Constructing Explicit RIP Matrices and the Square-Root Bottleneck

Ryan Cinoman

July 18, 2018

Norbert Wiener Center for Harmonic Analysis and Applications

Outline

Norbert Wiener Center
for Harmonic Analysis and Applications

Compressed Sensing

- The motivating problem: finding sparse solutions of underdetermined equations.
- The foundational result: yes, it is possible, and we can do so with linear programming!
- The catch: there are limitations on the sensing matrix that make the theory difficult to apply in practice.

Compressed Sensing

The Motivating Problem

- Let $\mathbf{x} \in \mathbb{R}^N$ be k-sparse. A measurement of $\mathbf{x}$ is $\mathbf{y} = A\mathbf{x}$ for an $m \times N$ matrix A. A is called the measurement matrix.
- We want to find a way to reconstruct $\mathbf{x}$ from the both the measurement $\mathbf{y}$, and the knowledge of its sparsity.
- We can frame this as a constrained optimization problem:

$$\mathbf{x}^{\#} = \mathrm{argmin}\|\mathbf{z}\|_{\ell_0} \quad \text{s.t. } A\mathbf{z} = \mathbf{y}. \tag{1}$$

Norbert Wiener Center
for Harmonic Analysis and Applications

Compressed Sensing

The Motivating Problem

- Problem (1) is computationally unrealistic, so we consider the convex relaxation of the problem, which has the convenient realization:

$$\mathbf{x}^{\#} = \mathrm{argmin}\|\mathbf{z}\|_{\ell_1} \quad \text{s.t. } A\mathbf{z} = \mathbf{y}. \tag{2}$$

Norbert Wiener Center for Harmonic Analysis and Applications

Compressed Sensing

The Foundational Result

- In 2004, Candes, Romberg and Tao published a series of papers on the problem (2) and its relationship to the motivating problem.
- The main result: not only does problem (2) have a unique solution, but it is guaranteed to recover $\mathbf{x}$ exactly as long as A is a satisfactory sensing matrix
- A randomly generated Gaussian matrix is satisfactory with high probability, provided it satisfies $m \gtrsim k \ln(eN/k)$.

- The catch: how do we know if a matrix is suitable for compressive sensing?
- To characterize matrices for compressive sensing is a big topic. I will focus on two properties which are most popular: NSP and RIP.

Null Space Property

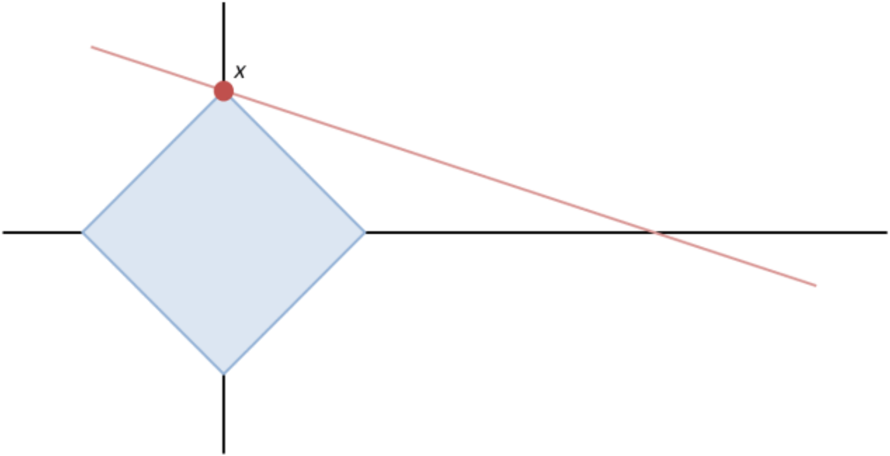
Definition

(Null Space Property) An $m \times N$ matrix A is said to have the NSP of order k if for any $\boldsymbol{\nu} \in \ker A \setminus \{0\}$, $S \subset \{1, \ldots, N\}$ with $|S| \leq k$, $\|\boldsymbol{\nu}_S\|_{\ell_1} < \|\boldsymbol{\nu}_{S^c}\|_{\ell_1}$.

- A matrix has NSP iff equation (2) recovers all k-sparse vectors

Norbert Wiener Center for Harmonic Analysis and Applications

Null Space Property



[0] https://dustingmixon.wordpress.com/

Null Space Property

- The Null Space Property (NSP) provides an exact characterization of the matrices which can recover k-sparse vectors.
- Difficult to work with in practice
- Not robust

Norbert Wiener Center
for Harmonic Analysis and Applications

Restricted Isometry Property

Definition

(Restricted Isometry Property) An $m \times N$ matrix is said to have the RIP of order k with constant $\delta \in (0,1)$ if for any k-sparse $\mathbf{x} \in \mathbb{R}^N$,

$$(1-\delta)\|\mathbf{x}\|_{\ell_2}^2 < \|A\mathbf{x}\|_{\ell_2}^2 < (1+\delta)\|\mathbf{x}\|_{\ell_2}^2 \,.$$

We say that δ_k is the restricted isometry constant of A if δ_k is the smallest $\delta > 0$ such that A satisfies RIP of order k.

- We say that δ_k is the restricted isometry constant of A if δ_k is the smallest $\delta > 0$ such that A satisfies RIP of order k.
- RIP is strictly stronger than NSP, but in return for the added restriction, we do get a robustness result.

Norbert Wiener Center
for Harmonic Analysis and Applications

Restricted Isometry Property

- Introduce an error term to the measurement: $\mathbf{y} = A\mathbf{x} + \mathbf{e}$.
- Likewise relax the constraint in (2) to:

$$\mathbf{x}^{\#} = \operatorname{argmin} \|\mathbf{z}\|_{\ell_1} \quad \text{s.t. } \|A\mathbf{z} - \mathbf{y}\|_{\ell_2} \leq \eta. \tag{3}$$

Theorem (Cai, Zhang)

If A has $\delta_{2k} < 1/\sqrt{2}$, then the solution $\mathbf{x}^{\#}$ to problem (3) satisfies

$$\left\|\mathbf{x} - \mathbf{x}^{\#}\right\|_{\ell_2} \leq \frac{C}{\sqrt{k}} \|\mathbf{x} - \mathbf{x}_k\|_{\ell_1} + D\|\mathbf{e}\|_{\ell_2}, \tag{4}$$

for some C and D depending only on δ_{2k}.

Random Gaussian Matrices

Norbert Wiener Center
for Harmonic Analysis and Applications

Random Gaussian Matrices

- The final foundational result of Candes, Romberg and Tao's original publications was that RIP matrices are in fact plentiful

Theorem

Let A be an $m \times N$ random matrix wherein each element is an independent Gaussian variable with mean zero and variance $1/m$. If $m \geq C_1 \delta^{-2} k \ln(eN/k)$ then with probability at least $1 - 2\exp(-C_2 \delta^2 m)$, $\delta_k < \delta$.

How many measurements must we take?

- Theorem (4) guarantees RIP matrices with $m \lesssim \delta^{-2} k \ln(eN/k)$.
- Fix $R = N/m$ and δ. Then the relationship becomes

$$m \lesssim k \ln(m/k).$$

- This in turn implies the existence of families of RIP matrices for $k = \Omega(m)$.
- However this is where the theory falters: explicit constructions of RIP matrices only manage $k = \Omega(m^{1/2})$!
- This discrepancy is the square-root bottleneck referenced in the title of this presentation.

Norbert Wiener Center for Harmonic Analysis and Applications

The Square-Root Bottleneck

- The explicit RIP problem is defined by Mixon as follows:

Definition

To solve ExRIP[z] is to find an explicit family of matrices with arbitrarily large aspect ratio N/m, such that each matrix satisfies RIP with constant δ of order k, where $k = \Omega(m^{z-\varepsilon})$ for all $\varepsilon > 0$ and $\delta < 1/3$.

- For many years after Candes, Romberg and Tao's foundational work, the best known result was ExRIP[1/2], thus the bottleneck.
- In 2011, Bourgain, et. al., managed to beat the above result by a small amount. Their result stands today as the best effort at solving the explicit RIP problem.

Norbert Wiener Center
for Harmonic Analysis and Applications

How many measurements must we take?

To gain insight into the difficulties for this problem, let's compare some existing methods of constructing RIP matrices.

Random Construction

- The following proof is paraphrased from work by Foucart [10].
- Outline of proof:
  1. A concentration inequality which quantifies the amount of vectors on which the sensing matrix A is not a near-isometry.
  2. A combinatorial argument to quantify the approximate number of degrees of freedom of the set of sparse unit vectors.
  3. Combine the above estimates to bound the probability that A is a near-isometry on the set of k-sparse vectors.

Norbert Wiener Center for Harmonic Analysis and Applications

Random Construction

The Concentration Inequality

- Let A be a random $m \times N$ matrix where each element is a Gaussian i.i.d. random variable with mean 0 and variance $1/m$.
- For a fixed $\mathbf{x}$, we have

$$(A\mathbf{x})_i = \sum_j A_{i,j} x_j = \frac{\|\mathbf{x}\|_2}{\sqrt{m}} g_i.$$

- Using this observation, we can find the likelihood that the energy of A is concentrated near $\mathbf{x}$.

Lemma

$$\mathbb{P}(\|A\mathbf{x}\|_2 - \|\mathbf{x}\|_2 > t\|\mathbf{x}\|_2) \leq 2\exp\left(-\frac{mt^2}{16}\right).$$

Random Construction

The Combinatorial Argument

- Consider an index set $S \subset \{1, \ldots, N\}$ of size k. We consider $\mathbb{R}^k$ to be the subset of $\mathbb{R}^N$ of vectors supported on S.

Lemma

The unit sphere in $\mathbb{R}^k$ can be covered by $n \leq (1 + 2/\rho)^k$ balls of radius ρ, with centers $\{\mathbf{u}_i\}$ on the sphere.

- We can combine the above lemma with the concentration inequality from before to apply it to the RIP.

Norbert Wiener Center for Harmonic Analysis and Applications

Random Construction

The Combinatorial Argument

- With the substitution $B = A_S^* A_S - I$, the concentration inequality reads:

$$\mathbb{P}(|\langle B\mathbf{x}, \mathbf{x}\rangle| > t) \leq 2\exp\left(-\frac{mt^2}{16}\right).$$

- Considering just the $\mathbf{u}_i$'s, calculate

$$\mathbb{P}\left(|\langle B\mathbf{u}_i, \mathbf{u}_i\rangle| > t \text{ for some } i\right) \leq 2\left(1 + \frac{2}{\rho}\right)^k \exp\left(-\frac{mt^2}{16}\right).$$

Norbert Wiener Center for Harmonic Analysis and Applications

Random Construction

The Combinatorial Argument

- If we assume that indeed $|\langle B\mathbf{u}_i, \mathbf{u}_i\rangle| \leq t$ for all i, then we can use the fact that any unit-norm vector supported on S is at most a distance ρ away from some $\mathbf{u}_i$ to put a bound on the operator norm of B,

$$\|B\| \leq \frac{t}{1-2\rho} = \delta,$$

for a choice $\rho = \frac{1}{4}$, $t = \frac{\delta}{2}$. Thus, we have an upper bound on the probability that A is not a near-isometry for any $\mathbf{x}$ supported on S!

Norbert Wiener Center for Harmonic Analysis and Applications

Random Construction

The Combinatorial Argument

- Lastly, generalize to any k-sparse vector by taking the union over all sets S.

$$\begin{aligned}\mathbb{P}(\delta_k > \delta) &\leq \binom{N}{k} 2\exp\left(\ln(9)k - \frac{m\delta^2}{64}\right) \\ &\leq 2\exp\left(k\ln(9e)\ln\left(\frac{eN}{k}\right) - \frac{m\delta^2}{64}\right).\end{aligned}$$

- So as long as $m \geq k\delta^{-2}\ln(9e)\ln(eN/k)/128$,

$$\mathbb{P}(\delta_k > \delta) \leq 2\exp\left(-\frac{m\delta^2}{128}\right)$$

Norbert Wiener Center
for Harmonic Analysis and Applications

Random Construction

Why doesn't this proof give us insight into explicit constructions?

- The proof hinges on a combinatorial argument: The number of vectors which are near to the null space of A and the degrees of freedom of the RIP are both small and unlikely to overlap.
- But the number of degrees of freedom of RIP is very large $(\binom{N}{k}(3/2)^k)$. It is unfeasible to explicitly prescribe this many values for even modest N and k.
- In addition, the problem of verifying RIP is known to be NP-hard.
- So any explicit construction must rely on some symmetry to reduce the degrees of freedom.

Norbert Wiener Center for Harmonic Analysis and Applications

The Coherence Method

- The standard method of explicitly constructing RIP matrices attempts to maximize the incoherence of the rows of A.
- The following is an equivalent characterization of RIP:

Definition

An $m \times N$ matrix A has the Restricted Isometry Property of order k with constant δ if for any $S \subset \{1, \ldots, N\}$ with $|S| \leq k$, every eigenvalue of $A_S^* A$ lies in the range $1 - \delta < \lambda < 1 + \delta$.

Norbert Wiener Center
for Harmonic Analysis and Applications

The Coherence Method

- Gershgorin's circle theorem gives us a method to bound the eigenvalues of a matrix in terms of its entries, i.e. the coherence of rows of A.

Theorem (Gershgorin's Circle Theorem)

Let A be $n \times n$ with entries $a_{i,j}$. Let $R_i = \sum_{j \neq i} |a_{i,j}|$ be the sums of the normed entries in row i of A. Then every eigenvalue of A falls in one of the discs $B(a_{i,i}, R_i)$.

- Prescribe A so that it has unit columns and maximum coherence μ. Then every eigenvalue falls within a distance of $(k-1)\mu$ of 1. If we can get $\mu \leq \delta/(k-1)$ then we're done.

Norbert Wiener Center for Harmonic Analysis and Applications

The Coherence Method

- But a bound from Welch [13] puts a bound on how small the coherence can be.

$$\mu \geq \sqrt{\frac{N-m}{m(N-1)}} \tag{5}$$

- This means that in order to get RIP using this method we need

$$\delta \geq (k-1)\sqrt{\frac{N-m}{m(N-1)}}.$$

If we again take N/m to be constant, this puts k on the order $O(m^{1/2})$ in order to control δ.

Norbert Wiener Center for Harmonic Analysis and Applications

Beating the Square-Root Bottleneck

There has been essentially one successful attempt to beat the bottleneck, pioneered by Bourgain, et al. I'll briefly go over a very abbreviated outline of his method, with help from an overview written by Dustin Mixon.

Beating the Square-Root Bottleneck

Definition

A is said to have Weak Flat RIP of order k with constant δ if for any disjoint $I, J \subset \{1, \dots, N\}$ with $|I|, |J| \leq k$,

$$\left| \left\langle \sum_{i \in I} \mathbf{a}_i, \sum_{j \in J} \mathbf{a}_j \right\rangle \right| \leq \delta k. \tag{6}$$

- Weak flat RIP, while weaker than restricting the maximum coherence, only implies RIP if we also put a bound on the coherence: $\mu \leq 1/k$, seeming to put us back inside the bottleneck.

Norbert Wiener Center
for Harmonic Analysis and Applications

Beating the Square-Root Bottleneck

- But all is not lost! If we can get the Weak Flat RIP constant very small, then we can apply the following lemma:

Lemma

If A has RIP of order k with constant δ, then it also has RIP of order sk with constant $2s\delta$, for any ≥ 1.

- We can scale δ and k simultaneously, so that a very small δ and a modest k can be turned into a modest δ and a larger k value.
- This is the approach Bourgain and his collaborators take: to find a sharp bound on the Weak Flat RIP using some sharp additive combinatorics, and then apply the approach above

Norbert Wiener Center for Harmonic Analysis and Applications

Beating the Square-Root Bottleneck

- In brief, the paper exploits a relationship between the additive energy of a set $S \subset \mathbb{F}_p$ and the complex exponential $\hat{\chi}_S$.
- Consider the family of vectors $\mathbf{u}_{a,b}(x) = p^{-1/2} e_p(ax^2 + bx)$, for some large prime b. Which has a nice expression for its mutual coherence:

$$\left\langle \mathbf{u}_{a_1,b_1}, \mathbf{u}_{a_2,b_2} \right\rangle = \frac{\sigma_p}{\sqrt{p}} \left(\frac{a_1 - a_2}{p} \right) e_p \left(-\frac{(b_1 - b_2)^2}{4(a_1 - a_2)} \right).$$

- Verifying weak flat RIP is then equivalent to finding a bound on a sum of complex exponentials

Beating the Square-Root Bottleneck

- The next step is to exploit the link between additive energy and the Fourier transform to bound the sum

$$\sum_{\substack{b_1 \in B_1 \\ b_2 \in B_2}} e_p(\theta(b_1 - b_2)^2)$$

  by the product of the additive energies and cardinalities of the sets B_1 and B_2.
- Last, the matrix A is defined to be $\mathbf{u}_{a,b}$ s.t. $(a, b) \in \mathcal{A}, \mathcal{B}$ for some $\mathcal{A}$, $\mathcal{B}$ carefully chosen to minimize additive energy.

Norbert Wiener Center
for Harmonic Analysis and Applications

Beating the Square-Root Bottleneck

- The above estimate still forces us to require that $k = \sqrt{p}$, but we get Weak Flat RIP with constant $p^{-2\varepsilon}$ for some $\varepsilon > 0$, which leads to RIP with constant $75p^{-\varepsilon}\ln p$.
- By applying the Lemma, we can convert this to RIP of order $k = p^{1/2+\varepsilon-\varepsilon'}$, for any $\varepsilon' > 0$, with constant $75p^{-\varepsilon'}\ln p < \sqrt{2} - 1$ for sufficiently large p.
- Hence the matrix construction indeed breaks the square-root bottleneck!

Norbert Wiener Center
for Harmonic Analysis and Applications

Beating the Square-Root Bottleneck

- Unfortunately, currently the best value we have for ε is on the order 10^{-24}.
- Followup work from Mixon sharpened Bourgain, et. al.'s estimates somewhat, but didn't yield any major insights.
- The state of the problem is unsatisfying: why is it so difficult to construct matrices with very little structure, especially when they're known to be plentiful?
- The facts we know about random matrices seem to virtually guarantee that improvement is possible in this regard.
- The connections with number theory and geometry suggest that future work could include some other deep mathematical insights.

References I

A. S. Bandeira, E. Dobriban, D. G. Mixon and W. F. Swain, Certifying the restricted isometry property is hard, *IEEE Transactions on Infromation Theory*, 2013.

J. Bourgain, S. Dilworth, K. Ford, S. Konyagin and D. Kutzarova, Explicit constructions of RIP matrices and related problems, *Duke Mathematical Journal*, 2011.

T. T. Cai and A. Zhang, Sparse representation of a polytope and recovery of sparse signals and low-rank matrices, *IEEE Transactions on Information Theory*, 2014.

E. Candes, The restricted isometry property and its implications for compressed sensing, *Comptes Rendus de l'Academie Sciences, Paris*, 2008.

E. Candes, J. Romberg and T. Tao, Robust uncertainty principles: exact signal reconstruction from highly incomplete frequency information, *IEEE Transactions on Information Theory*, 2006.

E. Candes, J. Romberg and T. Tao, Stable signal recovery from incomplete and inaccurate measurements, *Communications on Pure and Applied Mathematics*, 2006.

Norbert Wiener Center
for Harmonic Analysis and Applications

References II

- E. Candes and T. Tao, Decoding by linear programming, *IEEE Transactions on Information Theory*, 2005.
- E. Candes and T. Tao, Near optimal signal recovery from random projections: universal encoding strategies? *IEEE Transactions on Information Theory*, 2006.
- R. A. DeVore, Nonlinear approximation, *Acta Numerica*, 1998.
- S. Foucart, Flavors of compressive sensing, **Approximation Theory XV: San Antonio 2016**, pp. 61-104, ed. G. Fasshauer and L. L. Schumaker, *Springer*, 2016.
- S. Foucart and H. Rauhut, **A Mathematical Introduction to Compressive Sensing**, *Birkhauser*, 2013.
- D. Mixon, Explicit matrices with the restricted isometry property: breaking the square-root bottleneck, **Compressed Sensing and its Applications**, pp. 389-418, ed. H. Boche, R. Calderbank, Gitta Kutyniok and J. Vybiral, *Birkhauser*, 2013
- L. R. Welch, Lower bounds on the maximum cross correlation of signals, *IEEE Transactions on Information Theory*, 1974.

Norbert Wiener Center
for Harmonic Analysis and Applications