Dispersion of point sets in high dimensions

Jan Vybíral
Czech Technical University
Prague, Czech Republic

joint work with **M. Ullrich** (U Linz, Austria)

Point Distributions Webinar
September 2021

Dispersion: definition

- Let $d \geq 2$ and let $[0,1]^d$ be the unit cube in $\mathbb{R}^d$
- Let $\mathcal{P}_n = \{x_1, \ldots, x_n\} \subset [0,1]^d$ be a set of n points

Dispersion: definition

- Let $d \geq 2$ and let $[0,1]^d$ be the unit cube in $\mathbb{R}^d$
- Let $\mathcal{P}_n = \{x_1, \dots, x_n\} \subset [0,1]^d$ be a set of n points
- Let $\mathcal{B}_{ax}^d$ be all boxes in $[0,1]^d$ with sides parallel to coordinate axes
- Then

$$\mathrm{disp}(\mathcal{P}_n, d) := \sup \Big\{ |B| : \ B \in \mathcal{B}_{ax}^d \text{ with } \mathcal{P}_n \cap B = \emptyset \Big\}$$

is the volume of the largest box not intersecting $\mathcal{P}_n$

Dispersion: definition

- Let $d \geq 2$ and let $[0,1]^d$ be the unit cube in $\mathbb{R}^d$
- Let $\mathcal{P}_n = \{x_1, \dots, x_n\} \subset [0,1]^d$ be a set of n points
- Let $\mathcal{B}^d_{ax}$ be all boxes in $[0,1]^d$ with sides parallel to coordinate axes
- Then

$$\mathrm{disp}(\mathcal{P}_n, d) := \sup \Big\{ |B| : \ B \in \mathcal{B}^d_{ax} \text{ with } \mathcal{P}_n \cap B = \emptyset \Big\}$$

is the volume of the largest box not intersecting $\mathcal{P}_n$

Dispersion of a point set:
volume of the largest box not intersecting this set

Dispersion: definition

- A well-spread point set should have a small dispersion - i.e. no big holes

$$\varepsilon := \operatorname{disp}(n,d) := \inf_{\substack{\mathcal{P}_n \subset [0,1]^d \\ \#\mathcal{P}_n = n}} \operatorname{disp}(\mathcal{P}_n, d).$$

...search for the optimally distributed point set.

Dispersion: definition

- A well-spread point set should have a small dispersion - i.e. no big holes

$$\varepsilon := \mathrm{disp}(n,d) := \inf_{\substack{\mathcal{P}_n \subset [0,1]^d \\ \#\mathcal{P}_n = n}} \mathrm{disp}(\mathcal{P}_n, d).$$

...search for the optimally distributed point set.

- $$N(\varepsilon, d) := \min\{n : \mathrm{disp}(n,d) \le \varepsilon\}$$

...size of the smallest set achieving dispersion at most ε

Dispersion: definition

- A well-spread point set should have a small dispersion - i.e. no big holes

$$\varepsilon := \mathrm{disp}(n,d) := \inf_{\substack{\mathcal{P}_n \subset [0,1]^d \\ \#\mathcal{P}_n = n}} \mathrm{disp}(\mathcal{P}_n, d).$$

  ...search for the optimally distributed point set.

- $$N(\varepsilon, d) := \min\{n : \mathrm{disp}(n,d) \le \varepsilon\}$$

  ...size of the smallest set achieving dispersion at most ε

- Three coupled parameters ε, d, N; estimates can be given in various forms

Dispersion: different bounds

- Trivial pigeonhole principle:

$$1 \geq \mathrm{disp}(n,d) \geq \frac{1}{n+1}$$

- Aistleitner, Hinrichs, and Rudolf (2017):

$$\mathrm{disp}(n,d) \geq \frac{\log_2(d)}{4(n+\log_2(d))}$$

- Larcher (unpublished):

$$\mathrm{disp}(n,d) \leq \frac{2^{7d+1}}{n}$$

Dispersion: different bounds

- Rudolf ($n > 2d$, 2018):

$$\mathrm{disp}(n,d) \leq \frac{4d}{n} \log_2\left(\frac{9n}{d}\right)$$

- Sosnovec (2018): Randomized point set with at most $c_\varepsilon \log_2(d)$ points in $[0,1]^d$ and dispersion at most $\varepsilon \in (0, \frac{1}{4}]$
- Ullrich, V. (2018): Refining the argument of Sosnovec

$$\#\mathcal{P} \leq 2^7 \frac{(1+\log_2(\varepsilon^{-1}))^2}{\varepsilon^2} \log_2(d)$$

i.e.

$$\mathrm{disp}(n,d) \leq c \log_2(n) \sqrt{\frac{\log_2(d)}{n}}$$

Dispersion: updates

- Calculation of the dispersion for well-known sets and lattices (Krieg 2018; Temlyakov 2018); further bounds by Buch, Chao, Litvak, Livshyts, and others
- Essentially - estimates good in n are not optimal in d and vice versa
- Conjecture - at least for a large range of parameters:

$$\mathrm{disp}(n,d) \approx \min\left\{1, \frac{\log_2(d)}{n}\right\} \quad ???$$

log(d): Random construction of Sosnovec

Crucial observations:

- The boundary and the inside of $[0,1]^d$ play a different role.
- If we do not sample close to the boundary, then large intervals cover all possible choices.

$\log(d)$: Random construction of Sosnovec

Crucial observations:

- The boundary and the inside of $[0,1]^d$ play a different role.
- If we do not sample close to the boundary, then large intervals cover all possible choices.

Start with parameters:
m (corresponds to ε roughly as $\varepsilon = 2^{-m}$)
d - dimension

Define

$$M_m := \left\{ \frac{1}{2^m}, \frac{2}{2^m}, \dots, \frac{2^m - 1}{2^m} \right\} \subset [0,1].$$

“Algorithm”: Sample randomly and uniformly from M_m^d until all large boxes get hit!

log(d): Random construction of Sosnovec

Active coordinates:

- Let $B = I_1 \times \cdots \times I_d$ with $|B| > 2^{-m}$
- If $|I_j| > 1 - 2^{-m}$, then $M_m \subset I_j$... **non-active coordinates**
- $|I_j| \leq 1 - 2^{-m}$... **active coordinates**

$\log(d)$: Random construction of Sosnovec

Active coordinates:

- Let $B = I_1 \times \cdots \times I_d$ with $|B| > 2^{-m}$
- If $|I_j| > 1 - 2^{-m}$, then $M_m \subset I_j$... **non-active coordinates**
- $|I_j| \leq 1 - 2^{-m}$... **active coordinates**

From

$$2^{-m} < |B| = \prod_{j=1}^{d} |I_j| \leq (1 - 2^{-m})^{\alpha}$$

the number of active coordinates α satisfies

$$\alpha \leq A_m := \frac{\log(2^{-m})}{\log(1 - 2^{-m})} \qquad \sim m2^m.$$

...independent of d...!

log(d): Random construction of Sosnovec

We plan to use the union bound - but we have infinitely many B's:

$\log(d)$: Random construction of Sosnovec

We plan to use the union bound - but we have infinitely many B's:

We split $\mathcal{B}^d_{ax}$ into finitely many groups with large intersections

$\log(d)$: Random construction of Sosnovec

We plan to use the union bound - but we have infinitely many B's:

We split $\mathcal{B}^d_{ax}$ into finitely many groups with large intersections

To each $B = I_1 \times \cdots \times I_d$ with $|B| > 2^{-m}$, we associate
$j = (j_1, \dots, j_{A_m})$ - set of active coordinates
$k = (k_{j_1}, \dots, k_{j_{A_m}})$ - one value lying in B, i.e. $k_{j_1} \in I_{j_1}$ etc.

$\log(d)$: Random construction of Sosnovec

We plan to use the union bound - but we have infinitely many B's:

We split $\mathcal{B}_{ax}^d$ into finitely many groups with large intersections

To each $B = I_1 \times \dots \times I_d$ with $|B| > 2^{-m}$, we associate
$j = (j_1, \dots, j_{A_m})$ - set of active coordinates
$k = (k_{j_1}, \dots, k_{j_{A_m}})$ - one value lying in B, i.e. $k_{j_1} \in I_{j_1}$ etc.

Put

$$B_j^k := J_1 \times \dots \times J_d, \quad \text{where} \quad J_\ell = \begin{cases} M_m & \ell \notin j \\ \{k_\ell\} & \ell \in j \end{cases}$$

At most $\binom{d}{A_m} \cdot (2^m)^{A_m}$ such data

log(d): Random construction of Sosnovec

$\mathcal{X} = \{x_1, \ldots, x_N\}$ randomly and uniformly sampled from M_m^d

$\log(d)$: Random construction of Sosnovec

$\mathcal{X} = \{x_1, \dots, x_N\}$ randomly and uniformly sampled from M_m^d

$$\mathbb{P}(\exists B : |B| > 2^{-m} \,\&\, \mathcal{X} \cap B = \emptyset) \leq \mathbb{P}(\exists j, k : \mathcal{X} \cap B_j^k = \emptyset)$$

$\log(d)$: Random construction of Sosnovec

$\mathcal{X} = \{x_1, \dots, x_N\}$ randomly and uniformly sampled from M_m^d

$$\mathbb{P}(\exists B : |B| > 2^{-m} \,\&\, \mathcal{X} \cap B = \emptyset) \le \mathbb{P}(\exists j, k : \mathcal{X} \cap B_j^k = \emptyset)$$
$$\le \sum_{j,k} \mathbb{P}(\mathcal{X} \cap B_j^k = \emptyset) = \sum_{j,k} \mathbb{P}(\forall \ell : x_\ell \notin B_j^k)$$

log(d): Random construction of Sosnovec

$\mathcal{X} = \{x_1, \dots, x_N\}$ randomly and uniformly sampled from M_m^d

$$\begin{aligned}\mathbb{P}(\exists B : |B| > 2^{-m} \,\&\, \mathcal{X} \cap B = \emptyset) &\le \mathbb{P}(\exists j, k : \mathcal{X} \cap B_j^k = \emptyset) \\ &\le \sum_{j,k} \mathbb{P}(\mathcal{X} \cap B_j^k = \emptyset) = \sum_{j,k} \mathbb{P}(\forall \ell : x_\ell \notin B_j^k) \\ &= \sum_{j,k} \mathbb{P}(x \notin B_j^k)^N \le \binom{d}{A_m} \cdot (2^m)^{A_m} \cdot \mathbb{P}(x \notin B_j^k)^N\end{aligned}$$

$\log(d)$: Random construction of Sosnovec

$\mathcal{X} = \{x_1, \dots, x_N\}$ randomly and uniformly sampled from M_m^d

$$\mathbb{P}(\exists B : |B| > 2^{-m} \,\&\, \mathcal{X} \cap B = \emptyset) \leq \mathbb{P}(\exists j, k : \mathcal{X} \cap B_j^k = \emptyset)$$

$$\leq \sum_{j,k} \mathbb{P}(\mathcal{X} \cap B_j^k = \emptyset) = \sum_{j,k} \mathbb{P}(\forall \ell : x_\ell \notin B_j^k)$$

$$= \sum_{j,k} \mathbb{P}(x \notin B_j^k)^N \leq \binom{d}{A_m} \cdot (2^m)^{A_m} \cdot \mathbb{P}(x \notin B_j^k)^N < 1?$$

$\log(d)$: Random construction of Sosnovec

$\mathcal{X} = \{x_1, \dots, x_N\}$ randomly and uniformly sampled from M_m^d

$$\mathbb{P}(\exists B : |B| > 2^{-m} \,\&\, \mathcal{X} \cap B = \emptyset) \le \mathbb{P}(\exists j, k : \mathcal{X} \cap B_j^k = \emptyset)$$
$$\le \sum_{j,k} \mathbb{P}(\mathcal{X} \cap B_j^k = \emptyset) = \sum_{j,k} \mathbb{P}(\forall \ell : x_\ell \notin B_j^k)$$
$$= \sum_{j,k} \mathbb{P}(x \notin B_j^k)^N \le \binom{d}{A_m} \cdot (2^m)^{A_m} \cdot \mathbb{P}(x \notin B_j^k)^N < 1?$$

For N large enough, this gets smaller than 1.
N depends logarithmically on d, super-exponentially on $\varepsilon = 2^{-m}$

Btw. $\mathbb{P}(x \notin B_j^k) = 1 - \mathbb{P}(x \in B_j^k) = 1 - 2^{-mA_m}$.

log(d): Random construction of Ullrich & V.

- Improved randomized construction, with better dependence in ε^{-1}, given by Ullrich, V. (2018)
- We split

$$\Omega_m := \left\{ B = I_1 \times \cdots \times I_d \subset [0,1]^d \ : \ |B| > \frac{1}{2^m} \right\}$$

into

$$\Omega_m(s,p) := \left\{ B \in \Omega_m : \forall \ell : \quad \frac{s_\ell}{2^m} < |I_\ell| \le \frac{s_\ell + 1}{2^m}, \right.$$
$$\left. p_\ell - \frac{1}{2^m} \le \inf I_\ell < p_\ell \right\}$$

$\log(d)$: Random construction of Ullrich & V.

- Improved randomized construction, with better dependence in ε^{-1}, given by Ullrich, V. (2018)
- We split

$$\Omega_m := \left\{ B = I_1 \times \cdots \times I_d \subset [0,1]^d \ : \ |B| > \frac{1}{2^m} \right\}$$

into

$$\Omega_m(s,p) := \Big\{ B \in \Omega_m : \forall \ell : \quad \frac{s_\ell}{2^m} < |I_\ell| \le \frac{s_\ell + 1}{2^m}, \\ p_\ell - \frac{1}{2^m} \le \inf I_\ell < p_\ell \Big\}$$

- and define

$$B_m(s,p) \ := \bigcap_{B \in \Omega_m(s,p)} B \ = \prod_{\ell=1}^{d} \Big[p_\ell, p_\ell + \frac{s_\ell - 1}{2^m} \Big].$$

log(d): Random construction of Ullrich & V.

- Let z be uniformly distributed in M_m^d. Then

$$\mathbb{P}\big(z \in B_m(s,p)\big) \geq \frac{1}{2^{m+4}}\,.$$

- We have better control of the probability that some point lies in all the cubes in the subgroup.

log(d): Random construction of Ullrich & V.

- Let z be uniformly distributed in M_m^d. Then

$$\mathbb{P}\big(z \in B_m(s,p)\big) \geq \frac{1}{2^{m+4}}\,.$$

- We have better control of the probability that some point lies in all the cubes in the subgroup.
- The aim is again to use an union bound.
- We use the same "algorithm", final estimate is

$$N := \#\mathcal{P}_N \leq 2^4 m 2^{2m} \log_2(2^{m+1} d).$$

$\log(d)$: Deterministic constructions - [S]

We "derandomize" the construction of Sosnovec!

Recall: $\varepsilon = 2^{-m}$, $M_m := \{1/2^m, \dots, (2^m - 1)/2^m\}$ and

$$\Omega_m := \left\{ B = I_1 \times \cdots \times I_d \subset [0,1]^d \,:\, |B| > \frac{1}{2^m} \right\}.$$

- Each $B \in \Omega_m$ has at most $A_m \sim m2^m$ active coordinates

$\log(d)$: Deterministic constructions - [S]

We "derandomize" the construction of Sosnovec!
Recall: $\varepsilon = 2^{-m}$, $M_m := \{1/2^m, \dots, (2^m - 1)/2^m\}$ and

$$\Omega_m := \left\{ B = I_1 \times \cdots \times I_d \subset [0,1]^d \,:\, |B| > \frac{1}{2^m} \right\}.$$

- Each $B \in \Omega_m$ has at most $A_m \sim m2^m$ active coordinates
- We need to find a subset of M_m^d, which intersects every B_j^k
- Sosnovec actually constructs randomly a set of points $x^1, \dots, x^N \in \{1, 2, \dots, 2^m - 1\}^d$ such that for every $\mathcal{A} \subset \{1, \dots, d\}$ with $|\mathcal{A}| = m2^m$ the set of restrictions $x^1|_{\mathcal{A}}, \dots, x^N|_{\mathcal{A}}$ contains all $(2^m - 1)^{m2^m}$ possible values.

$\log(d)$: Deterministic constructions - [S]

- M. Naor, L. J. Schulman, and A. Srinivasan, Splitters and near-optimal derandomization, 1995
- (n, k)-**universal sets:** a set $T \subset \{0, 1\}^n$, such that for any index set $S \subset \{1, 2, \ldots, n\}$ with $|S| = k$, the projection of T on S contains all possible 2^k configurations.
- (n, k, m)-**universal sets:** a set $T \subset \{0, 1, \ldots, m-1\}^n$, such that for any index set $S \subset \{1, 2, \ldots, n\}$ with $|S| = k$, the projection of T on S contains all possible m^k configurations.

$\log(d)$: Deterministic constructions - [S]

- If $m = 2$: $(n, k, 2)$-universal sets are (n, k)-universal sets
- If $m = 2^\mu$ is dyadic, (n, k, m)-universal sets can be obtained from $(\mu n, \mu k)$-universal sets. Indeed, we represent each $x \in \{0, 1, \dots, m-1\}^n = \{0, 1, \dots, 2^\mu - 1\}^d$ by an $\tilde{x} \in \{0, 1\}^{\mu n}$ just by writing each coordinate x_j in the dyadic representation.
- For the derandomization of Sosnovec's proof, we need an $(d, m2^m, 2^m - 1)$-universal set

$\log(d)$: Deterministic constructions - [S]

- There are deterministic constructions of (n,k)-universal sets of size $2^k k^{O(\log k)} \log(n)$
- Therefore, there is a deterministic construction of an $(n,k,m)=(n,k,2^\mu)$-universal set with size of

$$2^{\mu k}(\mu k)^{O(\log(\mu k))}\log(\mu n) = m^k (k\log(m))^{O(\log(k\log(m)))}\log(n\log(m)).$$

- Size of an $(d, m2^m, 2^m-1)$-universal set:

$$N = 2^{m^2 2^m}(m^2 2^m)^{O(\log(m^2 2^m))}\log(md).$$

- A logarithmic dependence on d, rather bad dependence in $2^m \approx \varepsilon^{-1}$

$\log(d)$: Deterministic constructions based on [UV]

For every (s, p), we need to hit

$$B_m(s,p) := \bigcap_{B \in \Omega_m(s,p)} B = \prod_{\ell=1}^{d} \Big[p_\ell, p_\ell + \frac{s_\ell - 1}{2^m} \Big].$$

$\log(d)$: Deterministic constructions based on [UV]

For every (s,p), we need to hit

$$B_m(s,p) := \bigcap_{B\in\Omega_m(s,p)} B = \prod_{\ell=1}^{d}\Big[p_\ell, p_\ell + \frac{s_\ell - 1}{2^m}\Big].$$

k-**restriction problems:** more flexible notion

- b,k,n,M: positive integers
- $\mathcal{C} = \{C_1,\dots,C_M : C_i \subset \{0,1,\dots,b-1\}^k\}$ - invariant under the permutations of $\{1,\dots,k\}$
- $T = \{x^1,\dots,x^N\} \subset \{0,1,\dots,b-1\}^n$ satisfies the k-restriction problem with respect to $\mathcal{C}$, if for every $S \subset \{1,\dots,n\}$ with $\#S = k$ and for every $j \in \{1,\dots,M\}$, there exists $x^\ell \in T$ with $x^\ell|_S \in C_j$.

- Size of the minimal restriction

$$c = c(\mathcal{C}) := \min_{1 \le j \le M} \#C_j$$

- Size of the minimal restriction

$$c = c(\mathcal{C}) := \min_{1 \le j \le M} \# C_j$$

- For $M = b^k$ and $C_1, \ldots, C_M$ the singleton subsets of $\{0, \ldots, b-1\}^k$, T satisfies the k-restriction problem with respect to $\mathcal{C}$ if it is a (n, k, b)-universal subset.

- Size of the minimal restriction

$$c = c(\mathcal{C}) := \min_{1 \le j \le M} \# C_j$$

- For $M = b^k$ and $C_1, \ldots, C_M$ the singleton subsets of $\{0, \ldots, b-1\}^k$, T satisfies the k-restriction problem with respect to $\mathcal{C}$ if it is a (n, k, b)-universal subset.
- Straightforward random construction:

$$N \ge \frac{b^k}{c} \log(n^k M)$$

- Deterministic construction for the same bound on N, again by M. Naor, L. J. Schulman, and A. Srinivasan

- Size of the minimal restriction

$$c = c(\mathcal{C}) := \min_{1 \le j \le M} \# C_j$$

- For $M = b^k$ and $C_1, \dots, C_M$ the singleton subsets of $\{0, \dots, b-1\}^k$, T satisfies the k-restriction problem with respect to $\mathcal{C}$ if it is a (n, k, b)-universal subset.
- Straightforward random construction:

$$N \ge \frac{b^k}{c} \log(n^k M)$$

- Deterministic construction for the same bound on N, again by M. Naor, L. J. Schulman, and A. Srinivasan
- Just put

$$C_m(s, p) := 2^m [B(s, p) \cap M_m^d]$$

- $(2^m - 1, A_m, d, 2^{2mA_m})$-restriction problem
- $\#P = O(m^2 2^{2m} \log(d))$, but large running time

Splitters:

- n, k, ℓ: positive integers
- An (n, k, ℓ)-splitter H is a family of functions from $\{1, \ldots, n\}$ to $\{1, \ldots, \ell\}$, such that for every $S \subset \{1, \ldots, n\}$ with $\#S = k$ there is $h \in H$, which splits S perfectly.
- It means that the sets $h^{-1}(\{j\}) \cap S$ are of the same size for all $j \in \{1, \ldots, \ell\}$ (or as similar as possible if $\ell \nmid k$).

Splitters:

- n, k, ℓ: positive integers
- An (n, k, ℓ)-splitter H is a family of functions from $\{1, \ldots, n\}$ to $\{1, \ldots, \ell\}$, such that for every $S \subset \{1, \ldots, n\}$ with $\#S = k$ there is $h \in H$, which splits S perfectly.
- It means that the sets $h^{-1}(\{j\}) \cap S$ are of the same size for all $j \in \{1, \ldots, \ell\}$ (or as similar as possible if $\ell \nmid k$).
- Porat and Rothschild (2008): There is an explicit (n, k, k^2)-splitter of size

$$\mathcal{O}\big(k^2 \log(n)\big)$$

that can be constructed in time $\mathcal{O}(k^2 n \log n)$.

- ... obtained from asymptotically good error correcting codes

- $A(m,d)$: an (d, A_m, A_m^2)-splitter
- $T(m) \subset \{1,2,\dots,2^m-1\}^{A_m^2}$: solution of the k-restriction problem with parameters $(2^m-1, A_m, A_m^2, 2^{2mA_m})$
- The solution to the restriction problem with parameters $(2^m-1, A_m, d, 2^{2mA_m})$ and the system $\mathcal{C}$ is given by

$$\begin{aligned} T^* = & \, T(m) \circ A(m,d) \\ := & \{\tau \circ a : \{1,\dots,d\} \to \{0,1,\dots,2^m-2\} : \\ & \tau \in T(m), a \in A(m,d)\} \end{aligned}$$

(concatenations of any splitter with any element of the solution to the restriction problem)

- $$\#P \le C \left(\frac{1+\log_2(\varepsilon^{-1})}{\varepsilon}\right)^4 \log(\max(d, 2/\varepsilon)),$$

$\log(d)$: Deterministic constructions

Used techniques:

- k-wise independent probability spaces
- "smart" greedy exhaustive search
- Splitters (see later)
- Error correcting codes (Wozencraft's ensemble, Justesen codes, expander graphs)

Thank You!