

## UPPER BOUNDS ON THE MAGNITUDE OF SOLUTIONS OF CERTAIN LINEAR SYSTEMS WITH INTEGER COEFFICIENTS\*

PEDRO J. FREITAS<sup>†</sup>, SHMUEL FRIEDLAND<sup>‡</sup>, AND GASPAR PORTA<sup>§</sup>

**Abstract.** In this paper we consider a linear homogeneous system of  $m$  equations in  $n$  unknowns with integer coefficients over the reals. Assume that the sum of the absolute values of the coefficients of each equation does not exceed  $k+1$  for some positive integer  $k$ . We show that if the system has a nontrivial solution then there exists a nontrivial solution  $\mathbf{x} = (x_1, \dots, x_n)^\top$  such that  $\frac{|x_j|}{|x_i|} \leq k^{n-1}$  for each  $i, j$  satisfying  $x_i x_j \neq 0$ . This inequality is sharp.

We also prove a conjecture of A. Tyszka related to our results.

**Key words.** Linear Systems, Upper Bounds

**AMS subject classifications.** 15A39, 15A45.

**1. Introduction.** In this paper we consider  $m$  homogeneous linear equations, with integer coefficients, in  $n$  variables; i.e., our system is  $A\mathbf{x} = \mathbf{0}$ , where  $A$  is an  $m \times n$  matrix with integer entries  $A = [a_{ij}] \in \mathbb{Z}^{m \times n}$ , and  $\mathbf{x} \in \mathbb{R}^n$ .

For a nonzero vector  $\mathbf{x} = (x_1, \dots, x_n)^\top \in \mathbb{R}^n$  we define the *relative magnitude* of  $\mathbf{x}$  as

$$(1.1) \quad \omega(\mathbf{x}) = \max \left\{ \frac{|x_j|}{|x_i|}, \mathbf{x} \in \mathbb{R}^n \setminus \{\mathbf{0}\}, x_i \neq 0 \right\}.$$

It is easy to check that if  $\mathbf{x}$  has no zero coordinates then the relative magnitude of a vector  $\mathbf{x}$  coincides with the classic condition number of the diagonal matrix that has the entries of  $\mathbf{x}$  in its diagonal.

Denote by  $N(A)$  the nullspace of  $A$  and let  $N'(A) := N(A) \setminus \{\mathbf{0}\}$ .

The *solution relative magnitude* of  $A$  is given by

$$(1.2) \quad \omega(A) = \inf \{ \omega(\mathbf{x}), \mathbf{x} \in N'(A) \}.$$

---

\*Received by the editors on January 17, 2012. Accepted for publication on May 5, 2012. Handling Editor: Oskar Maria Baksalary.

<sup>†</sup>Centro de Estruturas Lineares e Combinatória, Av Prof Gama Pinto 2, P-1649-003 Lisboa and Departamento de Matemática da Faculdade de Ciências, Campo Grande, Edifício C6, piso 2, P-1749-016 Lisboa. Universidade de Lisboa (pedro@ptmat.fc.ul.pt).

<sup>‡</sup> Department of Mathematics, Statistics, and Computer Science, University of Illinois at Chicago, Chicago, Illinois 60607-7045, USA (friedlan@uic.edu).

<sup>§</sup> Washburn University, 1700 SW College Ave. Topeka, KS 66621 (gaspar.porta@washburn.edu).

We agree that if  $A\mathbf{x} = \mathbf{0}$  has the unique solution  $\mathbf{x} = \mathbf{0}$  then  $\omega(A) = 0$ .

The aim of this paper is to establish a sharp upper bound on  $\omega(A)$  in terms of  $\|A\|_\infty := \max_i \sum_{j=1}^n |a_{ij}|$ . Namely, we have the following result.

**Theorem 1.1.** *Consider a nonzero matrix  $A \in \mathbb{Z}^{m \times n}$ . If  $\|A\|_\infty = 1, 2$ , then  $\omega(A) = 0$  or  $\omega(A) = 1$ . For  $\|A\|_\infty \geq 3$ , we have the following sharp upper bound for the solution relative magnitude of  $A$ :*

$$(1.3) \quad \omega(A) \leq (\|A\|_\infty - 1)^{\text{rank } A}.$$

It is quite easy to see the sharpness of (1.3). Let  $A\mathbf{x} = \mathbf{0}$  be the system of  $n - 1$  homogeneous equations  $kx_i - x_{i+1} = 0$  for  $i = 1, \dots, n - 1$  for a given  $k \in \mathbb{N}$ . Then  $\|A\|_\infty = k + 1$ ,  $\text{rank } A = n - 1$  and  $\omega(A) = \omega(\mathbf{x}) = \frac{x_n}{x_1} = k^{n-1}$  for any  $\mathbf{x} \neq \mathbf{0}$  in the null space of  $A$ .

The cases  $\|A\|_\infty = 1, 2$  are simple (cf. Proposition 2.2). The case  $\|A\|_\infty \geq 3$  is deduced from the following result.

**Theorem 1.2.** *Fix an integer  $k \geq 2$ . Consider a linear system in which all equations are of one of the following two types:*

$$(1.4) \quad x_i = \pm 1$$

$$(1.5) \quad \pm x_{i_1} \pm x_{i_2} \pm \dots \pm x_{i_{l+1}} = 0,$$

where  $l$  is a non-negative integer satisfying  $l \leq k$ . (The indices  $i_1, \dots, i_{l+1}$  are not necessarily distinct and the integer  $l$  may depend on the equation.) Assume that this system is solvable. Then there exists a rational solution such that  $|x_j| \leq k^{n-1}$  for each variable  $x_j$ , with  $n$  being the rank of the system. The above bound is sharp.

For the system  $x_1 = 1$  and  $kx_i - x_{i+1} = 0$  for  $i = 1, \dots, n - 1$  our theorem is sharp. Our main tool is the Hadamard-Fischer determinant inequalities combined with graph theoretical arguments.

The case  $k = 2$  of Theorem 1.2 proves one of the conjectures of A. Tyszkowski presented in [5, 6]. See §2.

We now survey briefly the contents of our paper. In §2 we discuss some properties of  $\omega(A)$  and show that Theorem 1.2 implies Theorem 1.1 and one of the above mentioned Tyszkowski's conjectures. In §3 we lay the ground for the proof of Theorem 1.2. We reduce the system (1.4)–(1.5) to a system of the same type of equations with a smaller number of variables satisfying the following conditions. First, the system has a unique solution; second, the system (1.4) is  $x_1 = 1$ ; third, no variable is equal to zero, fourth,  $x_i \neq \pm x_j$  for  $i \neq j$ . These conditions allow us to split the new system of equations (1.5) into a

finite number of maximal chain equations of the form  $kx_{i_{j+1}} = \pm x_{i_j}$  for  $j = 1, \dots, t$ . We show that no two maximal chains have a common variable. In §4 we estimate the determinants of certain tridiagonal matrices related to a maximal chain, and the Euclidean norm of the coefficients of each equation in (1.5) which does not appear in any maximal chain. In §5 we conclude the proof of Theorem 1.2.

**2. On Relative Magnitudes.** For  $m \in \mathbb{N}$  denote  $[m] := \{1, \dots, m\}$ . Let  $\mathbf{x} = (x_1, \dots, x_n)^\top \in \mathbb{R}^n \setminus \{\mathbf{0}\}$ . Define the *relative magnitude* of  $\mathbf{x}$  by (1.1). Let

$$\phi(\mathbf{x}) = \min\{|x_i|, x_i \neq 0\}, \quad \Phi(\mathbf{x}) = \max\{|x_i|, i \in [n]\}.$$

So  $\omega(\mathbf{x}) = \frac{\Phi(\mathbf{x})}{\phi(\mathbf{x})}$ . Let  $A \in \mathbb{R}^{m \times n}$ . If  $\text{rank } A \leq n - 1$  we define the solution relative magnitude of  $A$  by (1.2), otherwise we let  $\omega(A) = 0$ . The following result shows that the infimum in (1.2) is attained.

**Proposition 2.1.** *Let  $A \in \mathbb{R}^{m \times n}$  and assume that  $\text{rank } A \leq n - 1$ . Then*

$$(2.1) \quad \omega(A) = \min\{\omega(\mathbf{x}), \mathbf{x} \in N'(A)\} \geq 1.$$

So, we can write

$$\omega(A) = \min_{\mathbf{x} \in N'(A)} \max_{x_i \neq 0} \frac{|x_j|}{|x_i|}.$$

*Proof.* Under these conditions, we must have a nonzero solution  $\mathbf{z}$  of the system  $A\mathbf{x} = \mathbf{0}$ . If  $z_i$  is a nonzero coordinate of  $\mathbf{z}$ , then  $\omega(\mathbf{z}) \geq |z_i|/|z_i| = 1$ . Since this holds for any nonzero solution,  $\omega(A) \geq 1$ .

Now let

$$\Sigma(A) := \left\{ \mathbf{x} = (x_1, \dots, x_n)^\top \in \mathbb{R}^n, \mathbf{x} \in N(A), \sum_{i=1}^n |x_i| = 1 \right\}.$$

We clearly have  $\omega(A) = \inf\{\omega(\mathbf{x}), \mathbf{x} \in \Sigma(A)\}$ , and  $\Sigma(A)$  is a compact set, but the function  $\omega$  is in general not continuous on  $\Sigma(A)$ . For instance, if  $A = [1 \ 0 \ 0]$  then  $(0, 1, 0)^\top, (0, 1-t, t)^\top \in \Sigma(A)$ ,  $(0, t, 1-t)^\top \in \Sigma(A)$  if and only if  $t \in [0, 1]$ , as  $|t| \leq 1$  and  $|1-t| \leq 1$ .

For  $t \in (0, \frac{1}{2})$ , we have  $\omega((0, 1, 0)^\top) = 1$ ,  $\omega((0, 1-t, t)^\top) = (1-t)/t$ ,  $\omega((0, t, 1-t)^\top) = \frac{t}{1-t}$ .

To circumvent this problem, we start by noticing that  $\omega(A) = \inf\{\omega(\mathbf{x}), \mathbf{x} \in \Sigma(A), \omega(\mathbf{x}) \leq 2\omega(A)\}$ . Assume that  $\mathbf{x} \in \Sigma(A), \omega(\mathbf{x}) \leq 2\omega(A)$ . As  $1 \leq (n-1)\Phi(\mathbf{x}) + \phi(\mathbf{x})$  we deduce that

$$\frac{1}{\phi(\mathbf{x})} \leq (n-1)\omega(\mathbf{x}) + 1 \leq 2(n-1)\omega(A) + 1,$$

and therefore

$$\phi(\mathbf{x}) \geq \alpha(A) := \frac{1}{2(n-1)\omega(A) + 1}.$$

Let  $\Sigma_1(A) := \{\mathbf{x} \in \Sigma(A), \phi(\mathbf{x}) \geq \alpha(A)\}$ . Clearly,  $\Sigma_1(A)$  is a compact set, and both  $\Phi(\mathbf{x})$  and  $\phi(\mathbf{x})$  are continuous on  $\Sigma_1(A)$ . Hence  $\omega(\mathbf{x})$  is a continuous function on  $\Sigma_1(A)$  and

$$\omega(A) = \inf\{\mathbf{x} \in \Sigma_1(A)\} = \min\{\mathbf{x} \in \Sigma_1(A)\}.$$

Hence, condition (2.1) holds.  $\square$

Define the *support* of a vector  $\mathbf{x}$  as the set  $I \subseteq [n]$  such that  $x_i \neq 0 \Leftrightarrow i \in I$ . Following [1], given a subspace  $W \leq \mathbb{R}^n$ , we say that a nonzero vector is *elementary* (in  $W$ ) if its support is minimal among all supports of nonzero vectors in  $W$ . In other words  $\mathbf{x} \in W$  is elementary if for  $\mathbf{y} \in W$ ,  $\mathbf{y} \neq \mathbf{0}$  and  $\text{supp } \mathbf{y} \subseteq \text{supp } \mathbf{x}$ , we have  $\text{supp } \mathbf{y} = \text{supp } \mathbf{x}$ . It is known that any subspace has a basis formed by elementary vectors (see [1] and [3, p. 528] for an algorithm to find such a basis).

**Proposition 2.2.** *Let  $A \in \mathbb{Z}^{m \times n}$  and assume that  $\|A\|_\infty = 1, 2$ . Then, for every elementary vector  $\mathbf{x} \in N(A)$ , we have  $\omega(\mathbf{x}) = 1$ . Therefore  $\omega(A) = 0$  or  $1$ , as stated in Theorem 1.1.*

*Proof.* Suppose first that  $\text{rank } A = n$ . Then  $\omega(A) = 0$ . Assume now that  $\text{rank } A \leq n - 1$ . Suppose first that  $\|A\|_\infty = 1$ . Then each nontrivial equation of  $A\mathbf{z} = \mathbf{0}$  is given by  $x_i = 0$  for some  $i \in [n]$ . Hence the general solution of  $A\mathbf{z} = \mathbf{0}$  is of the following form. The set of free variables is a nonempty strict subset  $S$  of  $[n]$  and all other dependent variables equal to zero. In particular, all elementary vectors have only one nonzero coordinate:  $A\mathbf{e}_j = \mathbf{0}$  for  $j \in S$ , hence  $\omega(\mathbf{x}) = \omega(A) = 1$ .

Assume now that  $\|A\|_\infty = 2$ . Then the nontrivial equations of  $A\mathbf{z} = \mathbf{0}$  are either of the form  $x_i = 0$  or  $x_i = \pm x_j$ . If  $A\mathbf{e}_j = \mathbf{0}$  for some  $j \in [n]$  then  $\omega(\mathbf{e}_j) = \omega(A) = 1$ . Assume that  $A\mathbf{e}_j \neq \mathbf{0}$  for each  $j \in [n]$ . It is easy to see that the general solution of this system has the following general form. One can partition the set  $[n]$  into

$$[n] = S \dot{\cup} T_1 \dot{\cup} \cdots \dot{\cup} T_l,$$

so that  $x_i = 0$  if  $i \in S$ , each  $T_p$  contains at least two elements, for  $p \in [l]$  and, for each pair of distinct indices  $i, j \in T_p$ , we have  $|x_i| = |x_j|$ . The value of all  $|x_i|, i \in T_p$ , can be prescribed arbitrarily for each  $p \in [l]$ ; in particular, elementary vectors  $\mathbf{x}$  satisfy  $x_i = 0$  for  $i \notin T_p$  and  $|x_j| \neq 0$  for  $j \in T_p$ , for some  $p \in [l]$ . Hence  $\omega(\mathbf{x}) = \omega(A) = 1$ .  $\square$

The following two results prove that, once Theorem 1.2 is proved, then Theorem 1.1 holds also for  $\|A\|_\infty \geq 3$ .

**Proposition 2.3.** *Let  $A \in \mathbb{Z}^{m \times n}$  with  $\text{rank } A \leq n - 1$  and  $\|A\|_\infty \geq 3$ , and let  $\mathbf{x} \neq \mathbf{0}$  be an elementary vector of  $N(A)$ , with support  $I$ . Then  $|I| - 1 \leq \text{rank } A$*

and

$$\omega(\mathbf{x}) \leq (\|A\|_\infty - 1)^{|I|-1}.$$

In particular,  $\omega(A) \leq (\|A\|_\infty - 1)^{\text{rank } A}$ , as stated in Theorem 1.1.

*Proof.* If  $|I| = 1$ , then the result holds. Now suppose  $|I| \geq 1$  and let  $\mathbf{y}$  be the vector of  $\mathbb{R}^{|I|}$  formed by the nonzero coordinates of  $\mathbf{x}$ ,  $\omega(\mathbf{x}) = \omega(\mathbf{y})$ . Let  $B$  be the matrix obtained from  $A$  by selecting the columns with indices in  $I$ . Then  $B\mathbf{y} = \mathbf{0}$  and we must have  $\text{rank } B = |I| - 1$ , otherwise we would be able to get a solution of  $A\mathbf{z} = \mathbf{0}$  with more zero coordinates (by setting one of the free variables to zero), contradicting the fact that  $\mathbf{x}$  is elementary. Clearly  $|I| - 1 \leq \text{rank } A$ . Set  $k + 1 := \|A\|_\infty \geq \|B\|_\infty$ ,  $k \geq 2$ .

Since the null space of  $B$  is spanned by one vector  $\mathbf{w} \neq \mathbf{0}$  we deduce that  $\omega(\mathbf{w}) = \omega(\mathbf{y})$ . Without loss of generality we may assume that  $\phi(\mathbf{w}) = 1$ , with  $w_j = 1$ . If we now consider the system  $B\mathbf{z} = \mathbf{0}$  along with the equation  $z_j = 1$ , its only solution will be  $\mathbf{y}$  and the system is of the type described in Theorem 1.2, with  $\text{rank } |I|$ . Once this result is proved, we get that  $\omega(\mathbf{y}) = \Phi(\mathbf{y}) \leq k^{|I|-1}$ , and hence

$$\omega(\mathbf{x}) = \omega(\mathbf{y}) \leq (\|A\|_\infty - 1)^{|I|-1}.$$

This proves the result.  $\square$

This previous result allows for an improvement of Theorem 1.2.

**Corollary 2.4.** Consider a nonzero matrix  $A \in \mathbb{Z}^{m \times n}$  with  $\|A\|_\infty \geq 2$ . Let  $t > 0$  be the least number of nonzero coordinates in a nonzero vector of  $N(A)$ . Then  $t - 1 \leq \text{rank } A$  and

$$\omega(A) \leq (\|A\|_\infty - 1)^{t-1} \leq (\|A\|_\infty - 1)^{\text{rank } A}.$$

One of the conjectures of A. Tyszkiewicz, presented in [5, 6], is:

**Conjecture 2.5.** Let  $n \geq 2$ . Assume that we have a solvable linear system of  $n$  equations of the following two types:

$$(2.2) \quad x_i + x_j = x_k \quad \text{and} \quad x_l = 1.$$

(The indices  $i, j$  and  $k$  are not necessarily distinct.) Then the above system has a rational solution with  $|x_i| \leq 2^{n-1}$  for all  $1 \leq i \leq n$ . This result is sharp.

Some partial results about this conjecture were obtained by Tyszkiewicz [5], with upper bound  $\sqrt{5}^{n-1}$ , and by Cipu [2], with upper bound  $2^n$ . Clearly, Theorem 1.2 for  $k = 2$  yields Conjecture 2.5. The rest of this paper is devoted to proving Theorem 1.2.

**3. Simplification.** Assume that our given system of equations has variables  $x_i, i \in [m]$ . We'll show that it is enough to prove Theorem 1.2 after the following simplifications.

1. Observe first that if all equations are of type (1.5), we have a homogeneous system, for which there is a zero solution. Now we assume we have a nonempty set of equations  $x_i = \pm 1$  for  $i \in S$ , and  $|S| \geq 2$ . We can replace this set by one equation  $x_i = \pm 1$  and equations  $x_j \pm x_i = 0$  for  $j \in S \setminus \{i\}$ .

2. If the system is indeterminate, let  $S \subset \mathbb{N}$  be a set of free variables, and for every  $x_j, j \in S$ , take  $x_j = 0$ . We get new equations of the type (1.5) with variables indexed in  $[m] \setminus S$ . Out of those, we select any  $m - |S| - 1$  linearly independent equations in  $m - |S|$  variables. Rename these variables so that the first equation is  $x_1 = \pm 1$  and the other variables are  $x_j, j = 2, \dots, m' = m - |S|$ . Hence the new system in  $m'$  variables has a unique solution  $\mathbf{x} = (x_1, \dots, x_{m'})^\top$ .

3. Let  $T \subset [m']$  be the set of all indices  $j$  for which  $x_j = 0$ . As above we replace the above system by a system with variables indexed in  $[m'] \setminus T$ , where the system has a unique solution and each  $x_j \neq 0$  for  $j \in [m'] \setminus T$ . Again we rename our variables so now we have  $n$  variables  $x_1, \dots, x_n$ , where the first equation is  $x_1 = \pm 1$  and all other  $n - 1$  equations are of the form (1.5). This system has a unique solution and each  $x_j \neq 0$ . Note that the rank of this system will be less than or equal to the rank of the original system.

4. Suppose that for this unique solution  $\mathbf{x}$  we have  $x_i = \pm x_j$  for  $i < j$ . Then we eliminate  $x_j$  in all equations (1.5) by substitution  $x_j = \pm x_i$ . Note that we still have a unique solution of the system with one variable less.

We repeat this process until we have a system  $x_1 = \pm 1$ , all other equations are of the form (1.5), such that this system has a unique solution and  $x_i \neq \pm x_j$  for  $i \neq j$ . So again we can assume that we have a system in which all equations are of the form (1.4) or (1.5).

5. Next we cancel terms in the left-hand side of (1.5). I.e., if we have a term  $x_j$  and a term  $-x_j$  in the left-hand side of (1.5), then we replace the two terms by 0. Thus we can assume that the left-hand side of (1.5) does not contain the same variable  $x_j$  with opposite signs. Since each solution  $x_j \neq 0$ , and the system has a unique solution, it means that after the cancelations each of  $n - 1$  equations of the form (1.5) contains at least two variables.

To summarize, we reduced our problem to the following system of  $n$  equations with  $n$  variables  $x_1, \dots, x_n$  one of them of the form  $x_1 = \pm 1$  and the other  $n - 1$

equations of the form

$$(3.1) \quad \sum_{j=1}^{l+1} \pm c_{i_j} x_{i_j} = 0, \quad 1 \leq i_1 < \cdots < i_{l+1} \leq n,$$

$$(3.2) \quad c_1, \dots, c_{l+1} \in \mathbb{N}, 1 \leq l \leq k, \text{ and } c_1 + \cdots + c_{l+1} \leq k + 1.$$

Furthermore the above system has a unique solution  $\mathbf{x} = (x_1, \dots, x_n)^\top$  with the following properties:  $x_i \neq 0$  for each  $i$  and  $x_i \neq \pm x_j$  for each  $i \neq j$ . We note that not all such systems are being considered. For instance, we are not allowing the case  $l = 1$  and  $c_1 = c_2$ , which would mean that  $x_{i_1} = \pm x_{i_2}$ .

We note that these reductions do not change the maximum value of  $|x_p|$ , but may reduce the rank of the original system. Hence it is enough to prove Theorem 1.2 for the system we have obtained.

6. Among the homogeneous equations, we now consider those of type  $kx_i = \pm x_j$ , with  $i \neq j$ . Call these *equations of type 2*. The equation  $x_1 = \pm 1$  is called of *type 1* and all others will be of *type 3*. By replacing  $x_1$  by  $\pm x_1$  we can assume that the equation of type 1 is  $x_1 = 1$ . We claim that we can organize equations of type 2 in maximal chains of the form

$$kx_{i_2} = \pm x_{i_1}, \quad kx_{i_3} = \pm x_{i_2}, \quad \dots \quad kx_{i_t} = \pm x_{i_{t-1}},$$

which we denote as the chain  $i_1 \rightarrow i_2 \rightarrow \dots \rightarrow i_t$ . So  $i_1$  and  $i_t$  are the *head* and the *tail* of the chain, respectively. We now claim that no variable appears in more than one maximal chain.

Assume to the contrary that two maximal chains intersect at  $i$ . Since these two chains are maximal  $i$  is not the tail of one chain and the head of another chain. Thus we only have the following two cases which are impossible in view of our assumptions.

- $kx_i = \pm x_j$  and  $kx_i = \pm x_p$ . Hence  $x_j = \pm x_p$ . This violates the condition that  $x_j \neq \pm x_p$
- $kx_j = \pm x_i$  and  $kx_p = \pm x_i$ . Hence  $x_j = \pm x_p$ . This violates the condition that  $x_j \neq \pm x_p$

We are now left with a system of equations with a unique solution, in which no chains intersect.

Reorder the variables so that in each chain we have

$$kx_{i_1} = \pm x_{i_1+1}, \quad kx_{i_1+1} = \pm x_{i_1+2}, \quad \dots \quad kx_{i_1+t-1} = \pm x_{i_1+t}.$$

After this, when we group all equations belonging to the same chain, the system matrix will have a block of rows that looks like this:

$$\begin{bmatrix} k & \pm 1 & & & & \\ & k & \pm 1 & & & \\ & & k & \pm 1 & & \\ & & & \ddots & \ddots & \\ & & & & k & \pm 1 \end{bmatrix}$$

This block is of type  $t \times (t + 1)$ , that is to say, the number of columns is the number of rows plus one.

Before these blocks of equations, we put first the equation of type 1. Since we have renumbered the variables, it is no longer necessary that the variable in the equation of type 1 has subindex 1, the equation now is  $x_s = 1$  for some index  $s$ . Moreover, it is possible that the variable  $x_s$  belongs to some maximal chain. Then we list all equations of type 3. We note that if we have  $r$  chains, we must have at least  $r - 1$  equations of type 3, so that we get a square matrix. Call the resulting matrix  $A$ .

We now estimate the magnitude of the only solution of our system using Cramer's rule. Let  $A_i$  be the matrix obtained by replacing column  $i$  of  $A$  by  $e_1$ , the column of independent terms. Since  $|\det A| \geq 1$ , we need to establish that  $|\det A_i| \leq k^{n-1}$ , where  $n$  is now the number of variables and equations of the system, and the size of the matrix  $A$ .

Let  $U_i$  be the  $(n - 1) \times (n - 1)$  submatrix obtained from  $A_i$  by deleting the first row and the  $i$ -th column. We claim that  $|\det A_i| = |\det U_i|$  for  $i = 1, \dots, n$ . Indeed, since the first equation in our system is  $x_s = 1$  then expanding  $\det A_s$  by the first row we obtain that  $\det A_s = \det U_s$ . Now take  $i \neq s$ . Subtract the column  $i$  in  $A_i$  from the column  $s$  to obtain the matrix  $V_i$ . Expand the determinant of  $V_i$  by the first row to deduce  $|\det A_i| = |\det V_i| = |\det U_i|$ .

**4. Principal minors of  $U_i U_i^\top$ .** To obtain a bound for  $|\det U_i|$ , we will consider  $\det W_i$ , where  $W_i = U_i U_i^\top$ , which is majorized by a product of its principal minors according to the Hadamard-Fischer inequality [4, Th. 7.8.3, p. 478]. In  $W_i$ , we consider the principal minors determined by the row blocks we have defined.

Assume that the block of rows corresponding to a chain with  $t$  rows is uncut by the column  $i$ ; i.e., the variable  $x_i$  is not participating in this chain. Then for  $t \geq 2$



the corresponding principal submatrix of  $W_i$  is a  $t \times t$  tridiagonal symmetric matrix:

$$B_t := \begin{bmatrix} k^2 + 1 & \pm k & & & \\ \pm k & k^2 + 1 & \pm k & & \\ & \pm k & k^2 + 1 & & \\ & & & \ddots & \pm k \\ & & & \pm k & k^2 + 1 \end{bmatrix} \in \mathbb{Z}^{t \times t}.$$

For  $t = 1$  we have  $B_1 = [k^2 + 1]$ .

We now consider the case where the deleted column  $i$  in  $A$  does cut through a block of rows in a chain. Then the corresponding principal submatrix of  $W_i$  is the direct sum of two blocks:

$$C_p := \begin{bmatrix} k^2 + 1 & \pm k & & & \\ \pm k & \ddots & & & \\ & & k^2 + 1 & \pm k & \\ & & \pm k & k^2 + 1 & \pm k \\ & & & \pm k & k^2 \end{bmatrix} \in \mathbb{Z}^{p \times p},$$

$$D_q := \begin{bmatrix} 1 & \pm k & & & \\ \pm k & k^2 + 1 & \pm k & & \\ & \pm k & k^2 + 1 & & \\ & & & \ddots & \pm k \\ & & & \pm k & k^2 + 1 \end{bmatrix} \in \mathbb{Z}^{q \times q},$$

where  $p, q \geq 2$  and  $p + q = t$ . Note that  $C_1 = k^2, D_1 = 1$ . It is possible that  $p$  or  $q$  is zero, i.e., we have only one block instead of a direct sum of two blocks. We define  $\det C_0 = \det D_0 = 1$ .

LEMMA 4.1. For  $t \in \mathbb{N}$  and a real number  $k$  the following equalities hold.

$$(4.1) \quad \det B_t = \frac{(k^2)^{t+1} - 1}{k^2 - 1},$$

$$(4.2) \quad \det C_t = k^{2t},$$

$$(4.3) \quad \det D_t = 1.$$

( $\det B_t = (t + 1)$  for  $k = \pm 1$ .)

*Proof.* It is enough to consider the case  $k \neq \pm 1$ . For  $t = 1, 2$  (4.1), (4.2) and (4.3) clearly hold. Assume that  $t \geq 3$ . Using the Laplace expansion by the first row

of  $B_t, D_t$  and by the last row of  $C_t$  we obtain

$$\begin{aligned}\det B_t &= (k^2 + 1) \det B_{t-1} - k^2 \det B_{t-2}, \\ \det C_t &= k^2 \det B_{t-1} - k^2 \det B_{t-2}, \\ \det D_t &= \det B_{t-1} - k^2 \det B_{t-2}.\end{aligned}$$

Consider first the recurrence equations for  $\det B_t$ ,  $t \geq 3$ . The roots of the characteristic polynomial of this recurrent system are 1 and  $k^2$ . So  $\det B_t = a_1(k^2)^t + a_0$ . Since (4.1) is of this form and it holds for  $t = 1, 2$ , (4.1) holds for all  $t \in \mathbb{N}$ .

Substitute the expression (4.1) in the expressions for  $\det C_t$  and  $\det D_t$  to deduce (4.2) and (4.3) respectively.  $\square$

We now discuss the equations of type 3.

LEMMA 4.2. *Let  $2 \leq k \in \mathbb{N}$ . Consider the equation (3.1) of type 3, i.e.  $l \in [2, k]$  and which is not of the form  $kx_i \pm x_j = 0, i \neq j$ . Then the  $\ell_2$  norm of the coefficient vector  $\|(\pm c_{i_1}, \dots, \pm c_{i_{l+1}})\|_2 = \sqrt{\sum_{j=1}^{l+1} c_{i_j}^2}$  is at most  $\min(\sqrt{k^2 - 1}, \sqrt{(k-1)^2 + 4}) \leq \sqrt{k^2 - 1}$ .*

*Furthermore, if we delete a variable  $x_{i_j}$  from (3.1) for some integer  $j \in [1, l+1]$  then the  $\ell_2$  norm of the coefficient vector of this new equation is bounded above by  $\sqrt{(k-1)^2 + 1}$ .*

*In particular, consider the diagonal element  $w_{tt}$  of  $U_i U_i^\top$  corresponding to the equation (3.1) of type 3. If the variable  $x_i$  does not appear in this equation,  $w_{tt} \leq k^2 - 1$ . If  $x_i$  appears in this equation then  $w_{tt} \leq (k-1)^2 + 1$ .*

*Proof.* Let  $x, y \geq 0$  and assume that  $z = x + y$ . Suppose that  $a \in [0, \frac{z}{2}]$  and  $x, y \geq a$ . It is straightforward to show that  $x^2 + y^2 \leq (z-a)^2 + a^2$ , with  $x^2 + y^2 = (z-a)^2 + a^2$  if and only if either  $x = z-a, y = a$  or  $x = a, y = z-a$ .

Suppose first that  $k = 2$ . Then the equation (3.1) of type 3 is either  $\pm x_{i_1} \pm x_{i_2} = 0$  or  $\pm x_{i_1} \pm x_{i_2} \pm x_{i_3} = 0$ . Then the  $\ell_2$  norm of the coefficient vector of this system is at most  $\sqrt{3} = \sqrt{k^2 - 1} < \sqrt{(k-1)^2 + 4}$ .

If we delete one of the variables in this equation, then the  $\ell_2$  norm of the coefficient vector of this new equation is at most  $\sqrt{2} = \sqrt{(k-1)^2 + 1}$ .

Suppose now that  $k \geq 3$ . Assume first that  $l \geq 2$ . Rename the indices so that  $c_{i_l}, c_{i_{l+1}} \in [1, k-2] \cap \mathbb{N}$ . Then  $c_{i_l}^2 + c_{i_{l+1}}^2 < (c_{i_l} + c_{i_{l+1}})^2 + 0^2$ . Hence

$$c := \sum_{j=1}^{l+1} c_{i_j}^2 < (c_{i_l} + c_{i_{l+1}})^2 + \sum_{j=1}^{l-1} c_{i_j}^2.$$

It now follows that the maximal value of  $c$  is achieved when the equation has only two variables, which correspond to the value  $l = 1$ . So the maximum value of  $c$  is achieved when the coefficient of one variable is  $\pm(k-1)$  and the coefficient of the other variable is  $\pm 2$ .

Assume now that in (3.1)  $l = 1$ , i.e.,  $\pm c_{i_1}x_{i_1} \pm c_{i_2}x_{i_2} = 0$ , where  $c_{i_1}, c_{i_2} \in [1, k-1] \cap \mathbb{N}$  and  $z = c_{i_1} + c_{i_2} \leq k+1$ . Using our observation in the beginning of the proof of this lemma it is straightforward to show that  $c_{i_1}^2 + c_{i_2}^2 \leq (k-1)^2 + 2^2$  (equality holds if and only if  $c_{i_1} = k-1, c_{i_2} = 2$  or  $c_{i_1} = 2, c_{i_2} = k-1$ ). For  $k > 2$ ,  $(k-1)^2 + 4 \leq k^2 - 1$ .

Assume now that we remove one variable from (3.1). By renaming the variables we may assume that it is the variable  $x_{i_{l+1}}$ . Thus we need to find an upper bound for  $\sum_{j=1}^l c_{i_j}^2$ , where  $c_{i_1}, \dots, c_{i_l} \in [1, k-1] \cap \mathbb{N}$  and  $\sum_{j=1}^l c_{i_j} \leq k$ . For  $l = 1$  the upper bound is  $(k-1)^2$ . For  $l \geq 2$ , from the above arguments, we deduce that this upper bound is  $(k-1)^2 + 1$ . Equality holds if and only if  $l = 2$ ,  $\{c_{i_1}, c_{i_2}, c_{i_3}\} = \{k-1, 1, 1\}$  and the coefficient of the deleted variable  $x_{i_j}$  is 1.  $\square$

## 5. Proof of Theorem 1.2.

*Proof.* Assume that after the reductions described in §2 we have a system of  $n$  linearly independent equations of the following forms. The first equation is of the form  $x_s = 1$ . The other  $n-1$  equations are of the form (3.1), which are of types 2 and 3.

The equations of type 2 are organized in groups. Each group of type 2 equations is a system of  $t$  equations in  $t+1$  variables of the form  $kx_{i_j} = \pm x_{i_{j+1}}, j = 1, \dots, t$ , which is called a chain of length  $t+1$ . No two distinct chains have a common variable.

The number of chains is  $r \geq 0$ . The equations of the type 3 are of the form (3.1), where  $c_{i_j} \in \mathbb{N}, c_{i_j} \leq k-1$  for  $j = 1, \dots, l+1 \leq k+1$ . Since we have a unique solution to our system, we must have at least  $r-1$  equations of type 3.

Hence our system of equations is of the form  $A\mathbf{x} = \mathbf{e}_1, A \in \mathbb{Z}^{n \times n}, |\det A| \geq 1$ . Recall that we denoted by  $U_i$  the submatrix of  $A$  obtained by deleting the first row and the column  $i$  for  $i = 1, \dots, n$ . Then Cramer's rule yields that  $|x_i| \leq |\det U_i|$ .

Also, recall that in one of our reductions we reduced the number of variables whenever we had the equality  $x_i = \pm x_j$  for  $i \neq j$  or  $x_i = 0$ . These reductions do not change the maximum value of  $|x_p|$ , but may reduce the rank of the original system. Hence it is enough to prove Theorem 1.2 in the above form.

We prove Theorem 1.2 by showing that

$$(5.1) \quad x_i^2 \leq \det W_i = \det U_i U_i^\top \leq k^{2(n-1)}, i = 1, \dots, n,$$

using the Hadamard-Fischer inequality and Lemmas 4.1 and 4.2. This is done by considering two cases.

**Case 1.** Column  $i$  cuts through a chain block of length  $t + 1$ . This block yields a principal minor in  $W_i$ , which will be less than or equal to  $k^{2t} = \det C_t \det D_0 \geq \det C_p \det D_q$ ,  $p + q = t$ , by Lemma 4.1.

Assume first that we have only one chain. If this chain is of length  $n$  we just showed that (5.1) holds. If  $t < n - 1$  then Lemma 4.2 and the Hadamard-Fischer inequality yield  $\det W_i \leq k^{2t}(k^2 - 1)^{n-1-t} < k^{2(n-1)}$ .

Assume now that we have  $r \geq 2$  chains. The lengths of the chains are  $t_1 + 1, \dots, t_{r-1} + 1, t_r + 1 = t + 1$ . Lemma 4.1 yields that the principal minor of  $W_i$  corresponding to the chain of length  $t_j + 1$  is bounded above by  $\frac{k^{2(t_j+1)} - 1}{k^2 - 1}$ . Therefore the product of the  $r - 1$  principal minors in  $W_i$  corresponding to chains  $1, \dots, r - 1$  is bounded above by

$$\prod_{j=1}^{r-1} \frac{k^{2(t_j+1)} - 1}{k^2 - 1} < \frac{k^{2\sum_{j=1}^{r-1} (t_j+1)}}{(k^2 - 1)^{r-1}}.$$

The number of the equations of type 3 is  $n - 1 - \sum_{j=1}^r t_j \geq r - 1$ . In view of Lemma 4.1 the product of the diagonal entries in  $W_i$  corresponding to the equations of type 3 is bounded above by  $(k^2 - 1)^{n-1-\sum_{j=1}^r t_j}$ . Combine all the above inequalities to deduce that

$$\begin{aligned} \det W_i &< k^{2t_r} \frac{k^{2\sum_{j=1}^{r-1} (t_j+1)}}{(k^2 - 1)^{r-1}} (k^2 - 1)^{n-1-\sum_{j=1}^r t_j} \\ &\leq k^{2(t_r + \sum_{j=1}^{r-1} (t_j+1))} (k^2 - 1)^{n-1-(r-1)-\sum_{j=1}^r t_j} \\ &\leq k^{2(n-1)}. \end{aligned}$$

**Case 2.** Column  $i$  cuts through a row of type 3 but not through any chain.

If we have only one chain of length  $t + 1$ , then

$$\det W_i \leq \frac{k^{2(t+1)} - 1}{k^2 - 1} (k^2 - 1)^{n-t-1} < k^{2(t+1)} k^{2(n-t-2)} = k^{2(n-1)}.$$

If there is no chain, the analysis is similar.

Assume now that we have  $r \geq 2$  chains of lengths  $t_1 + 1, \dots, t_r + 1$ . Hence we have  $n - 1 - \sum_{j=1}^r t_j \geq r - 1$  equations of type 3.

One equation of type 3 discussed above must contain the variable  $x_i$ . Since  $(k - 1)^2 + 1 \leq k^2 - 2$  for  $k \geq 2$ , we can conclude, using Lemma 4.2, that the diagonal

entry of  $W_i$  corresponding to this equation does not exceed  $k^2 - 2$ . Hence  $\det W_i$  is less than or equal to

$$\frac{k^{2(t_1+1)} - 1}{k^2 - 1} \frac{k^{2(t_2+1)} - 1}{k^2 - 1} (k^2 - 2) \left( \prod_{j=3}^r \frac{k^{2(t_j+1)} - 1}{k^2 - 1} \right) (k^2 - 1)^{n-2-\sum_{j=1}^r t_j}.$$

Recall that  $n - 2 - \sum_{j=1}^r t_j \geq r - 2$ . Hence as in the Case 1 we get

$$\begin{aligned} & \left( \prod_{j=3}^r \frac{k^{2(t_j+1)} - 1}{k^2 - 1} \right) (k^2 - 1)^{n-2-\sum_{j=1}^r t_j} \\ & \leq k^{2\sum_{j=3}^r (t_j+1)} (k^2 - 1)^{n-2-\sum_{j=1}^r t_j - (r-2)} \leq k^{2(n-2-(t_1+t_2))}. \end{aligned}$$

As  $(k^2 - 1)^2 > k^2(k^2 - 2)$  it follows that

$$\begin{aligned} \frac{k^{2(t_1+1)} - 1}{k^2 - 1} \frac{k^{2(t_2+1)} - 1}{k^2 - 1} (k^2 - 2) & < \frac{(k^{2(t_1+1)} - 1)(k^{2(t_2+1)} - 1)}{k^2(k^2 - 2)} (k^2 - 2) \\ & < \frac{k^{2(t_1+1)} k^{2(t_2+1)}}{k^2} = k^{2(t_1+t_2)+2}. \end{aligned}$$

Combine the above inequalities to deduce  $\det W_i \leq k^{2(n-1)}$ .  $\square$

We thank the referees for their useful remarks.

#### REFERENCES

- [1] R. A. Brualdi, S. Friedland and A. Pothen, The sparse basis problem and multilinear algebra. *SIAM J. Matrix Anal. Appl.*, 16 (1995), 1–20.
- [2] M. Cipu, Small solutions to systems of polynomial equations with integer coefficients, *An. St. Univ. Ovidius Constanta* 19 (2011), no. 2, 89–100.
- [3] T. F. Coleman, A. Pothen, The null space problem I. Complexity, *SIAM J. Alg. Disc. Math.*, (1986), 527–537.
- [4] R. A. Horn and C. R. Johnson, *Matrix Analysis*, Cambridge University Press, 1990.
- [5] A. Tyszk, Some conjectures on addition and multiplication of complex (real) numbers, *Int. Math. Forum* 4 (2009), no. 9–12, 521–530.
- [6] A. Tyszk, Two conjectures on the arithmetic in  $\mathbb{R}$  and  $\mathbb{C}$ . *Mathematical Logic Quarterly*, 56 (2010), 175–184. doi: 10.1002/malq.200910004