Homework 2

EE 263 Stanford University

Summer 2017

1. **Color perception.** Human color perception is based on the responses of three different types of color light receptors, called *cones*. The three types of cones have different spectral-response characteristics, and are called L, M, and, S because they respond mainly to long, medium, and short wavelengths, respectively. In this problem we will divide the visible spectrum into 20 bands, and model the cones' responses as follows:

$$L_{\text{cone}} = \sum_{i=1}^{20} l_i p_i, \qquad M_{\text{cone}} = \sum_{i=1}^{20} m_i p_i, \qquad S_{\text{cone}} = \sum_{i=1}^{20} s_i p_i,$$

where p_i is the incident power in the ith wavelength band, and l_i, m_i and s_i are nonnegative constants that describe the spectral responses of the different cones. The perceived color is a complex function of the three cone responses, *i.e.*, the vector $(L_{\text{cone}}, M_{\text{cone}}, S_{\text{cone}})$, with different cone response vectors perceived as different colors. (Actual color perception is a bit more complicated than this, but the basic idea is right.)

   a) *Metamers.* When are two light spectra, p and $\tilde{p}$, visually indistinguishable? (Visually identical lights with different spectral power compositions are called *metamers*.)

   b) *Visual color matching.* In a color matching problem, an observer is shown a test light, and is asked to change the intensities of three primary lights until the sum of the primary lights looks like the test light. In other words, the observer is asked the find a spectrum of the form

$$p_{\text{match}} = a_1 u + a_2 v + a_3 w,$$

   where u, v, w are the spectra of the primary lights, and a_i are the intensities to be found, that is visually indistinguishable from a given test light spectrum p_{test}. Can this always be done? Discuss briefly.

   c) *Visual matching with phosphors.* A computer monitor has three phosphors, R, G, and B. It is desired to adjust the phosphor intensities to create a color that looks like a reference test light. Find weights that achieve the match or explain why no such weights exist. The data for this problem is in `color_perception_data.json`, which contains the vectors `wavelength`, `B_phosphor`, `G_phosphor`, `R_phosphor`, `L_coefficients`, `M_coefficients`, `S_coefficients`, and `test_light`.

   d) *Effects of illumination.* An object's surface can be characterized by its reflectance (*i.e.*, the fraction of light it reflects) for each band of wavelengths. If the object is illuminated with a light spectrum characterized by I_i, and the reflectance of the object is r_i (which is between 0 and 1), then the reflected light spectrum is given by $I_i r_i$, where $i = 1, \dots, 20$ denotes the wavelength band. Now consider two objects illuminated (at different times) by two different light sources, say an incandescent bulb and sunlight. Sally argues that if the two objects look identical when illuminated by a tungsten bulb, then they will look identical when illuminated by sunlight. Beth disagrees: she says that two objects can appear identical when illuminated by a tungsten bulb, but look different when lit by sunlight. Who is right? If Sally is right, explain why. If Beth is right give an example of two objects that appear identical under one light source and different under another. You can use the vectors `sunlight` and `tungsten` defined in the data file as the light sources.

Remark. Spectra, intensities, and reflectances are all nonnegative quantities, which the material of EE263 doesn't address. So just ignore this while doing this problem. These issues can be handled using the material of EE364a, however.

2. Geometric analysis of navigation. Let $(x, y) \in \mathbb{R}^2$ be the unknown coordinates of a point in the plane, let $(p_i, q_i) \in \mathbb{R}^2$ be the known coordinates of a beacon for $i = 1, \dots, n$, and let ρ_i be the measured distance between (x, y) and the ith beacon. The linearized navigation equations near a point $(x_0, y_0) \in \mathbb{R}^2$ are

$$\delta\rho \approx A \begin{bmatrix} \delta x \\ \delta y \end{bmatrix},$$

where we define the matrix $A \in \mathbb{R}^{n \times 2}$ such that

$$A_{i1} = \frac{x_0 - p_i}{\|(x_0, y_0) - (p_i, q_i)\|} \qquad \text{and} \qquad A_{i2} = \frac{y_0 - q_i}{\|(x_0, y_0) - (p_i, q_i)\|}.$$

Find the conditions under which A has full rank. Describe the conditions geometrically (*i.e.*, in terms of the relative positions of the unknown coordinates and the beacons).

3. Projection matrices. A matrix $P \in \mathbb{R}^{n \times n}$ is called a *projection matrix* if $P = P^\mathsf{T}$ and $P^2 = P$.

a) Show that if P is a projection matrix then so is $I - P$.

b) Suppose that the columns of $U \in \mathbb{R}^{n \times k}$ are orthonormal. Show that UU^T is a projection matrix. (Later we will show that the converse is true: every projection matrix can be expressed as UU^T for some U with orthonormal columns.)

c) Suppose $A \in \mathbb{R}^{n \times k}$ is full rank, with $k \le n$. Show that $A(A^\mathsf{T}A)^{-1}A^\mathsf{T}$ is a projection matrix.

d) If $S \subseteq \mathbb{R}^n$ and $x \in \mathbb{R}^n$, the point y in S closest to x is called the *projection of x on S*. Show that if P is a projection matrix, then $y = Px$ is the projection of x on $\text{range}(P)$. (Which is why such matrices are called projection matrices ...)

4. Groups of equivalent statements. In the list below there are 11 statements about two square matrices A and B in $\mathbb{R}^{n\times n}$.

a) $\text{range}(B) \subseteq \text{range}(A)$.

b) there exists a matrix $Y \in \mathbb{R}^{n\times n}$ such that $B = YA$.

c) $AB = 0$.

d) $BA = 0$.

e) $\text{rank}(\begin{bmatrix} A & B \end{bmatrix}) = \text{rank}(A)$.

f) $\text{range}(A) \perp \text{null}(B^\mathsf{T})$.

g) $\text{rank}(\begin{bmatrix} A \\ B \end{bmatrix}) = \text{rank}(A)$.

h) $\text{range}(A) \subseteq \text{null}(B)$.

i) there exists a matrix $Z \in \mathbb{R}^{n\times n}$ such that $B = AZ$.

j) $\text{rank}(\begin{bmatrix} A & B \end{bmatrix}) = \text{rank}(B)$.

k) $\text{null}(A) \subseteq \text{null}(B)$.

Your job is to collect them into (the largest possible) groups of equivalent statements. Two statements are equivalent if each one implies the other. For example, the statement 'A is onto' is equivalent to '$\text{null}(A) = \{0\}$' (when A is square, which we assume here), because every square matrix that is onto has zero nullspace, and vice versa. Two statements are not equivalent if there exist (real) square matrices A and B for which one holds, but the other does not. A group of statements is equivalent if any pair of statements in the group is equivalent.

We want *just* your answer, which will consist of lists of mutually equivalent statements; we do not need any justification.

Put your answer in the following specific form. List each group of equivalent statements on a line, in (alphabetic) order. Each new line should start with the first letter not listed above. For example, you might give your answer as

a, c, d, h
b, i
e
f, g, j, k.

This means you believe that statements a, c, d, and h are equivalent; statements b and i are equivalent; and statements f, g, j, and k are equivalent. You also believe that the first group of statements is not equivalent to the second, or the third, and so on.

5. Fitting a quadratic form to data. A quadratic form is a function $f : \mathbb{R}^n \to \mathbb{R}$ of the form

$$f(x) = \sum_{i=1}^{n}\sum_{j=1}^{n} A_{ij}x_ix_j = x^\mathsf{T} Ax.$$

We can assume without loss of generality that A is symmetric because

$$x^\mathsf{T} Ax = x^\mathsf{T}\left(\tfrac{1}{2}(A + A^\mathsf{T})\right)x,$$

where $\frac{1}{2}(A + A^\mathsf{T})$ is a symmetric matrix (called the symmetric part of A). This observation follows from the fact that

$$x^\mathsf{T} Ax = (x^\mathsf{T} Ax)^\mathsf{T} = x^\mathsf{T} A^\mathsf{T} x,$$

where the first step follows from the fact that any scalar is equal to its own transpose, and the second step is an application of the identity that the transpose of a product is equal to the product of the transposes in the reverse order. Suppose you are given data points $(x_1, y_1), \dots, (x_N, y_N) \in \mathbb{R}^2$, where y_i is a noisy measurement of $f(x_i)$.

a) Explain how to find a symmetric matrix $A \in \mathbb{R}^{n\times n}$ that minimizes the mean squared fitting error:

$$J = \sum_{i=1}^{N}(x_i^\mathsf{T} Ax_i - y_i)^2.$$

b) Apply your method to the data given in `fit_quadratic_form_data.m`. Report your matrix A, and the corresponding relative fitting error:

$$e_{\text{rel}} = \frac{\sqrt{\sum_{i=1}^{N}(x_i^\mathsf{T} Ax_i - y_i)^2}}{\sqrt{\sum_{i=1}^{N} y_i^2}}.$$

6. Least-squares model fitting. In this problem you will use least-squares to fit several different types of models to a given set of input/output data. The data consist of a scalar input sequence u, and a scalar output sequence y, for $t = 1, \dots, N$. You will develop several different models that relate the signals u and y.

- *Memoryless models.* In a memoryless model, the output at time t, *i.e.*, $y(t)$, depends only the input at time t, *i.e.*, $u(t)$. Another common term for such a model is *static*.

$$\begin{aligned}
\text{constant model:}\quad & y(t) = c_0\\
\text{static linear:}\quad & y(t) = c_1u(t)\\
\text{static affine:}\quad & y(t) = c_0 + c_1u(t)\\
\text{static quadratic:}\quad & y(t) = c_0 + c_1u(t) + c_2u(t)^2
\end{aligned}$$

- *Dynamic models.* In a dynamic model, $y(t)$ depends on $u(s)$ for some $s \neq t$. We consider some simple time-series models (see problem 2 in the reader), which are linear dynamic models.

$$\begin{aligned}
\text{moving average (MA):}\quad & y(t) = a_0u(t) + a_1u(t-1) + a_2u(t-2)\\
\text{autoregressive (AR):}\quad & y(t) = a_0u(t) + b_1y(t-1) + b_2y(t-2)\\
\text{autoregressive moving average (ARMA):}\quad & y(t) = a_0u(t) + a_1u(t-1) + b_1y(t-1)
\end{aligned}$$

Note that in the AR and ARMA models, $y(t)$ depends indirectly on all previous inputs, $u(s)$ for $s < t$, due to the recursive dependence on $y(t-1)$. For this reason, the AR and ARMA models are said to have *infinite memory*. The MA model, on the other hand, has a *finite memory*: $y(t)$ depends only on the current and two previous inputs. (Another term for this MA model is 3-tap system, where taps refer to taps on a delay line.)

Each of these models is specified by its parameters, *i.e.*, the scalars c_i, a_i, b_i. For each of these models, find the least-squares fit to the given data. In other words, find parameter values that minimize the sum-of-squares of the residuals. For example, for the ARMA model, pick a_0, a_1, and b_1 that minimize

$$\sum_{t=2}^{N} \left(y(t) - a_0 u(t) - a_1 u(t-1) - b_1 y(t-1)\right)^2 .$$

(Note that we start the sum at $t = 2$ which ensures that $u(t-1)$ and $y(t-1)$ are defined.) For each model, give the root-mean-square (RMS) residual, *i.e.*, the squareroot of the mean of the optimal residual squared. Plot the output $\hat{y}$ predicted by your model, and plot the residual (which is $y - \hat{y}$). The data for this problem are available from the class web page in the file `uy_data.json`. This file contains the vectors u and y and the scalar N (the length of the vectors). Now you can plot u, y, etc. *Note:* the dataset u, y is *not* generated by any of the models above. It is generated by a nonlinear recursion, which has infinite memory.

7. Identifying a system from input/output data. We consider the standard setup:

$$y = Ax + v,$$

where $A \in \mathbb{R}^{m \times n}$, $x \in \mathbb{R}^n$ is the input vector, $y \in \mathbb{R}^m$ is the output vector, and $v \in \mathbb{R}^m$ is the noise or disturbance. We consider here the problem of estimating the matrix A, given some input/output data. Specifically, we are given the following:

$$x^{(1)}, \ldots, x^{(N)} \in \mathbb{R}^n, \qquad y^{(1)}, \ldots, y^{(N)} \in \mathbb{R}^m.$$

These represent N samples or observations of the input and output, respectively, possibly corrupted by noise. In other words, we have

$$y^{(k)} = Ax^{(k)} + v^{(k)}, \quad k = 1, \ldots, N,$$

where $v^{(k)}$ are assumed to be small. The problem is to estimate the (coefficients of the) matrix A, based on the given input/output data. You will use a least-squares criterion to form an estimate $\hat{A}$ of A. Specifically, you will choose as your estimate $\hat{A}$ the matrix that minimizes the quantity

$$J = \sum_{k=1}^{N} \|Ax^{(k)} - y^{(k)}\|^2$$

over A.

a) Explain how to do this. If you need to make an assumption about the input/output data to make your method work, state it clearly. You may want to use the matrices $X \in \mathbb{R}^{n \times N}$ and $Y \in \mathbb{R}^{m \times N}$ given by

$$X = \left[\, x^{(1)} \;\; \cdots \;\; x^{(N)} \,\right], \qquad Y = \left[\, y^{(1)} \;\; \cdots \;\; y^{(N)} \,\right]$$

in your solution.

b) On the course web site you will find some input/output data for an instance of this problem in the file `sysid_data.json`. Executing this Julia file will assign values to m, n, and N, and create two matrices that contain the input and output data, respectively. The $n \times N$ matrix variable `X` contains the input data $x^{(1)}, \ldots, x^{(N)}$ (*i.e.*, the first column of `X` contains $x^{(1)}$, etc.). Similarly, the $m \times N$ matrix `Y` contains the output data $y^{(1)}, \ldots, y^{(N)}$. You must give your final estimate $\hat{A}$, your source code, and also give an explanation of what you did.