THEORY OF COMPUTING, Volume 15 (8), 2019, pp. 1–7
*www.theoryofcomputing.org*

NOTE

# Matrix Rigidity and the Croot-Lev-Pach Lemma

Zeev Dvir[*] Benjamin L. Edelman

*Received September 19, 2017; Revised October 2, 2018; Published October 15, 2019*

**Abstract:** Matrix rigidity is a notion put forth by Valiant (1977) as a means for proving arithmetic circuit lower bounds. A matrix is rigid if it is far, in Hamming distance, from any low-rank matrix. Despite decades of effort, no explicit matrix rigid enough to carry out Valiant's plan has been found. Recently, Alman and Williams (STOC'17) showed that, contrary to common belief, the Walsh–Hadamard matrices cannot be used for Valiant's program as they are not sufficiently rigid.

Our main result is a similar non-rigidity theorem for *any* $q^n \times q^n$ matrix $M$ of the form $M(x,y) = f(x+y)$, where $f : \mathbb{F}_q^n \to \mathbb{F}_q$ is any function and $\mathbb{F}_q$ is a fixed finite field of $q$ elements ($n$ goes to infinity). The theorem follows almost immediately from a recent lemma of Croot, Lev and Pach (2017) which is also the main ingredient in the recent solution of the famous cap-set problem by Ellenberg and Gijswijt (2017).

[*]Supported by NSF CAREER award DMS-1451191 and NSF grant CCF-1523816.

**ACM Classification:** F.2.2, F.1.3

**AMS Classification:** 68Q17, 68Q15

**Key words and phrases:** complexity theory, complexity, combinatorics, additive combinatorics, algebraic complexity, circuit complexity, arithmetic circuits, lower bounds, rank, polynomials, matrix rigidity, polynomial method, hamming distance


DOI: 10.4086/toc.2019.v015a008

# 1 Introduction

We begin by defining the notion of matrix rigidity—a property of matrices that combines combinatorial conditions (Hamming distance) with algebraic ones (matrix rank). Recall that the Hamming distance between two vectors $x, y \in \Sigma^n$ over some alphabet $\Sigma$ is equal to the number of entries $i \in [n]$ for which $x_i \neq y_i$.

**Definition 1.1** (Matrix rigidity)**.** The rank-$r$ rigidity of a matrix $M$ over a field $\mathbb{F}$, denoted $\mathcal{R}_M^{\mathbb{F}}(r)$, is defined as the minimum Hamming distance between $M$ and any matrix of rank at most $r$. In other words, $\mathcal{R}_M^{\mathbb{F}}(r)$ is equal to the smallest number of entries in $M$ that one needs to change in order to reduce the rank of $M$ to $r$.

Specifying the field is important since there are $(0,1)$-matrices that have high rank over the integers but low rank over, say, $\mathbb{F}_2$.

The notion of matrix rigidity was introduced by Valiant [10] in the context of studying the arithmetic circuit complexity of linear transformations. A *linear circuit* is a model of computation in which the inputs represent the basic linear function $x_1, \ldots, x_n$ and each gate takes two previously computed linear forms and outputs some linear combination of them with coefficients in the field. We measure the size of a linear circuit by counting the wires, and the depth by the longest path from input to output. A linear circuit with $n$ inputs and $n$ outputs computes a linear map $T : \mathbb{F}^n \to \mathbb{F}^n$ and many important linear maps (e. g., Fourier transform) can be computed efficiently in this model. One can even show that any use of multiplication gates can be eliminated (with negligible cost) when computing a linear map over $\mathbb{C}$ [7].

One of the most important problems in theoretical computer science is to prove unconditional complexity lower bounds for realistic models of computation. Despite decades of attempts, we are still unable to prove super-linear circuit lower bounds (in any realistic model) for logarithmic depth circuits.[1] In an early attempt to bridge this gap Valiant [10] proved the following theorem.

**Theorem 1.2** (Valiant)**.** *Let $\{M_N\}_{N\in\mathbb{N}}$ be a family of matrices with $M_N$ being of dimensions $N \times N$ over a field $\mathbb{F}$. If*

$$\mathcal{R}_{M_N}^{\mathbb{F}}(N/\log\log N) \geq \Omega(N^{1+\varepsilon})$$

*for some $\varepsilon > 0$ then $M_N$ cannot be computed by linear circuits of size $O(N)$ and depth $O(\log(N))$ (asymptotically, as $N$ grows).*

We say that a matrix is *Valiant-rigid* if it satisfies the rigidity parameters in the above theorem. It is straightforward to check that for any matrix $M$ and field $\mathbb{F}$, $\mathcal{R}_M^{\mathbb{F}}(r) \leq (N-r)^2$ for any $r$. Valiant proved that almost all matrices achieve this maximum rigidity: for almost all matrices $M$, $\mathcal{R}_M^{\mathbb{F}}(r) = (N-r)^2$ if $\mathbb{F}$ is infinite and $\mathcal{R}_M^{\mathbb{F}}(r) = \Omega((N-r)^2/\log N)$ if $\mathbb{F}$ is finite. However, since Valiant's original paper, it remains an open problem to find an explicit Valiant-rigid matrix. By "explicit" we mean a matrix that can be produced in polynomial in $N$ time by a Turing machine given $N$ as input.

The current best rigidity lower bound for any explicit matrix is $\mathcal{R}_M^F(r) = \Omega((N^2/r)\log(N/r))$ [5, 8]. Until recently, the $2^n \times 2^n$ Walsh–Hadamard matrix $H = ((-1)^{\langle x,y\rangle})_{x,y\in\{0,1\}^n}$ was conjectured to be Valiant-rigid over the rational numbers [7]. A recent surprising result of Alman and Williams [1] showed

[1] One exception being arithmetic circuits computing polynomials of high degree, e. g., [9].

that in fact the Hadamard matrix is not sufficiently rigid. Denoting $N = 2^n$, they showed that for every $\varepsilon > 0$ there exists $\varepsilon' > \Omega(\varepsilon^2/\log(1/\varepsilon))$ such that $\mathcal{R}_H^{\mathbb{Q}}(N^{1-\varepsilon'}) \leq N^{1+\varepsilon}$.

The purpose of this note is to observe another "non-rigidity" phenomenon for a related (large) family of matrices. The hope is that by understanding the reasons for this non-rigidity we can perhaps get closer to proving stronger rigidity results. Our main result is the following.

**Theorem 1.3.** *Let $\mathbb{F}_q$ be any finite field and let $f : \mathbb{F}_q^n \to \mathbb{F}_q$ be any function. Let $M$ be the $q^n \times q^n$ matrix defined by $M_{x,y} = f(x+y)$ for $x, y \in \mathbb{F}_q^n$. Denoting $N = q^n$ we have that for any $\varepsilon > 0$, there exists*

$$\varepsilon' = \frac{\log q}{4(q-1)^2} \cdot \frac{\varepsilon^2}{(\log \varepsilon^{-1})^2}(1 - o(1))$$

*such that $\mathcal{R}_M^{\mathbb{F}_q}(N^{1-\varepsilon'}) \leq N^{1+\varepsilon}$. The asymptotic notation is for fixed $q$ as $\varepsilon \to 0$.*

One should note that, unlike Hadamard matrices, these matrices are over a finite field and not over the rational numbers. Having non-rigid matrices over a finite field is a bit less surprising since there are more "ways" for the rank to be low. It is an interesting open problem to decide if Theorem 1.3 still holds if one is allowed to take a function $f : \mathbb{F}_q^n \to \mathbb{F}$ where $\mathbb{F}$ is the rational numbers (or even the complex numbers). A positive answer will imply the results of [1] since the Walsh–Hadamard matrix can be written over the complex numbers as

$$(-1)^{\langle x,y\rangle} = (-1)^{|x|/2}(-1)^{|y|/2}(-1)^{|x\oplus y|/2},$$

where $|\cdot|$ represents the Hamming weight.

Another interesting question is that of replacing the group $\mathbb{F}_q^n$ indexing the rows/columns with other groups. Subsequent to this paper being circulated, Dvir and Liu [3] proved that $M_{x,y} = f(x+y)$ is not rigid for any $f : G \to \mathbb{C}$ where $G$ is an abelian group. In particular, complex-valued Toeplitz matrices and Fourier matrices are not Valiant-rigid.

## 1.1 The Croot-Lev-Pach (CLP) lemma

A *cap set* is a subset of $\mathbb{F}_q^n$ with no non-trivial three-term arithmetic progressions. We think of $q > 2$ as fixed and $n$ going to infinity. The cap-set problem asks how the size of the largest possible cap set (denoted $r(n)$) grows in terms of $n$. It was an open question whether $r(n) \leq c^n$ for some $c < q$. Croot, Lev, and Pach [2] used a variant of the polynomial method to solve the corresponding problem for $\mathbb{Z}_4^n$ (the ring mod 4) in the affirmative, proving a bound of $c^n$ for some $c < 4$, and soon afterwards Ellenberg and Gijswijt [4] adapted the CLP result to provide a positive answer to the cap set problem in $\mathbb{F}_q$ for all $q > 2$. At the core of [2] is a lemma saying that, if $P : \mathbb{F}_q^n \to \mathbb{F}_q$ is a polynomial of not too high degree, then the $q^n \times q^n$ matrix $M = (P(x+y))_{x,y\in\mathbb{F}_q^n}$ has very low rank (see below for the exact parameters). We observe that, since *any* function can be well approximated by such a polynomial, the matrix $f(x+y)$ can be changed in a small number of entries to give the low-rank matrix $P(x+y)$. This is the crux of the proof of our main theorem. We give the full details of the proof in the next section.

## 2 Proof of Theorem 1.3

Let $\mathcal{F}(q,n)$ denote the set of functions $f : \mathbb{F}_q^n \to \mathbb{F}_q$. Then, $\mathcal{F}(q,n)$ is an $\mathbb{F}_q$-vector space of dimension $q^n$. A basis for this vector space is given by the set of $q^n$ monomials

$$\mathcal{M}(q,n) = \{x_1^{a_1} \cdots x_n^{a_n} \mid 0 \le a_i \le q-1\}.$$

Let us denote by $\mathcal{M}_d(q,n)$ the set of monomials in $\mathcal{M}(q,n)$ of total degree at most $d$ and by $\mathcal{F}_d(q,n)$ the set of polynomials of degree at most $d$. Note that $\mathcal{M}_d(q,n)$ is a basis of $\mathcal{F}_d(q,n)$. Let $m_d(q,n)$ denote the size of $\mathcal{M}_d(q,n)$ or equivalently the dimension of $\mathcal{F}_d(q,n)$.

We start by stating the precise form of the CLP lemma [2]. For completeness we include a short sketch of the proof.

**Lemma 2.1** (CLP). *Let $P \in \mathcal{F}_d(q,n)$ and let $M$ denote the $q^n \times q^n$ matrix with entries $M_{x,y} = P(x+y)$ for $x,y \in \mathbb{F}_q^n$. Then* $\operatorname{rank}(M) \le 2 \cdot m_{\lfloor d/2 \rfloor}(q,n)$.

*Proof sketch.* To prove the claim we will show that $P(x+y) = \sum_{i=1}^{R} f_i(x) g_i(y)$ with $R \le 2 \cdot m_{\lfloor d/2 \rfloor}(q,n)$. To see how to do this observe that, for each monomial $m(x) = x_1^{a_1} \cdots x_n^{a_n}$ of degree at most $d$, the terms in the expression $m(x+y)$ all have degree $\le \lfloor d/2 \rfloor$ in either $x$ or $y$. Writing $P$ as a sum of monomials and grouping together terms with the same low-degree parts (in $x$ first and then in $y$) gives the desired decomposition. □

The main power of the CLP lemma comes from the following quantitative observation. For a fixed $q$ and sufficiently large $n$, the numbers $m_d(q,n)$ behave approximately like the CDF of a normal distribution when we increase $d$ from 0 to $(q-1)n$ (the largest possible degree). Most of the mass will be concentrated around the middle, $(q-1)n/2$, with the tails decaying exponentially fast. We will use the following (weak) estimate.

**Claim 2.2.** *For any prime power $q$ and any $\varepsilon > 0$ there exists*

$$\delta = \frac{\log q}{q-1} \cdot \frac{\varepsilon}{\log \varepsilon^{-1}} (1 - o(1))$$

*such that, for sufficiently large $n$, we have*

$$m_{(1-\delta)(q-1)n}(q,n) \ge q^n - q^{\varepsilon n}.$$

*Proof.* By symmetry it is enough to bound $m_{\delta(q-1)n}(q,n) \le q^{\varepsilon n}$. We reduce this problem to the binary alphabet case. We claim that $m_d(q,n) \le m_d(2, n(q-1))$ for all $d$. To see this, consider the injective mapping from $\mathcal{M}_d(q,n)$ into $\mathcal{M}_d(2, n(q-1))$ sending $x_i^{a_i}$ to the multilinear monomial $x_{i1} x_{i2} \cdots x_{ia_i}$. For the binary case we can use the standard tail bounds for the Binomial distribution to get that

$$m_{\delta(q-1)n}(2, n(q-1)) \le 2^{H(\delta)(q-1)n}$$

where $H(\delta)$ is the binary entropy function. Hence, we need $\delta$ to satisfy $2^{H(\delta)(q-1)} \le q^{\varepsilon}$. For small $x$, $H(\delta) \le x$ is achieved by some

$$\delta = \frac{x}{\log x^{-1}} (1 - o(1)).$$

Plugging in
$$x = \frac{\log q}{q-1}\varepsilon, \qquad \text{we obtain} \qquad \delta = \frac{\log q}{q-1} \cdot \frac{\varepsilon}{\log \varepsilon^{-1}}(1-o(1)). \qquad \square$$

The following claim and corollary show that any function can be approximated well by a polynomial of sufficiently high degree.

**Claim 2.3.** *Let $W$ be a subspace of finite codimension $r$ in the vector space $V$. Let $\mathcal{B}$ be a basis for $V$. Then, for any vector $v \in V$, we can modify $\leq r$ of the coordinates of $v$ (in the basis $\mathcal{B}$) to produce a vector that lies in $W$.*

*Proof.* By the Steinitz exchange lemma, one can select $r$ vectors, $b_1, \ldots, b_r \in \mathcal{B}$, such that $V = W \oplus U$ where $U$ is the span of the $b_i$. Now write $v$ as $v = w + u$ where $w \in W$ and $u \in U$. The $\mathcal{B}$-coordinates of $v$ and $w$ differ only at the $b_i$. $\square$

**Corollary 2.4.** *Let $f \in \mathcal{F}(q,n)$ be any function. Then, for all $d \leq n$, there exists a polynomial $P \in \mathcal{F}_d(q,n)$ such that*
$$|\{x \in \mathbb{F}_q^n \mid f(x) \neq P(x)\}| \leq q^n - m_d(q,n).$$

*Proof.* This follows from the previous claim and from the fact that $\dim(\mathcal{F}_d(q,n)) = m_d(q,n)$. $\square$

We are now ready to prove our main result.

*Proof of Theorem 1.3.* Let $f : \mathbb{F}_q \to \mathbb{F}_q$ be as in the theorem and let $\varepsilon > 0$. Using Claim 2.2 and Corollary 2.4, we can find $\delta > 0$ and a polynomial $P$ of degree at most $d = (1-\delta)(q-1)n$ such that $P$ agrees with $f$ on all but $q^{\varepsilon n} = N^{\varepsilon}$ values in $x \in \mathbb{F}_q^n$. Let $M$ denote the $q^n \times q^n$ matrix with entries $M_{x,y} = f(x+y)$ and let $L$ denote the matrix of the same dimensions with entries $L_{x,y} = P(x+y)$. Then, $M$ and $L$ differ in at most $N^{\varepsilon}$ entries in each row and in at most $N^{1+\varepsilon}$ entries altogether. Now, by Lemma 2.1 (the CLP lemma) we have that $\mathrm{rank}(L) \leq m_{\lfloor d/2 \rfloor}(q,n)$. But $d/2 = (1/2 - \delta/2)(q-1)n$ and so, by Hoeffding's inequality [6] (on the probability that a random monomial will have degree at most $d/2$), we have
$$m_{\lfloor d/2 \rfloor}(q,n) \leq e^{-\frac{\delta^2 n}{4}} q^n = N^{1-\delta^2/(4\log q)}.$$

This is at most $N^{1-\varepsilon'}$ as long as $\varepsilon' \leq \frac{\delta^2}{4\log q}$. Plugging in
$$\delta = \frac{\log q}{q-1} \cdot \frac{\varepsilon}{\log \varepsilon^{-1}}(1-o(1)),$$
we obtain
$$\varepsilon' = \frac{\log q}{4(q-1)^2} \cdot \frac{\varepsilon^2}{(\log \varepsilon^{-1})^2}(1-o(1)).$$
This concludes the proof. $\square$

## References


[1] JOSH ALMAN AND RYAN WILLIAMS: Probabilistic rank and matrix rigidity. In *Proc. 49th STOC*, pp. 641–652. ACM Press, 2017. [doi:10.1145/3055399.3055484, arXiv:1611.05558] 2, 3

[2] ERNIE CROOT, VSEVOLOD F. LEV, AND PÉTER PÁL PACH: Progression-free sets in $\mathbb{Z}_4^n$ are exponentially small. *Ann. of Math*, 185(1):331–337, 2017. [doi:10.4007/annals.2017.185.1.7, arXiv:1605.01506] 3, 4

[3] ZEEV DVIR AND ALLEN LIU: Fourier and circulant matrices are not rigid. In *Proc. 34th Conf. Comput. Complexity (CCC'19)*, pp. 17:1–17:23. Leibniz-Zentrum fuer Informatik, 2019. [doi:10.4230/LIPIcs.CCC.2019.17] 3

[4] JORDAN S. ELLENBERG AND DION GIJSWIJT: On large subsets of $\mathbb{F}_q^n$ with no three-term arithmetic progression. *Ann. of Math*, 185(1):339–343, 2017. [doi:10.4007/annals.2017.185.1.8, arXiv:1605.09223] 3

[5] JOEL FRIEDMAN: A note on matrix rigidity. *Combinatorica*, 13(2):235–239, 1993. [doi:10.1007/BF01303207] 2

[6] WASSILY HOEFFDING: Probability inequalities for sums of bounded random variables. *J. Amer. Statistical Assoc.*, 58(301):13–30, 1963. [doi:10.1080/01621459.1963.10500830] 5

[7] SATYANARAYANA V. LOKAM: Complexity lower bounds using linear algebra. *Found. Trends Theor. Comput. Sci.*, 4(1–2):1–155, 2009. [doi:10.1561/0400000011] 2

[8] M. AMIN SHOKROLLAHI, DANIEL A. SPIELMAN, AND VOLKER STEMANN: A remark on matrix rigidity. *Inform. Process. Lett.*, 64(6):283–285, 1997. [doi:10.1016/S0020-0190(97)00190-7] 2

[9] VICTOR SHOUP AND ROMAN SMOLENSKY: Lower bounds for polynomial evaluation and interpolation problems. *Comput. Complexity*, 6(4):301–311, 1996. Preliminary version in FOCS'91. [doi:10.1007/BF01270384] 2

[10] LESLIE G. VALIANT: Graph-theoretic arguments in low-level complexity. In *Proc. 6th Internat. Symp. Math. Found. Comput. Sci. (MFCS'77)*, pp. 162–176. Springer, 1977. [doi:10.1007/3-540-08353-7_135] 2

## AUTHORS

Zeev Dvir
Associate professor
Department of Mathematics and
Department of Computer Science
Princeton University, Princeton, NJ
zdvir@princeton.edu
http://www.cs.princeton.edu/~zdvir

Benjamin L. Edelman
Ph. D. student
Department of Computer Science
Harvard University, Cambridge, MA
bedelman@g.harvard.edu
http://benjaminedelman.com

## ABOUT THE AUTHORS

ZEEV DVIR was born in Jerusalem, Israel. He received his Ph. D. from the Weizmann Institute in Israel in 2008. His advisors were Ran Raz and Amir Shpilka. He has a broad interest in theoretical computer science and mathematics and especially in computational complexity, pseudorandomness, coding theory and discrete mathematics.

BENJAMIN L. EDELMAN is a Ph. D. student in computer science at Harvard University. He is advised by Leslie Valiant. His research interests include computational complexity, learning theory, and game theory. While an undergraduate at Princeton University, he did algebraic complexity research with Zeev Dvir and Ran Raz. His interest in theoretical computer science can be traced to PACT, a high school summer program run by the extraordinary Rajiv Gandhi.