UC DAVIS, BIG DATA (MAT280), SPRING 2016

Homework #2

Sam Fleischer

May 2, 2016

Problem 1 2

Problem 2 2

Problem 3 3

Problem 4 4

Problem 5 5

Problem 1

Let $i = \sqrt{-1}$ and set

$$A = \begin{bmatrix} i & 0 & -i \\ 0 & i & -i \end{bmatrix}.$$

Using the null space property, show that ℓ_1-minimization will recover any 1-sparse vector x, given $Ax = y$.

Proof. Given $x = [x_1, x_2, x_3]^T \in \mathbb{C}^3$, $Ax = i[x_1 - x_3, x_2 - x_3]^T = 0$ if $x_1 = x_2 = x_3$. Thus $\text{null} A = \text{span}\left([1,1,1]^T\right)$. Let $h \in \text{null} A$. Then $h = [a, a, a]^T$ for some $a \in \mathbb{C}$. Then choose $S_i = \{i\}$ for $i = 1, 2, 3$. Then

$$h_{S_1} = \begin{bmatrix} a \\ 0 \\ 0 \end{bmatrix}, \qquad h_{S_2} = \begin{bmatrix} 0 \\ a \\ 0 \end{bmatrix}, \qquad \text{and} \qquad h_{S_3} = \begin{bmatrix} 0 \\ 0 \\ a \end{bmatrix}$$

which gives

$$h_{S_1^C} = \begin{bmatrix} 0 \\ a \\ a \end{bmatrix}, \qquad h_{S_2^C} = \begin{bmatrix} a \\ 0 \\ a \end{bmatrix}, \qquad \text{and} \qquad h_{S_3^C} = \begin{bmatrix} a \\ a \\ 0 \end{bmatrix}$$

Clearly $\left\|h_{S_i}\right\|_1 = |a|$ and $\left\|h_{S_i^C}\right\|_1 = 2|a|$ for $i = 1, 2, 3$. Thus $\|h_S\|_1 \le \left\|h_{S^C}\right\|_1$ for all $h \in \text{null} A$ and all $S \subset \{1, 2, 3\}$ with $|S| = 1$. This shows the null space property holds and hence ℓ_1-minimization will recover any 1-sparse vector x. ☐

Problem 2

On the connection between (in)coherence parameter μ and restricted isometry constant δ_s: Show that $\delta_1 = 0$, $\delta_2 = \mu$, and $\delta_s \le (s-1)\mu$.

Proof. Let $A \in \mathbb{R}^{k \times d}$ where each column is normalized, i.e. $\left\|A_j\right\|_2 = 1$ for each $j = 1, 2, \ldots, d$. Let x be a 1-sparse vector of norm 1, that is, $|x_k| = 1$ for some $k \in \{1, 2, \ldots, d\}$ and $x_\ell = 0$ for $\ell \in \{1, 2, \ldots, d\} \setminus \{k\}$. Then $Ax = x_k A_k$, where A_k denotes the k^{th} column of A. Since each column is normalized,

$$\|Ax\|_2^2 = \left\|A_j\right\|_2^2 = 1 = \|x\|_2^2$$

Since this holds for all 1-sparse vector, $\delta_1 = 0$.
Now set x be a 2-sparse vector of norm 1, that is $\left|x_{k_1}\right|^2 + \left|x_{k_2}\right|^2 = 1$ for some $k_1, k_2 \in \{1, 2, \ldots, d\}$ and $x_\ell = 0$ for $\ell \in \{1, 2, \ldots, d\} \setminus \{x_{k_1}, x_{k_2}\}$. Then $Ax = x_{k_1} A_{k_1} + x_{k_2} A_{k_2}$. Thus,

$$\begin{aligned}
\|Ax\|_2^2 = \left\|x_{k_1} A_{k_1} + x_{k_2} A_{k_2}\right\|_2^2 &= \left\|x_{k_1} A_{k_1}\right\|_2^2 + 2\left\langle x_{k_1} A_{k_1}, x_{k_2} A_{k_2}\right\rangle_2 + \left\|x_{k_2} A_{k_2}\right\|_2^2 \\
&= \left|x_{k_1}\right|^2 \cancelto{1}{\left\|A_{k_1}\right\|_2^2} + 2 x_{k_1} x_{k_2} \left\langle A_{k_1}, A_{k_2}\right\rangle_2 + \left|x_{k_2}\right|^2 \cancelto{1}{\left\|A_{k_2}\right\|_2^2} \\
&= \cancelto{1}{\left|x_{k_1}\right|^2 + \left|x_{k_2}\right|^2} + 2 x_{k_1} x_{k_2} \left\langle A_{k_1}, A_{k_2}\right\rangle_2 \\
&\le 1 + \left\langle A_{k_1}, A_{k_2}\right\rangle_2
\end{aligned}$$

since $\displaystyle\max_{|a|^2 + |b|^2 = 1} [ab] = \frac{1}{2}$. Then since $\left|\left\langle A_{k_1}, A_{k_2}\right\rangle\right| < \mu$,

$$\left(1 - \mu\right)\|x\|_2^2 = 1 - \mu \le \|Ax\|_2^2 \le 1 + \mu = \left(1 + \mu\right)\|x\|_2^2.$$

Since this holds for all 2-sparse vectors, $\delta_2 \le \mu$. Since we can choose a 2-sparse vectors such that $x_{k_1} = x_{k_2} = \frac{\sqrt{2}}{2}$, then $\delta_2 = \mu$.
To show $\delta_s = (s-1)\mu$, first note, as our base case, that $\delta_2 = (2-1)\mu$. Then assume $\delta_s = (s-1)\mu$. It suffices to show $\delta_{s+1} = ((s+1)-1)\mu$, and then we will conclude, by induction that this holds for all integers s. Also note that the following calculations only make sense for $s \le d$. Let x be an $s+1$-sparse vector of norm 1, that is $\left|x_{k_1}\right|^2 + \cdots + \left|x_{k_s}\right|^2 + \left|x_{k_{s+1}}\right|^2 = 1$ for some $k_1, \dots, k_s, k_{s+1} \in \{1, 2, \dots, d\}$ and $x_\ell = 0$ for $\ell \in \{1, 2, \dots, d\} \setminus \{k_1, \dots, k_s, k_{s+1}\}$. Then $Ax = \sum_{i=1}^{s+1} x_{k_i} A_{k_i}$, and thus

$$\begin{aligned}
\|Ax\|_2^2 = \left\|\sum_{i=1}^{s+1} x_{k_i} A_{k_i}\right\|_2^2 &= \left\|\sum_{i=1}^{s} \left[x_{k_i} A_{k_i}\right] + x_{k_{s+1}} A_{k_{s+1}}\right\|_2^2 \\
&= \left\|\sum_{i=1}^{s} x_{k_i} A_{k_i}\right\|_2^2 + 2\left\langle \sum_{i=1}^{s} x_{k_i} A_{k_i}, x_{k_{s+1}} A_{k_{s+1}} \right\rangle_2 + \left\|x_{k_{s+1}} A_{k_{s+1}}\right\|_2^2 \\
&= \left\|\sum_{i=1}^{s} x_{k_i} A_{k_i}\right\|_2^2 + 2\sum_{i=1}^{s} \left(\left\langle x_{k_i} A_{k_i}, x_{k_{s+1}} A_{k_{s+1}} \right\rangle_2\right) + \left|x_{k_{s+1}}\right|^2 \cancelto{1}{\left\|A_{k_{s+1}}\right\|_2^2} \\
&= \left\|\sum_{i=1}^{s} x_{k_i} A_{k_i}\right\|_2^2 + 2\sum_{i=1}^{s} x_{k_i} x_{k_{s+1}} \left(\left\langle A_{k_i}, A_{k_{s+1}} \right\rangle_2\right) + \left|x_{k_{s+1}}\right|^2
\end{aligned}$$

Since $\sum_{i=1}^{s} x_{k_i} A_{k_i} = A\hat{x}$ for some s-sparse vector $\hat{x}$, then by the induction hypothesis,

$$\big(1-(s-1)\mu\big)\|\hat{x}\|_2^2 \le \|A\hat{x}\|_2^2 \le \big(1+(s-1)\mu\big)\|\hat{x}\|_2^2$$

Since $\|\hat{x}\|_2^2 \le \|x\|_2^2$, then $\|\hat{x}\| \le 1$, and thus

$$\big(1-(s-1)\mu\big) \le \left\|\sum_{i=1}^{s} x_{k_i} A_{k_i}\right\|_2^2 \le \big(1+(s-1)\mu\big)$$

Since $\max\limits_{\sum_{i=1}^{s+1}|a_i|^2=1} \left[\sum_{i=1}^{s+1} \left[a_i a_{s+1}\right]\right] = 1$ (which occurs when $a_i = \frac{\sqrt{s+1}}{s+1}$ for $i = 1, \dots, s+1$), and since $\left|\left\langle A_{k_i}, A_{k_{s+1}} \right\rangle_2\right| \le \mu$, then

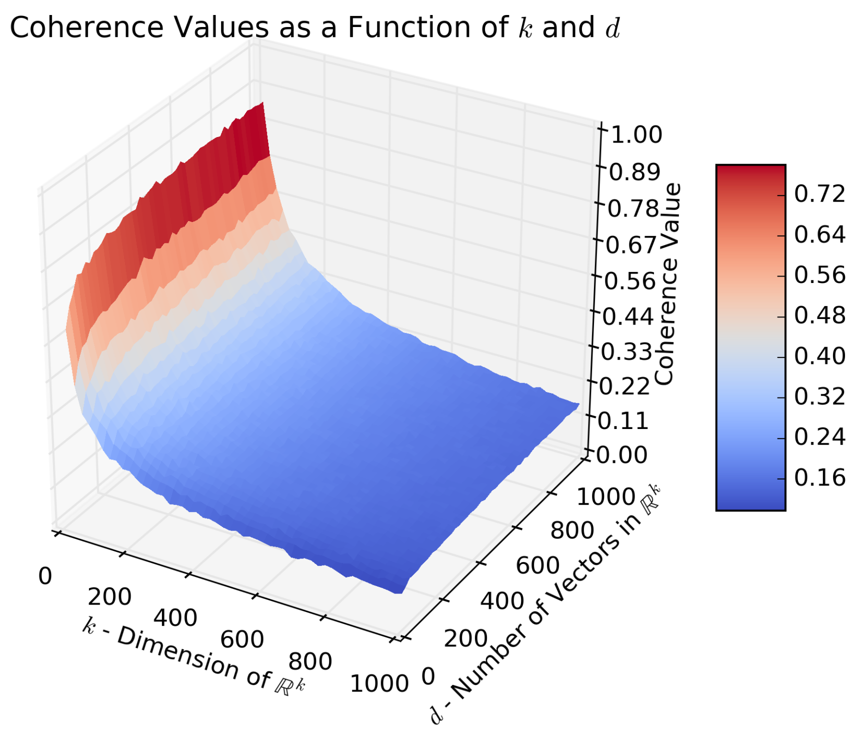
$$\big(1-s\mu\big)\|x\|_2^2 = 1 - s\mu = 1-(s-1)\mu - \mu \le \|Ax\|_2^2 \le 1 + (s-1)\mu + \mu = 1 + s\mu = \big(1+s\mu\big)\|x\|_2^2$$

Since this holds for arbitrary $(s+1)$-sparse vectors, then $\delta_{s+1} \le s\mu$. However, we can choose x such that $\|Ax\|_2^2 = \big(1+s\mu\big)\|x\|_2^2$ by choosing $x_{k_i} = \frac{\sqrt{s+1}}{s+1}$ for $i = 1, 2, \dots, s+1$. Then $\delta_{s+1} = s\mu$. By induction $\delta_s = (s-1)\mu$ for all integers s $(\le d)$. □

Problem 3

Let $A = \mathbb{R}^{k\times d}$ be a Gaussian random matrix. Given an estimate for the coherence μ of A.

Proof. The following surface was generated by taking the coherence of matrices for each value of k and d between 20 and 1000, in intervals of 20. In particular, for each k and d, the average coherence value over 10 Gaussian random matrices of size $k \times d$ was calculated. The figure shows that coherence, in general, increases as the dimension decreases, and as the number of vectors increases. This is intuitive since, as we saw from Homework 1 Problem 1, as the dimension decreases and/or the number of vectors increases, the probability that two vectors are close to orthogonal decreases. Equivalently, the probability that two vectors are close to parallel increases. If a matrix has two parallel columns, the coherence is 1. □



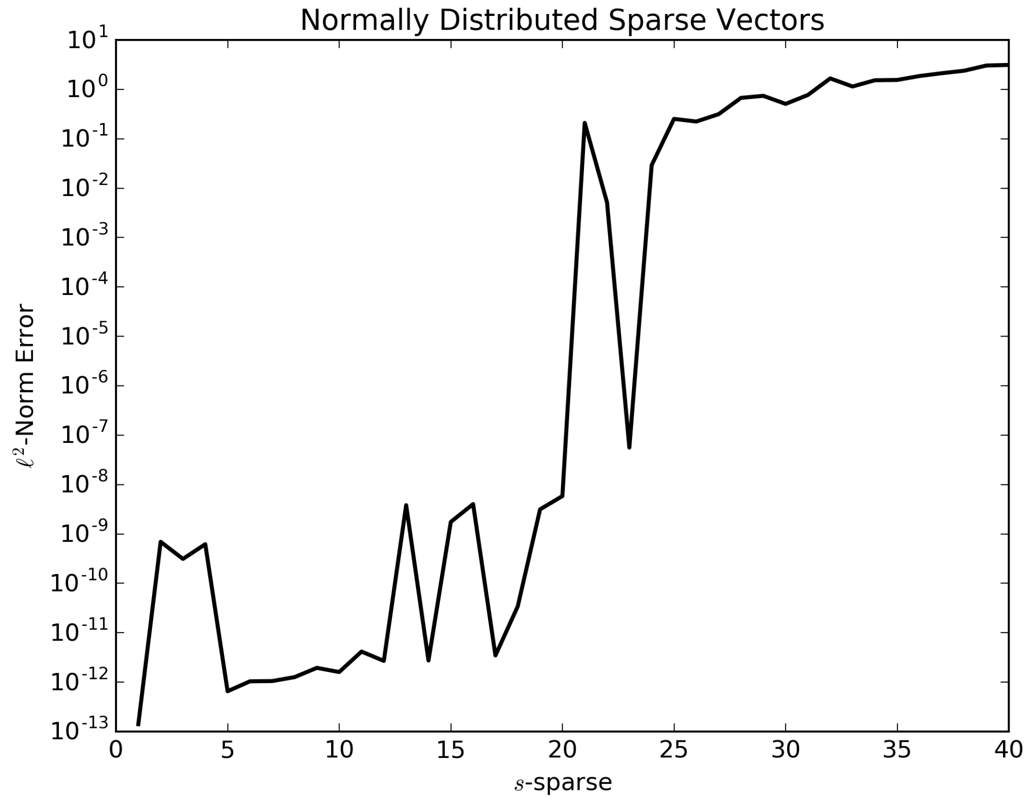
Problem 4

Consider $y = Ax$, where A is a 100×400 Gaussian random matrix and x is a s-sparse vector of length 400. The locations of the non-zero entries of x are chosen uniformly at random and the non-zero coefficients of x are normal-distributed. For $s = 1, 2, \ldots,$ solve

$$\min_z \|z\|_1 \qquad \text{subject to } Az = y,$$

(e.g. using the toolbox CVX). For each fixed s repeat the experiment 10 times. Create a graph plotting s versus the relative reconstruction error (averaged over the ten experiments for each s). Starting with which value of s approximately does ℓ_1-minimization fail to recover x?

Proof. The following graph shows the mean ℓ^2-norm errors of 10 experiments at each s for $s = 1, 2, \ldots, 40$. The non-zero entries of x are normally distributed around 0 with standard deviation of 1. In this experiment, this method failed for $s \geq 21$. □



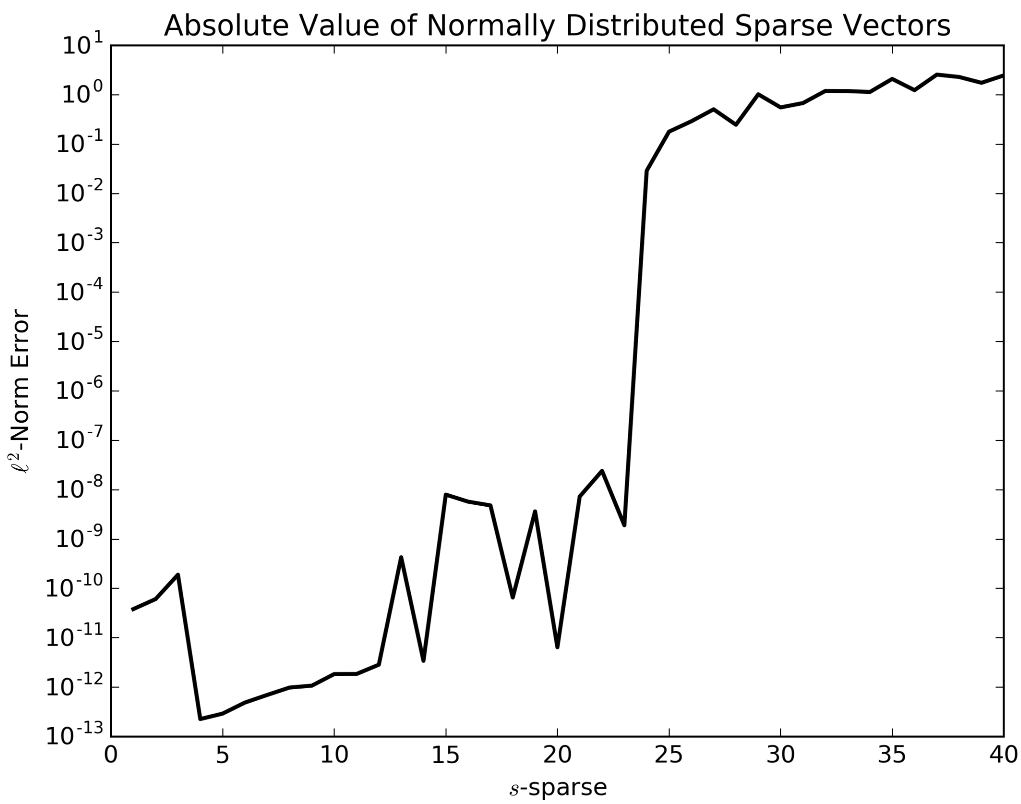
Problem 5

Same setup as in Problem 4, but now the non-zero entries of x are non-negative. Taking this information into account, we now solve

$$\min_z \|z\|_1 \qquad \text{subject to } Ax = y \text{ and } z \geq 0$$

(here, $z \geq 0$ is meant entrywise, i.e., for each k, $z_k \geq 0$). (The positivity constraint is easy to include in CVX). Repeat the simulations as described in Problem 4. Compare your findings to the results from your experiments of Problem 4 and try to quantify the difference regarding the range for s for which recovery is still possible in this case.

Proof. The following graph shows the mean ℓ^2-norm errors of 10 experiments at each s for $s = 1, 2, \ldots, 40$. The non-zero entries of x are the absolute value of a normal distribution around 0 with standard deviation of 1. In this experiment, this method failed for $s \geq 24$. $\square$


Absolute Value of Normally Distributed Sparse Vectors
ℓ^2-Norm Error
s-sparse
10^{1}
10^{0}
10^{-1}
10^{-2}
10^{-3}
10^{-4}
10^{-5}
10^{-6}
10^{-7}
10^{-8}
10^{-9}
10^{-10}
10^{-11}
10^{-12}
10^{-13}
0
5
10
15
20
25
30
35
40