An extremal problem for integer sparse recovery

Sergei Konyagin∗ Benny Sudakov†

Abstract

Motivated by the problem of integer sparse recovery we study the following question. Let $A$ be an $m \times d$ integer matrix whose entries are in absolute value at most $k$. How large can be $d = d(m, k)$ if all $m \times m$ submatrices of $A$ are non-degenerate? We obtain new upper and lower bounds on $d$ and answer a special case of the problem by Brass, Moser and Pach on covering $m$-dimensional $k \times \cdots \times k$ grid by linear subspaces.

1 Introduction

Compressed sensing is a relatively new mathematical paradigm that shows a small number of linear measurements are enough to efficiently reconstruct a large dimensional signal under the assumption that the signal is sparse (see, e.g., [4] and its references). That is, given a signal $x \in \mathbb{R}^d$, the goal is to accurately reconstruct $x$ from its noisy measurements $b = Ax + e$. Here, $A$ is an underdetermined matrix $A \in \mathbb{R}^{m \times d}$, where $m$ is much smaller than $d$, and $e \in \mathbb{R}^m$ is a vector modeling noise in the system. Since the system is highly underdetermined, it is ill-posed until one imposes additional constraints, such as the signal $x$ obeying a sparsity constrain. We say $x$ is $s$-sparse when it has at most $s$ nonzero entries. Clearly, any matrix $A$ that is one-to-one on 2s-sparse signals will allow reconstruction in the noiseless case when $e = 0$. However, compressed sensing seeks the ability to reconstruct efficiently and robustly even when one allows presence of noise. Motivated by this problem Fukshansky, Needell and Sudakov [5] considered the following extremal problem, which is of independent interest.

Problem 1.1. Given an integers $k, m$ what is the maximum integer $d$ such that there exists $m \times d$ matrix $A$ with integer entries satisfying $|a_{ij}| \leq k$ such that all $m \times m$ submatrices of $A$ are non-degenerate.

To see the connection of this question with integer sparse recovery let $s \leq m/2$ and consider $s$-sparse signal $x \in \mathbb{Z}^d$. We denote by $\|b\|$ the Euclidean norm of a vector $b = (b_1, \ldots, b_m) \in \mathbb{R}^m$ and by $\|b\|_\infty$ its $l_\infty$-norm: $\|b\|_\infty = \max_{i=1,\ldots,m} |b_i|$. Suppose we wish to decode $x$ from the noisy measurements $b = Ax + e$ where $\|e\|_\infty < \frac{1}{2}$ (in particular, this holds if $\|e\| < \frac{1}{2}$). Note that by definition of matrix $A$ we have that for any $m$-sparse vector $z$, $Az \neq 0$ and therefore being integer vector has $l_\infty$-norm at least one. So to decode $x$ we can select the $s$-sparse signal $y \in \mathbb{Z}^d$ minimizing

∗Laboratory “High-dimensional approximation and applications”, Department of Mechanics and Mathematics, Moscow State University, Moscow, Russia; Steklov Mathematical Institute of Russian Academy of Sciences, Moscow, Russia. Email: konyagin@mi-ras.ru. Research supported by the grant of the Government of the Russian Federation (project 14.W03.31.0031).

†Department of Mathematics, ETH Zurich, Switzerland. Email: benjamin.sudakov@math.ethz.ch. Research supported in part by SNSF grant 200021-175573.
\[ \| \mathbf{b} - \mathbf{A} \mathbf{y} \|. \] Then since \( \mathbf{x} \) satisfies \( \| \mathbf{b} - \mathbf{A} \mathbf{x} \|_{\infty} = \| \mathbf{e} \|_{\infty} < \frac{1}{2} \), it must be that the decoded vector \( \mathbf{y} \) satisfies this inequality as well. Therefore, \( \| \mathbf{A} \mathbf{y} - \mathbf{A} \mathbf{x} \|_{\infty} \leq \| \mathbf{b} - \mathbf{A} \mathbf{y} \|_{\infty} + \| \mathbf{b} - \mathbf{A} \mathbf{x} \|_{\infty} < 1. \) Since \( s \leq m/2 \), by definition \( \mathbf{x} - \mathbf{y} \) is an \( m \)-sparse vector, which guarantees that \( \mathbf{y} = \mathbf{x} \) so our decoding was successful. Note that if instead of error 1/2 we want to allow larger error \( C \) we can simply multiply all entries of \( \mathbf{A} \) by a factor of \( 2C \).

Fukshansky, Needell and Sudakov [5] showed that a matrix \( \mathbf{A} \in \mathbb{R}^{m \times d} \) with integer entries \( |a_{ij}| \leq k \) and all \( m \times m \) submatrices having full rank must satisfy \( d = O(k^2 m) \). They also proved that such matrices exist when \( d = \Omega(\sqrt{km}) \). Their upper bound was improved by Konyagin [6] who showed that \( d \) must have order at most \( O(k(\log k) m) \) (all logarithms here and later in the paper are in base \( e \)) for \( m \geq \log k \) and at most \( O(k^{m/(m-1)} m^2) \) for \( 2 \leq m < \log k \). Improving these results further, in this paper we obtain the following new upper bound.

**Theorem 1.2.** Let \( \mathbf{A} \) be an \( m \times d \) integer matrix such that \( |a_{ij}| \leq k \), \( m \geq \log k \), and all \( m \times m \) submatrices of \( \mathbf{A} \) have a full rank. If \( k \) is sufficiently large, then \( d \leq 100k\sqrt{\log k} m \) for \( m \geq \log k \) and \( d \leq 400k^{m/(m-1)} m^{3/2} \) for \( 2 \leq m < \log k \).

The lower bound construction for Problem 1 uses random matrices and is based on a deep result of Bourgain, Vu and Wood [3] which estimates the probability that a random \( m \times m \) matrix with integer entries from \([-k, k]\) is singular. It is expected that their result is not tight and the probability of singularity for such matrix has order \( k^{-(1-o(1))m} \) as \( k \to \infty \). If this is the case then \( m \times d \) matrices, satisfying Problem 1 exist for \( d \) close to \( km \). This suggests that our new bound for \( m \geq \log k \) is not far from being optimal.

On the other hand, we get the following result.

**Theorem 1.3.** Let \( k \in \mathbb{N}, m \in \mathbb{N}, m \geq 2, \) and

\[ m < d \leq \max(k + 1, k^{m/(m-1)}/2). \]

Then there is an \( m \times d \) integer matrix \( \mathbf{A} \) such that \( |a_{ij}| \leq k \) and all \( m \times m \) submatrices of \( \mathbf{A} \) have a full rank.

We observe that this theorem improves the lower bound from [5] for \( m = o(\sqrt{k}) \). Moreover, the existence of required matrices in [5] was proven by using probabilistic arguments, but the matrices in Theorem 1.3 are explicit and easily computable. Also we notice that upper and lower estimate for \( d \) in Theorems 1.2 and 1.3 differ by a factor \( O(m^{3/2}) \) depending on \( m \) only.

Our result can be also used to answer a special case of the problem by Brass, Moser and Pach. In [2] (Chapter 10.2, Problem 6) they asked, what is the minimum number \( M \) of \( s \)-dimensional linear subspaces necessary to cover \( m \)-dimensional \( k \times \cdots \times k \) grid \( K = \{x \in \mathbb{Z}^m : \|x\|\leq k\} \). Balko, Cibulko and Valtr [1] studied this problem and obtained upper and lower bounds for \( M \). In particular in the case when \( s = m - 1 \) they proved that

\[ k^{m/(m-1) - o(1)} \leq M \leq C_m k^{m/(m-1)}. \]

Using Theorem 1.3 we obtain a new lower bound which is tight up to a constant factor.

**Corollary 1.4.** For \( k \geq m \geq 2 \) we have

\[ M \geq k^{m/(m-1)/(2m - 2)}. \]

Indeed, suppose that we cover \( K \) by \( M \) hypersubspaces \( P_1, \ldots, P_M \). We consider the columns of the matrix \( \mathbf{A} \) constructed in Theorem 1.3. Since any \( m \) of them are linearly independent, every subspace \( P_i \) contains at most \( m - 1 \) of these columns. Thus, \( d \leq M(m - 1) \), and the corollary follows.
2 Proof of Theorem 1.2

Let \( t = \lfloor \log k \rfloor \) for \( m \geq \log k \) and \( t = m \) for \( 2 \leq m < \log k \). In the first case we suppose that \( d > 100k \sqrt{tm} \) and in the second case we suppose that \( d > 400k^{m/(m-1)}m^{3/2} \). Let \( v_1, \ldots, v_t \) be the first \( t \) rows of the matrix \( A \). Take \( \Lambda = 9 \) if \( m \geq \log k \) and \( \Lambda = \lceil 25k^{1/(m-1)} \rceil \) otherwise. Given a vector of integer coefficients \( \lambda = (\lambda_1, \ldots, \lambda_t) \) such that \( 0 \leq \lambda_i \leq \Lambda \) denote by \( v_\lambda \) a linear combination \( \sum_i \lambda_i v_i \). Our goal is to find two combinations \( \lambda \neq \lambda' \) such that corresponding vectors \( v_\lambda \) and \( v'_\lambda \) agree on at least \( m \) coordinates. This will show that a linear combination of first \( t \) rows of matrix \( A \) with coefficients \( \lambda - \lambda' \neq 0 \) has at least \( m \) zeros and therefore the \( m \times m \) submatrix of \( A \) whose columns correspond to these zeros is degenerate, since its first \( t \) rows are linearly dependent.

Consider \( \lambda \) chosen uniformly at random out of \((\Lambda + 1)^t\) possible vectors and look on a value of a fixed coordinate \( j \) of the vector \( v_\lambda \). This value is a random variable \( X \) which is a sum of the \( t \) independent random variable \( X_i \), where \( X_i \) is a value of the \( j \)-th coordinate of \( \lambda_i v_i \). Since \( |a_{ij}| \leq k \), we have that \( |X_i| \leq \Lambda k \) and therefore its variance \( \text{Var}(X_i) \leq \mathbb{E}(X_i^2) \leq \Lambda^2 k^2 \). This implies that \( \text{Var}(X) = \sum_i \text{Var}(X_i) \leq \Lambda^2 k^2 t \). Thus, by Chebyshev’s inequality, with probability at least \( 3/4 \) the value of \( X \) belongs to an interval \( I \) of length \( 4 \sqrt{\text{Var}(X)} \leq 4 \Lambda k \sqrt{t} \). Hence there are at least \( 0.75 \cdot (\Lambda + 1)^t \) linear combinations \( v_\lambda \) whose \( j \)-th coordinate belongs to \( I \). For every integer \( s \) let \( h_j(s) \) be the number of linear combinations \( v_\lambda \) whose \( j \)-th coordinate is \( s \) and let \( h_j \) be the number of ordered pairs \( \lambda \neq \lambda' \) such that \( v_\lambda \) and \( v'_\lambda \) agree on \( j \)-th coordinate. By definition \( 0.75 \cdot (\Lambda + 1)^t \leq \sum_{s \in I} h_j(s) \leq (\Lambda + 1)^t \) and \( h_j = \sum_{s} h_j(s)(h_j(s) - 1) \).

If \( m \geq \log k \) and \( \Lambda = 9 \) then using a Cauchy-Schwarz inequality, together with the fact that \( 10^t > k^2 \gg k \sqrt{\log k} = k \sqrt{t} \) for sufficiently large \( k \), we have

\[
\begin{align*}
  h_j &= \sum_{s} h_j(s)(h_j(s) - 1) = \sum_{s} h_j^2(s) - \sum_{s} h_j(s) \\
  &\geq \frac{1}{4 \Lambda k \sqrt{t}} \left( \sum_{s \in I} h_j(s) \right)^2 - 10^t \\
  &\geq \frac{1}{80k \sqrt{t}} 10^{2t} - 10^t \\
  &\geq \frac{1}{100k \sqrt{t}} 10^{2t}.
\end{align*}
\]

Since the number of ordered pairs \( \lambda \neq \lambda' \) is at most \( 10^{2t} \) and the number of coordinates \( j \) is \( d \), by averaging we obtain that there is a pair \( \lambda \neq \lambda' \) which agrees on at least

\[
\frac{\sum_{j=1}^{d} h_j}{10^{2t}} \geq \frac{d}{100k \sqrt{t}} \geq m
\]

coordinates. As we explain above, this implies that \( A \) has an \( m \times m \) degenerate submatrix.

Now we consider the case \( 2 \leq m < \log k \). Then, due to the inequality

\[
\frac{(\Lambda + 1)^m}{8 \Lambda k \sqrt{m}} \geq \frac{25^m k^{m/(m-1)} \sqrt{m}}{200k^{m/(m-1)} \sqrt{m}} > 2,
\]

we have
\[ h_j = \sum_s h_j(s)(h_j(s) - 1) = \sum_s h_j^2(s) - \sum_s h_j(s) \geq \sum_{s \in I} h_j^2(s) - (\Lambda + 1)^m \]
\[
\geq \frac{1}{4Ak\sqrt{m} + 1} \left( \sum_{s \in I} h_j(s) \right)^2 - (\Lambda + 1)^m
\]
\[
\geq \frac{1}{4Ak\sqrt{m} + 1} \left( 0.75 \cdot (\Lambda + 1)^m \right)^2 - (\Lambda + 1)^m
\]
\[
\geq (\Lambda + 1)^m \left( \frac{(\Lambda + 1)^m}{8Ak\sqrt{m}} - 1 \right) \geq (\Lambda + 1)^{\frac{m}{16Ak\sqrt{m}}}.
\]

Since the number of ordered pairs \( \lambda \neq \lambda' \) is at most \((\Lambda + 1)^{2m}\) and the number of coordinates \( j \) is \( d \), by averaging we obtain that there is a pair \( \lambda = \lambda' \) which agrees on at least
\[
\frac{\sum_{j=1}^d h_j}{(\Lambda + 1)^{2m}} \geq \frac{d}{16Ak\sqrt{m}} \geq m
\]
coordinates as required. This completes the proof of the theorem. \( \square \)

3 Proof of Theorem 1.3

We have to construct required \( m \times d \) matrices with \( d \geq k + 1 \) provided that \( k \geq m \) and \( d \geq k^{m/(m-1)/2} \) provided that \( k^{m/(m-1)/2} > m \).

First we will construct an \( m \times d \) matrix with \( d \geq k + 1 \). There exists an odd prime \( d \) with \( k + 1 \leq d \leq 2k + 1 \). We define the matrix \( A \) by taking \( a_{i,j} \equiv j^{-1}( \text{mod } d) \) with \( |a_{i,j}| \leq (d-1)/2 \leq k \). Considering the matrix \( A \) modulo \( d \) we find that any submatrix of \( A \) of size \( m \times m \) is a Vandermonde matrix modulo \( d \). Hence, its determinant is not zero modulo \( d \). This implies that this submatrix has a full rank.

Next we will construct an \( m \times d \) matrix with \( d \geq k^{m/(m-1)/2} \). We can consider that \( k^{m/(m-1)/2} > k + 1 \) and, in particular, \( k \geq 3 \). There exists a prime \( d \) with \( k^{m/(m-1)/2} \leq d < k^{m/(m-1)} \). For \( u \in \mathbb{R} \) we denote by \( \lfloor u \rfloor \) the distance from \( u \) to the nearest integer. We consider the \( m \times d \) matrix \( A' \) with entries \( a'_{i,j} = j^{-1} \). Again, the determinant of any \( m \times m \) submatrix of \( A' \) is not zero modulo \( d \). The idea is to multiply the columns of \( A' \) by appropriate integers not divisible by \( d \) and to replace all entries by integers congruent modulo \( d \) with absolute values bounded by \( k \). Clearly, these operations preserve the above mentioned property of submatrices of the matrix.

Using Dirichlet’s theorem on simultaneous approximations (see, e.g., [7], Chapter 2, Theorem 1A and the remark after it), we find that for every \( j = 1, \ldots, d \) there is a positive integer \( l_j < d \) such that \( \lfloor l_j j^{-1} / d \rfloor \leq d^{-1/m} \) for \( i = 1, \ldots, m \). Hence, for any \( i \) there is an integer \( a_{i,j} \) such that \( a_{i,j} \equiv l_j j^{-1} (\text{mod } d) \) and \( |a_{i,j}| \leq d^{1-1/m} \leq k \) as required. This completes the proof of Theorem 1.3 \( \square \)

References

[1] M. Balko, J. Cibulka, and P. Valtr, Covering lattice points by subspaces and covering point–hyperplane incidences, Discrete Comput. Geom. 61 (2019), 325–354
[2] P. Brass, W. Moser and J. Pach, *Research problems in discrete geometry*, Springer, New York, 2005.

[3] J. Bourgain, V. Vu, P. M. Wood, On the singularity probability of discrete random matrices, *J. Funct. Anal.* 258 (2010), 559–603.

[4] S. Foucart and H. Rauhut, *A Mathematical Introduction to Compressive Sensing*, Springer, New York, 2013.

[5] L. Fukshansky, D. Needell, and B. Sudakov, An algebraic perspective on integer sparse recovery, *Applied Mathematics and Computation* 340 (2019), 31–42.

[6] S. V. Konyagin, On the recovery of an integer vector from linear measurements, *Mathematical Notes*, 104 (2018), 859–865.

[7] W. M. Schmidt, Diophantine approximation, *Lecture Notes in Mathematics*, 785, Springer–Verlag, Berlin–Heidelberg–New York, 1980.