L. El Ghaoui

Midterm Solutions

1. (*6 points*) Find the projection z of the vector $x = (2, 1)$ on the line that passes through $x_0 = (1, 2)$ and with direction given by the vector $u = (1, 1)$.
   Solution: The line is the set

   $$\mathcal{L} = \{x_0 + tu \ : \ t \in \mathbf{R}\} = \left\{ \begin{pmatrix} 2+t \\ 1+t \end{pmatrix} \ : \ t \in \mathbf{R} \right\}.$$

   The problem reads

   $$\min_{z \in \mathcal{L}} \|z - x\|_2 = \min_t \|x_0 + tu - x\|_2^2.$$

   Expanding the squares, we express the objective function as

   $$\|x_0 + tu - x\|_2^2 = t^2 u^T u - 2t(x - x_0)^T u + (x - x_0)^T (x - x_0) = u^T u (t - \alpha)^2 + \text{constant},$$

   where $\alpha = (x - x_0)^T u / u^T u = 0$. The optimal t is thus $t^* = 0$, hence the projection is

   $$z = x_0 + t^* u = x_0.$$

2. (*12 points*) Consider the 2×2 matrix

$$A = \frac{1}{\sqrt{6}} \begin{pmatrix} 2 \\ 1 \end{pmatrix} \begin{pmatrix} 1 & -1 \end{pmatrix} + \frac{2}{\sqrt{6}} \begin{pmatrix} -1 \\ 2 \end{pmatrix} \begin{pmatrix} 1 & 1 \end{pmatrix}.$$

(a) What is an SVD of A? Express it as $A = USV^T$, 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

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

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 $\tilde{A} = [A, 0] \in \mathbf{R}^{2\times 3}$?

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

Solution:

(a) We have

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

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

$$u_1 = \frac{1}{\sqrt{3}} \begin{pmatrix} -1 \\ 2 \end{pmatrix}, \ \ u_2 = \frac{1}{\sqrt{3}} \begin{pmatrix} 2 \\ 1 \end{pmatrix}, \ \ v_1 = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 \\ 1 \end{pmatrix}, \ \ v_2 = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 \\ -1 \end{pmatrix}.$$

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 ($U^T U = V^T V = I_2$). 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 $\sqrt{6}$ into $\sqrt{2} \cdot \sqrt{3}$.) In addition, $u_1^T u_2 = v_1^T v_2 = 0$. Thus, U, V are orthogonal, as claimed.

(b) We have, for every x, $y := Ax = US(V^T x)$ hence $y = U\bar{y}$, with $\bar{y} = S\bar{x}$ and $\bar{x} = V^T x$. 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

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

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

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

is the same as

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

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

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

The set $\mathcal{E}(\tilde{A})$ is the set of points of the form $(y, 0) \in \mathbf{R}^3$ 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 $\mathbf{R}^3$, entirely contained on the plane defined by the first two unit vectors in $\mathbf{R}^3$.

3. (*12 points*) We consider the index fund tracking problem that arises in finance. The goal is to track a certain time-series containing the returns (the % gain if we invest one dollar at the beginning of period t, and collect it at the end of the period) of a certain index, for example the SP 500 index. The index by itself may not be tradable, or it may be very costly to trade it because it has too many assets in it. However we can try to invest in a few tradable assets, such as stocks, and hope to closely follow the index so that the return of our portfolio matches that of the index. In order to minimize the transaction costs involved in trading each asset, we would like to limit the number of assets present in the portfolio.

   Mathematically, we are given a real-valued time-series $s(t)$ with time t from $\{1, \ldots, T\}$. Here, $s(t)$ is the return of the index at time t. We also have a universe of n possible tradable assets, with, $s_i(t)$, $t = 1, \ldots, T$ the return time-series for asset i, $i = 1, \ldots, n$. Investing an amount x_i in the i-th asset at the beginning of period t produces the return $x_i s_i(t)$ at the end of period t.

   Here our variable is the vector of amounts invested in each asset, $x = (x_1, \ldots, x_n)$, which we refer to as the "portfolio vector". We plan to find the vector $x \in \mathbf{R}^n$ based on the available data, and hold it for some time in the future.

   (a) Express the time series of returns of the portfolio vector x, $s(x)$, as a linear function of x. (Here the t-th coordinate of $s(x)$ is the return of the portfolio at time t.) Precisely you will express the time series as a T-dimensional vector $s(x) = Ax$, with $x \in \mathbf{R}^n$ and A a $T \times n$ matrix that you will express in terms of problem data.

   (b) We would like to minimize the tracking error, defined as the average squared error between the return of the index and that of the portfolio. Formulate this problem as a least-squares problem. In this first model, "shorting" (selling assets) is allowed, which means that x is allowed to have negative components.

   (c) Now we do not allow shorting. Show how to modify the least-squares problem accordingly, and formulate the new problem as a QP. Make sure to put the QP in standard form, and state precisely the variables, objective function and constraints.

   (d) Now we add our concern for transaction costs, which means that we would like to trade-off the tracking error against the number of assets that have to be traded. Formulate a QP that accomplishes this. (Shorting is still not allowed.) You answer does not have to be exact, and may involve an approximation. Again, make sure to put the QP in standard form, and state precisely the variables, objective function and constraints. *Hint:* think about the l_1-norm. *Note:* you can answer this question even if you did not have the correct problem formulation in the previous parts.

   Solution:

(a) Let $s = (s(1), \ldots, s(T)) \in \mathbf{R}^T$ and $s_i = (s_i(1), \ldots, s_i(T)) \in \mathbf{R}^T$, $i = 1, \ldots, n$. We have

$$s(x) = \sum_{i=1}^{n} x_i s_i = Ax,$$

with $A = [s_1, \ldots, s_n]$ a $T \times n$ matrix.

(b) The tracking error writes

$$\frac{1}{T} \sum_{t=1}^{T} \left(s(t) - \sum_{i=1}^{n} x_i s_i(t) \right)^2 = \frac{1}{T} \|Ax - s\|_2^2.$$

The problem of minimizing the average tracking error by choice of x writes as the least-squares problem

$$\min_x \ \|Ax - s\|_2^2.$$

(c) When shorting is not allowed we add the constraints $x \geq 0$ to the least-squares problem, resulting in

$$\min_x \ \|Ax - s\|_2^2 \ : \ x \geq 0.$$

The above is a QP. Indeed, it can be written as

$$\min_x \ x^T Q x + c^T x \ : \ Cx \leq d,$$

which is a QP is standard form, with $C = -I_n$, $d = 0$, $c = -2A^T s$, and $Q := A^T A$ a positive semi-definite matrix.

(d) To control the transaction costs we need to minimize the number of non-zero components of x, since a zero value in x_i means the i-th asset will not have to be traded. We use the "l_1-trick" to minimize the number of non-zero components. This leads to the following modification to the least-squares problem

$$\min_x \ \|Ax - s\|_2^2 + \lambda \|x\|_1 \ : \ x \geq 0$$

with $\lambda > 0$ a parameter that is used to trade-off tracking error against our proxy for the number of non-zero elements. This also writes

$$\min_x \ \|Ax - s\|_2^2 + \lambda \mathbf{1}^T x \ : \ x \geq 0,$$

which can be again written as the QP in standard form above, with c replaced by $c + \lambda \mathbf{1}$.