Combinatorial designs and compressed sensing

Padraig Ó Catháin

Monash University

29 September 2014

Joint work with: Darryn Bryant, Daniel Horsley, Charlie Colbourn.

Three hard problems

- Find the sparsest solution x to the linear system $Ax = b$ (given A and b).
- Given a subset of the entries of a matrix, find the completion with lowest rank.
- Express a given matrix M as $L + S$ where the rank of L is small and S is sparse.

All three problems are in NP.

Convex relaxation

- Each problem can be expressed as a linear programming problem, where the objective function involves minimising the solution under some suitable norm.
- The optimal solution of the linear programming problem can be found efficiently, but may or may not be an optimal solution to the original problem.
- The main result of compressed sensing is that, under weak conditions, the solution of the linear program is optimal with high probability.

A compressed sensing result

- Let x be an s-sparse vector in $\mathbb{R}^N$ (standard basis e_i).
- How many measurements do we need to take to recover x (with high probability)?
- What type of measurement should we take?

Theorem (Candès-Tao)

- *Let* $A = \{a_1, \ldots, a_n\}$ *be a set of measurements (linear functionals). Define the incoherence of* A *to be*

$$\mu_A = \max_j |\langle \sum_i a_i, e_j \rangle|^2$$

- *Then the number of measurements required to recover* x *is* $O(\mu_A s \log(N))$.
- *This result is best possible (no alternative sampling strategy can be asymptotically better).*
- *Via probabilistic constructions, measurement sets* A *exist with* $\mu_A = 1$.

Compressed sensing as linear algebra

Data	$\iff$	points in $\mathbb{R}^N$
Measurement	$\iff$	linear functional
'Most' data	$\iff$	Sparse vectors

$$\Phi x = b$$

- Under the assumption that x is sparse, how many measurements are required if $N = 1000$, say?
- Candes-Tao is asymptotic - no explicit bounds...
- What about deterministically constructing such a matrix?

Compressed sensing as linear algebra

$$\Phi x = b$$

Lemma

The matrix Φ allows recovery of all t-sparse vectors if and only if each t-sparse vector lies in a different coset of the nullspace of Φ.

- But no-one knows how to (deterministically) build (useful) matrices with this property.
- Without further assumptions, recovery of t-sparse vectors is an NP-Hard problem. (And furthermore is computationally infeasible in practice.)

Proxies for the null-space condition

Definition

The matrix Φ has the (ℓ_1, t)-property if, for any vector v of sparsity at most t, the ℓ_1-minimal solution of the system $\Phi x = \Phi v$ is equal to v.

Lemma

The matrix Φ has the (ℓ_1, t)-property if and only if, for every non-zero v in the null-space of Φ, the sum of the t largest entries of v is less than half of $\|v\|_1$.

(A statement about ℓ_1-norms but still computationally infeasible!)

Proxies for the nullspace condition

Say that Φ has the restricted Isometry property (t, δ)-RIP if the following inequality holds for all t-sparse vectors.

$$(1 - \delta) \leq \frac{\|\Phi x\|_2^2}{\|x\|_2^2} \leq (1 + \delta)$$

Theorem (Candes, Tao)

If Φ has (t, δ)-RIP with $\delta \leq \sqrt{2} - 1$, then Φ has the $(\ell_1, \frac{t}{2})$-property.

With overwhelming probability an $n \times N$ matrix with entries drawn from a Gaussian $(0, 1)$-rv has the $(\ell_1, n/\log(N))$-property, and this is optimal.

Sufficient conditions for deterministic constructions

$$\mu_\Phi = \max_{i \neq j} \left| \frac{\langle c_i, c_j \rangle}{|c_i||c_j|} \right|$$

Theorem (Donoho)

The following is sufficient (but not necessary) for Φ to have the (ℓ_1, t)-property.

$$t \leq \frac{1}{2\mu_\Phi} + \frac{1}{2}$$

So we want to construct matrices (with more columns than rows) where all inner products of columns are small.

The bottleneck

Theorem (Welch)

For any $n \times N$ matrix Φ, $\mu_\Phi \geq \mu_{n,N} = \sqrt{\frac{N-n}{n(N-1)}} = \frac{1}{\sqrt{n}}\sqrt{\frac{N-n}{N-1}}$.

- Donoho's method: Φ has the (ℓ_1, t)-property for all $t \leq \frac{1}{2\mu_\Phi} + \frac{1}{2}$.
- The Welch bound: $\mu_\Phi \geq \mu_{n,N} \geq \frac{1}{\sqrt{n}}$.
- The obvious conclusion: Donoho's method is limited to establishing the (ℓ_1, t)-property for

$$t \leq \frac{\sqrt{n}}{2} + \frac{1}{2} \sim O(\sqrt{n}).$$

- In contrast, Tao et al. give probabilistic constructions where $t \sim O(\frac{n}{\log(n)})$.

Ideally, we would like deterministic constructions which overcome this 'square-root bottleneck'.

- Unfortunately - overcoming the square-root bottleneck is hard.
- One construction from 2010: by Bourgain et al.
- 60 pages of hard additive combinatorics allows them to recover $O(n^{\frac{1}{2}+\epsilon})$-sparse vectors. (And this comes with restrictions on which parameters are constructible.)
- Instead, for any α we will give a construction for compressed sensing matrices with parameters $n \times \alpha n$ for all $n > C_{\alpha}$. All of these matrices recover $O(\sqrt{n})$-sparse vectors.

Equiangular frames

1. A frame is a collection of vectors (a generalisation of a basis in harmonic analysis). We write the vectors as columns in a matrix.
2. A frame is equiangular if for all columns c_i and c_j, there exists fixed α with
$$\mu(c_i, c_j) = \left| \frac{\langle c_i, c_j \rangle}{\|c_i\| \|c_j\|} \right| = \alpha.$$
3. If α meets the Welch bound, then such a matrix meets the square-root bottleneck exactly.

Definition

An *equiangular tight frame* (ETF) is a matrix in which $\mu(c_i, c_j) = \sqrt{\frac{N-n}{n(N-1)}}$ for every pair of columns c_i, c_j.

ETFs exist, but not very often. An ETF recovers $\frac{\sqrt{n}}{2}$-sparse vectors, and this result is best possible in the mutual incoherence framework.

Lemma

Let Φ be a frame and let $\mu_{n,N}$ be the Welch bound for Φ. Suppose that

$$(1-\epsilon)\mu_{n,N} \leq \left| \frac{\langle c_i, c_j \rangle}{|c_i||c_j|} \right| \leq (1+\epsilon)\mu_{n,N}$$

for all columns $c_i \neq c_j$ of Φ. Then Φ has the (ℓ_1, t)-property for

$$t \leq \frac{1}{2(1+\epsilon)\mu_{n,N}} + \frac{1}{2} \approx \frac{\sqrt{n}}{2(1+\epsilon)}.$$

Call such a frame ϵ-equiangular. We give constructions for 1-equiangular frames, and hence matrices with the (ℓ_1, t)-property for $t \leq \frac{\sqrt{n}}{4}$.

Definition

Let K be a set of integers. An incidence structure Δ on v points is a *pairwise balanced design* if every block of Δ has size contained in K, and every pair of points occurs in a single block. We denote such a design by $\mathrm{PBD}(v, K)$.

Example

A $\mathrm{PBD}(11, \{3, 5\})$:

$$\{abcde, 01a, 02b, 03c, 04d, 05e,$$

$$25a, 31b, 42c, 53d, 14e, 34a, 45b, 15c, 12d, 23e\}.$$

We construct a 1-equiangular frame Φ as follows:

- Let A be the incidence matrix of a PBD(v, K), Δ, with rows labelled by blocks and columns by points.
- So the inner product of a pair of columns is 1 (since any pair of points is contained in a unique block).
- For each column c, of A, we construct $|c|$ columns of Φ as follow:
  - Let H_c be a complex Hadamard matrix of order $|c|$.
  - If row i of c is 0, so is row i of each of the $|c|$ columns of Φ.
  - If row i of c is 1, row i of the $|c|$ columns of Φ is a row of $\frac{1}{|c|}H_c$.
  - No row of H_c is repeated.

Theorem (Bryant, Ó C., 2014)

Suppose there exists a $\mathrm{PBD}(v, K)$ *with*

- *n blocks*
- $\sum_{b \in B} |b| = N$
- $\max(K) \leq \sqrt{2} \min(K)$

Then there exists an $n \times N$ 1-equiangular frame. Equivalently, this is a compressed sensing matrix with the (ℓ_1, t)-property for all $t \leq \frac{\sqrt{n}}{4}$.

- This is a generalisation of a construction Fickus, Mixon and Tremain for Steiner triple systems.
- More generally, for any infinite family of PBDs with fixed K, we get $O(\sqrt{n})$-recovery.
- Our results can be improved in many directions: e.g. ϵ-equiangularity for $\epsilon < 1$ is possible, as is adding additional columns to the construction using MUBS, etc.

$$\Phi x = b$$

- So how do we **actually** find x?
- We could use the simplex algorithm, or basis pursuit or some algorithm for solving linear programming problems.
- Noise, negative entries in signal, which Hadamard matrices, etc.
- Example - a 2-$(73, 9, 1)$ (from a Singer difference set). Dimensions 146×1314. Lower bound on performance $\frac{\sqrt{146}}{4} \approx 3$. Upper bound on performance $2r - 1 = 17$.

Sample LP recovery results

Sparsity	Fourier	Fourth Roots	Gaussian
28	100	100	100
30	100	98	100
32	96	95	99
34	98	89	92
36	92	83	80
38	85	65	61
40	69	55	48
Time	101	73	654

(Time in seconds for 1000 recoveries.)

$$\Phi x = b$$

- A more efficient recovery algorithm:
- For each point (set of columns) in the original design, construct an estimate for the corresponding entries in x (Fourier transform).
- Choose the cn columns with estimates of largest absolute value, for suitable c.
- Solve the $n \times cn$ reduced system of linear equations for x.
- Run time is competitive with LP, and complexity of recovery should be $O(n \log n)$, with suitable assumptions.

Let Φ be a matrix constructed from a PBD and a Hadamard matrix H.

- If $\mathrm{Null}(\Phi)$ contains a $2t$-sparse vector then there exist t-sparse vectors u and v with $\Phi u = \Phi v$.
- Suppose Φ is constructed from Hadamard matrices of order r. Then $\mathrm{Null}(\Phi)$ contains $2r$-sparse vectors. (Though a small proportion of the total.)
- But are there any sparser ones?

$$\left(\begin{array}{cc|cc} \mathbf{1} & \mathbf{1} & \mathbf{1} & \mathbf{1} \\ \mathbf{1} & \mathbf{-1} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{1} & \mathbf{-1} \end{array}\right) \left(\begin{array}{c} \mathbf{1} \\ \mathbf{1} \\ \mathbf{-1} \\ \mathbf{-1} \end{array}\right) = 0.$$

Lemma (Ó C, 2014)

Let H be a complex Hadamard matrix of order n, and v a linear combination of $k \leq \frac{n}{t}$ rows of H. Then v contains at least t non-zero entries. Furthermore, if v contains exactly t non-zero entries, $t \mid n$ and v is a linear combination of exactly $\frac{n}{t}$ rows.

Example

Suppose Φ is constructed with real Hadamard matrices of order r. Then, provided the PBD contains three non-collinear points, the following is an element of the null-space of Φ of sparsity $\frac{3}{2}r$:

$$\left(\begin{array}{cc|cc|cc} \mathbf{1} & \mathbf{1} & \mathbf{1} & \mathbf{1} & \mathbf{0} & \mathbf{0} \\ \mathbf{1} & \mathbf{-1} & \mathbf{0} & \mathbf{0} & \mathbf{1} & \mathbf{1} \\ \mathbf{0} & \mathbf{0} & \mathbf{1} & \mathbf{-1} & \mathbf{1} & \mathbf{-1} \end{array}\right) \left(\begin{array}{c} \mathbf{0} \\ \mathbf{1} \\ \mathbf{0} \\ \mathbf{-1} \\ \mathbf{0} \\ \mathbf{1} \end{array}\right) = 0.$$

Lemma

Let Φ be a CS matrix built from normalised Hadamard matrices of order r and a PBD. The minimal support of an element of $\mathrm{Null}(\Phi)$ is

- $\frac{3}{2}r$ *if H contains a ± 1 row.*
- *More generally $\frac{k+1}{k}r$ if H contains k orthogonal rows of k^{th} roots of unity.*
- *Otherwise, $2r$.*

Fourier matrices over p^{th} roots of unity are optimal in this framework.
Question: What are other constructions for families of Hadamard matrices in which no linear combination of t rows vanishes in more than t positions?

- Given a PBD, we know how to construct a compressed sensing matrix with the $(\ell_1, \frac{\sqrt{n}}{4})$-property.
- For which (n, N) does there exist a PBD with n blocks in which the sum of the block sizes is N?

Theorem (Wilson)

Let K be a set of integers with $\gcd\{k-1 \mid k \in K\} = \gcd\{k(k-1) \mid k \in K\} = 1$. Then there exists a constant C such that, for every $v > C$, there exists a $\mathrm{PBD}(v, K, 1)$.

A necessary condition for existence of a PBD with block sizes K is that there exists a solution to the equation

$$\sum_{k \in K} \alpha_k \binom{k}{2} = \binom{v}{2}$$

Say that a solution to this equation is *realisable* if there exists a PBD with α_k blocks of size k for each $k \in K$. Wilson states that for each sufficiently large v, some solution is realisable.
We want more.

Say that a set of graphs $\mathcal{F}$ is *good* if, for every $G \in \mathcal{F}$, the gcd of the vertex degrees of G is 1.

Theorem (Caro-Yuster)

Let $\mathcal{F}$ be a good family of graphs. Denote by α_G the number of edges in G. Then exists a constant C such that for all $v > C$, every solution of the equation

$$\sum_{G \in \mathcal{F}} \alpha_G |G| = \binom{v}{2}$$

is realisable.

Decompositions into complete graphs $\Leftrightarrow$ PBDs
But: a family of complete graphs is never *good*...

- Consider
  $\mathcal{F} = \{F_1 = K_{n-1} + K_n, F_2 = K_n + K_{n+1}, F_3 = K_{n-1} + K_n + K_{n+1}\}$.
- Suppose K_v is $\mathcal{F}$-totally-decomposable.
- Decomposing the F_i into blocks, what decompositions can we obtain?
- Observe: decomposing into β_j copies of F_j, we obtain α_i blocks of size i.

$$\begin{pmatrix} 1 & 0 & 1 \\ 1 & 1 & 1 \\ 0 & 1 & 1 \end{pmatrix} \begin{pmatrix} \beta_1 \\ \beta_2 \\ \beta_3 \end{pmatrix} = \begin{pmatrix} \alpha_{n-1} & \alpha_n & \alpha_{n+1} \end{pmatrix}.$$

$$\begin{pmatrix} 0 & 1 & -1 \\ -1 & 1 & 0 \\ 1 & -1 & 1 \end{pmatrix} \begin{pmatrix} \alpha_{n-1} \\ \alpha_n \\ \alpha_{n+1} \end{pmatrix} = \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix}.$$

Theorem

For all sufficiently large v, and every choice of α_i satisfying the following conditions, there exists a $\mathrm{PBD}(v, \{n-1, n, n+1\}, 1)$ *with α_i blocks of size i.*

$$\begin{aligned} \alpha_n &\geq \alpha_{n-1} \\ \alpha_n &\geq \alpha_{n+1} \\ \alpha_{n+1} + \alpha_{n-1} &\geq \alpha_n \\ \alpha_{n-1}\binom{n-1}{2} + \alpha_n\binom{n}{2} + \alpha_{n+1}\binom{n+1}{2} &= \binom{v}{2} \end{aligned}$$

Using arguments of this type, we obtain:

Theorem

Let $h \in \mathbb{Q}$, and let $K = \{\lfloor h \rfloor - 1, \lfloor h \rfloor, \lfloor h \rfloor + 1\}$.

- *There exists a constant C_h, depending only on h*
- *For **every** $n > C_h$, there exists some $v \in \mathbb{N}$*
- *Such that there exists a $\mathrm{PBD}(v, K, 1)$ with **n blocks and average block size h**.*

Corollary

For any $h \in \mathbb{Q}$ and all sufficiently large n, there exists an $n \times \lfloor hn \rfloor$ compressed sensing matrix with the $(\ell_1, O(\sqrt{n}))$ property.