007) proved, by giving a probabilistic construction of an abstract set system, that the upper bound of Theorem 41 is asymptotically tight if $d_{VC}(\mathcal{S}) \geq 5$ and that for $d_{VC}(\mathcal{S}) = 1$, an upper bound of k holds which is tight. A few years later, Eisenstat (2009) filled the gap by showing that there exists a set system (X, \mathcal{S}) of VC dimension 2 such that $d_{VC}(\mathcal{S}^{k\cup}) = \Omega(d_{VC}(\mathcal{S}) \cdot k \log k)$.

For $d \geq 4$, the current best upper-bound for the k -fold union and the k -fold intersection of half-spaces in \mathbb{R}^d is still the one given by Theorem 41 almost 30 years ago, while the lower-bound has remained $\Omega(d_{VC}(\mathcal{S}) \cdot k)$. We refer the reader to the PhD thesis of Johnson (2008) for a summary of the bounds on VC dimensions of these basic combinatorial and geometric set systems.

The resolution of the VC dimension of k -fold unions and intersections of half-spaces is left as one of the main open problems in the thesis.

For some time now, it has generally been expected that

$$d_{VC}(\mathcal{S}^{k\cup}) = d_{VC}(\mathcal{S}^{k\cap}) = O(dk)$$

for the k -fold unions and intersections of half-spaces. This upper-bound indeed holds for a related notion: the *primal shattering dimension* of the k -fold unions and intersections

of half-spaces is $O(dk)$. In fact, as it was pointed out by [Bachem \(2018\)](#), several papers in learning theory assume the same for VC dimension. Likewise for computational geometry literature: for example, the coresset size bounds in the constructions of [Feldman and Langberg \(2011\)](#), [Balcan et al. \(2013\)](#), and [Lucic et al. \(2016\)](#) would require an additional $\log k$ factor in the coresset size—if the upper-bound of Theorem 41 was tight for the k -fold intersection of half-spaces.

The main result of this chapter is a lower-bound construction, which proves that the $O(d \cdot k \log k)$ upper bound of Theorem 41 is in fact tight.

Theorem 42. *Let k be a given positive integer and $d \geq 4$ be an integer. Then there exists a set P of points in \mathbb{R}^d such that the set system \mathcal{S} induced on P by half-spaces satisfies*

$$\begin{aligned} \text{a) } d_{\text{VC}}(\mathcal{S}^{k\cup}) &= \Omega(d_{\text{VC}}(\mathcal{S}) \cdot k \log k) = \Omega(d \cdot k \log k), \\ \text{b) } d_{\text{VC}}(\mathcal{S}^{k\cap}) &= \Omega(d_{\text{VC}}(\mathcal{S}) \cdot k \log k) = \Omega(d \cdot k \log k), \\ \text{c) } d_{\text{VC}}(\mathcal{S}^{k\oplus}) &= \Omega(d_{\text{VC}}(\mathcal{S}) \cdot k \log k) = \Omega(d \cdot k \log k). \end{aligned}$$

The proof of Theorem 42 is presented in Section 5.1. Our construction is deterministic, therefore it also provides a non-probabilistic proof of the lower-bound of [Eisenstat and Angluin \(2007\)](#). Note that a set of points is shattered by (the k -fold union of) closed half-spaces if and only if it is shattered by (the k -fold union of) open half-spaces. Thus Theorem 42 also holds for open half-spaces.

Simplices. The following set system is fundamental in computational geometry. Given a set \mathcal{H} of hyperplanes in \mathbb{R}^d , define

$$\Delta(\mathcal{H}) = \left\{ \mathcal{H}' \subseteq \mathcal{H} : \exists \text{ an open } d\text{-dimensional simplex } S \text{ in } \mathbb{R}^d \text{ such that} \right. \\ \left. H \in \mathcal{H}' \text{ if and only if } H \text{ intersects } S \right\}. \quad (5.1)$$

Its importance derives from the fact that it is the set system underlying the construction of *cuttings* via random sampling (see [Chazelle and Friedman 1990](#) as well as the recent survey of [Mustafa and Varadarajan 2017](#)). Cuttings are the key tool for fast point-location algorithms and were studied in detail recently by [Ezra et al. \(2020\)](#). They provided the best bounds so far for the VC dimension of $\Delta(\mathcal{H})$.

Lemma 43 ([Ezra et al. 2020](#)). *For $d \geq 9$, we have*

$$d(d+1) \leq d_{\text{VC}}(\Delta(\mathcal{H})) \leq 5 \cdot d^2 \log d.$$

Using Theorem 42, we show an asymptotically optimal bound on the VC dimension of $\Delta(\mathcal{H})$, improving the bound of Ezra et al. (2020) and resolving a question that was studied in the computational geometry community starting in the 1980s. Our result holds for a more general set system on hyperplanes induced by the ranges

$$\Delta_k(\mathcal{H}) = \left\{ \mathcal{H}' \subseteq \mathcal{H} : \exists \text{ an open } k\text{-dimensional simplex } S \text{ in } \mathbb{R}^d \text{ such that } H \in \mathcal{H}' \text{ if and only if } H \text{ intersects } S \right\}.$$

In Section 5.2, we prove the following lower bound on the VC dimension of $(\mathcal{H}, \Delta_k(\mathcal{H}))$.

Theorem 44. *For any integer $d \geq 4$ and $k \leq d$, there exists a set \mathcal{H} of $\Omega(dk \log k)$ hyperplanes in \mathbb{R}^d such that $|\Delta_k(\mathcal{H})| = 2^{|\mathcal{H}|}$, that is, we have*

$$d_{\text{VC}}(\Delta_k(\mathcal{H})) = \Omega(d \cdot k \log k).$$

5.1 Lower bound for unions of half-spaces

In this section, we present the proof of Theorem 42. Let Q a set of points in \mathbb{R}^d and \mathcal{R} be a collection of sets in \mathbb{R}^d and define $\overline{\mathcal{R}} := \{\mathbb{R}^d \setminus R : R \in \mathcal{R}\}$. Observe that $d_{\text{VC}}(Q, \overline{\mathcal{R}}) = d_{\text{VC}}(Q, \mathcal{R})$ and that

$$d_{\text{VC}}(Q, \mathcal{R}^{k\cap}) = d_{\text{VC}}(Q, \overline{\mathcal{R}^{k\cap}}) = d_{\text{VC}}(\overline{\mathcal{R}}^{k\cup}).$$

holds by the De Morgan laws. Since for half-spaces $\overline{\mathcal{R}} = \mathcal{R}$, part a) of Theorem 42 implies part b).

We prove parts a) and c) for d even, starting from $d = 4$. The asymptotic lower-bound for odd values of d follows from the one in \mathbb{R}^{d-1} . The starting point of the proof is the following lemma.

Lemma 45 (Kupavskii et al., 2016, Lemma 2). *Let $n, d' \geq 2$ be integers. Then there exists a set $\mathcal{B}_{n,d'}$ of axis-parallel boxes in $\mathbb{R}^{d'}$, with $|\mathcal{B}_{n,d'}| = (d' - 1)(n + 1)2^{n-2}$, such that for any subset $\mathcal{S} \subseteq \mathcal{B}_{n,d'}$, there exists a set \mathcal{Q} of points in $\mathbb{R}^{d'}$ such that $|\mathcal{Q}| = 2^{n-1}$ and*

(i) $|\mathcal{Q} \cap B| = 1$ for any $B \in \mathcal{B}_{n,d'} \setminus \mathcal{S}$, and

(ii) $\mathcal{Q} \cap B = \emptyset$ for any $B \in \mathcal{S}$.

Remark. In Kupavskii et al. (2016) this lemma is stated in a weaker form, however the above stronger statement is implicit in their proof.

Let $d' = \lfloor \frac{d}{2} \rfloor$ and apply Lemma 45 with $n = \lfloor \log k \rfloor + 1$ in $\mathbb{R}^{d'}$ to get a set $\mathcal{B} = \mathcal{B}_{n,d'}$ of boxes in $\mathbb{R}^{d'}$. By translation, we can assume that all coordinates of points lying in each box in \mathcal{B} are positive. Now we will construct the following mappings:

$$\begin{array}{ccccc} \text{set of boxes } \mathcal{B} \text{ in } \mathbb{R}^{d/2} & \xrightarrow{\pi} & \text{points in } \mathbb{R}^d, \\ \text{points in } \mathbb{R}^{d/2} & \xrightarrow{\beta} & \text{boxes in } \mathbb{R}^d & \xrightarrow{\gamma} & \text{half-spaces in } \mathbb{R}^d. \end{array}$$

We will then prove the key property of these mappings, that for any $q \in \mathbb{R}^{d/2}$ and $B \in \mathcal{B}$

$$q \in B \xLeftrightarrow{(5.2)} \pi(B) \in \beta(q) \xLeftrightarrow{(5.3)} \pi(B) \in \gamma(\beta(q)).$$

We first define mappings π and β , with $\pi : \mathcal{B} \rightarrow \mathbb{R}^d$ and with β mapping points in $\mathbb{R}^{d'}$ to axis-parallel boxes in \mathbb{R}^d , such that for any $B \in \mathcal{B}$ and $q \in \mathbb{R}^{d'}$, we have

$$q \in B \iff \pi(B) \in \beta(q). \quad (5.2)$$

- Let $B \in \mathcal{B}$ be defined as the product of d' intervals:

$$B = [x_1, x'_1] \times [x_2, x'_2] \times \cdots \times [x_{d'}, x'_{d'}], \quad \text{with } x_i, x'_i > 0 \text{ for each } i \in [d'].$$

Then π maps B in $\mathbb{R}^{d'}$ to the following point in \mathbb{R}^d (see [Pach and Tardos, 2013](#))

$$\pi(B) = \left(x_1, \frac{1}{x'_1}, x_2, \frac{1}{x'_2}, \dots, x_{d'}, \frac{1}{x'_{d'}} \right) \in \mathbb{R}^d,$$

and we let $\pi(\mathcal{B}) := \{\pi(B) : B \in \mathcal{B}\}$.

- For any point $q = (q_1, q_2, \dots, q_{d'}) \in \mathbb{R}^{d'}$, define $\beta(q)$ to be the box

$$\beta(q) = [0, q_1] \times \left[0, \frac{1}{q_1}\right] \times \cdots \times [0, q_{d'}] \times \left[0, \frac{1}{q_{d'}}\right] \subset \mathbb{R}^d.$$

Proposition 46. *The mappings π and β satisfy (5.2).*

Proof. Let $q = (q_1, q_2, \dots, q_{d'})$ be a point in $\mathbb{R}^{d'}$ and $B = [x_1, x'_1] \times [x_2, x'_2] \times \cdots \times [x_{d'}, x'_{d'}]$ be an axis-parallel box in $\mathbb{R}^{d'}$. Then $q \in B$ if and only if $x_i \leq q_i \leq x'_i$ for all $i \in [d']$.

On the other hand, $\pi(B)$ lies in $\beta(q)$ if and only if $0 \leq x_i \leq q_i$ and $0 \leq 1/x'_i \leq 1/q_i$ for each $i \in [d']$ —or equivalently, $0 \leq x_i \leq q_i$ and $q_i \leq x'_i$ for each $i \in [d']$.

Note that these two conditions are exactly the same, implying (5.2). \square

Recall that $\mathcal{B} = \mathcal{B}_{n,d'}$ is the set of boxes provided by Lemma 45 with parameters $d' = \lfloor \frac{d}{2} \rfloor$ and $n = \lfloor \log k \rfloor + 1$. Thus we have

$$|\mathcal{B}| = \left(\left\lfloor \frac{d}{2} \right\rfloor - 1 \right) \lfloor \log k \rfloor 2^{\lfloor \log k \rfloor - 1}.$$

Proposition 47 (‘Lifted dual version’ of Lemma 45). *$\pi(\mathcal{B})$ is a set of*

$$\left(\left\lfloor \frac{d}{2} \right\rfloor - 1 \right) \lfloor \log k \rfloor 2^{\lfloor \log k \rfloor - 1}$$

points in \mathbb{R}^d such that for any subset $\mathcal{T} \subseteq \pi(\mathcal{B})$, there is a set \mathcal{Q}^ of at most k axis-parallel boxes in \mathbb{R}^d such that*

- (i) each point of \mathcal{T} is contained in exactly one box in \mathcal{Q}^* , and*
- (ii) no point of $\pi(\mathcal{B}) \setminus \mathcal{T}$ is contained in any box of \mathcal{Q}^* .*

In particular, $\pi(\mathcal{B})$ is shattered by the set system induced by the k -fold union of axis-parallel boxes in \mathbb{R}^d and also shattered by the set system induced by the k -fold symmetric difference of axis-parallel boxes in \mathbb{R}^d .

Proof. Let $\mathcal{S} = \{\pi^{-1}(p) : p \in \pi(\mathcal{B}) \setminus \mathcal{T}\}$. By Lemma 45, there is a set \mathcal{Q} of at most k points in $\mathbb{R}^{d'}$ such that (i), (ii) of Lemma 45 hold for \mathcal{S} and \mathcal{Q} . Letting $\mathcal{Q}^* = \{\beta(q) : q \in \mathcal{Q}\}$, the claim follows from (5.2). \square

Next we define the function $\gamma(\cdot)$ mapping boxes in \mathbb{R}^d to half-spaces in \mathbb{R}^d such that for any point $p \in \pi(\mathcal{B})$ and box $B = \beta(q)$, we have

$$p \in B \iff p \in \gamma(B). \quad (5.3)$$

For every $i \in [d]$, let $0 < \alpha_{i,1} < \alpha_{i,2} < \dots$ denote the sequence of distinct values of the x_i -coordinates of the elements of $\pi(\mathcal{B})$. Every such sequence has length at most $|\pi(\mathcal{B})|$. By re-scaling the coordinates, we can assume that

$$\text{for each } i \in [d] \text{ and } j \leq |\pi(\mathcal{B})|, \quad \frac{\alpha_{i,j+1}}{\alpha_{i,j}} > d. \quad (5.4)$$

Denote the resulting point set by P . Note that scaling along each coordinate does not change incidences with respect to axis-parallel boxes, thus Proposition 47 still holds if we replace $\pi(\mathcal{B})$ by P and that $|P| = |\pi(\mathcal{B})| = |\mathcal{B}|$.

We claim that P is shattered by the set system induced by the k -fold union of half-spaces in \mathbb{R}^d and also shattered by the set system induced by the k -fold symmetric difference of half-spaces in \mathbb{R}^d . To see that, let P' be any subset of P . Let \mathcal{Q}^* be the set of axis-parallel boxes corresponding to P' provided by Proposition 47.

For each box $B \in \mathcal{Q}^*$, we can re-scale B if necessary, without changing its intersection with P so that B is of the form

$$B = [0, b_1] \times [0, b_2] \times \cdots \times [0, b_d],$$

where each b_i is equal to α_{i,j_i} , for a suitable j_i . Now for each box $B \in \mathcal{Q}^*$, we define a half-space $\gamma(B)$ as the set of points $(x_1, \dots, x_d) \in \mathbb{R}^d$ satisfying

$$\frac{x_1}{b_1} + \frac{x_2}{b_2} + \cdots + \frac{x_d}{b_d} \leq d. \quad (5.5)$$

Clearly for any point (x_1, \dots, x_d) contained in the box B , we have $x_i \in [0, b_i]$ and thus each term on the left-hand side of the Equation (5.5) is at most 1. This implies that $B \subset \gamma(B)$ and so any point in the box B lies in the half-space $\gamma(B)$.

On the other hand, for any point $(x_1, \dots, x_d) \in P \setminus B$, there exists an index $i \in [d]$ such that $x_i > b_i$. By (5.4), we have that $x_i/b_i > d$, and thus (x_1, \dots, x_d) cannot lie in the half-space $\gamma(B)$.

Consider the set of at most k half-spaces defined as $\mathcal{H} = \{\gamma(B) : B \in \mathcal{Q}^*\}$. Now by Proposition 47 and (5.3), we have

- (i) each point of P' is contained in exactly one half-space in \mathcal{H} , and
- (ii) no point of $P \setminus P'$ is contained in any half-space of \mathcal{H} .

In other words, the union as well as the symmetric difference of the half-spaces in \mathcal{H} contains precisely the set P' . As this is true for any $P' \subseteq P$, the k -fold union of half-spaces in \mathbb{R}^d shatters P and the same holds for the k -fold symmetric difference of half-spaces in \mathbb{R}^d . Finally, we have

$$|P| = |\mathcal{B}| = \left(\left\lfloor \frac{d}{2} \right\rfloor - 1 \right) \lfloor \log k \rfloor 2^{\lfloor \log k \rfloor - 1} = \Omega(d \cdot k \log k),$$

as desired. □

5.2 Lower bound for simplices

In this section, we prove Theorem 44. Let P be the set of $\Omega(dk \log k)$ points in \mathbb{R}^d provided by Theorem 42. Using point-hyperplane duality (see eg. Matoušek, 2002), map each point $p \in P$ to a hyperplane $\alpha(p)$ by

$$p = (p_1, \dots, p_d) \longmapsto \alpha(p) := \{(x_1, \dots, x_d) : p_1 x_1 + p_2 x_2 + \cdots + p_{d-1} x_{d-1} - x_d = -p_d\}.$$

Our desired set \mathcal{H} of hyperplanes will simply be

$$\mathcal{H} = \{\alpha(p) : p \in P\}.$$

It is easy to check that the mapping α is injective and thus $|\mathcal{H}| = |P| = \Omega(dk \log k)$. We claim that \mathcal{H} is shattered by the set system induced by open k -dimensional simplices; in other words, for any $\mathcal{H}' \subseteq \mathcal{H}$, there exists a k -dimensional simplex S such that the interior of S intersects each hyperplane of \mathcal{H}' , and no hyperplane of $\mathcal{H} \setminus \mathcal{H}'$.

Fix any $\mathcal{H}' \subseteq \mathcal{H}$ and let $P' = \alpha^{-1}(\mathcal{H}')$ be the corresponding points of P . Since P is shattered by the k -fold union of open half-spaces there, exists a set $\mathfrak{H}(P')$ of k open half-spaces whose union contains all points in P' and no point in $P \setminus P'$. From Equation (5.5), it follows that each half-space in $\mathfrak{H}(P')$ is of the form

$$\frac{x_1}{b_1} + \frac{x_2}{b_2} + \cdots + \frac{x_d}{b_d} < d,$$

where b_1, \dots, b_d are positive reals. Map each half-space $H \in \mathfrak{H}(P')$ to the point $\delta(H)$, given by

$$H = \left\{ (x_1, \dots, x_d) : \frac{x_1}{b_1} + \frac{x_2}{b_2} + \cdots + \frac{x_d}{b_d} < d \right\} \quad \mapsto \quad \delta(H) := \left(\frac{b_d}{b_1}, \dots, \frac{b_d}{b_{d-1}}, d \cdot b_d \right).$$

It is easy to verify that for a point $p \in \mathbb{R}^d$ and the half-space H , we have

$$\begin{aligned} p \in H &\iff \frac{p_1}{b_1} + \frac{p_2}{b_2} + \cdots + \frac{p_d}{b_d} < d \\ &\iff p_1 \cdot \frac{b_d}{b_1} + p_2 \cdot \frac{b_d}{b_2} + \cdots + p_{d-1} \cdot \frac{b_d}{b_{d-1}} + p_d < d \cdot b_d \\ &\iff \text{point } \left(\frac{b_d}{b_1}, \frac{b_d}{b_2}, \dots, \frac{b_d}{b_{d-1}}, d \cdot b_d \right) \text{ lies strictly above the hyperplane} \\ &\quad p_1 x_1 + p_2 x_2 + \cdots + p_{d-1} x_{d-1} - x_d = -p_d \\ &\iff \text{the point } \delta(H) \text{ lies strictly above the hyperplane } \alpha(p). \end{aligned} \tag{5.6}$$

Here we crucially needed the fact that all half-spaces in $\mathfrak{H}(P')$ are downward facing, that is, each half-space in $\mathfrak{H}(P')$ contains the origin, which lies below (with respect to the x_d -coordinate) its bounding hyperplane.

Now consider the k open half-spaces in $\mathfrak{H}(P')$ and let

$$\Delta' = \{\delta(H) : H \in \mathfrak{H}(P')\}$$

be k points in \mathbb{R}^d . From the relation (5.6), it follows that

- As each point $p \in P'$ lies in *some* half-space $H \in \mathfrak{H}(P')$, the point $\delta(H)$ lies strictly above the hyperplane $\alpha(p)$ in \mathcal{H} —or equivalently, the hyperplane $\alpha(p)$ has at least one of the k points in the set Δ' lying strictly above it.
- For each point $p \in P \setminus P'$, all the k points in Δ' lie on or below the hyperplane $\alpha(p) \in \mathcal{H}$.

Then, by the above discussion, $H \in \mathcal{H}'$ if and only if one of these is true:

1. H intersects the interior of $\text{conv}(\Delta')$ and so at least one vertex of Δ' lies strictly above H , or
2. H does not intersect $\text{conv}(\Delta')$, but then *all* vertices of Δ' lie strictly above H .

Finally consider the k -dimensional simplex

$$S = \text{conv} \left(\Delta' \cup (0, \dots, 0, -\infty) \right).$$

Clearly, a hyperplane $H \in \mathcal{H}$ intersects the interior of S if and only if $H \in \mathcal{H}'$. Note that the point $(0, \dots, 0, -\infty)$ can be any point $(0, \dots, 0, t)$ for a small-enough value of $t \in \mathbb{R}$. This concludes the proof.

Chapter 6

Epsilon-nets

6.1 Previous results

Given set system (X, \mathcal{S}) and a parameter $0 < \varepsilon < 1$, a set $N \subseteq X$ is an ε -net of (X, \mathcal{S}) if any range $S \in \mathcal{S}$ with size $|S| > \varepsilon|X|$ contains at least one element of N . In other words, an ε -net is a hitting set for the ε -heavy ranges of \mathcal{S} .

The notion of ε -nets was introduced by [Haussler and Welzl \(1987\)](#), who observed that, just like for ε -approximations, the key property of (X, \mathcal{S}) which dictates the size of an ε -net is the VC-dimension. In particular, they proved the following statement.

Theorem 48. *Let (X, \mathcal{S}) be a set system with VC-dimension at most d and $0 < \varepsilon, \delta < 1$ be two parameters. Then a uniform sample N of*

$$\max \left\{ \frac{8d}{\varepsilon} \ln \frac{8d}{\varepsilon}, \frac{4}{\varepsilon} \ln \frac{2}{\delta} \right\}$$

points from X is an ε -net of (X, \mathcal{S}) with probability at least $1 - \delta$.

They also provided the following lower bound on the ε -net sizes in set systems with bounded VC-dimension: for any $d \geq 1$ and $0 < \varepsilon < 1$, there is a set system with VC-dimension at most d in which any ε -net has size at least

$$\left\lfloor \frac{d}{2} \right\rfloor \frac{1}{\varepsilon} - 1.$$

Let $f_d(\varepsilon)$ be the smallest integer such that every set system of VC-dimension d has an ε -net of size at most $f_d(\varepsilon)$. Using this notation, we can summarize the results of [Haussler and Welzl \(1987\)](#) as

$$\left\lfloor \frac{d}{2} \right\rfloor \frac{1}{\varepsilon} - 1 \leq f_d(\varepsilon) \leq \left\lceil \frac{8d}{\varepsilon} \ln \frac{8d}{\varepsilon} \right\rceil.$$

Later Komlós et al. (1992) showed that if $d = 1$, $f_1(\varepsilon) = \max\{2, \lceil 1/\varepsilon \rceil - 1\}$. In case of $d \geq 2$, they proved almost matching upper and lower bounds for $f_d(\varepsilon)$. We summarize their results in the next theorem.

Theorem 49.

Let (X, \mathcal{S}) be a set system with VC-dimension d and $0 < \varepsilon < \varepsilon_0$, where $\varepsilon_0 < 1$ is a parameter sufficiently small in terms of d . Then a uniform sample N of

$$\frac{d}{\varepsilon} \left[\ln \left(\frac{1}{\varepsilon} \right) + 2 \ln \ln \left(\frac{1}{\varepsilon} \right) + 3 \right]$$

points from X is an ε -net for (X, \mathcal{S}) with large probability (roughly $1 - e^{-d}$). Moreover, for any $d \geq 2$ and any real $\gamma < 2/(d+2)$ there exists $\varepsilon_0(d, \gamma)$ such that for all $\varepsilon \leq \varepsilon_0(d, \gamma)$ there is a set system (X, \mathcal{S}) of VC-dimension at most d which does not have an ε -net smaller than $(d - 2 + \gamma)/\varepsilon \ln(1/\varepsilon)$.

Summing up, they established the bounds

$$d - 2 + \frac{2}{d+2} \leq \liminf_{\varepsilon \rightarrow 0} \frac{f_d(\varepsilon)}{1/\varepsilon \ln(1/\varepsilon)} \leq d$$

for any $d \geq 2$.

6.2 Our results

We provide an upper bound on $f_d(\varepsilon)$ of

$$\frac{d}{\varepsilon} \ln \frac{1}{\varepsilon} + o\left(\frac{d}{\varepsilon} \ln \frac{1}{\varepsilon}\right)$$

while preserving the feature of arbitrarily large success probability.

Theorem 50. Let (X, \mathcal{S}) be a set system of VC-dimension d , $0 < \delta < 1$, and $0 < \varepsilon < \varepsilon_0$ for some ε_0 small enough in terms of d and δ . Let

$$m = \frac{d}{\varepsilon} \left(\ln \frac{1}{\varepsilon} + \frac{1}{d} \ln \frac{1}{\delta} + 2 \ln \left(\ln \frac{1}{\varepsilon} + \frac{1}{d} \ln \frac{1}{\delta} \right) + 5 \right),$$

and let N be a uniform sample of m i.i.d. points from X . Then N is an ε -net for (X, \mathcal{S}) with probability at least $1 - \delta$.

In addition, we provide a lemma that states a novel connection between the VC-dimension of the k -fold union $\mathcal{S}^{k\cup}$ and lower bounds for the sizes of ε -nets for \mathcal{S} .

Lemma 51. *Let $0 < \varepsilon < 1$ be a parameter, (X, \mathcal{S}) be a set system and $f : \mathbb{N} \rightarrow \mathbb{R}$ be a function that satisfies*

$$d_{\text{VC}}(X, \mathcal{S}^{k \cup}) \geq d_{\text{VC}}(X, \mathcal{S}) \cdot k \cdot f(k).$$

Then there is a subset X' of X so that any ε -net for $(X', \mathcal{S}|_{X'})$ must have size at least

$$\frac{d_{\text{VC}}(X', \mathcal{S}|_{X'})}{4\varepsilon} \cdot f\left(\frac{1}{2\varepsilon}\right).$$

Proof of Theorem 50

Without loss of generality, we can assume that $|S| > \varepsilon|X|$ holds for each range $S \in \mathcal{S}$. We shall approximate the probability of failure. By definition,

$$\mathbb{P}[N \text{ is not an } \varepsilon\text{-net}] = \mathbb{P}[\exists S \in \mathcal{S} : S \cap N = \emptyset].$$

The following claim was proved in [Komlós et al. \(1992\)](#) for a different choice of m . Although the proof for our choice of m is similar, we include it for the sake of completeness.

Claim 5. *Let $M > m$, then*

$$\mathbb{P}[\exists S \in \mathcal{S} : S \cap N = \emptyset] \leq 2 \max_{|Y|=M} |\mathcal{S}|_Y \cdot \left(1 - \frac{m}{M}\right)^{(M-m)\varepsilon-1}. \quad (6.1)$$

Proof. Let Y' be a sample of $M - m$ points from X drawn uniformly and independently from the first sample N and consider the (multi-)set $Y = N \cup Y'$ of M points. We will bound the probabilities of the following two events:

$$\begin{aligned} E_0 &:= \exists S \in \mathcal{S} \text{ such that } S \cap N = \emptyset; \\ E_\mu &:= \exists S \in \mathcal{S} \text{ such that } S \cap N = \emptyset \text{ and } |S \cap Y'| > \mu, \end{aligned}$$

where μ is the median¹ of the multi-set $\{|Y' \cap S| : S \in \mathcal{S}\}$. We show that for this choice of μ , $\mathbb{P}[E_0] \leq 2 \cdot \mathbb{P}[E_\mu]$ holds. To this end, consider the conditional probabilities $\mathbb{P}[E_0 | N]$ and $\mathbb{P}[E_\mu | N]$ for given N and Y' random.

If N is an ε -net, then $S \cap N \neq \emptyset$ for all $S \in \mathcal{S}$ and thus

$$\mathbb{P}[E_0 | N] = \mathbb{P}[E_\mu | N] = 0.$$

¹In case of even cardinality n , we take the $(n/2)^{\text{th}}$ smallest element.

If N is not an ε -net, then there is a range $S \in \mathcal{S}$ for which $S \cap N = \emptyset$, that is, $\mathbb{P}[E_0 \mid N] = 1$. Fix one such range and denote it by S_N . Clearly,

$$\mathbb{P}[E_\mu \mid N] \geq \mathbb{P}[|S_N \cap Y'| > \mu].$$

Since μ is the median of the multi-set $\{|Y' \cap S| : S \in \mathcal{S}\}$, we have

$$\min_{S \in \mathcal{S}} \mathbb{P}[|S \cap Y'| > \mu] \geq 1/2.$$

Thus, we conclude that

$$\mathbb{P}[E_\mu \mid N] \geq \mathbb{P}[|S_N \cap Y'| > \mu] \geq \min_{S \in \mathcal{S}} \mathbb{P}[|S \cap Y'| > \mu] \geq \frac{1}{2} = \frac{1}{2} \cdot \mathbb{P}[E_0 \mid N].$$

Hence for any choice of N ,

$$\mathbb{P}[E_\mu \mid N] \geq \frac{1}{2} \cdot \mathbb{P}[E_0 \mid N].$$

Taking total expectations we obtain

$$\mathbb{P}[E_0] \leq 2 \cdot \mathbb{P}[E_\mu]. \quad (6.2)$$

Now we count $\mathbb{P}[E_\mu]$ differently. Instead of sampling first the elements of N and then (independently) the elements of Y' , we first sample the multi-set Y of M elements and then sample m elements from Y to form N and define Y' as $Y \setminus N$.

First we bound the probability of the event $\{S \cap N = \emptyset \text{ and } |S \cap Y'| > \mu\}$ for a fixed set $S \in \mathcal{S}$ conditioned on the choice of Y :

$$\begin{aligned} \mathbb{P}\left[S \cap N = \emptyset \text{ and } |S \cap Y'| > \mu \mid Y\right] &\leq \mathbf{1}_{\{|S \cap Y| > \mu\}} \cdot \frac{\binom{M-m}{|S \cap Y|}}{\binom{M}{|S \cap Y|}} \\ &\leq \mathbf{1}_{\{|S \cap Y| > \mu\}} \left(1 - \frac{m}{M}\right)^{|S \cap Y|} \leq \left(1 - \frac{m}{M}\right)^\mu, \end{aligned}$$

where $\mathbf{1}_{\mathcal{A}}$ denotes the indicator random variable of the event \mathcal{A} . Applying the union bound over all different ranges in $\mathcal{S}|_Y$, we get

$$\mathbb{P}\left[\exists S \in \mathcal{S} : S \cap N = \emptyset \text{ and } |S \cap Y'| > \mu \mid Y\right] \leq \max_{|Y|=M} |\mathcal{S}|_Y \cdot \left(1 - \frac{m}{M}\right)^\mu \quad (6.3)$$

Recall that we set μ to be the median of the multi-set $\{|Y' \cap S| : S \in \mathcal{S}\}$. As the median of a binomial distribution is within 1 of the mean, $\mu \geq (M - m)\varepsilon - 1$. We conclude the proof by putting together the inequalities (6.2), (6.3) and the bound on μ .

□

The right-hand side of (6.1) can be reformulated as

$$2 \max_{|Y|=M} |\mathcal{S}|_Y \cdot \left(1 - \frac{m}{M}\right)^{(M-m)\varepsilon-1} = 2 \max_{|Y|=M} |\mathcal{S}|_Y \cdot \exp \left[[(M-m)\varepsilon - 1] \ln \left(1 - \frac{m}{M}\right) \right]. \quad (6.4)$$

Since $0 < m/M < 1$ and since $\ln(1-x) = -\sum_{k=1}^{\infty} (x^k/k)$ for all positive $x \in \mathbb{R}$, we have $\ln(1-m/M) < -m/M$ and thus the exponent can be bounded as

$$[(M-m)\varepsilon - 1] \ln \left(1 - \frac{m}{M}\right) < [(M-m)\varepsilon - 1] \cdot \left(-\frac{m}{M}\right) = \left(\frac{m^2}{M} - m\right) \varepsilon + \frac{m}{M} \quad (6.5)$$

By the Sauer-Shelah lemma (Lemma 6),

$$\max_{|Y|=M} |\mathcal{S}|_Y \leq \left(\frac{eM}{d}\right)^d. \quad (6.6)$$

Now putting together (6.1), (6.4), (6.5), and (6.6) we have

$$\begin{aligned} \mathbb{P}[\exists S \in \mathcal{S} : S \cap N = \emptyset] &\leq 2 \left(\frac{eM}{d}\right)^d \cdot \exp \left[\left(\frac{m^2}{M} - m\right) \varepsilon + \frac{m}{M} \right] \\ &= 2 \cdot \exp \left[d + d \ln \frac{M}{d} + \frac{\varepsilon m^2}{M} - \varepsilon m + \frac{m}{M} \right] \end{aligned} \quad (6.7)$$

Let

$$M = \frac{4d}{\varepsilon} \left(\ln \frac{1}{\varepsilon} + \frac{1}{d} \ln \frac{1}{\delta} \right)^2 = \frac{4d}{\varepsilon} \left(\ln \frac{1}{\varepsilon \sqrt[d]{\delta}} \right)^2$$

and recall that

$$\begin{aligned} m &= \frac{d}{\varepsilon} \left(\ln \frac{1}{\varepsilon} + \frac{1}{d} \ln \frac{1}{\delta} + 2 \ln \left(\ln \frac{1}{\varepsilon} + \frac{1}{d} \ln \frac{1}{\delta} \right) + 5 \right) \\ &= \frac{d}{\varepsilon} \left(\ln \frac{1}{\varepsilon \sqrt[d]{\delta}} + 2 \ln \left(\ln \frac{1}{\varepsilon \sqrt[d]{\delta}} \right) + 5 \right). \end{aligned}$$

Clearly, $M > m$ for small enough values of ε .

Now we would like to bound the exponent in (6.7). We start with bounding the sum $\varepsilon m^2/M + m/M$.

If $\ln \frac{1}{\varepsilon \sqrt[d]{\delta}} > 2 \ln \left(\ln \frac{1}{\varepsilon \sqrt[d]{\delta}} \right) + 5$, then

$$\frac{\varepsilon m^2}{M} + \frac{m}{M} = \frac{\varepsilon \left[\frac{d}{\varepsilon} \left(\ln \frac{1}{\varepsilon \sqrt[d]{\delta}} + 2 \ln \left(\ln \frac{1}{\varepsilon \sqrt[d]{\delta}} \right) + 5 \right) \right]^2}{\frac{4d}{\varepsilon} \left(\ln \frac{1}{\varepsilon \sqrt[d]{\delta}} \right)^2} + \frac{\frac{d}{\varepsilon} \left(\ln \frac{1}{\varepsilon \sqrt[d]{\delta}} + 2 \ln \left(\ln \frac{1}{\varepsilon \sqrt[d]{\delta}} \right) + 5 \right)}{\frac{4d}{\varepsilon} \left(\ln \frac{1}{\varepsilon \sqrt[d]{\delta}} \right)^2}$$

$$\begin{aligned}
&= \frac{d \left[\ln \frac{1}{\varepsilon \sqrt[d]{\delta}} + 2 \ln \left(\ln \frac{1}{\varepsilon \sqrt[d]{\delta}} \right) + 5 \right]^2}{\left(2 \ln \frac{1}{\varepsilon \sqrt[d]{\delta}} \right)^2} + \frac{\left[\ln \frac{1}{\varepsilon \sqrt[d]{\delta}} + 2 \ln \left(\ln \frac{1}{\varepsilon \sqrt[d]{\delta}} \right) + 5 \right]}{\left(2 \ln \frac{1}{\varepsilon \sqrt[d]{\delta}} \right)^2} \\
&\leq \frac{d \left[2 \ln \frac{1}{\varepsilon \sqrt[d]{\delta}} \right]^2}{\left(2 \ln \frac{1}{\varepsilon \sqrt[d]{\delta}} \right)^2} + \frac{\left[2 \ln \frac{1}{\varepsilon \sqrt[d]{\delta}} \right]}{\left(2 \ln \frac{1}{\varepsilon \sqrt[d]{\delta}} \right)^2} = d + \frac{1}{2 \ln \frac{1}{\varepsilon \sqrt[d]{\delta}}} \leq \frac{3d}{2}
\end{aligned} \tag{6.8}$$

Now the whole exponent of (6.7) simplifies as

$$\begin{aligned}
&d \ln \frac{M}{d} - m\varepsilon + \frac{\varepsilon m^2}{M} + \frac{m}{M} + d \leq \\
&\stackrel{(6.8)}{\leq} d \ln \left(\frac{4}{\varepsilon} \left(\ln \frac{1}{\varepsilon \sqrt[d]{\delta}} \right)^2 \right) - d \left[\ln \frac{1}{\varepsilon \sqrt[d]{\delta}} + 2 \ln \left(\ln \frac{1}{\varepsilon \sqrt[d]{\delta}} \right) + 5 \right] + \frac{5d}{2} = \\
&= d \ln \left(\frac{4}{\varepsilon} \left(\ln \frac{1}{\varepsilon \sqrt[d]{\delta}} \right)^2 \right) - d \left[\ln \frac{1}{\varepsilon \sqrt[d]{\delta}} + 2 \ln \left(\ln \frac{1}{\varepsilon \sqrt[d]{\delta}} \right) + \frac{5}{2} \right] = \\
&= d \left[\ln \frac{1}{\varepsilon} + \ln 4 + 2 \ln \left(\ln \frac{1}{\varepsilon} + \frac{1}{d} \ln \frac{1}{\delta} \right) \right] \\
&\quad - d \left[\ln \frac{1}{\varepsilon} + \frac{1}{d} \ln \frac{1}{\delta} + 2 \ln \left(\ln \frac{1}{\varepsilon} + \frac{1}{d} \ln \frac{1}{\delta} \right) + \frac{5}{2} \right] = \\
&= -\ln \frac{1}{\delta} + \left(\ln 4 - \frac{5}{2} \right) d \leq -\ln \frac{1}{\delta} - d
\end{aligned}$$

After simplifying the exponent, we get the bound for all $d \geq 1$

$$\mathbb{P}[\exists S \in \mathcal{S} : S \cap N = \emptyset] \leq 2 \cdot e^{-\ln(1/\delta) - d} = 2 \cdot \delta \cdot e^{-d} < \delta$$

This concludes the proof of Theorem 50. □

Proof of Lemma 51

Denote $d_{VC}(X, \mathcal{S})$ by d and let

$$\ell = \frac{1}{2\varepsilon}.$$

As $d_{VC}(\mathcal{S}^{\ell \cup}) \geq d \cdot \ell \cdot f(\ell)$, there is a subset X' of X of size $|X'| = d \cdot \ell \cdot f(\ell)$ which is shattered by $\mathcal{S}^{\ell \cup}$. We will show that if N is an ε -net for $(X', \mathcal{S}|_{X'})$, then

$$|N| \geq \frac{|X'|}{2} = \frac{d \cdot \ell \cdot f(\ell)}{2} = \frac{d}{4\varepsilon} \cdot f\left(\frac{1}{2\varepsilon}\right).$$

Suppose that $N < \frac{|X'|}{2}$. Since X' is shattered by $\mathcal{S}^{\ell \cup}$, there is a range in $\mathcal{S}^{\ell \cup}$ containing precisely the elements in $X' \setminus N$. In particular, we can find ℓ sets $S_1, \dots, S_\ell \in \mathcal{S}$ such that

$$(S_1 \cup \dots \cup S_\ell) \cap X' = X' \setminus N.$$

Note that each set S_i contains no point of N . On the other hand, by the pigeonhole principle, one of the sets $S_1|_{X'}, \dots, S_\ell|_{X'}$ must have size at least

$$\frac{|X' \setminus N|}{\ell} \geq \frac{|X'|/2}{\ell} = \frac{|X'|/2}{\frac{1}{2\varepsilon}} = \varepsilon|X'|.$$

Therefore there is a range in $\mathcal{S}|_{X'}$ of size at least $\varepsilon|X'|$ that is not hit by N , a contradiction to the fact that N was an ε -net for $(X', \mathcal{S}|_{X'})$. This concludes the proof of Lemma 51. \square

Chapter 7

Simple proof of optimal approximations

*The proof method presented in this chapter was published in the article
Optimal approximations made easy
Information Processing Letters.
It is a joint work with Nabil Mustafa.*

In this chapter, we present a simple, self-contained proof of the influential result of [Li et al. \(2001\)](#) for finite set systems. Their result was stated for (ε, δ) -samples: we say that A is an (ε, δ) -sample for (X, \mathcal{S}) if for any $S \in \mathcal{S}$,

$$d_\varepsilon \left(\frac{|S|}{|X|}, \frac{|S \cap A|}{|A|} \right) < \delta, \quad \text{where} \quad d_\varepsilon(a, b) = \frac{|a - b|}{a + b + \varepsilon}.$$

The notion of (ε, δ) -samples is asymptotically equivalent to the following more combinatorial measure. Given a set system (X, \mathcal{S}) with $n = |X|$ and parameters $0 < \varepsilon, \delta < 1$, a set A of size t is a *relative (ε, δ) -approximation* for (X, \mathcal{S}) if for all $S \in \mathcal{S}$,

$$\left| \frac{|S|}{n} - \frac{|A \cap S|}{t} \right| \leq \delta \cdot \max \left\{ \frac{|S|}{n}, \varepsilon \right\}.$$

The equivalence of relative approximations and samples was shown in [Har-Peled and Sharir \(2011\)](#), in particular, they proved that an (ε, δ) -sample is a relative $(\varepsilon, 4\delta)$ -approximation and a relative (ε, δ) -approximation is an (ε, δ) -sample for (X, \mathcal{S}) . Using this notion of approximations, the main result of [Li et al. \(2001\)](#) can be formulated as follows.

Theorem 52 ([Li et al. \(2001\)](#)). *There exists an absolute constant $c \geq 1$ such that the following holds. Let (X, \mathcal{S}) be a set system such that $|\mathcal{S}|_Y \leq (e|Y|/d)^d$ for all $Y \subseteq X$ with $|Y| \geq d$, and let $0 < \delta, \varepsilon, \gamma < 1$ be given parameters. Then for any integer t at least*

$$c \cdot \left(\frac{d}{\varepsilon \delta^2} \log \frac{1}{\varepsilon} + \frac{1}{\varepsilon \delta^2} \log \frac{1}{\gamma} \right)$$

a uniform random sample $A \subseteq X$ of size t is a relative (ε, δ) -approximation for (X, \mathcal{S}) with probability at least $1 - \gamma$.

Remarks.

1. Note that by the Sauer-Shelah lemma (Lemma 6), $d_{VC}(X, \mathcal{S}) \leq d$ implies that $|\mathcal{S}|_Y \leq (e|Y|/d)^d$ for any $Y \subseteq X$ (see e.g. Matoušek, 2002, Lemma 10.2.5). Moreover, as a relative $(1/2, \delta)$ -approximation is a δ -approximation, Theorem 52 implies Theorem 4.
2. Li et al. (2001) showed that this bound is asymptotically tight.

The original proof of Theorem 52 uses two probabilistic techniques:

Symmetrization. To prove that a random sample A satisfies the required properties, one takes another random sample G , sometimes called a ‘ghost sample’. Properties of A are then proven by comparing it with G . Note that G is not used in the algorithm or its construction—it is solely a method of analysis, a ‘thought experiment’ of sorts.

Chaining. The idea is to analyze the interaction of the sets in \mathcal{S} with a random sample by partitioning each $S \in \mathcal{S}$ into a logarithmic number of smaller sets, each belonging to a distinct ‘level’. The number of sets increases with increasing level while the size of each set decreases. The overall sum turns out to be a geometric series, which then gives the optimal bounds (Kolmogorov and Tikhomirov, 1959; Talagrand, 2016).

What makes the proof of Theorem 52 in Li et al. (2001) difficult is that it combines chaining and symmetrization intricately. All the tail bounds are stated in their ‘symmetrized’ forms and symmetrization is carried through the entire proof. It is not an easy proof to explain to undergraduate or even graduate students in computer science, as it is difficult to see what is really going on in terms of the significance and intuition of these two ideas. In fact, even the proofs of simpler statements involving just symmetrization, as given in textbooks¹—e.g., see Kearns and Vazirani (1994); Devroye et al. (1996); Matoušek (1999); Chazelle (2000); Matoušek (2002); Anthony and Bartlett (2009); Har-Peled (2011); Alon and Spencer (2016)—often come with the caveat that the idea is ingenious but difficult to understand intuitively (e.g., “one might be tempted to believe that it works by some magic” (Matoušek, 2002, Section 10.2)).

¹Also used in teaching; to pick two arbitrary examples, see [here](#) for an example from the perspective of statistics/learning and [here](#) from the algorithmic side.

In this chapter, we give a modular, self-contained, intuitive proof of Theorem 52 for finite set systems. The only ingredient we assume is the standard Chernoff's concentration bound.

Just as we did in the Preliminaries (Chapter 2) for ε -approximations, we start the study of relative approximations with a basic guarantee which follows from Chernoff's bound.

Theorem 53. *Let X be a set of n elements and A be a uniform random sample of X of size t . Then for any $S \subseteq X$ and $\eta > 0$,*

$$\mathbb{P} \left[|A \cap S| \notin \left(\frac{|S|t}{n} - \eta, \frac{|S|t}{n} + \eta \right) \right] \leq 2 \exp \left(-\frac{\eta^2 n}{2|S|t + \eta n} \right).$$

In particular, setting $\eta = \delta t \max \left\{ \frac{|S|}{n}, \varepsilon \right\}$, a uniform random sample A of size t fails to be a relative (ε, δ) -approximation for a fixed $S \in \mathcal{S}$ with probability at most $2 \exp \left(-\frac{\varepsilon \delta^2 t}{3} \right)$.

Theorem 53 in conjunction with the union bound gives the following upper-bound on relative (ε, δ) -approximation sizes for *any* finite set system.

Theorem 54. *Let (X, \mathcal{S}) be a finite set system and $0 < \varepsilon, \delta, \gamma < 1$ be given parameters. Then for any integer $t \geq \frac{3}{\varepsilon \delta^2} \ln \frac{2|\mathcal{S}|}{\gamma}$, a uniform random sample $A \subseteq X$ of size t is a relative (ε, δ) -approximation for \mathcal{S} with probability at least $1 - \gamma$.*

Proof. By Theorem 53, a uniform random sample A of size t fails to be a relative (ε, δ) -approximation for a fixed $S \in \mathcal{S}$ with probability at most $2 \exp \left(-\frac{\varepsilon \delta^2 t}{3} \right)$. By the union bound,

$$\begin{aligned} & \mathbb{P} \left[\exists S \in \mathcal{S} \text{ s.t. } |A \cap S| \notin \left(\frac{|S|t}{n} - \delta t \max \left\{ \frac{|S|}{n}, \varepsilon \right\}, \frac{|S|t}{n} + \delta t \max \left\{ \frac{|S|}{n}, \varepsilon \right\} \right) \right] \\ & \leq |\mathcal{S}| \cdot 2 \exp \left(-\frac{\varepsilon \delta^2 t}{3} \right) \leq \gamma. \end{aligned}$$

Therefore, with probability at least $1 - \gamma$, A is a relative (ε, δ) -approximation for any set $S \in \mathcal{S}$. \square

We show that in fact one can separate the roles of chaining and symmetrization, giving two separate statements which together immediately imply Theorem 52.

The role of symmetrization is to get a bound on relative (ε, δ) -approximations that is independent of $|\mathcal{S}|$:

Theorem 55. *There exists an absolute constant c_1 such that the following holds. Let (X, \mathcal{S}) be a set system such that $|\mathcal{S}|_Y \leq (e|Y|/d)^d$ for all $Y \subseteq X$, $|Y| \geq d$, and let $0 < \delta, \varepsilon, \gamma < 1/2$ be given parameters. Then for any integer t at least*

$$\frac{c_1}{\varepsilon \delta^2} \cdot \left(d \ln \frac{1}{\varepsilon \delta} + \ln \frac{1}{\gamma} \right),$$

a uniform random sample $A \subseteq X$ of size t is a relative (ε, δ) -approximation for (X, \mathcal{S}) with probability at least $1 - \gamma$.

Remark. *The proof of Theorem 55 is standard using symmetrization. In Section 7.2, we present a different proof, which in fact shows that symmetrization is not really necessary for finite set systems² and can be replaced by a more intuitive argument that makes it obvious, pedagogically, why the bound is independent of $|\mathcal{S}|$.*

On the other hand, the role of chaining is to get rid of logarithmic factors that arise when applying union bound, by more carefully analyzing the failure probability for a collection of events. When separated from symmetrization, chaining provides an upper bound which improves logarithmic factors but depends on $|\mathcal{S}|$:

Theorem 56. *There exists an absolute constant c_2 such that the following holds. Let (X, \mathcal{S}) be a set system such that $|\mathcal{S}|_Y \leq (e|Y|/d)^d$ for all $Y \subseteq X$, $|Y| \geq d$, and let $0 < \delta, \varepsilon, \gamma < 1/2$ be given parameters. Then for any integer t at least*

$$c_2 \max \left\{ \frac{1}{\varepsilon \delta} \ln \frac{|\mathcal{S}|}{\gamma}, \frac{1}{\varepsilon \delta^2} \ln \left(\frac{1}{\varepsilon^d \gamma} \right) \right\},$$

a uniform random sample $A \subseteq X$ of size t is a relative (ε, δ) -approximation for (X, \mathcal{S}) with probability at least $1 - \gamma$.

Remark. *The proof of Theorem 56 is given in Section 7.1.*

The above two statements imply Theorem 52: given (X, \mathcal{S}) , Theorem 55 guarantees that a uniform sample $A_1 \subseteq X$ of size

$$O \left(\frac{1}{\varepsilon \delta^2} \ln \frac{1}{\varepsilon^d \delta^d \gamma} \right)$$

²This is typically the case in its use in algorithms, computational geometry, combinatorics. The infinite case can usually be reduced to the finite case by a sufficiently fine grid, see [Matoušek et al. \(1991\)](#).

is a relative $(\varepsilon, \frac{\delta}{3})$ -approximation with probability at least $1 - \gamma/2$. Now apply Theorem 56 to the set system $(A_1, \mathcal{S}|_{A_1})$ to get $A_2 \subseteq A_1$ of size

$$O \left(\max \left\{ \frac{1}{\varepsilon \delta} \ln \frac{\left(\frac{e}{d \varepsilon \delta^2} \ln \frac{1}{\varepsilon^d \delta^d \gamma} \right)^d}{\gamma}, \frac{1}{\varepsilon \delta^2} \ln \left(\frac{1}{\varepsilon^d \gamma} \right) \right\} \right) = O \left(\frac{1}{\varepsilon \delta^2} \cdot \left(d \ln \frac{1}{\varepsilon} + \ln \frac{1}{\gamma} \right) \right).$$

which is a relative $(\varepsilon, \frac{\delta}{3})$ -approximation of $\mathcal{S}|_{A_1}$ with probability at least $1 - \gamma/2$. We claim that A_2 is a relative (ε, δ) -approximation of (X, \mathcal{S}) . Indeed, with probability at least $1 - \gamma$,

$$\begin{aligned} \left| \frac{|S|}{n} - \frac{|A_2 \cap S|}{|A_2|} \right| &= \left| \frac{|S|}{n} - \frac{|A_1 \cap S|}{|A_1|} + \frac{|A_1 \cap S|}{|A_1|} - \frac{|A_2 \cap S|}{|A_2|} \right| \\ &\leq \left| \frac{|S|}{n} - \frac{|A_1 \cap S|}{|A_1|} \right| + \left| \frac{|A_1 \cap S|}{|A_1|} - \frac{|A_2 \cap S|}{|A_2|} \right| \\ &\leq \frac{\delta}{3} \cdot \max \left\{ \frac{|S|}{n}, \varepsilon \right\} + \frac{\delta}{3} \cdot \max \left\{ \frac{|A_1 \cap S|}{|A_1|}, \varepsilon \right\} \\ &\leq \frac{\delta}{3} \cdot \max \left\{ \frac{|S|}{n}, \varepsilon \right\} + \frac{\delta}{3} \cdot \max \left\{ \frac{|S|}{n} + \frac{\delta}{3} \max \left\{ \frac{|S|}{n}, \varepsilon \right\}, \varepsilon \right\} \\ &\leq \frac{\delta}{3} \cdot \max \left\{ \frac{|S|}{n}, \varepsilon \right\} + \frac{\delta}{3} \cdot \max \left\{ \frac{|S|}{n}, \varepsilon \right\} + \frac{\delta^2}{9} \cdot \max \left\{ \frac{|S|}{n}, \varepsilon \right\} \\ &\leq \delta \cdot \max \left\{ \frac{|S|}{n}, \varepsilon \right\} \end{aligned}$$

Thus A_2 is a relative (ε, δ) -approximation of (X, \mathcal{S}) of the required size.

7.1 Chaining

In this section, we prove Theorem 56 using the probabilistic method of chaining. The key observation is that Theorem 53 provides a bound on the probability of failure for a set $S \in \mathcal{S}$ which *decreases* as the size of S decreases. We take advantage of this by partitioning each $S \in \mathcal{S}$ into a logarithmic number of smaller sets, each belonging to a distinct level, such that the levels strike a proper balance—the number of sets (arising from partitioning every $S \in \mathcal{S}$) increase each level, but their size across levels decreases geometrically. This way one gets an improved bound by applying the union bound separately to sets of different levels. To define suitable sets and levels, we use the following consequence of Theorem 55.

Lemma 57. *There is an absolute constant c_3 such that the following holds. Let $\alpha \geq 2$ and let $\mathcal{P} \subseteq \mathcal{S}$ be an α -packing of \mathcal{S} ; that is, for any $S, S' \in \mathcal{P}$, the symmetric difference of S and S' , denoted by $\Delta(S, S')$, has size at least α . Then $|\mathcal{P}| \leq \left(\frac{c_3 n}{\alpha} \right)^{2d}$.*

Remark. Better bounds exist (see e.g. Lemma 40 in Chapter 4), however the one stated in Lemma 57 suffices for our needs while keeping our proof self-contained.

Proof of Lemma 57. Let $\mathcal{G} = \{\Delta(S, S') : S, S' \in \mathcal{P}\}$. By Theorem 55, there exists a relative $(\frac{\alpha}{n}, \frac{1}{2})$ -approximation A' for \mathcal{G} of size

$$|A'| = \frac{c_1}{\frac{\alpha}{n} \cdot \frac{1}{4}} \left(d \ln \frac{2n}{\alpha} + \ln \frac{2n}{\alpha} \right) \leq \frac{8c_1 dn}{\alpha} \cdot \ln \frac{2n}{\alpha} \leq \frac{8c_1 dn^2}{\alpha^2},$$

where we set $\gamma = \frac{\alpha}{2n}$ (note that we could set any positive value for γ as we only use the existence of such approximations). Then for any $S, S' \in \mathcal{P}$, we get

$$|\Delta(S, S') \cap A'| \geq \frac{|\Delta(S, S')| |A'|}{n} - \frac{|A'|}{2} \cdot \max \left\{ \frac{|\Delta(S, S')|}{n}, \frac{\alpha}{n} \right\} = \frac{1}{2} \cdot \frac{|\Delta(S, S')| |A'|}{n} > 0.$$

This implies that $A' \cap S \neq A' \cap S'$ for any $S, S' \in \mathcal{P}$, and so we have that $|\mathcal{P}| = |\mathcal{P}|_{A'}$. Finally, we use that $\mathcal{P} \subset \mathcal{S}$ and thus

$$|\mathcal{P}| = |\mathcal{P}|_{A'} \leq |\mathcal{S}|_{A'} \leq \left(\frac{8ec_1 n^2}{\alpha^2} \right)^d \leq \left(\frac{\sqrt{8ec_1} n}{\alpha} \right)^{2d}.$$

Setting $c_3 = \sqrt{8ec_1}$ concludes the proof. \square

Levels and their approximations.

Set $k = \lceil \log \frac{1}{\delta} \rceil$ and for $i \in [0, k]$, let \mathcal{P}_i be a maximal $\frac{\varepsilon n}{2^i}$ -packing of \mathcal{S} and set $\mathcal{P}_{k+1} = \mathcal{S}$. For any $S \in \mathcal{P}_{i+1} \setminus \mathcal{P}_i$ there exists a set $F_S \in \mathcal{P}_i$ such that $|\Delta(S, F_S)| < \frac{\varepsilon n}{2^i}$. Define

$$\mathcal{A}_i = \{S \setminus F_S : S \in \mathcal{P}_{i+1} \setminus \mathcal{P}_i\} \quad \text{and} \quad \mathcal{B}_i = \{F_S \setminus S : S \in \mathcal{P}_{i+1} \setminus \mathcal{P}_i\}.$$

The members of $\mathcal{A}_i \cup \mathcal{B}_i$ are the sets of level i and the ranges in \mathcal{P}_0 are the ‘base level’ sets. Lemma 57 implies that

$$|\mathcal{A}_i|, |\mathcal{B}_i| \leq |\mathcal{P}_{i+1}| \leq \left(\frac{c_3 \cdot 2^i}{\varepsilon} \right)^{2d}.$$

The next step is to show that each level is well-approximated by A .

Claim 6. Let $\varepsilon_i = \sqrt{(i+1)/2^i} \varepsilon$. With probability $1 - \gamma$, A is simultaneously

- (i) a relative (ε, δ) -approximation for $\mathcal{A}_k \cup \mathcal{B}_k$, and
- (ii) a relative (ε_i, δ) -approximation for $\mathcal{A}_i \cup \mathcal{B}_i$ for all $i \in [0, k-1]$, and
- (iii) a relative (ε, δ) -approximation for \mathcal{P}_0 .

Proof.

- (i) Each set in $\mathcal{A}_k \cup \mathcal{B}_k$ has size less than $\frac{\varepsilon n}{2^k} \leq \varepsilon n \delta \leq \varepsilon n$. Therefore, we apply Theorem 53 with $\eta = \delta t \varepsilon$ and take the union bound over $|\mathcal{A}_k \cup \mathcal{B}_k| \leq 2|\mathcal{S}|$ sets which gives that for a large-enough value of c_2 , A fails to be an (ε, δ) -approximation for $\mathcal{A}_k \cup \mathcal{B}_k$ with probability at most

$$2|\mathcal{S}| \cdot 2 \exp \left(-\frac{\delta^2 t^2 \varepsilon^2 \cdot n}{2\varepsilon n \delta \cdot t + \delta t \varepsilon \cdot n} \right) = 2|\mathcal{S}| \cdot 2 \exp \left(-\frac{\delta \varepsilon t}{3} \right) \leq \frac{\gamma}{3}.$$

- (ii) For a fixed $S \in \mathcal{A}_i \cup \mathcal{B}_i$, we have $|S| \leq \frac{\varepsilon n}{2^i} \leq \varepsilon_i n$. Thus, applying Theorem 53 with $\eta = \delta t \varepsilon_i$ implies that the probability of failure for a fixed set $S \in \mathcal{A}_i \cup \mathcal{B}_i$ is at most

$$2 \exp \left(-\frac{\delta^2 t^2 \varepsilon_i^2 n}{2|S|t + \delta \varepsilon_i t n} \right) \leq 2 \exp \left(-\frac{\delta^2 t \varepsilon^2 (i+1)/2^i}{2\varepsilon/2^i + \delta \varepsilon \sqrt{(i+1)/2^i}} \right) \leq 2 \exp \left(-\frac{\varepsilon \delta^2 t (i+1)}{4} \right).$$

Hence, by the union bound, the overall probability of failure is at most

$$\begin{aligned} \sum_{i=0}^{k-1} |\mathcal{A}_i \cup \mathcal{B}_i| \cdot 2 \exp \left(-\frac{\varepsilon \delta^2 t (i+1)}{4} \right) &\leq \sum_{i=0}^{k-1} 2 \left(\frac{c_3 \cdot 2^i}{\varepsilon} \right)^{2d} 2 (\varepsilon^d \gamma)^{c_2(i+1)/4} \\ &\leq \gamma \sum_{i=0}^{k-1} \frac{4 (c_3 \cdot 2^{i-1})^{2d}}{2^{(d+1)c_2(i+1)/4}} \leq \gamma \sum_{i=1}^{\infty} \frac{1}{5^i} \leq \frac{\gamma}{3} \end{aligned}$$

$$\text{for } c_2 = 8 \log_2 c_3 + 18 \geq 8 \left(\log_2 c_3 + \frac{\log_2(5)}{2d+2} + 1 \right).$$

- (iii) Since $|\mathcal{P}_0| \leq \left(\frac{c_3}{\varepsilon} \right)^{2d}$, Theorem 54 implies that this failure probability is at most $\frac{\gamma}{3}$ if

$$t \geq \frac{3}{\varepsilon \delta^2} \ln \frac{2(c_3/\varepsilon)^{2d}}{\gamma/3}.$$

□

Observe that for any set $S \in \mathcal{S}$, there exists a set $S_k \in \mathcal{P}_k$, with $A_k = S \setminus S_k \in \mathcal{A}_k$ and $B_k = S_k \setminus S \in \mathcal{B}_k$, such that $S = (S_k \setminus B_k) \cup A_k$. Similarly, one can express S_k in terms of $S_{k-1} \in \mathcal{P}_{k-1}$, $A_{k-1} \in \mathcal{A}_{k-1}$, $B_{k-1} \in \mathcal{B}_{k-1}$ and so on until we reach $S_0 \in \mathcal{P}_0$.

Thus Claim 6 implies that with probability at least $1 - \gamma$,

$$\begin{aligned} \left| \frac{|S|}{n} - \frac{|A \cap S|}{t} \right| &= \left| \frac{|S_k|}{n} - \frac{|B_k|}{n} + \frac{|A_k|}{n} - \left(\frac{|A \cap S_k|}{t} - \frac{|A \cap B_k|}{t} + \frac{|A \cap A_k|}{t} \right) \right| \\ &\stackrel{(i)}{\leq} \left| \frac{|S_k|}{n} - \frac{|A \cap S_k|}{t} \right| + \delta \max \left\{ \varepsilon, \frac{|A_k|}{n} \right\} + \delta \max \left\{ \varepsilon, \frac{|B_k|}{n} \right\} \end{aligned}$$

$$\begin{aligned}
&= \left| \frac{|S_k|}{n} - \frac{|A \cap S_k|}{t} \right| + 2\delta\varepsilon \leq \dots \\
&\stackrel{(ii)}{\leq} \left| \frac{|S_0|}{n} - \frac{|A \cap S_0|}{t} \right| + 2\delta \sum_{j=0}^{k-1} \varepsilon_j + 2\delta\varepsilon \\
&\stackrel{(iii)}{\leq} \delta \max \left\{ \varepsilon, \frac{|S_0|}{n} \right\} + 14\delta\varepsilon \leq \delta \frac{|S|}{n} + 16\delta\varepsilon \leq 2\delta \max \left\{ \frac{|S|}{n}, 16\varepsilon \right\},
\end{aligned}$$

where the second-last step uses the fact that $|S_0| \leq |S| + \sum_{j=0}^k |B_j| \leq |S| + \sum_{j=0}^{\infty} \frac{\varepsilon n}{2^j} \leq |S| + 2\varepsilon n$.

Therefore, A is a relative $(16\varepsilon, 2\delta)$ -approximation of \mathcal{S} with probability at least $1 - \gamma$. Repeating the same arguments with $\delta' = \delta/2$ and $\varepsilon' = \varepsilon/16$, we get a relative (ε, δ) -approximation of \mathcal{S} , as required. \square

7.2 Iterative sampling

Now we prove Theorem 55. The proof uses an argument similar to the discrepancy-based halving method for ε -approximations (see Section 2.2.1). However, it is somewhat simpler as it does not need discrepancy, and it applies to the more general notion of a relative (ε, δ) -approximation.

To see the intuition, observe that since $|\mathcal{S}| \leq (e|X|/d)^d$, the bound of Theorem 54 depends only on $|X|$ —in particular that a random sample $A_1 \subseteq X$ of size

$$O\left(\frac{1}{\varepsilon\delta^2} \ln |X|^d\right) = O\left(\frac{d}{\varepsilon\delta^2} \ln |X|\right)$$

is a relative (ε, δ) -approximation. The size of A_1 is much smaller than that of X , therefore applying Theorem 54 again to $\mathcal{S}|_{A_1}$ gives a relative (ε, δ) -approximation $A_2 \subseteq A_1$ for $\mathcal{S}|_{A_1}$, with

$$|A_2| = O\left(\frac{1}{\varepsilon\delta^2} \ln |A_1|^d\right) = O\left(\frac{d}{\varepsilon\delta^2} \ln \left(\frac{d}{\varepsilon\delta^2} \ln |X|\right)\right) = O\left(\frac{d}{\varepsilon\delta^2} \ln \frac{d}{\varepsilon\delta} + \frac{d}{\varepsilon\delta^2} \ln \ln |X|\right).$$

The size of A_2 is again much smaller than that of A_1 . Furthermore, it follows immediately from the definition of relative (ε, δ) -approximations that A_2 is a relative $(\varepsilon, 3\delta)$ -approximation for \mathcal{S} . With each successive application of Theorem 54, the size of the set decreases rapidly, while the error of approximation increases only linearly, giving the required bound that is independent of $|\mathcal{S}|$.

Now we turn to the formal proof of Theorem 55. Let $T(\varepsilon, \delta, \gamma)$ be the smallest integer such that a uniform random sample of size at least $T(\varepsilon, \delta, \gamma)$ from X is a relative (ε, δ) -approximation for \mathcal{S} with probability at least $1 - \gamma$. Note that $T(\varepsilon, \delta, \gamma) \leq |X|$. Further

define $\delta_0 = 0$ and $\delta_i = \frac{3^{i-1}}{\sqrt{|X|}}$ for $i = 1, \dots, \left\lceil \frac{1}{2} \log_3(\sqrt{|X|}) \right\rceil + 1$. We prove that for all i , for all $\varepsilon, \gamma \in (0, 1/2)$ and for all $\delta \in (\delta_{i-1}, \delta_i]$, it holds that $T(\varepsilon, \delta, \gamma) \leq \frac{c_1}{\varepsilon \delta^2} \cdot \left(d \ln \frac{1}{\varepsilon \delta} + \ln \frac{1}{\gamma} \right)$, which is equivalent to the desired statement. The proof is by induction on i .

Base case ($i = 1$): When $\delta \in (0, \delta_1]$, we have $|X| \leq \frac{1}{\delta^2}$ and thus $T(\varepsilon, \delta, \gamma)$ is upper-bounded as required for any $\varepsilon, \gamma \in (0, 1/2)$.

Inductive hypothesis ($j \leq i$): Assume that the statement holds for all $j \leq i$, that is, for any $\delta \in (0, \delta_i]$ and $\varepsilon, \gamma \in (0, 1/2)$, we have $T(\varepsilon, \delta, \gamma) \leq \frac{c_1}{\varepsilon \delta^2} \cdot \left(d \ln \frac{1}{\varepsilon \delta} + \ln \frac{1}{\gamma} \right)$.

Inductive step ($i \rightarrow i+1$): Let $\delta \in (\delta_i, \delta_{i+1}]$. Since $\frac{\delta}{3} \in (0, \delta_i]$, the inductive hypothesis gives that a random sample $A' \subseteq X$ of size $T(\varepsilon, \frac{\delta}{3}, \frac{\gamma}{2}) \leq \frac{9c_1}{\varepsilon \delta^2} \cdot \left(d \ln \frac{3}{\varepsilon \delta} + \ln \frac{2}{\gamma} \right)$, is a relative $(\varepsilon, \frac{\delta}{3})$ -approximation for \mathcal{S} with probability at least $1 - \frac{\gamma}{2}$. By Theorem 54, a uniform random sample A of A' of size

$$\frac{3}{\varepsilon(\delta/3)^2} \ln \frac{2|\mathcal{S}|_{A'}}{(\gamma/2)}$$

is a relative $(\varepsilon, \frac{\delta}{3})$ -approximation for $\mathcal{S}|_{A'}$ with probability $1 - \frac{\gamma}{2}$. Thus A is a uniform random sample of X that is a relative (ε, δ) -approximation for \mathcal{S} with probability at least $1 - \gamma$, implying the recurrence

$$T(\varepsilon, \delta, \gamma) \leq |A| = \frac{3}{\varepsilon(\delta/3)^2} \ln \frac{2|\mathcal{F}|_{A'}}{(\gamma/2)} \leq \frac{27}{\varepsilon \delta^2} \ln \left(\frac{4}{\gamma} \left(\frac{e T(\varepsilon, \frac{\delta}{3}, \frac{\gamma}{2})}{d} \right)^d \right).$$

The required bound on $T(\varepsilon, \delta, \gamma)$ now follows by the inductive hypothesis. As

$$\left(1 + \frac{1}{d} \ln \frac{2}{\gamma} \right)^d \leq \frac{2}{\gamma},$$

we have

$$\begin{aligned} \frac{27}{\varepsilon \delta^2} \ln \left(\frac{4}{\gamma} \left(\frac{e \frac{9c_1}{\varepsilon \delta^2} \left(d \ln \frac{3}{\varepsilon \delta} + \ln \frac{2}{\gamma} \right)}{d} \right)^d \right) &\leq \frac{27}{\varepsilon \delta^2} \ln \left(\frac{4}{\gamma} \left(\frac{e 27 c_1}{\varepsilon^2 \delta^3} \right)^d \left(1 + \frac{1}{d} \ln \frac{2}{\gamma} \right)^d \right) \\ &\leq \frac{c_1}{\varepsilon \delta^2} \ln \frac{1}{(\varepsilon \delta)^d \gamma}, \end{aligned}$$

for any constant $c_1 \geq 318$, which concludes the proof of Theorem 55. \square

7.3 Sensitive approximations

We say that A is a *sensitive ε -approximation* of (X, \mathcal{S}) if for all $Y \in \mathcal{S}$,

$$\left| \frac{|Y|}{|X|} - \frac{|Y \cap A|}{|A|} \right| \leq \frac{\varepsilon}{2} \left(\sqrt{\frac{|Y|}{|X|}} + \varepsilon \right).$$

In particular, a sensitive ε -approximation is simultaneously an ε^2 -net, an ε -approximation, and a sensitive $(\delta\sqrt{\varepsilon})$ -approximation is a relative (ε, δ) -approximation (Har-Peled, 2011). The proof of Theorem 55 can be adapted to a stronger notion of approximations resulting in the following theorem.

Theorem 58. *There exists an absolute constant c_4 such that the following holds. Given a set system (X, \mathcal{S}) such that $|\mathcal{S}|_Y \leq (e|Y|/d)^d$ for all $Y \subseteq X$, $|Y| \geq d$, and parameters $0 \leq \varepsilon, \gamma \leq 1$, a uniform random sample $A \subseteq X$ of size*

$$\frac{c_4}{\varepsilon^2} \cdot \left(d \ln \frac{1}{\varepsilon} + \ln \frac{1}{\gamma} \right)$$

is a sensitive ε -approximation for \mathcal{S} with probability at least $1 - \gamma$.

Proof. Again, we prove the statement by induction on ε and let $T(\varepsilon, \gamma)$ be the smallest integer such that a uniform random sample of size $T(\varepsilon, \gamma)$ from X is a sensitive ε -approximation for \mathcal{S} . When $\varepsilon \leq 1/\sqrt{|X|}$, we have $T(\varepsilon, \gamma) \leq \frac{1}{\varepsilon^2} \cdot \left(d \ln \frac{1}{\varepsilon} + \ln \frac{1}{\gamma} \right)$. When $\varepsilon > 1/\sqrt{|X|}$, let $A' \subset X$ be a random sample of size $T(\frac{\varepsilon}{4}, \frac{\gamma}{2})$. By induction, A' is a sensitive $\frac{\varepsilon}{4}$ -approximation of (X, \mathcal{S}) with probability at least $1 - \frac{\gamma}{2}$. One can verify using Theorem 53 that a sample A of

$$\frac{8}{(\varepsilon/4)^2} \ln \frac{2|\mathcal{S}|_{A'}}{\gamma/2}$$

elements of A' is a sensitive $\frac{\varepsilon}{4}$ -approximation of $(A', \mathcal{S}|_{A'})$ with probability at least $1 - \frac{\gamma}{2}$.

Since A is a sensitive $(\frac{\varepsilon}{2} + 2 \cdot \frac{\varepsilon}{4})$ -approximation of (X, \mathcal{S}) (Brönnimann et al., 1993) with probability at least $1 - \gamma$, we get the following recursion on $T(\varepsilon, \gamma)$

$$T(\varepsilon, \gamma) \leq |A| = \frac{3}{(\varepsilon/4)^2} \ln \frac{2|\mathcal{S}|_{A'}}{(\gamma/2)} \leq \frac{48}{\varepsilon^2} \ln \left(\frac{4}{\gamma} \left(\frac{e T(\frac{\varepsilon}{2}, \frac{\gamma}{2})}{d} \right)^d \right) \leq \frac{1}{\varepsilon^2} \ln \left(\frac{1}{\varepsilon^d \gamma} \right)^{c_4},$$

for a large-enough c_4 .

□

Note that this bound is optimal for sensitive ε -approximations and it also implies an optimal bound for ε -nets—therefore, the bound of Theorem 58 cannot be improved by chaining! However, for ε -approximations and relative (ε, δ) -approximations, it provides suboptimal bounds (by Theorem 52).

Chapter 8

Conclusion

In this thesis, we studied approximations of set systems, focusing on two main construction methods:

1. uniform sampling and
2. non-uniform sampling along the edges of a low-crossing matching.

While the first one is simple and efficient, it does not lead to optimal-sized approximations. On the other hand, the second method can be used to improve the uniform sampling guarantee, but it is much more involved algorithmically. The crucial parameter appearing in the approximation guarantees is the (dual) VC-dimension.

8.1 Non-uniform sampling

The main algorithmic contribution of the thesis is an improved algorithm for constructing matchings with low crossing numbers, which is the bottleneck algorithmic step in constructing ε -approximations of sub-quadratic size. The key property that implies the correctness of the algorithm is a ‘hereditary type’ condition on the the existence of low crossing matchings:

Assumption (MAINASSUMPTION(a, b, γ)). *(X, \mathcal{S}) is a finite set system with $m \geq n$, $m \geq 34$, and any $Y \subseteq X$ has a matching with crossing number at most $a|Y|^\gamma + b$ with respect to \mathcal{S} .*

In particular, if (X, \mathcal{S}) is a set system with dual VC-dimension d , our method constructs a matching of X with crossing number $\tilde{O}(|X|^{1-1/d})$ in time $O(|X|^{1/d}|\mathcal{S}|)$, improving upon the previous-best construction of time $\tilde{O}(|X|^2|\mathcal{S}|)$.

I note that our method can easily be adapted¹ to construct spanning trees, or more specifically, spanning paths with the same asymptotic guarantees:

Theorem 59. *Let (X, \mathcal{S}) be a set system, $n = |X|$, $m = |\mathcal{S}|$ with $m \geq n$, which satisfies that any $Y \subseteq X$ has a spanning tree (resp. spanning path) with crossing number at most $a|Y|^\gamma + b$ with respect to \mathcal{S} . Then there is a randomized algorithm that returns a spanning tree (resp. spanning path) with expected crossing number*

$$O\left(\frac{a}{\gamma}n^\gamma + b \log n + \ln(mn) \log n\right)$$

with an expected number of Oracle calls

$$O\left(\min\left\{\frac{n^{3-\gamma} \ln n}{a} + \frac{mn^{1-\gamma} \ln mn}{a} \cdot \min\left\{\frac{1}{1-\gamma}, \log n\right\}, n^3 + mn\right\}\right).$$

8.1.1 Applicability - hereditary condition for different graphs

It is intriguing that we do not have a good understanding on the monotonicity of crossing numbers of optimal matchings. Even the simplest question appears to be open: if X has a perfect matching with crossing number κ with respect to \mathcal{S} , does any $Y \subseteq X$ has a perfect matching² with crossing number at most κ with respect to $\mathcal{S}|_Y$? The known monotonicity properties of optimal crossing numbers of spanning trees, paths, and matchings and the followings:

Spanning paths. If a point-set X has a spanning path of crossing number κ , then any $Y \subset X$ has a spanning path with crossing number at most κ —we can simply short-cut the original path. Thus any set system satisfies the conditions of Theorem 59 with $a = 1$, $b = \kappa$, and $\gamma = 0$, and so our algorithm provides a spanning path with crossing number $O(\kappa \log n)$.

Spanning trees. In this case, the problem is already less trivial. On one hand, it was shown that the crossing number is not monotone (Mulzer and Obenaus, 2020). On the other hand, the crossing number cannot increase too much: if X has a spanning tree with crossing number κ , then any $Y \subset X$ has a spanning tree (even a spanning path) with crossing number at most 2κ (Har-Peled, 2009).

Matchings. I am unaware of any result on the monotonicity of the minimum crossing number of matchings.

¹We only need to change the way we remove vertices.

²We define the perfect matching of an odd set as a graph which is a matching plus one isolated vertex.

To connect the above problems a bit better, it can instructive to state the relations of the optimal crossing numbers of trees, paths, and matchings denoted as $\text{OPT}_{\text{tree}}(X, \mathcal{S})$, $\text{OPT}_{\text{path}}(X, \mathcal{S})$, and $\text{OPT}_{\text{matching}}(X, \mathcal{S})$.

The optimums over trees and paths are asymptotically equivalent, in particular,

$$\text{OPT}_{\text{tree}}(X, \mathcal{S}) \leq \text{OPT}_{\text{path}}(X, \mathcal{S}) \leq 2 \cdot \text{OPT}_{\text{tree}}(X, \mathcal{S}),$$

where the first inequality is trivial and the proof of the second inequality can be found in [Pach and Agarwal \(2011\)](#).

As for matchings, we trivially have $\text{OPT}_{\text{matching}}(X, \mathcal{S}) \leq \text{OPT}_{\text{path}}(X, \mathcal{S})$. It seems likely that $\text{OPT}_{\text{matching}}(X, \mathcal{S})$ cannot be much smaller than $\text{OPT}_{\text{path}}(X, \mathcal{S})$, but I am unaware of such bound. For instance, the monotonicity of matchings would imply that and that $\text{OPT}_{\text{path}}(X, \mathcal{S}) \leq \log n \cdot \text{OPT}_{\text{matching}}(X, \mathcal{S})$.

8.1.2 Relation to Primal-Dual reweighing algorithms

Our algorithm fits in the line of several problems where ‘primal-dual reweighing’ could replace ‘exponential reweighing along an extremum’. These examples appear in various fields, such as matrix games ([Grigoriadis and Khachiyan, 1995](#); [Freund and Schapire, 1999](#)), approximate LP solvers ([Plotkin et al., 1995](#)), or geometric algorithms ([Agarwal and Pan, 2014](#)); for more examples see the survey of [Arora et al. \(2012\)](#).

I tried to apply some tools developed for primal-dual reweighing algorithms, e.g., non-uniform increments, to speed up our algorithm, but they did not lead to an immediate improvement.

On the other hand, one could try to apply ideas developed in our algorithm to other problems. For instance, I find it plausible that the pre-processing steps of the geometric hitting set algorithm of [Agarwal and Pan \(2014\)](#) could be replaced by random updates during the re-weighting steps, hence simplifying the algorithm.

Another interesting direction is to study whether the primal-dual reweighing approach can lead to improved algorithms for other geometric problems. For instance, one can consider simplicial partitions (a central notion in computational geometry that generalizes matchings with low crossing numbers) or polynomial partitions (another important tool, whose construction is based on the ham-sandwich theorem in high-dimension).

Finally, our algorithm is randomized, so one can think about bounding its variance (which is empirically low, see [Figure 4.4](#)) or de-randomizing—however this is only interesting if one can get below the $O(mn^3)$ running time of the deterministic algorithm of [Chazelle and Welzl \(1989\)](#).

8.1.3 Matchings with low relative crossing number

Another interesting problem is the following:

Given a set system (X, \mathcal{S}) with dual shatter function $\pi^*(k) \leq c \cdot k^d$, is there a perfect matching M of X such that any $S \in \mathcal{S}$ crosses $O(|S|^{1-1/d})$ edges of M ?

Initially, I came to this question while thinking whether a finer probabilistic analysis (in particular, chaining) could remove the logarithmic term in our $O\left(\sqrt{n^{1-1/d} \ln m}\right)$ bound for low discrepancy colorings. Technically, it would require a bound on the failure probability which depends on the range size and, intuitively, such bound can arise from the existence of matchings with low relative crossing number.

Bounded dual VC-dimension. In this general setup, our bound on discrepancy is tight for set systems with bounded dual VC-dimension (Matoušek, 1995). This makes me wonder: could the chaining argument and the discrepancy lower bound be used together to show that it is not possible to have matchings with $O(|S|^{1-1/d})$ relative crossing number for arbitrary set systems with bounded dual VC-dimension?

Half-spaces. In this special case, the discrepancy bound can be improved to $O\left(\sqrt{n^{1-1/d}}\right)$, so the above general argument cannot be applied in this setting. There have also been a dedicated studies of spanning trees with low relative crossing number with respect to half-planes. The notion was introduced by Har-Peled and Sharir (2011), who showed that any set P of n points in \mathbb{R}^2 has a spanning tree such that any set S induced by a half-plane crosses

$$O\left(\sqrt{|S|} \log\left(\frac{n}{|S|}\right)\right)$$

edges of the spanning tree. Later Obenaus (2019) showed an $\omega\left(\sqrt{|S|}\right)$ lower bound for the same problem. I note that these results imply that there are perfect matching with relative crossing number $O\left(\sqrt{|S|} \log\left(\frac{n}{|S|}\right)\right)$, but the lower bound does not translate immediately to matchings.

8.1.4 Algorithmic complexity

It would also be interesting to study whether the problem can be solved with less than $mn^{1/d}$ membership Oracle calls, or the hardness of the discrepancy problem for set systems with polynomially bounded dual shatter function. It is known that for general set systems, it is NP-hard to distinguish between discrepancy 0 and $\Omega(\sqrt{n})$, and similarly, for set

systems with primal shatter function $\pi(k) = O(k^d)$, it is NP-hard to decide whether the discrepancy is 0 or $\Omega(\sqrt{n^{1-1/d}})$ (Charikar et al., 2011).

8.2 VC-dimension

In this thesis, I showed that the set system induced by k -fold union of half-spaces in \mathbb{R}^d have VC dimension $\Omega(dk \log k)$ if $d \geq 4$. This settles the open problem of Blumer et al. (1989), studied by Eisenstat and Angluin (2007); Eisenstat (2009).

Furthermore, I show that asymptotic lower bounds on the VC dimension of k -fold unions of set systems imply asymptotic lower bounds on ε -net sizes (Lemma 51). For instance, the combination of the above two results immediately yields a lower bound of $\Omega(d/\varepsilon \log(1/\varepsilon))$ on the ε -net sizes in set systems induced by half-spaces, recovering the result of Pach and Tardos (2013).

A challenging open problem in this area is to provide asymptotically tight bounds on the VC-dimension of k -fold unions of balls in \mathbb{R}^3 . In dimensions other than 3, the problem is well understood: in \mathbb{R}^2 , the VC-dimension grows as $\Theta(k)$ and in dimensions at least 4, my work presented in Chapter 5 implies a growth of $\Theta(k \log k)$. The importance of this question lies in the fact that (via Lemma 51) an $\Omega(k \log k)$ lower bound on the VC-dimension of k -fold unions of balls would settle a long-standing open problem on whether the ε -net bound of $O(1/\varepsilon \log 1/\varepsilon)$ is tight for balls in \mathbb{R}^3 .

Bibliography

- Agarwal, P. K. (2017). Simplex range searching. In *Journey Through Discrete Mathematics*, pages 1–30. Springer.
- Agarwal, P. K., Matoušek, J., and Sharir, M. (2013). On range searching with semialgebraic sets. II. *SIAM Journal on Computing*, 42(6):2039–2062.
- Agarwal, P. K. and Pan, J. (2014). Near-linear algorithms for geometric hitting sets and set covers. In *Symposium on Computational Geometry*, page 271.
- Alexander, R. (1990). Geometric methods in the study of irregularities of distribution. *Combinatorica*, 10(2):115–136.
- Alon, N., Haussler, D., and Welzl, E. (1987). Partitioning and geometric embedding of range spaces of finite Vapnik-Chervonenkis dimension. In *SoCG '87*.
- Alon, N., Moran, S., and Yehudayoff, A. (2016). Sign rank versus VC dimension. In *COLT*.
- Alon, N., Rónyai, L., and Szabó, T. (1999). Norm-graphs: Variations and applications. *Journal of Combinatorial Theory, Series B*, 76(2):280 – 290.
- Alon, N. and Spencer, J. H. (2016). *The probabilistic method*. John Wiley & Sons.
- Anthony, M. and Bartlett, P. L. (2009). *Neural network learning: Theoretical foundations*. Cambridge University Press.
- Arora, S., Hazan, E., and Kale, S. (2012). The multiplicative weights update method: a meta algorithm and applications. *Theory of Computing [electronic only]*.
- Axenovich, M. and Csikós, M. (2019). Induced saturation of graphs. *Discrete Mathematics*, 342(4):1195–1212.
- Bachem, O. (2018). *Sampling for Large-Scale Clustering*. PhD thesis, ETH Zürich.

- Balcan, M.-F., Ehrlich, S., and Liang, Y. (2013). Distributed k -means and k -median clustering on general topologies. In *Proceedings of Neural Information Processing Systems (NIPS)*.
- Bansal, N. (2010). Constructive algorithms for discrepancy minimization. In *Proceedings of Symposium on Foundations of Computer Science, FOCS*, pages 3–10. IEEE Computer Society.
- Bansal, N., Dadush, D., Garg, S., and Lovett, S. (2018). The Gram-Schmidt walk: a cure for the Banaszczyk blues. In *Proceedings of the Symposium on Theory of Computing, STOC*, pages 587–597.
- Bartlett, P. L., Boucheron, S., and Lugosi, G. (2002). Model selection and error estimation. *Machine Learning*, 48(1):85–113.
- Blumer, A., Ehrenfeucht, A., Haussler, D., and Warmuth, M. K. (1989). Learnability and the Vapnik-Chervonenkis dimension. *J. ACM*, 36(4):929–965.
- Brönnimann, H., Chazelle, B., and Matoušek, J. (1993). Product range spaces, sensitive sampling, and derandomization. *Proc. Symposium on Foundations of Computer Science*, pages 400–409.
- Chan, T. M. (2012). Optimal partition trees. *Discrete Comput. Geom.*, 47(4):661–690.
- Chan, T. M. (2018). Improved deterministic algorithms for linear programming in low dimensions. *ACM Transactions on Algorithms (TALG)*, 14(3):1–10.
- Charikar, M., Newman, A., and Nikolov, A. (2011). Tight hardness results for minimizing discrepancy. In *Proceedings of the twenty-second annual ACM-SIAM symposium on Discrete Algorithms*, pages 1607–1614. SIAM.
- Chazelle, B. (2000). *The Discrepancy Method: Randomness and Complexity*. Cambridge University Press, New York, NY, USA.
- Chazelle, B. and Friedman, J. (1990). A deterministic view of random sampling and its use in geometry. *Combinatorica*, 10(3):229–249.
- Chazelle, B. and Welzl, E. (1989). Quasi-optimal range searching in spaces of finite VC-dimension. *Discrete Comput. Geom.*, page 467–489.
- Chekuri, C. and Quanrud, K. (2018). Randomized MWU for positive LPs. In *Proceedings of ACM-SIAM Symposium on Discrete Algorithms, SODA '18*, page 358–377.

- Chen, W., Srivastav, A., and Travaglini, G. (2014). *A Panorama of Discrepancy Theory*. Springer.
- Clarkson, K. L. (1987). New applications of random sampling in computational geometry. *Discrete & Computational Geometry*, 2(2):195–222.
- Csikós, M. and Mustafa, N. (2022+). Practical algorithm for colorings with optimal discrepancy. *Preprint*.
- Csikós, M. and Mustafa, N. H. (2021). Escaping the Curse of Spatial Partitioning: Matchings with Low Crossing Numbers and Their Applications. In *37th International Symposium on Computational Geometry (SoCG 2021)*, volume 189, pages 28:1–28:17.
- Csikós, M. and Mustafa, N. H. (2022). Optimal approximations made easy. *Information Processing Letters*.
- Csikós, M., Mustafa, N. H., and Kupavskii, A. (2019). Tight lower bounds on the vc-dimension of geometric set systems. *Journal of Machine Learning Research*, 20(81):1–8.
- Devroye, L., Györfi, L., and Lugosi, G. (1996). *A Probabilistic Theory of Pattern Recognition*. Springer, Berlin.
- Dobkin, D. P. and Gunopulos, D. (1995). Concept learning with geometric hypotheses. In *Proceedings of Computational Learning Theory (COLT)*, pages 329–336.
- Ducocfe, G., Habib, M., and Viennot, L. (2020). Diameter computation on h-minor free graphs and graphs of bounded (distance) VC-dimension. In *Proceedings of the ACM-SIAM Symposium on Discrete Algorithms (SODA)*, pages 1905–1922.
- Eisenstat, D. (2009). k -fold unions of low-dimensional concept classes. *Information Processing Letters*, 109(23-24):1232–1234.
- Eisenstat, D. and Angluin, D. (2007). The VC dimension of k -fold union. *Information Processing Letters*, 101(5):181 – 184.
- Ezra, E., Har-Peled, S., Kaplan, H., and Sharir, M. (2020). Decomposing arrangements of hyperplanes: VC-dimension, combinatorial dimension, and point location. *Discret. Comput. Geom.*, 64(1):109–173.
- Fekete, S. P., Lübbecke, M. E., and Meijer, H. (2008). Minimizing the stabbing number of matchings, trees, and triangulations. *Discrete & Computational Geometry*, 40(4):595–621.

- Feldman, D. and Langberg, M. (2011). A unified framework for approximating and clustering data. In *Proceedings of Symposium on the Theory of Computing (STOC)*, pages 569–578.
- Freund, Y. and Schapire, R. E. (1999). Adaptive game playing using multiplicative weights. *Games and Economic Behavior*, 29(1):79 – 103.
- Giannopoulos, P., Konzack, M., and Mulzer, W. (2014). Low-crossing spanning trees: an alternative proof and experiments. In *Proceedings of EuroCG*.
- Grigoriadis, M. D. and Khachiyan, L. G. (1995). A sublinear-time randomized approximation algorithm for matrix games. *Operations Research Letters*, 18(2):53 – 58.
- Har-Peled, S. (2000). Constructing planar cuttings in theory and practice. *SIAM J. Comput.*, 29:2016–2039.
- Har-Peled, S. (2009). Approximating spanning trees with low crossing number. *arXiv*, abs/0907.1131.
- Har-Peled, S. (2011). *Geometric Approximation Algorithms*. American Mathematical Society, Boston, MA, USA.
- Har-Peled, S. and Sharir, M. (2011). Relative (p, ϵ) -Approximations in Geometry. *Discrete & Computational Geometry*, 45(3):462–496.
- Haussler, D. (1995). Sphere Packing Numbers for Subsets of the Boolean n -Cube with Bounded Vapnik-Chervonenkis Dimension. *J. Comb. Theory, Ser. A*, 69(2):217–232.
- Haussler, D. and Welzl, E. (1987). ϵ -nets and simplex range queries. *Discrete & Computational Geometry*, 2:127–151.
- Johnson, H. (2008). *Definable Families of Finite Vapnik Chernonenkis Dimension*. PhD thesis, University of Maryland.
- Kearns, M. J. and Vazirani, U. V. (1994). *An Introduction to Computational Learning Theory*. MIT Press.
- Kolmogorov, A. N. and Tikhomirov, V. M. (1959). ϵ -entropy and ϵ -capacity. *Uspekhi Mat. Nauk*, 14:3–86.
- Komlós, J., Pach, J., and Woeginger, G. (1992). Almost tight bounds for ϵ -nets. *Discrete & Computational Geometry*, 7:163–173.

- Koufogiannakis, C. and Young, N. E. (2014). A nearly linear-time ptas for explicit fractional packing and covering linear programs. *Algorithmica*, 70(4):648–674.
- Kupavskii, A., Mustafa, N. H., and Pach, J. (2016). New lower bounds for epsilon-nets. In *Proceedings of Symposium on Computational Geometry (SoCG)*, pages 54:1–54:16.
- Levy, A., Ramadas, H., and Rothvoss, T. (2017). Deterministic discrepancy minimization via the multiplicative weight update method. In *Integer Programming and Combinatorial Optimization (IPCO)*, pages 380–391.
- Li, Y., Long, P. M., and Srinivasan, A. (2001). Improved Bounds on the Sample Complexity of Learning. *J. Comput. Syst. Sci.*, 62(3):516–527.
- Lovett, S. and Meka, R. (2015). Constructive discrepancy minimization by walking on the edges. *SIAM J. Comput.*, 44(5):1573–1582.
- Lucic, M., Bachem, O., and Krause, A. (2016). Strong Coresets for Hard and Soft Bregman Clustering with Applications to Exponential Family Mixtures. In *Proceedings of Artificial Intelligence and Statistics (AISTATS)*.
- Matheny, M. and Phillips, J. M. (2018). Practical low-dimensional halfspace range space sampling. In *Annual European Symposium on Algorithms (ESA)*, volume 112, pages 62:1–62:14.
- Matoušek, J. (1992). Efficient partition trees. *Discrete & Computational Geometry*, 8(3):315–334.
- Matoušek, J. (1995). Tight upper bounds for the discrepancy of half-spaces. *Discret. Comput. Geom.*, 13:593–601.
- Matoušek, J. (1997). On discrepancy bounds via dual shatter function. *Mathematika*, 44(1):42–49.
- Matoušek, J. (1999). *Geometric Discrepancy: An Illustrated Guide*. Springer Berlin Heidelberg.
- Matoušek, J. (2002). *Lectures in Discrete Geometry*. Springer-Verlag, New York, NY.
- Matoušek, J. (2013). *Lectures on discrete geometry*, volume 212. Springer Science & Business Media.

- Matoušek, J., Welzl, E., and Wernisch, L. (1991). Discrepancy and approximations for bounded VC-dimension. In *Proceedings 32nd Annual Symposium of Foundations of Computer Science*, pages 424–430. IEEE.
- Mulzer, W. and Obenaus, J. (2020). The tree stabbing number is not monotone. *CoRR*, abs/2002.08198.
- Mustafa, N. H. and Varadarajan, K. (2017). Epsilon-approximations and epsilon-nets. In Goodman, J. E., O’Rourke, J., and Tóth, C. D., editors, *Handbook of Discrete and Computational Geometry*. CRC Press LLC.
- Obenaus, J. (2019). *Spanning trees with low (shallow) stabbing number*. PhD thesis, ETH Zurich.
- Pach, J. and Agarwal, P. K. (2011). *Combinatorial geometry*, volume 37. John Wiley & Sons.
- Pach, J. and Tardos, G. (2013). Tight lower bounds for the size of epsilon-nets. *J. Amer. Math. Soc.*, 26(3):645–658.
- Plotkin, S., Shmoys, D., and Tardos, É. (1995). Fast approximation algorithms for fractional packing and covering problems. *Mathematics of Operations Research*, 20(2):257–301.
- Rothvoss, T. (2017). Constructive discrepancy minimization for convex sets. *SIAM J. Comput.*, 46(1):224–234.
- Sauer, N. (1972). On the Density of Families of Sets. *J. Comb. Theory, Ser. A*, 13(1):145–147.
- Shalev-Shwartz, S. and Ben-David, S. (2014). *Understanding machine learning: From theory to algorithms*. Cambridge university press.
- Shelah, S. (1972). A combinatorial problem; stability and order for models and theories in infinitary languages. *Pacific Journal of Mathematics*, 41:247–261.
- Spencer, J. (1985). Six standard deviations suffice. *Transactions of the American Mathematical Society*, 289:679–706.
- Talagrand, M. (2016). *Upper and Lower Bounds for Stochastic Processes: Modern Methods and Classical Problems*. Springer Berlin Heidelberg.

- Turán, P. (1941). On an external problem in graph theory. *Mat. Fiz. Lapok*, 48:436–452.
- Vapnik, V. N. and Chervonenkis, A. Y. (1971). On the uniform convergence of relative frequencies of events to their probabilities. *Theory of Probability and its Applications*, 16(2):264–280.
- Vershynin, R. (2018). *High-dimensional probability: An introduction with applications in data science*, volume 47. Cambridge University Press.
- Warren, H. E. (1968). Lower bounds for approximation by nonlinear manifolds. *Transactions of the American Mathematical Society*, 133(1):167–178.
- Welzl, E. (1988). Partition trees for triangle counting and other range searching problems. In *Proceedings of Annual Symposium on Computational Geometry (SoCG)*, page 23–33.
- Welzl, E. (1992). On spanning trees with low crossing numbers. In *Data Structures and Efficient Algorithms, Final Report on the DFG Special Joint Initiative*, page 233–249.

List of Figures

1.1	Matchings on 5000 points with different input and query types.	7
3.1	Couplages de 5000 points avec différents types d'entrée et d'ensemble. Colonne de gauche: notre méthode avec des ensembles de demi-plans. Colonne du milieu: notre méthode avec des ensembles de disques. Colonne de droite: échantillonnage aléatoire.	33
4.1	Comparing the quality and time of our algorithm with different settings of α on Grid input type.	70
4.2	Matchings created by our method on 5000 points (Grid input type) with different pre-sampling parameters α	71
4.3	Mean and variance of our discrepancy error and runtime on input type Grid . The shaded area denote ± 1 standard deviation.	73
4.4	Mean and variance of our method and random sampling on input type Grid . The shaded areas denote ± 1 standard deviation.	73

List of Tables

4.1	Summary of our results for set systems (X, \mathcal{S}) with $n = X $, $m = \mathcal{S} $, $n \leq m$, and $d \geq 2$. We use the notation $\pi_{\mathcal{S}}^*(\cdot)$ for the dual shatter function of (X, \mathcal{S}) , \mathcal{H}_d for half-spaces in \mathbb{R}^d , \mathcal{B}_d for balls in \mathbb{R}^d , and $\Gamma_{d,\Delta,s}$ for semi-algebraic ranges in \mathbb{R}^d described by at most s equations of degree at most Δ (see Section 4.7).	41
4.2	Testing crossing number of matchings with $n \log n$ and $n \log^d n$ random planes.	67
4.3	Experimental results for set systems induced by half-spaces.	72