EECS 127/227AT 3/8/2018
L. El Ghaoui

Midterm: Solutions

1. *Linear functions of images.* In this problem we consider linear functions of an image with 2×2 pixels shown below.

3	7
8	5

This given image can be represented as the 4-vector $\begin{bmatrix} 3 \\ 7 \\ 8 \\ 5 \end{bmatrix}$.

Each of the operations described below defines a linear transformation $y = f(x)$, where the 4-vector x represents the original image, and the 4-vector y represents the resulting or transformed image. For each of these operations, provide the 4×4 matrix A for which $y = Ax$. **Also** in each case, determine the rank of the matrix A.

(a) Reflect the original image x across the vertical (i.e. bottom-to-top) axis.

(b) Rotate the original image x clockwise $90°$.

Name: ________________________

SID: ________________________

(c) Rotate the original image x clockwise by $180°$.

(d) Set each pixel value y_i to be the average of the neighbors of pixel i in the original image. We define neighbors, to be the pixels immediately above and below and to the left and right. For the 2×2 matrix, every pixel has 2 neighbors.

Solution:

(a) For $y = Ax$, we have the reflection (across the vertical axis) matrix A:

$$\begin{bmatrix} 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} 3 \\ 7 \\ 8 \\ 5 \end{bmatrix} = \begin{bmatrix} 7 \\ 3 \\ 5 \\ 8 \end{bmatrix}$$

Where y is the flipped image x. Here $\text{rank}(A) = 4$ since all the columns are linearly independent.

(b) For $y = Ax$, we have the $90°$ rotation matrix A:

$$\begin{bmatrix} 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 1 & 0 & 0 \end{bmatrix} \begin{bmatrix} 3 \\ 7 \\ 8 \\ 5 \end{bmatrix} = \begin{bmatrix} 8 \\ 3 \\ 5 \\ 7 \end{bmatrix}$$

Where y is the rotated image x. Here $\text{rank}(A) = 4$ since all the columns are linearly independent.

(c) For $y = Ax$, we have the 180° rotation A:

$$\begin{bmatrix} 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} 3 \\ 7 \\ 8 \\ 5 \end{bmatrix} = \begin{bmatrix} 5 \\ 8 \\ 7 \\ 3 \end{bmatrix}$$

Where y is the rotated image x. Here $\text{rank}(A) = 4$ since all the columns are linearly independent.

(d) For $y = Ax$, we have the matrix A:

$$\begin{bmatrix} 0 & \frac{1}{2} & \frac{1}{2} & 0 \\ \frac{1}{2} & 0 & 0 & \frac{1}{2} \\ \frac{1}{2} & 0 & 0 & \frac{1}{2} \\ 0 & \frac{1}{2} & \frac{1}{2} & 0 \end{bmatrix} \begin{bmatrix} 3 \\ 7 \\ 8 \\ 5 \end{bmatrix} = \begin{bmatrix} 7.5 \\ 4 \\ 4 \\ 7.5 \end{bmatrix}$$

Where y_i is the average of the neighbors of pixel i. Here $\text{rank}(A) = 2$ since there are only 2 columns that are linearly independent.

Name: ______________________________

SID: ______________________________

2. *Fun with the SVD.* Consider the 4×3 matrix

$$A = \begin{bmatrix} a_1 & a_2 & a_3 \end{bmatrix} \tag{1}$$

where a_i for $i \in \{1, 2, 3\}$ form a set of *orthogonal* vectors satisfying $\|a_1\|_2 = 3$, $\|a_2\|_2 = 2$, $\|a_3\|_2 = 1$.

(a) What is the SVD of A? Express it as $A = USV^\top$, with S the diagonal matrix of singular values ordered in decreasing fashion, and explicitly describe U and V.

(b) Write A as a sum of 3 rank-one matrices.

(c) What is the dimension of the null space, $\dim(\text{null}(A))$?

Name: ________________________

SID: ________________________

(d) What is the rank of A, $\text{rank}(A)$? Provide an orthonormal basis for the range of A.

(e) Find the maximum "gain" of A (the amount that A can "expand" an input vectors ℓ_2 norm). More formally, what is the value of $\max_{x:\|x\|_2=1} \frac{\|Ax\|_2}{\|x\|_2}$?

(f) If I_3 denotes the 3×3 identity matrix, consider the matrix $\tilde{A} = \begin{bmatrix} A \\ I_3 \end{bmatrix} \in \mathbb{R}^{7\times 3}$? What are the singular values of $\tilde{A}$ (in terms of the singular values of A)?

Solution:

(a) The SVD of $A = USV^\top$. Due to the orthogonality of the a_i we have that

$$A^\top A = VS^2V = \begin{bmatrix} 9 & 0 & 0 \\ 0 & 4 & 0 \\ 0 & 0 & 1 \end{bmatrix} \tag{2}$$

Thus $V = I$ and $S = \mathrm{diag}(3, 2, 1)$. Finally we have that $U = AS^{-1}$ which becomes

$$A = \begin{bmatrix} \frac{a_1}{3} & \frac{a_2}{2} & \frac{a_3}{1} \end{bmatrix} \tag{3}$$

(b) If we let e_i denote the standard basis vectors in $\mathbb{R}^3$ then we immediately obtain that

$$A = a_1 e_1^\top + a_2 e_2^\top + a_3 e_3^\top. \tag{4}$$

Expanding out the SVD of the matrix is also a valid answer.

(c) From part (a) all of the singular values of the A are non-zero. So the dimension of the null space is 0. Alternatively, all the columns of A are orthogonal – so no (non-zero) linear combination of them can equal zero.

(d) The rank of A is simply the number of non-zero singular values. So $\mathrm{rank}(A) = 3$. The columns of U (defined above) provide an orthonormal basis for the range of A.

(e) As defined the maximum "gain" of A (also knowns as it spectral norm) is simply given by its largest singular value. Hence the maximum "gain" of A is 3.

(f) We have that $\tilde{A}^\top \tilde{A} = A^\top A + I_3 = V(S^2 + I_3)V^\top$. Hence if we denote σ_i as the singular values of A then the singular values of $\tilde{A}$ are $\tilde{\sigma}_i = \sqrt{\sigma_i^2 + 1}$ which are $\sqrt{10}, \sqrt{5}, \sqrt{2}$.

It's fine if they write either the algebraic expression or the explicit values for this question.

Name: __________________________

SID: __________________________

3. *Regression and Applications.* We first consider the regularized least-squares problem,

$$w_\lambda := \arg\min_w \|y - Xw\|_2^2 + \lambda\|w\|_2^2, \tag{5}$$

and subsequently an application to modeling time series. To begin, we investigate several fundamental properties of regression. Here, $X \in \mathbb{R}^{n,p}$ is the data matrix (with one data point per row), $y \in \mathbb{R}^n$ is the response vector, and $\lambda > 0$ is a "ridge" regularization parameter.

(a) Assume, only for this part, that $n < p$. Is $X^\top X$ invertible? Explain your reasoning.

(b) Now assume no relation between n and p. Is $X^\top X + \lambda I$ invertible? Explain your reasoning.

(c) Show that the solution to the full problem can be written as

$$w_\lambda = (X^\top X + \lambda I)^{-1} X^\top y \tag{6}$$

Name: ______________________

SID: ______________________

(d) Now suppose we would like find w that minimizes,

$$w = \arg\min_w \{\sum_{i=1}^{k} \lambda_i \|y_i - X_i w\|_2^2\} \tag{7}$$

(so it jointly fits k different linear regression objectives). Explain how to reformulate this problem as a *single* least-squares problem with augmented $\tilde{X}$ and $\tilde{y}$ in an objective $\|\tilde{y} - \tilde{X}w\|_2^2$, *and* find the solution w to the aforementioned problem[1].

(e) What is the computational complexity (in big-O notation) of computing the solution the previous question in terms of n, p, k? Assume each $X_i \in \mathbb{R}^{n \times p}$ and $y_i \in \mathbb{R}^n$ and once again that the relevant square matrices are invertible.

[1]you may assume the relevant square matrices are invertible.

Name: ______________________

SID: ______________________

Periodic Time Series. We now consider an application to the problem of modeling a periodic time series z_t which we approximate by a sum of K sinusoids:

$$z_t \approx \hat{z}_t = \sum_{k=1}^{K} a_k \cos(\omega_k t - \phi_k) \quad t = 1, 2, \ldots, T \tag{8}$$

The coefficient $a_k \geq 0$ are the amplitudes, ω_k the frequencies, and ϕ_k the phases. In many applications (and the one we consider) the frequencies ω_k are apriori known and **fixed**. We wish to find $a_1, \ldots, a_K$ and $\phi_1, \ldots, \phi_K$ to ensure the means-squared value of the approximation error $(\hat{z}_1 - z_1, \ldots, \hat{z}_T - z_t)$ is small.

(f) Explain how to solve the aforementioned problem using (regularized) least squares to estimate $a_1, \ldots, a_K$ and $\phi_1, \ldots, \phi_K$. Be explicit in the mappings between the values $z_t, a_k, \omega_k, \phi_k$ in the original formulation and the standard regression parametrization y, X, w_λ (detailed in the beginning of the question), and the dimensions of the relevant vectors/matrices. *Hint: Recall the identity* $a\cos(\omega t - \phi) = \alpha\cos(\omega t) + \beta\sin(\omega t)$ *for* $\alpha = a\cos\phi$ *and* $\beta = a\sin\phi$, *with* $a = \sqrt{\alpha^2 + \beta^2}$ *and* $\phi = \arctan(\beta/\alpha)$.

Solution:

(a) Since $n < p$ we have that $X^\top X \in \mathbb{R}^{p\times p}$ has rank at most n and hence cannot be invertible (although it is positive-semi-definite).

(b) $X^\top X + \lambda I$ is always invertible. We can see this by using the SVD of X which gives that $X^\top X + \lambda I = V(S + \lambda I)V^\top$. Since $S \succeq 0$ (because $X^\top X$ is positive semi-definite) the minimum eigenvalue of $X^\top X + \lambda I$ is bounded below by $\lambda > 0$, so the matrix must be invertible.

(c) We can rewrite the ridge regression objective as an augmented least-squares problem as $\min_w \|\tilde{y} - \tilde{X}w\|_2^2$ with $\tilde{X} = \begin{bmatrix} X \\ \sqrt{\lambda} I_p \end{bmatrix}$ and $\tilde{y} = \begin{bmatrix} y \\ 0 \end{bmatrix}$. Writing down and expanding the normal equations $w_\lambda = (\tilde{X}^\top\tilde{X})^{-1}\tilde{X}^\top\tilde{y} = (X^\top X + \lambda I)^{-1}X^\top y$.

Alternatively, The first-order stationary conditions give that

$$X^\top(X^\top w_\lambda - y) + \lambda w_\lambda = 0 \implies (X^\top X + \lambda I)w_\lambda = X^\top y \implies w_\lambda = (X^\top X + \lambda I)^{-1} X^\top y.$$

(d) We can define $\tilde{X} = \begin{bmatrix} \sqrt{\lambda_1} X_1 \\ \vdots \\ \sqrt{\lambda_k} X_k \end{bmatrix}$ and $\tilde{y} = \begin{bmatrix} \sqrt{\lambda_1} y_1 \\ \vdots \\ \sqrt{\lambda_k} y_k \end{bmatrix}$ for which $\min_w \|\tilde{y} - \tilde{X}w\|_2^2$ will have the same solution.

We can expand out the solution from the single least-squares problem (from the normal equations) as,

$$w = (\sum_{i=1}^{k} \lambda_i X_i^\top X)^{-1} (\sum_{i=1}^{k} \lambda_i X_i^\top y_i) \tag{9}$$

(It's fine if students did not explicitly write the expansion if their reformulation was correct and they wrote the solution to a single least-squares problem).

(e) The dominant computational cost is computing the matrix inverse. First computing $\sum_{i=1}^{k} \lambda_i X_i^\top X$ takes $O(knp^2)$ time. Inverting it takes an additional $O(p^3)$ time.

(f) Using the identity we have that,

$$z_t \approx \sum_{k=1}^{K} a_k \cos(\omega_k t - \phi_k) = \sum_{k=1}^{K} \alpha_k \cos(\omega_k t) + \beta_k \sin(\omega_k t) \tag{10}$$

We can formulate as a linear regression where $y \in \mathbb{R}^T$ is the concatenated vector of z_t's, the tth row of X (the analogue of datapoint) is the row vector of $(\cos(\omega_k t), \sin(\omega_k t)) \in \mathbb{R}^{2K}$, and the concatenated vector of (α_k, β_k) corresponds to $w_\lambda \in \mathbb{R}^{2K}$. More explicitly, we can map the sinusoid data into the regression parametrization as

$$y = \begin{bmatrix} z_1 \\ \vdots \\ z_T \end{bmatrix}, \tag{11}$$

and that,

$$X = \begin{bmatrix} \cos(\omega_1 \cdot 1) & \dots & \cos(\omega_k 1) & \sin(\omega_1 \cdot 1) & \dots & \sin(\omega_k 1) \\ \vdots & \ddots & & & & \\ \cos(\omega_1 \cdot T) & \dots & \cos(\omega_k T) & \sin(\omega_1 \cdot T) & \dots & \sin(\omega_k T) \end{bmatrix}, \tag{12}$$

with

$$w_\lambda = \begin{bmatrix} \alpha_1 \\ \vdots \\ \alpha_k \\ \beta_1 \\ \vdots \\ \beta_k \end{bmatrix}, \tag{13}$$

corresponding to the solution of the regression.

From this estimate of w_λ we can recover estimates of a_k and ϕ_k using the identities $a = \sqrt{a^2 + b^2}$ and $\phi = \arctan(\beta/\alpha)$. Explicitly, we will estimate,

$$a_k = \sqrt{(w_\lambda)_k^2 + (w_\lambda)_{k+K})^2} \tag{14}$$

and

$$\phi_k = \arctan\left[\frac{(w_\lambda)_{k+K}}{(w_\lambda)_k}\right] \tag{15}$$

(It's fine if students did not explicitly write down this last mapping if their previous formulation was correct and they explained how to recover a_k, ϕ_k from the regression).

Note the ordering of the features in the matrix X and w_λ we have chosen here is not essential – other orderings are equally valid.

Name: ______________________

SID: ______________________

4. *Positive-Definite Matrices and Hessians.*

Let $C \in \mathbb{R}^{n\times n}$ by a real, symmetric positive-definite matrix. Consider the function

$$f_\lambda(x) = \|C - xx^\top\|_F^2 + 2\lambda\|x\|_2^2$$

where $x \in \mathbb{R}^n$.

(a) Compute the Hessian matrix of the function $f_\lambda(x)$ with respect to x, $\nabla^2 f_\lambda(x)$.
Hint: Note that $\|C - xx^\top\|_F^2 = \|C\|_F^2 + \|x\|_2^4 - 2x^\top Cx$.

Name: ______________________________

SID: ______________________________

(b) When is the Hessian matrix (which depends on x), positive semi-definite at *all* points x? You should derive an "if and only if" condition expressed in terms of λ and a function of the matrix C.

Solution:

(a) We can express the function as $f_\lambda(x) = (\|x\|_2^2 + \lambda)^2 - 2x^\top Cx + \|C\|_F^2 - \lambda^2$. The gradient and Hessian of $f_\lambda(x)$ at a point $x \in \mathbb{R}^n$ are

$$\frac{1}{4}\nabla f_\lambda(x) = (\|x\|_2^2 + \lambda)x - Cx, \quad \frac{1}{4}\nabla^2 f_\lambda(x) = (\|x\|_2^2 + \lambda)I_n + 2xx^\top - C.$$

(b) The Hessian is PSD everywhere if and only if

$$\forall\, x,\ v,\ \|v\|_2 = 1 \ :\ \|x\|_2^2 + \lambda \geq v^\top(C - 2xx^\top)v$$

The above is equivalent to

$$\begin{aligned} \lambda &\geq \max_{x,\, v\,:\, \|v\|_2=1} v^\top Cv - 2(v^\top x)^2 - x^\top x \\ &= \max_{v\,:\,\|v\|_2=1} v^\top Cv - \min_x \left(x^\top x + 2(v^\top x)^2\right) \\ &= \lambda_{\max}(C). \end{aligned}$$

since $\min_x \left(x^\top x + 2(v^\top x)^2\right) = 0$.