EE 127 2/6/14
L. El Ghaoui

Homework Assignment #2

Due date: 2/20/14, in class.

Exercise 1 (Least norm estimation on traffic flow networks)
You want to estimate the traffic (in San Francisco for example, but we'll start with a smaller example). You know the road network as well as the historical average of flows on each road segment.

1. We call q_i the flow of vehicles on each road segment $i \in I$. Write down the linear equation that corresponds to the conservation of vehicles at each intersection $j \in J$. *Hint:* think about how you might represent the road network in terms of matrices, vectors, etc.

2. The goal of the estimation is to estimate the traffic flow on each of the road segment. The flow estimates should satisfy the conservation of vehicles exactly at each intersection. Among the solutions that satisfy this constraint, we are searching for the estimate that is the closest to the historical average, $\overline{q}$, in the l_2-norm sense. The vector $\overline{q}$ has size I and the i-th element represent the average for the road segment i. Pose the optimization problem.

3. Explain how to solve this problem mathematically. Detail your answer (do not only give a formula but explain where it comes from).

4. Formulate the problem for the small example of Figure 1 and solve it using the historical average given in Table 1. What is the flow that you estimate on road segments 1, 3, 6, 15 and 22?

5. Now, assume that besides the historical averages, you are also given some flow measurements on some of the road segments of the network. You assume that these flow measurements are correct and want your estimate of the flow to match these measurements perfectly (besides matching the conservation of vehicles of course). The right column of Table 1 lists the road segments for which we have such flow measurements. Do you estimate a different flow on some of the links? Give the difference in flow you estimate for road segments 1,3, 6, 15 and 22. Also check that you estimate gives you the measured flow on the road segments for which you have measured the flow.

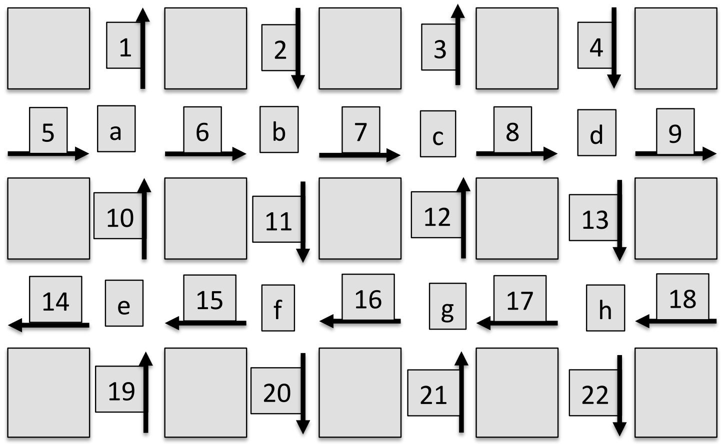

Figure 1: Example of traffic estimation problem. The intersections are labeled a to h. The road segments are labeled 1 to 22. The arrows indicate the direction of traffic.

segment	average	measured
1	2047.6	2028
2	2046.0	2008
3	2002.6	2035
4	2036.9	
5	2013.5	2019
6	2021.1	
7	2027.4	
8	2047.1	
9	2020.9	2044
10	2049.2	
11	2015.1	
12	2035.1	
13	2033.3	
14	2027.0	2043
15	2034.9	
16	2033.3	
17	2008.9	
18	2006.4	
19	2050.0	2030
20	2008.6	2025
21	2001.6	
22	2028.1	2045

Table 1: Table of flows: historical averages $\overline{q}$ (center column), and some measured flows (right column).

Exercise 2 (Exploiting structure in linear equations) Consider the linear equation in $x \in \mathbb{R}^n$

$$Ax = y$$

where $A \in \mathbb{R}^{m,n}$, $y \in \mathbb{R}^m$. Answer the following questions to the best of your knowledge.

1. The time required to solve the general system depends on the sizes m, n and the entries of A. Provide a rough estimate of that time as a function of m, n only. You may assume that m, n are of the same order.
2. Assume now that $m = n$, $A = D + uv^\top$, where D is diagonal, invertible, and $u, v \in \mathbb{R}^n$. How would you exploit this structure to solve the above linear system, and what is a rough estimate of the complexity of your algorithm?
3. What if A is upper-triangular?

Exercise 3 (Formulating problems as LPs or QPs)
Formulate the problem

$$p_j^* \doteq \min_x \ f_j(x),$$

for different functions f_j, $j = 1, \ldots, 5$, with values given in Table 2, as QPs or LPs, or, if you cannot, explain why. In our formulations, we always use $x \in \mathbb{R}^n$ as the variable, and assume that $A \in \mathbb{R}^{m,n}$, $y \in \mathbb{R}^m$, and $k \in \{1, \ldots, m\}$ are given. If you obtain an LP or QP formulation, make sure to put the problem in standard form, stating precisely what the variables, objective and constraints are.

$$\begin{aligned}
f_1(x) &= \|Ax - y\|_\infty + \|x\|_1 \\
f_2(x) &= \|Ax - y\|_2^2 + \|x\|_1 \\
f_3(x) &= \|Ax - y\|_2^2 - \|x\|_1 \\
f_4(x) &= \|Ax - y\|_2^2 + \|x\|_1^2 \\
f_5(x) &= \textstyle\sum_{i=1}^k |Ax - y|_{[i]} + \|x\|_2^2
\end{aligned}$$

Table 2: Table of the values of different functions f. $|z|_{[i]}$ denotes the element in a vector z that has the i-th largest magnitude.

Exercise 4 (A slalom problem) A two-dimensional skier must slalom down a slope, by going through n parallel gates of known position (x_i, y_i), and of width c_i, $i = 1, \ldots, n$. The initial position (x_0, y_0) is given, as well as the final one, (x_{n+1}, y_{n+1}). Here, the x-axis represents the direction down the slope, from left to right.

1. Find the path that minimizes the total length of the path. Your answer should come in the form of an optimization problem.

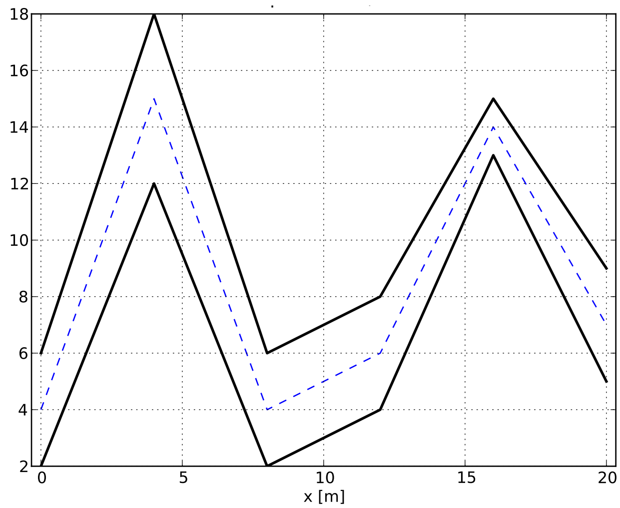

Figure 2: Slalom problem with $n = 5$ obstacles. "Uphill" (resp. "downhill") is on the left (resp. right) side. Middle path is dashed, initial and final positions are not shown.

i	x_i	y_i	c_i
0	0	4	N/A
1	4	5	3
2	8	4	2
3	12	6	2
4	16	5	1
5	20	7	2
6	24	4	N/A

Table 3: Problem data for Exercise 4.

2. Try solving the problem numerically, with the data given in Table 3.

Exercise 5 (LS with uncertain A matrix) Consider a linear leastsquares problem where the matrix involved is random. Precisely, the residual vector is of the form $A(\delta)x - b$, where the $m \times n$ A matrix is affected by stochastic uncertainty. In particular, assume that

$$A(\delta) = A_0 + \sum_{i=1}^{p} A_i \delta_i,$$

where the random variables δ_i are independently distributed with zero mean and variance σ_i^2, $i = 1, \ldots, p$. The standard least-squares objective function $\|A(\delta)x - b\|_2^2$ is now random, since it depends on δ. We seek to determine x such that the expected value (with respect to the random variable δ) of $\|A(\delta)x - b\|_2^2$ is minimized. Is such a problem convex? If yes, to which class does it belong to (LP, LS, QP, etc.)?