This homework is due Monday January 25, 2016, at Noon.

1. Homework process and study group

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

Solution: I worked on this homework with...

I first worked by myself for 2 hours, but got stuck on Problem 5 so I went to office hours on...
Then I went to homework party for a few hours, where I finished the homework.

2. Circuits and Gaussian Elimination

![Example Circuit](image)

Figure 1: Example Circuit

(a) Find a system of linear equations that could be solved to find the node voltages.

Solution:

\[
V_1 = V \\
\frac{V_2 - V_1}{R_1} + \frac{V_2}{R_2} + \frac{V_2 - V_3}{R_3} = 0 \\
\frac{V_3 - V_2}{R_3} + \frac{V_3}{R_4} + \frac{V_3 - V_C}{R_5} = 0
\]

Also, since in steady state no current flows through a capacitor, the current through \(R_5\) must be zero and so there is no voltage drop across it and so \(V_C = V_3\). This means that we can get rid of the third term of the last equation.

(b) Given that the component values are \(R_1 = 500\, \Omega\), \(R_2 = 3\, k\Omega\), \(R_3 = 1\, k\Omega\), \(R_4 = 2\, k\Omega\), and \(R_5 = 4\, k\Omega\), solve the circuit equations using Gaussian elimination.

Solution: With the provided component values, the system of equations is:

\[
\frac{V_2 - V_1}{500} + \frac{V_2}{3000} + \frac{V_2 - V_3}{1000} = \frac{V_2}{300} - \frac{V_1}{500} - \frac{V_3}{1000} = 0 \\
\frac{V_3 - V_2}{1000} + \frac{V_3}{2000} = \frac{3V_3}{2000} - \frac{V_2}{1000} = 0
\]
The augmented matrix for the above system of equations is:

\[
\begin{pmatrix}
1 & 0 & 0 & V \\
\frac{1}{500} & \frac{1}{1000} & -\frac{1}{2000} & 0 \\
0 & -\frac{1}{1000} & \frac{1}{2000} & 0
\end{pmatrix}
\]

This row-reduces down to:

\[
\begin{pmatrix}
1 & 0 & 0 & V \\
0 & 1 & 0 & \frac{3}{4}V \\
0 & 0 & 1 & \frac{1}{2}V
\end{pmatrix}
\]

(c) What’s the voltage \( V_C \) across the capacitor?

**Solution:** \( V_C = V_3 = \frac{1}{2}V \)

(d) How would you check your work? Do so.

**Solution:** Make sure KCL is satisfied at every node. For example, to check that KCL is satisfied at \( V_2 \), use Ohm’s Law to calculate the three currents flowing out of the node and make sure they balance:

\[
0 = \frac{V_2 - V_1}{R_1} + \frac{V_2}{R_2} + \frac{V_2 - V_3}{R_3}
\]

\[
= \frac{V}{4} + \frac{3V}{4} + \frac{V}{4}
\]

\[
= -2V + V + V = 0
\]

This is effectively just plugging your answers back into your original system of equations to make sure they work out. (Note that you can’t do this at \( V_1 \) because the current through a voltage source can be whatever is necessary to create the desired voltage.)

3. **Solving Recurrence Relations**

For this problem, we’ll work with a sequence defined by the following recurrence relation, where \( S[n] \) is the \( n \)th number in the sequence:

\[
S[n+1] = 3S[n] - 2S[n-1]
\]

\[
S[0] = 0
\]

\[
S[1] = 1
\]

You can probably see how this could be computed recursively or iteratively, but let’s try a linear-algebraic approach and see where it takes us.

(a) Starting from the definition of the sequence, determine the matrix \( A \) to calculate \( S[n+1] \) such that:

\[
A \begin{pmatrix}
S[n] \\
S[n-1]
\end{pmatrix} = \begin{pmatrix}
S[n+1] \\
S[n]
\end{pmatrix}
\]

**Solution:**

\[
A = \begin{pmatrix}
3 & -2 \\
1 & 0
\end{pmatrix}
\]
(b) Find the eigenvalues $\lambda_+$ and $\lambda_-$ of the matrix $A$.

**Solution:** Derive the characteristic polynomial of this matrix by solving $\det(A - \lambda I) = 0$. At this point, we can use the 2x2 determinant formula to find the characteristic polynomial $2 - \lambda (3 - \lambda) = 2 - 3\lambda + \lambda^2 = (1 - \lambda)(2 - \lambda)$ and set this to zero. So the eigenvalues are $\lambda_+ = 2$ and $\lambda_- = 1$.

(c) Find the eigenvectors associated with $\lambda_+$ and $\lambda_-$ from above.

**Solution:**

We know that eigenvalues $\lambda$ satisfy the relation $Av = \lambda v$, so:

$$
\begin{pmatrix}
3 & -2 \\
1 & 0
\end{pmatrix}
\begin{pmatrix}
v_1 \\
v_2
\end{pmatrix}
= \lambda
\begin{pmatrix}
v_1 \\
v_2
\end{pmatrix}
$$

Reading off the second equation (using both equations will just give back the characteristic equation from above - try it!), we get:

$$v_1 = \lambda v_2$$

Notice that if $v_1 = S[n]$ and $v_2 = S[n - 1]$, then $\lambda$ is the ratio of successive items in the sequence! (Isn’t that sweet? Do you think this is something special for our case of this particular recurrence or do you think it happens generally? Think about the structure of the second row in such $A$ matrices . . .)

This implies:

$$v =
\begin{pmatrix}
v_1 \\
v_2
\end{pmatrix}
= \begin{pmatrix}
\lambda v_2 \\
v_2
\end{pmatrix}
= v_2
\begin{pmatrix}
\lambda \\
1
\end{pmatrix}
$$

Since $v_2$ is just an arbitrary constant:

$$
\begin{pmatrix}
\lambda \\
1
\end{pmatrix}
$$

Reading off the solutions for both eigenvalues ($\lambda_+ = 2$ and $\lambda_- = 1$), we get:

$$v_+ =
\begin{pmatrix}
2 \\
1
\end{pmatrix}
$$

$$v_- =
\begin{pmatrix}
1 \\
1
\end{pmatrix}
$$

(d) How would you check your work? Do so. (Hint: Based on the eigenvalues you determined, how should this sequence behave for different initial conditions? Does it?)

**Solution:** We can check this using the two initial conditions given by the eigenvectors above, which were:
\[
\begin{pmatrix}
2 \\
1
\end{pmatrix}
\]
\[
\begin{pmatrix}
1 \\
1
\end{pmatrix}
\]

Let the eigenvector equal the initial conditions of the sequence, like so:

\[
\begin{pmatrix}
S[1] \\
S[0]
\end{pmatrix}
\]

Case 1 (\(v_+\)):
We have the sequence:

\[
\begin{align*}
S[0] &= 1 \\
S[1] &= 2 \\
S[2] &= 3S[1] - 2S[0] = 3(2) - 2(1) = 4 \\
\end{align*}
\]

As expected, the successive ratio between items is \(\lambda = 2\), so we know our positive eigenvalue and eigenvector are correct for case 1.

Case 2 (\(v_-\)):
We have the sequence:

\[
\begin{align*}
S[0] &= 1 \\
S[1] &= 1 \\
S[2] &= 3S[1] - 2S[0] = 3(1) - 2(1) = 1 \\
\end{align*}
\]

As expected, the successive ratio between items is \(\lambda = 1\), so we know our positive eigenvalue and eigenvector are correct for case 2. Case 2 is not particularly interesting - this initial condition just reaches a steady-state of persistence, since the recurrence relationship simply reproduces the initial values.

This foreshadows later on in the semester when we do state-space control, as we will see the significance of eigenvalue and eigenvector analysis again when matrices represent control systems.

(e) Using your previous results, diagonalize \(A\).

\textbf{Solution:}
Form the \(P\) matrix using the eigenvectors from the previous part:

\[
P = \begin{pmatrix}
v_+ & v_-
\end{pmatrix}
\]
\[ P = \begin{pmatrix} 2 & 1 \\ 1 & 1 \end{pmatrix} \]

We can also calculate the inverse \( P^{-1} \) here:

\[ P^{-1} = \begin{pmatrix} 1 & -1 \\ -1 & 2 \end{pmatrix} \]

The \( D \) matrix is just a diagonal matrix of the eigenvalues, which is:

\[ D = \begin{pmatrix} \lambda_+ & 0 \\ 0 & \lambda_- \end{pmatrix} \]

Finally, assemble it all together:

\[ A = PDP^{-1} \]

\[ A = \begin{pmatrix} 2 & 1 \\ 1 & 1 \end{pmatrix} \begin{pmatrix} 2 & 0 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & -1 \\ -1 & 2 \end{pmatrix} \]

(f) How can you use this new information to more efficiently compute any arbitrary \( S[n] \) without using any iteration or recursion? (Hint: if a matrix \( M = PDP^{-1} \) where \( D \) is a diagonal matrix, then think about \( M^2 = MM, M^3, \) and even \( M^n \).)

**Solution:**

First observe:

\[
\begin{pmatrix} S[1] \\ S[0] \end{pmatrix} = \begin{pmatrix} 1 \\ 0 \end{pmatrix} \\
\begin{pmatrix} S[2] \\ S[1] \end{pmatrix} = A \begin{pmatrix} 1 \\ 0 \end{pmatrix} \\
\begin{pmatrix} S[3] \\ S[2] \end{pmatrix} = A^2 \begin{pmatrix} 1 \\ 0 \end{pmatrix} \\
\begin{pmatrix} S[4] \\ S[3] \end{pmatrix} = A^3 \begin{pmatrix} 1 \\ 0 \end{pmatrix} \\
\begin{pmatrix} S[n] \\ S[n-1] \end{pmatrix} = A^{n-1} \begin{pmatrix} 1 \\ 0 \end{pmatrix}
\]

Since \( A^n = PDP^{-1} \) for a diagonal matrix (can you prove this using the rules of matrix multiplication and inverse matrices?), it follows that:
\[ A^n = PD^nP^{-1} \]

\[ A^n = \begin{pmatrix} 2 & 1 \\ 1 & 1 \end{pmatrix} \begin{pmatrix} 2^n & 0 \\ 0 & 1^n \end{pmatrix} \begin{pmatrix} 1 & -1 \\ -1 & 2 \end{pmatrix} \]

\[ A^n = \begin{pmatrix} 2 & 1 \\ 1 & 1 \end{pmatrix} \begin{pmatrix} 2^n & 0 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & -1 \\ -1 & 2 \end{pmatrix} \]

\[ A^{n-1} = \begin{pmatrix} 2 & 1 \\ 1 & 1 \end{pmatrix} \begin{pmatrix} 2^{n-1} & 0 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & -1 \\ -1 & 2 \end{pmatrix} \]

\[ A^{n-1} = \begin{pmatrix} 2^n - 1 & 2 - 2^n \\ 2^{n-1} - 1 & 2 - 2^{n-1} \end{pmatrix} \]

Substituting this result into our original result for \( S[n] \):

\[
\begin{pmatrix} S[n] \\ S[n-1] \end{pmatrix} = A^{n-1} \begin{pmatrix} 1 \\ 0 \end{pmatrix} = \begin{pmatrix} 2^n - 1 & 2 - 2^n \\ 2^{n-1} - 1 & 2 - 2^{n-1} \end{pmatrix} \begin{pmatrix} 1 \\ 0 \end{pmatrix}
\]

\[ (g) \] Finally, using your results from above, derive a closed-form expression with no summations, no recursion, and no matrix multiplications for \( S[n] \).

\textbf{Solution:} Starting off from the first line from the previous derivation:

\[
\begin{pmatrix} S[n] \\ S[n-1] \end{pmatrix} = \begin{pmatrix} 2^n - 1 & 2 - 2^n \\ 2^{n-1} - 1 & 2 - 2^{n-1} \end{pmatrix} \begin{pmatrix} 1 \\ 0 \end{pmatrix}
\]

Reading off the first row of the matrix equation to solve for \( S[n] \):

\[ S[n] = (2^n - 1)1 + (2 - 2^n)0 \]

\[ S[n] = 2^n - 1 \]

4. Show it

(a) Suppose \( \lambda_1, \ldots, \lambda_m \) are distinct eigenvalues of a matrix \( T \) and \( \vec{v}_1, \ldots, \vec{v}_m \) are the corresponding eigenvectors. Show that \( \vec{v}_1, \ldots, \vec{v}_m \) must be linearly independent.

\textbf{Solution:} Linear independence is fundamentally a concept that is about not being linearly dependent. Because of its nature, proofs by contradiction are the most natural way to proceed in showing linear independence from scratch. (As opposed to showing that some vectors are linearly independent because they satisfy another known set of sufficient conditions for linear independence.) This solution is in considerable detail — more than we expect from you.

So, we begin with assuming the statement that we actually believe is false.

Assume that \( \vec{v}_1, \ldots, \vec{v}_m \) are not linearly independent. This means that they are linearly dependent.
Now we have a tactical choice. Which linear dependence condition do we want? For proofs by contradiction, we tend to want the strongest possible condition — because its strength makes it rigid and hence easier to break later in the proof.

So, we will go with the idea that if they are linearly dependent, then there must be the smallest $k$ such that $\mathbf{v}_k$ is a linear combination of $\mathbf{v}_i$ for $i$ from 1 to $k - 1$. But the set $\{\mathbf{v}_i\}_{i=1}^{k-1}$ is linearly independent. Such a smallest $k$ must exist because the sets are nested and are linearly independent when $k = 1$ (unless the first vector is itself the zero vector, which is not allowed as an eigenvector), and are linearly dependent (by assumption) when $k = m$. So it must switch somewhere and we call that minimal number $k$.

So $\mathbf{v}_k = c_1 \mathbf{v}_1 + \cdots + c_{k-1} \mathbf{v}_{k-1}$, where at least one $c_k$ is nonzero.

This is something pretty strong. And now we want to force a contradiction. We clearly have to use $T$ somehow, so we might as well multiply both sides by it.

Then we have:

$$T \mathbf{v}_k = \lambda_k \mathbf{v}_k = \lambda_k (c_1 \mathbf{v}_1 + \cdots + c_{k-1} \mathbf{v}_{k-1})$$

$$= T (c_1 \mathbf{v}_1 + \cdots + c_{k-1} \mathbf{v}_{k-1}) = \lambda_1 c_1 \mathbf{v}_1 + \cdots + \lambda_{k-1} c_{k-1} \mathbf{v}_{k-1}$$

Now, we have two expressions that are equal to each other and we can go in for the “kill” and show a contradiction. Subtract one from the other and we have:

$$0 = (\lambda_k c_1 \mathbf{v}_1 + \cdots + \lambda_k c_{k-1} \mathbf{v}_{k-1}) - (\lambda_1 c_1 \mathbf{v}_1 + \cdots + \lambda_{k-1} c_{k-1} \mathbf{v}_{k-1})$$

$$= (\lambda_k - \lambda_1) c_1 \mathbf{v}_1 + \cdots + (\lambda_k - \lambda_{k-1}) c_{k-1} \mathbf{v}_{k-1}$$

So, now we know by the linear independence of $\mathbf{v}_1, \cdots, \mathbf{v}_{k-1}$ that all of the coefficients have to be zero. We also know that at least one of the $c_j$ is not zero. So consider that $j$. Then $(\lambda_k - \lambda_j) c_j = 0$ but $c_j \neq 0$. So $\lambda_k = \lambda_j$. But this is a contradiction since we assumed that all of the $\lambda_i$ were distinct.

Therefore, the one weak step in our chain of reasoning must be flawed and the assumption we made was wrong. So the $\{\mathbf{v}_i\}$ must be linearly independent.

(b) Show that if a vector $\mathbf{x}$ is in the null space of a matrix $A$, $\mathbf{x}$ must be orthogonal to all vectors in the column space of $A^*$. (Hint: does $\mathbf{x}$ have to be orthogonal to the columns of $A^*$? Remember, the $\ast$ means conjugate-transpose.)

**Solution:** Let $A$ be an $n$ by $m$ complex matrix, and let $\mathbf{x}$ be a vector of length $m$.

If we want to show that $\mathbf{x}$ is orthogonal to all vectors in the column space of $A^*$, it suffices to show that it is orthogonal to the columns of $A^*$. After all, the inner product is linear in each vector and so a sum of zeros is still zero.

Let $\mathbf{a}_1^T, \ldots, \mathbf{a}_n^T$ be the rows of $A$.

Because $\mathbf{x}$ is in the null space of $A$, we have $A \mathbf{x} = \mathbf{0}$. In other words, for each vector $\mathbf{a}_k$, we know by the nature of matrix multiplication that $\mathbf{a}_k^T \mathbf{x} = 0$.

We are almost there now. Let us just look from the other side.

Now let’s take a look at the matrix $A^*$: its column vectors are the complex conjugates of $\mathbf{a}_1, \ldots, \mathbf{a}_n$. Call these complex conjugates $\overline{\mathbf{c}_1}, \ldots, \overline{\mathbf{c}_n}$. Since taking a complex conjugate twice undoes itself, if we take the conjugate transpose of $\overline{\mathbf{c}_i}$ we just get $\overline{\overline{\mathbf{c}_i}} = \mathbf{c}_i^T$ — just the transpose.

So, we know that $\langle \overline{\mathbf{c}_i}, \mathbf{x}\rangle = \overline{\mathbf{c}_i^T \mathbf{x}} = \mathbf{c}_i^T \mathbf{x} = 0$ and we have shown orthogonality. Since $\mathbf{x}$ is orthogonal to all column vectors of $A^*$, we know that $\mathbf{x}$ is orthogonal to all vectors in the column space of $A^*$ as well. And we are done.
5. Getting Rich (or not) with Linear Regression

In this problem, we will use the linear least-squares models we learned previously for stock market analysis, and find the best model to do so. This problem is meant to brush up your skills with Linear Regression as well as IPython numpy, which we will use extensively throughout the course.

Use the provided ipython notebook file accompanying this homework.

(a) How many weeks of data have we given to you in nasdaq1.csv? Plot it. **Solution:** 86 weeks

(b) What are the coefficients of the linear regression model for the data in the previous part? And what is the norm of the residual error? Plot the straight-line model together with the data.

**Solution:**

For linear regression, we are trying to approximate a vector using the form of $ax_i + b = y_i$. As such, we are searching for the two coefficients $a$ and $b$.

We setup a matrix that consists of 1s (to find $b$) and $x_i$ (to find $a$) and apply standard linear least squares.

$$[a, b] = [12.61690943, 5143.1091197]$$

To calculate the error norm, we first find the error vector by subtracting the vector we wish to approximate $\vec{b}$ and our approximation $\vec{y}$. As such, the error vector is $\vec{e} = \vec{b} - \vec{y}$ and then you take its norm.

Error Norm: 1228.59364581

For plot, please look at ipython solution notebook

(c) nasdaq5.csv contains similar data, but since January 2010. Plot the NASDAQ composite index over the past 5 years along with the prediction of your model and calculate the norm of the error vector. What is the norm of the error vector for 5 years of stock data when predicted using the linear regression model found using nasdaq1.csv?

**Solution:**

Error Norm: 5816.24572278

For plots, please look at ipython solution notebook

(d) Repeat parts b) and c), but use a quadratic fitting model instead of a linear one. Use the 86 weeks of NASDAQ data to produce a model, find the coefficients for the quadratic regression, and extrapolate the model to 5 years of data. What is the norm of the error vector for both the 5 years of stock data and the 86 weeks of stock data? Plot the quadratic model together with the data for both 86 weeks and 5 years.

**Solution:**

For quadratic regression, we are trying to approximate a vector using the form of $ax_i^2 + bx_i + c = y_i$. As such, we are searching for the three coefficients $a,b$, and $c$.

We setup a matrix that consists of 1s (to find $c$) and $x_i$ (to find $b$) and $x_i^2$ (to find $a$) and apply standard linear least squares.

$$[a, b, c] = [-2.64149121e-02, 1.03716419e+01, 5.11167537e+03]$$

86 weeks Error Norm: 1221.15484512

5 year Error Norm: 16611.3755783

Notice that the 86-week error got mildly better (we fitted a more complicated model and so the fit could only get better) but the generalization error on the 5-year data got worse. This is a classic warning sign of a worse model.

For plots, please look at ipython solution notebook
(e) Repeat parts b) and c) again, but this time use an exponential fitting model instead of a linear one. Using the 86 weeks of data, find the coefficients for the exponential regression, and extrapolate the model to 5 years of data. What is the norm of the error vector for both the 5 years of stock data and the 86 weeks of stock data? Plot the exponential model together with the data for both 86 weeks and 5 years.

Solution:
For logarithmic regression, we are trying to approximate a vector using the form of \( ax_i + b = \log(y_i) \), which is actually coming from looking to fit \( \beta x^i = y_i \) and taking logs of both sides to get it into linear form.
As such, we are searching for the two coefficients \( a \) and \( b \). Depending on which log base you worked in, the coefficient may be different.
Using Base 10, these are the coefficients you get:

\[
[a, b] = [ 1.19291371e-03, 3.71291756 ]
\]

For calculating the error norm using this model, do not forget to take your approximation and take it to the power of the base you used. The answer should look like \( \| \hat{b} - 10^b 10^{ax^i} \| \), if you used the base 10.
86 weeks Error Norm: 1242.54722354
5 year Error Norm: 4260.35797068
For plots, please look at ipython solution notebook

(f) Which model had the lowest error? Why do you think that is?

Solution: The answer to this question depends on which error we are discussing.
If we are comparing the 86 weeks error, then the one with the lowest error is the quadratic regression. Though, the quadratic regression is best by a slim amount and the better fit is largely due to just having more knobs to fiddle with — parameters to estimate.
If we are comparing the 5 year error, then the one with the lowest error is the logarithmic regression. This error is different from the 86 weeks error because we did not perform least squares of the 5 year dataset. We used the 86 week dataset to find the least squares coefficients and then extrapolated the stock prices for the 5 year dataset. As such, the regression that would work best is the one that captures something closer to the true underlying model because it would hold outside the range of the 86 weeks.
Hence, the underlying model is probably closer to exponential and the logarithmic regression captures it the best. (This makes sense given that people in finance talk about growth rate using interest rates and exponential behavior.)

(g) Can these models be used to predict future behavior? Why or why not?

Solution:
This is a philosophical question and the right answer here is that we don’t know enough to answer this or even to make a reasoned guess. The right thing to do would be to try to make a prediction into the future and to see how well it works. Do an experiment to verify.

Reasons for why they can:
- If the true underlying structure were captured by the regression and it does not change in the future, then we can predict future behavior since the structure won’t change.

Reasons for why they cannot:
- There is no guarantee that the underlying structure will not change in the future. As such, the estimated model derived from these regressions would be inaccurate at that point.
There is insufficient data to properly estimate the true underlying structure. As such, the estimated models would be inaccurate. Although this second reason is true, in engineering contexts it is often unsatisfying. We never have enough data to completely estimate the true underlying structure. We have to do the best we can with what we have got. This is one of the reasons why there are many interesting signal-processing techniques that deal with learning and prediction.

6. Your Own Problem

Write your own problem related to this week’s material and solve it. You may still work in groups to brainstorm problems, but each student must submit a unique problem. What is the problem? How to formulate it? How to solve it? What is the solution?