Proof of the GM-MDS conjecture

Shachar Lovett
UC San Diego

Analytic Techniques in Theoretical Computer Science

Proof of the GM-MDS conjecture

Shachar Lovett
UC San Diego

Algebraic
~~Analytic~~ Techniques in Theoretical Computer Science

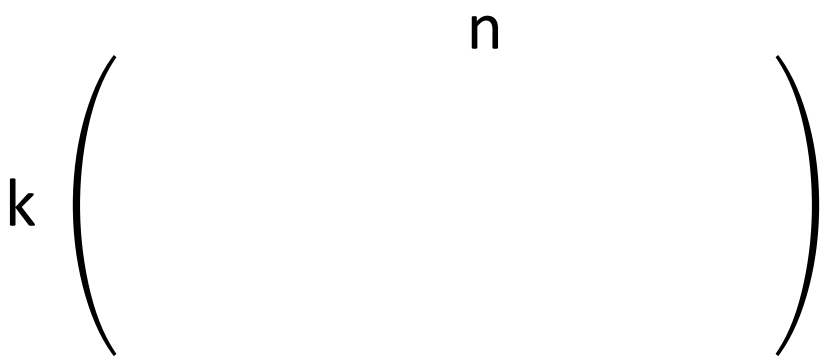
Overview

- MDS matrices
- MDS matrices with specific zeros
- GM-MDS conjecture
- Algebraic GM-MDS conjecture
- Proof (very briefly)
- General family of problems

MDS matrices

MDS matrices

- A $k \times n$ matrix is an MDS matrix if any k columns are linearly independent



- The name comes from coding theory, as their rows generate MDS (Maximum Distance Separable) codes
- Arise in many other contexts, for example k-wise independence

Construction of MDS matrices

- Standard construction: Vandermonde matrix (aka Reed-Solomon code)

- Let $a_1, \dots, a_n$ be distinct field elements

$$V = \begin{pmatrix} 1 & 1 & \cdots & 1 \\ a_1 & a_2 & \cdots & a_n \\ \vdots & \vdots & & \vdots \\ a_1^{k-1} & a_2^{k-1} & \cdots & a_n^{k-1} \end{pmatrix}$$

- Requires field of size $|\mathbb{F}| \geq n$
- Known: if $n \geq k + 2$ then any $k \times n$ MDS matrix requires $|\mathbb{F}| \geq \mathrm{n}/2$

(closing this gap is the "MDS conjecture")

MDS matrices with zeros

MDS matrices with zeros

- Goal: MDS matrices with a specific zero pattern
- Question: What are necessary / sufficient conditions on the locations of zeros?

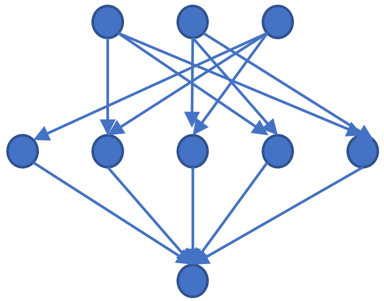
- For example, can the following matrix be completed to an MDS matrix?

$$\begin{pmatrix} 0 & * & 0 & * & * \\ 0 & 0 & * & * & * \\ * & * & * & 0 & 0 \end{pmatrix}$$

- Motivation: coding theory
  - Multiple access networks
  - Secure data exchange
  - Distributed Reed-Solomon codes

Application: distributed Reed-Solomon codes [Halbawi-Yao-Duursma 2014]

- 3 sources send information to single receiver via 5 relay nodes



- Goal: Code which protects against 2 malicious relay nodes
- Solution: MDS matrix with the following zero pattern

$$\begin{pmatrix} 0 & * & 0 & * & * \\ 0 & 0 & * & * & * \\ * & * & * & 0 & 0 \end{pmatrix}$$

Matrix completion problem

$$\begin{pmatrix} 0 & * & 0 & * & * \\ 0 & 0 & * & * & * \\ * & * & * & 0 & 0 \end{pmatrix}$$

- Goal: replace $*$ with field elements so that any k columns are linearly independent
- Equivalently: all $k \times k$ minor should be nonsingular

Necessary condition

- Consider the zero locations in a $k \times n$ MDS matrix

$$\begin{pmatrix} 0 & * & 0 & * & * \\ 0 & 0 & * & * & * \\ * & * & * & 0 & 0 \end{pmatrix}$$

- Any row can have $\leq k - 1$ zeros
- Any 2 rows can have $\leq k - 2$ common zeros
- Any 3 rows can have $\leq k - 3$ common zeros

…

- **Rectangle condition / MDS condition**:
  there are no $a \times b$ combinatorial rectangles of zeros with $a + b > k$

Necessary condition

- **Rectangle condition**:
  there are no $a \times b$ combinatorial rectangles of zeros with $a + b > k$

- Satisfied in this example (k=3, n=5)

$$\begin{pmatrix} 0 & * & 0 & * & * \\ 0 & 0 & * & * & * \\ * & * & * & 0 & 0 \end{pmatrix}$$

Necessary condition

- **Rectangle condition**:
  there are no $a \times b$ combinatorial rectangles of zeros with $a + b > k$

- Satisfied in this example (k=3, n=5)

a=1, b=2

$$\begin{pmatrix} 0 & * & 0 & * & * \\ \boxed{0 \quad 0} & & * & * & * \\ * & * & * & 0 & 0 \end{pmatrix}$$

Necessary condition

- **Rectangle condition**:
  there are no $a \times b$ combinatorial rectangles of zeros with $a + b > k$

- Satisfied in this example (k=3, n=5)

a=2, b=1

$$\begin{pmatrix} 0 & * & 0 & * & * \\ 0 & 0 & * & * & * \\ * & * & * & 0 & 0 \end{pmatrix}$$

Sufficient condition

- The rectangle condition is also sufficient, over large enough fields
- If we replace $*$ with variables, then the determinant of any $k \times k$ minor is not identically zero

$$\begin{pmatrix} 0 & * & 0 & * & * \\ 0 & 0 & * & * & * \\ * & * & * & 0 & 0 \end{pmatrix} \to \begin{pmatrix} 0 & x_1 & 0 & x_2 & x_3 \\ 0 & 0 & x_4 & x_5 & x_6 \\ x_7 & x_8 & x_9 & 0 & 0 \end{pmatrix}$$

Sufficient condition

$$\begin{pmatrix} 0 & x_1 & 0 & x_2 & x_3 \\ 0 & 0 & x_4 & x_5 & x_6 \\ x_7 & x_8 & x_9 & 0 & 0 \end{pmatrix}$$

- All $k \times k$ determinants are not identically zero

- Next step: replace variables by field elements
- Consider polynomial which is the product of all $k \times k$ determinants
- Apply Schwartz-Zippel lemma

- Problem: this requires huge field size
- As there are $\binom{n}{k}$ minors, individual degrees are $\binom{n}{k}$, so this requires $|\mathbb{F}| \geq \binom{n}{k}$

Summary so far

- Goal: MDS matrices with specific zeros
- **Rectangle condition**:
  there are no $a \times b$ combinatorial rectangles of zeros with $a + b > k$
- Necessary condition over any field
- Sufficient condition, but only over very large fields: $|\mathbb{F}| \geq \binom{n}{k}$
- Question: can we **decrease the field size**?

Why should we hope for small field size?

- Consider the problem of constructing $k \times n$ MDS matrices (without any zero constraints)
- Probabilistic construction still requires field $|\mathbb{F}| \geq \binom{n}{k}$
- Algebraic construction exists in any field with $|\mathbb{F}| \geq n$
- Question: can we hope for an algebraic construction even with the zero constraints?

The GM-MDS conjecture

GM-MDS conjecture

- **GM-MDS Conjecture (**[Dau-Song-Yuen ’14]):
  The rectangle condition is **sufficient** over fields of size $|\mathbb{F}| \geq n + k - 1$

- Recall: naïve construction requires $|\mathbb{F}| \geq \binom{n}{k}$, so this is a huge improvement.

- This was not a shot in the dark; Dau et al. gave an algebraic conjecture which implies the GM-MDS conjecture with these bounds

- We prove this algebraic conjecture, and hence the GM-MDS conjecture

Algebraic approach

- Recall: standard construction of MDS matrices is “algebraic”, eg Vandermonde. We can also allow for a change of basis on the rows. This gives a family of “algebraic” constructions.

- Any $k \times n$ matrix $M = TV$ is a MDS matrix, where:
  - T is $k \times k$ full rank matrix
  - V is $k \times n$ Vandermonde matrix

$$M = \begin{pmatrix} \quad T \quad \end{pmatrix}\begin{pmatrix} \qquad V \qquad \end{pmatrix}$$

- **Algebraic GM-MDS conjecture (informal version 1):** The GM-MDS conjecture can be solved by such “algebraic” constructions

Algebraic approach

- Let M=TV with: T= $\begin{pmatrix} T_{1,1} & T_{1,2} & \cdots & T_{1,k} \\ T_{2,1} & T_{2,2} & \cdots & T_{2,k} \\ \vdots & \vdots & & \vdots \\ T_{k,1} & T_{k,2} & \cdots & T_{k,k} \end{pmatrix}$, V= $\begin{pmatrix} 1 & 1 & \cdots & 1 \\ a_1 & a_2 & \cdots & a_n \\ \vdots & \vdots & & \vdots \\ a_1^{k-1} & a_2^{k-1} & \cdots & a_n^{k-1} \end{pmatrix}$

- View rows of T as coefficients of k univariate polynomials of degree $\leq k-1$:

$$f_i(x) = T_{i,1} + T_{i,2} \cdot x + T_{i,3} \cdot x^2 + \cdots + T_{i,k} \cdot x^{k-1}, \qquad i = 1 \ldots k$$

- The entries of M=TV are then given by: $M_{i,j} = f_i(a_j)$

Algebraic approach

- We want:
  - k univariate polynomials $f_1, \dots, f_k$ of degrees $\leq k-1$ (matrix T)
  - n distinct field elements $a_1, \dots, a_n$ (matrix V)

  Such that:

(1) The polynomials are linearly independent ($\equiv$ T is full rank)

(2) If we need $M_{i,j} = 0$ then $f_i(a_j) = 0$

- **Algebraic GM-MDS conjecture (informal version 2):** under the rectangle condition on the locations of zeros, this is possible

Algebraic approach

- Assume wlog that we have exactly k-1 zeros in each row
- In this case, polynomials $f_1, \dots, f_k$ are <u>uniquely defined</u> by their zeros

- Let $S_i = \{j \in [n]: M_{i,j} = 0\}$ denote the locations of zeros in the i-th row
- We require that $f_i(a_j) = 0$ for $j \in S_i$
- f_i is a polynomial of degree $\leq k-1$, and $|S_i| = k-1$
- So we must have:

$$f_i(x) = \prod_{j \in S_i} (x - a_j)$$

The algebraic GM-MDS conjecture

- Let $S_1, \ldots, S_k \subset [n]$ be the required zero locations, where $|S_i| = k - 1$
- Let $a_1, \ldots, a_n$ be formal variables over $\mathbb{F}$
- Define $f_i(x) = \prod_{j \in S_i} (x - a_j)$

- **Algebraic GM-MDS conjecture ([Dau-Song-Yuen '14]):** if $S_1, \ldots, S_k$ satisfy the rectangle condition, then $f_1, \ldots, f_k$ are linearly independent over $\mathbb{F}(a_1, \ldots, a_n)$

- Interpretation:
  - If the rectangle condition is false, then $f_1, \ldots, f_k$ are linearly dependent
  - Conjecture: this is the only case (if $a_1, \ldots, a_n$ are "generic")

Why field size $n + k - 1$?

- **Algebraic GM-MDS conjecture:**
  if $S_1, \dots, S_k$ satisfy the rectangle condition, then $f_1, \dots, f_k$ are linearly independent over $\mathbb{F}(a_1, \dots, a_n)$
- Assume the conclusion holds. We need to replace $a_1, \dots, a_n$ with distinct field elements such that $f_1, \dots, f_k$ remain linearly independent
- Express as a nonzero polynomial in $a_1, \dots, a_n$ with individual degrees $n + k - 2$:
- $f_1, \dots, f_k$ remain linearly independent → individual degree $k - 1$
- distinct field elements → individual degree $n - 1$
- By Schwartz-Zippel, has solution exists whenever $|\mathbb{F}| \geq n + k - 1$

Proof of algebraic GM-MDS conjecture

Proof of algebraic GM-MDS conjecture (very briefly)

- Proof is by an induction on the structure of zeros
- Requires a generalized conjecture, which allows for multiple zeros at a special point
- Several previous works proved special cases of the conjecture
- Hassibi-Yildiz proved it independently at about the same time

General family of problems

General matrix completion problem

- Consider a matrix with $0/*$ entries without any assumptions

$$\begin{pmatrix} 0 & * & 0 & * & * \\ 0 & 0 & * & * & * \\ * & * & 0 & 0 & 0 \end{pmatrix}$$

- Goal: replace $*$ with field elements, so that every minor that can be nonsingular will be
- Solution over large fields $|\mathbb{F}| \geq \binom{n}{k}$ always possible

- Question: are large fields necessary?

Known lower bounds

- Question arises in Maximally Recoverable (MR) codes, where the $0/*$ pattern depends on the code topology
- Meta conjecture: for any $0/*$ pattern, either there are algebraic constructions, or exponential field size is needed
- GM-MDS conjecture: family of patterns where algebraic constructions exist
- Exponential lower bounds on field size are known in two specific topologies [Kane-L-Rao '17, Gopi-Guruswami-Yekhanin '17]
- However, proofs are ad-hoc to the specific topology being studied

Open problem

- Let M be a random $k \times n$ matrix with $0/*$ entries, where

$$\Pr[M_{i,j} = 0] = \Pr[M_{i,j} = *] = \frac{1}{2}$$

- Goal: replace $*$ with field elements, so that every minor that can be nonsingular will be
- Intuition: random pattern should disallow algebraic solutions
- Conjecture: w.h.p an exponential field size is needed: $|\mathbb{F}| \geq \binom{n}{k}^{\Omega(1)}$
- We currently have no proof techniques to show anything like that

Summary

- GM-MDS conjecture: MDS matrices with zero pattern that satisfies the rectangle condition exist over small fields
- Construction is algebraic: change of basis to a Vandermonde matrix
- More general problems (arising in MR codes) are wide open
- General phenomena: when algebraic constructions fail, sometime combinatorial / probabilistic constructions have much worse parameters
- Examples: local codes, Zarankiewicz problem, high dimensional expenders

Thank you