Homework 2 solutions

2.90. Matrices and signal flow graphs.

a) Find $A \in \mathbb{R}^{2 \times 2}$ such that $y = Ax$ in the system below:

\[
\begin{align*}
&x_1 \quad 2 \quad y_1 \\
&0.5 \\
&x_2 \quad y_2
\end{align*}
\]

b) Find $B \in \mathbb{R}^{2 \times 2}$ such that $z = Bx$ in the system below:

\[
\begin{align*}
&x_1 \quad 2 \quad 2 \quad 2 \quad 2 \quad z_1 \\
&5 \quad .5 \quad .5 \quad .5 \quad z_2 \\
&x_2 \quad y_1 \quad y_2
\end{align*}
\]

Do this two ways: first, by expressing the matrix $B$ in terms of $A$ from the previous part (explaining why they are related as you claim); and second, by directly evaluating all possible paths from each $x_j$ to each $z_i$.

Solution.

a) By evaluating path gains we have

- Gain from $x_1$ to $y_1$. There is only one path with gain 2.
- Gain from $x_1$ to $y_2$. There is only one path with gain 0.5.
- Gain from $x_2$ to $y_1$. There are no paths and therefore the gain is 0.
- Gain from $x_2$ to $y_2$. There is only one path with gain 1.
and therefore
\[ A = \begin{bmatrix} 2 & 0 \\ 0.5 & 1 \end{bmatrix}. \]

b) Clearly \( B = A^4 \). Carrying out the multiplication gives
\[ B = \begin{bmatrix} 16 & 0 \\ 7.5 & 1 \end{bmatrix}. \]

Now by directly evaluating all possible path gains we get
- Gain from \( x_1 \) to \( z_1 \). There is only one path with gain \( 2 \times 2 \times 2 \times 2 = 16 \)
- Gain from \( x_1 \) to \( z_2 \). There are 4 possible paths. These paths have gains 0.5, \( 2 \times 0.5 \), \( 2 \times 2 \times 0.5 \) and \( 2 \times 2 \times 2 \times 0.5 \) that sum up to 7.5.
- Gain from \( x_2 \) to \( z_1 \). There are no paths and therefore the gain is 0.
- Gain from \( x_2 \) to \( z_2 \). There is only one path with gain 1.

and therefore we get the same \( B \) as expected.

2.210. **Express the following statements in matrix language.** You can assume that all matrices mentioned have appropriate dimensions. Here is an example: “Every column of \( C \) is a linear combination of the columns of \( B \)” can be expressed as “\( C = BF \) for some matrix \( F \)”.

There can be several answers; one is good enough for us.

a) Suppose \( Z \) has \( n \) columns. For each \( i \), row \( i \) of \( Z \) is a linear combination of rows \( i, \ldots, n \) of \( Y \).

b) \( W \) is obtained from \( V \) by permuting adjacent odd and even columns \( (i.e., \ 1 \ and \ 2, \ 3 \ and \ 4, \ldots) \).

c) Each column of \( P \) makes an acute angle with each column of \( Q \).

d) Each column of \( P \) makes an acute angle with the corresponding column of \( Q \).

e) The first \( k \) columns of \( A \) are orthogonal to the remaining columns of \( A \).

**Solution.**

a) \( Z = UY \), where \( U \) is upper triangular, \( i.e., \ U_{ij} = 0 \) for \( i > j \).

b) \( W = VS \), where \( S \) is the odd-even switch matrix, defined as
\[ S = \begin{bmatrix} e_2 & e_1 & e_4 & e_3 & \cdots & e_m & e_{m-1} \end{bmatrix}. \]
c) All entries of the matrix $P^T Q$ are positive.

d) The diagonal entries of the matrix $P^T Q$ are positive.

e) $A^T A$ has the block diagonal form

$$A^T A = \begin{bmatrix} B_{11} & 0 \\ 0 & B_{22} \end{bmatrix},$$

where $B_{11} \in \mathbb{R}^{k \times k}$.

3.250. 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:

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

where $p_i$ is the incident power in the $i$th 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 $(L_{cone}, M_{cone}, S_{cone})$, 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 to find a spectrum of the form

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

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 $i = 1, \ldots, 20$ 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 \texttt{sunlight} and \texttt{tungsten} defined in the data file as the light sources.

\textit{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.

\textbf{Solution.}

a) Let

$$A = \begin{bmatrix} l_1 & l_2 & l_3 & \cdots & l_{20} \\
m_1 & m_2 & m_3 & \cdots & m_{20} \\
s_1 & s_2 & s_3 & \cdots & s_{20} \end{bmatrix}.$$ 

Now suppose that $c = A p$ is the cone response to the spectrum $p$ and $\tilde{c} = A \tilde{p}$ is the cone response to spectrum $\tilde{p}$. If the spectra are indistinguishable, then $c = \tilde{c}$ and $A p = A \tilde{p}$. Solving the last expression for zero gives $A(p - \tilde{p}) = 0$. In other words, $p$ and $\tilde{p}$ are metamers if $(p - \tilde{p}) \in \text{null}(A)$.

b) In symbols, the problem asks if it is always possible to find nonnegative $a_1$, $a_2$, and $a_3$ such that

$$\begin{bmatrix} m_1 \\ m_2 \\ m_3 \end{bmatrix} = A p_{\text{test}} = A \begin{bmatrix} u & v & w \end{bmatrix} \begin{bmatrix} a_1 \\ a_2 \\ a_3 \end{bmatrix}.$$ 

Let $P = \begin{bmatrix} u & v & w \end{bmatrix}$ and let $B = AP$. If $B$ is invertible, then

$$\begin{bmatrix} a_1 \\ a_2 \\ a_3 \end{bmatrix} = B^{-1} \begin{bmatrix} m_1 \\ m_2 \\ m_3 \end{bmatrix}.$$ 

However, $B$ is not necessarily invertible. For example, if rank($A$) < 3 or rank($P$) < 3 then $B$ will be singular. Physically, $A$ is full rank if the L, M, and S cone responses are linearly independent, which they are. The matrix $P$ is full rank if and only if the spectra of the primary lights are independent. Even if both $A$ and $P$ are full rank, $B$ could still be singular. Primary lights that generate an invertible $B$ are called \textit{visually independent}. If $B$ is invertible, $a_1$, $a_2$, and $a_3$ exist that satisfy

$$A p_{\text{test}} = A \begin{bmatrix} u & v & w \end{bmatrix} \begin{bmatrix} a_1 \\ a_2 \\ a_3 \end{bmatrix}.$$ 

4
but one or more of the $a_i$ may be negative in which case in the experimental setup described, no match would be possible. However, in a more complicated experimental setup that allows the primary lights to be combined either with each other or with $p_{test}$, a match is always possible if $B$ is invertible. In this case, if $a_i < 0$, the $i$th light should be mixed with $p_{test}$ instead of the other primary lights. For example, suppose $a_1 < 0$, $a_2, a_3 \geq 0$ and $b_1 = -a_1$, then

$$A(b_1 u + p_{test}) = A(a_2 v + a_3 w),$$

and each spectrum has a nonnegative weight.

c) Weights can be found as described above. The R, G, and B phosphors should be weighted by 0.4226, 0.0987, and 0.5286 respectively.

The following Julia code illustrates the steps.

```
# Extraction of the data

include("readJSON263.jl");
mydata = readJSON263("color_perception.json");

L_coefficients = mydata["L_coefficients"]["data"]; M_coefficients = mydata["M_coefficients"]["data"]; S_coefficients = mydata["S_coefficients"]["data"]; R_phosphor = mydata["R_phosphor"]["data"]; G_phosphor = mydata["G_phosphor"]["data"]; B_phosphor = mydata["B_phosphor"]["data"]; test_light = mydata["test_light"]["data"]; tungsten = mydata["tungsten"]["data"]; sunlight = mydata["sunlight"]["data"];

A = [L_coefficients; M_coefficients; S_coefficients]; B = A*[R_phosphor' G_phosphor' B_phosphor']; weights = B'\A*test_light
```

Equivalently, the following matlab code illustrates the steps.

```
close all; clear all;
color_perception;
A = [L_coefficients; M_coefficients; S_coefficients]; B = A*[R_phosphor' G_phosphor' B_phosphor'];
weights = inv(B)*A*test_light;
```

d) Beth is right. Let $r$ and $\tilde{r}$ be the reflectances of two objects and let $p$ and $\tilde{p}$ be two
spectra. Let $A$ be defined as before. Then, the objects will look identical under $p$ if

$$
A \begin{bmatrix}
    r_1 & 0 & \cdots & 0 \\
    0 & r_2 & \cdots & 0 \\
    \vdots & \vdots & \ddots & \vdots \\
    0 & 0 & \cdots & r_{20}
\end{bmatrix} \quad \text{and} \quad p = A \begin{bmatrix}
    \tilde{r}_1 & 0 & \cdots & 0 \\
    0 & \tilde{r}_2 & \cdots & 0 \\
    \vdots & \vdots & \ddots & \vdots \\
    0 & 0 & \cdots & \tilde{r}_{20}
\end{bmatrix}.
$$

This is equivalent to saying $(R - \tilde{R})p \in \text{null}(A)$. The objects will look different under $\tilde{p}$ if, additionally, $AR\tilde{p} \neq A\tilde{R}\tilde{p}$ which means that $(R - \tilde{R})\tilde{p} \notin \text{null}(A)$. The following code shows how to find reflectances $r_1$ and $r_2$ for two objects such that the objects will have the same color under tungsten light and will have different colors under sunlight.

```matlab
n = N[:,1];
n = n*10;
for i in 1:20
    n[i] = n[i]/tungsten[i];
end

r1 = [0; 0.2; 0.3; 0.7; 0.7; 0.8; 0.7; 0.8; 0.2; 0.9; 0.8; 0.2; 0.8; 0.9; 0.2; 0.8; 0.7; 0.2; 0.4];
r2 = r1 - n;

t1 = zeros(20);
t2 = zeros(20);
for i in 1:20
    t1[i] = r1[i]*tungsten[i];
t2[i] = r2[i]*tungsten[i];
end

color1_tungsten = A*t1'
color2_tungsten = A*t2'

for i in 1:20
    s1[i] = r1[i]*sunlight[i];
s2[i] = r2[i]*sunlight[i];
end

color1_sunlight = A*s1'
color2_sunlight = A*s2'
```

Or, in Matlab:

```matlab
close all; clear all;
color_perception;
```
A = [L_coefficients; M_coefficients; S_coefficients]; N = null(A);
n = N(:,1);
n = n*10;
for i = 1:20
n(i) = n(i)/tungsten(i);
end
r1 = [0; .2; .3; .7; .7; .8; .8; .9; .8; .2; .8; .8; .9; .2; .8; .3; .8; .7; .2; .4];
r2 = r1-n;
for i = 1:20
t1(i) = r1(i)*tungsten(i);
t2(i) = r2(i)*tungsten(i);
end color1_tungsten = A*t1'; color2_tungsten = A*t2';
for i = 1:20
s1(i) = r1(i)*sunlight(i);
s2(i) = r2(i)*sunlight(i);
end color1_sun = A*s1'; color2_sun = A*s2';
253.5187

3.260. Halfspace. Suppose \( a, b \in \mathbb{R}^n \) are two given points. Show that the set of points in \( \mathbb{R}^n \) that are closer to \( a \) than \( b \) is a halfspace, i.e.:

\[
\{ x \mid ||x-a|| \leq ||x-b|| \} = \{ x \mid c^T x \leq d \}
\]

for appropriate \( c \in \mathbb{R}^n \) and \( d \in \mathbb{R} \). Give \( c \) and \( d \) explicitly, and draw a picture showing \( a, b, c, \) and the halfspace.

Solution. It is easy to see geometrically what is going on: the hyperplane that goes right between \( a \) and \( b \) splits \( \mathbb{R}^n \) into two parts; the points closer to \( a \) (than \( b \)) and the points closer to \( b \) (than \( a \)). More precisely, the hyperplane is normal to the line through \( a \) and \( b \), and intersects that line at the midpoint between \( a \) and \( b \). Now that we have the idea, let’s try to derive it algebraically. Let \( x \) belong to the set of points in \( \mathbb{R}^n \) that are closer to \( a \) than \( b \). Therefore \( ||x-a|| < ||x-b|| \) or \( ||x-a||^2 < ||x-b||^2 \) so

\[
(x-a)^T(x-a) < (x-b)^T(x-b).
\]

Expanding the inner products gives

\[
x^T x - x^T a - a^T x + a^T a < x^T x - x^T b - b^T x + b^T b
\]

or

\[
-2a^T x + a^T a < -2b^T x + b^T b
\]

and finally

\[
(b-a)^T x < \frac{1}{2}(b^T b - a^T a).
\]
Thus (1) is in the form $c^T x < d$ with $c = b - a$ and $d = \frac{1}{2}(b^T b - a^T a)$ and therefore we have shown that the set of points in $\mathbb{R}^n$ that are closer to $a$ than $b$ is a halfspace. Note that the hyperplane $c^T x = d$ is perpendicular to $c = b - a$.

3.420. Relative deviation between vectors. Suppose $a$ and $b$ are nonzero vectors of the same size. The relative deviation of $b$ from $a$ is defined as the distance between $a$ and $b$, divided by the norm of $a$,

$$\eta_{ab} = \frac{\|a - b\|}{\|a\|}.$$

This is often expressed as a percentage. The relative deviation is not a symmetric function of $a$ and $b$; in general, $\eta_{ab} \neq \eta_{ba}$.

Suppose $\eta_{ab} = 0.1$ (i.e., 10%). How big and how small can be $\eta_{ba}$ be? How big and how small can $\angle(a, b)$ be? Explain your reasoning. For bounding $\angle(a, b)$, you can just draw some pictures; you don’t have to give a formal argument.

Solution. We’ll work out a more general case. We have

$$\|a - b\| = \eta_{ab}\|a\|.$$
We need to get upper and lower bounds on $\|b\|$. We can use the triangle inequality to get an upper bound:

\[
\|b\| = \|a + (-a + b)\| \\
\leq \|a\| + \|a - b\| \\
= (1 + \eta_{ab})\|a\|.
\]

This inequality is tight, if $a$ and $a - b$ are anti-aligned, which is the same as $a$ and $b$ being aligned. Now we can say that

\[
\eta_{ba} = \frac{\|a - b\|}{\|b\|} \geq \frac{\eta_{ab}\|a\|}{(1 + \eta_{ab})\|a\|} = \frac{\eta_{ab}}{1 + \eta_{ab}}.
\]

This is a general bound, and it is tight when $a$ and $b$ are aligned. For $\eta_{ab} = 0.1$, we find that $\eta_{ba} \geq 0.1/1.1 = 0.0909$.

Now let’s get a lower bound on $\|b\|$, again using the triangle inequality:

\[
\|a\| = \|b + (a - b)\| \\
\leq \|b\| + \|a - b\| \\
= \|b\| + \eta_{ab}\|a\|.
\]

This inequality is tight if $a$ and $a - b$ are aligned, which is the same as $a$ and $b$ being anti-aligned. Subtracting, we get

\[
(1 - \eta_{ab})\|a\| \leq \|b\|.
\]

Assuming that $\eta_{ab} < 1$ (which is the case for $\eta_{ab} = 0.1$), we then have

\[
\eta_{ba} = \frac{\|a - b\|}{\|b\|} \leq \frac{\eta_{ab}\|a\|}{(1 - \eta_{ab})\|a\|} = \frac{\eta_{ab}}{1 - \eta_{ab}}.
\]

This is a general bound, tight when $a$ and $b$ are anti-aligned. For $\eta_{ab} = 0.1$, we find that $\eta_{ba} \geq 0.1/0.9 = 0.1111$.

In summary, when $\eta_{ab} = 0.1$, $\eta_{ba}$ can range between 0.0909 and 0.1111. The lower limit occurs when $a$ and $b$ are aligned; the upper limit occurs when $a$ and $b$ are aligned.

Now let’s look at the angle. We first give a geometric argument. Let’s look at the plane spanned by $a$ and $b$. Then vector $b$ must be in a ball of radius $\eta_{ab}\|a\|$, centered in $a$, as shown
below. Assuming that $\eta_{ab} < 1$ (which is the case here), the ball does not include the origin.

Now, we look on how small and how large the angle between $a$ and $b$ can be, as $b$ varies over the ball. When $a$ and $b$ are aligned, $\angle (a, b) = 0$. Now let’s see how large the angle can be. The largest angle is obtained when $b$ and $a - b$ are orthogonal; in this case $(0, a, b)$ are the vertices of a right triangle. In this case we have $\angle (a, b) = \arcsin \eta_{ab}$. For $\eta_{ab}$, we find that $\angle (a, b) = 0.1002$. Therefore $\angle (a, b)$ can take values in the interval $[0, 0.1002]$.

3.430. Single sensor failure detection and identification. We have $y = Ax$, where $A \in \mathbb{R}^{m \times n}$ is known, and $x \in \mathbb{R}^n$ is to be found. Unfortunately, up to one sensor may have failed (but you don’t know which one has failed, or even whether any has failed). You are given $\tilde{y}$ and not $y$, where $\tilde{y}$ is the same as $y$ in all entries except, possibly, one (say, the $k$th entry). If all sensors are operating correctly, we have $y = \tilde{y}$. If the $k$th sensor fails, we have $\tilde{y}_i = y_i$ for all $i \neq k$.

The file one_bad_sensor.m, available on the course web site, defines $A$ and $\tilde{y}$ (as $A$ and $\tilde{y}$). Determine which sensor has failed (or if no sensors have failed). You must explain your method, and submit your code.

For this exercise, you can use the matlab code \texttt{rank([F g]==rank(F) to check if $g \in \text{range}(F)$. (We will see later a much better way to check if $g \in \text{range}(F).$)

**Solution.** Let $y^{(i)}$ be the measurement vector $y$ with the $i$th entry removed. Likewise, let $A^{(i)}$ be the measurement matrix with the $i$th row of $A$ removed. This corresponds to the system without the $i$th sensor.

If the $i$th sensor is faulty, we will almost surely have $y \notin \text{range}(A)$ (unless the sensor failure happens to give the same response $y_i$ as that predicted by $A$, which is highly unlikely). However, once we remove its faulty measurement, we will certainly have $y^{(i)} \in \text{range}(A^{(i)})$.

To test if a vector $z$ is in range($C$), we can use matlab and compare \texttt{rank([C z]) == rank(C).} If they are equal, $z \in \text{range}(C)$. Otherwise \texttt{rank([C z]) == rank(C) + 1.} To find a faulty sensor, we remove one row of $A$ at a time, and use the above test.

The following matlab code solves the problem one_bad_sensor
for k=1:m
    withoutk=[1:k-1 k+1:m];
    Atent = A(withoutk,:);
    ytent = ytilde(withoutk);
    if rank([ Atent ytent ]) == rank(Atent)
        k
    end
end

The 11th sensor is faulty.