Random Matrices and Erasure Robust Frames

Yang Wang

Abstract  Data erasure can often occur in communication. Guarding against erasures involves redundancy in data representation. Mathematically this may be achieved by redundancy through the use of frames. One way to measure the robustness of a frame against erasures is to examine the worst case condition number of the frame with a certain number of vectors erased from the frame. The term numerically erasure-robust frames was introduced in Fickus and Mixon (Linear Algebra Appl 437:1394–1407, 2012) to give a more precise characterization of erasure robustness of frames. In the paper the authors established that random frames whose entries are drawn independently from the standard normal distribution can be robust against up to approximately 15 % erasures, and asked whether there exist frames that are robust against erasures of more than 50 %. In this paper we show that with very high probability random frames are, independent of the dimension, robust against erasures as long as the number of remaining vectors is at least $1 + \delta_0$ times the dimension for some $\delta_0 > 0$. This is the best possible result, and it also implies that the proportion of erasures can be arbitrarily close to 1 while still maintaining robustness. Our result depends crucially on a new estimate for the smallest singular value of a rectangular random matrix with independent standard normal entries.

Keywords  Random matrices · Singular values · Numerically erasure robust frame (NERF) · Condition number · Restricted isometry property

Mathematics Subject Classification  Primary 42C15

Communicated by Hans G. Feichtinger.

Yang Wang
yangwang@hkust.hk

1 Department of Mathematics, The Hong Kong University of Science and Technology, Clear Water Bay, Kowloon, Hong Kong
1 Introduction

Let $\mathbf{H}$ be a Hilbert space. A set of elements $\mathcal{F} = \{f_j\}$ in $\mathbf{H}$ (counting multiplicity) is called a frame if there exist two positive constants $C_*$ and $C^*$ such that for any $v \in \mathbf{H}$ we have
\[
C_* \|v\|^2 \leq \sum_j |\langle v, f_j \rangle|^2 \leq C^* \|v\|^2.
\]
In this way, the constants $C_*$ and $C^*$ are called the lower frame bound and the upper frame bound, respectively. A frame is called a tight frame if $C_* = C^*$. In this paper we focus mostly on real finite dimensional Hilbert spaces with $\mathbf{H} = \mathbb{R}^n$ and $\mathcal{F} = \{f_j\}_{j=1}^N$, although we shall also discuss the extendability of the results to the complex case. Let $\mathcal{F} = [f_1, f_2, \ldots, f_N]$. It is called the frame matrix for $\mathcal{F}$. It is well known that $\mathcal{F}$ is a frame if and only if the $n \times N$ matrix $\mathcal{F}$ has rank $n$. Furthermore, the optimal frames bounds are given by
\[
C_* = \sigma_1^2(F), \quad C^* = \sigma_n^2(F),
\]
where $\sigma_1 \geq \sigma_2 \geq \cdots \sigma_n > 0$ are the singular values of $F$. Throughout this paper we shall identify without loss of generality a frame $\mathcal{F}$ by its frame matrix.

The main focus of the paper is on the erasure robustness property for a frame. This property arise in applications such as communication where data can be lost or corrupted in the process of transmission. Suppose that we have a frame $\mathcal{F}$ that is full spark in the sense that every $n$ columns of $\mathcal{F}$ span $\mathbb{R}^n$, it is theoretically possible to erase up to $N - n$ data from the full set of data $\{\langle v, f_j \rangle\}_{j=1}^N$ while still reconstruct the signal $v$. This is a simple consequence of the property that with the remaining available data $\{\langle v, f_j \rangle\}_{j \in S}$ with $|S| \geq n$, $v$ is uniquely determined because $\text{span}(\{f_j : j \in S\}) = \mathbb{R}^n$. In practice, however, the condition number of the matrix $[f_j]_{j \in S}$ could be so poor that the reconstruction is numerically unstable against the presence of additive noise in the data. Thus robustness against data loss and erasures is a highly desirable property for a frame. There have been a number of studies that aim to address this important issue.

Among the first studies of erasure-robust frames was given in [10]. It was shown in subsequent studies that that unit norm tight frames are optimally robust against one erasure [6] while Grassmannian frames are optimally robust against two erasures [11,16]. The literature on erasure robustness for frames is quite extensive, see e.g. also [12,13,18]. In general, the robustness of a frame $\mathcal{F}$ against $q$-erasures, where $q \leq N - n$, is measured by the maximum of the condition numbers of all $n \times (N - q)$ submatrices of $\mathcal{F}$. More precisely, let $S \subseteq \{1, 2, \ldots, N\}$ and let $F_S$ denote the $n \times |S|$ submatrix of $\mathcal{F}$ with columns $f_j$ for $j \in S$ (in its natural order, although the order of the columns is irrelevant). Then the robustness against $q$-erasures of $\mathcal{F}$ is measured by
\[
R(\mathcal{F}, q) := \max_{|S| = N - q} \frac{\sigma_1(F_S)}{\sigma_n(F_S)}.
\]
$F$ is $(K, \alpha, \beta)$-NERF if

$$\alpha \leq \sigma_n(F_S) \leq \sigma_1(F_S) \leq \beta \quad \text{for any } S \subseteq \{1, 2, \ldots, N\}, \ |S| = K.$$

Thus in this case $R(F, N - K) \leq \beta/\alpha$. Note that for any full spark $n \times N$ frame matrix $F$ and any $n \leq K \leq N$ there always exist $\alpha, \beta > 0$ such that $F$ is $(K, \alpha, \beta)$-NERF. The main goal is to find classes of frames where the bounds $\alpha, \beta$, and more importantly, $R(F, N - K) = \beta/\alpha$, are independent of the dimension $n$ while allowing the proportion of erasures $1 - \frac{K}{N}$ as large as possible. The authors studied in [9] the erasure robustness of $F = \frac{1}{\sqrt{n}} A$, where the entries of $A$ are independent random variables of the standard normal $\mathcal{N}(0, 1)$ distribution. It was shown that with high probability such a matrix can be good NERFs provided that $K$ is no less than approximately $85\%$ of $N$. The authors also proved that equiangular frame $F$ in $\mathbb{C}^n$ with $N = n^2 - n + 1$ vectors is a good NERF against up to about $50\%$ erasures. As far as the proportion of erasures is concerned this was the best known result for NERFs. However, the frame requires almost $n^2$ vectors. The authors posed as an open question whether there exist NERFs with $K < N/2$. A more recent paper [8] explored a deterministic construction based on certain group theoretic techniques. The approach offers more flexibility in the frame design than the far more restrictive equiangular frames.

In this paper we revisit the robustness of random frames. We provide a much stronger result for Gaussian random frames. Set $\lambda = \frac{N}{n}$. This is the aspect ratio of the matrix $A$, and for a frame we always have $\lambda \geq 1$. The question raised by Fickus and Mixon is whether we can reduce the aspect ratio of by more than a half and still maintain erasure robustness. Our main result proves that for random Gaussian frames one can reduce the aspect ratio to arbitrarily close to $1$ while still maintaining erasure robustness with high probability. In other words, for any $0 < \delta \leq \lambda$, with very high probability the frame $F = \frac{1}{\sqrt{n}} A$ is a $((1 + \delta)n, \alpha, \beta)$-NERF where $\alpha, \beta$ depend only on $\delta$ and $\lambda$. One version of our result is given by the following theorem.

**Theorem 1.1** Let $F = \frac{1}{\sqrt{n}} A$ where $A$ is $n \times N$ whose entries are independent Gaussian random variables of $\mathcal{N}(0, 1)$ distribution. Let $\lambda = \frac{N}{n} > 1$. Then for any $0 < \delta_0 < \lambda - 1$ and $\tau_0 > 0$ there exist $\alpha, \beta > 0$ depending only on $\delta_0, \lambda$ and $\tau_0$ such that the frame $F$ is a $((1 + \delta_0)n, \alpha, \beta)$-NERF with probability at least $1 - 3e^{-\tau_0 n}$.

Theorem 1.1 shows that for random Gaussian frame erasure robustness will be maintained with high probability even as the number of erasures approaches the theoretical limit of $(\lambda - 1)n = N - n$ where $\lambda = N/n$. Thus as a corollary the conjecture in [9] that robustness cannot be achieved with more than $50\%$ erasures is false – It can in fact be arbitrarily close to $100\%$.

Later in the paper we shall provide more explicit estimates for $\alpha, \beta$ that will allow us to easily compute them numerically. Note that our result is essentially the best possible, as we cannot go to $\delta_0 = 0$. A corollary of the theorem is that for random Gaussian frames the proportion of erasures $1 - \frac{K}{N}$ can be made arbitrary close to $1$ while the frames still maintain robustness with overwhelming probability. For example, as we shall see later, for an $n \times N$ random Gaussian frame where $N = 50n$, with probability
at least $1 - 3e^{-0.25n}$, the condition number is no more than 26.52 after 50% erasures and 1462.4 after 90% erasures.

Our theorem depends crucially on a refined estimate on the smallest singular value of a random Gaussian matrix. There is a wealth of literature on random matrices. The study of singular values of random matrices has been particularly intense in recent years due to their applications in compressive sensing for the construction of matrices with the so-called restricted isometry property (see e.g. [1,2,4,5]). Random matrices have also been employed for phase retrieval [3], which aims to reconstruct a signal from the magnitudes of its samples. For a very informative and comprehensive survey of the subject we refer the readers to [15,19], which also contains an extensive list of references (among the notable ones [7,14,17]).

For the $n \times N$ Gaussian random matrix $A$ the expected value of $\sigma_1(A)$ and $\sigma_n(A)$ are asymptotically $\sqrt{N} + \sqrt{n}$ and $\sqrt{N} - \sqrt{n}$, respectively. Many important results, such as the NERF analysis of random matrices in [9] as well as results on the restricted isometry property in compressive sensing, often utilize known estimates of $\sigma_1(A)$ and $\sigma_n(A)$ based on Hoeffding-type inequalities. One classical such estimate that is ubiquitous in the random matrix literature is

$$
P \left( \sigma_n(A) < \sqrt{N} - \sqrt{n} - t \right) \leq e^{-\frac{t^2}{2}},
$$

(1.3) see [19]. The problem with this estimate is that even by taking $t = \sqrt{N} - \sqrt{n}$ we only get a bound of $e^{-((\sqrt{\lambda} - 1)^2 n/2)}$ even though the probability in this case is 0. Another one of the best results in this direction is the estimate

$$
P \left( \sigma_n(A) < \epsilon \left( \sqrt{N} - \sqrt{n} \right) \right) \leq cN + (C\epsilon)^{N-n}
$$

(1.4)

for sub-Gaussian random $n \times N$ matrices, where $c, C > 0$ depend only on the distribution of the sub-Gaussian, see [19]. But again the decay is capped from below by $c^n$.

The problem is that when decay rate is capped it will also limit the maximal rate of erasure for the establishment of erasure robustness. Thus for this study, estimates such as (1.3) and (1.4) that cap the decay rate are inadequate for extending the robustness to larger proportion of erasures. To see why this is case, we consider the case of using (1.3) to establish erasure robustness. Assume for simplicity that we start off with a Gaussian random matrix $A$ that is $n \times N$ with $N = 4n$ and the frame $F = \frac{1}{\sqrt{n}} A$. We want to know whether after the erasure of $2n$ columns we are still left with a robust frame, namely whether $F$ is $(2N, \alpha, \beta)$-NERF for some $\alpha, \beta > 0$ independent of $n$. Thus we will need to show that both

$$
P \left( \min_{|S|=2n} \left( \sigma_n(F_S) < \epsilon \sqrt{n} \right) \right) \quad \text{and} \quad P \left( \max_{|S|=2n} \left( \sigma_1(F_S) > C \sqrt{n} \right) \right)
$$

are small for some $\epsilon, C$ independent of $n$. The above probability bound for $\sigma_1(F_S)$ is much easier to obtain, and in fact (1.3) is sufficient. However, the best we can get
using (1.3) for a given $F_S$ is

$$\mathbb{P} \left( \sigma_n(F_S) < \varepsilon \sqrt{n} \right) \leq e^{- \frac{(\sqrt{N} - \sqrt{n})^2}{2}} = e^{-\frac{n}{2}},$$

because in (1.3) we must have $t \leq \sqrt{N} - \sqrt{n} = \sqrt{4N} - \sqrt{n} = \sqrt{n}$. Now the number of subsets $S \subset \{1, 2, \ldots, N = 4n\}$ with $|S| = 2n$ is $\binom{N}{2n} = \binom{4n}{2n}$, which is approximately $2^{4n}/2\sqrt{n\pi}$ by Stirling’s Formula. The standard union bound technique now yields

$$\mathbb{P} \left( \min_{|S|=2n} \sigma_n(F_S) < \varepsilon \sqrt{n} \right) = \mathbb{P} \left( \sigma_n(F_S) < \varepsilon \sqrt{n} \text{ for all } S \text{ with } |S| = 2n \right) \leq \sum_{|S|=2n} \mathbb{P} \left( \sigma_n(F_S) < \varepsilon \sqrt{n} \right) \leq \left( \frac{4n}{2n} \right) e^{-\frac{n}{2}}.$$

Note that $\left( \frac{4n}{2n} \right) e^{-\frac{n}{2}}$ is approximately $2^{4n} e^{-\frac{n}{2}}/2\sqrt{n\pi}$, which is anything but small. It grows exponentially to infinity. Thus this standard technique will not allow us to establish erasure robustness in this case where the erasure rate is 50 %. This was precisely the difficulty encountered in [9]. It is also the difficulty we try to overcome in this paper.

To go further we must prove an estimate that will allow the probability in (1.3) to decay much faster. We achieve this goal by proving the following theorem:

**Theorem 1.2** Let $A$ be $n \times N$ whose entries are independent random variables of standard normal $N(0, 1)$ distribution. Let $\lambda = \frac{N}{n} > 1$. Then there exists a constant $C_\lambda$ depending only on $\lambda$ such that

$$\mathbb{P} \left( \sigma_n(A) \leq \varepsilon \sqrt{n} \right) \leq 3 \left( C_\lambda \varepsilon \right)^{(\lambda - 1)n} \ln^n \left( \frac{1}{\varepsilon} \right) = 3 \left( C_\lambda \varepsilon \right)^{N-n} \ln^n \left( \frac{1}{\varepsilon} \right).$$

More specifically,

$$\mathbb{P} \left( \sigma_n(A) \leq \varepsilon \sqrt{n} \right) \leq 3 \left( B_\lambda \varepsilon \right)^{N-n} \left( 1 + \sqrt{\lambda} + \sqrt{(\lambda - 1) \ln(1/\varepsilon)} \right)^n$$

where

$$B_\lambda := \frac{\lambda}{\lambda - 1} 2^{\frac{1}{\lambda - 1}} (e\lambda)^{\frac{\lambda}{2(\lambda - 1)}}.$$  

We remark that if we replace the condition $\lambda = \frac{N}{n} > 1$ by $\frac{N}{n} \geq \lambda > 1$ then the first part of the inequality in (1.8) still holds. Namely

$$\mathbb{P} \left( \sigma_n(A) \leq \varepsilon \sqrt{n} \right) \leq 3 \left( C_\lambda \varepsilon \right)^{(\lambda - 1)n} \ln^n \left( \frac{1}{\varepsilon} \right).$$

This is easily seen because $\mathbb{P} \left( \sigma_n(A) \leq \varepsilon \sqrt{n} \right)$ only decreases as we increase $N$. 


Naturally one may ask whether an analogous estimate holds for sub-Gaussian random matrices, e.g. Bernoulli random matrices. If it does then we will be able to establish similar erasure robustness results for sub-Gaussian random frames. Unfortunately the answer is no. An indirect way of looking at it is that for any Bernoulli random matrices 50\% erasure robustness cannot be achieved as observed in [9], because you can always remove no more than half of the columns in any Bernoulli matrix so that the first two rows are linearly dependent. As a result an analogous estimate for sub-Gaussian random matrices cannot hold for all sub-Gaussian distributions. However, it does not preclude the existence of analogous estimates for some sub-Gaussian distributions.

Theorem 1.3 below is actually what we shall first prove, and it will be used to derive Theorem 1.2.

**Theorem 1.3** Let $A$ be $n \times N$ whose entries are independent random variables of standard normal $N(0, 1)$ distribution. Let $\lambda = \frac{N}{n} > 1$. Then for any $\mu > 0$ there exist constants $c, C > 0$ depending only on $\mu$ and $\lambda$ such that

$$
P(c \sqrt{n} \leq \sigma_n(A) \leq \sigma_1(A) \leq C \sqrt{n}) \geq 1 - 3e^{-\mu n}. 
$$

(1.11)

Furthermore, we may take $C = 1 + \sqrt{\lambda} + \sqrt{2\mu}$ and $c = \sup_{0 < t < 1} \varphi(t)$ where

$$
\varphi(t) = \frac{t^2}{L} - \frac{2Ct}{1 - t}, \quad \text{where} \quad L = \sqrt{\frac{e^3 - e^{\mu}}{\lambda}}. 
$$

(1.12)

In particular,

$$
c \geq \left( \frac{1}{2C\lambda L^2} \right)^{\lambda^{-1}} \left( 1 - \frac{1}{\lambda} \right). 
$$

(1.13)

2 Smallest Singular Value of a Random Matrix: Nonasymptotic Estimate

We begin with estimates on the extremal singular values of a random matrix $A$ whose entries are independent standard normal random variables. We shall assume throughout the section that $A$ is $n \times N$ where $\frac{N}{n} = \lambda > 1$. One of the very important estimates is

$$
P(\sigma_1(A) > \sqrt{N} + \sqrt{n} + t) \leq e^{-\frac{t^2}{2}}, 
$$

(2.1)

see [19]. Our main goal of this section is to prove the estimates for smallest singular value $\sigma_n(A)$ stated in Theorem 1.3. An equivalent formulation of (2.1) is

$$
P(\sigma_1(A) > C \sqrt{n}) \leq e^{-\frac{(C-1-\sqrt{\lambda})^2 n}{2}}, \quad C \geq 1 + \sqrt{\lambda}. 
$$

(2.2)

Observe that

$$
\sigma_n(A) = \min_{v \in S^{n-1}} \|A^* v\|,
$$
where $S^{n-1}$ denotes the unit sphere in $\mathbb{R}^n$.

**Lemma 2.1** Let $c > 0$. For any $v \in S^{n-1}$ the probability $\mathbb{P}(\|A^*v\| \leq c)$ is independent of the choice of $v$. We have

$$\mathbb{P}(\|A^*v\| \leq \sqrt{\delta n}) \leq \left(\frac{e\delta}{\lambda}\right)^{\frac{n}{2}}$$

(2.3)

for any $\delta > 0$.

**Proof** The fact that $\mathbb{P}(\|A^*v\| \leq c)$ is independent of the choice of $v$ is a well-known fact, which stems from the fact that the entries of $PA$ are again independent standard normal random variables for any orthogonal $n \times n$ matrix $P$. In particular, one can always find an orthogonal $P$ such that $Pv = e_1$. Thus we may without loss of generality take $v = e_1$. In this case $\|A^*v\|^2 = a_{11}^2 + \cdots + a_{1N}^2$ where $[a_{11}, \ldots, a_{1N}]$ denotes the first row of $A$. Denote $Y_N = a_{11}^2 + \cdots + a_{1N}^2$. Then $Y_N$ has the $\Gamma\left(\frac{n}{2}, \frac{1}{2}\right)$ distribution, which has the density function

$$\rho(t) = \frac{1}{2^{\frac{n}{2}} \Gamma\left(\frac{n}{2}\right)} e^{-\frac{t}{2} \frac{n}{2}} \cdot t^{\frac{n}{2} - 1}, \quad t > 0.$$

Denote $m = \frac{N}{2}$. It follows that

$$\mathbb{P}(\|A^*v\| \leq \sqrt{\delta n}) = \mathbb{P}(Y_N \leq \delta n)$$

$$= \frac{1}{2^m \Gamma(m)} \int_0^{\delta n} e^{-\frac{t}{2} \frac{n}{2}} t^{m-1} dt$$

$$\leq \frac{1}{2^m \Gamma(m)} \int_0^{\delta n} t^{m-1} dt$$

$$= \frac{\delta^m n^m}{2^m m!}.$$

Note that $m! \geq \left(\frac{m}{e}\right)^m$ by Stirling’s formula. The theorem now follows from $\frac{N}{n} = \lambda$ and $m = \frac{N}{2}$. \hfill \Box

The above lemma does not hold for sub-Gaussian random matrices in general, which is why the main results of the paper cannot be generalized to the sub-Gaussian setting. Other arguments used in the paper do work in the sub-Gaussian setting.

A ubiquitous tool in the study of random matrices is an $\varepsilon$-net. For any $\varepsilon > 0$ an $\varepsilon$-net for $S^{n-1}$ is a set in $S^{n-1}$ such that any point on $S^{n-1}$ is no more than $\varepsilon$ distance away from the set. The following result is known and can be found in [19]:

**Lemma 2.2** For any $\varepsilon > 0$ there exists an $\varepsilon$-net $\mathcal{N}_\varepsilon$ in $S^{n-1}$ with cardinality no larger than $(1 + 2\varepsilon^{-1})^n$. 
Proof of Theorem 1.3 Assume that \( \sigma_n(A) = b\sqrt{n} \). Then there exists a \( v_0 \in S^{n-1} \) such that \( \|A^*v_0\| = b\sqrt{n} \). Let \( N_\varepsilon \) be an \( \varepsilon \)-net for \( S^{n-1} \) and take \( u \in N_\varepsilon \) that is the closest to \( v_0 \). So \( \|u - v_0\| \leq \varepsilon \). Thus

\[
\|A^*u\| \leq \|A^*v_0\| + \|A^*(u - v_0)\| \leq b\sqrt{n} + \varepsilon \sigma_1(A). \tag{2.4}
\]

Hence

\[
P(\sigma_n(A) \leq c\sqrt{n}) \leq \sum_{u \in N_\varepsilon} P(\|A^*u\| \leq c\sqrt{n} + \varepsilon \sigma_1(A)). \tag{2.5}
\]

Note that

\[
P(\|A^*u\| \leq c\sqrt{n} + \varepsilon \sigma_1(A)) = P(\|A^*u\| \leq c\sqrt{n} + \varepsilon \sigma_1(A), \sigma_1(A) \leq C\sqrt{n}) + P(\|A^*u\| \leq c\sqrt{n} + \varepsilon \sigma_1(A), \sigma_1(A) > C\sqrt{n}). \]

By Lemma 2.1 the first term on the right hand side is bounded from above by

\[
P(\|A^*u\| \leq c\sqrt{n} + \varepsilon \sigma_1(A), \sigma_1(A) \leq C\sqrt{n}) \leq \left( \frac{e(c + \varepsilon C)^2}{\lambda} \right)^{\frac{N}{2}}. \tag{2.6}
\]

By (2.2) the second term on the right hand side is bounded from above by

\[
P(\|A^*u\| \leq c\sqrt{n} + \varepsilon \sigma_1(A), \sigma_1(A) > C\sqrt{n}) \leq e^{-\frac{(c-1-\sqrt{\lambda})^2}{2} n}. \tag{2.7}
\]

Thus combining these two upper bounds we obtain the estimate

\[
P(\sigma_n(A) \leq c\sqrt{n}) \leq \left( 1 + \frac{2}{e} \right)^n \left( \left( \frac{e(c + \varepsilon C)^2}{\lambda} \right)^{\frac{N}{2}} + e^{-\frac{(c-1-\sqrt{\lambda})^2}{2} n} \right). \tag{2.6}
\]

We would like to bound \( P(\sigma_n(A) \leq c\sqrt{n}) \) by \( 2e^{-\mu n} \). All we need then is to choose \( \varepsilon, c, C > 0 \) so that both upper bound terms in (2.6) are bounded by \( e^{-\mu n} \). Note that \( \frac{N}{2} = \frac{1}{2} n \). Hence we only need

\[
-\mu \geq \ln(1 + 2\varepsilon^{-1}) + \frac{\lambda}{2} \left( 1 - \ln \lambda + 2 \ln(c + \varepsilon C) \right), \tag{2.7}
\]

\[
-\mu \geq -\frac{1}{2} (C - 1 - \sqrt{\lambda})^2. \tag{2.8}
\]

The equation (2.8) leads to the condition

\[
C \geq \sqrt{2\mu} + \sqrt{\lambda} + 1. \tag{2.9}
\]
To meet condition (2.7) we set $c = r\varepsilon$. Then $\ln(c + \varepsilon C) = -\ln \varepsilon^{-1} + \ln(r + C)$. Thus (2.7) becomes

$$(\lambda - 1) \ln(\varepsilon^{-1}) \geq \mu + \ln(2 + \varepsilon) + \frac{\lambda}{2} \ln\left(\frac{(r + C)^2}{\lambda}\right). \quad (2.10)$$

Clearly, once we fix $C$ and $r$, say, take $C = \sqrt{2\mu} + \sqrt{\lambda} + 1$ and $r = 1$, $\ln \varepsilon^{-1}$ will be greater than the right hand side of (2.10) for small enough $\varepsilon$ because of the condition $\lambda > 1$. Both $C$, $c$ only depend on $\lambda$ and $\mu$. The existence part of the theorem is thus proved.

While we have already a good explicit estimate $C = \sqrt{2\mu} + \sqrt{\lambda} + 1$, it remains to establish the explicit formula for $c$. For any fixed $r$ the largest $\varepsilon$ is achieved when (2.10) is an equality, namely

$$(\lambda - 1) \ln(\varepsilon^{-1}) = \mu + \ln(2 + \varepsilon) + \frac{\lambda}{2} \ln\left(\frac{(r + C)^2}{\lambda}\right),$$

which one can rewrite as

$$\ln(r + C) = -(1 - p) \ln \varepsilon - p \ln(2 + \varepsilon) - \ln L,$$

where $p = \lambda^{-1}$ and $L = \sqrt{\frac{\mu}{\lambda}} e^\frac{\mu}{\lambda}$. It follows that

$$r\varepsilon = \frac{1}{L} \left(\frac{\varepsilon}{2 + \varepsilon}\right)^p - C\varepsilon = \frac{1}{L} \frac{1}{t^\frac{1}{\lambda}} - \frac{2Ct}{1 - t},$$

where $t = \frac{\varepsilon}{2 + \varepsilon}$. Note that $0 < t < 1$. Now we can take $c$ to be the supreme value of $r\varepsilon$, which yields

$$c = \sup_{0 < t < 1} \left\{ \frac{1}{L} \frac{1}{t^\frac{1}{\lambda}} - \frac{2Ct}{1 - t} \right\}. \quad (2.11)$$

Finally, (1.11) follows from $P\left(\sigma_n(A) \leq c\sqrt{n}\right) \leq 2e^{-\mu n}$ and (2.2). The proof of the theorem is now complete.

**Remark** Although there does not seem to exist an explicit formula for $c$ given in (2.11), there is a very good explicit approximation of it. In general, the $t$ that maximize $\varphi(t)$ is rather small. So we may approximate $\frac{2Ct}{1 - t}$ simply by $2Ct$ and find the maximum of

$$\tilde{\varphi}(t) = \frac{1}{L} \frac{1}{t^\frac{1}{\lambda}} - 2Ct. \quad (2.12)$$

The maximum of $\tilde{\varphi}(t)$ is obtained at $t_0 = (2C\lambda L)^{-\frac{1}{\lambda - 1}}$. This $t_0$ is very close to the actual $t$ that maximizes $\varphi(t)$. Thus

$$\tilde{c} := \varphi(t_0) = \left(\frac{1}{2C\lambda L^\frac{1}{\lambda}}\right)^{\frac{1}{\lambda - 1}} \left(1 - \frac{1}{\lambda}\right) \quad (2.13)$$
has $\tilde{c} \leq c$ and it is a close approximation of the optimal $c$. This leads to the following Corollary:

**Corollary 2.3** Let $A$, $C$, $L$ and $\lambda$ be as in Theorem 1.3 for any given $\mu > 0$. Then

$$\mathbb{P}(\tilde{c}\sqrt{n} \leq \sigma_n(A) \leq \sigma_1(A) \leq C\sqrt{n}) \geq 1 - 3e^{-\mu n} \quad (2.14)$$

where $\tilde{c}$ is given by (2.13).

**Proof of Theorem 1.2** We use Corollary 2.3 and in (2.14) we set $\epsilon = \tilde{c}$ while solving for $\mu$. It follows that

$$\epsilon^{\lambda-1} = \frac{1}{2C\lambda L^2} \left(1 - \frac{1}{\lambda}\right)^{\lambda-1}. \quad (2.15)$$

Substituting in the explicit values of $C$, $L$ from Theorem 1.3 and $B_\lambda$ in (1.10) and simplifying the equation yield

$$\left(1 + \sqrt{\lambda} + \sqrt{\mu}\right) e^{\mu \epsilon^{\lambda-1}} = \frac{1}{2(e\lambda)^{\frac{1}{2}}} \left(1 - \frac{1}{\lambda}\right)^{\lambda-1} = B_\lambda^{1-\lambda}. \quad (2.15)$$

Thus

$$e^{\mu} < \left(1 + \sqrt{\lambda} + \sqrt{2\mu}\right) e^{\mu} = \left(B_\lambda \epsilon\right)^{1-\lambda}$$

and since $B_\lambda > 1$,

$$\mu < (\lambda - 1) \ln \left(\frac{1}{\epsilon}\right) - (\lambda - 1) \ln B_\lambda \leq (\lambda - 1) \ln \left(\frac{1}{\epsilon}\right).$$

Consequently

$$e^{-\mu} = (1 + \sqrt{\lambda} + \sqrt{2\mu}) \epsilon^{\lambda-1} B_\lambda^{\lambda-1} < \left(1 + \sqrt{\lambda} + \sqrt{(\lambda - 1) \ln (1/\epsilon)}\right) \epsilon^{\lambda-1} B_\lambda^{\lambda-1}.$$ 

Since $(\lambda - 1)n = N - n$, (1.9) of the theorem now follows. Clearly (1.8) follows from (1.9). \qed

Although Theorem 1.3 is for real Gaussian random matrices, a complex version of it can also be proved with minor modifications. A complex random variable $Z = X + iY$ has the complex standard normal distribution if $X$ and $Y$ are independent and both have real normal distribution $\mathcal{N}(0, 1)$. Theorem 1.3 extends to the following theorem for the complex case:
Theorem 2.4 Let $A$ be an $n \times N$ matrix whose entries are independent random variables of complex standard normal $N(0, 1)$ distribution. Let $\lambda = \frac{N}{n} > 1$. Then for any $\mu > 0$ there exist constants $c, C > 0$ depending only on $\mu$ and $\lambda$ such that

$$
P \left( c \sqrt{n} \leq \sigma_n(A) \leq \sigma_1(A) \leq C \sqrt{n} \right) \geq 1 - 3e^{-\mu n}. \quad (2.16)
$$

Furthermore, we may take $C = \sqrt{2} + 2\sqrt{\lambda} + 2\sqrt{\mu}$ and $c = \sup_{0 < t < 1} \phi(t)$ where

$$
\phi(t) = \frac{t^\frac{1}{2}}{L} - \frac{2Ct}{1 - t}, \quad \text{where} \quad L = \sqrt{e - e^{\frac{\mu}{\lambda}}}.
$$

(2.17)

Proof The proof follows the same argument as in the real case so we only sketch the proof here. In particular we point out the places where the estimates need to be modified.

Write $A = AR + iA_I$ and set $B = [AR, A_I]$. Then $B$ is an $n \times 2N$ matrix whose entries are independent real standard normal random variables. It is easy to check that $\sigma_1(A) \leq \sqrt{2}\sigma_1(B)$. Thus by taking $C = 2\sqrt{\lambda} + \sqrt{2} + 2\sqrt{\mu}$ we have via (2.1) that

$$
P \left( \sigma_1(A) \leq C \sqrt{n} \right) \leq e^{-\mu n}.
$$

(2.18)

The estimate for $\sigma_n(A)$ follows from the same strategy as in the real case. First of all, just like the real case for any $n \times n$ unitary matrix $U$ the entries of $UA$ are still independent complex standard normal random variables. As a result the probability $\mathbb{P} (\|A^*v\| \leq \sqrt{\delta n})$ where $v \in \mathbb{C}^n$ is a unit vector does not depend on the choice of $v$. By taking $v = e_1$ we see that $(\|A^*v\|^2$ has the $\Gamma(N, 1)$ distribution (as opposed to the $\Gamma(N, 1)$ distribution in the real case). Applying Lemma 2.1 we obtain the equivalent result for the complex case in

$$
P (\|A^*v\| \leq \sqrt{\delta n}) \leq \left( \frac{e^{\delta \lambda}}{\lambda} \right)^N.
$$

(2.19)

Next for the $\varepsilon$-net, we observe that the unit sphere in $\mathbb{C}^n$ is precisely the unit sphere in $\mathbb{R}^{2n}$ if we identify $\mathbb{C}^n$ as $\mathbb{R}^{2n}$. Thus we can find an $\varepsilon$-net $\mathcal{N}_\varepsilon$ of cardinality no more than $(1 + 2\varepsilon^{-1})^{2n}$. The proof of Theorem 1.3 now goes through with some minor modifications. The most important one is that with (2.18) and (2.19) the inequality condition (2.7) now becomes

$$
-\frac{\mu}{2} \geq \ln \left( 1 + 2\varepsilon^{-1} \right) + \frac{\lambda}{2} \left( 1 - \ln \lambda + 2 \ln(c + \varepsilon C) \right),
$$

where the constant $C$ is changed to $C = 2\sqrt{\lambda} + \sqrt{2} + 2\sqrt{\mu}$. Substituting this $C$ and $\frac{\mu}{2}$ for $\mu$ we prove the theorem.

□
Again as in the real case we may approximate $\frac{2Ct}{1-t}$ simply by $2Ct$, and the maximum of the approximation function is obtained at $t_0 = (2C\lambda L)^{-\frac{1}{\lambda}}$. Thus

$$\tilde{c} := \varphi(t_0) = \left(\frac{1}{2C\lambda L^\lambda}\right)^{\frac{1}{\lambda-1}} \left(1 - \frac{1}{\lambda}\right)$$  \hspace{1cm} (2.20)

has $\tilde{c} \leq c$ and it is a close approximation of the optimal $c$. Therefore Theorem 2.4 yields a more explicit estimate for a complex standard Gaussian random matrix $A$:

$$\mathbb{P}(\tilde{c}\sqrt{n} \leq \sigma_n(A) \leq \sigma_1(A) \leq C\sqrt{n}) \geq 1 - 3e^{-\mu n}$$  \hspace{1cm} (2.21)

where $\tilde{c}$ is given by (2.20). We can now derive the complex version of Theorem 1.2.

**Theorem 2.5** Let $A$ be $n \times N$ whose entries are independent complex random variables of standard normal distribution. Let $\lambda = \frac{N}{n} > 1$. Then there exists a constant $C_\lambda$ depending only on $\lambda$ such that

$$\mathbb{P}(\sigma_n(A) \leq \varepsilon\sqrt{n}) \leq 3 \left( C_\lambda \varepsilon \right)^{2(N-n)} \ln^n \left(\frac{1}{\varepsilon}\right).$$  \hspace{1cm} (2.22)

More specifically,

$$\mathbb{P}(\sigma_n(A) \leq \varepsilon\sqrt{n}) \leq 3 \left( B_\lambda \varepsilon \right)^{2(N-n)} \left(\sqrt{2} + 2\sqrt{\lambda} + 2\sqrt{2(\lambda - 1) \ln(1/\varepsilon)}\right)^{2n}$$  \hspace{1cm} (2.23)

where

$$B_\lambda := \frac{\lambda}{\lambda-1} 2^{\frac{1}{\lambda-1}} (e\lambda)^{\frac{2}{\lambda-1}}. \hspace{1cm} (2.24)$$

**Proof** We use the same steps as in the real case. Set $\varepsilon = \tilde{c}$ where $\tilde{c}$ is given by (2.20) and solve for $\mu$. It follows that

$$\varepsilon^{\lambda-1} = \frac{1}{2C_\lambda L^\lambda} \left(1 - \frac{1}{\lambda}\right)^{\lambda-1}.$$  

Substituting in the explicit values of $C$, $L$ from Theorem 2.4 and $B_\lambda$ in (2.24) yield

$$\left(\sqrt{2} + 2\sqrt{\lambda} + 2\sqrt{\mu}\right) e^{\frac{\mu}{2\varepsilon^{\lambda-1}}} = \frac{1}{2(e\lambda)^{\frac{1}{\lambda}}} \left(1 - \frac{1}{\lambda}\right)^{\lambda-1} = B_\lambda^{1/\lambda - \lambda}. \hspace{1cm} (2.25)$$

Thus

$$e^{\frac{\mu}{2\varepsilon^{\lambda-1}}} < \left(\sqrt{2} + 2\sqrt{\lambda} + 2\sqrt{\mu}\right) e^{\frac{\mu}{2\varepsilon^{\lambda-1}}} = \left(B_\lambda \varepsilon\right)^{1-\lambda}. \hspace{1cm} (2.25)$$

Since $B_\lambda > 1$ we have

$$\mu < 2(\lambda - 1) \ln \left(\frac{1}{\varepsilon}\right) - 2(\lambda - 1) \ln B_\lambda < 2(\lambda - 1) \ln \left(\frac{1}{\varepsilon}\right).$$
Consequently
\[
e^{-\mu n} = \left(\sqrt{2} + 2\sqrt{\lambda} + 2\sqrt{\mu}\right) e^{\lambda-1} B_\lambda^{\lambda-1}
< \left(\sqrt{2} + 2\sqrt{\lambda} + 2\sqrt{2(\lambda - 1) \ln(1/\varepsilon)}\right) e^{\lambda-1} B_\lambda^{\lambda-1}.
\]
Since \((\lambda - 1)n = N - n\), (2.22) of the theorem now follows. Clearly (2.23) follows from (2.22). □

3 Random Frames as NERFs

Our goal in this section is to establish the robustness of random frames against erasures by proving Theorem 1.1. Here we restate Theorem 1.1 in a different form for the benefit of simpler notation in the proof.

**Theorem 3.1** Let \(F = \frac{1}{\sqrt{n}} A\) where \(A\) is \(n \times N\) whose entries are drawn independently from the standard normal \(N(0, 1)\) distribution. Let \(\lambda = \frac{N}{n} > 1\) and \(K = p N = p \lambda n\) where \(\lambda^{-1} < p \leq 1\). For any \(\tau_0 > 0\) there exist constants \(\alpha, \beta > 0\) depending only on \(\lambda, p\) and \(\tau_0\) such that \(F\) is a \((K, \alpha, \beta)\)-NERF with probability at least \(1 - 3e^{-\tau_0 n}\).

**Proof** There exists exactly \(\frac{N!}{K!(N-K)!}\) subsets \(S \subseteq \{1, 2, \ldots, N\}\) of cardinality \(|S| = K\). It is well known that
\[
\frac{N!}{K!(N-K)!} \leq \frac{N^N}{K^K (N-K)^{N-K}},
\]
which can be shown easily by Stirling’s Formula or induction on \(N\). Set \(s_p = p \ln p^{-1} + (1 - p) \ln(1 - p)^{-1}\), which has \(0 \leq s_p \leq \ln 2\). We have then
\[
\frac{N!}{K!(N-K)!} \leq \left(p^{-p} (1 - p)^{p-1}\right)^N = e^{\lambda s_p n}.
\]

Now we set \(\mu = \lambda s_p + \tau_0\). Let \(C = \sqrt{2\mu} + \sqrt{p\lambda} + 1\) and \(c = \sup_{0 < t < 1} \varphi(t)\) where \(\varphi(t)\) is given in (1.12). Let the columns of \(A\) be \(\{a_j\}_{j=1}^N\). For any \(S \subseteq \{1, 2, \ldots, N\}\) we denote by \(A_S\) the submatrix of \(A\) whose columns are \(\{a_j : j \in S\}\). Then for \(|S| = K = p \lambda n\) we have
\[
\mathbb{P} \left(c \sqrt{n} \leq \sigma_n(A_S) \leq \sigma_1(A_S) \leq C \sqrt{n}\right) \geq 1 - 3e^{-\mu n}.
\]
by Theorem 1.3. It follows that
\[
\mathbb{P} \left(\sigma_n(A_S) \leq c \sqrt{n} \text{ or } \sigma_1(A_S) \geq C \sqrt{n}\right) \\ \leq \sum_{|S|=K} \mathbb{P} \left(\sigma_n(A_S) \leq c \sqrt{n} \text{ or } \sigma_1(A_S) \geq C \sqrt{n}\right) \\ \leq 3e^{(\lambda s_p - \mu) n} = 3e^{-\tau_0 n}.
\]
It follows that
\[
\mathbb{P}\left(c \sqrt{n} \leq \sigma_n(A_S) \leq \sigma_1(A_S) \leq C \sqrt{n} \text{ for all } S \text{ with } |S| = K \right) \geq 1 - 3e^{-\tau_0 n}.
\]

This implies that, by setting \(\alpha = c\) and \(\beta = C\), \(F = \frac{1}{\sqrt{n}} A\) is a \((K, \alpha, \beta)\)-NERF with probability at least \(1 - 3e^{-\tau_0 n}\). \(\square\)

Theorems 1.1 and 3.1 state that random Gaussian frames can be robust with high probability against erasures of an arbitrary proportion of data from the original data, at least in theory, as long as the number of remaining vectors is at least \((1 + \delta_0)n\) for some \(\delta_0 > 0\). In practice one may ask how good the condition numbers are if the erasures reach a high proportion, say, 90 % of the data. We show some numerical results below.

**Example 1** Let \(F = \frac{1}{\sqrt{n}} A\) where \(A\) is \(n \times N\) whose entries are independent standard normal random variables. Set \(\tau_0 = 0.4\). In this experiment we fix \(K = 2n\) and \(K = 5n\), respectively, and let \(N\) vary. As \(N\) increases from \(N = K\) to \(N = 100K\) the proportion of erasure \(s = 1 - \frac{K}{N}\) increases from 0 to 99 %. We shall use \(\beta/\alpha\) as a measure of robustness since it is an upper bound for the condition number. Clearly, as \(s\) increases we should expect \(\beta/\alpha\) to increase. Because the frame is normalized so that each column is on average a unit norm vector, it also makes sense to use the smallest singular value as a measurement of robustness. The left plot in Fig. 1 shows \(\log_2(\beta/\alpha)\) (top curve) and \(\log_2(1/\alpha)\) (bottom curve) respectively against \(s\) for \(K = 2n\) (top curve). The right plot in Fig. 1 shows the same two curves, but for \(K = 5n\) (bottom curve). Our numerical results show that in the case \(K = 2n\), with probability at least \(1 - 3e^{-0.4n}\), the condition number is no more than 592.50 for 50 % erasures and no more than 406030 for 90 % erasures. In the case \(K = 5n\), the corresponding numbers are 109.3 and 1523.8, respectively. In fact, even with 99 % erasures the condition number is no more than 35931.

**Fig. 1** Left \(\log_2(\beta/\alpha)\) (top curve) and \(\log_2(1/\alpha)\) (bottom curve) against the proportion of erasures when \(N\) varies from \(K\) to \(100K\) while \(K\) is fixed at \(K = 2n\). Right Same curves as in the left figure, but for \(K = 5n\).
Example 2  Again we let $F = \frac{1}{\sqrt{n}} A$ where $A$ is $n \times N$ whose entries are independent standard normal random variables, and let $t_0 = 0.25$. In this experiment we fix $N = 200n$ and $N = 50n$, respectively, and let $K$ vary so the proportion of erasures $s = 1 - \frac{K}{N}$ varies from 0 to 99% ($N = 200n$) and 0 to 97% ($N = 50n$), respectively. Again we should expect the robustness to go down as we increase $s$. The left plot in Fig. 2 shows $\log_2(\beta/\alpha)$ (top curve) and $\log_2(1/\alpha)$ (bottom curve) respectively against $s$ for $N = 50n$ (top curve) and $N = 200n$ (bottom curve). The right plot in Fig. 2 shows the same two curves, but for $N = 200n$. Our numerical results show that in the case $N = 50n$, with probability at least $1 - 3e^{-0.25n}$, the condition number is no more than 26.52 for 50% erasures and 1462.4 for 90% erasures. In the case $N = 200n$, the corresponding numbers are 20.72 and 272.1, respectively. Even with 95% erasures the condition number is no more than 1110.3.

Acknowledgements  This research was completed while the author was at Department of Mathematics, Michigan State University. Part of this research was supported in part by the National Science Foundation grant DMS-1043032 and AFOSR grant FA9550-12-1-0455.

References

1. Baraniuk, R., Davenport, M., DeVore, R., Wakin, M.: A simple proof of the restricted isometry property for random matrices. Const. Approx. 28(3), 253–263 (2008)
2. Candès, E.J.: The restricted isometry property and its implications for compressed sensing. C. R. Math. 346(9), 589–592 (2008)
3. Candes, E.J., Eldar, Y., Strohmer, T., Voroninski, V.: Phase retrieval via matrix completion. SIAM Rev. 57(2), 225–251 (2015)
4. Candes, E.J., Tao, T.: Decoding by linear programming. IEEE Trans. Inf. Theory 51(12), 4203–4215 (2005)
5. Candes, E.J., Tao, T.: Near-optimal signal recovery from random projections: Universal encoding strategies? IEEE Trans. Inf. Theory 52(12), 5406–5425 (2006)
6. Casazza, P.G., Kovačević, J.: Equal-norm tight frames with erasures. Adv. Comput. Math. 18(2), 387–430 (2003)
7. Edelman, A.: Eigenvalues and condition numbers of random matrices. SIAM J. Matrix Anal. Appl. 9(4), 543–560 (1988)
8. Fickus, M., Jasper, J., Mixon, D.G., Peterson, J.: Group-theoretic constructions of erasure-robust frames. Linear Algebra Appl. 479, 131–154 (2015)
9. Fickus, M., Mixon, D.G.: Numerically erasure-robust frames. Linear Algebra Appl. 437(6), 1394–1407 (2012)
10. Goyal, V.K., Kovačević, J., Kelner, J.A.: Quantized frame expansions with erasures. Appl. Comput. Harmon. Anal. 10(3), 203–233 (2001)
11. Holmes, R.B., Paulsen, V.I.: Optimal frames for erasures. Linear Algebra Appl. 377, 31–51 (2004)
12. Kovacevic, J., Dragotti, P.L., Goyal, V.K.: Filter bank frame expansions with erasures. IEEE Trans. Inf. Theory 48(6), 1439–1450 (2002)
13. Puschel, M., Kovacevic, J.: Real, tight frames with maximal robustness to erasures. In: Proceedings of the Data Compression Conference (DCC) 2005, pp. 63–72. IEEE
14. Rudelson, M., Vershynin, R.: Smallest singular value of a random rectangular matrix. Commun. Pure Appl. Math. 62(12), 1707–1739 (2009)
15. Rudelson, M., Vershynin, R.: Non-asymptotic theory of random matrices: extreme singular values (2010). arXiv preprint arXiv:1003.2990
16. Strohmer, T., Heath, R.W.: Grassmannian frames with applications to coding and communication. Appl. Comput. Harmon. Anal. 14(3), 257–275 (2003)
17. Tao, T., Vu, V.: Random matrices: the distribution of the smallest singular values. Geom. Funct. Anal. 20(1), 260–297 (2010)
18. Vershynin, R.: Frame expansions with erasures: an approach through the non-commutative operator theory. Appl. Comput. Harmon. Anal. 18(2), 167–176 (2005)
19. Vershynin, R.: Introduction to the non-asymptotic analysis of random matrices (2010). arXiv preprint arXiv:1011.3027