# Factorization Norms and Hereditary Discrepancy

**Jiří Matoušek, Aleksandar Nikolov, Kunal Talwar**
*Presented by Martin Böhm*
*Spring School 2015*

## Discrepancy

**D**(Discrepancy): We work in a universe $U = [n]$, and we are given a list of sets $\mathcal{F}$, where each $f \in \mathcal{F}$ is a subset of $U$. Our goal is to color the universe by two colors so that all sets in $\mathcal{F}$ are as balanced as possible.

Formally, given a coloring $x \in \{-1,1\}^n$, we have $\mathrm{disc}(\mathcal{F}, x) := \max_{F \in \mathcal{F}} |\sum_{i \in F} x_i|$ and our minimization goal $\mathrm{disc}(\mathcal{F}) = \min_{x \in \{-1,1\}^n} \mathrm{disc}(\mathcal{F}, x)$.

**Problem:** Assuming $P \neq NP$ and assuming that $m = O(n)$, we cannot distinguish between $\mathcal{F}$ with discrepancy 0 and $\mathcal{F}$ with discrepancy $\sqrt{n}$.

**D**(Hereditary discrepancy): $\mathrm{herdisc}(\mathcal{F}) = \max_{J \subseteq U} \mathrm{disc}(\mathcal{F}|_J)$.

**D**(Discrepancy for matrices): Given a matrix $A \in \mathbb{R}^{m \times n}$, we define $\mathrm{disc}(A) = \min_{x \in \{-1,1\}^n} \|Ax\|_\infty$. We also define $\mathrm{herdisc}(A) = \max_{J \subseteq [n]} \mathrm{disc}(A_J)$.

## Previous work

**D**(Detlb): A *determinant lower bound* for a matrix $A \in \mathbb{R}^{m \times n}$ is

$$\mathrm{detlb}(A) = \max_k \max_{B \in \mathbb{R}^{k \times k}, B \subseteq A} |\det B|^{1/k}.$$

**T:** $\mathrm{herdisc}(A) \geq (1/2) \cdot \mathrm{detlb}(A)$.

**T:** $\mathrm{herdisc}(A) \leq O(\log(mn)\sqrt{\log n})\mathrm{detlb}(A)$.

**Problem 1:** The function detlb is not a norm.

**Problem 2:** Nobody knows how to compute detlb.

## Norms

**D**(Norm): Given a vector space $V$ say over $\mathbb{C}$, a *norm* is a function $n : V \to \mathbb{R}$ such that the following holds:

1. $n(av) = an(v)$ for a vector $v$ and a scalar $a$,
2. $n(u+v) \leq n(u) + n(v)$ for a pair of vectors $u, v$,
3. if $n(u) = 0$ then $u$ is a zero vector.

Some useful norms:

1. $\|u\|_p = (\sum_i |u_i|^p)^{1/p}$ – an $l_p$ norm.
2. $\|A\|_{p \to q} = \max_{\|x\|_q = 1} \|Ax\|_p$ – $l_p \to l_q$ operator norm.
3. $\|A\|_* = \sum_{i=1}^m \sigma_i$ – the *nuclear norm*, where $\sigma_i$ is a *singular value* of $A$.

**D**(SV decomposition): Let $M \in \mathbb{R}^{m \times n}$. Then there exists a decomposition $M = U \Sigma V^T$, where $U, V$ are orthogonal matrices and $\Sigma$ is a diagonal matrix with non-negative real entries named *singular values*.

## New results

**D**(Gamma-2): We define the $\gamma_2$ function from $A \in \mathbb{R}^{m \times n}$ to $\mathbb{R}$ as follows

$$\gamma_2(A) = \min_{A=BC} \|B\|_{2 \to \infty} \|C\|_{1 \to 2}.$$

**T**(Known.): $\gamma_2(A)$ is a norm.

**T**(Known.): $\gamma_2(A)$ can be computed using a semidefinite program of size polynomial to $A$.

**T**(Main theorem 1): $\mathrm{herdisc}(A) \geq \gamma_2(A)/c \log m$

**T**(Main theorem 2): $\mathrm{herdisc}(A) \leq \gamma_2(A) \cdot c\sqrt{\log m}$

**Note:** Both inequalities are asymptotically tight.

**Other results**: Applications of the previous bounds in data structure lower bounds, new bounds on combinatorial discrepancy of axis-parallel rectangles in $\mathbb{R}^d$, easier proofs of previously-known bounds, and more.

## Some properties of $\gamma_2$

**O:** $\|B\|_{2 \to \infty}$ is equal to $\max_{r \text{ row of } B} \{\|r\|_2\}$.

**O:** $\|C\|_{1 \to 2}$ is equal to $\max_{c \text{ column of } C} \{\|c\|_2\}$.

**T:**

$$\gamma_2(A) = \min_{A=BC} \{ \max_{r \text{ row of } B} \{\|r\|_2\} \cdot \max_{c \text{ column of } C} \{\|c\|_2\}\}.$$

**O:** $\gamma_2(A_{I,J}) \leq \gamma_2(A)$.

**O:** $\gamma_2(A^T) = \gamma_2(A)$.

**T**(Non-trivial): $\gamma_2(A+B) \leq \gamma(A) + \gamma(B)$.

## Graphical interpretation

**D**(Ellipsoid): A zero-centered ellipsoid is defined as $\{x \in \mathbb{R}^n | x^T M x \leq 1\}$ for a positive definite matrix $M$ (with a positive square root $M^{1/2}$).

Alternately, a zero centered ellipsoid is a continuous image of a ball, i.e. $\{Bx; \|x\|_2 \leq r\}$ for a radius $r$ and a matrix $B$.

**D:** For a zero-centered ellipsoid $E$, we define $\|E\|_\infty$ to be the the maximum $\|u\|_\infty$ over all points $u \in E$.

**T:** $\gamma_2(A) = \min\{\|E\|_\infty$ for $E$ ellipsoid, $E$ contains columns of A.$\}$

## A semidefinite program

$$\begin{aligned} \gamma_2(A) &= \min t \\ \forall i \in \{1,2,\ldots,m+n\} &: X_{ii} \leq t \\ \forall i \in \{1,\ldots,m\}, j \in \{1,\ldots,n\} & X_{i,m+j} = a_{ij} \\ \forall i \in \{1,\ldots,m\}, j \in \{1,\ldots,n\} & X_{n+i,j} = a_{ji} \\ X &\succeq 0 \end{aligned}$$

**O:** The variable $t$ becomes tight in at least two rows, one corresponding to $B$ and one to $C$.

From the dual we can get the following characterization:

**T**(Known.): $\gamma_2(A) = \max\{\|P^{1/2} A Q^{1/2}\|_*$ for $P, Q$ diagonal, nonnegative, with $Tr(P) = Tr(Q) = 1\}$.

## Finally, a theorem with proof

**T:** For any $m \times n$ matrix $A$ of rank $r$,

$$\mathrm{detlb}(A) \leq \gamma_2(A) \leq O(\log r)\mathrm{detlb} A.$$

In the proof we use a variant of Binet Cauchy formula:

**T**(Binet-Cauchy): Let $A \in \mathbb{R}^{m \times n}$ be a matrix, and $B \in \mathbb{R}^{m \times n}$ be a matrix. Then we have

$$\det(AB) = \sum_{S \in \binom{[n]}{m}} \det(A_{[m],S}) \det(B_{S,[m]}).$$

**C:** For some choice of columns $J$, $\det(A_J)^2 \geq \frac{1}{\binom{n}{k}} \det(AA^T)$.

**T**(Weighted Binet-Cauchy): Let $A$ be a $k \times n$ matrix, and let $W$ be a nonnegative diagonal unit-trace $n \times n$ matrix. Then there exists a $k$-element set $J \subseteq [n]$ such that

$$|det A_J|^{1/k} \geq \sqrt{k/e} \cdot |\det(AWA^T)|^{1/2k}.$$