CS6220 Lecture Notes: Basis Pursuit De-noising, LASSO, and Compressed Sensing

Sebastian Ament

October 31, 2017

1 Basis Pursuit De-noising and LASSO

1.1 Introduction

Both LASSO (least absolute shrinkage and selection operator) and BPDN (Basis Pursuit De-noising) are methods which deal with the following problem. Let

$$A = [I \ F], \tag{1}$$

where I is the identity and F is the Fourier transform matrix. If $b = Ax$, where x is sparse, how do we recover this sparse solution, given the observations b and that A is over-complete?

It turns out that solving

$$\begin{aligned} &\min_x \|x\|_1 \\ &\text{s.t. } \ Ax = b \end{aligned} \tag{2}$$

can recover the sparse x. This is the basis of both LASSO and BPDN, which are similar methods, but were developed by different research communities. LASSO was developed by the statistics community, while BPDN was developed by the signal processing community.

In real-world applications, the observations b might be noisy. Therefore, it would be better to solve

$$\begin{aligned} &\min_x \|x\|_1 \\ &\text{s.t. } \ \|Ax - b\|_2 \leq \epsilon \end{aligned} \tag{3}$$

Note that this problem is equivalent to

$$\min_x \|Ax - b\|_2 + \lambda \|x\|_1 \tag{4}$$

for some choice of λ. A number of optimization packages have been developed to deal with these types of problems. For example:

1. ASP by Friedlander and Saunders
2. CVX

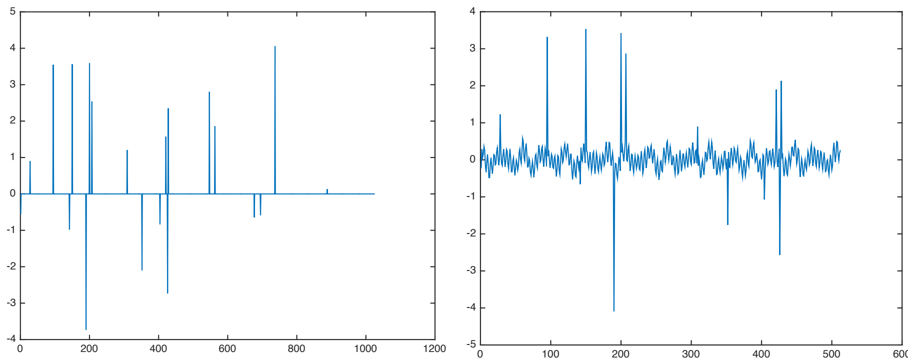

Figure 1: Sparse x (left), and Ax (right)

1.2 Demonstration and Discussion of BPDN

In the following, we will demonstrate the basis pursuit algorithm. The Matlab code for this demonstration, which uses ASP, can be found here.

Without noise perfect recovery is possible. Indeed, the left side of 1 shows a sparse x in the basis $A = [I, T]$, where T is the discrete cosine transform. The right side of 1 shows Ax, clearly not a sparse signal. However, solving (2) recovers x to arbitrary precision, depending on the stopping condition of the optimization method.

As the noise increases, the corresponding ϵ in the constraint of (4) has to be adjusted. If ϵ is below the noise level, the resulting x will be less sparse, as a tight fit to the data has to be enforced. If we adjust ϵ, the solution will become sparse again, though small coefficients might get pushed to zero. This is because the error incurred by missing a small coefficient is likely to be insignificant to the fit for a large ϵ. Intriguingly, the recovery of x is still successful up to the noise level.

1.3 Comparison of BPDN to Competing Methods

[1] compares basis pursuit to other decomposition methods into over-complete bases. A notable one is the method of frames (MOF). This method relies on solving

$$\begin{aligned} &\min_x \|x\|_2 \\ &\text{s.t. } Ax = b. \end{aligned} \tag{5}$$

The advantage of this method is that its solution is available in closed form. That is, $x = A^T(AA^T)^{-1}b$. However, the solution to (5) is in general not sparse, if A is over-complete.

Matching pursuit (MP), another method, works by greedily adding non-zero coefficients to x one at a time, based on which basis vector is most correlated with the current residual. However, this method is also not guaranteed to be sparsity preserving, if A is not orthogonal, which is the case for any over-complete basis. If the columns of A are in fact orthogonal, this algorithm is also called orthogonal matching pursuit (OMP). OMP can be shown to recover the sparsity pattern of a k-sparse x if the mutual incoherence μ of A is smaller than $\frac{1}{2k-1}$ (see [2]). Recall that the mutual incoherence μ of A is $\mu(A) = \max_{i \neq j} (A_{\cdot i})^T A_{\cdot j}$.

Best orthogonal basis (BOB), attempts to adaptively select an orthogonal basis out of an over-complete basis, in order to represent the signal. This works well for signals which are generated by orthogonal basis functions. However, it does not yield sparse representations for signals which are generated by non-orthogonal elements of the basis at hand.

In summary, BPDN delivers superior performance with regards to recovering sparse signals for a wide class of problems, compared to the other methods outlined above.

1.4 BPDN and Quadratic Programming

Consider $y = Ax + \sigma z$, where $z \sim \mathcal{N}(0, 1)$. If we relax $\min \|x\|_1$ s.t. $Ax = b$ to $\min \frac{1}{2}\|Ax - b\|_2^2 + \lambda\|x\|_1$, this can we written as a quadratic program. In particular,

$$\begin{aligned} \min_{u,v,r} \; & \lambda 1^T(u+v) + \frac{1}{2} r^T r \\ \text{s.t.} \; & \begin{bmatrix} A & -A \end{bmatrix} \begin{bmatrix} u \\ v \end{bmatrix} + r = b \\ & u, v \geq 0 \end{aligned} \tag{6}$$

Now suppose that there exists a feasible x_0 of the BPDN problem

$$\begin{aligned} \min_{u,v,r} \; & \|x\|_1 \\ \text{s.t.} \; & \|Ax - b\|_2 \leq \epsilon \end{aligned} \tag{7}$$

such that $\|x_0\|_0 \leq \frac{1}{4}(1 + \frac{1}{\mu(A)})$. Then $\hat{x}$, the solution to (6) satisfies

$$\|\hat{x} - x_0\|_2^2 \leq \frac{4\epsilon^2}{1 - \mu(A)(4\|x\|_0 - 1)}. \tag{8}$$

2 Next Lecture: Compressed Sensing

Let f be a signal that is sparse in some basis. That is,

$$f = [I \; \mathrm{T}]x, \tag{9}$$

for some sparse x, where T is the discrete cosine transform. Compressed sensing deals with the problem of recovering f with some small number of measurements $y_i^T f$ for $i = 1, 2, ..., m$. It turns out, this is feasible! But given f, with sparse x s.t. $f = \Phi x$, how many linear measurement y_i ($\hat{f}_i = y_i^T f$) do we need to exactly recover f? This depends on 2 key things.

1. Sparsity of x
2. Incoherence between y and Φ.

To be continued on November 2nd.

References

[1] Scott Shaobing Chen, David L. Donoho, and Michael A. Saunders. Atomic decomposition by basis pursuit. *SIAM Rev.*, 43(1):129–159, January 2001.

[2] T. T. Cai and L. Wang. Orthogonal matching pursuit for sparse signal recovery with noise. *IEEE Transactions on Information Theory*, 57(7):4680–4688, July 2011.