Prof. Ming Gu, 861 Evans, tel: 2-3145
Email: mgu@math.berkeley.edu
http://www.math.berkeley.edu/∼mgu/MA170Spring2020

Math170: Mathematical Methods for Optimization
Final Project

In this project we solve the 1−*norm regression problem*:

$$\min_{\mathbf{x}} \; \|A\,\mathbf{x}-\mathbf{b}\|_1 \,. \tag{1}$$

In this problem, the matrix $A \in \mathcal{R}^{m\times n}$ and vector $b \in \mathcal{R}^m$ are given, with $m > n$. Our work is to find the optimal solution vector $\mathbf{x} \in \mathcal{R}^n$ that minimizes the 1-norm $\|A\,\mathbf{x}-\mathbf{b}\|_1$. For any given vector $\mathbf{y} = (y_1, \cdots, y_m)^\top$, its 1-norm is defined as

$$\|\mathbf{y}\|_1 = \sum_{j=1}^{m} |y_j|.$$

The problem (1) can be recast as a linear program as

$$\begin{array}{ll} \min\limits_{\mathbf{x},\,\mathbf{u}} & \mathbf{e}^\top \mathbf{u} \\ s.t. & -\mathbf{u} \le A\,\mathbf{x}-\mathbf{b} \le \mathbf{u}, \end{array}$$

where $\mathbf{e} = (1, \cdots, 1)^\top \in \mathcal{R}^m$. This linear program, in turn, has a dual

$$\begin{array}{ll} \max\limits_{\mathbf{u}} & -\mathbf{b}^\top \mathbf{y} \\ s.t. & A^\top \mathbf{y} = 0, \\ & |\mathbf{y}| \le \mathbf{e}. \end{array}$$

While problem (1) is equivalent to a linear program, it is typically much more efficient to solve it with a specialized simplex-type method. Below we discuss such a method, under the **Non-degeneracy assumptions:**

- The matrix $A\,(\mathcal{B},:)$ is invertible for every index set $\mathcal{B} \subset \{1, \cdots, m\}$ with exactly n indexes.
- There does not exist an index set $\mathcal{B}$ with more than n indexes such that $A(\mathcal{B},:)\,\mathbf{x} = \mathbf{b}\,(\mathcal{B})$.

Under these assumptions, there exists a unique index set $\mathcal{B}^{\mathbf{opt}} \subset \{1, \cdots, m\}$ with n indexes such that $\mathbf{x}^{\mathbf{opt}} = A\left(\mathcal{B}^{\mathbf{opt}}, :\right)^{-1} \mathbf{b}\left(\mathcal{B}^{\mathbf{opt}}\right)$ solves the problem (1).

To describe an algorithm for solving problem (1), we start with any given index set $\mathcal{B} \subset \{1, \cdots, m\}$ with n indexes. Let $\overline{\mathcal{B}} = \{1, \cdots, m\} \backslash \mathcal{B}$ be the complement set of $\mathcal{B}$. Choosing $\mathbf{x} = A(\mathcal{B}, :)^{-1} \mathbf{b}(\mathcal{B})$, we reach objective value $\left\| A\left(\overline{\mathcal{B}}, :\right) \mathbf{x} - \mathbf{b}\left(\overline{\mathcal{B}}\right) \right\|_1$ in problem (1). Below we exlpain a procedure to update $\mathcal{B}$ in a fashion similar to the simplex method to reach a lower objective value in problem (1). Just like simplex method, we then repeat this procedure until we eventually reach the optimal index set $\mathcal{B}^{\mathbf{opt}}$ and therefore the optimal solution $\mathbf{x}^{\mathbf{opt}}$.

Define

$$\mathbf{x} = A(\mathcal{B}, :)^{-1} \mathbf{b}(\mathcal{B}) \quad \text{and} \quad \mathbf{h} = A\,\mathbf{x} - \mathbf{b}.$$

It follows that $\mathbf{h}\left(\overline{\mathcal{B}}\right) = A\left(\overline{\mathcal{B}}, :\right) \mathbf{x} - \mathbf{b}\left(\overline{\mathcal{B}}\right)$. By the non-degeneracy assumptions none of the components in $\mathbf{h}\left(\overline{\mathcal{B}}, :\right)$ is exactly zero. Now define $\mathbf{y} \in \mathcal{R}^m$ as

$$\begin{aligned} \mathbf{y}\left(\overline{\mathcal{B}}\right) &= \mathbf{sign}\left(\mathbf{h}\left(\overline{\mathcal{B}}\right)\right), \\ \mathbf{y}(\mathcal{B}) &= -A(\mathcal{B}, :)^{-\top} A\left(\overline{\mathcal{B}}, :\right)^{\top} \mathbf{y}\left(\overline{\mathcal{B}}\right), \end{aligned}$$

where **sign** is the sign function, so $\mathbf{y}\left(\overline{\mathcal{B}}\right)$ contains the signs of the $\mathbf{h}\left(\overline{\mathcal{B}}\right)$ components. The components of $|\mathbf{y}\left(\overline{\mathcal{B}}\right)|$ are all 1.

If all components of $|\mathbf{y}(\mathcal{B})|$ are less than or equal to 1, then $\mathbf{y}$ is a feasible solution to the dual problem, and by the equilibrium conditions $\mathbf{x}$ and $\mathbf{y}$ are optimal solutions to problem (1) and the dual, respectively.

If, on the other hand, some components of $|\mathbf{y}(\mathcal{B})|$ are greater than 1, then $\mathbf{y}$ is not dual feasible, and now we proceed to reduce the objective value in problem (1) as follows.

Choose an index $j_s \in \mathcal{B}$ such that $|y_s| > 1$ (j_s is the s-th entry in $\mathcal{B}$.) Define

$$\begin{aligned} \mathbf{t}\left(\overline{\mathcal{B}}\right) &= -\left(\mathbf{sign}(y_s)\right)\left(\mathbf{y}\left(\overline{\mathcal{B}}\right)\right) .* \left(A\left(\overline{\mathcal{B}}, :\right) A(\mathcal{B}, :)^{-1} \mathbf{e}_s\right), \\ r &= \underset{j}{\operatorname{argmin}}\{\frac{|h_j|}{t_j}, \quad | \quad j \in \overline{\mathcal{B}} \quad \text{and} \quad t_j > 0, \} \end{aligned}$$

where $\mathbf{e}_s$ is the vector which is 0 everywhere except the s-th entry, which is 1. In other words, $A(\mathcal{B}, :)^{-1} \mathbf{e}_s$ is the s-th column of $A(\mathcal{B}, :)^{-1}$. Then the new index set is

$$\widehat{\mathcal{B}} = \mathcal{B} \backslash \{j_s\} \cup \{r\}.$$

The new solution $\widehat{\mathbf{x}} = A\left(\widehat{\mathcal{B}}, :\right)^{-1} \mathbf{b}\left(\widehat{\mathcal{B}}\right)$ will lead to a reduced objective value in problem (1). Notice that j_s is the s-th entry in $\mathcal{B}$, while r refers to r-th row of A that is currently indexed in $\overline{\mathcal{B}}$. You need to test your code carefully for the correct indexing in $\mathcal{B}$ and $\overline{\mathcal{B}}$.

Our job in this project is to develop the above idea into a simplex-type algorithm for solving problem (1) under the non-degeneracy assumptions. You do not need to use the Simplex tableau. You just need to following the simplex method (without the tableau) to find the optimal solution.

Note that the Phase I calculations for this problem consists of picking up any initial index set $\mathcal{B}$ with n indexes and computing $M^{-1} = A(\mathcal{B}, :)^{-1}$ and $A(\mathcal{B}, :)^{-1} \mathbf{b}(\mathcal{B})$.

You should turn in a .m file OneNormLPxxx.m which contains a matlab function of the form

```
function [data, info] = OneNormLP(A,b)
```

to solve a given 1−norm regression problem in (1). Here xxx is your student id. On output (case sensitive):

- If `info.run = Failure`, then
  - `info.msg`: Explain where and how the failure occured
    (failure due to arithmetic exceptions or degeneracy)
- If `info.run = Success`
  - `data.Opt` = the optimal objective value
  - `data.loop` = # of iterations to solve for optimal solution.
  - `data.Initx, data.Optx` = initial and optimal solutions as column vectors.
  - `data.Opty` = optimal dual solution as column vector.

Along with the signature page, due 23:59PM, Thursday, May 14, 2020 on `gradescope`.