

On Concentration Inequalities for Sparse Vectors

Kai Zhang

Computer and Information Science and Engineering
University of Florida
Gainesville, Florida
Email: zhangkai6@ufl.edu

Alireza Entezari

Computer and Information Science and Engineering
University of Florida
Gainesville, Florida
Email: entezari@ufl.edu

Abstract—Concentration inequalities play a key role in establishing the suitability of random matrices for compressed sensing as well as dimensionality reduction through random projections. In particular, the Restricted Isometry Property (RIP) for random Gaussian matrices [1], [2] and its relationship to Johnson-Lindenstrauss lemma can be established directly from concentration inequality [3]. While concentration inequality is usually established for general vectors, the distortion (in length) to sparse vectors by random matrices is usually obtained by a combinatorial argument together with union bounds. In this work, we study concentration inequalities specific to sparse vectors when projected by random matrices from compactly supported (e.g., uniform) distributions. From this approach we naturally obtain sharper bounds compared to generic concentration inequalities. These results suggest the superiority of such distributions over the Gaussian distribution for random projection of sparse vectors. Our experiments show this improvement in concentration bound for a special case of sparse binary signals and the results are further corroborated by a higher rate of recovery of general (non-binary) sparse signals from random projections.

Given a matrix $\mathbf{A} \in \mathbb{R}^{m \times N}$, where each element is i.i.d. and is drawn the Normal distribution $A_{i,j} \sim \mathcal{N}(0, \frac{1}{m})$, any vector, $\mathbf{x} \in \mathbb{R}^N$, with a unit length is projected to a vector, $\mathbf{y} \in \mathbb{R}^m$, by $\mathbf{y} = \mathbf{A}\mathbf{x}$ will satisfy the inequality:

$$\Pr(\|\mathbf{y}\|_2^2 > 1 + \delta) \leq e^{-(\frac{\delta^2}{2} - \frac{\delta^3}{3})m}, \quad (1)$$

for $\delta \in (0, 1)$. This shows that it is possible to represent any N - D signal in m - D ($m \ll N$) while preserving its energy (mostly) with high probability. For recovering a class of signals (e.g., sparse signals), usually we want δ to be as small as possible to establish this random projection a bijection for that class. This is possible when m or the number of measurements is sufficient with respect to properties of \mathbf{A} (e.g., RIP). While the majority of work on this topic Gaussian and Bernoulli distributions are considered, we show that the uniform distribution offers a tighter bound via concentration inequality specifically for sparse vectors due to its compact support.

As described above, if each entry of \mathbf{A} is i.i.d. and follows the Gaussian distribution with 0-mean and a variance of $\frac{1}{m}$, then we have:

$$y_i \sim \mathcal{N}(0, \frac{\|\mathbf{x}\|_2^2}{m}), \quad 0 \leq i \leq m$$

such that $E[\|\mathbf{y}\|_2^2] = \|\mathbf{x}\|_2^2$. Without loss of generality, we assume $\|\mathbf{x}\|_2 = 1$ for simplicity of discussion.

When the matrix \mathbf{A} is drawn from a Uniform distribution with mean 0 and variance $\frac{1}{m}$ (i.e., $A_{i,j} \sim \mathcal{U}(-\sqrt{\frac{3}{m}}, \sqrt{\frac{3}{m}})$) the mean and variance of $\|\mathbf{y}\|_2^2$ remain the same (due to linearity). However, the distribution of y_i is no longer a Gaussian, but is a compactly supported distribution (a piecewise polynomial). The key point, driving our work, is that the support of this distribution depends on the ℓ_1 norm of \mathbf{x} . As \mathbf{y} can be viewed as a linear combination of columns of \mathbf{A} weighted by elements of \mathbf{x} , y_i , as a random variable, is a linear

combination of random variables $\mathbf{A}_{i,1}, \mathbf{A}_{i,2}, \dots, \mathbf{A}_{i,N}$ weighted by the elements of \mathbf{x} . Since entries of \mathbf{A} are drawn independently the distribution of y_i is the convolution of its constituent random variables. With some abuse of notation (for brevity), one can show that:

$$p(y_i) = p(x_1 A_{i,1}) * p(x_2 A_{i,2}) * \dots * p(x_N A_{i,N}), \quad (2)$$

where $*$ is convolution operator. Since $p(x_i A_{i,j})$ is a (scaled) box function with support $[-x_i \sqrt{\frac{3}{m}}, x_i \sqrt{\frac{3}{m}}]$, $p(y_i)$ is a univariate (centered) box spline [4] described by a vector of directions which is precisely $(\sqrt{3/m})\mathbf{x}^T$. The distribution of y_i is the box spline $M_{(\sqrt{3/m})\mathbf{x}^T}(y)$ whose support is limited to $-\sqrt{3/m}\|\mathbf{x}\|_1 \leq y \leq \sqrt{3/m}\|\mathbf{x}\|_1$.

When \mathbf{x} is sparse, the support of the distribution is limited by its sparsity and hence, one can obtain tighter bounds for concentration inequality. When the elements in \mathbf{x} are binary (i.e., 0 and $1/\sqrt{k}$ for a k -sparse signal), this box spline simplifies to a uniform B-spline [5]. This distribution, in the statistics literature is also known as Irwin-Hall distribution. While we derive the concentration inequality (using the Chernoff bound) for the entire class of box splines for generic k -sparse signals, we here briefly discuss the simple case of uniform B-splines that apply to binary k -sparse signals.

Figure 1 shows the of the distribution of $\|\mathbf{y}\|_2^2$ for a k -sparse binary signal (with $k = 4$) with the uniform (solid lines) and Gaussian (dash lines) distributions with different δ and m . This experiment shows that for large m , both distributions converge to 0 exponentially (as expected); however, with the uniform distribution has a faster decay as its decay depends on the sparsity of the underlying signal.

Another experiment in 2 shows the comparison of the recovery rate of general sparse signals from different random measurements that are drawn form uniform distribution and Gaussian distribution. In this experiment, the signal dimension N is 200. We vary k from 1 to 70 and measure times m from 40 to 90, and use FISTA algorithm. For each parameter setting, we repeat 1000 times. Its obvious to notice that by using the sensing matrix that is drawn from Uniform distribution has a little higher recovery rate than from Gaussian distribution. Also, this improvement is more noticeable when m is not too big (e.g., $40 \leq m \leq 70$), which is consistent with our result from Figure 1.

Although the experiment shows a little improvement on recovery rate than Gaussian sensing matrix, due to the difficulty of deducing the close form bound for Uniform distribution on Concentration Inequality, in the future work, we will pay attention on solving this problem in another norm space and research the RIP for Uniform random matrices. Also, we will design new algorithms which can benefit from Uniform distributed random matrices.

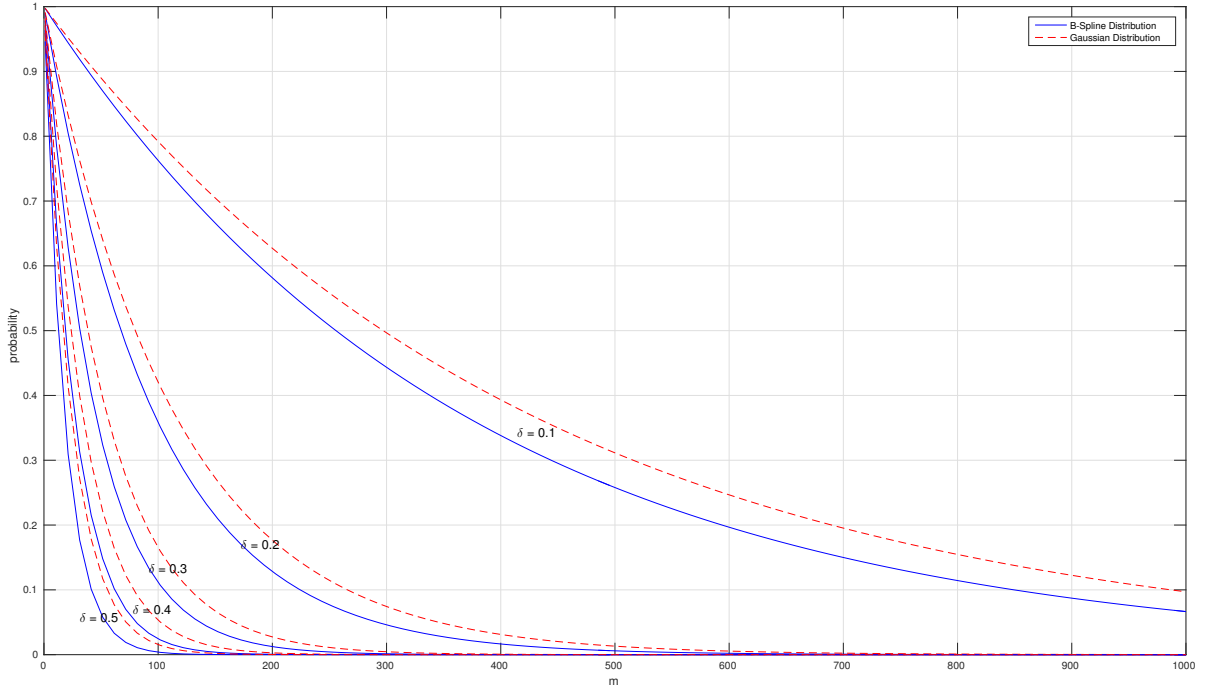


Fig. 1. the solid lines show $\Pr(\|\mathbf{y}\|_2^2 > 1 + \delta)$ when $k = 4$, δ changes from 0.1 to 0.5 and \mathbf{A} follows the Uniform distribution with mean 0, variance $\frac{1}{m}$, the dash lines show $\Pr(\|\mathbf{y}\|_2^2 > 1 + \delta)$ with any k and \mathbf{A} follows the Gaussian distribution with mean 0, variance $\frac{1}{m}$.

ACKNOWLEDGEMENT

This research was supported in part by the NSF grant IIS-1617101 and the ONR grant N00014-14-1-0762.

REFERENCES

- [1] E. J. Candes and T. Tao, "Decoding by linear programming," *IEEE transactions on information theory*, vol. 51, no. 12, pp. 4203–4215, 2005.
- [2] E. J. Candes, J. K. Romberg, and T. Tao, "Stable signal recovery from incomplete and inaccurate measurements," *Communications on pure and applied mathematics*, vol. 59, no. 8, pp. 1207–1223, 2006.
- [3] R. Baraniuk, M. Davenport, R. DeVore, and M. Wakin, "A simple proof of the restricted isometry property for random matrices," *Constructive Approximation*, vol. 28, no. 3, pp. 253–263, 2008.
- [4] C. de Boor, K. Höllig, and S. Riemenschneider, *Box splines*. Springer-Verlag, 1993, vol. 98.
- [5] C. de Boor, *A practical guide to splines*. Springer-Verlag New York, 1978, vol. 27.

Uniform distribution - Gaussian distribution

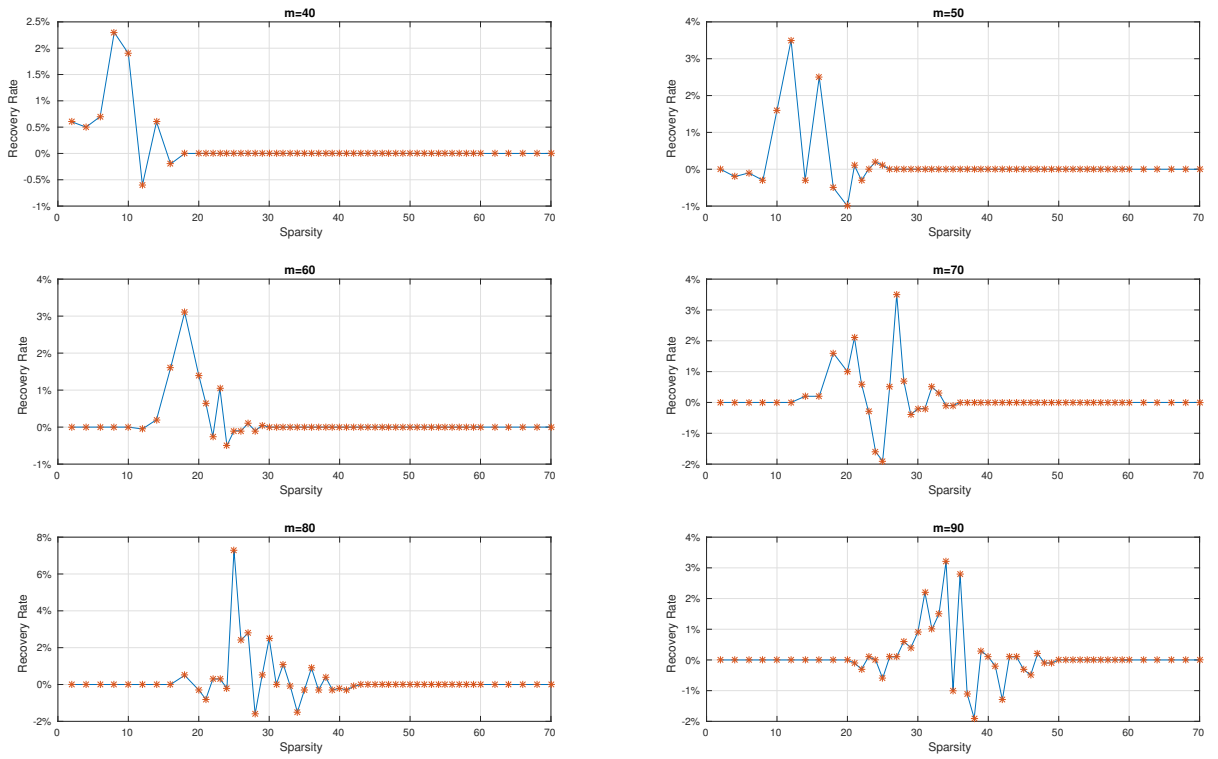


Fig. 2. shows the difference of recover rate between Uniform distributed and Gaussian distributed sensing matrices.