18.S096: Compressed Sensing and Sparse Recovery

Topics in Mathematics of Data Science (Fall 2015)

Afonso S. Bandeira
bandeira@mit.edu
http://math.mit.edu/~bandeira

November 10, 2015

These are lecture notes not in final form and will be continuously edited and/or corrected (as I am sure it contains many typos). Please let me know if you find any typo/mistake. Also, I am posting the open problems on my Blog, see [Ban15].

6.1 Sparse Recovery

Most of us have noticed how saving an image in JPEG dramatically reduces the space it occupies in our hard drives (as oppose to file types that save the pixel value of each pixel in the image). The idea behind these compression methods is to exploit known structure in the images; although our cameras will record the pixel value (even three values in RGB) for each pixel, it is clear that most collections of pixel values will not correspond to pictures that we would expect to see. This special structure tends to exploited via sparsity. Indeed, natural images are known to be sparse in certain bases (such as the wavelet bases) and this is the core idea behind JPEG (actually, JPEG2000; JPEG uses a different basis).

Let us think of $x \in \mathbb{R}^N$ as the signal corresponding to the image already in the basis for which it is sparse. Let's say that x is s-sparse, or $\|x\|_0 \leq s$, meaning that x has, at most, s non-zero components and, usually, $s \ll N$. The ℓ_0 norm[1] $\|x\|_0$ of a vector x is the number of non-zero entries of x. This means that, when we take a picture, our camera makes N measurements (each corresponding to a pixel) but then, after an appropriate change of basis, it keeps only $s \ll N$ non-zero coefficients and drops the others. This motivates the question: "If only a few degrees of freedom are kept after compression, why not measure in a more efficient way and take considerably less than N measurements?". This question is in the heart of Compressed Sensing [CRT06a, CRT06b, CT05, CT06, Don06, FR13]. It is particularly important in MRI imaging [**?**] as less measurements potentially means less measurement time. The following book is a great reference on Compressed Sensing [FR13].

More precisely, given a s-sparse vector x, we take $s < M \ll N$ linear measurements $y_i = a_i^T x$ and the goal is to recover x from the underdetermined system:

[1] The ℓ_0 norm is not actually a norm though.

$$\begin{bmatrix} \\ y \\ \\ \end{bmatrix} = \begin{bmatrix} & & \\ & A & \\ & & \end{bmatrix} \begin{bmatrix} \\ x \\ \\ \\ \end{bmatrix}.$$

Last lecture we used Gordon's theorem to show that, if we took random measurements, on the order of $s \log\left(\frac{N}{s}\right)$ measurements suffice to have all considerably different s-sparse signals correspond to considerably different sets of measurements. This suggests that $\approx s \log\left(\frac{N}{s}\right)$ may be enough to recover x, we'll see (later) in this lecture that this intuition is indeed correct.

Since the system is underdetermined and we know x is sparse, the natural thing to try, in order to recover x, is to solve

$$\begin{array}{ll} \min & \|z\|_0 \\ s.t. & Az = y, \end{array} \tag{1}$$

and hope that the optimal solution z corresponds to the signal in question x. Unfortunately, (1) is known to be a computationally hard problem in general. Instead, the approach usually taken in sparse recovery is to consider a convex relaxation of the ℓ_0 norm, the ℓ_1 norm: $\|z\|_1 = \sum_{i=1}^{N} |z_i|$. Figure 1 depicts how the ℓ_1 norm can be seen as a convex relaxation of the ℓ_0 norm and how it promotes sparsity.

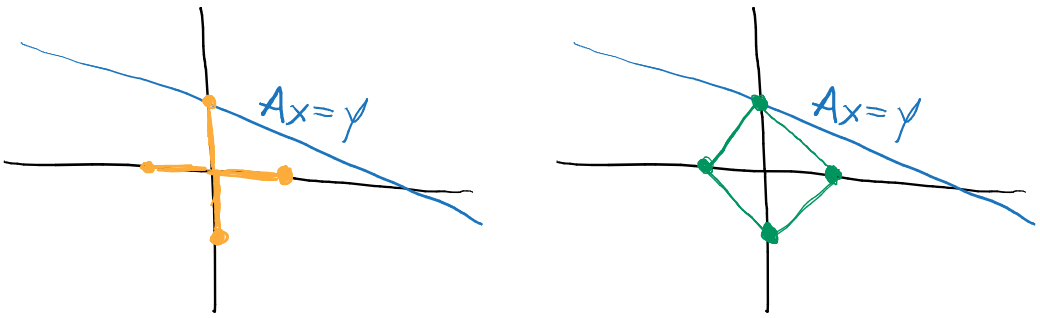

Figure 1: A two-dimensional depiction of ℓ_0 and ℓ_1 minimization. In ℓ_1 minimization (the picture of the right), one inflates the ℓ_1 ball (the diamond) until it hits the affine subspace of interest, this image conveys how this norm promotes sparsity, due to the pointy corners on sparse vectors.

This motivates one to consider the following optimization problem (surrogate to (1)):

$$\begin{array}{ll} \min & \|z\|_1 \\ s.t. & Az = y, \end{array} \tag{2}$$

In order for (2) we need two things, for the solution of it to be meaningful (hopefully to coincide with x) and for (2) to be efficiently solved.

We will formulate (2) as a Linear Program (and thus show that it is efficiently solvable). Let us think of ω^+ as the positive part of x and ω^- as the symmetric of the negative part of it, meaning that $x = \omega^+ - \omega^-$ and, for each i, either ω_i^- or ω_i^+ is zero. Note that, in that case,

$$\|x\|_1 = \sum_{i=1}^{N} \omega_i^+ + \omega_i^- = \mathbf{1}^T \left(\omega^+ + \omega^-\right).$$

Motivated by this we consider:

$$\begin{array}{ll} \min & \mathbf{1}^T \left(\omega^+ + \omega^-\right) \\ s.t. & A\left(\omega^+ - \omega^-\right) = y \\ & \omega^+ \geq 0 \\ & \omega^- \geq 0, \end{array} \tag{3}$$

which is a linear program. It is not difficult to see that the optimal solution of (3) will indeed satisfy that, for each i, either ω_i^- or ω_i^+ is zero and it is indeed equivalent to (2). Since linear programs are efficiently solvable [VB04], this means that (2) efficiently.

6.2 Duality and exact recovery

The goal now is to show that, under certain conditions, the solution of (2) indeed coincides with x. We will do this via duality (this approach is essentially the same as the one followed in [Fuc04] for the real case, and in [Tro05] for the complex case.)

Let us start by presenting duality in Linear Programming with a game theoretic view point. The idea will be to reformulate (3) without constraints, by adding a dual player that wants to maximize the objective and would exploit a deviation from the original constraints (by, for example, giving the dual player a variable u and adding to to the objective $u^T\left(y - A\left(\omega^+ - \omega^-\right)\right)$). More precisely consider the following

$$\min_{\substack{\omega^+ \\ \omega^-}} \max_{\substack{u \\ v^+ \geq 0 \\ v^- \geq 0}} \mathbf{1}^T \left(\omega^+ + \omega^-\right) - \left(v^+\right)^T \omega^+ - \left(v^-\right)^T \omega^- + u^T \left(y - A\left(\omega^+ - \omega^-\right)\right). \tag{4}$$

Indeed, if the primal player (picking ω^+ and ω^- and attempting to minimize the objective) picks variables that do not satisfy the original constraints, then the dual player (picking u, v^+, and v^- and trying to maximize the objective) will be able to make the objective value as large as possible. It is then clear that (3) = (4).

Now image that we switch the order at which the players choose variable values, this can only benefit the primal player, that now gets to see the value of the dual variables before picking the primal variables, meaning that (4) $\geq$ (5), where (5) is given by:

$$\max_{\substack{u \\ v^+ \geq 0 \\ v^- \geq 0}} \min_{\substack{\omega^+ \\ \omega^-}} \mathbf{1}^T \left(\omega^+ + \omega^-\right) - \left(v^+\right)^T \omega^+ - \left(v^-\right)^T \omega^- + u^T \left(y - A\left(\omega^+ - \omega^-\right)\right). \tag{5}$$

Rewriting

$$\max_{\substack{u \\ v^+ \geq 0 \\ v^- \geq 0}} \min_{\substack{\omega^+ \\ \omega^-}} \left(\mathbf{1} - v^+ - A^T u\right)^T \omega^+ + \left(\mathbf{1} - v^- + A^T u\right)^T \omega^- + u^T y \tag{6}$$

With this formulation, it becomes clear that the dual players needs to set $\mathbf{1} - v^+ - A^T u = 0$, $\mathbf{1} - v^- + A^T u = 0$ and thus (6) is equivalent to

$$\begin{array}{ll} \max\limits_{u} & u^T y \\ v^+ \geq 0 & \\ v^- \geq 0 & \\ \mathbf{1} - v^+ - A^T u = 0 & \\ \mathbf{1} - v^- + A^T u = 0 & \end{array}$$

or equivalently,

$$\begin{array}{ll} \max_u & u^T y \\ s.t. & -\mathbf{1} \leq A^T u \leq \mathbf{1}. \end{array} \tag{7}$$

The linear program (7) is known as the dual program to (3). The discussion above shows that (7) $\leq$ (3) which is known as weak duality. More remarkably, strong duality guarantees that the optimal values of the two programs match.

There is a considerably easier way to show weak duality (although not as enlightening as the one above). If ω^- and ω^+ are primal feasible and u is dual feasible, then

$$0 \leq \left(\mathbf{1}^T - u^T A\right)\omega^+ + \left(\mathbf{1}^T + u^T A\right)\omega^- = 1^T\left(\omega^+ + \omega^-\right) - u^T\left[A\left(\omega^+ - \omega^-\right)\right] = 1^T\left(\omega^+ + \omega^-\right) - u^T y, \tag{8}$$

showing that (7) $\leq$ (3).

6.3 Finding a dual certificate

In order to show that $\omega^+ - \omega^- = x$ is an optimal solution[2] to (3), we will find a dual feasible point u for which the dual matches the value of $\omega^+ - \omega^- = x$ in the primal, u is known as a *dual certificate* or *dual witness*.

From (8) it is clear that u must satisfy $\left(\mathbf{1}^T - u^T A\right)\omega^+ = 0$ and $\left(\mathbf{1}^T + u^T A\right)\omega^- = 0$, this is known as *complementary slackness*. This means that we must have the entries of $A^T u$ be $+1$ or -1 when x is non-zero (and be $+1$ when it is positive and -1 when it is negative), in other words

$$\left(A^T u\right)_S = \operatorname{sign}\left(x_S\right),$$

where $S = \operatorname{supp}(x)$, and $\left\|A^T u\right\|_\infty \leq 1$ (in order to be dual feasible).

Remark 6.1 *It is not difficult to see that if we further ask that $\left\|\left(A^T u\right)_{S^c}\right\|_\infty < 1$ any optimal primal solution would have to have its support contained in the support of x. This observation gives us the following Lemma.*

Lemma 6.2 *Consider the problem of sparse recovery discussed this lecture. Let $S = \operatorname{supp}(x)$, if A_S is injective and there exists $u \in \mathbb{R}^M$ such that*

$$\left(A^T u\right)_S = \operatorname{sign}\left(x_S\right),$$

[2]For now we will focus on showing that it is an optimal solution, see Remark 6.1 for a brief discussion of how to strengthen the argument to show uniqueness

and

$$\left\|\left(A^T u\right)_{S^c}\right\|_\infty < 1,$$

then x is the unique optimal solution to the ℓ_1 minimization program 2.

Since we know that $\left(A^T u\right)_S = \operatorname{sign}(x_S)$ (and that A_S is injective), we'll try to construct u by least squares and hope that it satisfies $\left\|\left(A^T u\right)_{S^c}\right\|_\infty < 1$. More precisely, we take

$$u = \left(A_S^T\right)^\dagger \operatorname{sign}(x_S),$$

where $\left(A_S^T\right)^\dagger = A_S\left(A_S^T A_S\right)^{-1}$ is the Moore Penrose pseudo-inverse of A_S^T. This gives the following Corollary.

Corollary 6.3 *Consider the problem of sparse recovery discussed this lecture. Let $S = \operatorname{supp}(x)$, if A_S is injective and*

$$\left\|\left(A_{S^c}^T A_S\left(A_S^T A_S\right)^{-1}\operatorname{sign}(x_S)\right)_{S^c}\right\|_\infty < 1,$$

then x is the unique optimal solution to the ℓ_1 minimization program 2.

Recall the definition of $A \in \mathbb{R}^{M\times N}$ satisfying the restricted isometry property from last Lecture.

Definition 6.4 (Restricted Isometry Property) *A matrix $A \in \mathbb{R}^{M\times N}$ is said to satisfy the (s,δ) restricted isometry property if*

$$(1-\delta)\|x\|^2 \le \|Ax\|^2 \le (1+\delta)\|x\|^2,$$

for all two s-sparse vectors x.

Last lecture (Lecture 5) we showed that if $M \gg s\log\left(\frac{N}{s}\right)$ and $A \in \mathbb{R}^{M\times N}$ is drawn with i.i.d. gaussian entries $\mathcal{N}\left(0,\frac{1}{M}\right)$ then it will, with high probability, satisfy the $(s,1/3)$-RIP. Note that, if A satisfies the (s,δ)-RIP then, for any $|S| \le s$ one has $\|A_S\| \le \sqrt{1+\frac{1}{3}}$ and $l\left(A_S^T A_S\right)^{-1}\| \le \left(1-\frac{1}{3}\right)^{-1} = \frac{3}{2}$, where $\|\cdot\|$ denotes the operator norm $\|B\| = \max_{\|x\|=1}\|Bx\|$.

This means that, if we take A random with i.i.d. $\mathcal{N}\left(0,\frac{1}{M}\right)$ entries then, for any $|S \le s|$ we have that

$$\|A_S\left(A_S^T A_S\right)^{-1}\operatorname{sign}(x_S)\| \le \sqrt{1+\frac{1}{3}}\frac{3}{2} = \sqrt{3}\sqrt{s},$$

and because of the independency among the entries of A, A_{S^c} is independent of this vector and so for each $j \in S^c$ we have

$$\operatorname{Prob}\left(\left|A_j^T A_S\left(A_S^T A_S\right)^{-1}\operatorname{sign}(x_S)\right| \ge \frac{1}{\sqrt{M}}\sqrt{3}\sqrt{s}t\right) \le 2\exp\left(-\frac{t^2}{2}\right),$$

where A_j is the j-th column of A.

Union bound gives

$$\operatorname{Prob}\left(\left\|A_S^T A_S\left(A_S^T A_S\right)^{-1}\operatorname{sign}(x_S)\right\|_\infty \ge \frac{1}{\sqrt{M}}\sqrt{3}\sqrt{s}t\right) \le 2N\exp\left(-\frac{t^2}{2}\right),$$

which implies

$$\text{Prob}\left(\left\|A_S^T A_S \left(A_S^T A_S\right)^{-1} \text{sign}\left(x_S\right)\right\|_\infty \geq 1\right) \leq 2N \exp\left(-\frac{\left(\frac{\sqrt{M}}{\sqrt{3s}}\right)^2}{2}\right) = \exp\left(-\frac{1}{2}\left[\frac{M}{3s} - 2\log(2N)\right]\right),$$

which means that we expect to exactly recover x via ℓ_1 minimization when $M \gg s\log(N)$, similarly to what was predicted by Gordon's Theorem last Lecture.

6.4 A different approach

Given x a sparse vector, we want to show that x is the unique optimal solution to

$$\begin{array}{ll} \min & \|z\|_1 \\ s.t. & Az = y, \end{array} \tag{9}$$

Let $S = \text{supp}(x)$ and suppose that $z \neq x$ is an optimal solution of the ℓ_1 minimization problem. Let $v = z - x$, it satisfies

$$\|v + x\|_1 \leq \|x\|_1 \quad \text{and} \quad A(v + x) = Ax,$$

this means that $Av = 0$. Also, $\|x\|_S = \|x\|_1 \geq \|v+x\|_1 = \|\left(v + x\right)_S\|_1 + \|v_{S^c}\|_1 \geq \|x\|_S - \|v_S\|_1 + \|v\|_{S^c}$, where the last inequality follows by triangular inequality. This means that $\|v_S\|_1 \geq \|v_{S^c}\|_1$, but since $|S| \ll N$ it is unlikely for A to have vectors in its nullspace that are this concentrated on such few entries. This motivates the following definition.

Definition 6.5 (Null Space Property) *A is said to satisfy the s-Null Space Property if, for all v in $\ker(A)$ (the nullspace of A) and all $|S| = s$ we have*

$$\|v_S\|_1 < \|v_{S^c}\|_1.$$

From the argument above, it is clear that if A satisfies the Null Space Property for s, then x will indeed be the unique optimal solution to (2). Also, now that recovery is formulated in terms of certain vectors not belonging to the nullspace of A, one could again resort to Gordon's theorem. And indeed, Gordon's Theorem can be used to understand the number of necessary measurements to do sparse recovery[3] [CRPW12]. There is also an interesting approach based on Integral Geometry [ALMT14].

As it turns out one can show that the $\left(2s, \frac{1}{3}\right)$-RIP implies s-NSP [FR13]. We omit that proof as it does not appear to be as enlightening (or adaptable) as the approach that was shown here.

6.5 Partial Fourier matrices satisfying the Restricted Isometry Property

While the results above are encouraging, rarely one has the capability of designing random gaussian measurements. A more realistic measurement design is to use rows of the Discrete Fourier Transform: Consider the random $M \times N$ matrix obtained by drawing rows uniformly with replacement from the $N \times N$ discrete Fourier transform matrix. It is known [CT06] that if $M = \Omega_\delta(K \,\text{polylog}\, N)$, then the resulting partial Fourier matrix satisfies the restricted isometry property with high probability.

[3] In these references the sets considered are slightly different than the one described here, as the goal is to ensure recovery of just one sparse vector, and not all of them simultaneously.

A fundamental problem in compressed sensing is determining the order of the smallest number M of random rows necessary. To summarize the progress to date, Candès and Tao [CT06] first found that $M = \Omega_\delta(K \log^6 N)$ rows suffice, then Rudelson and Vershynin [RV08] proved $M = \Omega_\delta(K \log^4 N)$, and more recently, Bourgain [Bou14] achieved $M = \Omega_\delta(K \log^3 N)$; Nelson, Price and Wootters [NPW14] also achieved $M = \Omega_\delta(K \log^3 N)$, but using a slightly different measurement matrix. The latest result is due to Haviv and Regev [HR] giving an upper bound of $M = \Omega_\delta(K \log^2 k \log N)$. As far as lower bounds, in [BLM15] it was shown that $M = \Omega_\delta(K \log N)$ is necessary. This draws a contrast with random Gaussian matrices, where $M = \Omega_\delta(K \log(N/K))$ is known to suffice.

Open Problem 6.1 *Consider the random $M \times N$ matrix obtained by drawing rows uniformly with replacement from the $N \times N$ discrete Fourier transform matrix. How large does M need to be so that, with high probability, the result matrix satisfies the Restricted Isometry Property (for constant δ)?*

6.6 Coherence and Gershgorin Circle Theorem

Last lectures we discussed the problem of building deterministic RIP matrices (building deterministic RIP matrices is particularly important because checking whether a matrix is RIP is computationally hard [BDMS13, TP13]). Despite suboptimal, coherence based methods are still among the most popular ways of building RIP matrices, we'll briefly describe some of the ideas involved here.

Recall the definition of the Restricted Isometry Property (Definition 6.4). Essentially, it asks that any $S \subset [N]$, $|S| \leq s$ satisfies:

$$(1-\delta)\|x\|^2 \leq \|A_S x\|^2 \leq (1+\delta)\|x\|^2,$$

for all $x \in \mathbb{R}^{|S|}$. This is equivalent to

$$\max_x \frac{x^T \left(A_S^T A_S - I\right) x}{x^T x} \leq \delta,$$

or equivalently

$$\left\|A_S^T A_S - I\right\| \leq \delta.$$

If the columns of A are unit-norm vectors (in $\mathbb{R}^M$), then the diagonal of $A_S^T A_S$ is all-ones, this means that $A_S^T A_S - I$ consists only of the non-diagonal elements of $A_S^T A_S$. If, moreover, for any two columns a_i, a_j, of A we have $\left|a_i^T a_j\right| \leq \mu$ for some μ then, Gershgorin's circle theorem tells us that $\left\|A_S^T A_S - I\right\| \leq \delta(s-1)$.

More precisely, given a symmetric matrix B, the Gershgorin's circle theorem [HJ85] tells that all of the eigenvalues of B are contained in the so called Gershgorin discs (for each i, the Gershgorin disc corresponds to $\left\{\lambda : |\lambda - B_{ii}| \leq \sum_{j \neq i} |B_{ij}|\right\}$. If B has zero diagonal, then this reads: $\|B\| \leq \max_i |B_{ij}|$.

Given a set of N vectors $a_1, \ldots, a_N \in \mathbb{R}^M$ we define its worst-case coherence μ as

$$\mu = \max_{i \neq j} \left|a_i^T a_j\right|$$

Given a set of unit-norm vectors $a_1, \ldots, a_N \mathbb{R}^M$ with worst-case coherence μ, if we form a matrix with these vectors as columns, then it will be $(s, \mu(s-1)\mu)$-RIP, meaning that it will be $\left(s, \frac{1}{3}\right)$- RIP for $s \leq \frac{1}{3}\frac{1}{\mu}$.

6.6.1 Mutually Unbiased Bases

We note that now we will consider our vectors to be complex valued, rather than real valued, but all of the results above hold for either case.

Consider the following $2d$ vectors: the d vectors from the identity basis and the d orthonormal vectors corresponding to columns of the Discrete Fourier Transform F. Since these basis are both orthonormal the vectors in question are unit-norm and within the basis are orthogonal, it is also easy to see that the inner product between any two vectors, one from each basis, has absolute value $\frac{1}{\sqrt{d}}$, meaning that the worst case coherence of this set of vectors is $\mu = \frac{1}{\sqrt{d}}$ this corresponding matrix $[I\ F]$ is RIP for $s \approx \sqrt{d}$.

It is easy to see that $\frac{1}{\sqrt{d}}$ coherence is the minimum possible between two orthonormal bases in $\mathbb{C}^d$, such bases are called unbiased (and are important in Quantum Mechanics, see for example [BBRV01]) This motivates the question of how many orthonormal basis can be made simultaneously (or mutually) unbiased in $\mathbb{C}^d$, such sets of bases are called mutually unbiased bases. Let $\mathcal{M}(d)$ denote the maximum number of such bases. It is known that $\mathcal{M}(d) \leq d+1$ and that this upper bound is achievable when d is a prime power, however even determining the value of $\mathcal{M}(6)$ is open [BBRV01].

Open Problem 6.2 *How many mutually unbiased bases are there in 6 dimensions? Is it true that* $\mathcal{M}(6) < 7$*?*[4]

6.6.2 Equiangular Tight Frames

Another natural question is whether one can get better coherence (or more vectors) by relaxing the condition that the set of vectors have to be formed by taking orthonormal basis. A tight frame (see, for example, [CK12] for more on Frame Theory) is a set of N vectors in $\mathbb{C}^M$ (with $N \geq M$) that spans $\mathbb{C}^M$ "equally". More precisely:

Definition 6.6 (Tight Frame) *$v_1, \ldots, v_N \in \mathbb{C}^M$ is a tight frame if there exists a constant α such that*

$$\sum_{k=1}^{N} |\langle v_k, x\rangle|^2 = \alpha \|x\|^2, \quad \forall_{x \in \mathbb{C}^M},$$

or equivalently

$$\sum_{k=1}^{N} v_k v_k^T = \alpha I.$$

The smallest coherence of a set of N unit-norm vectors in M dimensions is bounded below by the Welch bound (see, for example, [BFMW13]) which reads:

$$\mu \geq \sqrt{\frac{N-M}{M(N-1)}}.$$

One can check that, due to this limitation, deterministic constructions based on coherence cannot yield matrices that are RIP for $s \gg \sqrt{M}$, known as the square-root bottleneck [BFMW13, Tao07].

[4] The author thanks Bernat Guillen Pegueroles for suggesting this problem.

There are constructions that achieve the Welch bound, known as Equiangular Tight Frames (ETFs), these are tight frames for which all inner products between pairs of vectors have the same modulus $\mu = \sqrt{\frac{N-M}{M(N-1)}}$, meaning that they are "equiangular". It is known that for there to exist an ETF in $\mathbb{C}^M$ one needs $N \leq M^2$ and ETF's for which $N = M^2$ are important in Quantum Mechanics, and known as SIC-POVM's. However, they are not known to exist in every dimension (see here for some recent computer experiments [SG10]). This is known as Zauner's conjecture.

Open Problem 6.3 *Prove or disprove the SIC-POVM / Zauner's conjecture: For any d, there exists an Equiangular tight frame with d^2 vectors in $\mathbb{C}^d$ dimensions. (or, there exist d^2 equiangular lines in $\mathbb{C}^d$).*

We note that this conjecture was recently shown [Chi15] for $d = 17$ and refer the reader to this interesting remark [Mix14] on the description length of the constructions known for different dimensions.

6.6.3 The Paley ETF

There is a simple construction of an ETF made of $2M$ vectors in M dimensions (corresponding to a $M \times 2M$ matrix) known as the Paley ETF that is essentially a partial Discrete Fourier Transform matrix. While we refer the reader to [BFMW13] for the details the construction consists of picking rows of the $p \times p$ Discrete Fourier Transform (with $p \cong 1 \mod 4$ a prime) with indices corresponding to quadratic residues modulo p. Just by coherence considerations this construction is known to be RIP for $s \approx \sqrt{p}$ but conjectured [BFMW13] to be RIP for $s \approx \frac{p}{\text{polylog}p}$, which would be predicted if the choice os rows was random (as discussed above)[5]. Although partial conditional (conditioned on a number theory conjecture) progress on this conjecture has been made [BMM14] no unconditional result is known for $s \ll \sqrt{p}$. This motivates the following Open Problem.

Open Problem 6.4 *Does the Paley Equiangular tight frame satisfy the Restricted Isometry Property pass the square root bottleneck? (even by logarithmic factors?).*

We note that [BMM14] shows that improving polynomially on this conjecture implies an improvement over the Paley clique number conjecture (Open Problem 8.4.)

6.7 The Kadison-Singer problem

The Kadison-Singer problem (or the related Weaver's conjecture) was one of the main questions in frame theory, it was solved (with a non-constructive proof) in the recent breakthrough of Marcus, Spielman, and Srivastava [MSS15b], using similar techniques to their earlier work [MSS15a]. Their theorem guarantees the existence of universal constants $\eta \geq 2$ and $\theta > 0$ s.t. for any tight frame $\omega_1, \ldots, \omega_N \in \mathbb{C}^M$ satisfying $\|\omega_k\| \leq 1$ and

$$\sum_{k=1}^{N} \omega_k \omega_k^T = \eta I,$$

[5]We note that the quadratic residues are known to have pseudorandom properties, and indeed have been leveraged to reduce the randomness needed in certain RIP constructions [BFMM14]

there exists a partition of this tight frame $S_1, S_2 \subset [N]$ such that each is "almost a tight frame" in the sense that,

$$\sum_{k \in S_j} \omega_k \omega_k^T \preceq (\eta - \theta)\, I.$$

However, a constructive prove is still not known and there is no known (polynomial time) method that is known to construct such partitions.

Open Problem 6.5 *Give a (polynomial time) construction of the tight frame partition satisfying the properties required in the Kadison-Singer problem (or the related Weaver's conjecture).*

References

[ALMT14] D. Amelunxen, M. Lotz, M. B. McCoy, and J. A. Tropp. Living on the edge: phase transitions in convex programs with random data. 2014.

[Ban15] A. S. Bandeira. Relax and Conquer BLOG: Ten Lectures and Forty-two Open Problems in Mathematics of Data Science. 2015.

[BBRV01] S. Bandyopadhyay, P. O. Boykin, V. Roychowdhury, and F. Vatan. A new proof for the existence of mutually unbiased bases. *Available online at arXiv:quant-ph/0103162*, 2001.

[BDMS13] A. S. Bandeira, E. Dobriban, D.G. Mixon, and W.F. Sawin. Certifying the restricted isometry property is hard. *IEEE Trans. Inform. Theory*, 59(6):3448–3450, 2013.

[BFMM14] A. S. Bandeira, M. Fickus, D. G. Mixon, and J. Moreira. Derandomizing restricted isometries via the Legendre symbol. *Available online at arXiv:1406.4089 [math.CO]*, 2014.

[BFMW13] A. S. Bandeira, M. Fickus, D. G. Mixon, and P. Wong. The road to deterministic matrices with the restricted isometry property. *Journal of Fourier Analysis and Applications*, 19(6):1123–1149, 2013.

[BLM15] A. S. Bandeira, M. E. Lewis, and D. G. Mixon. Discrete uncertainty principles and sparse signal processing. *Available online at arXiv:1504.01014 [cs.IT]*, 2015.

[BMM14] A. S. Bandeira, D. G. Mixon, and J. Moreira. A conditional construction of restricted isometries. *Available online at arXiv:1410.6457 [math.FA]*, 2014.

[Bou14] J. Bourgain. An improved estimate in the restricted isometry problem. *Lect. Notes Math.*, 2116:65–70, 2014.

[Chi15] T.-Y. Chien. *Equiangular lines, projective symmetries and nice error frames.* PhD thesis, 2015.

[CK12] P. G. Casazza and G. Kutyniok. *Finite Frames: Theory and Applications.* 2012.

[CRPW12] V. Chandrasekaran, B. Recht, P.A. Parrilo, and A.S. Willsky. The convex geometry of linear inverse problems. *Foundations of Computational Mathematics*, 12(6):805–849, 2012.

[CRT06a] E. J. Candès, J. Romberg, and T. Tao. Robust uncertainty principles: exact signal reconstruction from highly incomplete frequency information. *IEEE Trans. Inform. Theory*, 52:489–509, 2006.

[CRT06b] E. J. Candès, J. Romberg, and T. Tao. Stable signal recovery from incomplete and inaccurate measurements. *Comm. Pure Appl. Math.*, 59:1207–1223, 2006.

[CT05] E. J. Candès and T. Tao. Decoding by linear programming. *IEEE Trans. Inform. Theory*, 51:4203–4215, 2005.

[CT06] E. J. Candès and T. Tao. Near optimal signal recovery from random projections: universal encoding strategies? *IEEE Trans. Inform. Theory*, 52:5406–5425, 2006.

[Don06] D. L. Donoho. Compressed sensing. *IEEE Trans. Inform. Theory*, 52:1289–1306, 2006.

[FR13] S. Foucart and H. Rauhut. *A Mathematical Introduction to Compressive Sensing.* Birkhauser, 2013.

[Fuc04] J. J. Fuchs. On sparse representations in arbitrary redundant bases. *Information Theory, IEEE Transactions on*, 50(6):1341–1344, 2004.

[HJ85] R. A. Horn and C. R. Johnson. *Matrix Analysis.* Cambridge University Press, 1985.

[HR] I. Haviv and O. Regev. The restricted isometry property of subsampled fourier matrices. *SODA 2016.*

[Mix14] D. G. Mixon. Short, Fat matrices BLOG: Gordon's escape through a mesh theorem. 2014.

[MSS15a] A. Marcus, D. A. Spielman, and N. Srivastava. Interlacing families i: Bipartite ramanujan graphs of all degrees. *Annals of Mathematics*, 2015.

[MSS15b] A. Marcus, D. A. Spielman, and N. Srivastava. Interlacing families ii: Mixed characteristic polynomials and the kadison-singer problem. *Annals of Mathematics*, 2015.

[NPW14] J. Nelson, E. Price, and M. Wootters. New constructions of RIP matrices with fast multiplication and fewer rows. *SODA*, pages 1515–1528, 2014.

[RV08] M. Rudelson and R. Vershynin. On sparse reconstruction from Fourier and Gaussian measurements. *Comm. Pure Appl. Math.*, 61:1025–1045, 2008.

[SG10] A. J. Scott and M. Grassl. Sic-povms: A new computer study. *J. Math. Phys.*, 2010.

[Tao07] T. Tao. What's new blog: Open question: deterministic UUP matrices. 2007.

[TP13] A. M. Tillmann and M. E. Pfefsch. The computational complexity of the restricted isometry property, the nullspace property, and related concepts in compressed sensing. 2013.

[Tro05] J. A. Tropp. Recovery of short, complex linear combinations via ℓ_1 minimization. *IEEE Transactions on Information Theory*, 4:1568–1570, 2005.

[VB04] L. Vanderberghe and S. Boyd. *Convex Optimization.* Cambridge University Press, 2004.