ESI 6420 : Fundamentals of Mathematical Programming

Jean-Philippe P. Richard

Fall 2015

Homework 2, due in class, Monday October 5^{th} 2015

Preliminaries

Throughout the semester, I would like you to learn how to use latex (which you will find helpful in the future). In particular, for homework i, I would like you to latex at least $\lfloor \frac{i}{2} \rfloor$ of your answers. For an introduction to latex, refer to `http://www.ctan.org/tex-archive/info/gentle/gentle.pdf`.

If you received some help to obtain the solution of a problem, you should acknowledge the source of help you received. In particular, for each question, I would like you to cite, if applicable, any book (other than the textbook) you consulted, any website you searched, or any individual you cooperated with. This information will not be used to adjust your homework score provided that help is limited to a reasonable portion of the homework.

Finally, I would like you to candidly assess the number of hours it took you to complete the homework.

Problem 1: Sparse linear separation, JPR

We are given two families of vectors $x^{(1)}, \ldots, x^{(m_1)}$ and $y^{(1)}, \ldots, y^{(m_2)}$ in $\mathbb{R}^n$, but we are allowed to remove up to a total of k of them. We are trying to determine whether the remaining points can be separated by a hyperplane, with the additional requirement that no more than p variable coefficients of the hyperplane are nonzero. Among such separating hyperplanes, we seek one for which the margin is the largest possible. This problem can be viewed as an extension of the basic hard-margin SVM model, in which we seek a hyperplane that does not depend on too many variables (thereby allowing prediction from few data features) in situations where an exact separation might not be possible.

1. Formulate this problem as a mathematical program.
2. Solve this problem with GAMS using the data provided in the file "hw2pb1.gms" for all combination of values $k = 4, 6, 8, 10, 12, 14$ and $p = 3, 4, 5, 6$. *(Note: you probably will need to use NEOS to solve this problem)*
3. Let $\mu(k, p)$ be the function representing the margin as a function of k and p. Represent $\mu(.)$ on a picture for the above values of k and p.

Problem 2: Basic concepts in linear algebra, P.

Prove the following results:

1. Let A and B be matrices of size $m \times n$ and $k \times n$, respectively, such that if $AX = 0$ for a certain column X, then $BX = 0$. Prove that $B = CA$, where C is a matrix of size $k \times m$.
2. Let $v_1, \ldots, v_m$ be vectors in $\mathbb{R}^n$. Prove that, if $m \geq n + 2$, then there exists scalars $\alpha_1, \ldots, \alpha_m$, not all equal to zero, such that $\sum_{i=1}^m \alpha_i v_i = 0$ and $\sum_{i=1}^m \alpha_i = 0$.
3. Let $A : U \mapsto V$ and $B : V \mapsto W$ be linear maps. Show that

$$\dim(\mathrm{R}[A] \cap \ker[B]) = \dim(\mathrm{R}[A]) - \dim(\mathrm{R}[BA]) = \dim(\ker[BA]) - \dim(\ker[A]).$$

Then use this result to prove that, if D is a linear operator

$$\dim(\ker[D^{n+1}]) = \dim(\ker[D]) + \sum_{j=1}^{n} \dim(\mathrm{R}[D^k] \cap \ker[D]).$$

4. Let A_1 and A_2 be matrices of same dimensions. Let V_1 and V_2 be the spaces spanned by the rows of A_1 and A_2, respectively. Let W_1 and W_2 be the spaces spanned by the columns of A_1 and A_2, respectively. Prove that the following conditions are equivalent:

   (a) $\operatorname{rank}(A_1 + A_2) = \operatorname{rank} A_1 + \operatorname{rank} A_2$;

   (b) $V_1 \cap V_2 = 0$;

   (c) $W_1 \cap W_2 = 0$.

   Use this result to show that if A and B are matrices of the same size and $B^\intercal A = 0$, then $\operatorname{rank}(A+B) = \operatorname{rank}(A) + \operatorname{rank}(B)$.

5. Let V be a n-dimensional real vector space. Let U and W be m-dimensional subspaces of V such that $u \perp W$ for $u \in U\backslash\{0\}$. Prove that $w \perp U$ for some $w \in W\backslash\{0\}$.

Problem 3: About norms, JPR

Let $M_n(\mathbb{R})$ be the real vector space of square $n \times n$ real matrices. A matrix norm $\langle\langle . \rangle\rangle$ on $M_n(\mathbb{R})$ is a norm on $M_n(\mathbb{R})$ with the additional property that

$$\langle\langle AB \rangle\rangle \leq \langle\langle A \rangle\rangle \langle\langle B \rangle\rangle$$

for all matrices A and B in $M_n(\mathbb{R})$.

1. Show that, for the identity matrix I, $\langle\langle I \rangle\rangle \geq 1$ for all matrix norms.

2. Show that the Frobenius norm defined as $\langle\langle A \rangle\rangle_F = \sqrt{trace(A^\intercal A)} = \sqrt{\sum_{i=1}^n \sum_{j=1}^n |a_{ij}|^2}$ is indeed a matrix norm

3. Given a vector norm $||.||$ on $\mathbb{R}^n$, we define the function $\langle\langle . \rangle\rangle$ on $M_n(\mathbb{R})$ as

$$\langle\langle A \rangle\rangle = \sup_{x \in \mathbb{R}^n \backslash \{0\}} \frac{||Ax||}{||x||}.$$

   [It can be shown that this supremum is finite.]

   (a) Show that $\langle\langle A \rangle\rangle = \sup_{x \in \mathbb{R}^n, ||x||=1} \frac{||Ax||}{||x||}$.

   (b) Show that $\langle\langle . \rangle\rangle$ is a matrix norm on $M_n(\mathbb{R})$. From here on, we refer to $\langle\langle . \rangle\rangle$ as the subordinate matrix norm to vector norm $||.||$.

   (c) Show that $\langle\langle A \rangle\rangle = \inf\{\lambda \in \mathbb{R} \,|\, ||Ax|| \leq \lambda ||x||, \forall x \in \mathbb{R}^n\}$.

   (d) Show that for a subordinate norm on $M_n(\mathbb{R})$, then $\langle\langle I \rangle\rangle = 1$, where I is the identity matrix.

   (e) Show that the subordinate norm arising from $||.||_2$ can be computed as $\sqrt{\lambda_{\max}(A^\intercal A)}$ where $\lambda_{\max}(.)$ is used to denote the largest eigenvalue of a matrix.

   (f) Show that the Frobenius norm is not a subordinate matrix norm.

Problem 4: Laplacians, C. & E.G. 4.5, pg 120

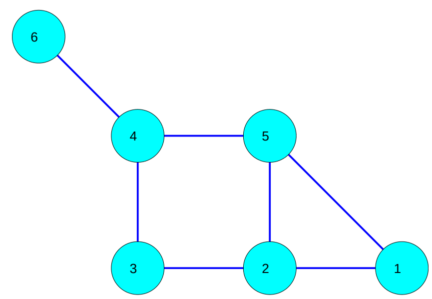

Figure 1: Example graph for Problem #4

We are given a graph as a set of vertices in $V = \{1, \ldots, n\}$, with an edge joining any pair of vertices in a set $E \subseteq V \times V$. We assume that the graph is undirected (without arrows), meaning that $(i, j) \in E$ implies $(j, i) \in E$. We define the Laplacian matrix by

$$L_{ij} = \begin{cases} -1 & \text{if } (i,j) \in E \\ d(i) & \text{if } i = j \\ 0 & \text{otherwise.} \end{cases}$$

Here, $d(i)$ is the number of edges adjacent to vertex i. For example, $d(4) = 3$, and $d(6) = 1$ for the graph of Figure 1.

1. Form the Laplacian for the graph shown in Figure 1.
2. Turning to a generic graph, show that the Laplacian L is symmetric.
3. Show that L is positive-semidefinite, proving the following identity, valid for any $u \in \mathbb{R}^n$:

$$u^\intercal L u = q(u) = \frac{1}{2} \sum_{(i,j) \in E} (u_i - u_j)^2.$$

4. Show that 0 is always an eigenvalue of L, and exhibit an eigenvector. *(Hint: consider a matrix square-root of L.)*
5. The graph is said to be connected if there is a path joining any pair of vertices. Show that if the graph is connected, then the zero eigenvalue is simple, that is, the dimension of the nullspace of L is 1. *(Hint: prove that if $u^\intercal L u = 0$, then $u_i = u_j$ for every pair $(i, j) \in E$.)*

Problem 5: About singular values, JPR

Consider a matrix $A \in \mathbb{R}^{m \times n}$ with $m \geq n$,

1. Show that $A^\intercal A$ is symmetric and positive semidefinite.
2. Let v_j be a basis of normed orthogonal eigenvectors of $A^\intercal A$. Denote by λ_j the eigenvalue associated with eigenvector v_j, and assume that $\lambda_1 \geq \ldots \geq \lambda_n \geq 0$. Define $\sigma_j = \sqrt{\lambda_j}$ for $j = 1, \ldots, n$. Further, for $j = 1, \ldots, n$ in this order, define

$$u_j = \begin{cases} \frac{1}{\sigma_j} A v_j & \text{if} \quad \sigma_j \neq 0 \\ w_j & \text{if} \quad \sigma_j = 0, \end{cases}$$

where w_j is any unit vector such that $w_j \in span\{u_1, \ldots, u_{j-1}\}^\perp$. Argue that $u_j^\intercal u_k = 0$ for $1 \leq j < k \leq n$ and that $||u_j|| = 1$ for $j = 1, \ldots, n$.

3. Show that there exists $U \in \mathbb{R}^{m \times n}$ and $V \in \mathbb{R}^{n \times n}$ such that $A = U\Sigma V^\intercal$ where (i) $VV^\intercal = I$, (ii) $V^\intercal V = I$, (iii) $U^\intercal U = I$, and Σ is a diagonal matrix whose diagonal entries are σ_i for $i = 1, \dots, n$. [Note that, in general, the relation $UU^\intercal = I$ does not hold.] This matrix factorization is known as the *reduced singular value decomposition*.

4. Define $r = \max\{i \in \{1, \dots, n\} \,|\, \sigma_i \neq 0\}$. Argue that

   (a) $\{u_1, \dots, u_r\}$ is a basis of the column space of A.

   (b) $\{u_{r+1}, \dots, u_m\}$ is a basis of the left-null space of A.

   (c) $\{v_1, \dots, v_r\}$ is a basis of the row space of A.

   (d) $\{v_{r+1}, \dots, v_n\}$ is a basis of the null space of A.

5. Prove that, for $1 \leq k \leq rank(A)$, then

$$\min_{X \in \mathbb{R}^{m \times n} \,|\, rank(X) = k} ||A - X||_2 = \sigma_{k+1},$$

   where $||.||_2$ is the subordinate matrix norm arising from vector norm $||.||_2$. Also show that the minimum is attained by

$$A_k = \sum_{j=1}^{k} \sigma_j u_j v_j^\intercal.$$

   In the above expressions, vectors u and v relate to the reduced singular value decomposition $A = U\Sigma V^\intercal$ of matrix A. In other words, the best rank-k approximation of A is obtained from the vectors associated with the k largest singular values of A. To prove this result, use the following steps:

   (a) Let $X \in \mathbb{R}^{m \times n}$ be any rank-k matrix. Argue that $ker(X) \cap span < v_1, \dots, v_{k+1} >$ contains a normed vector z.

   (b) Use z to show that $||A - X||_2 \geq \sigma_{k+1}$.

   (c) Argue that $||A - A_k||_2 = \sigma_{k+1}$.

6. The above result has many applications, including the compression of digital images. In this application, we are given a $m \times n$ matrix A with integer entries between 0 and 255. These integers represent the shades of each pixel of the image. If m and n are large, storing the image might require a large amount of space. To reduce this amount, one could store an optimal rank-k approximation of the matrix A (by storing the associated vectors u_j, v_j and singular values σ_j for $j = 1, \dots, k$) for a sufficiently small value of k. For the images displayed in Figure 2 and available on Canvas under the name "durer.jpg" and "escher.jpg", use MATLAB to obtain a reduced singular value decomposition of A, and display an optimal rank-5, rank-15 and rank-30 approximations of A.

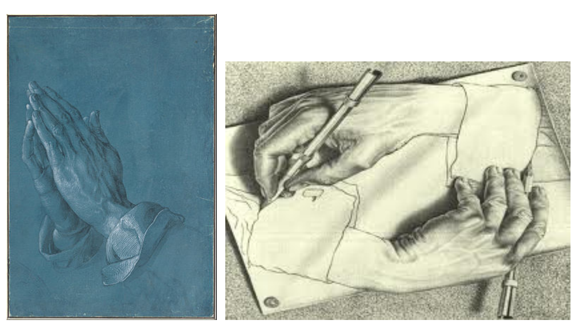

Figure 2: "Praying Hands" (Durer) and "Drawing Hands" (Escher)

Implementation suggestions:

1. To read a jpg picture into matlab, you may use the commands:
```
I = rgb2gray(imread('durer.jpg'));
I2 = double(I);
image(I2);
colormap(gray(256));
```
where the first line reads the picture in, the second line converts the entries of the matrix into double, the third line displays the image, and the fourth uses grayscale in the plot.

2. To compute a reduced singular value decomposition of a matrix A, use the MATLAB command:
```
[Uhat, Sighat, V] = svd(A,0);
```

Problem 6: Basic concepts in real analysis, R. P.

Prove the following statements:

1. Let $U, V \in \mathbb{R}$ be open sets. Consider the product set

$$U \times V = \{(x, y) \mid x \in U, y \in V\} \subset \mathbb{R}^2.$$

Show that $U \times V$ is open.

2. Series (partial sums of sequences) form an important class of sequences. One example is

$$S_n = \sum_{k=0}^{n} \frac{1}{k!} = \frac{1}{0!} + \frac{1}{1!} + \frac{1}{2!} + \frac{1}{3!} + \cdots + \frac{1}{n!}.$$

Show that S_n converges. *(Hint: You might find it useful to use the fact that $n! \geq 2^{n-1}$.)*

3. Consider the sequence of partial sums given by

$$S_n = \sum_{k=1}^{n} \frac{1}{k^2}.$$

Show that S_n converges by showing it is Cauchy.

4. Answer the following questions regarding the Mean Value Theorem:

   (a) Let $f(x)$ be any quadratic polynomial $f(x) = \alpha x^2 + \beta x + \gamma$. Consider the secant line joining the points $(t_1, f(t_1))$ and $(t_2, f(t_2))$. What is the slope of this secant line (in terms of α, β, γ, and t_i)? Simplify as much as possible.

   (b) For the function f in (a), the Mean Value Theorem guarantees the existence of some $c \in (t_1, t_2)$ such that $f'(c)$ is equal to the above slope. For this particular f, what is this point c?

   (c) Use the Mean Value Theorem to deduce the following inequality for all x, y:

$$|\sin y - \sin x| \leq |y - x|.$$

5. Let $f : [a, b] \mapsto \mathbb{R}$ be a function that is continuous on $[a, b]$ and second differentiable (*i.e.*, f'' exists) on (a, b). Assume that the line segment joining the points $A = (a, f(a))$ and $B = (b, f(b))$ intersect the graph of f in a third point different from A and B. Show that $f''(c) = 0$ for some $c \in (a, b)$.