Q1

Euclidean distance matrices. A matrix $X \in \mathbf{S}^n$ is a *Euclidean distance matrix* if its elements x_{ij} can be expressed as

$$x_{ij} = \|p_i - p_j\|_2^2, \quad i,j = 1, \ldots, n,$$

for some vectors $p_1, \ldots, p_n$ (of arbitrary dimension). In this exercise we prove several classical characterizations of Euclidean distance matrices, derived by I. Schoenberg in the 1930s.

(a) Show that X is a Euclidean distance matrix if and only if

$$X = \mathbf{diag}(Y)\mathbf{1}^T + \mathbf{1}\,\mathbf{diag}(Y)^T - 2Y \tag{33}$$

for some matrix $Y \in \mathbf{S}_+^n$ (the symmetric positive semidefinite matrices of order n). Here, $\mathbf{diag}(Y)$ is the n-vector formed from the diagonal elements of Y, and $\mathbf{1}$ is the n-vector with all its elements equal to one. The equality (33) is therefore equivalent to

$$x_{ij} = y_{ii} + y_{jj} - 2y_{ij}, \quad i,j = 1, \ldots, n.$$

Hint. Y is the Gram matrix associated with the vectors $p_1, \ldots, p_n$, *i.e.*, the matrix with elements $y_{ij} = p_i^T p_j$.

(b) Show that the set of Euclidean distance matrices is a convex cone.

(c) Show that X is a Euclidean distance matrix if and only if

$$\mathbf{diag}(X) = 0, \qquad X_{22} - X_{21}\mathbf{1}^T - \mathbf{1}X_{21}^T \preceq 0. \tag{34}$$

The subscripts refer to the partitioning

$$X = \begin{bmatrix} x_{11} & X_{21}^T \\ X_{21} & X_{22} \end{bmatrix}$$

with $X_{21} \in \mathbf{R}^{n-1}$, and $X_{22} \in \mathbf{S}^{n-1}$.

Hint. The definition of Euclidean distance matrix involves only the distances $\|p_i - p_j\|_2$, so the origin can be chosen arbitrarily. For example, it can be assumed without loss of generality that $p_1 = 0$. With this assumption there is a unique Gram matrix Y for a given Euclidean distance matrix X. Find Y from (33), and relate it to the lefthand side of the inequality (34).

(d) Show that X is a Euclidean distance matrix if and only if

$$\mathbf{diag}(X) = 0, \qquad (I - \frac{1}{n}\mathbf{1}\mathbf{1}^T)X(I - \frac{1}{n}\mathbf{1}\mathbf{1}^T) \preceq 0. \tag{35}$$

Hint. Use the same argument as in part (c), but take the mean of the vectors p_k at the origin, *i.e.*, impose the condition that $p_1 + p_2 + \cdots + p_n = 0$.

(e) Suppose X is a Euclidean distance matrix. Show that the matrix $W \in \mathbf{S}^n$ with elements

$$w_{ij} = e^{-x_{ij}}, \quad i,j = 1, \ldots, n,$$

is positive semidefinite.

Hint. Use the following identity from probability theory. Define $z \sim \mathcal{N}(0, I)$. Then

$$\mathbf{E}\, e^{iz^T x} = e^{-\frac{1}{2}\|x\|_2^2}$$

for all x, where $i = \sqrt{-1}$ and $\mathbf{E}$ denotes expectation with respect to z. (This is the characteristic function of a multivariate normal distribution.)

Q2

Feature selection and sparse linear separation. Suppose $x^{(1)}, \dots, x^{(N)}$ and $y^{(1)}, \dots, y^{(M)}$ are two given nonempty collections or classes of vectors in $\mathbf{R}^n$ that can be (strictly) separated by a hyperplane, *i.e.*, there exists $a \in \mathbf{R}^n$ and $b \in \mathbf{R}$ such that

$$a^T x^{(i)} - b \geq 1, \quad i = 1, \dots, N, \qquad a^T y^{(i)} - b \leq -1, \quad i = 1, \dots, M.$$

This means the two classes are (weakly) separated by the slab

$$S = \{z \mid |a^T z - b| \leq 1\},$$

which has thickness $2/\|a\|_2$. You can think of the components of $x^{(i)}$ and $y^{(i)}$ as *features*; a and b define an affine function that combines the features and allows us to distinguish the two classes.

To find the thickest slab that separates the two classes, we can solve the QP

$$\begin{array}{ll} \text{minimize} & \|a\|_2 \\ \text{subject to} & a^T x^{(i)} - b \geq 1, \quad i = 1, \dots, N \\ & a^T y^{(i)} - b \leq -1, \quad i = 1, \dots, M, \end{array}$$

with variables $a \in \mathbf{R}^n$ and $b \in \mathbf{R}$.

In this problem we seek (a, b) that separate the two classes with a thick slab, and also has a sparse, *i.e.*, there are many j with $a_j = 0$. Note that if $a_j = 0$, the affine function $a^T z - b$ does not depend on z_j, *i.e.*, the jth feature is not used to carry out classification. So a sparse a corresponds to a classification function that is parsimonious; it depends on just a few features. So our goal is to find

an affine classification function that gives a thick separating slab, and also uses as few features as possible to carry out the classification.

This is in general a hard combinatorial (bi-criterion) optimization problem, so we use the standard heuristic of solving

$$\begin{array}{ll} \text{minimize} & \|a\|_2 + \lambda \|a\|_1 \\ \text{subject to} & a^T x^{(i)} - b \geq 1, \quad i = 1, \dots, N \\ & a^T y^{(i)} - b \leq -1, \quad i = 1, \dots, M, \end{array}$$

where $\lambda \geq 0$ is a weight vector that controls the trade-off between separating slab thickness and (indirectly, through the ℓ_1 norm) sparsity of a.

Get the data in `sp_ln_sp_data.m`, which gives $x^{(i)}$ and $y^{(i)}$ as the columns of matrices `X` and `Y`, respectively. Find the thickness of the maximum thickness separating slab. Solve the problem above for 100 or so values of λ over an appropriate range (we recommend log spacing). For each value, record the separation slab thickness $2/\|a\|_2$ and $\mathbf{card}(a)$, the cardinality of a (*i.e.*, the number of nonzero entries). In computing the cardinality, you can count an entry a_j of a as zero if it satisfies $|a_j| \leq 10^{-4}$. Plot these data with slab thickness on the vertical axis and cardinality on the horizontal axis.

Use this data to choose a set of 10 features out of the 50 in the data. Give the indices of the features you choose. You may have several choices of sets of features here; you can just choose one. Then find the maximum thickness separating slab that uses only the chosen features. (This is standard practice: once you've chosen the features you're going to use, you optimize again, using only those features, and without the ℓ_1 regularization.)

Q3

Triangulation from multiple camera views. A projective camera can be described by a linear-fractional function $f : \mathbf{R}^3 \to \mathbf{R}^2$,

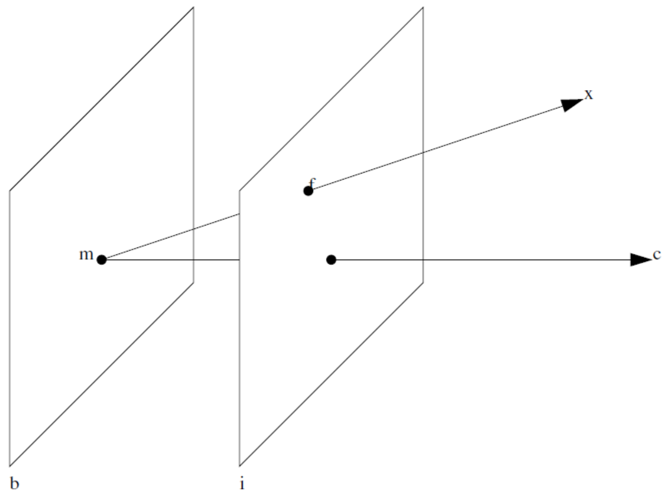
$$f(x) = \frac{1}{c^T x + d}(Ax + b), \qquad \mathbf{dom}\, f = \{x \mid c^T x + d > 0\},$$

with

$$\mathbf{rank}(\begin{bmatrix} A \\ c^T \end{bmatrix}) = 3.$$

The domain of f consists of the points in front of the camera.

Before stating the problem, we give some background and interpretation, most of which will not be needed for the actual problem.



The 3×4-matrix

$$P = \begin{bmatrix} A & b \\ c^T & d \end{bmatrix}$$

is called the *camera matrix* and has rank 3. Since f is invariant with respect to a scaling of P, we can normalize the parameters and assume, for example, that $\|c\|_2 = 1$. The numerator $c^T x + d$ is then the distance of x to the plane $\{z \mid c^T z + d = 0\}$. This plane is called the *principal plane*. The point

$$x_c = -\begin{bmatrix} A \\ c^T \end{bmatrix}^{-1} \begin{bmatrix} b \\ d \end{bmatrix}$$

lies in the principal plane and is called the *camera center*. The ray $\{x_c + \theta c \mid \theta \geq 0\}$, which is perpendicular to the principal plane, is the *principal axis*. We will define the *image plane* as the plane parallel to the principal plane, at a unit distance from it along the principal axis.

The point x' in the figure is the intersection of the image plane and the line through the camera center and x, and is given by

$$x' = x_c + \frac{1}{c^T(x - x_c)}(x - x_c).$$

Using the definition of x_c we can write $f(x)$ as

$$f(x) = \frac{1}{c^T(x - x_c)} A(x - x_c) = A(x' - x_c) = Ax' + b.$$

This shows that the mapping $f(x)$ can be interpreted as a projection of x on the image plane to get x', followed by an affine transformation of x'. We can interpret $f(x)$ as the point x' expressed in some two-dimensional coordinate system attached to the image plane.

In this exercise we consider the problem of determining the position of a point $x \in \mathbf{R}^3$ from its image in N cameras. Each of the cameras is characterized by a known linear-fractional mapping f_k and camera matrix P_k:

$$f_k(x) = \frac{1}{c_k^T x + d_k}(A_k x + b_k), \qquad P_k = \begin{bmatrix} A_k & b_k \\ c_k^T & d_k \end{bmatrix}, \qquad k = 1, \ldots, N.$$

The image of the point x in camera k is denoted $y^{(k)} \in \mathbf{R}^2$. Due to camera imperfections and calibration errors, we do not expect the equations $f_k(x) = y^{(k)}$, $k = 1, \ldots, N$, to be exactly solvable. To estimate the point x we therefore minimize the maximum error in the N equations by solving

$$\text{minimize} \quad g(x) = \max_{k=1,\ldots,N} \|f_k(x) - y^{(k)}\|_2. \tag{38}$$

(a) Show that (38) is a quasiconvex optimization problem. The variable in the problem is $x \in \mathbf{R}^3$. The functions f_k (*i.e.*, the parameters A_k, b_k, c_k, d_k) and the vectors $y^{(k)}$ are given.

(b) Solve the following instance of (38) using CVX (and bisection): $N = 4$,

$$P_1 = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix}, \qquad P_2 = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & -1 & 0 & 10 \end{bmatrix},$$

$$P_3 = \begin{bmatrix} 1 & 1 & 1 & -10 \\ -1 & 1 & 1 & 0 \\ -1 & -1 & 1 & 10 \end{bmatrix}, \qquad P_4 = \begin{bmatrix} 0 & 1 & 1 & 0 \\ 0 & -1 & 1 & 0 \\ -1 & 0 & 0 & 10 \end{bmatrix},$$

$$y^{(1)} = \begin{bmatrix} 0.98 \\ 0.93 \end{bmatrix}, \qquad y^{(2)} = \begin{bmatrix} 1.01 \\ 1.01 \end{bmatrix}, \qquad y^{(3)} = \begin{bmatrix} 0.95 \\ 1.05 \end{bmatrix}, \qquad y^{(4)} = \begin{bmatrix} 2.04 \\ 0.00 \end{bmatrix}.$$

You can terminate the bisection when a point is found with accuracy $g(x) - p^\star \leq 10^{-4}$, where $p^\star$ is the optimal value of (38).