Midterm Solutions

1. (*4 points*) Consider the set in $\mathbf{R}^3$, defined by the equation

$$\mathcal{P} := \left\{ x \in \mathbf{R}^3 \ : \ x_1 + 2x_2 + 3x_3 = 1 \right\}.$$

(a) Show that the set $\mathcal{P}$ is an affine subspace of dimension 2. To this end, express it as $x^0 + \mathbf{span}(x^1, x^2)$, where $x^0 \in \mathcal{P}$, and x^1, x^2 are independent vectors.

(b) Find the minimum Euclidean distance from 0 to the set $\mathcal{P}$. Find a point that achieves the minimum distance. (*Hint:* either apply a formula if you know it, or prove that the minimum-distance point is proportional to the vector $a := (1, 2, 3)$.)

Solutions:

(a) The affine subspace $\mathcal{P}$ is of dimension 2 in $\mathbf{R}^3$, it is a (hyper-) plane. To show this, we solve the equation for one of the variables, say x_1:

$$x_1 = 1 - 2x_2 - 3x_3.$$

This shows that any vector $x \in \mathcal{P}$ can be expressed as

$$x = \begin{pmatrix} 1 - 2x_2 - 3x_3 \\ x_2 \\ x_3 \end{pmatrix} = \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix} + x_2 \begin{pmatrix} -2 \\ 1 \\ 0 \end{pmatrix} + x_3 \begin{pmatrix} -3 \\ 0 \\ 1 \end{pmatrix},$$

with x_2, x_3 free parameters. Thus, $\mathcal{P} = x^0 + \mathbf{span}(x^1, x^2)$, with

$$x^0 = \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}, \quad x^1 = \begin{pmatrix} -2 \\ 1 \\ 0 \end{pmatrix}, \quad x^2 = \begin{pmatrix} -3 \\ 0 \\ 1 \end{pmatrix}.$$

We check that the two vectors x^1, x^2 are indeed independent, since $\lambda x^1 + \mu x^2 = 0$ implies $\lambda = \mu = 0$.

(b) The minimum distance to the affine set $\{x \ : \ Ax = b\}$, with $b \in \mathbf{R}^m$, and $A \in \mathbf{R}^{m \times n}$ full row rank (that is, AA^T is positive-definite), is given by the formula $x^* = A^T(AA^T)^{-1}b$. Applying this formula to $A = a^T$, $b = 1$, yields

$$x^* = \frac{a}{a^T a} = \frac{1}{14} \begin{pmatrix} 1 \\ 2 \\ 3 \end{pmatrix}.$$

The minimum distance is $\|x^*\|_2 = 1/\sqrt{a^T a} = 1/\sqrt{14}$.

Alternatively, we notice that any vector $x \in \mathbf{R}^3$ can be decomposed as $x = ta + z$, with $t \in \mathbf{R}$, and $z \in \mathbf{R}^3$, with $z^T a = 0$. The condition $x \in \mathcal{P}$ then implies $t = 1/a^T a$. Since $\|x\|_2^2 = t^2 a^T a + z^T z = (1/a^T a) + z^T z$, the objective function of the minimum Euclidean distance problem

$$\min_{x \in \mathcal{P}} \|x\|_2 = \min_z \sqrt{\frac{1}{a^T a} + z^T z} \;:\; a^T z = 0$$

is minimal when $z = 0$. This shows that $x^* = a/(a^T a)$, as claimed.

2. (*8 points*) Consider the operation of finding the point symmetric to a given point about a given line $\mathcal{L}$ in $\mathbf{R}^n$.

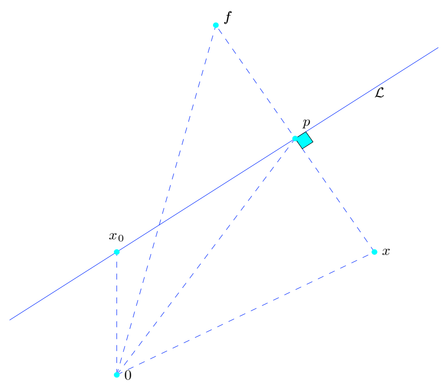

Figure 1: A point and its symmetric about the line $\mathcal{L}$.

We define the line as $\mathcal{L} := \{x_0 + tu, t \in \mathbf{R}\}$, where x_0 is a point on the line, and u its direction, which we assume is normalized: $\|u\|_2 = 1$. For a given point $x \in \mathbf{R}^n$, We denote by $f(x) \in \mathbf{R}^n$ the point that is symmetric to x about the line. (See Fig 1.) That is, $f(x) = 2p(x) - x$ where $p(x)$ is the projection of x on the line:

$$p(x) = \arg\min_{p \in \mathcal{L}} \|p - x\|_2$$

(a) Show that the mapping f is affine. Describe it in terms of a $n \times n$ matrix A and a $n \times 1$ vector b, such that $f(x) = Ax + b$ for every x. (It will be useful to use the notation $P := uu^T$.)

(b) What is the geometric interpretation of the vector b?

(c) Show that the mapping f is linear if and only if the line passes through 0.

(d) Show that $f(f(x)) = x$ for every x. What is the geometric meaning of this property?

(e) What is the range and nullspace of the matrix A? What is the rank of A? Is A invertible?

(f) Show that A is symmetric, find its eigenvalue decomposition (EVD). *Hint:* define $u_2, \ldots, u_n$ to be an orthonormal basis for the subspace orthogonal to u, and show that the orthogonal matrix $U := [u, u_2, \ldots, u_n]$ contains eigenvectors of A.

(g) Find an SVD decomposition of A. What is the relationship between the EVD of A with its SVD?

(h) Assume that the input is bounded: $\|x\|_2 \leq 1$. Find a bound on the Euclidean norm of the output $f(x)$. Find an input x that achieves the bound.

Solutions:

(a) The minimum of the function with values

$$\begin{aligned} h(t) = \|tu + x_0 - x\|_2^2 &= t^2 - 2tu^T(x - x_0) + \|x - x_0\|_2^2 \\ &= (t - u^T(x - x_0))^2 + \text{constant} \end{aligned}$$

is obtained with $t(x) := u^T(x - x_0)$. Thus, we have

$$p(x) = t(x)u + x_0 = (u^T(x - x_0))u + x_0 = P(x - x_0) + x_0, \quad P := uu^T.$$

Hence

$$f(x) = 2p(x) - x = (2P - I)(x - x_0) + x_0 = Ax + b,$$

where $A = 2P - I$, $b = 2(I - P)x_0$.

(b) Since $f(0) = b$, the latter is simply the symmetric to the origin about the line.

(c) The mapping is linear if only if $b = 0$, that is, when x_0 satisfies $Px_0 = x_0$. Hence, $x_0 = uu^T x_0 = (u^T x_0)u$ is proportional to u. In that case, the line goes through 0, since $0 = x_0 + tu$, with $t = -(u^T x_0)$.

(d) We have $Pu = u$. Further, $P^2 = (uu^T)(uu^T) = (u^T u)uu^T = uu^T = P$. The latter implies

$$A^2 = (2P - I)(2P - I) = 4P^2 - 4P + I = I.$$

In addition, $Pb = 2P(I - P)x_0 = 0$. We obtain $Ab = (2P - I)b = -b$. We thus obtain that

$$f(f(x)) = A(Ax + b) + b = A^2 x + Ab + b = x - b + b = x.$$

The geometry of this is simply that the symmetric to the symmetric is itself.

(e) The nullspace of A is the set of vectors with $Ax = 0$, meaning $2Px = x$. Thus, $x = 2(u^T x)u$ is proportional to u. Since $Ax = 0$, but $Au = u \neq 0$, we must have $u^T x = 0$, hence $x = 2(u^T x)u = 0$. We conclude that the nullspace is $\{0\}$, the range is $\mathbf{R}^n$, and A is full rank, hence invertible since it is also square.

(f) Since $A = 2uu^T - I$, it is symmetric.

Let $u_2, \ldots, u_n$ be an orthonormal basis for the subspace orthogonal to u; we have $u_i^T u = 0$, $i = 2, \ldots, n$. We have $Au = u$, $Au_i = -u_i$, $i = 2, \ldots, n$. Hence the vector u is an eigenvector associated with the eigenvalue 1, and the u_i, $i = 2, \ldots, n$, are eigenvectors all associated with the eigenvalue -1. Writing the previous conditions compactly as $AU = U\Lambda$, with $U = [u, u_2, \ldots, u_n]$ an orthogonal matrix, $\Lambda = \mathbf{diag}(1, -1, \ldots, -1)$, we obtain that A admits the symmetric eigenvalue decomposition $A = U\Lambda U^T$.

(g) We have $Pu_i = (u^T u_i)u = 0$, $i = 2, \ldots, n$. With $U := [u, u_2, \ldots, u_n]$, we get $PU = [Pu, Pu_2, \ldots, Pu_n] = [u, 0, \ldots, 0]$, therefore

$$AU = [Au, Au_2, \ldots, Au_n] = [u, -u_2, \ldots, -u_n] =: V.$$

Both U, V are orthogonal matrices. Post-multiplying the above relation by $U^T = U^{-1}$, we obtain $A = VU^T$, which is the SVD of A, with V the left singular vectors, and U the right singular vectors. Note that every singular value of A is one, which is consistent with $A^2 = AA^T = A^T A = I$.

The relationship with the SVD is simply that the eigenvectors $u, u_2, \ldots, u_n$ are the right singular vectors as well. Flipping the signs on the last $n-1$ eigenvectors provides the left singular vectors.

(h) We want to solve

$$\max_{x \,:\, \|x\|_2 \le 1} \|Ax + b\|_2.$$

Using the SVD of $A = UV^T$, we reduce the problem to

$$\max_{\tilde{x} \,:\, \|\tilde{x}\|_2 \le 1} \|\tilde{x} + U^T b\|_2,$$

where $\tilde{b} = U^T b$, and $\tilde{x} = V^T x$. The solution is obvious: simply choose a unit-norm vector in the same direction as $U^T b$:

$$\tilde{x} = \frac{U^T b}{\|U^T b\|_2} = \frac{U^T b}{\|b\|_2}.$$

We obtain

$$x = V\tilde{x} = \frac{A^T b}{\|b\|_2} = \frac{Ab}{\|b\|_2} = -\frac{b}{\|b\|_2},$$

where $b = f(0) = 2(I - P)x_0$ is the point symmetric to 0 about the line $\mathcal{L}$.

In other words, the worst-case input in the ball $\{x \,:\, \|x\|_2 \le 1\}$ is simply the one that extends away from the line in the direction opposite to the projection of 0 on the line.

3. (*6 points*) We are given m of points $x_1, \ldots, x_m$ in $\mathbf{R}^n$. To a given normalized direction $w \in \mathbf{R}^n$ ($\|w\|_2 = 1$), we associate the line with direction w passing through the origin, $\mathcal{L}(w) = \{tw \,:\, t \in \mathbf{R}\}$.

We then consider the projection of the points x_i, $i = 1, \ldots, m$, on the line $\mathcal{L}(w)$, and look at the associated coordinates of the points on the line. These *projected values* are given by $t_i(w) := \arg\min_t \|tw - x_i\|_2$, $i = 1, \ldots, m$.

We assume that for any w, the empirical average $\hat{t}(w)$ of the projected values $t_i(w)$, $i = 1, \ldots, m$, and their empirical variance $\sigma^2(w)$, are both constant, independent of the direction w (wih $\|w\|_2 = 1$). Denote by $\hat{t}$ and σ^2 the (constant) empirical average and variance. Justify your answer to the following as carefully as you can.

(a) Show that $t_i(w) = x_i^T w$, $i = 1, \ldots, m$.

(b) Show that the empirical average of the data points,

$$\hat{x} := \frac{1}{m} \sum_{i=1}^{m} x_i,$$

is zero.

(c) Show that the empirical covariance matrix of the data points,

$$\Sigma := \frac{1}{m} \sum_{i=1}^{m} (x_i - \hat{x})(x_i - \hat{x})^T,$$

is of the form $\sigma^2 \cdot I$, where I is the identity matrix of order n. (*Hint:* the largest eigenvalue $\lambda_{\max}$ of the matrix Σ can be written as: $\lambda_{\max} = \max_w \{w^T \Sigma w : w^T w = 1\}$, and a similar expression holds for the smallest eigenvalue.)

Solutions:

(a) For a given $i = 1, \ldots, m$, we have

$$t_i(w) = \arg\min_t \|tw - x_i\|_2.$$

Let us drop i for a moment, and solve the least-squares problem with variable $t \in \mathbf{R}$:

$$p^* := \min_t \|tw - x\|_2^2.$$

One can apply the closed-form solution for least-squares problem, in which the matrix involved is the full column-rank matrix w. This leads to $t(w) = (w^T w)^{-1} w^T x = w^T x$. (Recall that $\|w\|_2 = 1$.)

Alternatively, we can solve the above problem directly, exploiting again $\|w\|_2 = 1$:

$$p^* = \min_t t^2 - 2(w^T x)t + \|x\|_2^2 = \min_t (t - (w^T x))^2 + C, \text{ with } C := \|x\|_2^2 - (w^T x)^2.$$

The quantity C is constant (independent of the variable t). The first term in the objective function above is non-negative, hence $p^* \geq C$. This lower bound is attained with $t = x^T w$.

(b) The empirical average of the numbers $t_i(w)$, $i = 1, \ldots, m$, is

$$\hat{t}(w) = \frac{1}{m}\sum_{i=1}^{m} t_i(w) = \frac{1}{m}\sum_{i=1}^{m} w^T x_i = w^T \hat{x},$$

where $\hat{x}$ is the empirical average of the data points. We obtain that there is a constant $\alpha \in \mathbf{R}$ such that

$$\forall\, w, \ \ \|w\|_2 = 1 \ : \ w^T \hat{x} = \alpha.$$

Expressing the condition above for both w and $-w$, we obtain that $\alpha = 0$. This means that $\hat{x}$ is orthogonal to any (unit-norm) vector, hence it is zero.

(c) The empirical variance of the numbers $t_i(w)$, $i = 1, \ldots, m$, is given by

$$\sigma^2(w) = \frac{1}{m}\sum_{i=1}^{m}(t_i(w) - \hat{t}(w))^2.$$

Exploiting $t_i(w) = w^T x_i$, $i = 1, \ldots, m$, and $\hat{t}(w) = w^T \hat{x} = 0$, we obtain

$$\sigma^2(w) = \frac{1}{m}\sum_{i=1}^{m}(w^T x_i)^2 = \frac{1}{m}\sum_{i=1}^{m}(w^T x_i)(x_i^T w) = w^T \left(\frac{1}{m}\sum_{i=1}^{m} x_i x_i^T\right) w = w^T \Sigma w,$$

where Σ is the empirical covariance matrix of the data points. The property of constant variance is thus equivalent to the fact that the quadratic form $w \to w^T \Sigma w$ is a constant function on the unit ball $\{w \, : \, \|w\|_2 = 1\}$. We have denoted by σ^2 this constant.

The largest and smallest eigenvalue of Σ admit the variational representation

$$\lambda_{\max}(\Sigma) = \max_{\|w\|_2 = 1} w^T \Sigma w, \ \ \lambda_{\min}(\Sigma) = \min_{\|w\|_2 = 1} w^T \Sigma w.$$

Since the objective function of these problem is the same constant function, we obtain that $\lambda_{\max}(\Sigma) = \lambda_{\min}(\Sigma) = \sigma^2$. Hence all eigenvalues of Σ are equal, to σ^2. That is, the diagonal matrix of eigenvalues is given by $\Lambda = \sigma^2 I$. The eigenvalue decomposition of Σ is of the form $\Sigma = U^T \Lambda U$, with U an orthogonal matrix of eigenvectors. Since $\Lambda = \sigma^2 I$, we obtain that $\Sigma = \sigma^2 U^T U = \sigma^2 I$ as well.