23rd International Symposium on Mathematical Theory of Networks and Systems
Hong Kong University of Science and Technology, Hong Kong, July 16-20, 2018

Sparse Regression using Compressive Sensing with Input Shaping

Vivek S. Borkar[1], Vikranth R. Dwaracherla[2] and Neeraja Sahasrabudhe[3]

***Abstract*— We present a compressive sensing approach for sparse regression problem and using the ideas from concentration of measure, extend it to case where the samples are generated from a Markov chain. This is achieved by constructing a matrix from samples such that it is suitable for reconstructing sparse parameter vectors via l_1-minimization.**

***Index Terms*— Input and excitation design, Recursive identification, Sparse Regression, Compressive Sensing, Subgaussianity.**

I. INTRODUCTION

We consider the problem of sparse linear regression where, given pair samples (X_n, Y_n), we aim to fit a model $Y_n = AX_n +$ noise, where A is sparse. We do so using ideas from compressive sensing. The main aim of compressive sensing is to reconstruct a sparse vector from linear measurements of the vector such that the number of observed measurements m is significantly smaller than the dimension n of the original vector (See [2], [3]). The problem of reconstruction can be formulated as an l_1-minimization problem. In classical compressive sensing, however, the paradigm is $y = Ax +$ *noise*, where x is the unknown sparse vector to be reconstructed, y the observed quantity and A the measurement matrix *chosen* to satisfy with high probability the so called '*Restricted Isometry Property (RIP)*'. In [1], [8], [9], RIP is proved for different types of matrices. We interchange the roles of A and x so that $\{X_n\}$ serves as a surrogate for measurement matrix and A the unknown sparse entity to be reconstructed. The key step is to aggregate $\{X_n\}$ suitably so that the subgaussian concentration effect kicks in and the desired variant of RIP follows from the known properties of subgaussian matrices. In other words, we use the input suitably to come up with subgaussian matrices X and then invoke the notions of compressive sensing to estimate the rows of matrix A.

We consider matrices such that the entries are either i.i.d. or are sampled from a Markov chain. In the latter case, the entries or even the rows (or the columns) of the matrix are not independent. Loosely speaking, one can prove that a function $f : \mathcal{R}^d \mapsto \mathcal{R}$ is concentrated around its mean if the fluctuations in every coordinate (with rest of the coordinates fixed) are bounded. This phenomenon has been studied in detail when f is a function of independent (or weakly dependent) variables. Recently in [4], [7], functions defined on irreducible aperiodic Markov chains satisfying certain conditions (viz. geometric ergodicity) have been studied and certain concentration results have been obtained. These can be used to prove RIP for such matrices.

While the idea of using compressive sensing for sparse regression is not new (See, e.g., [6]), our main contribution is the observation that a judicious aggregation of input samples to shape their distribution can yield significant performance enhancement, in both IID and Markov samples.

II. BASIC SET-UP

Suppose $Y = AX$ where $A \in \mathcal{R}^{m\times n}$ is a sparse matrix - by this we mean that each row of A (denoted by A_i for $i = 1, \ldots, m$) is sparse, X is a vector in $\mathcal{R}^n$ and $Y \in \mathcal{R}^m$. Consider the classical linear regression problem

$$\min_{A\in\mathcal{R}^{m\times n}} \|Y - AX\|^2$$

We use ideas from compressive sensing to estimate A_i for $i = 1, \ldots, m$.

We can write $Y^T = X^T A^T$ (where M^T denotes the transpose of a matrix M). By y_j we denote the j^{th} entry of Y^T. The problem of estimating the parameters A now transforms into that of estimating sparse vectors A_j from the m equations

$$y_j = X^T A_j \quad \text{for } j = 1, \ldots, m. \tag{1}$$

Let $[m]$ denote the set $\{1, \ldots, m\}$. For each $j \in [m]$, given a sufficiently large number of noisy samples of $\{y_j, X\}$, we would like to find a sparse vector A_j^* such that

$$y_j \approx XA_j^* \ \ \forall j$$

where y_j denotes the vector $(y_j^1, \ldots, y_j^k)$ of k samples of j^{th} entry of Y^T and by abuse of notation X also denotes the matrix in $\mathcal{R}^{k\times n}$ such that each row is a sample of the vector X.

Throughout the paper $\|\cdot\|_1$ denotes the l_1 norm and $\|\cdot\|$ denotes the l_2 norm over $\mathcal{R}$.

III. COMPRESSIVE SENSING APPROACH

Fix $j \in [m]$. In the above scenario, suppose we have a large sample ($l \times k$ for $l, k > 0$) of $\mathcal{S} = \{y_j^i, X^i\}$ and $i = 1, \ldots, lk$. Let $Y_j = (\bar{y}_j^1, \ldots, \bar{y}_j^k)$, where each entry is averaged over l sample points from $\mathcal{S}$.

Let $\bar{X}$ denote the matrix with k rows where each row is obtained by averaging over l samples from $\mathcal{S}$. Assume that

[1] Department of Electrical Engineering, Indian Institute of Technology, Powai, Mumbai - 400076. borkar.vs@gmail.com
[2] Department of Electrical Engineering, Stanford, USA. vikranth.dwaracherla@gmail.com
[3] Department of Mathematical Sciences, Indian Institute of Science Education and Research, Mohali - 140306. neeraja.budhey@gmail.com

the samples are drawn from distributions with zero mean and variance I.

Given that that A_j is sparse for every j, according to the theory of compressive sensing, the above problem of estimating A_j can be framed as an l_1-minimization:

$$\underset{z\in\mathcal{R}^n}{\text{minimize}} \ \|z\|_1 \ \text{ subject to } \|Y_j - \bar{X}z\| \leq \eta \qquad (2)$$

where $\eta > 0$ denotes the bound on measurement error. For $\bar{X}$ to be suitable for reconstruction of A_j's via compressive sensing, it suffices to show that the matrix $\bar{X}$ satisfies the restricted isometry property.

Remark: We assume, without loss of generality, Y_j, X to be zero-mean and X with identity covariance matrix. Our analysis will be under these conditions. More generally, deviation from this will affect only the constants in the estimates and not how they scale with the number of samples. Based on above discussion, we propose the following algorithm:

Algorithm 1 Sparse regression via compressing sensing

Given: $\{Y_i, X_i\} \leftarrow$ **large sample set.**

- $X^j \in \mathcal{R}^{k\times n},\ 1 \leq j \leq l$**: Obtained by putting** X_i **together as rows of a matrix.**
- $\bar{X} = \frac{1}{l}\sum_{j=1}^{l} X^j$ **(Averaging the matrix** X **over** l **samples).**
- l_1**-minimization:**

  $\underset{z\in\mathcal{R}^n}{\textbf{minimize}} \ \|z\|_1$ **subject to** $Y_j = \bar{X}z$**.**

Output: Estimated vector A_j^***.**

A. Main result

In this section, we prove some theoretical results to justify the reconstruction scheme illustrated via simulations in the next section.

Definition 3.1: Let $M \in \mathcal{R}^{k\times n}$ be a random matrix such that the entries of M are independent mean-zero subgaussian random variables with variance 1 and common subgaussian parameters α, β, that is

$$P(|M_{ij}| \geq t) \leq \beta e^{-\alpha t^2} \qquad \forall\ t > 0, i \in [k], j \in [n]$$

then M is called a subgaussian random matrix.

It is known that for a suitable choice of k, subgaussian random matrices satisfy the restricted isometry property with high probability.

Definition 3.2 (Subgaussian Random Vector): If, for all $v \in \mathcal{R}^n$ with $\|v\| = 1$, the random variable $\langle Y, v\rangle$ is subgaussian with subgaussianity parameters being independent of v, then Y is called a subgaussain random vector.

From the theory of compressive sensing, we know that a matrix $M \in \mathcal{R}^{k\times n}$ is "good" for reconstruction of an s-sparse vector in $\mathcal{R}^n$ if it satisfies the restricted isometry property (RIP) with high probability:

Definition 3.3 (Restricted Isometry Property): The sth restricted isometry constant $\delta_s = \delta_s(A)$ of a matrix $A \in \mathcal{R}^{m\times n}$ is the smallest $\delta \geq 0$ such that

$$(1-\delta)\|x\|_2^2 \leq \|Ax\|_2^2 \leq (1+\delta)\|x\|_2^2$$

for all s-sparse vectors $x \in \mathcal{R}^n$. We say that A satisfies the restricted isometry property if δ_s is small for moderately large s.

Theorems 9.9 and 9.11 in [5] imply that to prove RIP property for a matrix M, it is sufficient to prove that it satisfies the following concentration inequality

$$P\left(|k^{-1}\|Mx\|^2 - \|x\|^2| \geq t\|x\|^2\right) \leq 2\exp(-ct^2k) \qquad (3)$$

A simple application of Bernstein's inequality shows that the above concentration inequality holds for a matrix with subgaussian rows.

B. Independent samples

We motivate our results by first considering the simple case of independent samples. Hoeffding lemma implies that if the entries of the matrix X are bounded, identically distributed and are sampled independently, then the resulting matrix is subgaussian and is therefore suitable for reconstruction of sparse vectors. However, averaging over the samples for each entry should improve the reconstruction error specially for high dimensional vectors. This is illustrated in Theorem 3.4 below. The sampling process ensures that we get subgaussian concentration for each entry and that leads us to conclude that $\bar{X}$ satisfies the restricted isometry property with high probability. A precise bound can be obtained by combining known results as follows:

Theorem 3.4: Let each sample of X_{ij} be i.i.d. (with zero mean and variance 1) such that $a \leq X_{ij} \leq b$. Note that by construction the random variables are zero mean and with variance 1. Then, there exist C_1, C_2 depending only on l, b, a such that if, for $\epsilon \in (0,1)$,

$$k \geq C_1 s\log(en/s) + C_2\ln(2\epsilon^{-1})$$

then with probability at least $1-\epsilon$ every s-sparse vector v is recovered from $y = \bar{X}v$ via l_1-minimization.

Proof: Recall that each entry of $\bar{X}$, denoted by $\bar{X}_{ij}$, is averaged over l samples (and is zero-mean by construction).

Since each sample is independent, by Hoeffding's inequality, we have,

$$P\left(|\bar{X}_{ij}| \geq t\right) \leq 2\exp\left(-\frac{2lt^2}{(b-a)^2}\right) \tag{4}$$

This means that $\bar{X}$ is a subgaussian random matrix. The statement then follows from Theorem 9.2 [5]. ∎

Remark: Averaging each entry over l samples helps in improving the subgaussian concentration bound by a factor of l. The improvement is illustrated in simulations in section IV.

In real problems, it is inconceivable to measure a signal with complete precision. This means that the measurement vector y is corrupted with some measurement error. However, from Theorem 9.13 [5], we know that if a matrix satisfies restricted isometry property, it is suitable for a robust recovery via l_1-minimization. Thus, $\bar{X}$ works as a recovery matrix even when $\|y - \bar{X}v\| \leq \eta$ for some $\eta > 0$.

By using generalizations of Hoeffding's inequality, we can obtain similar results under alternative hypotheses. For example, if the sample generated for $\bar{X}$ comes from a martingale difference sequence, we can use Azuma inequality to prove a similar reconstruction result.

C. Dependent Samples

Definition 3.5 (Isotropic random vector): Given a random vector $Y \in \mathcal{R}^n$ with $E[Y] = 0$, Y is called isotropic if $E|\langle Y, v\rangle|^2 = \|v\|^2$ for all $v \in \mathcal{R}^n$

In Theorem 3.4, since we have independent entries with zero mean and variance 1, the isotropic property comes for free. More generally, the results below hold modulo a scale factor for constants as pointed at earlier. This is because any distribution with finite second moment can be rendered isotropic by an affine transformation.

Theorem 3.6 (Theorem 5.71, [9]): Let $B = (b_{ij})$ be an $k \times n$ matrix whose rows B_i are independent isotropic random vectors in $\mathcal{R}^n$. Let $R > 0$ be such that all entries $|b_{ij}| \leq R$ almost surely. Then the normalized matrix $\bar{B} = \frac{1}{\sqrt{k}}B$ satisfies the following for $k \leq n$, for every sparsity level $1 < s \leq n$ and every number $\delta \in (0, 1)$:

$$\text{if } \; k \geq C\delta^{-2}s\log n\log^2(s)\log(\delta^{-2}s\log n\log^2 s)$$

then $E[\delta_s(\bar{B})] \leq \delta$. Here, $C = C_R > 0$ may depend only on R.

Note that there is no averaging here. This means that matrices with rows sampled independently such that they are bounded and isotropic would work for recovery via l_1-minimization with suitable k without the averaging step. However, the results are expected to improve with averaging because of improvement in subgaussian parameters for tails.

D. Markovian sampling

It is well-known that under suitable conditions a Markov chain exhibits rapid mixing and hence a fast convergence to the stationary distribution π. The concentration of measure phenomenon has also been studied in this context and it turns out that under suitable conditions Markov chains also display strong concentration of measure in the stationary distribution. We assume that rows of $\bar{X}$ are generated via averaging over l samples from a finite state space Markov process. Suppose V_n denotes an irreducible aperiodic Markov chain. That is, $\bar{X}_{ij} = \frac{1}{l}\sum_{k=r}^{r+l-1} V_k$, where $r = ((i-1)n + j)l + 1$. We first state a general theorem for concentration of an irreducible aperiodic Markov chain.

Theorem 3.7 (Theorem 2, [4]): Let (X_n) be an irreducible aperiodic Markov chain which is geometrically ergodic on a space S. Let C be a small set such that the tails of the return time to C are exponential. Let $f : \mathcal{R}^n \mapsto \mathcal{R}$ be separately bounded. That is, for $i = 1, \ldots, n$ and some constants L_i:

$$|f(X_1, \ldots, X_{i-1}, X_i, X_{i+1}, \ldots, X_n) - f(X_1, \ldots, X_{i-1}, X_i', X_{i+1}, \ldots, X_n)| \leq L_i$$

Let π be the stationary distribution of $\{X_n\}$. Then there exists M_0 (depending on C) such that for all $t > 0$,

$$P_\pi(|f(X_1, \ldots, X_n) - E_\pi f(X_1, \ldots, X_n)| > t) \leq 2e^{-M_0 t^2/\sum L_i^2}$$

Using this Theorem, we prove the following:

Proposition 3.8: Let $\bar{X} = (\bar{x}_{ij})$ be an $k \times n$ matrix whose rows $\bar{X}_i$ are isotropic random vectors in $\mathcal{R}^n$ averaged over samples x_{ij} generated from a geometrically ergodic irreducible aperiodic Markov chain satisfying the conditions in theorem 3.7. Let $R > 0$ be such that all entries $|x_{ij}| < R$ almost surely. Then for all $u \in \mathcal{R}^n$ and every $t \in (0, 1)$, $\bar{X}$ satisfies the concentration inequality of the form (3).

Proof: Without loss of generality assume that $\|x\| = 1$. We show that the concentration inequality of the form (3) is satisfied for matrix $M := \bar{X} \in \mathcal{R}^{k\times n}$ when: (i) M has independent (and isotropic) rows with each row averaged over samples from the given Markov chain, and (ii) M has (dependent but isotropic) rows where each row is averaged over samples from the given Markov chain. Let M_i denote the rows of M for $1 \leq i \leq k$.

1) Assume that M_i are independent and isotropic. Define f, a real-valued function on $\mathcal{R}^l$ by $f(x_{i_1}, x_{i_2}, \ldots, x_{i_l}) = \frac{1}{l}\sum_{j=1}^{l} x_{i_j}$. where x_{i_j} are sampled from the given Markov chain. Then,

$$|f(x_{i_1}, \ldots, x_{i_j}, \ldots, x_{i_l}) - f(x_{i_1}, \ldots, x_{i_j}', \ldots, x_{i_l})| \leq 2R/l.$$

Thus f is separately bounded. Observe that each entry of M is of the form $\frac{1}{l}\sum_{j=1}^{l} x_{i_j}$. Thus, there exists M_0 such that: $P_\pi(|M_{ij}| > t) \leq 2e^{-M_0 l t^2/4R^2}$. This implies that each entry of M is subgaussian. This means that $|\langle M_i, x\rangle|$ is subgaussian with parameters $\beta = 2, \kappa = \frac{M_0 l}{4nR^2}$, for every x such that $\|x\| = 1$. Then, using the same idea as in the proof of Lemma 9.8 in [5], define:

$$Z_i = |\langle M_i, x\rangle|^2 - \|x\|_2^2, \;\; i \in \{1, \ldots, k\}.$$

Since M_i is isotropic, we have $E_\pi Z_i = 0$. Further, Z_i is subexponential because $\langle M_i, x\rangle$ is subgaussian, that is, $P_\pi(|Z_i| \geq r) \leq \beta \exp(-\kappa r)$ for all $r > 0$ and $\beta = 2, \kappa = \frac{M_0 l}{4nR^2}$. By independence of the M_i, Z_i are also independent. Now,

$$\frac{1}{k}\sum_{i=1}^{k} Z_i = \frac{1}{k}\sum_{i=1}^{k}(|\langle M_i, x\rangle|^2 - \|x\|_2^2).$$

By Bernstein inequality for subexponential random variables and the fact that $t \in (0, 1)$, we get:

$$P_\pi\left(\left|k^{-1}\sum_{i=1}^{k} Z_i\right| \geq t\right) \leq 2\exp\left(-\frac{\kappa^2}{4\beta + 2\kappa} k t^2\right).$$

where, $\beta = 2, \kappa = \frac{M_0 l}{4nR^2}$.

2) Assume that M_i are isotropic but not independent. Represent the matrix $M \in \mathcal{R}^{k\times n}$ as a vector in $\mathcal{R}^{lkn}$ such that each entry is sampled from the given Markov chain. By abuse of notation we again denote this vector by M. For every $x \in \mathcal{R}^n$, with $\|x\| = 1$, define $f_x : \mathcal{R}^{lkn} \mapsto \mathcal{R}$ by $f_x(M) = \frac{1}{k}\sum_{i=1}^{k} |\langle M_i, x\rangle|^2$, where the $f_x(M)$ is obtained by first sequentially averaging over batches of l n-dimensional vectors in M to get $(M_1, M_2, \ldots, M_k) \in \mathcal{R}^{kn}$. Let M' denote a vector in $\mathcal{R}^{lkn}$ such that it differs from M at one entry. Then, we have

$$|f_x(M) - f_x(M')| \leq \frac{4R^2}{lk}$$

Since $\|Mx\| = \sum_{i=i}^{k} |\langle M_i, x\rangle|^2$, by Theorem 3.7, there exists M_0' such that:

$$P_\pi\left(\frac{1}{k}\|Mx\| - E_\pi\|Mx\|| \geq t\right) \leq 2e^{-\frac{M_0' l k t^2}{16nR^4}}$$

holds for every $x \in \mathcal{R}^n$ and $t \geq 0$. The inequality in (3) follows from the fact that M_is are isotropic.

■

Combining this with Theorem 9.2 from [5], we have the following reconstruction theorem for markovian samples.

Theorem 3.9: Let entries x_{ij} of $X \in \mathcal{R}^{k\times n}$ be sampled from an irreducible aperiodic Markov chain which is geometrically ergodic on a space S satisfying the conditions in theorem 3.7, with stationary distribution π. Assume that $|x_{ij}| \leq R$ almost surely for some $R > 0$. Then there exist constants $C_1, C_2 > 0$ depending on R, M_0 (as above) for $i = 1, \ldots, n$ such that if for $\epsilon \in (0, 1)$

$$k \geq C_1 s \log(en/s) + C_2 \log(2\epsilon^{-1}),$$

then with probability $1 - \epsilon$ every s-sparse vector $v \in \mathcal{R}^n$ is recovered from $y = \bar{X}v$ via l_1-minimization.

IV. Simulations

Simulations are carried out for the measurement matrix obtained by both averaging over samples as well as without averaging. Figures show 5 subgraphs that illustrate - error between the actual dependency and one estimated from our method with averaging (i.e. using $\bar{X}$ as measurement matrix), the actual dependency vector, the estimated dependency vector, error between the actual and estimated sparse dependency without averaging (i.e. using X as measurement matrix) and the estimated dependency vector using our method without averaging respectively.

A. Independent samples

Each element of the sample X is an i.i.d. random variable. We consider a vector with dimension $n = 10000$ and sparsity $s = 7$. Number of samples considered are $lk = 7500$, i.e., we are averaging over $l = 100$ rows to obtain a single row of $\bar{X} \in \mathcal{R}^{75\times 10000}$. A small noise was introduced into the system with variance σ^2. For the simulations below, we take $\sigma^2 = 0.1$.

In Figure 1, samples are obtained from F-distribution with 8 degrees of freedom in the numerator and 5 degrees of freedom in the denominator. Increasing k (number of observations) helps in reducing the error. At higher values of k, the reconstruction without averaging is almost as good as the reconstruction with averaging. This is illustrated in Figure 2. This means that there is a trade-off between choosing k and l. However, as we shall see later, this is not the case for dependent (markovian) samples and in those cases averaging is necessary.

B. Markov model perturbed with nonlinearity

we generate entries x_{ij} of X from the following:

$$X_{n+1} = \alpha X_n + f(X_n) + \zeta_n \tag{5}$$

where $\alpha \in [0, 1)$, $f : \mathcal{R} \mapsto \mathcal{R}$ is a nonlinear function and ζ_n is a random noise.

1) Independent Rows: Each row is generated independently by choosing the first value of the row independently. We consider reconstruction by both averaging over the samples and without averaging and observe that the improvement is marginal. This is consistent with the theoretical results. For simulation purposes, $\alpha = 0.9$, $f(x) = \sin(x) + \log sig(x) + 0.1e^{-|x|}$ and ζ_n is a random variable were chosen, where $\log sig$ denotes the log-sigmoid function.

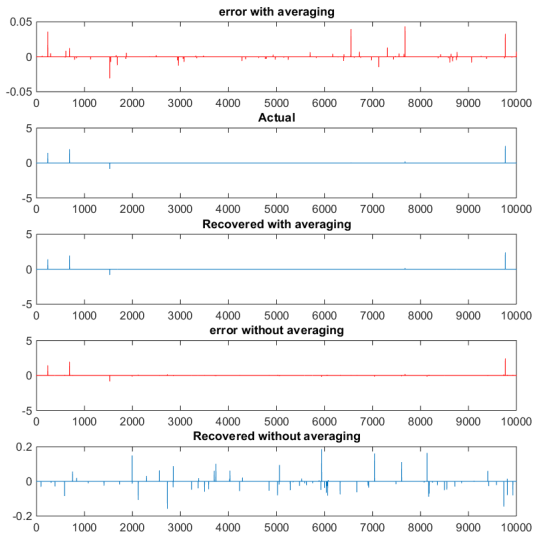

Fig. 1. Sparse recovery with $k = 75$ using elements from F-distribution.

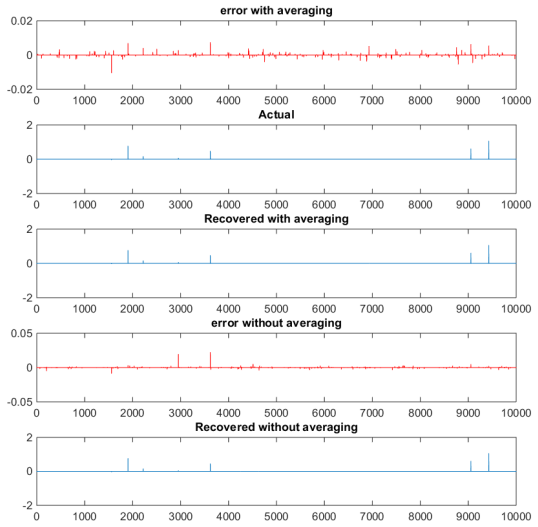

Fig. 2. Sparse recovery with $k = 250$ using elements from F-distribution.

Figure 3 shows and example where the random noise is generated from F-distribution with 8 degrees of freedom in the numerator and 5 degrees of freedom in the denominator.

2) Dependent Rows: We generate all the elements from the markov chain described in the equation (5). For generating each element the previous value is used. So, we generate a large vector with these values and rearrange them in the form of rows of X. We look at the case of averaging over the samples as well as without averaging by considering the first m samples for reconstruction.

Figure 4 illustrates an example where the random noise

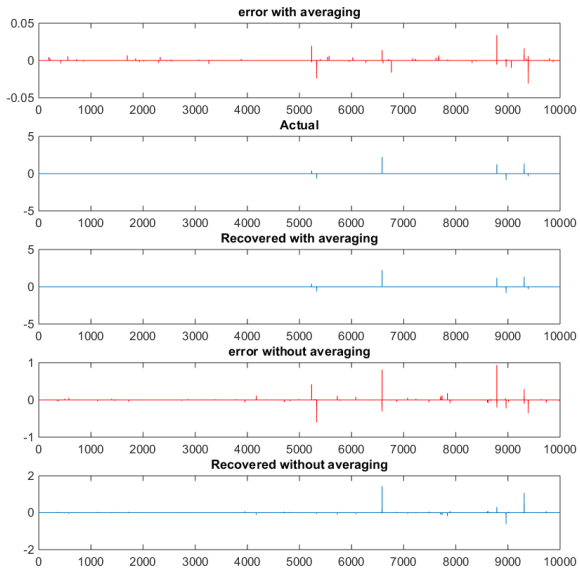

Fig. 3. Sparse recovery using the random noise in the Markov process as F-random variable

is generated from random variable generated by linear combination of Laplacian with mean 15 and variance 5 and F-distribution with 8 degrees of freedom in the numerator and 5 degrees of freedom in the denominator.

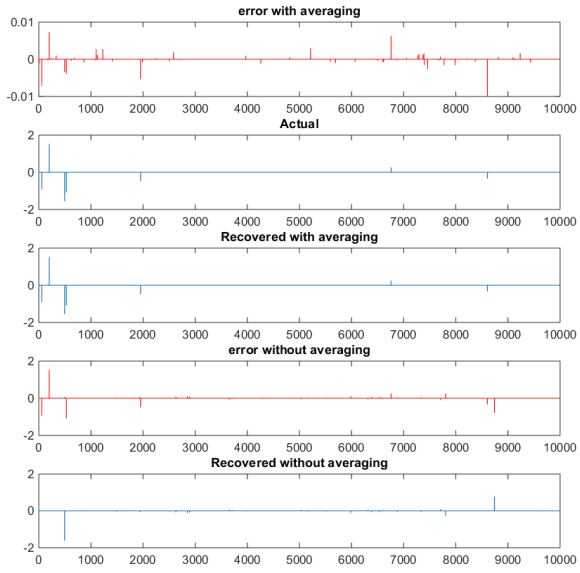

Fig. 4. Sparse recovery using the random noise in the markov process as combination of Laplacian and F-random variable.

We also consider an example with random noise with arbitrary distribution shown in Figure 5. This is illustrated in Figure 6.

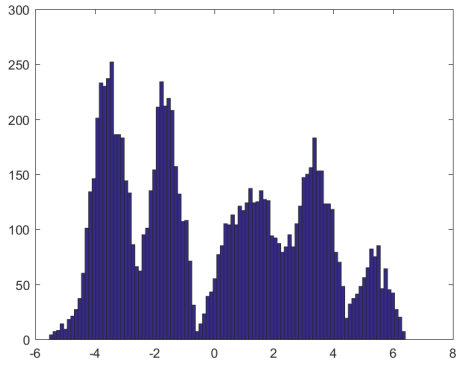

Fig. 5. Distribution of random variable used as noise in the markov process.

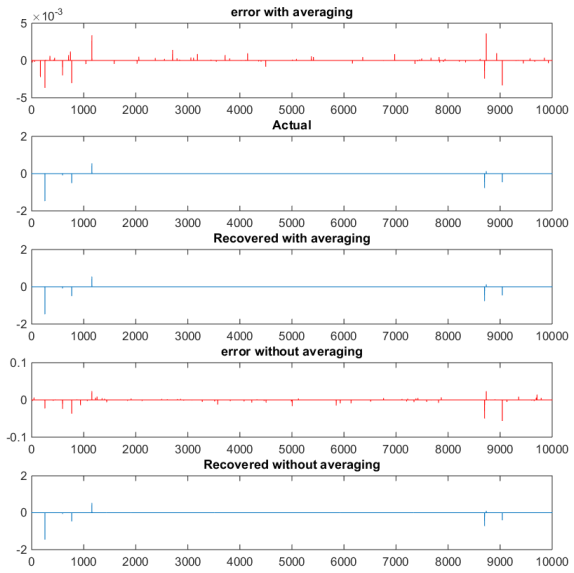

Fig. 6. Sparse recovery using the random noise in the markov process as shown in Figure 5.

C. Comparison with LASSO: Indoor localization

In this section, we test our method on real data and compare it with plain LASSO method. For this purpose, UCI Machine Learning database has been used. It is available online at: *https://archive.ics.uci.edu/ml/datasets/UJIIndoorLoc*.

This database can be used for testing regression methods to estimate indoor location (latitude and longitude) using data on wireless protocols. In the following simulation, $n = 450$, number of samples $lk = 20000$ and $l = 200$ samples are averaged to get a single row in $\bar{X}$ (this means, $k = 100$). So, $\bar{X} \in \mathcal{R}^{100\times 450}$.

We have compared our result with the LASSO method available in glmnet package (*http://web.stanford.edu/hastie/glmnet/glmnet_beta.html*).

Comparisons were done for A_j (for $j = 1, 2$) obtained using the training Set, with the validation set available. Here, A_1 corresponds to longitude and A_2 corresponds to latitude. Figures 7 and 8 show the coefficients of A_1 and the fractional error in construction respectively. Similarly, Figures 9 and 10 show the coefficients of A_2 and the fractional error in construction respectively.

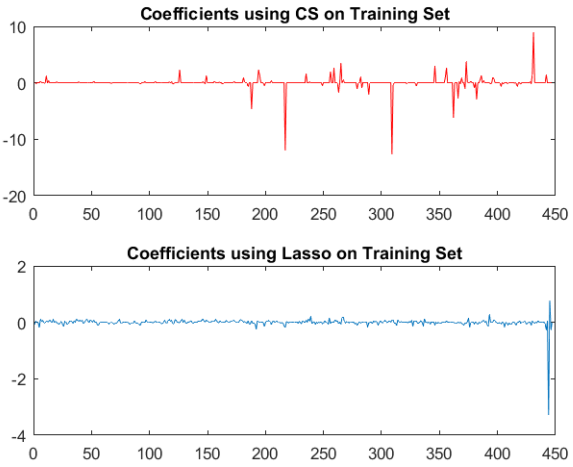

Fig. 7. Coefficients of A_1 (corresponding to longitude).

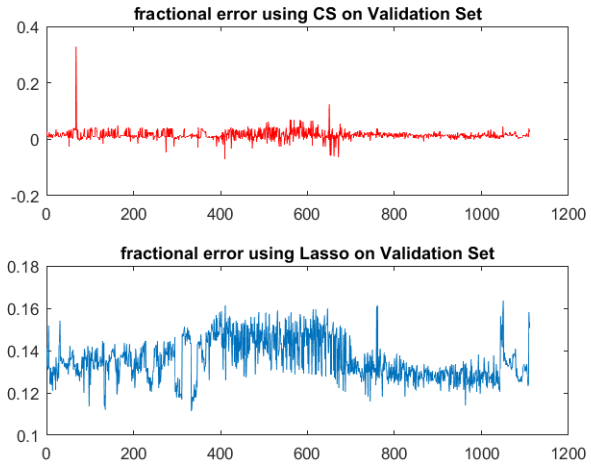

Fig. 8. Fractional error in reconstruction.

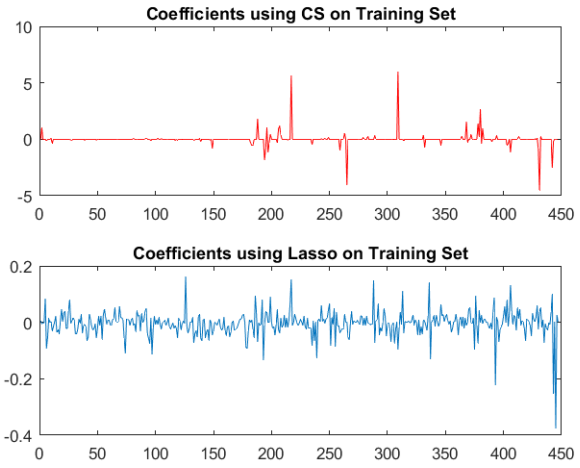

Fig. 9. Coefficients of A_2 (corresponding to latitude).

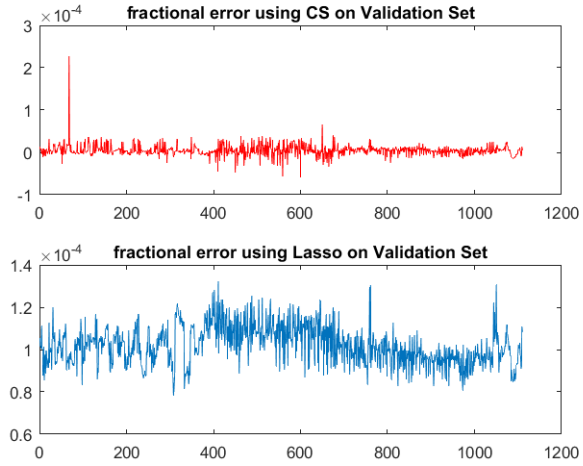

Fig. 10. Fractional error in reconstruction.

V. Concluding remarks

The simulations confirm that while averaging does not necessarily give significant advantage when entries of X are independent, it leads to visible improvement when the entries are not independent. As mentioned earlier, this is because after sufficient averaging, owing to the concentration of measure phenomenon, $\bar{X}$ behaves like a subgaussian matrix.

Acknowledgements

This work was supported in part by CEFIPRA grant No. IFC/DST-Inria-2016-01/448 “Machine Learning for Network Analytics”. The work of Vivek S. Borkar was supported in part by J. C. Bose Fellowship and a grant ‘Approximation of High Dimensional Optimization and Control Problems’ by Department of Science and Technology, Government of India. The work of Neeraja Sahasrabudhe was supported in part by the DST-INSPIRE Faculty Fellowship from the Government of India.

References

[1] R. Adamczak, A. Litvak, A. Pajor, N. Tomczak-Jaegermann, Restricted isometry property of matrices with independent columns and neighborly polytopes by random sampling, Constructive Approximation, 2010.
[2] E. Candés, J. Romberg and T. Tao, Robust uncertainty principles: exact signal reconstruction from highly incomplete frequency information, IEEE Trans. Inform. Theor. 52(2), pp. 489-509, 2006.
[3] D. L. Donoho, Compressed sensing, IEEE Trans. Inform. Theor. 52(4), pp. 1289-1306, 2006.
[4] J. Dedecker and S. Gouëzel, Subgaussian concentration inequalities for geometrically ergodic Markov chains, Electronic Communications in Probability, 20(64), pp. 1-12, 2015.
[5] S. Foucart and H. Rauhut, A Mathematical Introduction to Compressive Sensing, Birkhäuser, New York, 2013.
[6] A. Emad and O. Milenkovic, CaSPIAN: A causal compressive sensing algorithm for discovering directed interactions in gene networks, PLoS ONE. 9(3), 2014.
[7] D. Paulin, Concentration inequalities for Markov chains by Marton couplings and spectral methods, Electronic Journal of Probability, 20, pp. 1-32, 2015.
[8] H. Rauhut, Compressive Sensing and Structured Random Matrices, Theoretical Foundations and Numerical Methods for Sparse Recovery, M. Radon Series Comp. Appl. Math XX, 9, pp. 1-92, 2010.
[9] R. Vershynin, Introduction to the non-asymptotic analysis of random matrices, Compressed Sensing, Theory and Applications, Edited by Y. Eldar and G. Kutyniok, Cambridge University Press, pp. 210-268, 2012.