Mathematical Methods for Optimization

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

$$\begin{aligned} \min_{\mathbf{x},\,\mathbf{u},\,\mathbf{v}} \quad & \mathbf{e}^\top\,(\mathbf{u}+\mathbf{v}) \\ s.t. \quad & A\,\mathbf{x} - \mathbf{b} = \mathbf{u} - \mathbf{v}, \quad \mathbf{u},\,\mathbf{v} \geq \mathbf{0} \end{aligned}$$

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

$$\begin{aligned} \max_{\mathbf{y}} \quad & \mathbf{b}^\top\,\mathbf{y} \\ s.t. \quad & A^\top\,\mathbf{y} = 0 \\ & -\mathbf{e} \leq \mathbf{y} \leq \mathbf{e} \end{aligned}$$

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 $\mathbf{x} = A\left(\mathcal{B}, :\right)^{-1} \mathbf{b}\left(\mathcal{B}\right)$, 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

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

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

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

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 $|\mathbf{y}\left(\mathcal{B}\right)|$ 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 $|\mathbf{y}\left(\mathcal{B}\right)|$ 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

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

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

$$\mathcal{B}^{\mathbf{new}} = \mathcal{B} \backslash \{j_s\} \cup \{r\}.$$

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 $\mathcal{B}^{\mathbf{new}}$.

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.

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

You do not need to use the Simplex tableau, but you will need to use the Sherman-Morrison formula to update the inverse matrix M.

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.obj` = the optimal objective value
  - `data.x` = optimal solution as column vector.
  - `data.loop` = # of iterations to solve for optimal solution.

Due 23:59PM, Monday, Nov. 22, 2021 on `gradescope`.