
EECS 16A Designing Information Devices and Systems I

Fall 2018 Homework 13

This homework is due November 30, 2018, at 23:59.

This is a long homework so you should start early.

Self-grades are due December 4, 2018, at 23:59.

Submission Format

Your homework submission should consist of **two** files.

- `hw13.pdf`: A single PDF file that contains all of your answers (any handwritten answers should be scanned) as well as your IPython notebook saved as a PDF.

If you do not attach a PDF of your IPython notebook, you will not receive credit for problems that involve coding. Make sure that your results and your plots are visible.

- `hw13.ipynb`: A single IPython notebook with all of your code in it.

In order to receive credit for your IPython notebook, you must submit both a “printout” and the code itself.

Submit each file to its respective assignment on Gradescope.

1. Recipe Reconnaissance

Engineering Edibles has been growing in size and popularity and has introduced two new cookies: Decadent Dwight and Heavenly Hearst. As a result of their popularity there is an increased interest in understanding their secret recipes.

The bakery produces **40 Decadent Dwight** and **50 Heavenly Hearst** cookies each day. Each cookie costs \$1. The bakery business is way overpriced, and everyone knows about 80 cents of the cost comes from profits (and labor); ingredients are only worth **about 20 cents**.

The team from Berkeley wants to figure out the recipes for the two new cookies, and they know the recipes are *different*. For the purpose of this problem, each cookie only contains three ingredients: eggs, sugar and butter (of course, you would use flour, water, etc. in real life!).

- (a) How many unknowns are there in the problem? How many linearly independent equations would you need to be able to solve for these unknowns (assuming a consistent system of equations)?
- (b) Unfortunately, the team is not able to find precise information about the ingredients used in the cookie making process.

They stake-out Engineering Edibles, and they see Bob the Baker buy a dozen eggs for \$2 **daily** and a 5 kg bag of sugar **every week**. They hire a master-taster, who tells them there is **exactly** 10 grams of butter in each of the cookies (for both Decadent Dwight and Heavenly Hearst). The team precisely knows how much butter is in the cookies. They know from the supermarket that 1 kg of sugar costs 5 dollars, and 100 grams of butter costs 1 dollar. (*The prices of sugar and butter are precise and is precisely known, as is the amount of butter from the master taster.*) However, the amount of eggs and sugar are imprecisely known.

All this is not enough for the team to unlock the secret recipe! (Why?)

One day, however, the team gets lucky and hears through the grapevine that Heavenly Hearst contains **about** 10 grams of sugar. (*This value might not be precise, since it's just gossip!*)

Let's also remember that the ingredients for each cookie of either variety would cost \$0.2.

Using the information above, set up the problem as a least-squares problem and find best estimate of the secret recipe. Identify what variables you are estimating. You are welcome to use a computer to solve this.

- (c) The Berkeley team decides to do scientific experiments to better their estimates of the recipe. They obtain a scale, and weigh the new cookies. Their (cheap and second-hand) scale is only accurate to the gram and easily affected by air currents, so they get noisy observations: Decadent Dwight is *about* 25 grams, and Heavenly Hearst is about 24 grams. Assume that 1 egg = 50 grams.

Update the least squares problem with this new information, and see how the values change.

Are the new values more accurate? (*Hint: Find the sum of squared errors, $\|e\|^2$ then divide $\|e\|$ by the number of measurements to find the **average error for data points** for both (b) and (c).*)

2. Constrained Least Squares Optimization

In this problem, you'll go through a process of guided discovery to solve the following optimization problem: Consider a matrix $\mathbf{A} \in \mathbb{R}^{M \times N}$, of full column rank, where $M > N$. Determine a unit vector \vec{x} that minimizes $\|\mathbf{A}\vec{x}\|$, where $\|\cdot\|$ denotes the 2-norm—that is,

$$\|\mathbf{A}\vec{x}\|^2 \triangleq \langle \mathbf{A}\vec{x}, \mathbf{A}\vec{x} \rangle = (\mathbf{A}\vec{x})^T \mathbf{A}\vec{x} = \vec{x}^T \mathbf{A}^T \mathbf{A} \vec{x}.$$

This is equivalent to solving the following optimization problem:

$$\text{Determine } \vec{x} = \operatorname{argmin}_{\vec{x}} \|\mathbf{A}\vec{x}\|^2 \quad \text{subject to the constraint } \|\vec{x}\|^2 = 1.$$

This task may *seem* like solving a standard least squares problem $\mathbf{A}\vec{x} = \vec{b}$, where $\vec{b} = \vec{0}$, but it isn't. As an example, notice $\vec{x} = \vec{0}$ is *not* a valid solution to our problem, because the zero vector does not have unit length. However, it could be the solution to a different least squares problem. Our optimization problem is a least squares problem with a constraint—hence the term *Constrained Least Squares Optimization*. The constraint is that the vector \vec{x} must lie on the unit sphere in \mathbb{R}^N . You'll tackle this problem in a methodical, step-by-step fashion.

Let $(\lambda_1, \vec{v}_1), \dots, (\lambda_N, \vec{v}_N)$ denote the eigenpairs (i.e., eigenvalue/eigenvector pairs) of $\mathbf{A}^T \mathbf{A}$. Assume that the eigenvalues are all real, distinct and indexed in an ascending fashion—that is,

$$\lambda_1 < \dots < \lambda_N.$$

Assume, too, that each eigenvector has been normalized to have unit length—that is, $\|\vec{v}_k\| = 1$ for all $k \in \{1, \dots, N\}$.

- (a) Show that $0 < \lambda_1$, i.e. all the eigenvalues are strictly positive.

Hint: Consider $\|\mathbf{A}\vec{v}\|^2$

- (b) Consider two eigenpairs (λ_k, \vec{v}_k) and $(\lambda_\ell, \vec{v}_\ell)$ corresponding to distinct eigenvalues of $\mathbf{A}^T \mathbf{A}$ —that is, $\lambda_k \neq \lambda_\ell$. Prove that the corresponding eigenvectors \vec{v}_k and \vec{v}_ℓ are orthogonal: $\vec{v}_k \perp \vec{v}_\ell$.

To help you get started, consider the two equations

$$\mathbf{A}^T \mathbf{A} \vec{v}_k = \lambda_k \vec{v}_k \quad (1)$$

and

$$\vec{v}_\ell^T \mathbf{A}^T \mathbf{A} = \lambda_\ell \vec{v}_\ell^T. \quad (2)$$

Premultiply Equation 1 with \vec{v}_ℓ^T , postmultiply Equation 2 with \vec{v}_k , compare the two, and explain how one may then infer that \vec{v}_k and \vec{v}_ℓ are orthogonal, i.e. $\langle \vec{v}_k, \vec{v}_\ell \rangle = 0$.

- (c) The results of part (b) implies that the N eigenvectors of $\mathbf{A}^T \mathbf{A}$ are mutually orthogonal—and each has unit length. A basis formed by vectors that are both (1) mutually orthogonal and (2) has unit length is called **an orthonormal basis**. Since the eigenvalues of $\mathbf{A}^T \mathbf{A}$ are distinct, the eigenvectors form a basis, and the extra properties imply they form an orthonormal basis. in \mathbb{R}^N . This means that we can express an arbitrary vector $\vec{x} \in \mathbb{R}^N$ as a linear combination of the eigenvectors $\vec{v}_1, \dots, \vec{v}_N$, as follows:

$$\vec{x} = \sum_{n=1}^N \alpha_n \vec{v}_n.$$

- i. Determine the n^{th} coefficient α_n in terms of \vec{x} and one or more of the eigenvectors $\vec{v}_1, \dots, \vec{v}_N$.
- ii. Suppose \vec{x} is a unit-length vector (i.e., a unit vector) in \mathbb{R}^N . Show that

$$\sum_{n=1}^N \alpha_n^2 = 1$$

where the α_n 's are the coefficients of \vec{x} in the basis defined earlier.

- (d) Now you're well-positioned to tackle the grand challenge of this problem—determine the unit vector \vec{x} that minimizes $\|\mathbf{A}\vec{x}\|$.

Note that the task is the same as finding a unit vector \vec{x} that minimizes $\|\mathbf{A}\vec{x}\|^2$.

Express $\|\mathbf{A}\vec{x}\|^2$ in terms of $\{\alpha_1, \alpha_2 \dots \alpha_N\}$, $\{\lambda_1, \lambda_2 \dots \lambda_N\}$, and $\{\vec{v}_1, \vec{v}_2 \dots \vec{v}_N\}$, and find an expression for \vec{x} such that $\|\mathbf{A}\vec{x}\|^2$ is minimized. You may *not* use any tool from calculus to solve this problem—so avoid differentiation of any flavor.

For the optimal vector \vec{x} that you determine—that is, the vector

$$\vec{x} = \operatorname{argmin}_{\vec{x}} \|\mathbf{A}\vec{x}\|^2 \quad \text{subject to the constraint} \quad \|\vec{x}\|^2 = 1,$$

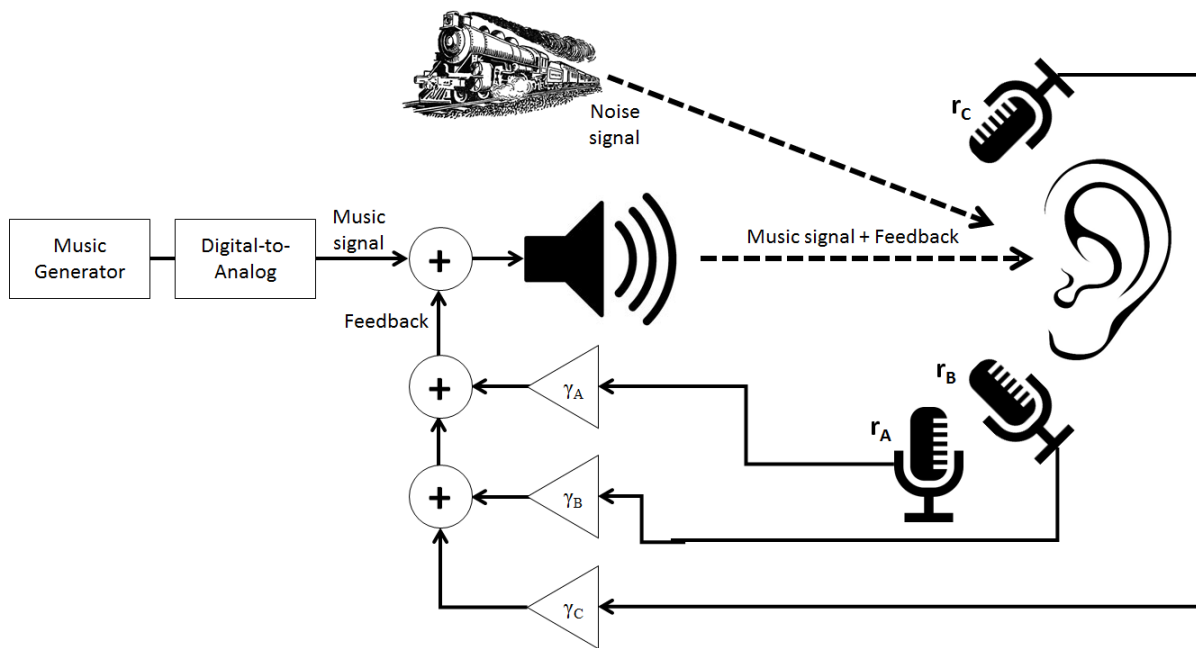
determine a simple, closed-form expression for the minimum value

$$\min_{\|\vec{x}\|=1} \|\mathbf{A}\vec{x}\| = \left\| \mathbf{A} \vec{x} \right\|.$$

3. Noise Cancelling Headphones

In this problem, we will explore a common design for noise cancellation, using noise-cancelling headphones as an example application. We will work with the model shown in the figure below.

A music signal is generated at a speaker and transmitted to the listener's ear. If there is noise in the environment (such as other people's voices, a train going by, or just any kind of noise), this noise signal will be



superimposed on the music signal and the listener will hear both. In order to cancel the noise, we will try to record the noise and subtract it directly from the transmitted signal, with the hope that we can achieve perfect cancellation of everything but the music. Since our system is imperfect, we'll have to solve a least squares problem.

The gain blocks marked by γ (Greek "gamma") represent scalar multiplication, and we will assume that they can take on any real number, positive or negative.

(a) First, consider a noise signal noted by \vec{n} ,

$$\vec{n} = \begin{bmatrix} n_1 \\ n_2 \\ n_3 \\ n_4 \\ n_5 \end{bmatrix}$$

We can use three microphones to record this signal, Mic A, Mic B, and Mic C. Each microphone records the noise, but they each have their own characteristics, so that they don't perfectly record the noise and that they are distinct recordings:

$$\vec{r}_A = \begin{bmatrix} a_1 \\ a_2 \\ a_3 \\ a_4 \\ a_5 \end{bmatrix}; \vec{r}_B = \begin{bmatrix} b_1 \\ b_2 \\ b_3 \\ b_4 \\ b_5 \end{bmatrix}; \vec{r}_C = \begin{bmatrix} c_1 \\ c_2 \\ c_3 \\ c_4 \\ c_5 \end{bmatrix}$$

We can arrange the recordings into a matrix \mathbf{R} and the microphone gains, γ , into a vector $\vec{\gamma}$

$$\mathbf{R} = \begin{bmatrix} a_1 & b_1 & c_1 \\ a_2 & b_2 & c_2 \\ a_3 & b_3 & c_3 \\ a_4 & b_4 & c_4 \\ a_5 & b_5 & c_5 \end{bmatrix}, \vec{\gamma} = \begin{bmatrix} \gamma_A \\ \gamma_B \\ \gamma_C \end{bmatrix}$$

For the system that is drawn in the figure above, and using matrix notation, write down the signal at the listener's ear. It should include,

- the music signal $\vec{m} = \begin{bmatrix} m_1 \\ m_2 \\ m_3 \\ m_4 \\ m_5 \end{bmatrix}$
- the noise signal \vec{n}
- the matrix of recorded noise signals \mathbf{R}
- the microphone gain vector $\vec{\gamma}$

You can assume that the microphones do not pick up the music signal.

- (b) Ideally, we would want to have a signal at the ear that matches the original music signal perfectly. In reality, this is not possible, so we will aim to minimize the effect of the noise. What quantity would we need to minimize to make sure this happens? Write your answer in terms of the matrix \mathbf{R} , the vector of mic gains $\vec{\gamma}$, and the noise vector \vec{n} .
- (c) We can solve minimization problems by the least squares method. In effect, if we have a problem, $\min_{\vec{x}} \|\mathbf{A}\vec{x} - \vec{b}\|$, then the \vec{x} that solves this problem is,

$$\vec{x} = (\mathbf{A}^T \mathbf{A})^{-1} \mathbf{A}^T \vec{b} \quad (3)$$

Implement this least squares method in the IPython Notebook helper function `doLeastSquares`.

- (d) For the given \vec{n} and the recordings, \vec{r}_A , \vec{r}_B , \vec{r}_C , below, report the γ 's that minimize the effect of noise.

$$\vec{n} = \begin{bmatrix} 0.10 \\ 0.37 \\ -0.45 \\ 0.068 \\ 0.036 \end{bmatrix}; \vec{r}_A = \begin{bmatrix} 0 \\ 0.11 \\ -0.31 \\ -0.012 \\ -0.018 \end{bmatrix}; \vec{r}_B = \begin{bmatrix} 0 \\ 0.22 \\ -0.20 \\ 0.080 \\ 0.056 \end{bmatrix}; \vec{r}_C = \begin{bmatrix} 0 \\ 0.37 \\ -0.44 \\ 0.065 \\ 0.038 \end{bmatrix}$$

The next few questions can be answered in the IPython notebook by running the associated cells.

- (e) We can use this least squares solution to train our algorithm for a given number of microphones and a training signal. Follow the instructions in the IPython notebook to load a music signal and some noise signals. Listen to the music signal and the two noise signals. Which ones are full of static and which ones are not.
- (f) **For Fun:**(ie. not graded) Use the IPython notebook to record the first noise signal using the `recordAmbientNoise` function and calculate a vector $\vec{\gamma}$. Create the noise cancellation signal by performing the multiplication $\mathbf{R}\vec{\gamma}$.
- (g) **For Fun:**(ie. not graded) Add the noise cancellation signal (with the correct sign) to the music signal to get the signal from the speaker and, finally, add the noise signal to the speaker signal. Play the noisy signal and the noise-cancelled signal. Can you hear a difference?
- (h) **For Fun:**(ie. not graded) Try adding the other noise signal to the music signal without re-calculating new values for $\vec{\gamma}$ (don't solve the least squares problem again). Add the noise-cancelling signal to your speaker signal and add the noise as well. Comment on the quality of the resulting noise-cancelled signal. Is it perfect or are there artifacts?

4. Image Analysis

Applications in medical imaging often require an analysis of images based on the pixels of the image. For instance, we might want to count the number of cells in a given sample. One way to do this is to “take a picture” of the cells and use the pixels to determine their locations and how many there are. Automatic detection of shape is useful in image classification as well (e.g. consider a robot trying to find out autonomously where a mug is in its field of vision).

Let us focus back on the medical imaging scenario. You are interested in finding the exact position and shape of a cell in an image. You will do this by finding the equation of the circle or ellipse that bounds the cell relative to a given coordinate system in the image. Your collaborator uses edge detection techniques to find a bunch of points that are approximately along the edge of the cell. We assume that the origin is in the center of the image with standard axes and collect the following points: $(0.3, -0.69)$, $(0.5, 0.87)$, $(0.9, -0.86)$, $(1, 0.88)$, $(1.2, -0.82)$, $(1.5, 0.64)$, $(1.8, 0)$.

Recall that a quadratic equation of the form

$$ax^2 + bxy + cy^2 + dx + ey = 1$$

can be used to represent an ellipse (if $b^2 - 4ac < 0$), and a quadratic equation of the form

$$a(x^2 + y^2) + dx + ey = 1$$

is a circle if $d^2 + e^2 - 4a > 0$. The circle has fewer parameters.

- How can you find the equation of a circle that surrounds the cell? First, provide a setup and formulate a set of matrix equations to do this, i.e. an equation of the form $\mathbf{A}\vec{x} = \vec{b} + \vec{e}$ where you attempt to find the unknown coefficients a , d , and e from your points and \vec{e} is the error vector. *Hint: $x^2 + y^2$, x , and y can be thought of as variables calculated from your data points*
- How can you find the equation of an ellipse that surrounds the cell? Provide a setup and formulate a set of matrix equations similar to in part a.
- In the IPython notebook, write a short program to fit a circle using these points. If you model your system of equations as $\mathbf{A}\vec{x} = \vec{b} + \vec{e}$, where \vec{e} is the error vector and the number of data points is N , what is $\frac{\|\vec{e}\|}{N}$? Plot your points and the best fit circle in IPython.
- Write a short program in IPython to fit an ellipse using these points. If you model your system of equations as $\mathbf{A}\vec{x} = \vec{b} + \vec{e}$, where \vec{e} is the error vector and the number of data points is N , what is $\frac{\|\vec{e}\|}{N}$? Plot your points and the best fit ellipse in IPython. How does this error compare to the one in the previous subpart? Which technique is better?

5. Pollster: Regularized Least Squares

Least squares techniques are useful for many different kinds of prediction problems (also called *regression*). Here, we’ll explore how least squares can be used for polling prediction.¹

Your job to predict how each county will vote, either for candidate A or candidate B. To do this you will build a linear model that predicts how each county will vote based on the (un)importance of several topics (the economy, healthcare, education, pineapple pizza, etc). Each topic is graded on an importance scale where a negative value means that the topic is not important and a positive value means that the topic is important.

¹The core ideas we learned in class have been extensively further developed—if you’re interested you **should take** EE127A (convex optimization), CS189 (machine learning), EE221 (linear systems)! In these classes you learn ideas that build off of the basic least squares problem for applications in finance, healthcare, advertising, image processing, control, and many other fields.

The magnitude of the number corresponds to how important/unimportant the topics are to members of the county.

Each county is represented by a “feature vector” of length 10. The value of each element of the vector captures the importance of that feature to the county. The dataset you are given contains the “feature vectors” for 100 counties as well as how those 100 counties will vote, a value of +1 if the county votes for candidate A and a value of −1 if the county votes for candidate B. We want you to find a “good” linear model using the first 90 counties (this is known as training data) and test the “good”ness of your model on the remaining 10 counties (testing data). Your job in this problem is to find the “good” linear model, with weights α_i , using the training data’s “feature vectors”, \vec{f} , and voting decision, b , to minimize the total prediction error of the testing data.

$$b = \alpha_1 f_1 + \alpha_2 f_2 + \dots + \alpha_{10} f_{10} = \vec{\alpha}^T \vec{f} \quad (4)$$

- Using only the training data**, set up the least squares problem ($\mathbf{A}\vec{x} = \vec{b}$). How are \mathbf{A} , \vec{x} , and \vec{b} constructed? What are the dimensions of \mathbf{A} , \vec{x} and \vec{b} ?
- Using IPython, build these matrices and solve for \hat{x} using linear least squares with the provided function `doLeastSquares(A, b)`. Evaluate what the total prediction error is on the **training** data using your linear model (i.e. the length of the error vector, $\|\vec{e}\| = \|\vec{b} - \mathbf{A}\hat{x}\|$). Also, evaluate what the total prediction error is on the **testing** data using your linear model.
- A real life problem when building models can be the data itself. In a country with two very polarizing candidates, knowing how a county feels about one topic allows us to predict how important they consider *every* topic. This issue makes the columns of \mathbf{A} almost linearly dependent—something we observed in the image stitching homework problem and in the imaging lab! Let us analyze it further by looking at the eigenvalues of $\mathbf{A}^T \mathbf{A}$ denoted λ_i . Show that the total prediction error

$$\|\vec{e}\| = \|\vec{b} - \mathbf{A}\hat{x}\| = \left\| \vec{b} - \mathbf{A} \left(\frac{\beta_1}{\lambda_1} \vec{v}_1 + \frac{\beta_2}{\lambda_2} \vec{v}_2 + \dots + \frac{\beta_N}{\lambda_N} \vec{v}_N \right) \right\|. \quad (5)$$

The β_i are the coordinates of the vector $\mathbf{A}^T \vec{b}$ in the eigenbasis of $(\mathbf{A}^T \mathbf{A})^{-1}$,

$$\mathbf{A}^T \vec{b} = \beta_1 \vec{v}_1 + \beta_2 \vec{v}_2 + \dots + \beta_N \vec{v}_N \quad (6)$$

(Hint: Consider how the eigenbasis for $(\mathbf{A}^T \mathbf{A})^{-1}$ is related to the eigenbasis of $\mathbf{A}^T \mathbf{A}$). The issue described above will make some of the $\lambda_i \approx 0$. What happens to the total prediction error when there exist eigenvalues $\lambda_i \approx 0$? For this problem assume all eigenvalues are distinct and unique.

- In IPython, plot the eigenvalues of $\mathbf{A}^T \mathbf{A}$. Do we encounter the problem described in part (c)?
- There are many solutions to this issue, but a common one involves including prior knowledge. We introduce now a value γ which we will use to try and rectify the error from part (c). We are given that the weights of our linear model tend to be small (close to zero). Written in equation form,

$$\sqrt{\gamma} \alpha_1 = 0 \dots \sqrt{\gamma} \alpha_{10} = 0 \quad (7)$$

and as a matrix,

$$\sqrt{\gamma} \mathbf{I} \vec{\alpha} = \vec{0}. \quad (8)$$

Let us concatenate the new equations to the bottom of our matrix equation, $\mathbf{A} \vec{x} = \vec{b}$, from the previous parts to create the augmented matrix, $\tilde{\mathbf{A}}$, and the augmented vector, $\tilde{\vec{b}}$. Our modified matrix equations will be,

$$\tilde{\mathbf{A}} \vec{x} = \tilde{\vec{b}} \rightarrow \begin{bmatrix} \mathbf{A} \\ \sqrt{\gamma} \mathbf{I} \end{bmatrix} \vec{x} = \begin{bmatrix} \vec{b} \\ \vec{0} \end{bmatrix} \quad (9)$$

What are the dimensions of $\tilde{\mathbf{A}}$, $\tilde{\vec{b}}$? Using IPython, create these matrices using `np.concatenate`.

- (f) Using the solution to the linear least squares problem derived in class, show that solution to the modified linear least squares problem is

$$\vec{\hat{x}} = (\tilde{\mathbf{A}}^T \tilde{\mathbf{A}})^{-1} \tilde{\mathbf{A}}^T \tilde{\vec{b}} = (\mathbf{A}^T \mathbf{A} + \gamma \mathbf{I})^{-1} \mathbf{A}^T \vec{b}. \quad (10)$$

Hint: Matrix-matrix multiplication can be handled in blocks! However, there are restrictions on the dimensions of the matrices when written in this block format, certain dimensions must agree. In this setting, $\mathbf{W} \in \mathbb{R}^{n \times m}$, $\mathbf{Y} \in \mathbb{R}^{m \times l}$, $\mathbf{X} \in \mathbb{R}^{n \times p}$, and $\mathbf{Z} \in \mathbb{R}^{p \times l}$. User beware: Remember that matrix-matrix multiplication does not, in general, commute.

$$\begin{bmatrix} \mathbf{W} & \mathbf{X} \end{bmatrix} \begin{bmatrix} \mathbf{Y} \\ \mathbf{Z} \end{bmatrix} = \mathbf{WY} + \mathbf{XZ}$$

- (g) Show that the new total prediction error

$$\|\vec{e}\| = \|\vec{b} - \mathbf{A} \vec{\hat{x}}\| = \left\| \vec{b} - \mathbf{A} \left(\frac{\beta_1}{\lambda_1 + \gamma} \vec{v}_1 + \frac{\beta_2}{\lambda_2 + \gamma} \vec{v}_2 + \dots + \frac{\beta_N}{\lambda_N + \gamma} \vec{v}_N \right) \right\|, \quad (11)$$

using our modified least squares solution, $\vec{\hat{x}} = (\mathbf{A}^T \mathbf{A} + \gamma \mathbf{I})^{-1} \mathbf{A}^T \vec{b}$, the eigenvalues λ_i of $\mathbf{A}^T \mathbf{A}$, and the eigenvectors \vec{v}_i of $(\mathbf{A}^T \mathbf{A} + \gamma \mathbf{I})^{-1}$. The β_i are the coordinates of the vector $\mathbf{A}^T \vec{b}$ in the eigenbasis of $(\mathbf{A}^T \mathbf{A} + \gamma \mathbf{I})^{-1}$,

$$\mathbf{A}^T \vec{b} = \beta_1 \vec{v}_1 + \beta_2 \vec{v}_2 + \dots + \beta_N \vec{v}_N \quad (12)$$

(Hint: Consider how the eigenbasis for $\mathbf{A}^T \mathbf{A} + \gamma \mathbf{I}$ is related to the eigenbasis of $\mathbf{A}^T \mathbf{A}$. Think about diagonalization of $\mathbf{A}^T \mathbf{A}$.) For this problem assume all eigenvalues are distinct and unique.

- (h) In IPython, in a **single plot** display the eigenvalues of $\mathbf{A}^T \mathbf{A} + \gamma \mathbf{I}$ for several values of γ (e.g. 0, 10, 100). What does this do to the eigenvalues of $\mathbf{A}^T \mathbf{A}$?
- (i) In IPython, let us now find the “best” γ to improve our testing total prediction error. Evaluate the total testing prediction error using different values of γ (use the list in the IPython Notebook). What is the best choice of γ (ie. which gamma minimizes the total testing prediction error)? How does the total testing prediction error using modified least squares with the best choice of γ compare with the total testing prediction error from part (b)?

6. OMP Exercise

- (a) Suppose we have a vector $\vec{x} \in \mathbb{R}^4$. We take 3 measurements of it, $b_1 = \vec{m}_1^T \vec{x} = 4$, $b_2 = \vec{m}_2^T \vec{x} = 6$, and $b_3 = \vec{m}_3^T \vec{x} = 3$, where \vec{m}_1 , \vec{m}_2 and \vec{m}_3 are some measurement vectors. In the general case when there are 4 unknowns in \vec{x} and we only have 3 measurements, i.e. system is underdetermined, it is not possible to solve for \vec{x} using Gaussian Elimination. Furthermore, there could be noise in the measurements. However, in this case, we are told that \vec{x} is sparse and has only 2 non-zero entries. In particular,

$$\begin{aligned} \mathbf{M}\vec{x} &\approx \vec{b} \\ \begin{bmatrix} - & \vec{m}_1^T & - \\ - & \vec{m}_2^T & - \\ - & \vec{m}_3^T & - \end{bmatrix} \vec{x} &\approx \vec{b} \\ \begin{bmatrix} 1 & 1 & 0 & -1 \\ 1 & 0 & -1 & 0 \\ 0 & 1 & 1 & 1 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{bmatrix} &\approx \begin{bmatrix} 4 \\ 6 \\ 3 \end{bmatrix} \end{aligned}$$

where exactly 2 of x_1 to x_4 are non-zero. Use Orthogonal Matching Pursuit to estimate x_1 to x_4 . In this problem, since we are looking for a solution to the equation $\mathbf{M}\vec{x} = \vec{b}$, we will not cross correlate \vec{b} with the columns of \mathbf{M} , instead we will just compute the inner product of \vec{b} with every column of \mathbf{M} .

- (b) We know that OMP works only when the vector \vec{x} is sparse, which means that it has very few non-zero entries. What if \vec{x} is not sparse in the standard basis but is only sparse in a different basis? What we can do is to change to the basis where \vec{x} is sparse, run OMP in that basis, and transform the result back into the standard basis.

- Suppose we have a $m \times n$ measurement matrix \mathbf{M} and a vector of measurements $\vec{b} \in \mathbb{R}^m$ where $\mathbf{M}\vec{x} = \vec{b}$ and we want to find $\vec{x} \in \mathbb{R}^n$. The basis that \vec{x} is sparse in is defined by basis vectors $\vec{a}_1 \cdots \vec{a}_n$, and we define:

$$\mathbf{A} = \begin{bmatrix} | & & | \\ \vec{a}_1 & \cdots & \vec{a}_n \\ | & & | \end{bmatrix}$$

such that $\vec{x} = \mathbf{A}\vec{x}_a$ and that \vec{x}_a is sparse.

- Suppose that we have an OMP function that can compute \vec{x}' for $\mathbf{M}'\vec{x}' = \vec{b}'$ **only when \vec{x}' is sparse**: $\vec{x}' = \text{OMP}(\mathbf{M}', \vec{b}')$.

Assuming that the **change of basis does not significantly affect the orthogonality of vectors**, describe how you would compute \vec{x} using the function OMP using necessary equations.

7. Sparse Imaging

Recall the imaging lab where we projected masks on an object to scan it to our computer using a single pixel measurement device, that is, a photoresistor. In that lab, we were scanning a 30×40 image having 1200 pixels. In order to recover the image, we took exactly 1200 measurements because we wanted our ‘measurement matrix’ to be invertible.

However, we saw that an iterative algorithm that does “matching and peeling” can enable reconstruction of a sparse vector while reducing the number of samples that need to be taken from it. In the case of imaging, the idea of sparsity corresponds to most parts of the image being black with only a small number of light pixels. In these cases, we can reduce the overall number of samples necessary. This would reduce the time required for scanning the image. (This is a real-world concern for things like MRI where people have to stay still while being imaged.)

In this problem, we have a 2D image I of size 91×120 pixels for a total of 10920 pixels. The image is made up of mostly black pixels except for 476 of them that are white.

Although the imaging illumination masks we used in the lab consisted of zeros and ones, in this question, we are going to have masks with real values — i.e. the light intensity is going to vary in a controlled way. Say that we have an imaging mask M_0 of size 91×120 . The measurements using the solar cell using this imaging mask can be represented as follows.

First, let us vectorize our image to \vec{i} which is a column vector of length 10920. Likewise, let us vectorize the mask M_0 to \vec{m}_0 which is a column vector of length 10920. Then the measurement using M_0 can be represented as

$$b_0 = \vec{m}_0^T \vec{i}.$$

Say we have a total of K measurements, each taken with a different illumination mask. Then, these measurements can collectively be represented as

$$\vec{b} = \mathbf{A} \vec{i},$$

where \mathbf{A} is an $K \times 10920$ size matrix whose rows contain the vectorized forms of the illumination masks, that is

$$\mathbf{A} = \begin{bmatrix} \vec{m}_1^T \\ \vec{m}_2^T \\ \vdots \\ \vec{m}_K^T \end{bmatrix}.$$

To show that we can reduce the number of samples necessary to recover the sparse image I , we are going to only generate 6500 masks. The columns of \mathbf{A} are going to be approximately uncorrelated with each other. The following question refers to the part of IPython notebook file accompanying this homework related to this question.

- (a) In the IPython notebook, we call a function `simulate` that generates masks and the measurements. You can see the masks and the measurements in the IPython notebook file. Complete the function `OMP` that does the OMP algorithm described in lecture.

Remark: Note that this remark is not important for solving this problem; it is about how such measurements could be implemented in our lab setting. When you look at the vector `measurements` you will see that it has zero average value. Likewise, the columns of the matrix containing the masks \mathbf{A} also have zero average value. To satisfy these conditions, they need to have negative values. However, in an imaging system, we cannot project negative light. One way to get around this problem is to find the smallest value of the matrix \mathbf{A} and subtract it from all entries of \mathbf{A} to get the actual illumination masks. This will yield masks with positive values, hence we can project them using our real-world projector. After obtaining the readings using these masks, we can remove their average value from the readings to get measurements as if we had multiplied the image using the matrix \mathbf{A} .

- (b) Run the code `rec = OMP((height, width), sparsity, measurements, A)` and see the image being correctly reconstructed from a number of samples smaller than the number of pixels of your figure. What is the image?
- (c) **PRACTICE:** We have supplied code that reads a PNG file containing a sparse image, takes measurements, and performs OMP to recover it. An example input image file is also supplied together with the code. Using `smiley.png`, generate an image of size 91×120 pixels of sparsity less than 400 and recover it using OMP with 6500 measurements.

You can answer the following parts of this question in very general terms. Try reducing the number of measurements. Does the algorithm start to fail in recovering your sparse image? Why do you think it fails?

8. Homework Process and Study Group

Who else did you work with on this homework? List names and student ID's. (In case of homework party, you can also just describe the group.) How did you work on this homework?