Faster Algorithms for Data Approximation

Alexandre Louvet

Laboratoire d'Informatique de Paris Nord

November 8^{th} 2024

Preliminaries

We consider a set system $(X, \mathcal{F})$ with VC-dimension $\leq d$ and denote $|X| = n, |\mathcal{F}| = m$.

Preliminaries

We consider a set system $(X, \mathcal{F})$ with VC-dimension $\leq d$ and denote $|X| = n, |\mathcal{F}| = m$.

The VC-dimension of a set system is defined as the cardinality d of $Y \subseteq X$ s.t.

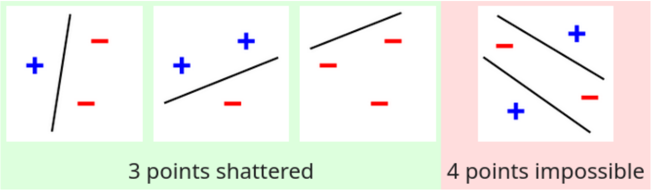
$$Y = \max_{Z \subseteq X} \left| \{F \cap Z | F \in \mathcal{F}\} \right| = 2^d$$

Preliminaries

We consider a set system $(X, \mathcal{F})$ with VC-dimension $\leq d$ and denote $|X| = n, |\mathcal{F}| = m$.

The VC-dimension of a set system is defined as the cardinality d of $Y \subseteq X$ s.t.

$$Y = \max_{Z \subseteq X} \left| \{F \cap Z | F \in \mathcal{F}\} \right| = 2^d$$



https://en.wikipedia.org/wiki/Vapnik%E2%80%93Chervonenkis_dimension

Combinatorial discrepancy

Combinatorial Discrepancy

We want to compute a 2-coloring $\chi : X \to \{-1, 1\}$ s.t.

$$\forall F \in \mathcal{F}, \chi(F) = \sum_{x \in F} \chi(x) \text{ is small.}$$

We denote:

$$\text{disc}(X, \mathcal{F}) = \min_{\chi} \max_{F \in \mathcal{F}} \chi(F)$$

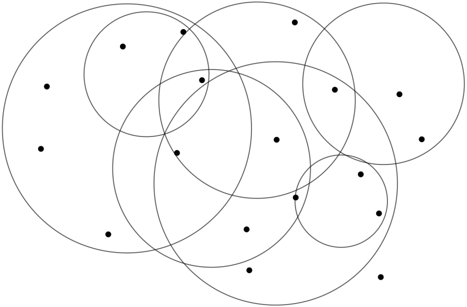
Combinatorial Discrepancy

We want to compute a 2-coloring $\chi : X \to \{-1, 1\}$ s.t.

$$\forall F \in \mathcal{F}, \chi(F) = \sum_{x \in F} \chi(x) \text{ is small.}$$

We denote:

$$\text{disc}(X, \mathcal{F}) = \min_{\chi} \max_{F \in \mathcal{F}} \chi(F)$$



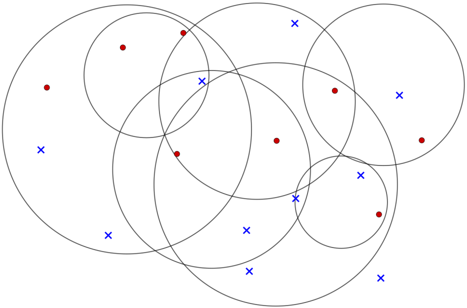
Combinatorial Discrepancy

We want to compute a 2-coloring $\chi : X \to \{-1, 1\}$ s.t.

$$\forall F \in \mathcal{F}, \chi(F) = \sum_{x \in F} \chi(x) \text{ is small.}$$

We denote:

$$\text{disc}(X, \mathcal{F}) = \min_{\chi} \max_{F \in \mathcal{F}} \chi(F)$$



Combinatorial Discrepancy

We want to compute a 2-coloring $\chi : X \to \{-1, 1\}$ s.t.

$$\forall F \in \mathcal{F}, \chi(F) = \sum_{x \in F} \chi(x) \text{ is small.}$$

We denote:

$$\operatorname{disc}(X, \mathcal{F}) = \min_{\chi} \max_{F \in \mathcal{F}} \chi(F)$$

[Spe85] proved that arbitrary set systems admit a coloring with $\operatorname{disc}(X, \mathcal{F}) = O(\sqrt{n \log(\frac{m}{n})})$.

Combinatorial Discrepancy

We want to compute a 2-coloring $\chi : X \to \{-1, 1\}$ s.t.

$$\forall F \in \mathcal{F}, \chi(F) = \sum_{x \in F} \chi(x) \text{ is small.}$$

We denote:

$$\text{disc}(X, \mathcal{F}) = \min_{\chi} \max_{F \in \mathcal{F}} \chi(F)$$

[Spe85] proved that arbitrary set systems admit a coloring with $\text{disc}(X, \mathcal{F}) = O(\sqrt{n \log(\frac{m}{n})})$.

[Mat95] proved that set systems with VC-dimension $\leq d$ admit a coloring with $\text{disc}(X, \mathcal{F}) = O(n^{1/2-1/2d})$.

Combinatorial Discrepancy

We want to compute a 2-coloring $\chi : X \to \{-1, 1\}$ s.t.

$$\forall F \in \mathcal{F}, \chi(F) = \sum_{x \in F} \chi(x) \text{ is small.}$$

We denote:

$$\text{disc}(X, \mathcal{F}) = \min_{\chi} \max_{F \in \mathcal{F}} \chi(F)$$

[Spe85] proved that arbitrary set systems admit a coloring with $\text{disc}(X, \mathcal{F}) = O(\sqrt{n \log(\frac{m}{n})})$.

[Mat95] proved that set systems with VC-dimension $\leq d$ admit a coloring with $\text{disc}(X, \mathcal{F}) = O(n^{1/2 - 1/2d})$.

Before [Ban10]'s breakthrough, there was no polynomial algorithm constructing these colorings.

A New Discrepancy Game

A New Discrepancy Game

Let Alice and Bob be two players competing in a T rounds game.

A New Discrepancy Game

Let Alice and Bob be two players competing in a T rounds game.

They are both given a set system $(X, \mathcal{F})$ with VC-dimension $\leq d$.

A New Discrepancy Game

Let Alice and Bob be two players competing in a T rounds game.

They are both given a set system $(X, \mathcal{F})$ with VC-dimension $\leq d$.

Each round t is as follows:

A New Discrepancy Game

Let Alice and Bob be two players competing in a T rounds game.

They are both given a set system $(X, \mathcal{F})$ with VC-dimension $\leq d$.

Each round t is as follows:

- Alice picks a coloring $\chi^{(t)}$ of X and sends it to Bob.

A New Discrepancy Game

Let Alice and Bob be two players competing in a T rounds game.

They are both given a set system $(X, \mathcal{F})$ with VC-dimension $\leq d$.

Each round t is as follows:

- Alice picks a coloring $\chi^{(t)}$ of X and sends it to Bob.
- Bob picks a set $F^{(t)} \in \mathcal{F}$ and sends it to Alice.

A New Discrepancy Game

Let Alice and Bob be two players competing in a T rounds game.

They are both given a set system $(X, \mathcal{F})$ with VC-dimension $\leq d$.

Each round t is as follows:

- Alice picks a coloring $\chi^{(t)}$ of X and sends it to Bob.
- Bob picks a set $F^{(t)} \in \mathcal{F}$ and sends it to Alice.

Alice's goal: minimize $\sum_{t=1}^{T} \chi^{(t)}(F^{(t)})$.

Bob's goal: maximize $\sum_{t=1}^{T} \chi^{(t)}(F^{(t)})$.

A New Discrepancy Game

Let Alice and Bob be two players competing in a T rounds game.

They are both given a set system $(X, \mathcal{F})$ with VC-dimension $\leq d$.

Each round t is as follows:

- Alice picks a coloring $\chi^{(t)}$ of X and sends it to Bob.
- Bob picks a set $F^{(t)} \in \mathcal{F}$ and sends it to Alice.

Alice's goal: minimize $\sum_{t=1}^{T} \chi^{(t)}(F^{(t)})$.

Bob's goal: maximize $\sum_{t=1}^{T} \chi^{(t)}(F^{(t)})$.

Random coloring: $\frac{1}{T} \sum_{t=1}^{T} \chi^{(t)}(F^{(t)}) = O(\sqrt{n \log(m)})$

Discrepancy Game

Discrepancy Game value

There is a strategy for Alice to pick colorings such that,

$$\sum_{t=1}^{T} \chi^{(t)}(F^{(t)}) \leq (4+\sqrt{2})\, T^{1-1/2d} \sqrt{n} \log^2(mn) + 3\,T \sqrt{n \ln\left(\frac{T}{n}\right)}$$

Discrepancy Game

Discrepancy Game value

There is a strategy for Alice to pick colorings such that,

$$\sum_{t=1}^{T} \chi^{(t)}(F^{(t)}) \leq (4+\sqrt{2})\, T^{1-1/2d} \sqrt{n} \log^2(mn) + 3T \sqrt{n \ln\left(\frac{T}{n}\right)}$$

In particular for an n rounds game, we obtain:

$$\frac{1}{n} \sum_{t=1}^{n} \chi^{(t)}(F^{(t)}) = O\left(n^{1/2-1/2d} \log^2(mn)\right)$$

The Lovett-Meka algorithm

[LM15] following [Ban10]'s breakthrough gave a simpler algorithm both in time and complexity realizing [Spe85]'s bound.

The Lovett-Meka algorithm

[LM15] following [Ban10]'s breakthrough gave a simpler algorithm both in time and complexity realizing [Spe85]'s bound.

[LM15] Partial Coloring Theorem

Let $c_{F_1}, \ldots, c_{F_m} \geq 0$ such that $\sum_{F \in F} \exp(-c_F^2) \leq n$. There exists an algorithm that finds w.h.p. a coloring χ such that:

- $\forall F \in \mathcal{F}, \chi(F) \leq c_F \sqrt{|F|}$
- $|\{1 \leq i \leq n : |x_i| = 1\}| \geq n/2$
- $\forall A \subseteq X, \chi(A) \leq \sqrt{|A| \log(mn)}$

This algorithm has runtime $O\left((n+m)^3 \log(mn)\right)$.

The Lovett-Meka algorithm

[LM15] following [Ban10]'s breakthrough gave a simpler algorithm both in time and complexity realizing [Spe85]'s bound.

[LM15] Partial Coloring Theorem

Let $c_{F_1}, \ldots, c_{F_m} \geq 0$ such that $\sum_{F \in F} \exp(-c_F^2) \leq n$. There exists an algorithm that finds w.h.p. a coloring χ such that:

- $\forall F \in \mathcal{F}, \chi(F) \leq c_F \sqrt{|F|}$
- $|\{1 \leq i \leq n : |x_i| = 1\}| \geq n/2$
- $\forall A \subseteq X, \chi(A) \leq \sqrt{|A| \log(mn)}$

This algorithm has runtime $O\left((n+m)^3 \log(mn)\right)$.

By iteratively running this algorithm on uncolored elements of X, we obtain a full coloring of X.

Outline of the proof of the discrepancy game value

Alice's strategy is to use Bob's strategy against him, in the long run, by using the sets he gives as constraints in [LM15]. She sets $\forall i < t, c_{F^{(i)}} = \sqrt{\ln\left(\frac{T}{n}\right)}$.

Outline of the proof of the discrepancy game value

Alice's strategy is to use Bob's strategy against him, in the long run, by using the sets he gives as constraints in [LM15]. She sets $\forall i < t, c_{F(i)} = \sqrt{\ln\left(\frac{T}{n}\right)}$.

Key property of discrepancy

Let χ be a coloring of X and F, F' subsets of X. We have:

$$\chi(F) \leq \chi(F') + \chi(\Delta(F, F'))$$

where $\Delta(F, F') = (F \cup F') \backslash (F \cap F')$.

Outline of the proof of the discrepancy game value

Alice's strategy is to use Bob's strategy against him, in the long run, by using the sets he gives as constraints in [LM15]. She sets $\forall i < t, c_{F^{(i)}} = \sqrt{\ln\left(\frac{T}{n}\right)}$.

Key property of discrepancy

Let χ be a coloring of X and F, F' subsets of X. We have:

$$\chi(F) \leq \chi(F') + \chi(\Delta(F, F'))$$

where $\Delta(F, F') = (F \cup F') \backslash (F \cap F')$.

At round t, since $\forall i < t, \chi^{(t)}(F^{(i)}) = \sqrt{|F^{(i)}| \ln\left(\frac{T}{n}\right)}$,

$$\chi^{(t)}(F^{(t)}) \leq \min_{i<t} \sqrt{|F^{(i)}| \ln\left(\frac{T}{n}\right)} + \chi^{(t)}(\Delta(F^{(t)}, F^{(i)}))$$

Outline of the proof of the discrepancy game value (part 2)

$$\chi^{(t)}(F^{(t)}) \leq \min_{i<t} \sqrt{|F^{(i)}| \ln\left(\frac{T}{n}\right)} + \chi^{(t)}(\Delta(F^{(t)}, F^{(i)}))$$

Outline of the proof of the discrepancy game value (part 2)

$$\chi^{(t)}(F^{(t)}) \leq \min_{i<t} \sqrt{|F^{(i)}| \ln \left(\frac{T}{n} \right)} + \chi^{(t)}(\Delta(F^{(t)}, F^{(i)}))$$

$$\Rightarrow \chi^{(t)}(F^{(t)}) \leq \sqrt{n \ln \left(\frac{T}{n} \right)} + \min_{i<t} \chi^{(t)}(\Delta(F^{(t)}, F^{(i)}))$$

Outline of the proof of the discrepancy game value (part 2)

$$\chi^{(t)}(F^{(t)}) \leq \min_{i<t} \sqrt{|F^{(i)}| \ln\left(\frac{T}{n}\right)} + \chi^{(t)}(\Delta(F^{(t)}, F^{(i)}))$$

$$\Rightarrow \chi^{(t)}(F^{(t)}) \leq \sqrt{n \ln\left(\frac{T}{n}\right)} + \min_{i<t} \chi^{(t)}(\Delta(F^{(t)}, F^{(i)}))$$

$$\Rightarrow \chi^{(t)}(F^{(t)}) \leq \sqrt{n \ln\left(\frac{T}{n}\right)} + \min_{i<t} \sqrt{|\Delta(F^{(t)}, F^{(i)})|} \log^2(mn)$$

Outline of the proof of the discrepancy game value (part 2)

$$\chi^{(t)}(F^{(t)}) \leq \min_{i<t} \sqrt{|F^{(i)}| \ln \left(\frac{T}{n}\right)} + \chi^{(t)}(\Delta(F^{(t)}, F^{(i)}))$$

$$\Rightarrow \chi^{(t)}(F^{(t)}) \leq \sqrt{n \ln \left(\frac{T}{n}\right)} + \min_{i<t} \chi^{(t)}(\Delta(F^{(t)}, F^{(i)}))$$

$$\Rightarrow \chi^{(t)}(F^{(t)}) \leq \sqrt{n \ln \left(\frac{T}{n}\right)} + \min_{i<t} \sqrt{|\Delta(F^{(t)}, F^{(i)})|} \log^2(mn)$$

[Hau95] packing theorem

Let $(X, \mathcal{F})$ be a set system with VC-dimension $\leq d$ with the property that $\exists \delta \leq n$ s.t. $\forall (F, F') \in \mathcal{F}, \Delta(F, F') \geq \delta$. Then $|\mathcal{F}| = O\left(\left(\frac{n}{\delta}\right)^d\right)$.

Outline of the proof of the discrepancy game value (part 2)

$$\chi^{(t)}(F^{(t)}) \leq \min_{i<t} \sqrt{|F^{(i)}| \ln\left(\frac{T}{n}\right)} + \chi^{(t)}(\Delta(F^{(t)}, F^{(i)}))$$

$$\Rightarrow \chi^{(t)}(F^{(t)}) \leq \sqrt{n \ln\left(\frac{T}{n}\right)} + \min_{i<t} \chi^{(t)}(\Delta(F^{(t)}, F^{(i)}))$$

$$\Rightarrow \chi^{(t)}(F^{(t)}) \leq \sqrt{n \ln\left(\frac{T}{n}\right)} + \min_{i<t} \sqrt{|\Delta(F^{(t)}, F^{(i)})|} \log^2(mn)$$

[Hau95] packing theorem

Let $(X, \mathcal{F})$ be a set system with VC-dimension $\leq d$ with the property that $\exists \delta \leq n$ s.t. $\forall (F, F') \in \mathcal{F}, \Delta(F, F') \geq \delta$. Then $|\mathcal{F}| = O\left(\left(\frac{n}{\delta}\right)^d\right)$.

[Hau95] packing theorem gives a bound on the number of sets that can *all* have big symmetric difference with each others.

Outline of the proof of the discrepancy game value (part 2)

$$\chi^{(t)}(F^{(t)}) \leq \min_{i<t} \sqrt{|F^{(i)}| \ln \binom{T}{n}} + \chi^{(t)}(\Delta(F^{(t)}, F^{(i)}))$$

$$\Rightarrow \chi^{(t)}(F^{(t)}) \leq \sqrt{n \ln \binom{T}{n}} + \min_{i<t} \chi^{(t)}(\Delta(F^{(t)}, F^{(i)}))$$

$$\Rightarrow \chi^{(t)}(F^{(t)}) \leq \sqrt{n \ln \binom{T}{n}} + \min_{i<t} \sqrt{|\Delta(F^{(t)}, F^{(i)})|} \log^2(mn)$$

[Hau95] packing theorem

Let $(X, \mathcal{F})$ be a set system with VC-dimension $\leq d$ with the property that $\exists \delta \leq n$ s.t. $\forall (F, F') \in \mathcal{F}, \Delta(F, F') \geq \delta$. Then $|\mathcal{F}| = O\left(\left(\frac{n}{\delta}\right)^d\right)$.

[Hau95] packing theorem gives a bound on the number of sets that can *all* have big symmetric difference with each others.

This allows us to bound the right hand side as if t is large, some pairs of sets given by Bob must have small symmetric difference as otherwise they would create a packing of size t.

Application: A Multiplicative Weight Update Algorithm

We design an algorithm as a direct application of the New Discrepancy Game:

Algorithm: MWU algorithm using the New Discrepancy Game

Input: $(X, \mathcal{F})$ a set system with VC-dimension $\leq d$

$y^{(0)} = (1, \dots, 1) \in \mathbb{R}^m$

for $t \leftarrow 1$ *to* T **do**

  $\chi^{(t)} \leftarrow$ [LM15] with $\forall i < t, c_{F^{(i)}} = \sqrt{\ln\left(\frac{T}{n}\right)}$

  $i^{(t)} \sim y^{(t)}$

  $F^{(t)} \leftarrow \mathcal{F}_{i^{(t)}}$

  $\forall j \leq m, y_j^{(t+1)} = y_j^{(t)}(1 + \eta\chi^{(t)}(\mathcal{F}_j))$

Application: A Multiplicative Weight Update Algorithm

We design an algorithm as a direct application of the New Discrepancy Game:

Algorithm: MWU algorithm using the New Discrepancy Game

Input: $(X, \mathcal{F})$ a set system with VC-dimension $\leq d$

$y^{(0)} = (1, \ldots, 1) \in \mathbb{R}^m$

for $t \leftarrow 1$ *to* T **do**

- $\chi^{(t)} \leftarrow$ [LM15] with $\forall i < t, c_{F^{(i)}} = \sqrt{\ln\left(\frac{T}{n}\right)}$
- $i^{(t)} \sim y^{(t)}$
- $F^{(t)} \leftarrow \mathcal{F}_{i^{(t)}}$
- $\forall j \leq m, y_j^{(t+1)} = y_j^{(t)}(1 + \eta\chi^{(t)}(\mathcal{F}_j))$

RT: $O(n^4 + mn^{3/2-1/2d})$

Application: A Multiplicative Weight Update Algorithm

We design an algorithm as a direct application of the New Discrepancy Game:

Algorithm: MWU algorithm using the New Discrepancy Game

Input: $(X, \mathcal{F})$ a set system with VC-dimension $\leq d$

```
y^(0) = (1, ..., 1) ∈ R^m
for t ← 1 to T do
    χ^(t) ← [LM15] with ∀i < t, c_{F^(i)} = sqrt(ln(T/n))
    i^(t) ~ y^(t)
    F^(t) ← F_{i^(t)}
    ∀j ≤ m, y_j^(t+1) = y_j^(t) (1 + η χ^(t)(F_j))
```

1. $y^{(0)} = (1, \ldots, 1) \in \mathbb{R}^m$
2. **for** $t \leftarrow 1$ *to* T **do**
3. $\quad \chi^{(t)} \leftarrow$ [LM15] with $\forall i < t, c_{F^{(i)}} = \sqrt{\ln\left(\frac{T}{n}\right)}$
4. $\quad i^{(t)} \sim y^{(t)}$
5. $\quad F^{(t)} \leftarrow \mathcal{F}_{i^{(t)}}$
6. $\quad \forall j \leq m, y_j^{(t+1)} = y_j^{(t)}(1 + \eta\chi^{(t)}(\mathcal{F}_j))$

RT: $O(n^4 + mn^{3/2-1/2d})$

This algorithm applies Alice's strategy of picking colorings using [LM15]'s algorithm and emulates Bob's behaviour by picking sets according to some weights that increase if the set associated has big discrepancy.

Application: A Multiplicative Weight Update Algorithm

The Multiplicative Weight Update analysis of the algorithm gives:

$$\forall F \in \mathcal{F}, \sum_{t=1}^{T} \chi^{(t)}(F) \leq \eta \sum_{t=1}^{T} \chi^{(t)}(F)^2 + \frac{\ln(m)}{\eta}$$

$$+ \sum_{t=1}^{T} \sum_{j=1}^{m} \frac{y_j^{(t)}}{\sum_{j=1}^{m} y_j^{(t)}} \chi^{(t)}(\mathcal{F}_j)$$

Application: A Multiplicative Weight Update Algorithm

The Multiplicative Weight Update analysis of the algorithm gives:

$$\forall F \in \mathcal{F}, \sum_{t=1}^{T} \chi^{(t)}(F) \leq \underbrace{\eta \sum_{t=1}^{T} \chi^{(t)}(F)^2 + \frac{\ln(m)}{\eta}}_{\text{error terms}} + \sum_{t=1}^{T} \sum_{j=1}^{m} \frac{y_j^{(t)}}{\sum_{j=1}^{m} y_j^{(t)}} \chi^{(t)}(\mathcal{F}_j)$$

Application: A Multiplicative Weight Update Algorithm

The Multiplicative Weight Update analysis of the algorithm gives:

$$\forall F \in \mathcal{F}, \sum_{t=1}^{T} \chi^{(t)}(F) \leq \underbrace{\eta \sum_{t=1}^{T} \chi^{(t)}(F)^2 + \frac{\ln(m)}{\eta}}_{\text{error terms}}$$

$$+ \underbrace{\sum_{t=1}^{T} \sum_{j=1}^{m} \frac{y_j^{(t)}}{\sum_{j=1}^{m} y_j^{(t)}} \chi^{(t)}(\mathcal{F}_j)}_{\text{bounded by the game}}$$

Application: A Multiplicative Weight Update Algorithm

The Multiplicative Weight Update analysis of the algorithm gives:

$$\forall F \in \mathcal{F}, \sum_{t=1}^{T} \chi^{(t)}(F) \le \underbrace{\eta \sum_{t=1}^{T} \chi^{(t)}(F)^2 + \frac{\ln(m)}{\eta}}_{\text{error terms}}$$
$$+ \underbrace{\sum_{t=1}^{T} \sum_{j=1}^{m} \frac{y_j^{(t)}}{\sum_{j=1}^{m} y_j^{(t)}} \chi^{(t)}(\mathcal{F}_j)}_{\text{bounded by the game}}$$

For n rounds: $\forall F \in \mathcal{F}, \frac{1}{n} \sum_{t=1}^{n} \chi^{(t)}(F) = O\left(n^{1/2-1/2d} \log^2(mn)\right)$

Application: A Multiplicative Weight Update Algorithm

The Multiplicative Weight Update analysis of the algorithm gives:

$$\forall F \in \mathcal{F}, \sum_{t=1}^{T} \chi^{(t)}(F) \leq \underbrace{\eta \sum_{t=1}^{T} \chi^{(t)}(F)^2 + \frac{\ln(m)}{\eta}}_{\text{error terms}} + \underbrace{\sum_{t=1}^{T} \sum_{j=1}^{m} \frac{y_j^{(t)}}{\sum_{j=1}^{m} y_j^{(t)}} \chi^{(t)}(\mathcal{F}_j)}_{\text{bounded by the game}}$$

For n rounds: $\forall F \in \mathcal{F}, \frac{1}{n} \sum_{t=1}^{n} \chi^{(t)}(F) = O\left(n^{1/2-1/2d} \log^2(mn)\right)$

Which implies $\exists t < n$ s.t.

$$|\{F \in \mathcal{F} : \chi^{(t)}(F) = O(n^{1/2-1/2d} \log^2(mn))\}| \geq \frac{m}{2}$$

Another approach: Computing minimal coverings

Set covering + [LM15]

Using [LM15]'s algorithm and a set covering algorithm constructs a coloring with discrepancy $O(n^{1/2-1/2d}\log^2(mn))$ in time $mn^{2/d}+n^3\log^{3d}(n)$.

Another approach: Computing minimal coverings

Set covering + [LM15]

Using [LM15]'s algorithm and a set covering algorithm constructs a coloring with discrepancy $O(n^{1/2-1/2d}\log^2(mn))$ in time $mn^{2/d}+n^3\log^{3d}(n)$.

Outline of the method:

Another approach: Computing minimal coverings

Set covering + [LM15]

Using [LM15]'s algorithm and a set covering algorithm constructs a coloring with discrepancy $O(n^{1/2-1/2d}\log^2(mn))$ in time $mn^{2/d}+n^3\log^{3d}(n)$.

Outline of the method:

- New result: We give an $O(mn^{2/d}+n^{2+2/d}\log(n))$ near-minimal covering algorithm with which we compute a $n^{1-1/d}$-covering $\mathcal{C}$ of $(X,\mathcal{F})$ that has size n.

Another approach: Computing minimal coverings

Set covering + [LM15]

Using [LM15]'s algorithm and a set covering algorithm constructs a coloring with discrepancy $O(n^{1/2-1/2d}\log^2(mn))$ in time $mn^{2/d}+n^3\log^{3d}(n)$.

Outline of the method:

- New result: We give an $O(mn^{2/d}+n^{2+2/d}\log(n))$ near-minimal covering algorithm with which we compute a $n^{1-1/d}$-covering $\mathcal{C}$ of $(X,\mathcal{F})$ that has size n.
- We then compute a coloring χ of X using [LM15] with constraints $\forall C\in\mathcal{C}, c_C=0$.

Another approach: Computing minimal coverings

Set covering + [LM15]

Using [LM15]'s algorithm and a set covering algorithm constructs a coloring with discrepancy $O(n^{1/2-1/2d}\log^2(mn))$ in time $mn^{2/d}+n^3\log^{3d}(n)$.

Outline of the method:

- New result: We give an $O(mn^{2/d}+n^{2+2/d}\log(n))$ near-minimal covering algorithm with which we compute a $n^{1-1/d}$-covering $\mathcal{C}$ of $(X,\mathcal{F})$ that has size n.
- We then compute a coloring χ of X using [LM15] with constraints $\forall C\in\mathcal{C}, c_C=0$.
- Because $\mathcal{C}$ is a covering, $\forall F\in\mathcal{F}, \exists C\in\mathcal{C}$ s.t. :

$$\chi(F)\leq\chi(C)+\chi(\Delta(F,C))\leq 0+\sqrt{n^{1-1/d}}\log^2(mn)=O(n^{1/2-1/2d}\log^2(mn))$$

Another approach: Computing minimal coverings

Set covering + [LM15]

Using [LM15]'s algorithm and a set covering algorithm constructs a coloring with discrepancy $O(n^{1/2-1/2d}\log^2(mn))$ in time $mn^{2/d}+n^3\log^{3d}(n)$.

Outline of the method:

- New result: We give an $O(mn^{2/d}+n^{2+2/d}\log(n))$ near-minimal covering algorithm with which we compute a $n^{1-1/d}$-covering $\mathcal{C}$ of $(X,\mathcal{F})$ that has size n.
- We then compute a coloring χ of X using [LM15] with constraints $\forall C\in\mathcal{C}, c_C=0$.
- Because $\mathcal{C}$ is a covering, $\forall F\in\mathcal{F}, \exists C\in\mathcal{C}$ s.t. :

$$\chi(F)\leq\chi(C)+\chi(\Delta(F,C))\leq 0+\sqrt{n^{1-1/d}}\log^2(mn)=O(n^{1/2-1/2d}\log^2(mn))$$

This can be further improved to $O(n^{1/2-1/2d})$ using chaining.

Simplicial partitions

Simplicial Partitions

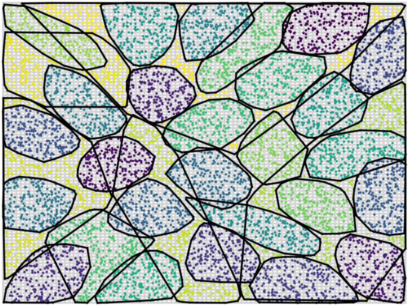
Given $(X, \mathcal{F})$ and a parameter $t \in [1..n]$, find t partitions $P_1, .., P_t$ of size $O\left(\frac{n}{t}\right)$ of X such as to minimize the *crossing number*:

$$\max_{F \in \mathcal{F}} \left|\{P \in \{P_1, ...P_t\} \text{ s.t. } P \text{ intersects } F\}\right|$$

Simplicial Partitions

Given $(X, \mathcal{F})$ and a parameter $t \in [1..n]$, find t partitions $P_1, .., P_t$ of size $O\left(\frac{n}{t}\right)$ of X such as to minimize the *crossing number*:

$$\max_{F \in \mathcal{F}} |\{P \in \{P_1, ...P_t\} \text{ s.t. } P \text{ intersects } F\}|$$



Simplicial Partitions

Given $(X, \mathcal{F})$ and a parameter $t \in [1..n]$, find t partitions $P_1, .., P_t$ of size $O\left(\frac{n}{t}\right)$ of X such as to minimize the *crossing number*:

$$\max_{F \in \mathcal{F}} |\{P \in \{P_1, ...P_t\} \text{ s.t. } P \text{ intersects } F\}|$$

[Mat92] proved that set systems spanned by halfspaces in $\mathbb{R}^d$ admit partition with crossing number $O(t^{1-1/d})$.

Simplicial Partitions

Given $(X, \mathcal{F})$ and a parameter $t \in [1..n]$, find t partitions $P_1, .., P_t$ of size $O\left(\frac{n}{t}\right)$ of X such as to minimize the *crossing number*:

$$\max_{F \in \mathcal{F}} \left|\{P \in \{P_1, ...P_t\} \text{ s.t. } P \text{ intersects } F\}\right|$$

[Mat92] proved that set systems spanned by halfspaces in $\mathbb{R}^d$ admit partition with crossing number $O(t^{1-1/d})$.

There are no fast algorithms in practice in dimension > 3 to compute small crossing number partitions.

A greedy approach

An Ordering Lemma

Let $(X, \mathcal{F})$ be a set system such that $\forall Y \subseteq X$ and $s \leq |Y|, (Y, \mathcal{F}|_Y)$ admits a partition of size s with crossing number $s^{1-1/d}$. $\mathcal{P} = \{P_1, \ldots, P_t\}$ be t disjoint subsets of X, where $|P_i| = n/t$ for all $i \in [t]$.

A greedy approach

An Ordering Lemma

Let $(X, \mathcal{F})$ be a set system such that $\forall Y \subseteq X$ and $s \leq |Y|$, $(Y, \mathcal{F}|_Y)$ admits a partition of size s with crossing number $s^{1-1/d}$. $\mathcal{P} = \{P_1, \ldots, P_t\}$ be t disjoint subsets of X, where $|P_i| = n/t$ for all $i \in [t]$.

Let $P_l \in \mathcal{P}$ be selected uniformly at random.

A greedy approach

An Ordering Lemma

Let $(X, \mathcal{F})$ be a set system such that $\forall Y \subseteq X$ and $s \leq |Y|$, $(Y, \mathcal{F}|_Y)$ admits a partition of size s with crossing number $s^{1-1/d}$. $\mathcal{P} = \{P_1, \ldots, P_t\}$ be t disjoint subsets of X, where $|P_i| = n/t$ for all $i \in [t]$.

Let $P_l \in \mathcal{P}$ be selected uniformly at random.

Then there exists an ordering of the elements of P_l, say $\langle x_1, x_2, \ldots, x_{n/t} \rangle$, such that w.h.p.,

A greedy approach

An Ordering Lemma

Let $(X, \mathcal{F})$ be a set system such that $\forall Y \subseteq X$ and $s \leq |Y|$, $(Y, \mathcal{F}|_Y)$ admits a partition of size s with crossing number $s^{1-1/d}$. $\mathcal{P} = \{P_1, \dots, P_t\}$ be t disjoint subsets of X, where $|P_i| = n/t$ for all $i \in [t]$.

Let $P_l \in \mathcal{P}$ be selected uniformly at random.

Then there exists an ordering of the elements of P_l, say $\langle x_1, x_2, \dots, x_{n/t} \rangle$, such that w.h.p.,

$\forall k \leq \frac{n}{t}$, the prefix set $\{x_1, \dots, x_k\}$ is crossed by at most $\frac{4mk^{1/d}}{n^{1/d}}$ sets.

A greedy approach

An Ordering Lemma

Let $(X, \mathcal{F})$ be a set system such that $\forall Y \subseteq X$ and $s \leq |Y|$, $(Y, \mathcal{F}|_Y)$ admits a partition of size s with crossing number $s^{1-1/d}$. $\mathcal{P} = \{P_1, \ldots, P_t\}$ be t disjoint subsets of X, where $|P_i| = n/t$ for all $i \in [t]$.

Let $P_l \in \mathcal{P}$ be selected uniformly at random.

Then there exists an ordering of the elements of P_l, say $\langle x_1, x_2, \ldots, x_{n/t} \rangle$, such that w.h.p.,

$\forall k \leq \frac{n}{t}$, the prefix set $\{x_1, \ldots, x_k\}$ is crossed by at most $\dfrac{4mk^{1/d}}{n^{1/d}}$ sets.

An algorithm that picks elements following this potential function builds partitions with crossing number $O(t^{1-1/d} \ln(t))$.

A greedy algorithm

Algorithm: Algorithm minimizing the potential function

$n \leftarrow |X|, m \leftarrow |\mathcal{F}|, \mathcal{P} \leftarrow \emptyset, \pi(F) \leftarrow 1 \; \forall F \in \mathcal{F}$.

for $i \leftarrow 1$ *to* t **do**

    $x_0 \leftarrow$ a random element of X, $P_i \leftarrow \{x_0\}, \omega(P_i) \leftarrow 0$.

    for $F \in \mathcal{F}$ **do**

        for $x \in X$ *with* F *crossing* $\{x_0, x\}$ **do**

            $\omega(x) \leftarrow \omega(x) + \pi(F)$

    for $k \leftarrow 2$ *to* $\frac{n}{t}$ **do**

        $y_k \leftarrow$ any element of X s.t. $\omega(P_i) + \omega(y_k) \leq \frac{2k^{1/d} \sum_{F \in \mathcal{F}} \pi(F)}{|X|^{1/d}}$.

        $X \leftarrow X \setminus \{y_k\}, \omega(P_i) \leftarrow \omega(P_i) + \omega(y_k)$.

        for $F \in (P_i, y_k)$ **do**

            for $x \in X$ *with* F *crossing* $\{x_0, x\}$ **do**

                $\omega(x) \leftarrow \omega(x) - \pi(F)$

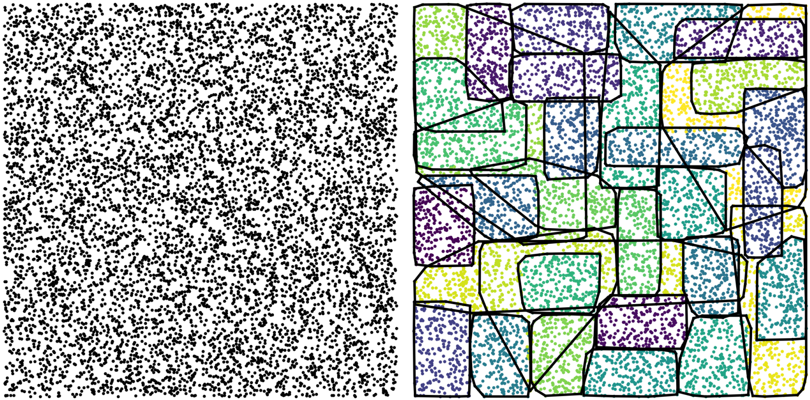
        $P_i \leftarrow P_i \cup \{y_k\}$

    $\forall F \in \mathcal{F}, \pi(F) \leftarrow \pi(F) \cdot 2^{I(P_i, F)}$

    $\mathcal{P} \leftarrow \mathcal{P} \cup \{P_i\}$

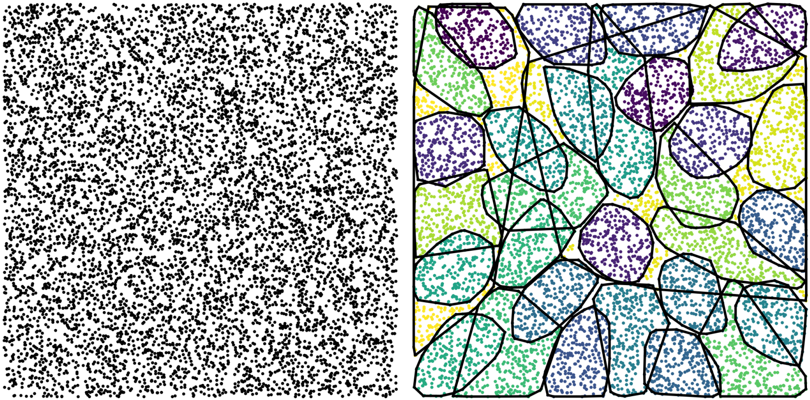
return $\mathcal{P}$

Experiments



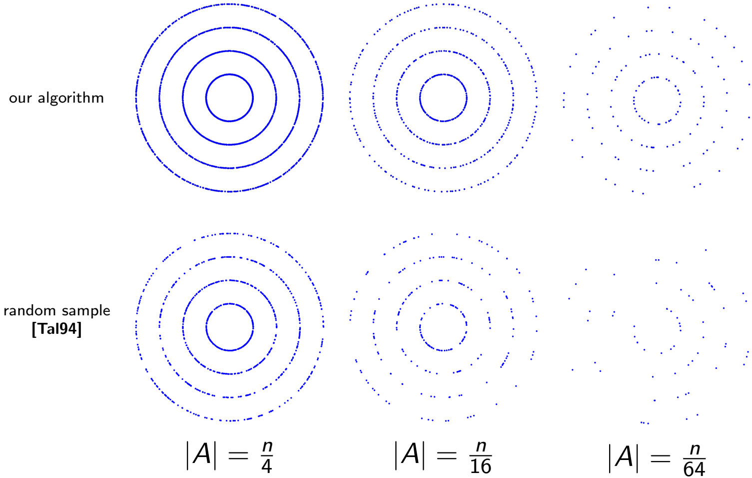
$n = 8192, t = 32$

Experiments



$n = 8192, t = 32$

ε-approximation experiments



Open problems

- Faster algorithms for discrepancy.

Open problems

- Faster algorithms for discrepancy.
- Faster set packing/covering algorithms.

Open problems

- Faster algorithms for discrepancy.
- Faster set packing/covering algorithms.
- Fast algorithms constructing small crossing number partitions with guarantees for set systems in dimension > 3.

Open problems

- Faster algorithms for discrepancy.
- Faster set packing/covering algorithms.
- Fast algorithms constructing small crossing number partitions with guarantees for set systems in dimension > 3.

Thank you!

References

References I

Nikhil Bansal.
Constructive algorithms for discrepancy minimization.
In *2010 IEEE 51st Annual Symposium on Foundations of Computer Science*, pages 3–10. IEEE, 2010.

David Haussler.
Sphere packing numbers for subsets of the boolean n-cube with bounded vapnik-chervonenkis dimension.
Journal of Combinatorial Theory, Series A, 69(2):217–232, 1995.

Shachar Lovett and Raghu Meka.
Constructive discrepancy minimization by walking on the edges.
SIAM Journal on Computing, 44(5):1573–1582, 2015.

References II

Jiří Matoušek.
Efficient partition trees.
Discrete & Computational Geometry, 8:315–334, 1992.

Jiří Matoušek.
Tight upper bounds for the discrepancy of half-spaces.
Discrete & Computational Geometry, 13:593–601, 1995.

Joel Spencer.
Six standard deviations suffice.
Transactions of the American mathematical society, 289(2):679–706, 1985.