On the Doubly Sparse Compressed Sensing Problem

Grigory Kabatiansky
Institute for Information Transmission Problems (IITP), Moscow, Russia;
Cedric Tavernier, Assystem AEOS, France;
Serge Vladuts, IITP and Aix-Marseille Universite, Marseille, France

IMACC-2015

The Compressed Sensing (CS) subject was born in two papers: Donoho,"Compressed sensing", and Candes & Tao, "Near-Optimal Signal Recovery From Random Projections: Universal Encoding Strategies? ", both published in IEEE-IT, 2006.
The CS problem is to reconstruct an n-dimensional vector $x \in \mathbb{R}^n$, which is t-sparse, i.e. $||x||_0 = wt_H(x) = |\{i : x_i \neq 0\}| \leq t$, by a few linear measurements $s_i = (h_i, x)$ even if measurements (h_i, x) are known with some errors e_i.
Let us form the $r \times n$ matrix H, whose rows are $h_1, \dots, h_r$. Then the goal of CS is to find a t-sparse solution x of the following equation

$$s = Hx^T + e, \tag{1}$$

if Euclidean length of vector e, which we call the syndrome error vector, is small enough.

Finding solution of Eq. (1) with minimal l_0 is NP-hard. Theory of CS employes the following popular "trick" (used, for instance, in Lasso method) replacing this hard problem on finding solution with minimal l_1 norm. Namely, to find *arg* $\min \sum |x_i|$ such that $||s - Hx^T|| \leq \varepsilon$.

This problem is LP problem! Moreover it was proved that if matrix H is *RIP-matrix* (next slide) then the solution x^* of LP problem is a good *approximation* to the solution x_0 of the original problem and the corresponding number of measurements r has the minimal possible *order* of number of measurements, namely,

$$r_{min} = O(t \log \frac{n}{t}) \qquad (2)$$

Restricted δ-Isometry Property(RIP) matrix H if

$$(1-\delta_D)||x||_2 \leq ||Hx^T||_2 \leq (1+\delta_D)||x||_2, \qquad (3)$$

for any vector $x \in \mathbb{R}^n : ||x||_0 \leq D$, where $0 < \delta_D < 1$.
Typical result looks like this
"for $\delta_{3t} + 3\delta_{4t} < 2$ the solution x^* of linear programming problem is unique and equal to x if $e = 0$"
and, additionally *" $||x^* - x||_2 \leq C\varepsilon$ for any perturbation e with $||e||_2 \leq \varepsilon$".*
We can rewrite it as $\delta_{4t} < 1/2$ or that for any z s.t. $||z||_0 \leq 4t$

$$\frac{1}{2}||z||_2 \leq ||Hz^T||_2 \leq \frac{3}{2}||z||_2$$

It was proved that such RIP-matrices exist if $r = O(t log(n/t))$

Let us consider $r \times n$ matrix H of Eq. 1 as a parity-check matrix of a linear $(n, n-r)$-code C_H, where

$$C_H = \{x \in K^n : Hx^T = 0\} \tag{4}$$

Then Eq. (1) is known as the *syndrome equation* and the only difference from Coding Theory is that we know the syndrome vector $s = (s_1, \ldots, s_r)$ not exactly!
Indeed, if $e = 0$ then we have ordinary coding theory problem, and it can be solved for $\mathbb{R}$ or $\mathbb{C}$, for instance, by Reed-Solomon codes and one gets exact solution with minimal possible $r = 2t$ and polynomial algorithm of x recovery. But if $e \neq 0$ we have a new problem of coding theory:-)

The CS problem is usually investigated under assumption that syndrome error vector $e = (e_1, \ldots, e_r)$ has relatively small Euclidean norm (length) $||e||_2$. Discrete case motivates us to consider another assumption, namely, that the vector e is also sparse, say $||e||_0 \leq L$. Let us call these assumptions: $||x||_0 \leq t$ and $||e||_0 \leq L$, as **double sparse**. It means that Euclidean norm of e can be arbitrary large and we will be able to find not an approximation but exact solution of Eq.(1)!
Our main result is that for double sparse CS problem

$$r_{min} = 2(t + L) \tag{5}$$

and we construct **optimal** measurement matrices which allow to recover x **exactly** and with polynomial complexity.

Definition

An $r \times n$ matrix H over field K called a (t, L)-double sparse compressed sensing (DSCS) matrix if

$$||Hx^T - Hy^T||_0 \geq 2L + 1 \quad (6)$$

for any two distinct vectors $x, y \in K^n$ such that $||x||_0 \leq t$ and $||y||_0 \leq t$.

This definition immediately leads to the following
Proposition A matrix H is a (t, L)-DSCS matrix iff

$$||Hz^T||_0 \geq 2L + 1 \quad (7)$$

for any nonzero vector $z \in K^n$ such that $||z||_0 \leq 2t$.

Consider the following measurement (or, *parity-check*) matrix $H = (h_{ij})$, where $h_{ij} = \alpha_j^{i-1}$ and $\alpha_1, \ldots, \alpha_n$ are distinct elements of the field K. Note that the field K can be finite and can be infinite, for instance, $\mathbb{R}$ or $\mathbb{C}$. If $e = 0$, i.e. one knows all sums

$$s_i = \sum_{j=1}^{n} x_j \alpha_j^{i-1} \tag{8}$$

exactly, then x can be recovered (and efficiently by Berlekamp-Massey algorithm) if $||x||_0 \leq r/2$.
But what will happen if some s_i are known with errors?!
In fact, it is very old question, which goes back to the time of French Revolution, see R.Prony in J. de Ecole Polytechnique 1, pp.24-76, 1795!

The modern solution was given in M.T. Comer, E.L. Kaltofen, C.Pernet "Sparse Polynomial Interpolation and Berlekamp-Massey Algorithms That Correct Outlier Errors in Input Values" (2012).
Namely, it was shown that it is possible to solve equation (1) by RS-code *iff* its redundancy $r \geq 2t(2L+1)$.
We shall show that it is too much expensive solution for double sparse CS-problem.

Consider matrix

$$H = G_B^T H_A, \tag{9}$$

where an $r' \times n$ matrix H_A is a parity-check matrix of an $(n, n - r')$-code A over field K, correcting t errors, and G_B is a generator matrix of an (r, r')-code B over K of length r, correcting L errors.

Saying in words, we encode columns of parity-check matrix H_A, which already capable to correct t errors, by a code, correcting L errors, in order to restore correctly syndrom of H_A.

Theorem

Matrix $H = G_B^T H_A$ *is a* (t, L)*-DSCS matrix.*

Proof. According to Proposition it is enough to prove that $||Hz^T||_0 \geq 2L+1$ for any nonzero vector $z \in K^n$ such that $||z||_0 \leq 2t$. Note that

$$Hz^T = G_B^T(H_A z^T) = (uG_B)^T$$

belongs to code B, where $u^T = H_B z^T$. Then $H_B z^T \neq 0$ since any $2t$ columns of H_B are linear independent. Hence $Hz^T = (uG_B)^T$ is a nonzero code vector from code B and its Hamming weight $wt(Hz^T) \geq 2L+1$, what concludes the proof. □

How to decode? First we decode vector $\hat{s} = s + e$ by a decoding algorithm of the code with generator matrix G_B. Since $||e||_0 \leq L$ this algorithm outputs the correct syndrome s. Then we form a syndrome $\tilde{s}$ by selecting first $\tilde{r}$ coordinates of s (we assume w.l.o.g. that G is a systematic encoding matrix) and apply syndrom decoding algorithm for the corresponding syndrom equation

$$\tilde{s} = H_A x^T. \tag{10}$$

Theorem

(Singleton bound)For any (t, L)*-DSCS* $r \times n$*-matrix*

$$r \geq 2(t + L). \tag{11}$$

Proof. Let H be a (t, L)-DSCS matrix of size $r \times n$, i.e., $||Hz^T||_0 \geq 2L + 1$ for any nonzero vector $z \in K^n : ||z||_0 \leq 2t$. And let H_{2t-1} be $(2t - 1) \times n$ matrix consisting of first $2t - 1$ rows of H. There exists a nonzero vector $\hat{z} = (\hat{z}_1, \ldots, \hat{z}_{2t}, 0, 0, \ldots, 0) \in K^n$ such that $H_{2t-1}\hat{z}^T = 0$ (a system of linear homogenious equations with the number of unknown variables larger than the number of equations has a nontrivial solution). Then $||H\hat{z}^T||_0 \leq r - (2t - 1)$ and finally $r \geq 2t + 2L$ since $||H\hat{z}^T||_0 \geq 2L + 1$. □

Now let us construct concatenated matrix $H = G_B^T H_A$ by taking RS-codes over $\mathbb{C}$ (or $\mathbb{R}$) as both constituent codes. Note, that the usual restriction on the code length for RS-codes is void over these fields. Therefore the resulting matrix H gives the optimal solution for the doubly sparse compressed sensing problem with the number of measurements just $2(t + L)$. Moreover, we can take as the corresponding recovery (or decoding) algorithm Berlekamp-Massey or Guruswami-Sudan algorithms.

Conclusion

TO FIND SOMETHING IN BETWEEN RIP-MATRICES AND PROPOSED CONCATENATED MATRICES.