New Constructions of RIP Matrices with Fast Multiplication and Fewer Rows

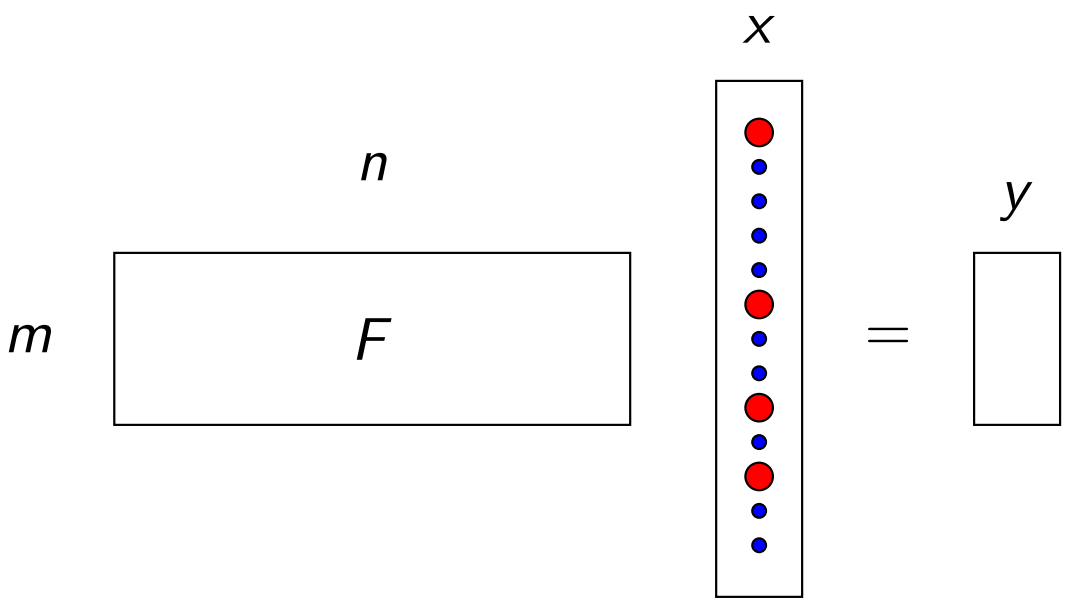
aka, **sparse** recovery from **Fourier**-like measurements with applications to **fast** Johnson-Lindenstrauss **transforms**, **etc.**

Jelani Nelson, Eric Price, and **Mary Wootters**

February 18, 2013

Compressed Sensing

Given: A few linear measurements of an (approximately) k-sparse vector $x \in \mathbb{R}^n$.
Goal: Recover x (approximately).



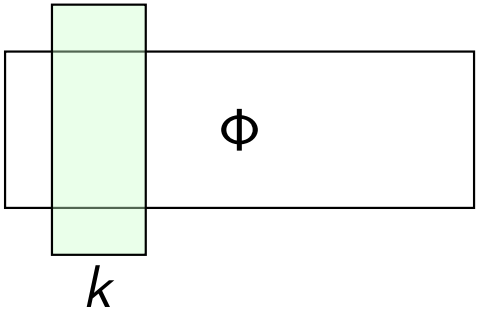
Algorithms for compressed sensing

- A lot of people use linear programming.
- Also Iterative Hard Thresholding, CoSaMP, OMP, StOMP, ROMP....
- For all of these:
  - the time it takes to multiply by Φ or Φ^* is the bottleneck.
  - the *Restricted Isometry Property* is a sufficient condition.

Restricted Isometry Property (RIP)

$$(1-\varepsilon)\|x\|_2^2 \leq \|\Phi x\|_2^2 \leq (1+\varepsilon)\|x\|_2^2$$

for all k-sparse $x \in \mathbb{R}^n$.



All of these submatrices are well conditioned.

Goal

Matrices Φ which have the RIP and support fast multiplication.

An open question

If the rows of Φ are random rows from a Fourier matrix, how many measurements do you need to ensure that Φ has the RIP?

- $m = O(k \log(n) \log^3(k))$ [CT06, RV08, CGV13].

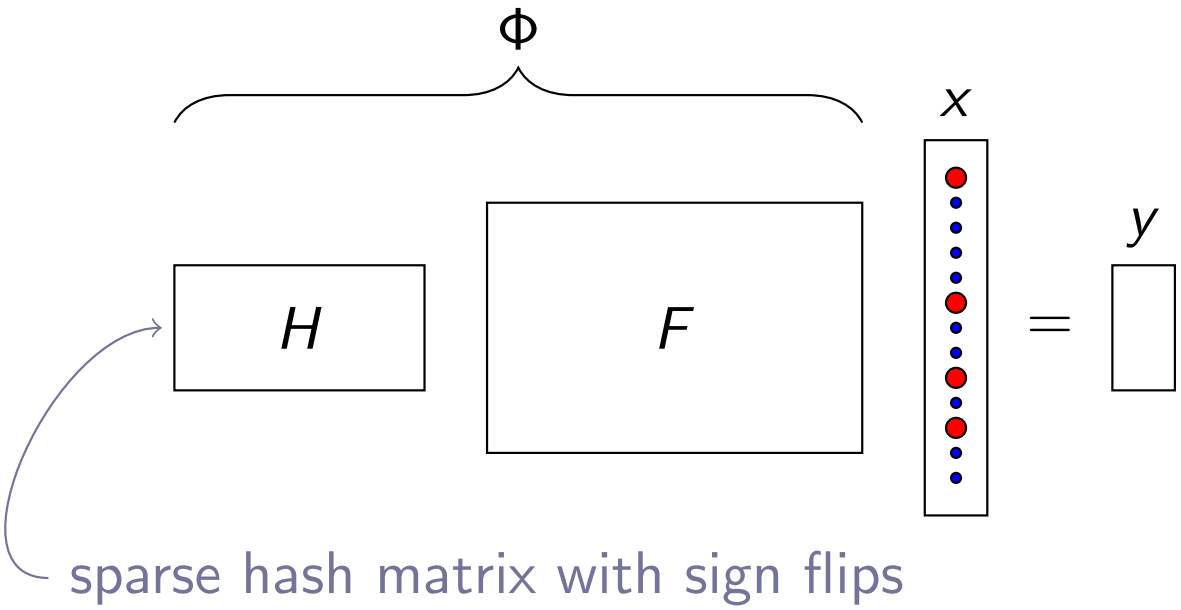
Ideal:

- $m = O(k \log(n/k))$.

(Related: how about partial circulant matrices?)

- $m = O(k \log^2(n) \log^2(k))$ [RRT12, KMR13].

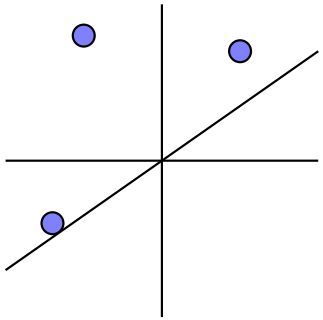
In this work



- Can still multiply by Φ quickly.
- Our result: has the RIP with

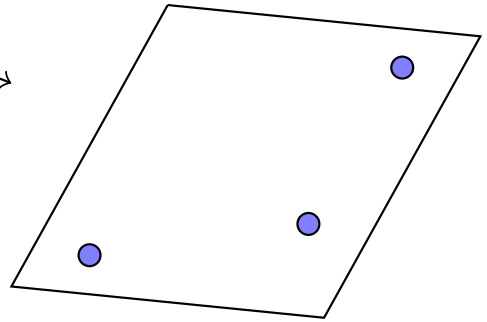
$$m = O(k \log(n) \log^2(k)).$$

Another motivation:
Johnson Lindenstrauss (JL) Transforms



High dimensional data
$S \subset \mathbb{R}^n$

Linear map, Φ



Φ preserves
the geometry of S

Low dimensional sketch
$\Phi(S) \in \mathbb{R}^m$

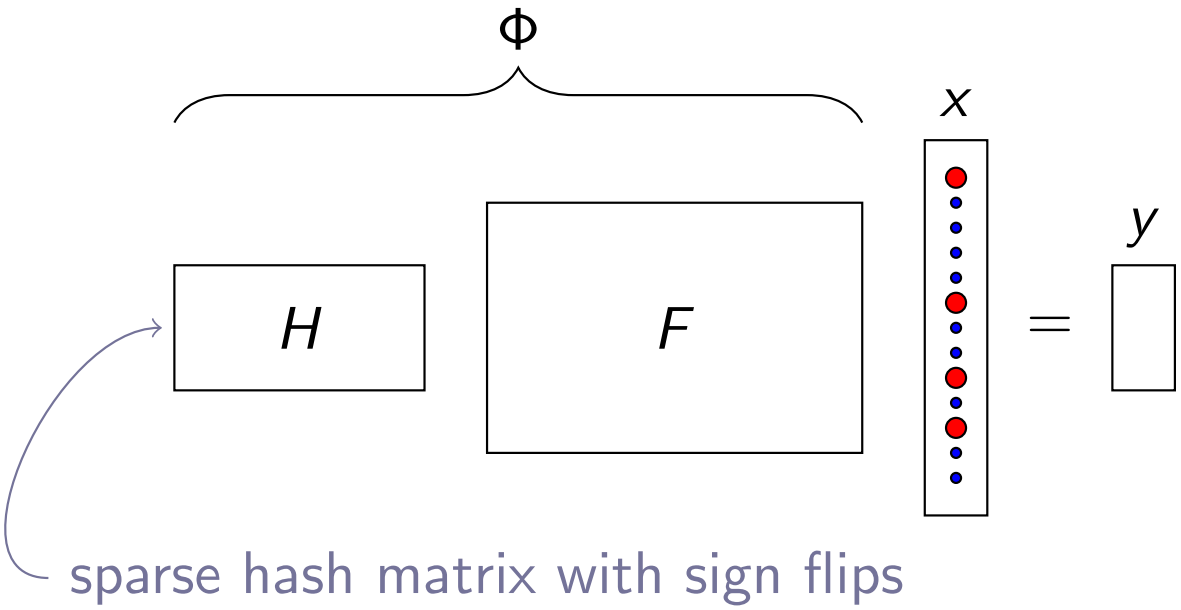
What do we want in a JL matrix?

- Target dimension should be small (like $\log(|S|)$).
- Fast multiplication.
  - Approximate numerical algebra problems (e.g., linear regression, low-rank approximation)
  - k-means clustering

How do we get a JL matrix?

- Gaussians will do.
- Best way known for *fast JL*: By [KW11], RIP $\Rightarrow$ JL.*
- So our result also gives fast JL transforms with the fewest rows known.

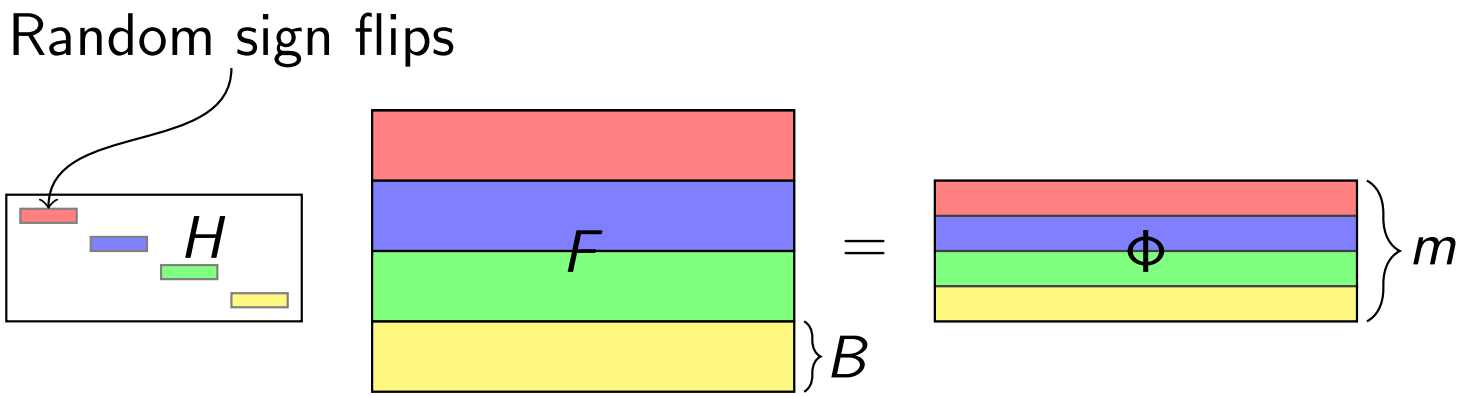
Our results



- Can still multiply by Φ quickly.
- Our result: has the RIP with

$$m = O(k \log(n) \log^2(k)).$$

More precisely



- If A has mB rows, then Φ has m rows.
- The "buckets" of H have size B.

Theorem

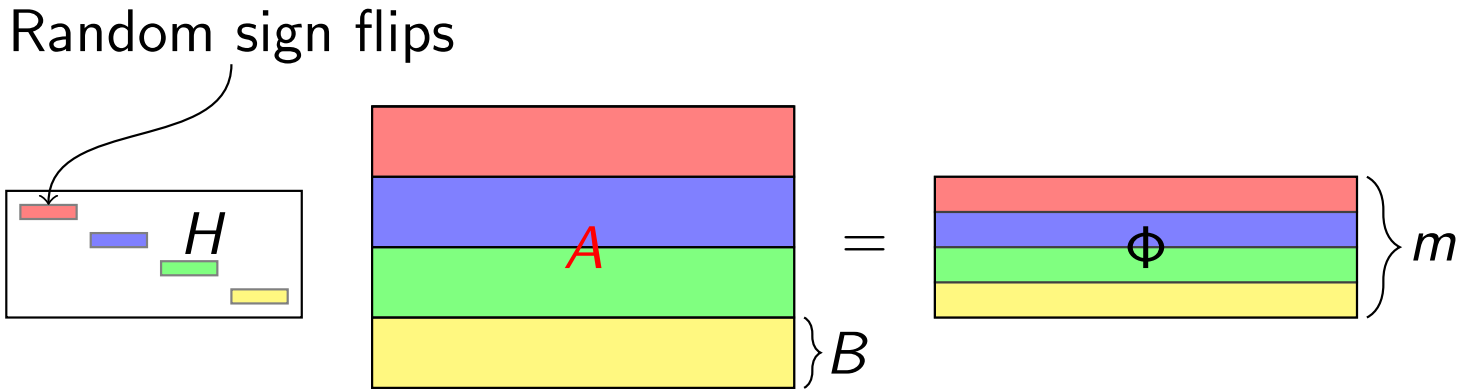
If $B \simeq \log^{2.5}(n)$, $m \simeq k \log(n) \log^2(k)$, and F is a random partial Fourier matrix, then Φ has the RIP with probability at least $2/3$.

Previous results

Construction	Measurements m	Multiplication Time	Notes
[AL09, AR13]	$\frac{k\log(n)}{\varepsilon^2}$	$n\log(n)$	as long as $k \leq n^{1/2-\delta}$
Sparse JL matrices [KN12]	$\frac{k\log(n)}{\varepsilon^2}$	εmn	
Partial Fourier [RV08, CGV13]	$\frac{k\log(n)\log^3(k)}{\varepsilon^2}$	$n\log(n)$	
Partial Circulant [KMR13]	$\frac{k\log^2(n)\log^2(k)}{\varepsilon^2}$	$n\log(m)$	
Hash / partial Fourier [NPW12]	$\frac{k\log(n)\log^2(k)}{\varepsilon^2}$	$n\log(n)$ + $m\text{polylog}(n)$	
Hash / partial circulant [NPW12]	$\frac{k\log(n)\log^2(k)}{\varepsilon^2}$	$n\log(m)$ + $m\text{polylog}(n)$	

Approach

Our approach is actually more general:



General result

If A is a "decent" RIP matrix:

- A has too many (mB) rows, but *does* have the RIP (whp).
- RIP-ness degrades gracefully as number of rows decreases.

Then Φ is a better RIP matrix:

- Φ has the RIP (whp) with fewer (m) rows.
- Time to multiply by Φ = time to multiply by $A + mB$.

Proof overview

We want

$$\mathbb{E} \sup_{x \in \Sigma_k} \left| \|\Phi x\|_2^2 - \|x\|_2^2 \right| < \varepsilon,$$

where Σ_k is unit-norm k-sparse vectors.

Proof overview I: triangle inequality

$$\begin{aligned}
&\mathbb{E} \sup_{x \in \Sigma_k} \left| \|\Phi x\|_2^2 - \|x\|^2 \right| \\
&\quad \leq \mathbb{E} \sup_{x \in \Sigma_k} \left| \|\Phi x\|_2^2 - \|Ax\|_2^2 \right| + \mathbb{E} \sup_{x \in \Sigma_k} \left| \|Ax\|_2^2 - \|x\|_2^2 \right| \\
&\quad \ldots \\
&\quad \leq \mathbb{E} \sup_{x \in \Sigma_k} \left| \|X_x \xi\|_2^2 - \mathbb{E}_\xi \|X_x \xi\|_2^2 \right| + (\text{RIP constant of } A),
\end{aligned}$$

where X_x is some matrix depending x and A, and ξ is the vector of random sign flips used in H.

Proof overview I: triangle inequality

$$\mathbb{E}\sup_{x\in\Sigma_k}\left|\|X_x\xi\|_2^2-\mathbb{E}_\xi\|X_A(x)\xi\|_2^2\right|+(\text{RIP constant of } A)$$

By assumption, this is small.
(Recall A has too many rows)

This is a *Rademacher Chaos Process.*
We have to do some work to show that it is small.

Proof overview II: probability and geometry

By [KMR13], it suffices to bound

$$\gamma_2(\Sigma_k, \|\cdot\|_A)$$

Some norm induced by A

Captures how "clustered" Σ_k is with respect to $\|\cdot\|_A$

We estimate this by bounding the *covering number* of Σ_k with respect to $\|\cdot\|_A$. $\square$

Open Questions

(1) How many random fourier measurements do you need for the RIP?

(2) Can you remove the other two log factors from our construction?

- It seems like doing this would remove two log factors from (1) as well.

(3) Can you come up with any ensemble of RIP matrices with $k \log(N/k)$ rows and fast multiplication?

(4) Can you come up with any ensemble JL matrices with $\log(|S|)$ rows supporting fast multiplication?

Thanks!

Nir Ailon and Edo Liberty.
Fast dimension reduction using Rademacher series on dual BCH codes.
Discrete Comput. Geom., 42(4):615–630, 2009.

N. Ailon and H. Rauhut.
Fast and RIP-optimal transforms.
Preprint, 2013.

Mahdi Cheraghchi, Venkatesan Guruswami, and Ameya Velingker.
Restricted isometry of Fourier matrices and list decodability of random linear codes.
In *Proceedings of the 24th Annual ACM-SIAM Symposium on Discrete Algorithms (SODA)*, pages 432–442, 2013.

Emmanuel J. Candès and Terence Tao.
Near-optimal signal recovery from random projections: universal encoding strategies?
IEEE Trans. Inform. Theory, 52:5406–5425, 2006.

F. Krahmer, S. Mendelson, and H. Rauhut.

Suprema of chaos processes and the restricted isometry property.
Comm. Pure Appl. Math., 2013.

Daniel M. Kane and Jelani Nelson.
Sparser Johnson-Lindenstrauss transforms.
In *Proceedings of the Twenty-Third Annual ACM-SIAM Symposium on Discrete Algorithms*, pages 1195–1206. SIAM, 2012.

Felix Krahmer and Rachel Ward.
New and improved Johnson-Lindenstrauss embeddings via the Restricted Isometry Property.
SIAM J. Math. Anal., 43(3):1269–1281, 2011.

Jelani Nelson, Eric Price, and Mary Wootters.
New constructions of rip matrices with fast multiplication and fewer rows.
arXiv preprint arXiv:1211.0986, 2012.

Holger Rauhut, Justin Romberg, and Joel A. Tropp.
Restricted isometries for partial random circulant matrices.
Appl. and Comput. Harmon. Anal., 32(2):242–254, 2012.

Mark Rudelson and Roman Vershynin.
On sparse reconstruction from Fourier and Gaussian measurements.
Communications on Pure and Applied Mathematics, 61(8):1025–1045, 2008.