Generalized Restricted Isometry Property for α-Stable Random Projections

D. Otero and Gonzalo R. Arce
Department of Electrical and Computer Engineering
University of Delaware
Email:arce@udel.edu

ICASSP.
Prague, Czech Republic

May 2011

Outline

- Motivation
- α-stable random projections
- Restricted$-p$ Isometry Property (RIP$_p$)
- Computer simulations
- Conclusions

In order to show that a sparse signal can be recovered exactly from a limited number of measurements, Candès and Tao introduced the restricted isometry property (RIP)[†]:

Let $\mathbf{R}$ be an $K \times N$ matrix, where $K < N$, and s a positive integer. Then δ_s is the smallest number such that

$$(1-\delta_s)\left\|x\right\|_2^2 \leq \left\|\mathbf{R}x\right\|_2^2 \leq (1+\delta_s)\left\|x\right\|_2^2$$

for all x such that $\left\|x\right\|_0 \leq s$.

Here, δ_s is a number that measures how much $\mathbf{R}$ acts like an isometry on $s-$sparse vectors.

[†] Candès E. and Tao T., "Decoding by linear programming", IEEE Trans. Inf. Theory 51, 4203-4215, 2005.

Chartrand and Staneva proposed a variant of the RIP to show that l_p minimization may be used to recover sparse signals[†]:

Let $\mathbf{R}$ be an $K \times N$ matrix, where $K < N$, L a positive integer and $0 < p \leq 1$. Then δ_L is the smallest number such that

$$(1 - \delta_L) \|x\|_2^p \leq \|\mathbf{R}x\|_2^p \leq (1 + \delta_L) \|x\|_2^p$$

for all x such that $\|x\|_0 \leq L$.

They used this result to show that l_p minimization requires fewer measurements than l_1 minimization. In this case, the entries of $\mathbf{R}$ are Gaussian realizations.

† R. Chartrand and V. Staneva, "Restricted isometry properties and nonconvex compressed sensing", Inverse Problems, vol. 24, no. 3, June 2008.

Motivation

The latter result can be extended to **R** having $\alpha-$stable distributions if $1 \leq \alpha \leq 2$:

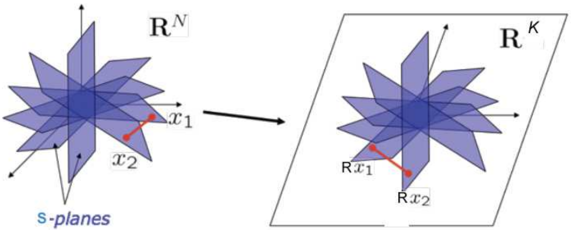
Let **R** be an $K \times N$ matrix, where $K < N$, s a positive integer and $p < \alpha/2$. Then, for a given δ

$$(1-\delta)\left\|x\right\|_\alpha^p < \left\|\mathbf{R}x\right\|_p^p < (1+\delta)\left\|x\right\|_\alpha^p$$

for all x such that $\left\|x\right\|_0 = s$.

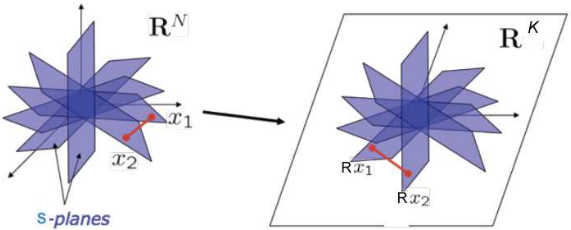
Despite that the variance of $\alpha-$stable variables is infinite, this RIP may hold with probability greater than $1 - 1/\binom{N}{s}$.

Motivation



ℓ_2 based dimensionality reduction:
$\|\mathbf{R}x_2 - \mathbf{R}x_1\|_{\ell_2}^2 \approx \|x_2 - x_1\|_{\ell_2}^2$

ℓ_1 based dimensionality reduction:
$\|\mathbf{R}x_2 - \mathbf{R}x_1\|_{\ell_p}^p \approx \|x_2 - x_1\|_{\ell_1}^p$; $p < 1/2$



ℓ_2 based dimensionality reduction:
$\|\mathbf{R}x_2 - \mathbf{R}x_1\|_{\ell_2}^2 \approx \|x_2 - x_1\|_{\ell_2}^2$

ℓ_1 based dimensionality reduction:
$\|\mathbf{R}x_2 - \mathbf{R}x_1\|_{\ell_p}^p \approx \|x_2 - x_1\|_{\ell_1}^p$; $p < 1/2$

ℓ_2 RIP:
$(1-\delta)\|x\|_{\ell_2}^2 \le \|\mathbf{R}x\|_{\ell_2}^2 \le (1+\delta)\|x\|_{\ell_2}^2$
$\mathbf{R} \sim$ Gaussian or subgaussian

ℓ_p RIP:
$(1-\delta)\|x\|_{\ell_1}^p \le \|\mathbf{R}x\|_{\ell_p}^p \le (1+\delta)\|x\|_{\ell_1}^p$
$\mathbf{R} \sim \alpha$-stable

Why?: ℓ_1 norm is more robust and sparse inducing!

† Adapted from Lecture Notes - R. Baraniuk

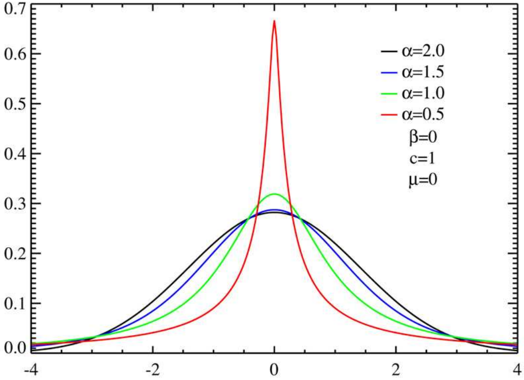
¿Why?

- α-stable distributions preserve the l_α norm. This one is more robust to noise, missing data, and outliers, than the l_2 norm.
- l_p minimization may be used using $\alpha-$stable random projections. Furthermore, l_p minimization requires fewer measurements as p tends to zero.
- $\alpha-$stable random projections are more robust to Gaussian noise and impulsive noise† than random projections with finite variance.

† If the impulsive noise has parameter α', and $\alpha < \alpha'$, this is true.

The characteristic function for symmetric $\alpha-$stable distributions ($S\alpha S$), with location parameter zero and dispersion γ

$$\varphi(t) = e^{-|\gamma t|^{\alpha}}$$



Let $r_i \sim S\alpha S(0, \alpha, \gamma_i)$ and y be the linear combination of n i.i.d. $S\alpha S$ random variables: $y = r_1 x_1 + r_2 x_2 + \cdots + r_n x_n$.
Stable law dictates that

$$y \sim S\alpha S(0, \alpha, \gamma)$$

$$\gamma = (|x_1\gamma_1|^\alpha + |x_2\gamma_2|^\alpha + \cdots + |x_n\gamma_n|^\alpha)^{1/\alpha}.$$

When the $r_i \sim S\alpha S(0, \alpha, 1)$, the dispersion of y is the l_α norm of the scalars of the linear combination.

$$y \sim S\alpha S(0, \alpha, \gamma)$$

$$\gamma = (\sum_{i=1}^{n} |x_i|^\alpha)^{\frac{1}{\alpha}} = \|x\|_\alpha$$

Thus, $\alpha-$stable distributions preserve the l_α norm.

$\alpha-$stable Random Projections. Cauchy random projections ($\alpha = 1$)

Cauchy random projections arise as a solution to ℓ_1 distance preservation in Dimensionality reduction applications.

Let $r_i \sim S\alpha S(0,1,1) \triangleq C(0,1)$ then

$$y = \sum_{i=1}^{n} r_i x_i \sim C\left(0, \sum_{i=1}^{n} |x_i|\right)$$

For two vectors: $y_1 = \sum_{i=1}^{n} r_{1,i} x_{1,i}$ and $y_2 = \sum_{i=1}^{n} r_{2,i} x_{2,i}$

$$u_j = y_1 - y_2 \sim C\left(0, \sum_{i=1}^{n} |x_{1,i} - x_{2,i}|\right) = C(0, \|x_1 - x_2\|)$$

In Cauchy random projections, ℓ_1 distance estimation reduces to estimate the scale parameter γ from k i.i.d. samples $u_j \sim C(0,\gamma)$

$S\alpha S$ do not have finite variance, but its dispersion may be estimated using the fractional moment of order p. If $p < \alpha$, this moment is finite and it is given by:

$$E(|x|^p) = \frac{2^{p+1}\Gamma(-p/\alpha)}{\alpha\sqrt{\pi}\Gamma(-p/2)}\Gamma\left(\frac{p+1}{2}\right)\gamma^p.$$

From the latter expression, the following non-linear estimator follows

$$E\left(\|X\|_p^p\right) = E\left(\frac{1}{k}\sum_{i=1}^{k}|x_i|^p\right) = k\frac{2^{p+1}\Gamma(-p/\alpha)}{\alpha\sqrt{\pi}\Gamma(-p/2)}\Gamma\left(\frac{p+1}{2}\right)\gamma^p$$

The l_p^p norm of a vector –which components are $\alpha-$stable realizations– is an estimate of the dispersion of its entries.

RESTRICTED$-p$ ISOMETRY PROPERTY (RIP$_p$)

Theorem. Let $\mathbf{R}$ be an $K \times N$ matrix whose elements are *i.i.d.* alpha-stable realizations with $1 \leq \alpha \leq 2$, which is multiplied by a scalar which depends only on K, α and p. Let $\|x\|_0 = s$, if $K \geq \; C_1 s \log(N/s) + C_2 s + C_3$ and $0 < p < \alpha/2$, then, for a given δ

$$(1-\delta)\,\|x\|_\alpha^p < \|\mathbf{R}x\|_p^p < (1+\delta)\,\|x\|_\alpha^p$$

holds with probability greater than $1 - 1/\binom{N}{s}$.

The proof of the theorem uses probabilistic methods. Rather than focusing on the values of the RIC, we choose a δ such that the RIP holds with high probability.

The proof will be divided in three lemmas and a proof that gathers the results obtained in the lemmas:

- **Lemma 1.** Probability that the RIP holds for a fixed x such that $x \in \mathbb{R}^s$ and $\|x\|_0 = s$.
- **Lemma 2.** Probability that the RIP fails for any $x \in \mathbb{R}^s$ having the same $\|x\|_\alpha^p$ such that $\|x\|_0 = s$.
- **Lemma 3.** Probability that the RIP fails for any submatrix of the matrix **R**.

Lemma 1. Let T be a set of indices with $|T| = s$ $(1 \leq s \leq N)$, denote by X_T the set of discrete signals in $\mathbb{R}^s$ that are zero outside of T. Let Ψ be an $K \times s$ submatrix of Φ formed with the columns whose indices are in T. If $0 < p < \alpha/2$ and for a $x \in X_T$, then, for a given $\eta > 0$

$$(1-\eta)KC_\mu(\alpha,p)\,\|x\|_\alpha^p < \|\Psi x\|_p^p < (1+\eta)KC_\mu(\alpha,p)\,\|x\|_\alpha^p$$

holds with probability lower-bounded by $1 - 2e^{-\frac{K\eta^2}{2(1+\eta)^2}C(\alpha,p)}$.

Proof : Let $y = \Psi x$, where x is a vector that belongs to X_T. The resultant vector y has K components, each one given by

$$y_i = \sum_{i=1}^{s} \psi_{ij} x_j \sim S\alpha S(0, \|x\|_\alpha).$$

The l_p^p norm of y is equal to

$$\|y\|_p^p = \|\Psi x\|_p^p = \sum_{i=1}^{K} \left| \sum_{j=1}^{k} \psi_{ij} x_j \right|^p = \sum_{i=1}^{K} |y_i|^p.$$

This implies that in average $\|\Psi x\|_p^p$ tends to

$$\begin{aligned} E(\|\Psi x\|_p^p) &= K \frac{2^{p+1}\Gamma(-p/\alpha)}{\alpha\sqrt{\pi}\Gamma(-p/2)} \Gamma\left(\frac{p+1}{2}\right) \|x\|_\alpha^p \\ &= K C_\mu(\alpha, p) \|x\|_\alpha^p. \end{aligned}$$

for $0 < p < \alpha$.

If $0 < p < \alpha/2$, the variance of the estimator $\|\Psi x\|_p^p$ is finite as well

$$
\begin{aligned}
Var(\|\Psi x\|_p^p) &= Var\left(\sum_{i=1}^{K} |y_i|^p\right) \\
&= KC_\mu(\alpha, 2p)\|x\|_\alpha^{2p} - KC_\mu(\alpha, p)^2\|x\|_\alpha^{2p} \\
&= KC_{\sigma^2}(\alpha, p)\|x\|_\alpha^{2p}.
\end{aligned}
$$

The distribution of $\|\Psi x\|_p^p$ may be approximated with a Gaussian, however, a much better approximation may be done using a distribution with the same support of $\|\Psi x\|_p^p$: the Inverse Gaussian.

Restricted−p Isometry Property (RIP$_p$)

If $\mathbf{X}$ follows an inverse Gaussian distribution, it can be shown that $\mathbf{X}$ has the following Chernoff bound:

$$P(|\mathbf{X} - \mu| < \eta\mu) \geq 1 - 2e^{-\frac{\eta^2}{2(1+\eta)}\left(\frac{\mu^2}{\sigma^2}\right)}.$$

It follows that

$$P(|\,\|\Psi x\|_p^p - \mu| < \eta\mu) \gtrapprox 1 - 2e^{-\frac{K\eta^2}{2(1+\eta)^2}C(\alpha,p)} = 1 - P_{\alpha,p,K}(\eta).$$

Where $C(\alpha, p) = C_\mu(\alpha, p)^2 / C_{\sigma^2}(\alpha, p)$. If η is chosen properly, the RIP holds with high probability for a fixed x.

Restricted−p Isometry Property (RIP$_p$)

Lemma 2. Let $0 < p < \alpha/2$, Ψ a $K \times s$ submatrix of Φ as in *Lemma* 1. Let $\delta > 0$ and choose $\epsilon > 0$ for a given η such that $\frac{\eta+\epsilon^p}{1-\epsilon^p} \leq \delta$. Then for any $x \in \mathbb{R}^s$

$$(1-\delta)KC_\mu(\alpha,p)\,\|x\|_\alpha^p < \|\Psi x\|_p^p < (1+\delta)KC_\mu(\alpha,p)\,\|x\|_\alpha^p \qquad (1)$$

holds with probability $\gtrapprox 1-(1+2/\epsilon)^s P_{\alpha,p,K}(\eta)$.

Lemma 3. Let Φ be an $K \times N$ matrix, $x \in \mathbb{R}^N$ and $\|x\|_0 = s$. Let $0 < p < \alpha/2$ and $\delta > 0$, then

$$(1-\delta)KC_\mu(\alpha,p)\,\|x\|_\alpha^p < \|\Phi x\|_p^p < (1+\delta)KC_\mu(\alpha,p)\,\|x\|_\alpha^p$$

with probability $\gtrapprox 1-(eN/s)^s(1+2/\epsilon)^s P_{\alpha,p,K}(\eta)$.

Proof of the Theorem. According to *Lemma* 3, we have

$$(1-\delta)KC_\mu(\alpha,p)\left\|x\right\|_\alpha^p < \left\|\Phi x\right\|_p^p < (1+\delta)KC_\mu(\alpha,p)\left\|x\right\|_\alpha^p$$
$$(1-\delta)\left\|x\right\|_\alpha^p < \left\|(KC_\mu(\alpha,p))^{-1/p}\Phi x\right\|_p^p < (1+\delta)\left\|x\right\|_\alpha^p .$$

Letting $\mathbf{R}$ equal to $(KC_\mu(\alpha,p))^{-1/p}\Phi$, we obtain

$$(1-\delta)\left\|x\right\|_\alpha^p < \left\|\mathbf{R}x\right\|_p^p < (1+\delta)\left\|x\right\|_\alpha^p .$$

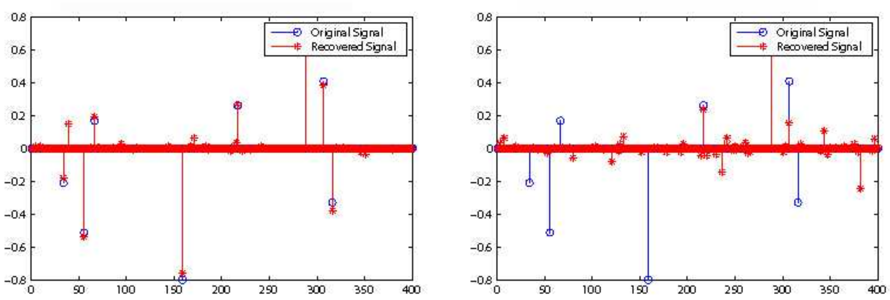
Then, the probability that the RIP fails is approximately less or equal than $\left(\frac{eN}{s}\right)^s\left(1+\frac{2}{\epsilon}\right)^s P_{\alpha,p,K}(\eta)$.

With a number of projections given by: $K \gtrapprox C_1 s\log(N/s) + C_2 s + C_3$

Sparse Signal Reconst. from Cauchy Random Proj. Noisy Case (α=1.2)

- Model: $y = Rx + \eta$; $f_\eta(x) = \frac{1}{2\pi}\int_{-\infty}^{\infty}\varphi(t;0,\alpha,\gamma)\exp^{-ixt}\,dt$.
- $n = 400$, $k = 100$, $s = 8$.
- $\ell_0 - LL$: ℓ_0-regularized coordinate-descent Myriad-based regression.
- $\ell_1 - \ell_s$: ℓ_1-regularized least squares.

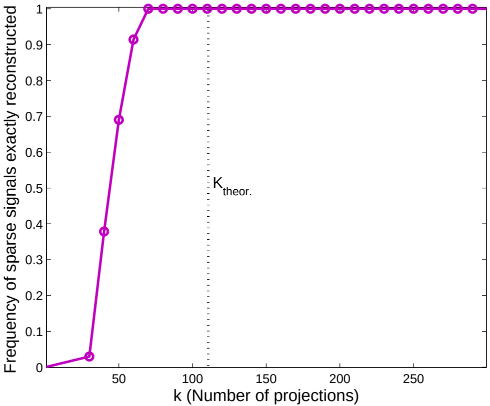
$\ell_0 - LL$ $\qquad\qquad$ $\ell_1 - \ell_s$



$\gamma = 0.5$, MSE=-40.1dB $\qquad\qquad$ $\gamma = 0.5$. MSE=-24.9dB

Number of required measurements. Noiseless case.

- Model: $y = Rx$.
- $n = 300$, $s = 6$.
- Reconstruction method: ℓ_0-regularized coordinate-descent Myriad-based regression.



Conclusions

- A new RIP for $S\alpha S$ random matrices with $1 \leq \alpha \leq 2$ has been formulated.
- Useful applications such as methods of robust statistics, reconstruction algorithms, among others.
- As $p \to 0$, less measurements are required for the RIP to hold with high probability.
- The restricted isometry constants (RIC) for this new RIP remain as an open problem.