# LINEAR TRANSFORMATIONS AND RESTRICTED ISOMETRY PROPERTY

*Leslie Ying and Yi Ming Zou*


## ABSTRACT

The Restricted Isometry Property (RIP) introduced by Candés and Tao is a fundamental property in compressed sensing theory. It says that if a sampling matrix satisfies the RIP of certain order proportional to the sparsity of the signal, then the original signal can be reconstructed even if the sampling matrix provides a sample vector which is much smaller in size than the original signal. This short note addresses the problem of how a linear transformation will affect the RIP. This problem arises from the consideration of extending the sensing matrix and the use of compressed sensing in different bases. As an application, the result is applied to the redundant dictionary setting in compressed sensing.


***Index Terms***— Compressed Sensing, Concentration Inequalities, Restricted Isometry Property

## 1. INTRODUCTION

In Compressed Sensing (CS), one considers the problem of recovering a vector (discrete signal) $x \in \mathbb{R}^N$ from its linear measurements $y$ of the form

$$y_i =< x, \varphi_i >,\ 1 \le i \le n, \tag{1}$$

with $n << N$. If $x$ is sparse, CS theory says that one can actually recover $x$ from $y$ which is much smaller in size than $x$ by solving a convex program with a suitably chosen set of sampling row vectors $\{\varphi_i | 1 \le i \le n\}$ [1][2][3]. The linear system (1) can be written in the form of matrix multiplication

$$y = \Phi x, \tag{2}$$

where $\Phi$ is an $n \times N$ matrix formed by the row vectors $\varphi_i$ called the sampling matrix. One of the Conditions that ensures the performance of the sampling matrix $\Phi$ is the RIP. A matrix $\Phi \in \mathbb{R}^{n\times N}$ is said to satisfy the RIP of order $k \in \mathbb{N}$ and *isometry constant* $\delta_k \in (0,1)$ if

$$(1-\delta_k)\|z\|_2^2 \le \|\Phi_T z\|_2^2 \le (1+\delta_k)\|z\|_2^2, \quad \forall z \in \mathbb{R}^{|T|}, \tag{3}$$

where $T \subset \{1, 2, \ldots, N\}$ satisfying $|T| \le k$, and $\Phi_T$ denotes the matrix obtained by retaining only the columns of $\Phi$ corresponding to the entries of $T$. Condition (3) is equivalent to the condition that all the matrices $\Phi_T^t \Phi_T$ have their eigenvalues in $[1-\delta_k, 1+\delta_k]$. For any matrix $X \in \mathbb{R}^{r\times s}$ and any $k \in \mathbb{N}$, we denote the corresponding isometry constant of $X$ by $\delta_k(X)$. If there is no confusion, we will just write $\delta_k$. In particular, we always use $\delta_k$ for the matrix $\Phi$.

A theorem due to Candés, Romberg, and Tao [4] says that if $\Phi$ satisfies the RIP of order $3k$, then the solution $\hat{x}$ of the following convex minimization problem

$$\min \|x\|_1 \quad \text{subject to } \Phi x = y, \tag{4}$$

satisfies (see also [5])

$$\|x - \hat{x}\|_2 \le \frac{C_2 \sigma_k(x)}{\sqrt{k}}, \tag{5}$$

where $\sigma_k(x)$ is the $\ell_1$ error of the best $k$-term approximation, and $C_2$ is a constant depending only on $\delta_{3k} \in (0,1)$.[1]

A condition that ensures a random matrix satisfies the RIP with high probability is given by the concentration of measure inequality. An $n \times N$ random matrix $\Phi$ is said to satisfy the concentration of measure inequality if for any $x \in \mathbb{R}^N$,

$$P(|\|\Phi x\|_2^2 - \|x\|_2^2| \ge \varepsilon \|x\|_2^2) \le 2e^{-nc_0(\varepsilon)}, \tag{6}$$

where $\varepsilon \in (0,1)$, and $c_0(\varepsilon)$ is a constant depending only on $\varepsilon$.

The random matrices $\Phi = (r_{ij})$ generated by the following probability distributions are known to satisfy the concentration of measure inequality with $c_0(\varepsilon) = \varepsilon^2/4 - \varepsilon^3/6$ [5]:

$$r_{ij} \sim N\left(0, \frac{1}{n}\right),$$

$$r_{ij} = \begin{cases} \frac{1}{\sqrt{n}} & \text{with probability } 1/2 \\ -\frac{1}{\sqrt{n}} & \text{with probability } 1/2 \end{cases}. \tag{7}$$

According to Theorem 5.2 in [5][2], for given integers $n$ and $N$, and $0 < \delta < 1$, if the probability distribution generating

L. Ying is with the Department of Electrical Engineering, University of Wisconsin, Milwaukee, WI 53201, USA email: leiying@uwm.edu.

Y. M. Zou is with the Department of Mathematical Sciences, University of Wisconsin, Milwaukee, WI 53201, USA email: ymzou@uwm.edu.

This work is supported in part by the University of Wisconsin System Applied Research Grant and the University of Wisconsin-Milwaukee Research Growth Initiative Grant.

[1] It should be noted that the RIP is only a sufficient condition for reconstruction. If $\Phi$ satisfying the RIP, $cA$ may not satisfy the RIP for $c \neq 0$. However, it is clear that both $A$ and $cA$ lead to similar sparse recovery using $\ell_1$ program. However, this issue is beyond the current scope [6].

[2] In the proof given in [5], the constant $c_1$ was first chosen such that $a := c_0(\varepsilon)\delta/2 - c_1[1 + (1 + \log\frac{12}{\delta})/\log\frac{N}{k}] > 0$, then the constant $c_2$ was chosen such that $0 < c_2 < a$. Thus the constants depend also on $\varepsilon$.

the $n \times N$ matrices $\Phi$ satisfies the concentration inequality (6), then there exist constants $c_1, c_2 > 0$ depending only on $\delta$ such that the RIP holds for $\Phi$ with the prescribed $\delta$ and any

$$k \leq c_1 n / \log(N/k) \tag{8}$$

with probability $\geq 1 - e^{-2c_2 n}$. Furthermore, this RIP for $\Phi$ is universal in sense that it holds with respect to any orthogonal basis used in the measurement.

There are also deterministic constructions of matrices satisfying the RIP [7][8][9][10].

For application purposes, one often needs to analyze the RIP constants of the products of a matrix $\Phi$ with known RIP constant $\delta$ and other matrices. For example, when one considers different bases or redundant dictionaries under which the signals of interest are sparse, matrices of the form $\Phi B$ needs to be analyzed [2][11], where $B$ is given by the basis or the dictionary. For another example, if the size of $\Phi$ is $n \times N$ with $n < N$, one would like to extend $\Phi$ to $A\Phi B$ of size $m \times q$ with $m < n < N < q$ if possible, since that gives a further reduction on the number of measurements one needs to collect: for $\Phi$, the number of measurements is $n$; while for $A\Phi B$, the number of measurements is $m$.

These situations can be formulated under a more general framework by asking the following question: If a matrix $\Phi$ of size $n \times N$ satisfies the RIP with a given isometry constant $0 < \delta < 1$ (with certain probability if $\Phi$ is random), and $A, B$ are given matrices of sizes $m \times n$ and $N \times q$ respectively, then what is the isometry constant of the matrix $A\Phi B$?

In section 2, we first show that if all $\Phi$, $A$, and $B$ are random and satisfy the concentration of measure inequality, then $A\Phi B$ satisfies the concentration of measure inequality, therefore it has RIP. Then we observe that if deterministic matrix is involved, the problem is more complicated, but it can still be analyzed by using the SVDs of $A$ and $B$. It is not possible to multiply by a deterministic $A$ from the left to achieve more reduction on the number of measurements without further assumption. Our result shows that it is possible to extend the matrix $\Phi$ by multiplying a deterministic $B$ from the right to extend $\Phi$ if $\Phi$ is random, though the isometry constant will be changed. This result can be applied to redundant dictionary setting to give a different approach for using CS with redundant dictionaries.

## 2. MAIN RESULT

We first consider the random case. Let $\Phi$ be an $n \times N$ matrix satisfying the concentration inequality (6) with constant $\varepsilon$, and let $A$ (respectively $B$) be a random matrix size $m \times n$ (respectively $N \times q$) satisfying the concentration inequality (6) with $\varepsilon_1$ (respectively, $\varepsilon_2$). Then we have:

**Theorem 2.1.** *Assume that all $\varepsilon, \varepsilon_1, \varepsilon_2 < 1/3$. The matrix $A\Phi$ satisfies the concentration inequality*

$$P(|\|A\Phi x\|_2^2 - \|x\|_2^2| \geq \varepsilon_3 \|x\|_2^2) \leq 2e^{-mc_0'},$$

*where $\varepsilon_3 = \varepsilon + \varepsilon_1(1+\varepsilon)$, and $c_0'$ is a constant that depends only on $c_0(\varepsilon)$ and $c_0(\varepsilon_1)$ (as defined in (6)). The same statement holds for $\Phi B$ with $\varepsilon_3 = \varepsilon + \varepsilon_2(1+\varepsilon)$ and $m$ replaced by $n$.*

*Proof.* We give the proof for the case of left multiplication by $A$, the proof for the case of right multiplication by $B$ is similar. By assumption, with probability $\geq 1 - 2e^{-mc_0(\varepsilon_1)}$, the matrix $A$ satisfies

$$(1-\varepsilon_1)\|y\|_2^2 < \|Ay\|_2^2 < (1+\varepsilon_1)\|y\|_2^2, \text{ for any } y \in \mathbb{R}^n.$$

Replacing $y$ by $\Phi x$ ($x \in \mathbb{R}^N$), we have

$$(1-\varepsilon_1)\|\Phi x\|_2^2 < \|A\Phi x\|_2^2 < (1+\varepsilon_1)\|\Phi x\|_2^2. \tag{9}$$

Again by assumption, with probability $\geq 1 - 2e^{-nc_0(\varepsilon)}$, the matrix $\Phi$ satisfies

$$(1-\varepsilon)\|x\|_2^2 < \|\Phi x\|_2^2 < (1+\varepsilon)\|x\|_2^2, \text{ for any } x \in \mathbb{R}^N. \tag{10}$$

Now the statement follows by combining (9) and (10). □

**Remark.** If $m \leq n$, the constant $c_0'$ in Theorem 2.1 can be roughly estimated by the inequality $c_0' \leq c_0(\varepsilon') - \log 2/m$, where $c_0(\varepsilon') = min\{c_0(\varepsilon_1), c_0(\varepsilon)\}$. This is obtained from

$$1 - (1 - 2e^{-mc_0(\varepsilon_1)})(1 - 2e^{-nc_0(\varepsilon)}) \leq 2e^{-m(c_0(\varepsilon') - \log 2/m)}.$$

More precise estimation can be carried out, but we are not concerning this point here.

Now we consider the cases when deterministic matrices are involved. We observe that it is not possible to multiply a deterministic matrix $A$ from the left to extend the sensing matrix to achieve further reduction on the number of measurements without other assumptions. To see this, we consider the SVD of $A$.

For any positive integer $d$, let $O(d)$ be the set of $d \times d$ orthogonal matrices. There exists $U \in O(n)$ such that

$$A^t A = U^t \begin{pmatrix} \sigma_1 & & & \\ & \sigma_2 & & \\ & & \ddots & \\ & & & \sigma_n \end{pmatrix} U, \tag{11}$$

where $\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_n \geq 0$. Since for any $T \subset \{1, 2, \ldots, N\}$, $(A\Phi)_T = A\Phi_T$, we have

$$\begin{aligned} (A\Phi)_T^t (A\Phi)_T &= \Phi_T^t A^t A \Phi_T \qquad (12) \\ &= \Phi_T^t U^t \begin{pmatrix} \sigma_1 & & & \\ & \sigma_2 & & \\ & & \ddots & \\ & & & \sigma_n \end{pmatrix} U \Phi_T \\ &= (U\Phi)_T^t \begin{pmatrix} \sigma_1 & & & \\ & \sigma_2 & & \\ & & \ddots & \\ & & & \sigma_n \end{pmatrix} (U\Phi)_T. \end{aligned}$$

If $m < n$, then $\sigma_{m+1} = \cdots = \sigma_n = 0$, and hence

$$\begin{pmatrix} \sigma_1 & & & \\ & \sigma_2 & & \\ & & \ddots & \\ & & & \sigma_n \end{pmatrix} (U\Phi)_T = \begin{pmatrix} A_1 \\ 0 \end{pmatrix}$$

for a suitable block matrix $A_1$. From the last matrix one can see immediately that RIP fails: any information carried by the last $m - n$ entries will be lost.

If $m \geq n$, then we can change $\Phi$ by multiplying $A$ from the left if $A$ has full column rank. Since under this assumption, all $\sigma_i > 0$. Note that $U\Phi$ has the same RIP as $\Phi$, so if $\delta_k$ is the RIP constant of $\Phi$ corresponding to all $T$ of size $k \leq N$, we can bound the RIP constant of $A\Phi$ by $\sigma_n(1-\delta_k)$ and $\sigma_1(1+\delta_k)$. In fact, for $z \in \mathbb{R}^k$, if we let $U\Phi_T z = y = (y_1, \ldots, y_n)^t$, then $\|y\|_2 = \|\Phi_T z\|_2$, and according to (12)

$$\sigma_n \|y\|_2^2 \leq \|A\Phi_T z\|_2^2 = \sum_{i=1}^n \sigma_i y_i^2 \leq \sigma_1 \|y\|_2^2. \tag{13}$$

Thus we have (use (3))

$$\sigma_n(1-\delta_k)\|z\|_2^2 \leq \|A\Phi_T z\|_2^2 \leq \sigma_1(1+\delta_k)\|z\|_2^2. \tag{14}$$

Note that the above analysis works whether $\Phi$ is random or deterministic.

Next, we consider the product $\Phi B$. In this case, we need to distinguish between random matrix $\Phi$ and deterministic matrix $\Phi$. Assume that $\Phi$ is a random matrix satisfying the concentration inequality (6) and hence satisfying the RIP inequality (3) with probability $\geq p$. Note that the concentration inequality is invariant under the right multiplication by an orthogonal matrix. That is, if $U \in O(N)$, then $\Phi U$ also satisfies (3) with probability $\geq p$.

Let $B$ be an $N \times q$ matrix. To make the argument clearer, we assume that $T \subset \{1, 2, \ldots, q\}$ with $|T| = k < N$ (note that this is sufficient for our purpose). We have $U \in O(N)$ and $V \in O(k)$ such that

$$B_T = U \begin{pmatrix} D \\ 0 \end{pmatrix}_{N \times k} V, \tag{15}$$

where

$$D = \begin{pmatrix} \lambda_1 & & & \\ & \lambda_2 & & \\ & & \ddots & \\ & & & \lambda_k \end{pmatrix}, \quad \lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_k \geq 0.$$

For $x \in \mathbb{R}^k$,

$$\Phi B_T x = \Phi U \begin{pmatrix} D \\ 0 \end{pmatrix}_{N \times k} Vx. \tag{16}$$

Let

$$z = \begin{pmatrix} D \\ 0 \end{pmatrix} Vx \in \mathbb{R}^N.$$

Then $z$ is $k$-sparse (the last $N-k$ entries are always 0). Thus, since $\Phi U$ has the same RIP as $\Phi$, we have

$$(1-\delta_k)\|z\|_2^2 \leq \|\Phi B_T x\|_2^2 = \|\Phi U z\|_2^2 \leq (1+\delta_k)\|z\|_2^2 \tag{17}$$

with probability $\geq p$.

Let $y = (y_1, \ldots, y_k)^t = Vx$, then $\|y\|_2 = \|x\|_2$, and

$$\begin{aligned} \|z\|_2^2 &= x^t V^t (D^t \ 0) \begin{pmatrix} D \\ 0 \end{pmatrix} Vx \\ &= y^t \begin{pmatrix} \lambda_1^2 & & & \\ & \lambda_2^2 & & \\ & & \ddots & \\ & & & \lambda_k^2 \end{pmatrix} y = \sum_{i=1}^k \lambda_i^2 y_i^2. \end{aligned}$$

Since

$$\lambda_k^2 \|y\|_2^2 \leq \sum_{i=1}^k \lambda_i^2 y_i^2 \leq \lambda_1^2 \|y\|_2^2,$$

by (17), we have

$$\lambda_k^2(1-\delta_k)\|x\|_2^2 \leq \|\Phi B_T x\|_2^2 \leq \lambda_1^2(1+\delta_k)\|x\|_2^2 \tag{18}$$

with probability $\geq p$.

If $\Phi$ is deterministic, then for arbitrary $U \in O(N)$, $\Phi U$ may not satisfy the same RIP as $\Phi$, and we do not have a good analysis of $\Phi B$ for this case at the moment. Summarize our discussion, we have:

**Theorem 2.2.** *Notation as before.*

*(1) If $A$ is deterministic, then regardless whether $\Phi$ is random or deterministic, $A\Phi$ has RIP if and only if $A$ has full column rank. If that is the case, the RIP constant of $A\Phi$ can be obtained from (14). If $\Phi$ is random, then the probability for $A\Phi$ to satisfy RIP is the same as that of $\Phi$ (with possible different RIP constant).*

*(2) If $\Phi$ is a random matrix satisfying the concentration inequality (6) (hence satisfying the RIP (3) with probability at least p), and $B$ is an $N \times q$ deterministic matrix such that $\delta_k(B) \in (0, \frac{2}{1+\delta_k})$, then with probability at least*

$$1 - \begin{pmatrix} q \\ k \end{pmatrix}(1-p), \tag{19}$$

*the matrix $\Phi B$ satisfies the RIP with the same order as that of $\Phi$ and a possible different RIP constant $\delta_k(\Phi B)$ determined by (18).*

## 3. REDUNDANT BASES IN COMPRESSED SENSING

In this section, we apply Theorem 2.2 to redundant bases setting in compressed sensing. From (8), we see that for given $N$ and $k$, the random matrices of size $n \times N$ generated by the distributions described in (7) satisfy the RIP with high probability as long as $n \geq Ck\log(N/k)$ for some constant $C$. Therefore it is desirable to reduce the integer $k$, i.e. to increase the sparsity level of the signal, by considering redundant bases (or redundant dictionaries). Recall that if a set of vectors $\mathbf{B}$ spans a vector space $V$, then we call $\mathbf{B}$ a basis if $\mathbf{B}$ is linearly independent and call $\mathbf{B}$ a redundant basis otherwise. To apply compressed sensing to a signal $y \in \mathbb{R}^N$ that has a sparse representation $x$ under a redundant basis $\mathbf{B}$ of size $q > N$, we need to consider how the combination of a good sensing matrix with a redundant basis affects the RIP.

Let $B$ be the matrix corresponds to the redundant basis $\mathbf{B}$. Then $B$ is of size $N \times q$ and $y = Bx$ with $x \in \mathbb{R}^q$ sparse. This problem has been considered in [2][11]. In particular, in [11], a detailed analysis of the situation was given. According to Theorem 2.2 in [11], if $\Phi$ satisfies the concentration inequality (6) with [3]

$$n \geq C\delta_k^{-2}[k(\log(\frac{N}{k}) + \log e(1 + \frac{12}{\delta_k})) + \log 2 + t], \quad (20)$$

for some $\delta_k \in (0,1)$ and $t > 0$, then with probability at least $1 - e^{-t}$, the restricted isometry constant of $\Phi B$ satisfies

$$\delta_k(\Phi B) \leq \delta_k(B) + \delta_k(1 + \delta_k(B)). \quad (21)$$

We now apply Theorem 2.2 to obtain a similar result.

**Theorem 3.1.** *Notation as above. With the isometry constant satisfying*

$$\delta_k(\Phi B) \leq \delta_k(B) + \delta_k(\Phi)(1 + \delta_k(B)) \quad (22)$$

*and the probability bound given by (19), the matrix $\Phi B$ satisfies the RIP with the same order as that of $\Phi$.*

*Proof.* One just needs to note that the numbers $\lambda_k$ and $\lambda_1$ which appear in (18) satisfy $1 - \delta_k(B) \leq \lambda_k^2 \leq \lambda_1^2 \leq 1 + \delta_k(B)$. □

For examples of redundant bases satisfying the condition in Theorem 3.1, we refer the readers to [11].

## 4. CONCLUSION AND DISCUSSION

We analyzed the problem of how the multiplication of a matrix to a good sensing matrix affects its RIP. This type of problems arise in CS when one wants to extend the sensing matrix by taking the product of the sensing matrix with another matrix. A particular interesting example is the application of CS under the redundant bases setting. Our result in this short note provides some basic theory for further investigation on the RIP and its applications in CS under different settings. Future work includes constructing good redundant bases, which is related to constructing good deterministic sensing matrices, and analyzing their properties under CS.

[3] There should be a factor $S$ (which is our $k$) for the term $\log(e(1+12/\delta))$ in the bound for $n$ given in [11]. This affects some later estimates in [11].

## 5. REFERENCES


[1] E. Candès, J. Romberg, and T. Tao, "Robust uncertainty principles: Exact sinal reconstruction from highly incomplete frequency information", *IEEE Trans. Inf. Theory 52*, no. 2, pp. 489-509, 2006.

[2] D. Donoho, "Compressed Sensing", *IEEE Trans. Information Theory 52*, no. 4, pp. 1289-1306, 2006.

[3] E. Candès and T. Tao, "Decoding by linear programming", *IEEE Trans. Inf. Theory 51*, no. 12, pp. 4203-4215, 2005.

[4] E. Candès, J. Romberg, and T. Tao, "Stable signal recovery from incomplete and inaccurate measurements", *Comm. Pure Appl. Math. 59*, no. 8, pp. 1207–1223, 2006.

[5] R. Baraniuk, M. Davenport, R. DeVore, and M. Wakin, "A Simple Proof of the Restricted Isometry Property for Random Matrices", (2007), to appear in Constructive Approximation.

[6] M. Davies and R. Gribonval, "Restricted isometry constants where $\ell^p$ sparse recovery can fail for $0 < p \leq 1$, preprint, 2008, available at http://www.compressedsensing.com/.

[7] R. DeVore, "Deterministic Constructions of Compressed Sensing Matrices", *J. of Complexity 23*, pp. 918-925, 2007.

[8] V. Saligrama, "Deterministic Designs with deterministic guarantees: Toeplitz compressed sensing matrices, sequence design and system identification", *arXiv:0806.4958v1*, 6/30/2008.

[9] M. A. Herman and T. Strohmer, "High-resolution radar via compressed sensing", *arXiv:0803.2257v1*, 3/14/2008.

[10] S. Gurevich, R. Hadani, and N. Sochen, "On some deterministic dictionaries supporting sparsity", to appear in the special issure of JFAA on sparsity, 9/1/2007.

[11] H. Rauhut, K. Schnass, and P. Vandergheynst, "Compressed sensing and redundant dictionaries", *IEEE Trans. inf. Theory 54*, no. 5, pp. 2210-2219, 2008.