Project Description

ESI6417 – Spring 2019

General Description (For Teams of Two)

This is a take-home project for groups of **two**. There are two problems in this project. Please solve all two of them and provide a report that contains your derivations, thoughts, and descriptions of your experiments, and answers. The report should be written in either WORD or LaTeX. For the latter, please only submit the generated PDF file.

Due date: April 19, 2019 before 11:59pm (before the reading days start).

Deliverables: Each team will submit the following items

- One report.
- Two ".mat" files that contain solutions to the problems. The file names should follow "Group_<Group Number>_Solution_<Problem Number>.mat".
- Programming codes for all two problems.

Problem 1: Intensity Modulated Radiation Treatment Planning

Intensity modulated radiation treatment (IMRT) is a common treatment for cancer. The IMRT uses linear accelerators to deliver radiation doses to a tumor while minimizing the doses to the surrounding normal tissues. The quality of the treatment depends highly on the plan of the treatment.

In this problem, we will consider applying linear programming for IMRT planning in finding the optimal configurations of radiation intensities from different radiation sources. For each source i of radiation, denote by x_i the corresponding radiation intensity. Let $\mathbf{x} := (x_i)$.

Let $\mathcal{O}$ be the index set of tissues (the tumor is also considered as a tissue in this problem). For simplicity, we discrete the continuous volume of a tissue into finitely many grids. Denote by d_v the dose level on the v-th grid. Let $\mathbf{d}_\sigma$ be the vector of doses on the grids that belong to the σ-th tissue ($\sigma \in \mathcal{O}$); namely, if $\mathcal{V}_\sigma$ denotes the index set for all the voxels of the σ-th tissue, then $\mathbf{d}_\sigma := (d_v : \sigma \in \mathcal{O})$.

Assume that there is a linear relationship between $\mathbf{d}_\sigma$ and $\mathbf{x}$:

$$\mathbf{A}_\sigma \mathbf{x} = \mathbf{d}_\sigma, \quad \sigma \in \mathcal{O}. \tag{1}$$

For each organ σ, assume that a prescribed dose level $d_\sigma^{pr} \in \Re$ is given and that a uniform upper bound on the dose levels U is known. Please solve the following optimization problem:

$$\begin{aligned}
\min_{\mathbf{x},\,(\mathbf{d}_\sigma)} \ & \sum_{\sigma\in\mathcal{O}} w_\sigma |d_\sigma^{pr} \cdot \mathbf{1} - \mathbf{d}_\sigma| \\
s.t.\ & \mathbf{A}_\sigma \mathbf{x} = \mathbf{d}_\sigma, \quad \sigma \in \mathcal{O} \\
& \mathbf{d}_\sigma \le U, \quad \sigma \in \mathcal{O} \\
& \mathbf{x} \ge 0.
\end{aligned} \tag{2}$$

1. Reformulate (2) into a linear programming problem.

2. Load problem data from "Problem_1.mat". The correspondence between parameters in (2) and the saved data are as below:
   - $\mathbf{A}_1$, $\mathbf{A}_2$, ..., are saved as A{1}, A{2},..., respectively.
   - d_1^{pr}, d_2^{pr}, ..., are saved as dpr{1}, dpr{2},..., respectively.
   - w_1, w_2, ..., are saved as w{1}, w{2},..., respectively.

   Implement the simplex method to solve (2) using Matlab or Python.

Problem 2: Linear Optimization in Dantzig Selector

In high-dimensional learning, the following problem (a.k.a., the Dantzig selector) is of great interest:

$$
\begin{aligned}
&\min_{\mathbf{x}} \ |\mathbf{x}| \\
&s.t. \ \ \frac{1}{m}\|\mathbf{A}^\top \mathbf{A}\mathbf{x} - \mathbf{A}^\top \mathbf{b}\|_\infty \leq \epsilon,
\end{aligned}
\tag{3}
$$

where $\mathbf{A} \in \Re^{m\times n}$ is the design matrix of a linear regression problem, $\mathbf{b} \in \Re^m$ is the vector of responses, n is the sample size, and $\epsilon > 0$ is a tuning parameter. Linear regression is used commonly as a data-driven estimation scheme. It works very well when sample size is large (and larger than the number of problem dimensions). However, in certain modern applications of linear regression, the sample size can be far less than the number of dimensions. In such a scenario, the Dantzig selector can substantially enhance the estimation quality under the assumption that the target of estimation is sparse. In our problem, we assume $N = 10 < n = 18$.

1. Reformulate (3) into a linear programming problem.
2. Load problem data from "Problem_2.mat". Implement the simplex method to solve (3) using Matlab or Python.
3. Import data from "Problem_2.mat', and solve the linear programming reformulation of (3) using GAMS.
4. Compare the three solution approaches in terms of solution quality and computational time.
5. Analyze how sensitive the model is when ϵ is perturbed.