GEORGIA INSTITUTE OF TECHNOLOGY
SCHOOL of ELECTRICAL and COMPUTER ENGINEERING

ECE 6250 Fall 2018
Problem Set #9

Assigned: 7-Nov-18
Due Date: 14-Nov-18

As stated in the syllabus, unauthorized use of previous semester course materials is strictly prohibited in this course.

All assignments will be due at the begining of class on the due date.

PROBLEM 9.1:

Using your class notes, prepare a 1-2 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 things you have learned here or in other classes?). The more insight you give, the better.

PROBLEM 9.2:

(a) Recall the definition of the Frobenius norm of an $M \times N$ matrix:

$$\|\boldsymbol{A}\|_F = \left(\sum_{m=1}^{M} \sum_{n=1}^{N} |A[m,n]|^2 \right)^{1/2}.$$

Show that

$$\|\boldsymbol{A}\|_F^2 = \operatorname{trace}(\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A}) = \sum_{r=1}^{R} \sigma_r^2,$$

where R is the rank of $\boldsymbol{A}$ and the $\{\sigma_r\}$ are the singular values of $\boldsymbol{A}$.

(b) The *operator norm* (sometimes called the *spectral norm*) of an $M \times N$ matrix is

$$\|\boldsymbol{A}\| = \max_{\boldsymbol{x}\in\mathbb{R}^N,\ \|\boldsymbol{x}\|_2=1} \|\boldsymbol{A}\boldsymbol{x}\|_2.$$

(This matrix norm is so important, it doesn't even require a designation in its notation — if somebody says "matrix norm" and doesn't elaborate, this is what they mean.) Show that

$$\|\boldsymbol{A}\| = \sigma_1,$$

where σ_1 is the largest singular value of $\boldsymbol{A}$. For which $\boldsymbol{x}$ does

$$\|\boldsymbol{A}\boldsymbol{x}\|_2 = \|\boldsymbol{A}\| \cdot \|\boldsymbol{x}\|_2 \ \ ?$$

(c) Prove that $\|\boldsymbol{A}\| \leq \|\boldsymbol{A}\|_F$. Give an example of an $\boldsymbol{A}$ with $\|\boldsymbol{A}\| = \|\boldsymbol{A}\|_F$.

PROBLEM 9.3:

Let $f(\boldsymbol{x})$ be a functional on $\mathbb{R}^N$; that is is, $f(\cdot)$ maps a vector to a real number. Recall the definition of the *gradient* of f at $\boldsymbol{x}$:

$$\nabla_x f = \begin{bmatrix} \frac{\partial f}{\partial x_1} \\ \frac{\partial f}{\partial x_2} \\ \vdots \\ \frac{\partial f}{\partial x_N} \end{bmatrix},$$

where the x_i are the components of $\boldsymbol{x}$.

(a) Calculate the gradient for $f(\boldsymbol{x}) = \|\boldsymbol{x}\|_2^2$.

(b) Calculate the gradient for $f(\boldsymbol{x}) = \|\boldsymbol{A}\boldsymbol{x}\|_2^2$, where $\boldsymbol{A}$ is an $M \times N$ matrix.

(c) Calculate the gradient for $f(\boldsymbol{x}) = \boldsymbol{y}^{\mathrm{T}}\boldsymbol{A}\boldsymbol{x}$.

(d) A necessary condition for $\boldsymbol{x}_0$ to be a minimizer of $f(\boldsymbol{x})$ is that $\nabla_{\boldsymbol{x}=\boldsymbol{x}_0} f = \boldsymbol{0}$. Show that a minimizer $\hat{\boldsymbol{x}}$ of

$$\min_{\boldsymbol{x}} \ \|\boldsymbol{y} - \boldsymbol{A}\boldsymbol{x}\|_2^2$$

must obey the so-called normal equations

$$\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A}\boldsymbol{x} = \boldsymbol{A}^{\mathrm{T}}\boldsymbol{y}.$$

(e) Having $\nabla_{\boldsymbol{x}=\boldsymbol{x}_0} f = \boldsymbol{0}$ is also a sufficient condition for $\boldsymbol{x}_0$ to be a minimizer when $f(\boldsymbol{x})$ is *convex*. If f is twice differentiable, this is the same as the Hessian matrix

$$\boldsymbol{H}_{\boldsymbol{x}} = \begin{bmatrix} \frac{\partial^2 f}{\partial x_1^2} & \frac{\partial^2 f}{\partial x_1 \partial x_2} & \cdots & \frac{\partial^2 f}{\partial x_1 \partial x_N} \\ \frac{\partial^2 f}{\partial x_2 \partial x_1} & \frac{\partial^2 f}{\partial x_2^2} & \cdots & \frac{\partial^2 f}{\partial x_2 \partial x_N} \\ \vdots & & \ddots & \vdots \\ \frac{\partial^2 f}{\partial x_N \partial x_1} & \frac{\partial^2 f}{\partial x_1 \partial x_2} & \cdots & \frac{\partial^2 f}{\partial x_N^2} \end{bmatrix}$$

being symmetric positive semi-definite. Show that this is indeed the case for $f(\boldsymbol{x}) = \|\boldsymbol{y} - \boldsymbol{A}\boldsymbol{x}\|_2^2$.

(f) Show that the minimizer $\hat{\boldsymbol{x}}$ of

$$\min_{\boldsymbol{x}} \ \|\boldsymbol{y} - \boldsymbol{A}\boldsymbol{x}\|_2^2 + \delta\|\boldsymbol{x}\|_2^2$$

is always

$$\hat{\boldsymbol{x}} = \left(\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A} + \delta\mathbf{I}\right)^{-1}\boldsymbol{A}^{\mathrm{T}}\boldsymbol{y},$$

no matter what $\boldsymbol{A}$ is. Make sure to include an explanation of why $\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A} + \delta\mathbf{I}$ is always invertible when $\delta > 0$.

PROBLEM 9.4:
Download the file `blocksdeconv.mat`. This file contains the vectors:

- `x`: the 512 × 1 "blocks" signal
- `h`: a 30 × 1 boxcar filter
- `y`: a 541 × 1 vector of observations of h convolved with x
- `yn`: a noisy observation of y. The noise is iid Gaussian with standard deviation .01.

(a) Write a function which takes an vector $\boldsymbol{h}$ of length L and a number N, and returns the $M \times N$ (with $M = N + L - 1$) matrix $\boldsymbol{A}$ such that for any $\boldsymbol{x} \in \mathbb{R}^N$, $\boldsymbol{Ax}$ is the vector of non-zero values of $\boldsymbol{h}$ convolved with $\boldsymbol{x}$.

(b) Use MATLAB's `svd()` command to calculate the SVD of $\boldsymbol{A}$. What is the largest singular value? What is the smallest singular value? Calculate $\boldsymbol{A}^{\dagger}\boldsymbol{y}$ and plot it ($\boldsymbol{y}$ is the noise-free data).

(c) Apply $\boldsymbol{A}^{\dagger}$ to the noisy `yn`. Plot the result. Calculate the mean-sqaure error $\|\boldsymbol{x} - \hat{\boldsymbol{x}}\|_2^2$ and compare to the measurement error $\|\boldsymbol{y} - \boldsymbol{yn}\|_2^2$.

(d) Form an approximation to $\boldsymbol{A}$ by truncating the last q terms in the singular value decomposition:

$$\boldsymbol{A}' = \sum_{k=1}^{R-q} \sigma_k \boldsymbol{u}_k \boldsymbol{v}_k^{\mathrm{T}}.$$

Apply the new pseudo-inverse $\boldsymbol{A}'^{\dagger}$ to `yn` and plot the result. Try a number of different values of q, and choose the one which "looks best" to turn in (indicate the value of q used). Calculate the mean-square reconstruction error.

(e) Now form another approximate inverse using Tikhonov regularization. Try a number of different values for δ and choose the one which "looks best" to turn in (indicate the value of δ used). Calculate the mean-square reconstruction error.

(f) Summarize your findings by comparing the MSE in parts (c), (d), and (e). Also include the error of doing nothing: $\|\boldsymbol{x} - \boldsymbol{yn}'\|_2^2$ where $\boldsymbol{yn}'$ is the appropriate piece of $\boldsymbol{yn}$.

PROBLEM 9.5:
Let

$$\boldsymbol{A} = \begin{bmatrix} 1.01 & 0.99 \\ 0.99 & 0.98 \end{bmatrix}$$

(a) Find the eigenvalue decomposition of $\boldsymbol{A}$. Recall that λ is an eigenvalue of $\boldsymbol{A}$ if for some $u[1], u[2]$ (entries of the corresponding eigenvector) we have

$$\begin{aligned} (1.01 - \lambda)u[1] + 0.99u[2] &= 0 \\ .99u[1] + (0.98 - \lambda)u[2] &= 0. \end{aligned}$$

Another way of saying this is that we want the values of λ such that $\boldsymbol{A} - \lambda\mathbf{I}$ (where $\mathbf{I}$ is the 2×2 identity matrix) has a non-trivial null space — there is a nonzero vector $\boldsymbol{u}$ such that $(\boldsymbol{A} - \lambda\mathbf{I})\boldsymbol{u} = 0$. Yet another way of saying this is that we want the values of λ such that $\det(\boldsymbol{A} - \lambda\mathbf{I}) = 0$. Once you have found the two eigenvalues, you can solve the 2×2 systems of equations $\boldsymbol{A}\boldsymbol{u}_1 = \lambda_1\boldsymbol{u}_1$ and $\boldsymbol{A}\boldsymbol{u}_2 = \lambda_2\boldsymbol{u}_2$ for $\boldsymbol{u}_1$ and $\boldsymbol{u}_2$.

(b) If $\boldsymbol{y} = \begin{bmatrix} 1 & 1 \end{bmatrix}^{\mathrm{T}}$, determine the solution to $\boldsymbol{A}\boldsymbol{x} = \boldsymbol{y}$.

(c) Now let $\boldsymbol{y} = \begin{bmatrix} 1.1 & 1 \end{bmatrix}^{\mathrm{T}}$ and solve $\boldsymbol{A}\boldsymbol{x} = \boldsymbol{y}$. Comment on how the solution changed.

(d) Suppose we observe

$$\boldsymbol{y} = \boldsymbol{A}\boldsymbol{x} + \boldsymbol{e}$$

with $\|\boldsymbol{e}\|_2 = 1$. We form an estimate $\tilde{\boldsymbol{x}} = \boldsymbol{A}^{-1}\boldsymbol{y}$. Which vector $\boldsymbol{e}$ (over all error vectors with $\|\boldsymbol{e}\|_2 = 1$) yields the maximum error $\|\tilde{\boldsymbol{x}} - \boldsymbol{x}\|_2^2$?

(e) Which (unit) vector $\boldsymbol{e}$ yields the minimum error?

(f) (Optional) Suppose the components of $\boldsymbol{e}$ are iid Gaussian:

$$e[i] \sim \mathrm{Normal}(0, 1).$$

What is the mean-square error $\mathrm{E}[\|\tilde{\boldsymbol{x}} - \boldsymbol{x}\|_2^2]$?

(g) (Optional) Verify your answer to the previous part in MATLAB by taking $\boldsymbol{A}\boldsymbol{x} = \begin{bmatrix} 1 & 1 \end{bmatrix}^{\mathrm{T}}$, and then generating $10,000$ different realizations of $\boldsymbol{e}$ using the `randn` command, and then averaging the results. Turn in your code and the results of your computation.