

Math Camp - Homework 7

1. MORE LINEAR ALGEBRA

Question 1: In this problem, we'll investigate matrix diagonalization.

To review from lecture, suppose we're given an $n \times n$ matrix A . If we can find an invertible matrix S and a diagonal matrix Λ such that $A = S\Lambda S^{-1}$, then we say that A is *diagonalizable*. Such a representation exists if and only if A has n linearly independent eigenvectors. In that case, the columns of the matrix S are the eigenvectors of A and the diagonal entries of Λ are the eigenvalues of A .

Part A: Why would this representation be useful? One reason is that it makes it easy to characterize the powers of a matrix. Prove that for a diagonalizable matrix A , the following equality holds for all natural numbers k :

$$A^k = S\Lambda^k S^{-1}$$

Part B: In this part, we'll see how matrix diagonalization can be applicable to a political science problem. Suppose we want to estimate the long-run effects of a policy on wealth distribution in a society. We have a simple model of the political economy where people are split into two groups, rich and poor. Let's suppose that under the policy in question, we have estimated the probability that a rich person will be rich (or poor) tomorrow and the probability that a poor person will be poor (or rich) tomorrow.

We can describe the model using the following "Markov matrix" that gives these transition probabilities:

$$M = \begin{array}{cc} & \begin{array}{cc} \text{Rich} & \text{Poor} \end{array} \\ \begin{pmatrix} .8 & .25 \\ .2 & .75 \end{pmatrix} & \begin{array}{cc} \text{Rich} \\ \text{Poor} \end{array} \end{array}$$

The entries represent the probability of someone who's in a column category today transitioning to the each row category tomorrow. That is, M tells us that there is a 0.8 probability of a rich person staying rich and a 0.2 probability of a rich person becoming poor, and so on. (Note that the columns add up to 1.)

Say the vector $\mathbf{v}_t = (v_{tR}, v_{tP})$ describes the proportion of rich and poor people at time t . In our model, the proportion of rich and poor people in time $t + 1$ is given by:

$$\mathbf{v}_{t+1} = M\mathbf{v}_t$$

This implies that $\mathbf{v}_{t+2} = M\mathbf{v}_{t+1} = MM\mathbf{v}_t = M^2\mathbf{v}_t$, and in general $\mathbf{v}_{t+k} = M^k\mathbf{v}_t$.

With this model, we can compute the long-run distribution of rich and poor (or the "steady state") by finding $\lim_{t \rightarrow \infty} \mathbf{v}_t$. By the property above, we can write this as $\lim_{t \rightarrow \infty} \mathbf{v}_t = \lim_{t \rightarrow \infty} M^t \mathbf{v}_0$, where \mathbf{v}_0 is some initial distribution of rich and poor. This limit looks really hard to compute, until we remember the tools of diagonalization!

It turns out that the eigenvalues for M are $\lambda_1 = .55$ and $\lambda_2 = 1$, and the eigenvectors are $\mathbf{x}_1 = (-1, 1)$ and $\mathbf{x}_2 = (1.25, 1)$. Using these facts, compute the steady-state distribution of rich and poor people, assuming that the initial value $\mathbf{v}_0 = (.5, .5)$. Are your results sensitive to the initial values?

2. MORE CALCULUS WITH MANY VARIABLES

Question 2: Someone tells you that there is a function $f(x, y)$ for which $f_x(x, y) = x + 4y$ and $f_y(x, y) = 3x - y$. Should you believe this assertion? Why or why not? (Integration is not necessary to answer this question. We can answer this using another concept we learned in class today. *Hint:* Does order of differentiation matter?)

Question 3: Find the Hessian H for the following functions.

(a) $g(x, y) = x^4 - 3x^2y^3$

(b) $f(x, y, z) = xyz - x^2$

Question 4: Find the local minimum values, local maximum values, and saddle point(s) of the function. Remember the process we discussed in class: Calculate the gradient, set it equal to zero to solve the system of equations, calculate the Hessian, and assess the Hessian at critical values. Be sure to show your work on each of these steps.

(a) $f(x, y) = x^4 + y^4 - 4xy + 2$

(b) $k(x, y) = (1 + xy)(x + y)$

Question 5: Calculate the following integrals:

(a) $\int_0^1 \int_2^3 x^2 y^3 \, dx dy$

(b) $\int_2^3 \int_0^1 x^2 y^3 \, dy dx$

(c) $\int_0^1 \int_0^{\sqrt{1-x^2}} 2x^3 y \, dy dx$

Question 6: Use implicit differentiation to solve for the derivative of y with respect to x .

(a) $3x^2 - 9y^2 = 12$

(b) $e^x + e^y = x^5$

Question 7. Suppose we were interested in learning about how years of schooling affect the probability that a person turns out to vote. Unlike yesterday, let's assume we now have many observations in our data.

Let Y_i be a vector containing many (n) observations of our dependent variable (whether or not a person i turned out to vote) and X_i be a vector containing many (n) observations of our independent variable, *education*, the number of years of schooling for individual i . We're also going to include an intercept term, β_0 in our model along with β_1 a coefficient that's associated with X_i .

This produces the following model for which we want to find the values of both β_0 and β_1 that minimize the sum of square errors.

$$Y_i = \beta_0 + \beta_1 X_i + \epsilon_i$$

where ϵ_i is an error term for person i .

Part A: Find the partial derivatives of the sum of squares with respect to β_0 and β_1 . Set these partial derivatives equal to 0 and determine $\hat{\beta}_0$ and $\hat{\beta}_1$.

$$\sum_{i=1}^n (\epsilon_i)^2 = \sum_{i=1}^n (Y_i - (\beta_0 + \beta_1 X_i))^2$$

Part B - Optional Challenge Problem: Use what we learned about multivariate optimization to find the hessian of $\hat{\beta}$. Does this hessian indicate that $\hat{\beta}_0$ and $\hat{\beta}_1$ minimize the sum of squared errors?

NOTE:

We've worked a lot on taking derivatives, but doing so with when summation is involved makes this different than what we've done before. Algebra tips for working with summation signs are below.

- (a) If you sum a constant n times, this is equivalent to multiplying that constant by n (e.g., $\sum_{i=1}^n 1 = n \times 1 = n$).
- (b) If you sum a variable that is also multiplied by a constant, you can move the constant outside the summation sign (e.g., $\sum_{i=1}^n 2 \times X_i = 2 \sum_{i=1}^n X_i$).
- (c) The summation sign is a linear operator so we can distribute it across variables we add together (e.g., $\sum_{i=1}^n (X_i + Y_i) = \sum_{i=1}^n X_i + \sum_{i=1}^n Y_i$)

Question 8 (Optional Challenge Problem): For continuous functions with only one variable, it is impossible to have two local maxima and no local minima. (In two dimensions, what goes up must come down before going up again!) However, for functions with two variables, such functions do exist. Show that the function

$$f(x, y) = -(x^2 - 1)^2 - (x^2y - x - 1)^2$$

has only two critical points, but that it has a local maxima at both of them. If it's hard to visualize how this is possible, go to WolframAlpha and try plotting it. To get a plot that will display the maxima, enter

`extrema -(x^2 - 1)^2 - (yx^2 - x - 1)^2`