How Much Does a Matrix of Rank k Weigh?

Theresa Migler, Kent E. Morrison, Mitchell Ogle

Department of Mathematics
California Polytechnic State University
San Luis Obispo, CA 93407 *

June 11, 2004

1 Introduction

Matrices with very few non-zero entries cannot have large rank. On the other hand matrices without any zero entries can have rank as low as 1. These simple observations lead us to our main question. For matrices over finite fields, what is the relationship between the rank of a matrix and the number of non-zero entries in the matrix? This question motivated a summer research project collaboration among the authors (two undergraduate students and their adviser), and although the question seems natural, we were unable to find any previously published work dealing with it.

We call the number of non-zero entries of a matrix A the **weight** of A and denote it by $\operatorname{wt} A$. For matrices over finite fields the weight of $A - B$ is the most natural way to define the distance between A and B. The weight of a matrix is then the distance from that matrix to the zero matrix. In coding theory the distance between vectors defined in this way is called the

*tmigler@calpoly.edu, kmorriso@calpoly.edu, mogle@calpoly.edu

"Hamming distance,"named after Richard Hamming, a pioneer in the field of error correcting codes.

The **rank** of A can also be regarded as a measure of the distance from A to the zero matrix. In fact it can be used to define the distance between matrices. (The key to showing that the distances defined by either rank or weight satisfy the triangle inequality is that $\operatorname{wt}(A+B) \leq \operatorname{wt} A + \operatorname{wt} B$ and that $\operatorname{rk}(A+B) \leq \operatorname{rk} A + \operatorname{rk} B$. Our fundamental questions is about the relationship between these two measures of distance to the origin in the space of $m \times n$ matrices over $\mathbf{F}_q$.

Sufficient background for sections 1 and 2 is contained in the standard undergraduate courses in linear algebra and abstract algebra. Although the first linear algebra course typically deals with real vector spaces, we expect that the student who has also studied abstract algebra will understand that the fundamental ideas of linear algebra–linear independence, basis, row reduction, rank–are valid for vector spaces over arbitrary fields and in particular over finite fields. The background needed for finite fields is minimal. We recall that there is a unique field (up to isomorphism) for each prime power q, which we denote by $\mathbf{F}_q$. However, nothing is needed about the structure of finite fields, and there is no problem in reading the rest of the paper with only the field $\mathbf{F}_2$ in mind. The third section assumes some probability background through the Central Limit Theorem and should be found in an upper-division probability course for mathematics majors.

Having restricted our investigation to matrices over finite fields, we restate the fundamental question in this way. Over $\mathbf{F}_q$ how many $m \times n$ matrices of rank k and weight w are there? In probabilistic terms we are asking for distribution of the weight for matrices of rank k. We do not have the complete answer to this question, and it seems we are far from the complete answer, so there is plenty of work left to be done. The main results we have are the average value of the weight with the rank fixed (section 2) and the complete description of the weight distribution for rank one matrices (section 3). In the remainder of the introduction we give some easy preliminary results and develop some background material.

We consider matrices of fixed size $m \times n$.

- The zero matrix is the only matrix of weight 0 and the only matrix of rank 0.
- Any matrix of weight 1 has rank 1.
- A matrix of rank k has weight at least k.

It is possible to count the matrices of very small rank and weight. For a matrix of rank 1 and weight 1 we choose a row and column in which to place one of the $q-1$ non-zero elements of the field. That gives us $mn(q-1)$ matrices of rank 1 and weight 1. We leave two more results as exercises for the reader.

- The number of rank 1 and weight 2 is

$$\frac{1}{2}mn(m+n-2)(q-1)^2.$$

- The number of rank 2 and weight 2 is

$$\frac{1}{2}mn(m-1)(n-1)(q-1)^2.$$

The number of matrices of weight w (regardless of rank) is easy to count. There are w locations to select and in each location there are $q-1$ non-zero elements to choose. Therefore, the number of $m \times n$ matrices of weight w is

$$\binom{mn}{w}(q-1)^w.$$

The probability that a matrix has weight w, assuming that each matrix is equally probable, is then

$$\frac{1}{q^{mn}}\binom{mn}{w}(q-1)^w = \binom{mn}{w}(1-1/q)^w(1/q)^{mn-w}.$$

The weight has a binomial distribution with parameters mn and $1-1/q$, which means that it is the same as the distribution of the number of heads in mn tosses of a biased coin with probability of heads being $1-1/q$.

The number of $m \times n$ matrices of rank k (regardless of weight) is more difficult to count. According to Lidl and Niederreiter [4] the result was first

proved by Landsberg for q prime in 1893. Let V be a k-dimensional subspace of the n-dimensional space $\mathbf{F}_q^n$. We can identify the matrices whose column space is V with the $k \times n$ matrices of rank k by using a basis for V. Hence, the number of $m \times n$ matrices of rank k is the product of the number of k-dimensional subspaces of $\mathbf{F}_q^m$ with the number of $k \times n$ matrices of rank k. Formulas 1 and 2 express those two results, which are then put together in Formula 3 for the number of matrices of rank k.

Formula 1 *The number of $k \times n$ matrices of rank k is*

$$\prod_{0 \le i \le k-1} (q^n - q^i) = (q^n - 1)(q^n - q) \cdots (q^n - q^{k-1}).$$

Proof The k rows must be linearly independent vectors of length n. The first row can be any non-zero vector; there are $q^n - 1$ such vectors. The second row must be independent of the first row. That means it cannot be any of the q scalar multiples of that row, but any other row vector is allowed. There are $q^n - q$ vectors to choose from. The third row can be any vector not in the span of the first two rows. There are q^2 linear combinations of the first two rows, and so there are $q^n - q^2$ possible vectors for row 3. We continue in this way with row $i + 1$ not allowed to be any of the q^i linear combinations of the first i rows. □

Now for the number of k-dimensional subspaces of a vector space of dimension m, we count the number of bases of all such subspaces and then divide by the number of bases that each subspace has. A basis is an ordered list of k linearly independent vectors lying in $\mathbf{F}_q^m$. Putting them into a $k \times m$ as the columns what we get is a matrix of rank k. We can use the formula above (with n replaced by m) to see that there are $\prod_{0 \le i \le k-1}(q^m - q^i)$ bases. Each subspace, however, is represented by multiple bases. In particular, the number of bases of a k-dimensional space is just the number of $k \times k$ matrices of rank k. Again, we use the formula (with n replaced by k) to see that there are $\prod_{0 \le i \le k-1}(q^k - q^i)$ bases of a particular subspace.

Formula 2 *The number of k-dimensional subspaces of an m-dimensional*

vector space over $\mathbf{F}_q$ *is*

$$\frac{\prod_{0\le i\le k-1}(q^m - q^i)}{\prod_{0\le i\le k-1}(q^k - q^i)}.$$

There is a well-developed analogy in the world of combinatorics between the subsets of a finite set and the subspaces of a finite dimensional vector space over a finite field. The number of k-dimensional subspaces of an m-dimensional vector space is analogous to the number of subsets of size k in a set of size m, which is given by the binomial coefficient $\binom{m}{k}$. So, we let

$$\binom{m}{k}_q$$

denote the number the number of k-dimensional subspaces of an m-dimensional vector space over $\mathbf{F}_q$ as given in Formula 2. This number is often called a **Gaussian binomial coefficient**. The full development of the subset-subspace analogy is not necessary for us, but an introductory survey can be found in [3].

With these formulas we have the two factors we need for the number of $m \times n$ matrices of rank k.

Formula 3 *The number of* $m \times n$ *matrices of rank* k *is*

$$\frac{\prod_{0\le i\le k-1}(q^n - q^i)\prod_{0\le i\le k-1}(q^m - q^i)}{\prod_{0\le i\le k-1}(q^k - q^i)}.$$

As one should expect the formula is symmetric in m and n.

2 The Average Weight of Rank k Matrices

As we have mentioned, the weight of a matrix A is the distance between A and the zero matrix, and we expect that in some way increasing weight is correlated with increasing rank. In this section we determine the average weight of the set of matrices of a fixed rank in terms of the the key parameters q, m, n, and k. Indeed we find that the average weight grows with k when the other parameters are held fixed.

We consider the weight as a random variable W, which is the sum $\sum_{i,j} W_{ij}$, where W_{ij} is the weight of the i,j entry. Hence, $W_{ij} = 1$ for a matrix whose i,j entry is non-zero and $W_{ij} = 0$ when the entry is 0. Then the average or expected value of W is the sum of the expected values of the W_{ij}. The expected value of W_{ij} is simply the probability that the i,j entry is non-zero. It is this probability that we will compute. An important observation is that this probability is the same for all i and j. In other words, the W_{ij} are identically distributed.

Theorem 1 *For $m \times n$ matrices of rank k the probability that the i,j entry is non-zero is the same for all i and j.*

Proof For a fixed row index i and column index j there is a bijection on the space of $m \times n$ matrices defined by switching row 1 with row i and switching column 1 with column j. This bijection preserves the rank and weight, and so it defines a bijective correspondence between the subset of matrices of rank k with non-zero 1,1 entry and the subset of matrices of rank k with non-zero i,j entry. □

With this result we know that the expected value of W is mn times the average weight of the 1,1 entry, so that we can focus our attention on the upper left entry. Our analysis depends on the reduced row echelon form. The definition can be found in any introductory linear algebra text. For completeness we reproduce the definition in Lay's book [2]. The leading entry of a row means the leftmost non-zero entry.

A rectangular matrix is in **reduced row echelon form** if it has the following properties:

1. All non-zero rows are above any rows of all zeros.
2. Each leading entry of a row is in a column to the right of the leading entry of the row above it.
3. All entries in a column below a leading entry are zero.
4. The leading entry in each non-zero row is 1.

5. Each leading 1 is the only non-zero entry in its column.

When an $m \times n$ matrix A of rank k is reduced to reduced row echelon form (by a sequence of elementary row operations) the result is an $m \times n$ matrix whose first k rows form a basis of the row space of A, and whose remaining $m-k$ rows are zero. Let R be the $k \times n$ matrix consisting of these first k rows. Then there is a unique $m \times k$ matrix C such that $A = CR$. The entries of C are the coefficients needed to express the rows of A as linear combinations of the rows of R. Note that C has rank k. Matrices A and B are said to be **row equivalent** if one can transform A into B by a sequence of elementary row operations. Other characterizations of row equivalence are:

- The reduced row echelon forms of A and B are identical.
- There is an invertible matrix P such that $A = PB$.
- The null spaces of A and B are identical.
- The row spaces of A and B are identical.

With the last description of row equivalence we see that there is a bijection between the $k \times n$ matrices of rank k in reduced row echelon form and the subspaces of dimension k in an n-dimensional vector space.

Using the factorization $A = CR$ we are able to express the set of rank k matrices as the Cartesian product of the set of $m \times k$ matrices of rank k with the set of $k \times n$ reduced row echelon matrices of rank k. This means that A can be selected randomly by independently choosing the factors C and R. Now the 1,1 entry of A is given by $a_{11} = c_{11}r_{11} + c_{12}r_{21} + \cdots + c_{1k}r_{k1}$. Since R is in reduced row echelon form, r_{11} is 0 or 1 and the rest of the first column $r_{21}, r_{31}, \ldots, r_{k1}$ are all 0. Therefore, $a_{11} = c_{11}r_{11}$, and so the probability that the 1,1 entry is non-zero is

$$\mathbf{P}(a_{11} \neq 0) = \mathbf{P}(c_{11} \neq 0)\mathbf{P}(r_{11} \neq 0).$$

Now C is $m \times k$ and has rank k. Thus, the first column of C is any non-zero vector of length m, of which there are $q^m - 1$. There are $q^{m-1} - 1$

of those vectors that have a zero in the top entry, and so there are $q^m - q^{m-1}$ that have a non-zero top entry. Then

$$\mathbf{P}(c_{11} \neq 0) = \frac{q^m - q^{m-1}}{q^m - 1}.$$

The choice of the reduced matrix R is the same as the choice of row space of A. If any of the vectors in the row space have a non-zero first entry, then the first column cannot be the zero column and then r_{11} is not 0. In order that $r_{11} = 0$ the row space of A must be entirely within the $n-1$ dimensional subspace of vectors of the form $(0, x_2, x_3, \ldots, x_n)$. The probability of that occurring is the ratio of the number of k-dimensional subspaces of a space of dimension $n-1$ to the number of k-dimensional subspaces of a space of dimension n:

$$\mathbf{P}(r_{11} = 0) = \frac{\binom{n-1}{k}_q}{\binom{n}{k}_q}.$$

Therefore, the complementary probability gives

$$\mathbf{P}(r_{11} \neq 0) = 1 - \frac{\binom{n-1}{k}_q}{\binom{n}{k}_q}.$$

Using Formula 2 we simplify this to get

$$\mathbf{P}(r_{11} \neq 0) = \frac{q^n - q^{n-k}}{q^n - 1}.$$

Putting these results together gives us

$$\mathbf{P}(a_{11} \neq 0) = \left(\frac{q^m - q^{m-1}}{q^m - 1}\right)\left(\frac{q^n - q^{n-k}}{q^n - 1}\right).$$

As a probability this is more easily analyzed in the following form:

$$\mathbf{P}(a_{11} \neq 0) = \frac{(1 - 1/q)(1 - 1/q^k)}{(1 - 1/q^m)(1 - 1/q^n)}.$$

For matrices with no condition on the rank the probability that a particular entry is non-zero is $1 - 1/q$. We see that as this is the approximately the case for m, n, and k large. One case of interest is that of invertible matrices. For

$n \times n$ invertible matrices we have $m = n$ and $k = n$, and so the probability that an entry is non-zero simplifies to

$$\frac{1 - 1/q}{1 - 1/q^n}$$

and so we see that it is slightly more likely that an invertible matrix has non-zero entries than an arbitrary matrix.

Having determined the probability that the 1,1 entry is non-zero and hence that the probability that the i, j entry is non-zero, we have proved the following theorem.

Theorem 2 *The average weight of an $m \times n$ matrix of rank k over the field of order q is*

$$mn\frac{(1 - 1/q)(1 - 1/q^k)}{(1 - 1/q^m)(1 - 1/q^n)}.$$

We also see that with m and n fixed the average weight increases as k increases. It is this formula that best expresses the intuitive idea that increasing rank is correlated with increasing weight.

3 The Weight of Rank One Matrices

We are able to analyze more completely the weight distribution for matrices of rank one. From Theorem 2 with $k = 1$ we see that the average weight of a rank one matrix is

$$mn\frac{(1 - 1/q)^2}{(1 - 1/q^m)(1 - 1/q^n)}.$$

For m and n large this average is just about $mn(1 - 1/q)^2$, whereas the average weight for all $m \times n$ matrices is $mn(1 - 1/q)$, and so rank 1 matrices tend to have a lot more zero entries than the average matrix. For $q = 2$ this effect is the most pronounced. On average one fourth of the entries are 1 in a large rank one matrix over $\mathbf{F}_2$, while an average of half the entries are 1 and half are 0 for all matrices.

In the factorization $A = CR$, where $\operatorname{rk} A = 1$, C is a non-zero column vector of length m and R is a non-zero row vector of length n whose leading

non-zero entry is 1. The entries of A are given by $a_{ij} = c_i r_j$, and so the weight of A is the product of the weights of C and R. The weight of C has a binomial distribution conditioned on the weight being positive

$$\mathbf{P}(\mathrm{wt}\, C = \mu) = \frac{\binom{m}{\mu}(q-1)^\mu q^{m-\mu}}{(q^m - 1)}$$

Likewise for R the weight distribution is given by

$$\mathbf{P}(\mathrm{wt}\, R = \nu) = \frac{\binom{n}{\nu}(q-1)^\nu q^{n-\nu}}{(q^n - 1)}$$

To select a random R, choose a random non-zero vector of length n and then scale it to make the leading non-zero entry 1. The scaling does not change the weight.

Immediately we see that there is a restriction on the possible weight of a matrix of rank one. For example, for 3×4 matrices the weight cannot be $5, 7, 10, or 11$ because those numbers are not products $\mu\nu$ with $1 \leq \mu \leq 3$ and $1 \leq \nu \leq 4$. All other weights between 1 and 12 are possible.

The weight of rank one matrices is the product of these two binomial random variables, each conditioned to be positive.

$$\begin{aligned}
\mathbf{P}(\mathrm{wt}\, A = \omega) &= \sum_{\mu\nu=\omega} \mathbf{P}(\mathrm{wt}\, C = \mu)\mathbf{P}(\mathrm{wt}\, R = \nu) \\
&= \sum_{\mu\nu=\omega} \binom{m}{\mu}\binom{n}{\nu} \frac{(q-1)^{m+n-\mu-\nu} q^{\mu+\nu}}{(q^m-1)(q^n-1)}
\end{aligned}$$

Because not all weights between 1 and mn occur for rank 1 matrices, plots of actual probability distributions show spikes and gaps. However, the plots of cumulative distributions are smoother and lead us to expect a limiting normal distribution as the size of the matrices goes to infinity. Figures 1 and 2 show this behavior quite well. (In order to plot the approximating normal distribution we numerically computed the standard deviation of the weight distribution for the given m, n, and q.)

Theorem 3 *As m or n goes to infinity, the weight distribution of rank one matrices approaches a normal distribution.*

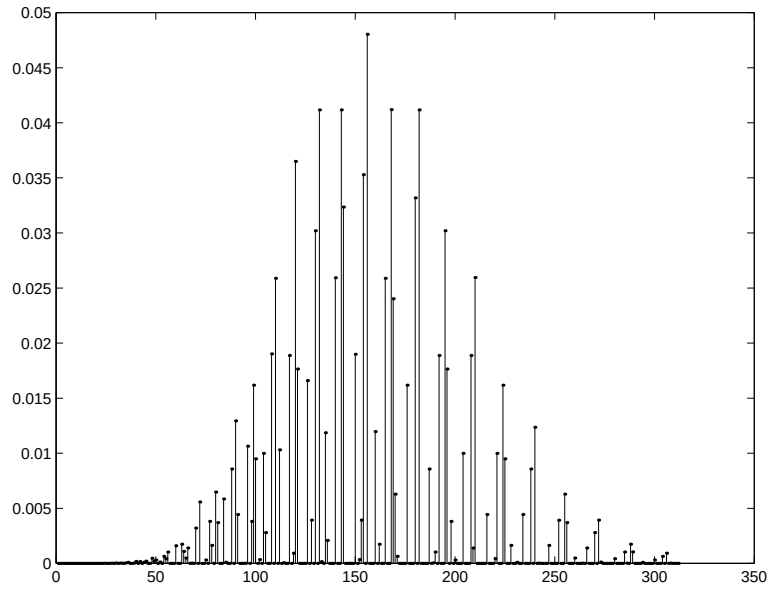

Figure 1: Distribution for the weight of rank 1 matrices, $m = n = 25$, $q = 2$.

Proof The weight random variable for rank one matrices of size $m \times n$ is the product of independent binomial random variables conditioned on being positive. Define $W = XY$, where $X = \sum_{1 \leq i \leq m} X_i$, $Y = \sum_{1 \leq j \leq n} Y_j$, and X_i and Y_j are independent Bernoulli random variables with probability $1/q$ of being 0. Then W is the sum of m independent identically distributed random variables $X_i Y$. Conditioning W on $W > 0$ is the weight of rank one matrices. By the Central Limit Theorem the distribution of W converges, as $m \to \infty$, to a normal distribution after suitable scaling. Now conditioning on W being positive does not change this result because the probability that $W > 0$ is $1 - q^{-m}$, which goes to 1 as $m \to \infty$. □

Therefore, when m and n are large we can use a normal distribution of mean $\mathbf{E}(W)$ and variance $\mathrm{var}\,(W)$ to approximate the weight distribution for rank 1 matrices. Note that this variance is not exactly the variance of the weight of rank 1 matrices because we have not conditioned on W being positive. However, the exact computation of that variance is rather complicated, and because of the theorem it is not any more illuminating

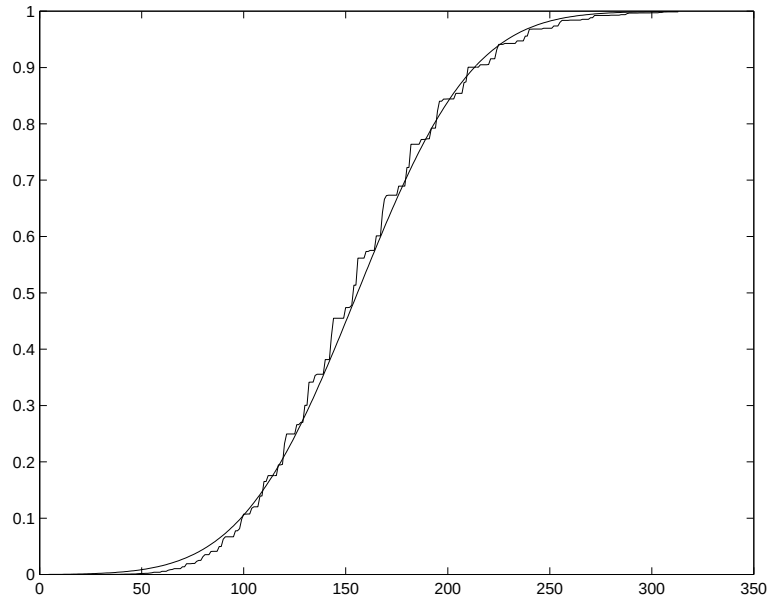

Figure 2: Cumulative frequency distribution for the weight of rank 1 matrices, $m = n = 25$, $q = 2$. The smooth curve is the normal cdf with the same mean ($\approx mn/4 = 156.25$) and standard deviation (≈ 44.63).

than the variance of the unconditioned random variable W. The variance of W is given by

$$\mathrm{var}\,(W) = \mathbf{E}(W^2) - \mathbf{E}(W)^2.$$

Since X and Y are independent binomial random variables,

$$\mathbf{E}(W) = \mathbf{E}(XY) = \mathbf{E}(X)\mathbf{E}(Y) = m(1-1/q)n(1-1/q) = mn(1-1/q)^2.$$

And we have

$$\mathbf{E}(W^2) = \mathbf{E}(X^2Y^2) = \mathbf{E}(X^2)\mathbf{E}(Y^2).$$

Then using the independence of the X_i and the fact that $X_i^2 = X_i$ we get

$$\begin{aligned}
\mathbf{E}(X^2) &= \mathbf{E}\left(\left(\sum X_i\right)^2\right) \\
&= \sum_i \mathbf{E}(X_i) + \sum_{i\neq j} \mathbf{E}(X_i)\mathbf{E}(X_j) \\
&= m(1-1/q) + m(m-1)(1-1/q)^2.
\end{aligned}$$

Likewise,

$$\begin{aligned}\mathbf{E}(Y^2) &= \mathbf{E}\left(\left(\sum Y_i\right)^2\right) \\ &= \sum_i \mathbf{E}(Y_i) + \sum_{i\neq j} \mathbf{E}(Y_i)\mathbf{E}(Y_j) \\ &= n(1-1/q) + n(n-1)(1-1/q)^2.\end{aligned}$$

From these it follows that

$$\begin{aligned}\mathbf{E}(W^2) &= mn\left(1-\frac{1}{q}\right)^2 + mn(m+n-2)\left(1-\frac{1}{q}\right)^3 \\ &\quad + m(m-1)n(n-1)\left(1-\frac{1}{q}\right)^4.\end{aligned}$$

Finally, the variance of W, which is $\mathbf{E}(W^2) - \mathbf{E}(W)^2$ can be simplified to give

$$\begin{aligned}\mathrm{var}\,(W) &= mn(1-n-m)\left(1-\frac{1}{q}\right)^4 \\ &\quad + mn(n+m-2)\left(1-\frac{1}{q}\right)^3 \\ &\quad + mn\left(1-\frac{1}{q}\right)^2\end{aligned}$$

From this we can see, for example, that for $m \approx n$, $m, n \to \infty$, the variance is of order n^3, and so the standard deviation is of order $n^{3/2}$.

Specializing to square matrices $(n \times n)$ over $\mathbf{F}_2$, the variance is

$$\frac{n^3}{8} + \frac{n^2}{16},$$

and so as n grows the standard deviation is asymptotic to $(n/2)^{3/2}$. In the example shown in Figures 1 and 2 the standard deviation of the actual weight distribution of rank one matrices was computed and turns out to be 44.63 (rounded to two places). The value of $(n/2)^{3/2}$ with $n = 25$ is 44.19 (also rounded to two places).

4 Further Questions

Analyzing the CR factorization for rank 2 matrices should conceivably allow us to find the weight distribution for rank 2, but the analysis is considerably more difficult, and for higher ranks the difficulty continues to increase. This suggests gathering some information by simulation, leading to the question of efficiently generating random matrices of fixed rank. In [1] there is a treatment of the related problem of randomly generating a subspace of fixed dimension over a finite field, and Wilf has suggested to us that a random rank k matrix could be generated by adding together k rank 1 matrices, which are easy to generate, and then keeping those of rank k. Alternatively, one might use the CR factorization. Selecting R is exactly the subspace selection problem in [1]. Selecting C can be done by generating a random $m \times k$ matrix and keeping those of rank k. Which approach is more efficient we leave as an open question, as well as the question of whether there are even better ways to generate matrices of a fixed rank.

We have focused on the weight of fixed rank matrices, but it would be interesting to look at the rank of fixed weight matrices. As an example, consider the $n \times n$ matrices of weight n. Those of rank 1 we have counted in section 2 and the result is in terms of the divisors of n. Those of rank n are generalizations of permutation matrices and there are $n!(q-1)^n$ of them. What about the other ranks? In particular, how many $n \times n$ matrices of weight n and rank $n-1$ are there over $\mathbf{F}_2$?

Since the weight of A is the distance from A to 0, it plays a role analogous to the norm of a real or complex matrix. In both cases it is the distance to the only matrix of rank 0. Now we may ask for the distance from A to the subset of matrices of rank 1, that is for the minimal distance from A to some matrix of rank 1. In general we may ask for the distance from A to the matrices of rank k. For real or complex matrices these distances (using the linear map norm) are given by the **singular values** and can easily be computed [2]. For matrices over finite fields can these distances (define by the weight) be computed in any other way than by exhaustive search?

References

[1] E. Calabi and H. S. Wilf. On the sequential and random selection of subspaces over a finite field. *J. Combinatorial Theory (A)*, **22** (1977) 107–109; MR **55** #5649

[2] D. C. Lay. *Linear Algebra and Its Applications*, second edition. Addison-Wesley, Reading, Massachusetts, 1997.

[3] J. P. S. Kung, The subset-subspace analogy, in *Gian-Carlo Rota on combinatorics*, 277–283, Birkhäuser, Boston, Boston, MA, 1995; MR 99b:01027

[4] R. Lidl and H. Niederreiter. *Finite Fields*, second edition. Cambridge Univ. Press, Cambridge, UK, 1997; MR 97i:11115