DATA LOSS RECOVERY FOR WIRELESS SENSOR NETWORKS BASED ON COMPRESSIVE SAMPLING TECHNIQUES

Yuequan Bao*
Harbin Institute of Technology, China

Hui Li
Harbin Institute of Technology, China

Jinping Ou
Harbin Institute of Technology;
Dalian University of Technology; China

ABSTRACT: In a wireless sensor network, data loss often occurs during the data transmission between wireless sensor nodes and the base station, which decreases the communication reliability in wireless sensor network applications. Errors caused by data loss inevitably affect the data analysis of the structure and subsequent decision making. This paper proposed an approach to recover lost data in a wireless sensor network based on the compressive sampling (CS) technique. The main idea in this approach is to project the transmitted data from x onto y, where y is the linear projection of x on a random matrix. The data vector y is permitted to lose part of the original data x in wireless transmissions between the sensor nodes and the base station. After the base station receives the imperfect data, the original data vector x can be reconstructed based on the data y using the CS method. The acceleration data collected from the vibration test of Shandong Harbin Sifangtai Bridge by wireless sensors is used to analyze the data loss recovery ability of the proposed method.

Corresponding author's email: baoyuequan@hit.edu.cn

DATA LOSS RECOVERY FOR WIRELESS SENSOR NETWORKS BASED ON COMPRESSIVE SAMPLING TECHNIQUES

Yuequan Bao[1], Hui Li[1], and Jinping Ou[1,2]

1 School of Civil Engineering, Harbin Institute of Technology, Harbin 150090, China
2 School of Civil and Hydraulic Engineering, Dalian University of Technology, Dalian 116024, China

ABSTRACT: In a wireless sensor network, data loss often occurs during the data transmission between wireless sensor nodes and the base station, which decreases the communication reliability in wireless sensor network applications. Errors caused by data loss inevitably affect the data analysis of the structure and subsequent decision making. This paper proposed an approach to recover lost data in a wireless sensor network based on the compressive sampling (CS) technique. The main idea in this approach is to project the transmitted data from *x* onto *y*, where *y* is the linear projection of *x* on a random matrix. The data vector *y* is permitted to lose part of the original data *x* in wireless transmissions between the sensor nodes and the base station. After the base station receives the imperfect data, the original data vector *x* can be reconstructed based on the data *y* using the CS method. The acceleration data collected from the vibration test of Shandong Harbin Sifangtai Bridge by wireless sensors is used to analyze the data loss recovery ability of the proposed method.

1 GENERAL INSTRUCTIONS

Wireless smart sensors and networks are the future trend for structural health monitoring (SHM) because they not only provide the same functionality as traditional systems at a much lower installed cost but can also process data autonomously using embedded software (Lynch and Loh, 2005). However, some issues such as the time synchronization error, data loss, long time energy supplying and limited bandwidth impede the application of smart wireless sensor and networks in long term monitoring of structures. In the traditional SHM systems, all the data is reliably transferred in a wired system which is no data loss problem. However, in the wireless system, the data loss should be considered.

The data loss during wireless communication has been reported by many researchers. Pei et al. (2005) have measured the range and amount of data loss of different wireless sensors operating on the unlicensed ISM bands in various structural monitoring applications. Kurata et al. (2004) found the data loss in the building risk monitoring using wireless sensor network. Meyer et al. (2010) also found the data loss using wireless sensor to measure the temperature data during some periods of time. They

concluded two reasons of data loss: 1) the stability issues in the communication software on the Motes which cause the communication link break down; 2) bugs in the software can render the base station irresponsive and block the reception of the data packets from the Motes. For the excitation tests conducted using the Motes installed upon the Austrian steel bridge section, data losses as high as 40% are reported (Casciati et al. 2003c).
For the influence of lost data on structural analysis, Nagayama and Spencer (2007) have investigated clearly by simulation and experiment. They showed that the impact of 0.5% data loss is similar with 5% noise addition on PSD estimation. Otherwise, data loss was found to introduce error into the modal analysis. With considering that a corresponding 5 to 10 percent observation noise is unexceptional in the monitoring of civil infrastructure, they indicated that a loss of 0.5% is acceptable.

Since the data loss can not be avoided completely in the data wireless communication procedure, why don not change the sent data packet in the sensor node? That is to change the sent data x to *y*, where the data vector y is part data loss allowed in communication between sensor nodes and base station. When the base station received the data, the original data vector *x* can be reconstructed exactly.

However, how to transform the data *x* to *y*? How to reconstruct the signal form the received data with part of it lost? Recently, a new data acquisition method called compressive sampling (CS) has been introduced (Candes 2006). CS is an alternative theory to Nyquist/Shannon sampling theory that can reconstruct images or signals accurately from far fewer measurements than what is usually considered necessary. The CS acquires the data by liner projection of the original signal to a random measurement matrix. If the signal is spare or near sparse, then the signal can be exactly reconstructed from the measurement. The idea of CS techniques provides a good solution for the ill-condition inverse problem. The inspiration of this paper is also from CS, i.e. the data loss recovery approach is a novel application of CS.

2 LOSS DATA RECOVERY APPROACH

Suppose that we acquire an $m\ (=n)$ length data vector $y \in \Re^m$ by a linear projection of $x \in \Re^n$:

$$y = \mathbf{\Phi} x \tag{1}$$

where x is the original data vector; $\mathbf{\Phi}$ is an $m \times n$ matrix called a *measurement matrix*. Considering the randomly data loss of the m length vector y, the received data $\hat{y} \in \Re^{m-n_l}$ is a $\hat{m} = m - n_l$ dimensional vector, where n_l is the number of lost data of y. The Equation (2) is became as

$$\hat{y} = \hat{\mathbf{\Phi}} x \tag{2}$$

where $\hat{\mathbf{\Phi}}$ is corresponding measurement matrix with $\hat{m}$ rows vectors lost.

Considering $\hat{m} < n$, since $\hat{\mathbf{\Phi}}$ is an $\hat{m} < n$ matrix, the problem of recovering the signal x from the measurements $\hat{y}$ is ill-posed. But if x is known to be sparse or nearly sparse in the basis $\mathbf{\Psi}$, i.e. the coefficients α are mostly zero or nearly zero, then under some conditions α can be reconstructed exactly by solving the convex optimization program (Candes 2006). Therefore, the problem of Equation (2) is became the CS

problem, which is introduced shortly as follows and more contents about CS can be found in the reference Candes (2006).

Any discrete-time signal $x(k)$, $k=1,\ldots,n$, in $\Re^n$ can be represented in terms of an orthonormal basis of $n\times 1$ vectors $\{\psi_i\}_{i=1}^n$. Introducing the $n\times n$ *basis matrix* $\mathbf{\Psi}=[\psi_1,\psi_2,\ldots,\psi_n]$ with the basis vectors ψ_i as columns, then $\mathbf{\Psi}^{-1}=\mathbf{\Psi}^T$ and the vector $x=\left[x(1),\cdots,x(n)\right]^T\in\Re^n$ of n sampled values of $x(k)$ can be expressed as:

$$x=\sum_{j=1}^{n}\alpha_j\psi_j \text{, or } x=\mathbf{\Psi}\alpha \tag{3}$$

where the basis, or weighting, coefficients α_j are given by $\alpha=\mathbf{\Psi}^T x$.

$$\hat{y}=\hat{\mathbf{\Phi}}x=\hat{\mathbf{\Phi}}\mathbf{\Psi}\alpha=\hat{\mathbf{\Theta}}\alpha \tag{4}$$

The basis coefficients α can be reconstructed by solving the l_1 optimization problem:

$$\text{(P1)}\quad \hat{\alpha}=\arg\min\left\|\tilde{\alpha}\right\|_1 \text{ such that } \hat{\mathbf{\Theta}}\tilde{\alpha}=y \tag{5}$$

The solution $\hat{\alpha}$ can recover α exactly under the following conditions:

(i) the signal x is *s-sparse* where a vector is defined to be *s*-sparse if it has at most s nonzero entries;

(ii) the matrix $\hat{\mathbf{\Theta}}$ obeys a so-called *restricted isometry property* (RIP) (Candes 2006), that is, there exists an isometry constant δ_s for the matrix $\hat{\mathbf{\Theta}}$, where δ_s is defined as the smallest number such that

$$(1-\delta_s)\left\|v_s\right\|_2^2\le\left\|\mathbf{\Theta}v_s\right\|_2^2\le(1+\delta_s)\left\|v_s\right\|_2^2 \tag{6}$$

holds for all s-sparse vectors v_s.

3 ROBUST OF THE DATA RECOVNERY APPROACH

Obviously, in the data wireless transmission procedure, the data vector $\hat{y}$ cannot be expected to be received without any noise corruption. If the received data $\hat{y}$ is corrupted with noise during transmission, consider the model

$$\hat{y}=\hat{\mathbf{\Phi}}x+e \tag{7}$$

The problem is became the robust CS problem, the Equation (9) also can be expressed as

$$\hat{y}=\hat{\mathbf{\Phi}}\mathbf{\Psi}\alpha+e=\hat{\mathbf{\Theta}}\alpha+e \tag{8}$$

where e is a bounded but uncertain error with. The basis coefficients α can be reconstructed by solving the convex optimization problem:

$$\text{(P2)}\quad \hat{\alpha}=\arg\min\left\|\tilde{\alpha}\right\|_1 \text{ such that } \left\|\mathbf{\Theta}\tilde{\alpha}-y\right\|_2\le\varepsilon \tag{9}$$

where ε is the chosen bound on the size of the measurement error. As for problem (P1) in Equation (3), problem (P2) also has a unique solution and the optimal basis coefficients $\hat{\alpha}$ may be used to represent the data. The original data can be recovered using the optimal basis coefficients $\hat{\alpha}$ as follows:

$$\hat{x}=\mathbf{\Psi}\alpha \tag{10}$$

where $\mathbf{\Psi}$ is the *basis matrix*.

4 THE PROCEDURE OF PROPOSED LOSS DATA RECOVERY APPROACH

The procedures of the loss data recovery approach can be presented as three steps. First, the original signal x is linearly projected to a random matrix $\mathbf{\Phi}$ to obtain the data $y = \mathbf{\Phi}x$. Second, the data y is sent. It may be lost part of data or corrupted by noise in wireless transmission. The received data in the base station will be changed to $\hat{y}$. Third, the original signal $\hat{x}$ is reconstructed from the received data $\hat{y}$. The procedure of proposed loss data recovery approach is shown in Figure. 1.

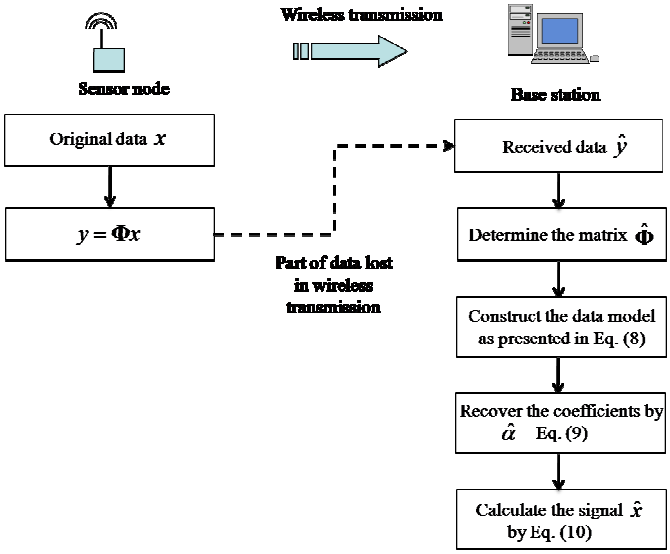


Figure 1. Procedure of the approach

5 INVESTIGATION OF THE VIBRATION DATA MEASURED FROM BRIDGE

5.1 Introduction on the Harbin Sifangtai Bridge

Harbin Sifangtai Bridge as shown in Figure 2 is a large-span cabled stay bridge in highway around the Harbin city.

Figure3. Sensor placement on bridge

Figure 2. Harbin Sifangtai Bridge

Figure4. Data acquisition system

A vibration field test based on the wireless sensor based is carried out for the bridge inspection. Figs. 3 and 4 show the wireless sensor node and the base station used in the vibration test of the bridge. A typical acceleration time history response data is shown in Figure 5.

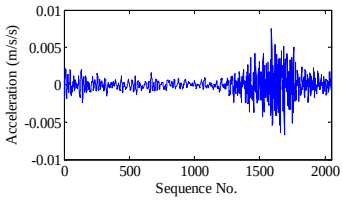


Figure 5. Acceleration response of bridge measured by wireless sensor

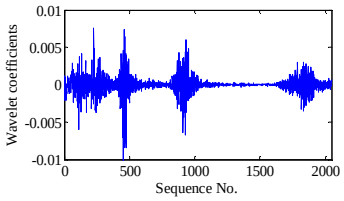


Figure 6. Wavelet coefficients of the acceleration data using the Haar wavelet basis

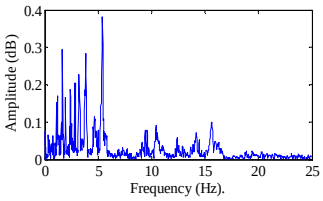


Figure 7. Fourier spectra of the acceleration data

The sampling frequency is 50HZ. The wavelet coefficients of the acceleration data when decomposed by the Haar wavelet basis with five resolution levels are shown in Figure 6. It can be seen that the wavelet coefficients are nearly sparse because part of the coefficients are near to zero. Figure 7 is the Fourier spectra of the acceleration data which shows the vibration of the bridge is low frequency and the nearly sparsity of signal in the Fourier basis.

5.2 Loss data recovery results

5.2.1 Random part of data loss

In this example, the length of original data x is $n=2048$. Keeping the length measurement same with the original data, that is $m=n=2048$. Therefore, there is no any increasing of data acquisition cost for the proposed loss data recovery approach. The dimension of the measurement matrix $\mathbf{\Phi}$ is 2048×2048. To investigate the loss data recovery ability, the data loss rate is $\gamma=(0\%, 2\%, 4\%, ..., 20\%)$. The reconstruction error of $\hat{x}$ is calculated as

$$\xi=\frac{\left\|\tilde{x}-x\right\|_2}{\left\|x\right\|_2} \tag{9}$$

where $\hat{x}$ and x are the reconstruction data and the original data, respectively. The data loss rate is defined as

$$\gamma=100\%\times\frac{n_l}{m} \tag{10}$$

where n_l is the number of lost data.

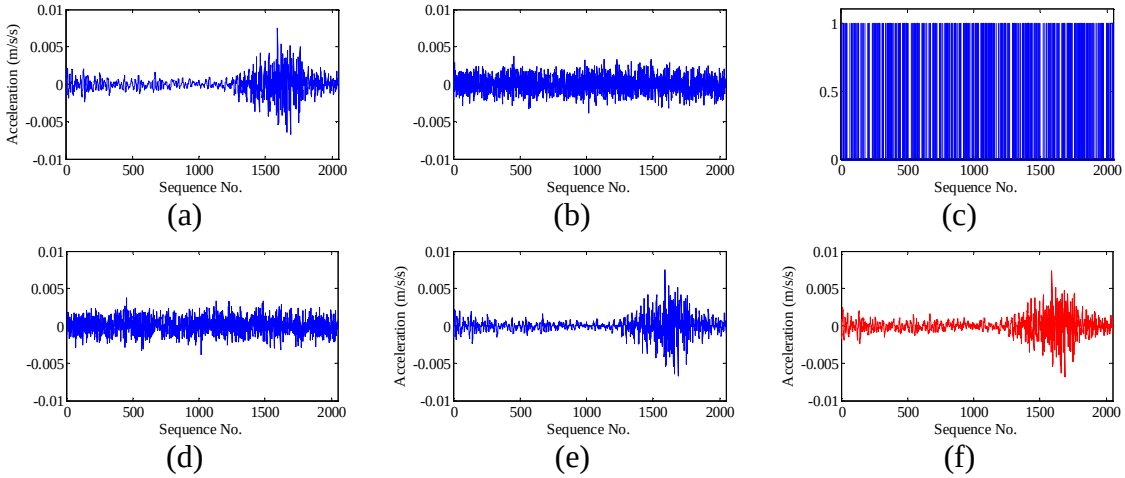
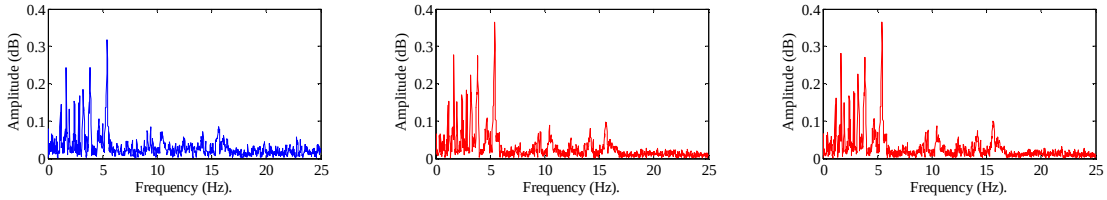


Figure 8. Data recovery with 20% random data loss: (a) original data; (b) sent data *y*; (c)

20% randomly lost data of *y*; (d) Received data on the base station $\hat{y}$; (e) original data *x* with 20% random data loss; (f) Recovered data $\hat{x}$, reconstruction error $\xi = 0.2005$

To illustrate the procedure of the loss data recovery, an example of data recovery with 20% data loss are shown in Figure 8. Figure 8(a) is the original data, the sent data *y* is calculated by $y = \mathbf{\Phi} x$, which is looked like a white noise because the matrix $\mathbf{\Phi}$ is a Gaussian random matrix with zero mean and unit variance. Figure 8(c) shows a vector of location *q* with 20% randomly data lost of *y*. Since part of data is lost, the length of received data $\hat{y}$ as shown in Figure 9(d) at the base station is $\hat{m} = 2048 - \lfloor 2048 \times 20\% \rfloor = 2048 - 409 = 1639$, where the $\lfloor \cdot \rfloor$ is a function that can round the elements to the nearest integers. Using the zero to replace the lost data, the received data $\hat{y} = y(k) \times (1 - q(k))$,$(k = 1, \cdots, n)$ is shown in Figure 9(d). For comparison, the signal original *x* with 20% random data loss of original signal *x* is also shown in Figure 8(e). The reconstructed data $\hat{x}$ can be calculated by Equation (3) and the result with reconstruction error $\xi = 0.2005$ is shown in Figure 8(f).

To better show the influence of data loss and data recovery effects, the Fourier spectra of the original data with 20% random data loss, the recovered data without noise and 5% noise are shown in Figure 10. The Fourier spectra of the recovered data are shown in Figures 10(b, c). There is a reduction in amplitude from the original data. However, the Fourier spectrum of the original data with 20% random data loss has changed a lot compared with the original data. There has not only amplitude reduction but also frequency components loss.

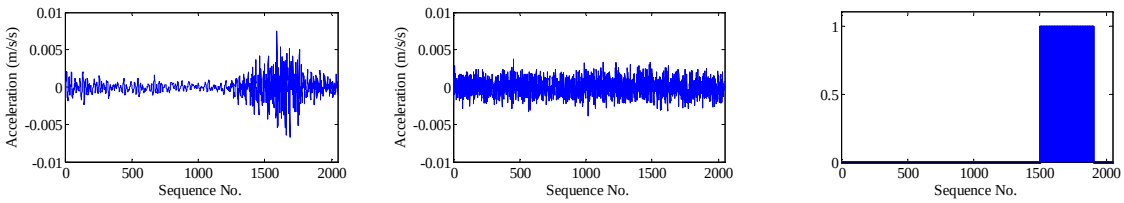


(a) (b) (c)

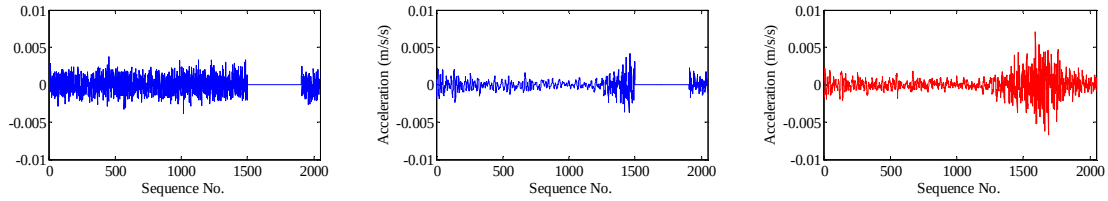
Figure 9. (a) Fourier spectra of original data *x* with 20% random data loss; (c) Fourier spectra of recovered data without noise; (d) Fourier spectra of recovered data without 5% noise

5.2.2 Continuous part of data loss

To consider the continuous data lost case, the data sequence number from 1500-1908 are lost, which the maximal response of the structural in this sequence are included. Therefore, loss of this part data will cause quite unfavourable influences on the analysis of the structural. Figure 11 shows the recovery of results with 20% continuous data lost. Figure 10(e) is original signal with 20% continuous data loss. Figure 10(f) is the recovered data from received signal $\hat{y}$ as shown in Figure 10(d). Figure 10(f) shows that the lost data can be recovered well although the reconstruction error is $\xi = 0.2254$.



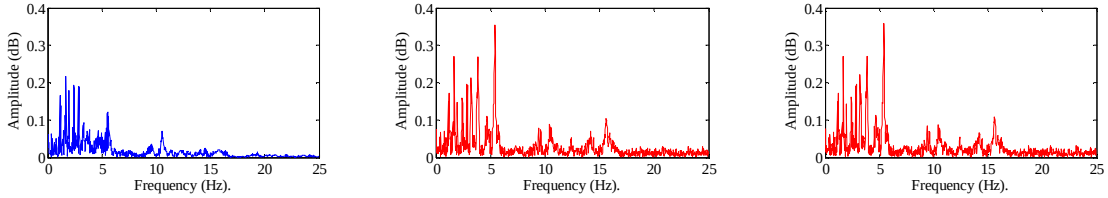
(a) (b) (c)



(d) (e) (f)

Figure 10. Data recovery with 20% continuous data loss: (a) original data; (b) sent data y; (c) locations of 20% selected lost data, the data sequence number from 1500-1908 are lost; (d) Received data on the base station $\hat{y}$; (e) original data x with 20% random data loss; (f) Recovered data $\hat{x}$, reconstruction error $\xi = 0.2254$.

In the frequency domain, the influence of 20% loss of data on the original x is remarkable as shown in Figure 11 (a), which indicates lots of information has lost. However, the recovered data as shown in Figure 11(b) and (c) is much better, which illustrate the good performance of the proposed approach.



(a) (b) (c)

Figure 11. (a) Fourier spectra of original data x with 20% continuous data loss; (b) Fourier spectra of recovered data without noise; (c) Fourier spectra of recovered data without 5% noise;

6 CONCLUSIONS

This paper proposed a novel approach for recovery of lost data of wireless sensor networks in SHM. An example of the wireless sensor data measured from a real bridge is carried to illustrate the data recovery ability of the proposed approach. Some conclusions can be drawn as follows.

The data loss is harmful and it not only changed amplitude of signal but also caused the lost of frequency components in frequency domain. Comparing with the random data loss, the continuous data will cause more unfavourable effects.

The proposed data recovery approach can recover data well in case of part of data lost without any increasing of data acquisition cost, even with additional noise corruption in data transform. The reconstruction error is unavoidable and it is increased with increasing of data loss rate. However reconstruction error is acceptable compared with bad effects caused by the data loss.

ACKNOWLEDGEMENTS

This research is supported by NSFC (Grant No.51008095), and Natural Scientific Research Innovation Foundation in Harbin Institute of Technology (Project No. HIT.NSRIF.2009097) and China Postdoctoral Science Foundation (Grant No. 20090460909).

REFERENCES

J.P. Lynch and K.J. Loh, A summary review of wireless sensors and sensor networks for structural health monitoring, Shock and Vibration Digest, 38(2): 91-128, 2005

Pei, JS, Kapoor, C, Graves-Abe, TL, Sugeng Y, and Lynch, JP. (2005). Critical design parameters and operating conditions of wireless sensor units for structural health monitoring, in Proc. of the 23rd International Modal Analysis Conference, Orlando, FL.

Kurata, N, Spencer BF, Jr, and Ruiz-Sandoval M. (2004). Building risk monitoring using wireless sensor network, Proc. of the 13th world conference on earthquake engineering, August 2-6, Vancouver, BC, Canada.

Meyer, J, Bischoff, RB, Feltrin G, and Motavalli, M. (2010). Wireless sensor network for long-term structural health monitoring, Smart Structures and Systems, 6(3): 263-275, 2010.

Casciati, F, Faraveli L, and Borghetti, F. (2003). Wireless links between sensor-device control stations in long span bridges, Proceedings of the SPIE, Vol. 5057, San Diego, CA, pp.1-7.

Nagayama, T and Spencer BF, Jr. (2007). Structural health monitoring using smart sensors, NSEL report, Series 001, http://hdl.handle.net/2142/3521.

Candès, EJ. (2006) Compressive Sampling, Proceedings of the International Congress of Mathematicians, Madrid, Spain: 1433-1452.