ECE 4803, Fall 2020

Homework #3

Due Thursday, September 10, at 9:30am

1. Prepare a one paragraph summary of what we talked about in class in the last week. I do not want just a bulleted list of topics, I want you to use complete sentences and establish context (Why is what we have learned relevant? How does it connect with other classes?). The more insight you give, the better.

2. Using the properties of an inner product, prove that the Pythagorean Theorem holds for any "induced norm". Specifically, show that if $\|\boldsymbol{x}\| = \sqrt{\langle \boldsymbol{x}, \boldsymbol{x} \rangle}$, then for any $\boldsymbol{x}, \boldsymbol{y}$ satisfying $\langle \boldsymbol{x}, \boldsymbol{y} \rangle = 0$, $\|\boldsymbol{x} + \boldsymbol{y}\|^2 = \|\boldsymbol{x}\|^2 + \|\boldsymbol{y}\|^2$.

3. Using the properties of an inner product, prove that the Cauchy-Schwarz inequality holds for any "induced norm".
   (a) Specifically, show that if $\|\boldsymbol{x}\| = \sqrt{\langle \boldsymbol{x}, \boldsymbol{x} \rangle}$, then for any $\boldsymbol{x}, \boldsymbol{y}$ we have $|\langle \boldsymbol{x}, \boldsymbol{y} \rangle| \leq \|\boldsymbol{x}\| \|\boldsymbol{y}\|$. [Hint: Consider writing $\boldsymbol{x}$ as a linear combination of $\boldsymbol{y}$ and the vector $\boldsymbol{z} = \boldsymbol{x} - \frac{\langle \boldsymbol{x}, \boldsymbol{y} \rangle}{\|\boldsymbol{y}\|^2} \boldsymbol{y}$. What is $\langle \boldsymbol{z}, \boldsymbol{y} \rangle$? What can we infer from the previous problem?]
   (b) *Optional: Show that equality holds if and only if $\boldsymbol{y} = a\boldsymbol{x}$ for some $a \in \mathbb{R}$.

4. Prove that if $\boldsymbol{A}$ and $\boldsymbol{B}$ are square $N \times N$ matrices, then if $\boldsymbol{AB} = \mathbf{I}$, we must also have $\boldsymbol{BA} = \mathbf{I}$. Some hints to help you get started:
   - An equivalent statement to $\boldsymbol{BA} = \mathbf{I}$ is that, for any $\boldsymbol{x} \in \mathbb{R}^N$, $\boldsymbol{BAx} = \boldsymbol{x}$.
   - Think about what can you say about $\mathcal{R}(\boldsymbol{AB})$.
   - Think about what you can say about the relationship between $\mathcal{R}(\boldsymbol{AB})$ and $\mathcal{R}(\boldsymbol{B})$.
   - Recall that for any $\boldsymbol{x} \in \mathcal{R}(\boldsymbol{B})$, we can write $\boldsymbol{x}$ as a linear combination of the columns of $\boldsymbol{B}$, i.e., $\boldsymbol{x} = \boldsymbol{Ba}$ for some vector $\boldsymbol{a} \in \mathbb{R}^N$.

5. Let $\boldsymbol{D}$ be a diagonal $R \times R$ matrix whose diagonal elements are positive. Show that the maximizer $\widehat{\boldsymbol{\beta}}$ to
$$\underset{\boldsymbol{\beta} \in \mathbb{R}^R}{\text{maximize}} \ \|\boldsymbol{D\beta}\|_2^2 \quad \text{subject to} \quad \|\boldsymbol{\beta}\|_2 = 1$$
   has a 1 in the entry corresponding to the largest diagonal element of $\boldsymbol{D}$, and is 0 elsewhere.

6. In the previous homework we encountered a problem where we were given noisy observations of a "bandlimited" signal that was assumed to have a Fourier series with coefficients that were zero beyond a certain point. This is not the most realistic model because it inherently assumes that the signal is composed only of *harmonic* sinusoids, i.e., sinusoids that have an integer number of cycles/periods within the observation window. Even a pure sinusoid, if it has a non-integer number of cycles within the observation window, will not be particularly well represented using a basis containing only harmonic sinusoids.

An alternative approach is to replace the model from before:

$$f(t) = \sum_{k=-B}^{B} \alpha_k e^{j2\pi kt},$$

with a model in which we take a finer "sampling" in the frequency domain:

$$f(t) = \sum_{k=-BS}^{BS} \alpha_k e^{j2\pi(k/S)t}, \tag{1}$$

By setting $S > 1$ our model now includes more frequencies than just those corresponding to harmonic sinusoids.

Suppose we observe M uniformly spaced samples between $[0, 1]$:

$$y[m] = f((m-1)/(M-1)), \quad m = 1, \dots, M. \tag{2}$$

(a) Write a Python script that takes a number of samples M and a dimension B and an oversampling factor S (that you may assume is an integer) as input, and computes an $M \times N$ matrix $\boldsymbol{A}$ such that when $\boldsymbol{A}$ is applied to the vector $\boldsymbol{\alpha}$ (as in (1)), it returns the sample values in (2).

(b) Set $M = 1024$, $B = 16$, and $S = 3$. Use your code from the previous part to form the matrix $\boldsymbol{A}$, and then compute the SVD of $\boldsymbol{A}$. Next construct a vector of samples $\boldsymbol{y}$ corresponding to the signal

$$g(t) = \cos(2\pi^2 t).$$

Compute $\widehat{\boldsymbol{\alpha}} = \boldsymbol{A}^\dagger \boldsymbol{y}$. Verify that $\boldsymbol{A}\widehat{\boldsymbol{\alpha}}$ provides a good reconstruction of $g(t)$.

(c) Show that this process might be highly unstable by constructing a seemingly small perturbation of $\boldsymbol{y}$ that leads to a very large change in $\boldsymbol{\alpha}$. (Note, most numerical linear algebra tools have built in safeguards that might prevent things from being quite as bad as predicted by the theory from our notes.)

(d) Does the perturbation found in part (c), which leads to a large error in $\boldsymbol{\alpha}$, also lead to a large error in the reconstruction provided by $\boldsymbol{A}\boldsymbol{\alpha}$? Explain why or why not?

7. In class (and the previous problem) we saw that when the smallest singular value σ_R of a matrix $\boldsymbol{A}$ gets very small, the least squares solution can be extremely sensitive to even small perturbations. One strategy for addressing this is via an approach called the "truncated SVD". In this case, rather than solving the least squares problem via the standard pseudo-inverse

$$\widehat{\boldsymbol{x}}_{\text{ls}} = \boldsymbol{V}\boldsymbol{\Sigma}^{-1}\boldsymbol{U}^{\mathrm{T}}\boldsymbol{y},$$

we instead use the modified formula

$$\widehat{\boldsymbol{x}}_{\text{trunc}} = \boldsymbol{V}_{R'}\boldsymbol{\Sigma}_{R'}^{-1}\boldsymbol{U}_{R'}^{\mathrm{T}}\boldsymbol{y}.$$

Here, $\boldsymbol{U}_{R'}$ and $\boldsymbol{V}_{R'}$ denotes the matrices formed using the first $R' \leq R$ columns of $\boldsymbol{U}$ and $\boldsymbol{V}$, and $\boldsymbol{\Sigma}_{R'}$ denotes the $R' \times R'$ matrix formed from the first R' columns and and rows of $\boldsymbol{\Sigma}$. We are essentially choosing to simply ignore some of the singular vectors corresponding to small singular values – this reduces our susceptibility to small perturbations, but has the cost of biasing us away from the true solution. Typically there is a "sweet spot" for R' that balances this tradeoff.

In this problem I would like you to explore using this idea to "stabilize" the reconstruction problem from problem 4(c) in Homework #2. Specifically, set $M = 21$ and form the system of equations $\boldsymbol{y} = \boldsymbol{X}\boldsymbol{a}$ required to interpolate a polynomial of degree 20. However, rather than computing $\widehat{\boldsymbol{a}} = \boldsymbol{X}^{\dagger}\boldsymbol{y} = \boldsymbol{X}^{-1}\boldsymbol{y}$, compute the SVD of $\boldsymbol{X}$ and then compute

$$\widehat{\boldsymbol{a}}_{\text{trunc}} = \boldsymbol{V}_{R'}\boldsymbol{\Sigma}_{R'}^{-1}\boldsymbol{U}_{R'}^{\mathrm{T}}\boldsymbol{y}$$

for choices of R' ranging from 1 all the way up to 21. Comment on which value of R' gives the best estimate of the original $f(x)$? How does this compare to the result you obtained in problem 4(c) in the last homework where you instead chose a smaller value of M? Look at the singular values of $\boldsymbol{X}$ (or the singular values for the $\boldsymbol{X}$ matrix formed when M is set to be smaller than 21) and comment if you observe anything interesting.