EECS 127/227AT 3/8/2018
L. El Ghaoui

Midterm: Solutions

1. Consider the 2×2 matrix

$$A = \frac{1}{\sqrt{10}} \begin{pmatrix} 2 \\ 1 \end{pmatrix} \begin{pmatrix} 1 & -1 \end{pmatrix} + \frac{2}{\sqrt{10}} \begin{pmatrix} -1 \\ 2 \end{pmatrix} \begin{pmatrix} 1 & 1 \end{pmatrix}.$$

   (a) What is an SVD of A? Express it as $A = USV^\top$, with S the diagonal matrix of singular values ordered in decreasing fashion. Make sure to check all the properties required for U, S, V.

   (b) Find the semi-axis lengths and principal axes of the ellipsoid

$$\mathcal{E}(A) = \left\{ Ax \ : x \in \mathbb{R}^2, \ \ \|x\|_2 \leq 1 \right\}.$$

   Hint: Use the SVD of A to show that every element of $\mathcal{E}(A)$ is of the form $y = U\bar{y}$ for some element $\bar{y}$ in $\mathcal{E}(S)$. That is, $\mathcal{E}(A) = \{U\bar{y} \ : \ \bar{y} \in \mathcal{E}(S)\}$. (In other words the matrix U maps $\mathcal{E}(S)$ into the set $\mathcal{E}(A)$.) Then analyze the geometry of the simpler set $\mathcal{E}(S)$.

   (c) What is the set $\mathcal{E}(A)$ when we append a zero vector after the last column of A, that is A is replaced with $\tilde{A} = [A, 0] \in \mathbb{R}^{2\times 3}$?

   (d) Same question when we append a row after the last row of A, that is, A is replaced with $\tilde{A} = [A^\top, 0]^\top \in \mathbb{R}^{3\times 2}$. Interpret geometrically your result.

Solution:

   (a) We have

$$A = \sigma_1 u_1 v_1^\top + \sigma_2 u_2 v_2^\top = USV^\top,$$

   where $U = [u_1, u_2]$, $V = [v_1, v_2]$ and $S = \operatorname{diag}\left(() \, \sigma_1, \sigma_2\right)$, with $\sigma_1 = 2$, $\sigma_2 = 1$, and

$$u_1 = \frac{1}{\sqrt{5}} \begin{pmatrix} -1 \\ 2 \end{pmatrix}, \ \ u_2 = \frac{1}{\sqrt{5}} \begin{pmatrix} 2 \\ 1 \end{pmatrix}, \ \ v_1 = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 \\ 1 \end{pmatrix}, \ \ v_2 = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 \\ -1 \end{pmatrix}.$$

   The triplet (U, S, V) is an SVD of A, since S is diagonal with non-negative elements on the diagonal, and U, V are orthogonal matrices ($U^\top U = V^\top V = I_2$). To check this, we first check that the Euclidean norm of u_1, u_2, v_1, v_2 is one. (This is why we factored the term $\sqrt{10}$ into $\sqrt{2} \cdot \sqrt{5}$.) In addition, $u_1^\top u_2 = v_1^\top v_2 = 0$. Thus, U, V are orthogonal, as claimed.

(b) We have, for every x, $y := Ax = US(V^\top x)$ hence $y = U\bar{y}$, with $\bar{y} = S\bar{x}$ and $\bar{x} = V^\top x$. Since V is orthogonal, $\|\bar{x}\|_2 = \|x\|_2$. In fact, when x runs the unit Euclidean sphere, so does $\bar{x}$. Thus every element of $\mathcal{E}(A)$ is of the form $y = U\bar{y}$ for some element $\bar{y}$ in $\mathcal{E}(S)$. To analyze $\mathcal{E}(A)$ it suffices to analyze $\mathcal{E}(S)$ and then transform the points of the latter set via the mapping $\bar{y} \to U\bar{y}$.

Since

$$\mathcal{E}(S) = \left\{\sigma_1 \bar{x}_1 e_1 + \sigma_1 \bar{x}_2 e_2 \ : \ \bar{x}_1^2 + \bar{x}_2^2 \leq 1\right\},$$

with e_1, e_2 the unit vectors, we have

$$\mathcal{E}(A) = \left\{\sigma_1 \bar{x}_1 u_1 + \sigma_1 \bar{x}_2 u_2 \ : \ \bar{x}_1^2 + \bar{x}_2^2 \leq 1\right\}.$$

In the coordinate system defined by the orthonormal basis (u_1, u_2) the set is an ellipsoid with semi-axis lengths (σ_1, σ_2), and principal axes given by the coordinate axes. In the original system the principal axes are u_1, u_2.

(c) When we append a zero column after the last column of A we are doing nothing to $\mathcal{E}(A)$. Indeed, the condition

$$y = Ax \text{ for some } x \in \mathbb{R}^2, \quad \|x\|_2 \leq 1$$

is the same as

$$y = \begin{pmatrix} A & 0 \end{pmatrix} z \text{ for some } z \in \mathbb{R}^3, \quad \|z\|_2 \leq 1.$$

Geometrically, the projection of a 3-dimensional unit sphere on the first two coordinates is the 2-dimensional unit sphere. Hence we loose nothing if the 2D sphere used to generate the points x is replaced by the projection of the 3D sphere.

(d) Here we append a row after the last row of A, replacing A with

$$\tilde{A} = \begin{pmatrix} A \\ 0 \end{pmatrix} \in \mathbb{R}^{3\times 2}.$$

The set $\mathcal{E}(\tilde{A})$ is the set of points of the form $(y, 0) \in \mathbb{R}^3$ where $y \in \mathcal{E}(A)$. This means that we are simply embedding the ellipsoid $\mathcal{E}(A)$ into a 3D space, instead of the original 2D one. The set $\mathcal{E}(\tilde{A})$ is now a degenerate (flat) ellipsoid in $\mathbb{R}^2$, entirely contained on the plane defined by the first two unit vectors in $\mathbb{R}^2$.

2. *Student scores and duality.* We are given a $n \times m$ matrix M that contains the scores of n sudents on an exam with m parts, so that $M_{i,j}$ is the score on student i on part j.

   (a) Someone hypothesizes that the score M_{ij} is simply the product of two variables, the i-th student overall academic ability a_i, and the difficulty level of part j, d_j. If that is the case, what is the rank and an SVD of M, in terms of the vectors a, d?

   (b) How would you test the above hypothesis, on real-world data, and estimate vectors a, d? Describe precisely the steps you would take.

   (c) We define an n-vector b, with b_i the largest score, across the whole exam, of student i. Similarly we define an m-vector s, with s_j the smallest score, across students, for part j. Show that for every i, j, $s_j \le b_i$. *Hint:* show that M_{ij} is between the two quantities.

   (d) Show that

$$d^* := \max_{1 \le j \le m} \min_{1 \le i \le n} M_{ij} \le \min_{1 \le i \le n} \max_{1 \le j \le m} M_{ij} := p^*.$$

   How does this relate to weak duality? Discuss.

   (e) Assume that M satisfies the hypothesis of part 2a, with positive vectors a, d. Does strong duality hold (that is, $p^* = d^*$) in that case? Justify your answer.

Solution:

(a) The hypothesis implies that $M = ad^\top$, hence M is rank-one. An SVD is

$$M = \sigma u v^\top,$$

with $\sigma = \|a\|_2 \cdot \|d\|_2$, $u = a/\|a\|_2$, $v = d/\|d\|_2$. Here, u is a left singular vector, and v is a right singular vector, both associated with the only non-zero singular value σ.

(b) We can solve the rank-one approximation problem

$$\min_{a,d} \|M - ad^\top\|_F$$

At optimum, a (resp. d) will be (proportional to) the left (resp. right) singular vector associated with the largest singular value of M. If the above quantity is small with respect to $\|M\|_F$, we can say that the hypothesis (approximately) hold.

(c) We have for every i, j

$$s_j = \min_k M_{kj} \le M_{ij} \le \max_h M_{ih} = b_i,$$

which proves the result. This is a special case of weak duality, applied to a function of two discrete variables $(i, j) \to M_{ij}$.

(d) The result follows immediately, since $s_j \leq b_i$ for every i, j then for every i:

$$d^* = \max_{1 \leq j \leq m} s_j \leq b_i.$$

Since the above is true for every i, we have

$$p^* = \min_{1 \leq i \leq n} p_i \geq d^*.$$

(e) If $M = ad^\top$, then

$$\begin{aligned} d^* &= \max_{1 \leq j \leq m} \min_{1 \leq i \leq n} M_{ij} \\ &= \max_{1 \leq j \leq m} \min_{1 \leq i \leq n} a_i d_j \\ &= \max_{1 \leq j \leq m} \left(d_j \cdot \min_{1 \leq i \leq n} a_i \right) \text{ (since } d > 0\text{)} \\ &= \left(\max_{1 \leq j \leq m} d_j \right) \cdot \left(\min_{1 \leq i \leq n} a_i \right). \end{aligned}$$

A similar reasoning shows that p^* is equal to the same quantity, hence strong duality holds.

3. *Optimization over a dome.* We are given two n-vectors a, y, and a scalar c, and consider the following optimization problem:

$$p^* := \max_x \ a^\top x \ : \ \|x\|_2 \leq 1, \ \ y^\top x \geq c. \tag{1}$$

(a) Is the problem convex, as stated? Justify your answer.

(b) Show that problem (1) is feasible if and only if $c \leq \|y\|_2$, which we henceforth assume.

(c) Interpret problem (1) geometrically; in particular, explain why $r^* = \|a\|_2$ when $c \leq -\|y\|_2$ and $r^* = +\infty$ (that is, the problem is infeasible) when $c > \|y\|_2$.

(d) Show that

$$p^* = \max_x \min_{\lambda \geq 0} \ a^\top x + \lambda(y^\top x - c) \ : \ \|x\|_2 \leq 1.$$

Hint: fix x and compute the minimum in the above.

(e) Assuming that strong duality holds, show that the problem can be reduced to a one-dimensional search:

$$p^* = \min_{\lambda \geq 0} \ \|a + \lambda y\|_2 - c\lambda.$$

Solution:

(a) As stated, the problem is convex, since it involves the maximization of a linear function over a convex set.

(b) The problem is feasible if and only if

$$c \geq \max_{x \,:\, \|x\|_2 \leq 1} \ y^\top x = \|y\|_2.$$

(c) The problem is to maximize a linear function over a set that is the intersection between a sphere (of radius one and center 0) and a half-space. The feasible set is a dome, and assuming a is normalized so that $\|a\|_2 = 1$, we are searching for the vector x in the dome with the smallest angle with a.

Due to the Cauchy-Shwartz inequality, we have

$$-|y\|_2 \leq x^\top y \leq \|y\|_2.$$

Hence, the dome is the full sphere when $c \leq -\|y\|_2$, in which case the problem reduces to

$$p^* = \max_{x \,:\, \|x\|_2 \leq 1} \ a^\top x = \|a\|_2.$$

On the other hand, when $c > \|y\|_2$ the problem is infeasible, because the dome is actually empty in that case.

(d) For fixed x, we have

$$\min_{\lambda \geq 0} \; a^\top x + \lambda(y^\top x - c) = \begin{cases} a^\top x & \text{if } y^\top x \geq c, \\ -\infty & \text{otherwise,} \end{cases}$$

which proves the result.

(e) Assuming strong duality:

$$\begin{aligned} p^* &= \min_{\lambda \geq 0} \; \max_x \; a^\top x + \lambda(y^\top x - c) \;\; : \;\; \|x\|_2 \leq 1 \\ &= \min_{\lambda \geq 0} \; g(\lambda), \end{aligned}$$

where g is the dual function:

$$\begin{aligned} g(\lambda) &= \max_x \; a^\top x + \lambda(y^\top x - c) \;\; : \;\; \|x\|_2 \leq 1 \\ &= \max_{\|x\|_2 \leq 1} \; x^\top (a + \lambda y) - c\lambda \\ &= \|a + \lambda y\|_2 - c\lambda, \end{aligned}$$

as claimed.