University of Wisconsin-Madison Lecturer: Benjamin Recht

CS838 Topics In Optimization:
Convex Geometry in High-Dimensional Data Analysis

February 2, 2010 Scribe: Matt Malloy

Lecture 5

1 Sparsity

Definition 1. $\mathbf{x} \in \mathbb{R}^D$ *is* s-sparse *if exactly* s *of the* D *components of* x *are non-zero.*

Definition 2. *The* support *of* $\mathbf{x}$ *is defined as:*

$$\text{supp}(\mathbf{x}) = \{i : x_i \neq 0\}. \tag{1}$$

Definition 3. *The* cardinality *of* $\mathbf{x}$*, denoted* $\text{card}(\mathbf{x})$ *is the number of elements in the support of* x.

A vector is s-sparse if and only if $\text{card}(\mathbf{x}) = s$.

1.1 Sparsity Maximization Problem

We will spend the next few weeks studying the following optimization problem

Problem: SPARSEST VECTOR Given $\epsilon \geq 0$, $\mathbf{A}$ a $[m \times D]$ matrix, and $\mathbf{b} \in \mathbb{R}^n$. What is the sparsest x satisfying $||\mathbf{Ax} - \mathbf{b}|| \leq \epsilon$?

That is, we seek to

$$\begin{array}{ll} \text{minimize} & \text{card}(\mathbf{x}) \\ \text{subject to} & \|\mathbf{Ax} - \mathbf{b}\| \leq \epsilon \end{array} . \tag{2}$$

Of particular interest is when $\epsilon = 0$. What is the sparsest solution of $\mathbf{Ax} = \mathbf{b}$?

There are many situations where we would like to find low cardinality solutions to optimization problems.

- **Model selection** Consider a list of attributes about a process - $\{x_1, x_2, ..., x_D\}$. Our goal is to predict a value, y, from these attributes. We can write our model as a linear combination of the attributes:

  $$y = \sum w_i x_i. \tag{3}$$

  Our goal is to minimize $\text{card}(\mathbf{w})$.

- **Compressed Sampling**. Most of the energy in a natural image lies in low order discrete cosine transform (DCT) components. Can we take a few specially coded samples from which we can perfectly reconstruct the image?

- **Astronomy** Sparsity already lies in the pixel basis. Consider an image of the nighttime sky.

- **Error correcting codes** Consider a signal in which a number of the elements have been corrupted. How can we create a code to correct these errors? We assume errors are sparse. If you know the signal is zero, you can find the errors only from the parity check.

1.2 Sparse Approximation is Hard

We will show that finding a sparse vector that approximates the linear system $\mathbf{Ax} = \mathbf{b}$ is NP-HARD. We do so by a reduction from the Exact Cover Problem, proposed by Natarajan in [1].

Problem: EXACT COVER Let S be a set with m elements. Let C be a collection of D subsets of S, each having exactly 3 elements. Does there exist a sub-collection, $\hat{C}$ from C such that every element of S is a member of exactly one set from $\hat{C}$?

To reduce EXACT COVER to a sparse approximation problem, enumerate the elements of S and C

$$\begin{aligned} S &= \{S_1, S_2, S_3, ..., S_m\} && (4)\\ C &= \{C_1, C_2, C_3, ..., C_D\} && (5) \end{aligned}$$

and then define the $m \times D$ matrix

$$A_{i,j} = \begin{cases} 1 & S_i \in C_j \\ 0 & else \end{cases} \tag{6}$$

Note A has exactly three 1's per column. Let $\mathbf{b} = [1 \quad 1 \quad ... \quad 1]^T$, $\mathbf{b} \in \mathbb{R}^m$ and $\epsilon = 1/2$.

Proposition 1. *Suppose we have an algorithm for the sparse approximation problem, and our routine returns* $\mathbf{x}$ *with* $\text{card}(\mathbf{x}) = s$. $\text{card}(\mathbf{x}) \leq \frac{m}{3}$ *if and only if an exact cover exists.*

Proof. Suppose an exact cover exists. Let

$$x_i = \begin{cases} 1 & C_i \in \text{exact cover} \\ 0 & \text{otherwise} \end{cases} \tag{7}$$

Then,

$$\mathbf{Ax} = \mathbf{b} \Rightarrow s \leq \frac{m}{3} \tag{8}$$

Conversely, suppose $s \leq \frac{m}{3}$. Then since $||\mathbf{Ax} - \mathbf{b}|| \leq \frac{1}{2}$, we have

$$\frac{1}{2} \leq \sum_j A_{i,j} x_j \leq \frac{3}{2} \quad \forall \ i\,. \tag{9}$$

To construct an exact cover, let $C_j \in \hat{C}$ if $j \in \text{support}(x)$.

$$\frac{1}{2} \leq \sum_j A_{i,j} \leq \sum_{C_j \in \hat{C}} A_{i,j} ||x||_\infty\,. \tag{10}$$

This means that $\sum_{C_j \in \hat{C}} A_{i,j} > 0$ for all i. In particular, since all of the entries of $\mathbf{A}$ are either 0 or 1, $\sum_{C_1 \in \hat{C}} A_{i,j} > 1$ for all $i = 1, \ldots m$. Since $|\hat{C}| \leq m/3$, and each column of $\mathbf{A}$ has exactly 3 non-zeros, this means that $\sum_{C_1 \in \hat{C}} A_{i,j} = 1$ for all i, and $\hat{C}$ is an exact cover.

□

1.3 Algorithm for finding the sparsest vector

What would be a simple algorithm for finding the sparsest vector?

```
for i=1,...,m
    for a subset I of {1,...,D} of size s
        try to solve A_I z = b
                %where A_I is the submatrix of A
                %consisting of the columns indexed by I.
        if a solution is found
            return x_i = z_i if i is in I, 0 otherwise
```

Unfortunately, there are $\binom{D}{s}$ subsets of size s. Obviously we would like to find a faster algorithm.

References

[1] B. K. Natarajan. Sparse approximate solutions to linear systems. *SIAM Journal of Computing*, 24(2):227–234, 1995.