Numerically Stable Real Number Codes Based on Random Matrices [⋆]

Zizhong Chen and Jack Dongarra

Computer Science Department, University of Tennessee, Knoxville
1122 Volunteer Blvd., Knoxville, TN 37996-3450, USA
{zchen, dongarra}@cs.utk.edu

Abstract. Error correction codes defined over real-number field have been studied and recognized as useful in many applications. However, most real-number codes in literature are quite suspect in their numerical stability. In this paper, we introduce a class of real-number codes based on random generator matrices over real-number fields. Codes over complex-number field are also discussed. Experiment results demonstrate our codes are numerically much more stable than existing codes in literature.

1 Introduction

Error correction codes are often defined over finite fields. However, in many applications, error correction codes defined over finite fields do not work. Instead, codes defined over real-number or complex-number fields have to be used to detect and correct errors. For example, in algorithm-based fault tolerance [2, 10, 11, 13] and fault tolerant dynamic systems [8], to provide fault tolerance in computing, data are first encoded using error correction codes and then algorithms are re-designed to operate (using floating point arithmetic) on the encoded data. Due to the impact of the floating-point arithmetic on the binary representation of these encoded data, codes defined over finite fields do not work. But codes defined over real-number and complex-number fields can be used in these applications to correct errors in computing by taking advantage of certain relationships, which are maintained only when real-number (or complex-number) codes are used.

However, most real-number and complex-number codes in literature are quite suspect in their numerical stability. Error correction procedures in most error correction codes involve solving linear system of equations. In computer floating point arithmetic where no computation is exact due to round-off errors, it is well known [7] that, in solving a linear system of equations, a condition number

[⋆] This research was supported in part by the Los Alamos National Laboratory under Contract No. 03891-001-99 49 and the Applied Mathematical Sciences Research Program of the Office of Mathematical, Information, and Computational Sciences, U.S. Department of Energy under contract DE-AC05-00OR22725 with UT-Battelle, LLC.

of 10^k for the coefficient matrix leads to a loss of accuracy of about k decimal digits in the solution. In the generator matrices of most existing real-number and complex-number codes, there exist ill-conditioned sub-matrices. Therefore, in these codes, when certain error patterns occur, an ill-conditioned linear system of equations has to be solved in the error correction procedure, which can cause the loss of precision of possibly all digits in the recovered numbers.

The numerical issue of the real-number and complex-number codes has been recognized and studied in some literature. In [2], Vandermonde-like matrix for the Chebyshev polynomials was introduced to relieve the numerical instability problem in error correction for algorithm-based fault tolerance. In [5, 6, 9, 12], the numerical properties of the Discrete Fourier Transform codes were analyzed and methods to improve the numerical properties were also proposed. To some extent, these efforts have alleviated the numerical problem of the real-number and complex-number codes. However, how to construct real-number and complex-number codes without numerical problem is still an open problem.

In this paper, we introduce a class of real-number and complex-number codes that are numerically much more stable than existing codes in literature. Our codes are based on random generator matrices over real-number and complex-number fields. The rest of this paper is organized as follow: Section 2 specifies the problem we focus on. In Section 3, we first study the properties of random matrices and then introduce our codes. Section 4 compares our codes with most existing codes in both burst error correction and random error correction. Section 5 concludes the paper and discusses the future work.

2 Problem Specification

Let $x = (x_1, x_2, ..., x_N)^T \in \mathcal{C^N}$ denote the original information, and G denote a M by N real or complex matrix. Let $y = (y_1, y_2, ..., y_M)^T \in \mathcal{C^M}$, where $M = N + K$, denote the encoded information of x with redundancy. The original information x and the encoded information y are related through

$$y = Gx. \tag{1}$$

Our problem is: how to choose the matrix G such that, after any no more than K erasures in the elements of the encoded information y, a good approximation of the original information x can still be reconstructed from y?

When there are at most K elements of y lost, there are at least N elements of y available. Let J denote the set of indexes of any N available elements of y. Let y_J denote a sub-vector of y consisting of the N available elements of y whose indexes are in J. Let G_J denote a sub-matrix of G consisting of the N rows whose indexes are in J. Then, from (1), we can get the following relationship between x and y_J:

$$y_J = G_J x. \tag{2}$$

When the matrix G_J is singular, there are infinite number of solutions to (2). But, if the matrix G_J is non-singular, then (2) has one and only one solution, which is the original information vector x.

In computer real-number and complex-number arithmetic where no computation is exact due to round-off errors, it is well known [7] that, in solving a linear system of equations, a condition number of 10^k for the coefficient matrix leads to a loss of accuracy of about k decimal digits in the solution. Therefore, in order to reconstruct a good approximation of the original information x, G_J has to be well-conditioned.

For any N by N sub-matrix G_J of G, there is a erasure pattern of y which requires to solve a linear system with G_J as the coefficient matrix to reconstruct an approximation of the original x. Therefore, to guarantee that a reasonably good approximation of x can be reconstructed after any no more than K erasures in y , the generator matrix G must satisfy: *any N by N sub-matrix of G is well-conditioned.*

3 Real Number Codes Based on Random Matrices

In this section, we will introduce a class of new codes that are able to reconstruct a very good approximation of the original information with high probability regardless of the erasure patterns in the encoded information. Our new codes are based on random matrices over real or complex number fields.

3.1 Condition Number of Random Matrices from Standard Normal Distribution

In this sub-section, we mainly focus on the probability that the condition number of a random matrix is large and the expectation of the logarithm of the condition number. Let $G(m,n)$ be an $m \times n$ real random matrix whose elements are independent and identically distributed standard normal random variables and $\widetilde{G}(m,n)$ be its complex counterpart.

Theorem 1. *Let κ denote the condition number of $G(n,n)$, $n > 2$, and $t \geq 1$, then*

$$\frac{0.13n}{t} < P(\kappa > t) < \frac{5.60n}{t}. \tag{3}$$

Moreover,

$$E(\log(\kappa)) = \log(n) + c + \epsilon_n, \tag{4}$$

where $c \approx 1.537$, $lim_{n\to\infty}\epsilon_n = 0$,

Proof. The inequality (3) is from Theorem 1 of [1]. The formula (4) can be obtained from Theorem 7.1 of [3]. □

Theorem 2. *Let $\widetilde{\kappa}$ denote the condition number of $\widetilde{G}(n,n)$, and $t \geq \sqrt{n}$, then*

$$1 - \left(1 - \frac{1}{t^2}\right)^{n^2-1} \leq P(\widetilde{\kappa} > t) \leq 1 - \left(1 - \frac{n}{t^2}\right)^{n^2-1}. \tag{5}$$

Moreover,

$$E(\log(\widetilde{\kappa})) = \log(n) + c + \epsilon_n, \tag{6}$$

where $c \approx 0.982$, $lim_{n\to\infty}\epsilon_n = 0$, *Proof.* Let $\widetilde{\kappa}_D$ denote the scaled condition number (see [4] for definition) of $\widetilde{G}(n, n)$, then

$$P(\frac{\widetilde{\kappa}_D}{\sqrt{n}} > t) \leq P(\widetilde{\kappa} > t) \leq P(\widetilde{\kappa}_D > t). \tag{7}$$

From Corollary 3.2 in [4], we have

$$P(\widetilde{\kappa}_D > t) = 1 - \left(1 - \frac{n}{t^2}\right)^{n^2-1}. \tag{8}$$

Therefore,

$$P(\frac{\widetilde{\kappa}_D}{\sqrt{n}} > t) = P(\widetilde{\kappa}_D > \sqrt{n}t) = 1 - \left(1 - \frac{1}{t^2}\right)^{n^2-1}. \tag{9}$$

The inequality (5) can be obtained from (7), (8) and (9). The formula (6) can be obtained from Theorem 7.2 of [3]. □

In error correction practice, all random numbers used are pseudo random numbers, which have to be generated through a random number generator. Fig.1 shows the empirical probability density functions of the condition numbers of the pseudo random matrix $G(100, 100)$ and $\widetilde{G}(100, 100)$, where $G(100, 100)$ is generated by $randn(100, 100)$ and $\widetilde{G}(100, 100)$ is generated by $randn(100, 100) + \sqrt{-1} * randn(100, 100)$ in MATLAB. From these density functions, we know that most pseudo random matrices also have very small condition numbers. And, for the same matrix size, the tail of the condition number for a complex random matrix is thinner than that of a real one.

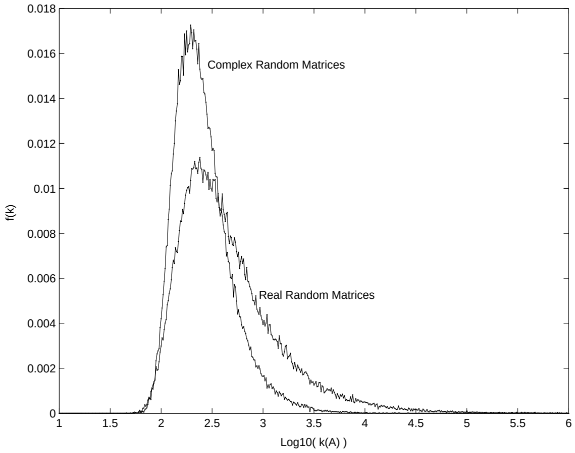

Fig. 1. The density functions of the condition numbers of $G(100, 100)$ and $\widetilde{G}(100, 100)$.

We have also tested some other random matrices. Experiments show a lot of other random matrices, for example, uniformly distributed pseudo random matrices, also have small condition numbers with high probability. For random matrices of non-normal distribution, we will report our experiments and some analytical proofs of their condition number properties in a further coming paper.

3.2 Real Number Codes Based on Random Matrices

In this sub-section, we introduce a class of new codes that are able to reconstruct a very good approximation of the original information with very high probability regardless of the erasure patterns in the encoded information.

In the real number case, we propose to use $G(M, N)$ or uniformly distributed M by N matrices with mean 0 (denote as $U(M, N)$) as our generator matrices G. In the complex number case, we propose to use $\widetilde{G}(M, N)$ or uniformly distributed M by N complex matrices with mean 0 (denote as $\widetilde{U}(M, N)$) as our generator matrices G.

Take the real-number codes based on random matrix $G(M, N)$ as an example. Since each element of the generator matrix $G(M, N)$ is a random number from the standard normal distribution, so each element of any $N \times N$ sub-matrix $(G_J)_{N\times N}$ of $G(M, N)$ is also a random number from the standard normal distribution. According to the condition number results in Subsection 3.1 , the probability that the condition number of $(G_J)_{N\times N}$ is large is very small. Hence, any N by N sub-matrix $(G_J)_{N\times N}$ of G is well-conditioned with very high probability. Therefore, no mater what erasure patterns occur, the error correction procedure is numerically stable with high probability.

We admit that our real-number and complex-number codes are not perfect. Due to the probability approach we used, the drawback of our codes is that, no matter how small the probability is, there is a probability that a erasure pattern may not be able to be recovered accurately.

However, compared with the existing codes in literature, the probability that our codes fail to recover a good approximation of the original information is negligible (see Section 4 for detail). Moreover, in the error correction practice, we may first generate a set of pseudo random generator matrices and then test each generator matrix until we find a satisfied one.

4 Comparison with Existing Codes

In the existing codes in literature, the generator matrices mainly include: Vandermonde matrix (Vander) [8], Vandermonde-like matrix for the Chebyshev polynomials (Chebvand) [2], Cauchy matrix (Cauchy), Discrete Cosine Transform matrix (DCT), Discrete Fourier Transform matrix (DFT) [6]. These generator matrices all contain ill-conditioned sub-matrices. Therefore, in these codes, when certain error patterns occur, an ill-conditioned linear system has to be solved to reconstruct an approximation of the original information, which can cause the loss of precision of possibly all digits in the recovered numbers. However, in our codes, the generator matrices are random matrices. Any sub-matrix of our generator matrices is still a random matrix, which is well-conditioned with very high probability. Therefore, no mater what erasure patterns occur, the error correction procedure is numerically stable with high probability. In this section, we compare our codes with existing codes in both burst erasure correction and random erasure correction.

4.1 Burst Erasure Correction

We compare our codes with existing codes in burst error correction using the following example.

Example 1. Suppose $x = (1, 1, 1, ..., 1)^T$ and the length of x is $N = 100$. G is a 120 by 100 generator matrix. $y = Gx$ is a vector of length 120. Suppose y_i, where $i = 101, 102, ...120$, are lost. We will use y_j, where $j = 1, 2, ...100$, to reconstruct x through solving (2) .

Table 1. The generator matrices of different codes

Name	The generator matrix $G = (g_{mn})_{120\times100}$
Vander	$\left((m+1)^{100-n-1}\right)_{120\times100}$
Chebvand	$(T_{m-1}(n))_{120\times100}$, where T_{m-1} is the chebyshev polynomial of degree $n-1$
Cauchy	$\left(\frac{1}{m+n}\right)_{120\times100}$
DCT	$\left(\sqrt{\frac{i}{120}}\cos\frac{\pi(2n+1)m}{240}\right)_{120\times100}$, where if $m = 0, i = 1$, and if $m \neq 0, i = 2$
DFT	$\left(e^{-j\frac{2\pi}{120}mn}\right)_{120\times100}$, where $j = \sqrt{-1}$
RandN	randn(120,100) in MATLAB
RandN-C	randn(120,100) + j * randn(120,100) in MATLAB, where $j = \sqrt{-1}$
RandU	rand(120,100) - 0.5 in MATLAB
RandU-C	rand(120,100) - 0.5 + j * (rand(120,100) - 0.5) in MATLAB,

Table 1 shows how the generator matrix of each code is generated. Table 2 reports the accuracy of the recovery for each code. All calculations are done using MATLAB. The machine precision is 16 digits. Table 2 shows our codes are able to reconstruct the original information x with much higher accuracy than the existing codes. The reconstructed x from all existing codes lost all of their 16 effective digits. However, the reconstructed x from the codes we proposed in the last section lost only about 2 effective digits.

4.2 Random Erasure Correction

For any N by N sub-matrix G_J of G, there is a erasure pattern of y which requires to solve a linear system with G_J as the coefficient matrix to reconstruct an

Table 2. Burst erasure recovery accuracy of different codes

Name	$\kappa(G_J)$	$\frac{\|x-\widehat{x}\|_2}{\|x\|_2}$	Accurate digits	Number of digits lost
Vander	3.7e+218	2.4e+153	0	16
Chebvand	Inf	1.7e+156	0	16
Cauchy	5.6e+17	1.4e+03	0	16
DCT	1.5e+17	2.5e+02	0	16
DFT	2.0e+16	1.6e+00	0	16
RandN	7.5e+2	3.8e-14	14	2
RandN-C	4.5e+2	6.8e-14	14	2
RandU	8.6e+2	3.7e-14	14	2
RandU-C	5.7e+2	2.6e-14	14	2

approximation of the original x. A random erasure actually results in a randomly picked N by N sub-matrix of G. In Table 3, we compare the proportion of 100 by 100 sub-matrices whose condition number is larger than 10^i, where $i = 4, 6, 8$, and 10, for different kind of generator matrices of size 150 by 100. All generator matrices are defined in Table 1. All results in Table 3 are calculated using MATLAB based on 1,000,000 randomly (uniformly) picked sub-matrices.

From Table 3, we can see, of the 1,000,000 randomly picked sub-matrices from any of our random generator matrices, there are 0.000% sub-matrices whose condition number is larger than 10^8. However, for all existing codes in literature that we have tested, there are at least 21.644% sub-matrices whose condition number is larger than 10^8. Therefore, our codes are much more stable than the existing codes in literature.

Table 3. Percentage of 100 by 100 sub-matrices (of a 150 by 100 generator matrix) whose condition number is larger than 10^i, where $i = 4, 6, 8$, and 10.

Name	$\kappa \geq 10^4$	$\kappa \geq 10^6$	$\kappa \geq 10^8$	$\kappa \geq 10^{10}$
Vander	100.000%	100.000%	100.000%	100.000%
Chebvand	100.000%	100.000%	100.000%	100.000%
Cauchy	100.000%	100.000%	100.000%	100.000%
DCT	96.187%	75.837%	48.943%	28.027%
DFT	92.853%	56.913%	21.644%	5.414%
RandN	1.994%	0.023%	0.000%	0.000%
RandN-C	0.033%	0.000%	0.000%	0.000%
RandU	1.990%	0.018%	0.000%	0.000%
RandU-C	0.036%	0.000%	0.000%	0.000%

5 Conclusion and Future Work

In this paper, we have introduced a class of real-number and complex-number codes based on random generator matrices over real-number and complex-number fields. we have compared our codes with existing codes in both burst erasure correction and random erasure correction. Experiment results demonstrate our codes are numerically much more stable than existing codes in literature.

For the future, we will compare real-number codes based on different random matrices with different probability distributions. we would also like to investigate what is the numerically optimal real number codes.

References

1. Azais, J. M. and Wschebor, M.: Upper and lower bounds for the tails of the distribution of the condition number of a gaussian matrix, submitted for publication, 2003
2. Boley, D. L., Brent, R. P., Golub, G. H. and Luk, F. T.: Algorithmic Fault Tolerance Using the Lanczos Method, SIAM Journal on Matrix Analysis and Applications, vol. 13, (1992), pp. 312-332.
3. Edelman, A.: Eigenvalues and Condition Numbers of Random Matrices, Ph.D. thesis, Dept. of Math., M.I.T., 1989.
4. Edelman, A.: On the distribution of a scaled condition number, Mathematics of Computation,vol. 58, (1992), pp. 185-190.
5. Ferreira, P.: Stability issues in error control coding in complex field, interpolation, and frame bounds, IEEE Signal Processing Letters, vol.7 No.3,(2000) pp.57-59.
6. Ferreira, P., Vieira, J.: Stable DFT codes and frames, IEEE Signal Processing Letters, vol.10 No.2,(2003) pp.50-53.
7. Golub, G. H. and Van Loan, C. F.: Matrix Computations, 2nd Ed., The John Hopkins University Press, 1989.
8. Hadjicostis, C. N. and Verghese, G. C.: Coding approaches to fault tolerance in linear dynamic systems, Submitted to IEEE Transactions on Information Theory.
9. Henkel, W.: Multiple error correction with analog codes, Proceedings of AAECC, Springer-Verlag, (1989), pp. 239-249.
10. Huang, H. and Abraham, J. A.: Algorithm-based fault tolerance for matrix operations, IEEE Transactions on Computers, vol. C-39, (1984) pp.300-304.
11. Luk, F. T. and Park, H.: An analysis of algorithm-based fault tolerance techniques, Journal of Parallel and Distributed Computing, vol. 5 (1988), pp. 1434-1438.
12. Marvasti, F., Hasan, M., Echhart, M. and Talebi, S.: Efficient algorithms for burst error recovery using FFT and other transform kernels, IEEE Transactions on Signal Processing, vol.47, No.4, (1999), pp. 1065-1075.
13. Nair, S. S. and Abraham, J. A.: Real-number codes for fault-tolerant matrix operations on processor arrays, IEEE Transactions on Computers, vol. C-39,(1990) pp.300-304.