Project Presentation

Dhanus M Lal

April 2022

Introduction

Suppose we have a signal $x \in \mathbb{R}^n$. The number of mesurements required to completely understand the signal is n. But when we have some additional information about x, say, all but the first coordinate of x is equal to 0, we only need a single measurement: we can completely understand x by making the measuremnt $y = \langle e_1, x \rangle$. However, if we relax the condition to be that at most one of the coordinates of the vector x is zero, it is not very apparent whether it is possible to recover x with fewer than n measurements, since the non-zero coordinate could be in any of the n possible locations. A surprising result is that such a signal can be recovered in $\mathcal{O}(\log(N))$ measurements.

Sparse Recovery Problem

Now we can state the sparse recovery problem as recovering x from the linear measurement $y = Ax$, when it is given that x is s-sparse (that is, all but at most s coordinates of x are non-zero). Clearly, the recovery problem does not have solution for every choice of measurement matrix A. There could be situations where multiple s sparse vectors have the same measurement. Now, the question is whether a measurement matrix A with $m < n$ exists that makes sparse recovery possible.

Let us take A to be an $m \times n$ matrix with $m > 2s$ and with every subset of $2s$ columns of A beings linearly independent. Then it can be shown that the sparse recovery problem has a unique solution.

In order to prove this, take x_1, x_2 to be two s-sparse vectors satisfyting $Ax_1 = Ax_2$. Then $A(x_1 - x_2) = 0$. Now, $x_1 - x_2$ is $2s$ sparse, so $A(x_1 - x_2) = 0$ would mean that A has a linearly dependent subset of size less than $2s$, which is a contradiction.

The proposition shows that there are measurement matrices that can solve sparse recovery problem with very few measurements than n.

If the solution is unique, the solution can be obtained by solving

$$x = \arg\min \|x'\|_0 \quad \text{s.t } Ax' = y.$$

However, finding the exact solution would require solving the linear system $A_I x = y$ over every subset I of size at most s. (A_I is the $m \times |I|$ submatrix of A with only indexes in I selected). Since there are at least $\binom{n}{s}$ such sets, performing this is not computationally feasible.

This problem can be slightly modified by replacing the non convex $\|\cdot\|_0$ by its closest ℓ_p norm, which is the ℓ_1 norm. The modified problem is:

$$x = \arg\min \|x'\|_1 \quad \text{s.t } Ax' = y.$$

This is a convex optimization problem and is called basis pursuit. It can be solved efficiently using standard convex optimization algorithms like linear programming.

Now we need to adress the following questions:

- How close is the solution of basis pursuit to the original solution?
- When does basis pursuit exactly recover the original signal?

Recovery Using Random Matrices

Let us now take the measurement matrix A to be a random matrix from the class $\mathcal{A}(m, n)$, which is the set of matrices with independent, isotropic, sub-gaussian rows. This is done because various results about this class of random matrices is known which would help in finding the recovery error and conditions for exact recovery.

A result which we will use is called the M^* bound, which states

$$\mathbb{E}\mathrm{diam}(T \cap E) \leq Cw(T)/\sqrt{m}$$

Using this, we can get a bound on the expected ℓ_2 recovery error for the solution of basis pursuit.

$$\mathbb{E} \sup_{x':Ax'=y} \|x - x'\|_2 \leq \|x\|_2 Cw(T)/\sqrt{m}$$

This is obtained by noting that $x/\|x\|_2$ and $\hat{x}/\|x\|_2$ both lie in the scaled ℓ_2 ball $\sqrt{s}B_1^n$ and the gaussian width of this set is bounded by $\sqrt{s \log n}$.

Exact Recovery for $A \in \mathcal{A}$

Another property of $\mathcal{A}(m, n)$ is the escape theorem, which states that if m is large, kernel of A misses subsets of the unit sphere with high probability. This gives theorem gives the number of measurements required for exact recovery with high probability using basis pursuit to be $\geq Cs \log n$ (larger than the square of $w(\sqrt{s} \log n)$) This is because if $h = \hat{x} - x$ is non-zero, it can be shown that $h/\|h\|_2$ lies inside $2\sqrt{s}B_1^n \cap S^{n-1}$, which is empty with high probability because of the condition on m.

RIP Implies Exact Recovery

Now we will fix the measurement matrix to be from a deterministic class of matrices. A deterministic condition for measurement matrices so that exact recovery becomes possible is Restricted Isometry Property. A matrix A satisfies RIP if for every $3s$-sparse vector, $\|Ax\|_2$ is almost close to $\|x\|_2$ (lies between $0.9\|x\|_2$ and $1.1\|x\|_2$). If a matrix satisfies RIP, then basis pursuit solves the sparse recovery problem for s–sparse vectors. The proof involves only triangle inequality and Cauchy-Schwarz inequality.

However, constructing matrices that satisfy RIP is difficult, since all the singular values of each sub-matrix of $3s$ columns should have all singular values between 0.9 and 1.1.

Surprisingly, the matrices from class $\mathcal{A}(m, n)$ satisfy RIP with high probability, if m is large enough. This is proved using the tail bounds on the singular values of the matrices in class $\mathcal{A}$ and then a union bound.

Further Reading

In some applications the measurement matrix is fixed. For example, say, the measurement can only be made in the frequence domain of a signal. In this case the measurement matrix is the Fourier matrix. Since the signal is sparse, we would like to reconstruct the signal from the measurement using only a subset of Fourier coefficients. A result proved by Candes and Tao in 2006 says that exact recovery using partial Fourier coefficients is possible when enough number of Fourier coefficients are sampled randomly.

For matrices, there is another notion of sparsity, which is rank. The measurements are made by taking inner-products with measurement matrices. An equivalent of basis pursuit exists in this case and the expected ℓ_2 error can be obtained using the M^* bound.

Numerical simulations can be run to test the validity of these results. In practical applications, the data is very high dimensional and require heavy computational power.