

Math Camp - Homework 7 Solutions

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:

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

Solution We proceed by induction. First, in our base case, let $k = 1$. We know from the supposition of the problem that $A^1 = A = S\Lambda S^{-1} = S\Lambda^1 S^{-1}$. So we have a basis for induction. Now, assume that $A^k = S\Lambda^k S^{-1}$. Then we have:

$$\begin{aligned} A^{k+1} &= AA^k \\ &= AS\Lambda^k S^{-1} && \text{(by the inductive hypothesis)} \\ &= S\Lambda S^{-1}S\Lambda^k S^{-1} \\ &= S\Lambda\Lambda^k S^{-1} \\ &= S\Lambda^k S^{-1} \\ &= S\Lambda^{k+1} S^{-1} \end{aligned}$$

Then, by the principle of mathematical induction, the statement holds generally. ■

Further, you might notice that because Λ is a diagonal matrix, its k^{th} power is simply each diagonal entry

raised to the k^{th} power: $\Lambda^k = \begin{bmatrix} \lambda_1^k & 0 & \dots & 0 \\ 0 & \lambda_2^k & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \lambda_n^k \end{bmatrix}$

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:

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

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?

Solution: First, let’s write the diagonalized representation of M :

$$\begin{aligned} M &= SAS^{-1} \\ M &= [\mathbf{x}_1 \quad \mathbf{x}_2] \begin{bmatrix} .55 & 0 \\ 0 & 1 \end{bmatrix} [\mathbf{x}_1 \quad \mathbf{x}_2]^{-1} \\ M &= \begin{bmatrix} -1 & 1.25 \\ 1 & 1 \end{bmatrix} \begin{bmatrix} .55 & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} -1 & 1.25 \\ 1 & 1 \end{bmatrix}^{-1} \\ M &= \begin{bmatrix} -1 & 1.25 \\ 1 & 1 \end{bmatrix} \begin{bmatrix} .55 & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} -4/9 & 5/9 \\ 4/9 & 4/9 \end{bmatrix} \end{aligned}$$

Now, using the results from Part A, we can find the limit:

$$\begin{aligned} \lim_{t \rightarrow \infty} M^t \mathbf{v}_0 &= \lim_{t \rightarrow \infty} SA^t S^{-1} \mathbf{v}_0 \\ &= \lim_{t \rightarrow \infty} \begin{bmatrix} -1 & 1.25 \\ 1 & 1 \end{bmatrix} \begin{bmatrix} .55^t & 0 \\ 0 & 1^t \end{bmatrix} \begin{bmatrix} -4/9 & 5/9 \\ 4/9 & 4/9 \end{bmatrix} \mathbf{v}_0 \\ &= \begin{bmatrix} -1 & 1.25 \\ 1 & 1 \end{bmatrix} \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} -4/9 & 5/9 \\ 4/9 & 4/9 \end{bmatrix} \mathbf{v}_0 \\ &= \begin{bmatrix} 5/9 & 5/9 \\ 4/9 & 4/9 \end{bmatrix} \mathbf{v}_0 \equiv M^* \mathbf{v}_0 \end{aligned}$$

(We’re denoting this new matrix M^* .) Given the initial values $\mathbf{v}_0 = (.5, .5)$, we have $M^* \mathbf{v}_0 = (5/9, 4/9) \equiv \mathbf{v}^*$. Thus, in the long run under the policy, the “steady state” result \mathbf{v}^* is that about 56% of the population is rich and 44% of the population is poor.

If you try different initial values, you’ll notice that your answer will be the same. This makes intuitive sense: since we’re thinking about what happens “infinity” time periods in the future, it shouldn’t really matter what we start off with today.

Finally, if you’re insanely observant, you might have noticed that the steady state vector \mathbf{v}^* is proportional to one of the eigenvectors — specifically, \mathbf{x}_2 , which corresponds to the eigenvalue $\lambda_2 = 1$. This is not a coincidence. In fact, every Markov matrix like M (where entries are between 0 and 1 and the columns sum to 1) has 1 as an eigenvalue. If we normalize the eigenvector corresponding to this eigenvalue, so that its components sum to 1, we get the steady state probability vector!

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? (Contrary to what you may believe, 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?)

Solution: You should not believe this assertion. According to Young’s Theorem, $f_{xy} = f_{yx}$ (where x and y can be any two generic but different variables). Let’s take these second derivatives using the first derivatives that we are provided to see whether this is the case.

$$f_{xy} = \frac{\partial}{\partial y}(x + 4y) = 4 \quad f_{yx} = \frac{\partial}{\partial x}(3x - y) = 3$$

These second derivatives are clearly not equal. We can therefore conclude that either no such $f(x, y)$ exists or that the person made a mistake.

Another approach that you may have first thought of is to find the antiderivative of these two first partial derivatives to see whether the two end up being the same. We haven't discussed multivariate antidifferentiation/integration by this point, but you may have figured that it still works in a very similar way to multivariate differentiation where you treat all other variables as constants. This is a case where you see the importance of including dx or dy in your calculations to indicate what variable you are antidifferentiating over.

$$\int (x + 4y)dx = \frac{1}{2}x^2 + 4xy + c \quad \int (3x - y)dy = 3xy - \frac{1}{2}y^2 + c$$

Since we are taking indefinite integrals, we technically need to include the $+c$ in our answers. But once again, these two functions are not equal.

This and a couple previous questions suggest that it's never a bad idea to find f_{xy} and f_{yx} separately. It is tempting to simply find one and assume that the other is equal, but doing both is a helpful way of testing whether your math is correct. If your two second partial derivatives are equal, chances are great that you did it correctly. If they are unequal, you have caught yourself making an error.

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

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

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

Solutions

- (a) In order to find any second partial derivatives, we need the first partial derivatives. These are straightforward. (Here, we'll use the other partial derivative notation just to keep things interesting and to get you accustomed to it.)

$$g_x = 4x^3 - 6xy^3 \quad g_y = -9x^2y^2$$

Now, we find the second partial derivatives, which are also pretty simple to find.

$$g_{xx} = 12x^2 - 6y^3 \quad g_{xy} = -18xy^2 \quad g_{yx} = -18xy^2 \quad g_{yy} = -18x^2y$$

Note that $f_{xy} = f_{yx}$, just as Young's Theorem stipulates. We now have everything we need for the Hessian.

$$H = \begin{bmatrix} 12x^2 - 6y^3 & -18xy^2 \\ -18xy^2 & -18x^2y \end{bmatrix}$$

- (b) Finding this Hessian initially seems daunting because it involves three variables. Fortunately, the second derivatives are all really simple. We start by finding first partial derivatives.

$$f_x = yz - 2x \quad f_y = xz \quad f_z = xy$$

Now, we look for the second partial derivatives. Just to keep things orderly, let's start with f_{xx} , f_{yy} , and f_{zz} .

$$f_{xx} = -2 \quad f_{yy} = 0 \quad f_{zz} = 0$$

Then we can look for the other second partial derivatives that involve two different variables.

$$f_{xy} = \frac{\partial}{\partial y}(yz - 2x) = z \quad f_{yz} = \frac{\partial}{\partial z}(xz) = x \quad f_{zx} = \frac{\partial}{\partial x}(xy) = y$$

$$f_{yx} = \frac{\partial}{\partial x}(xz) = z \quad f_{zy} = \frac{\partial}{\partial y}(xy) = x \quad f_{xz} = \frac{\partial}{\partial z}(yz - 2x) = y$$

We have found all second derivatives, so we can now produce the Hessian.

$$H = \begin{bmatrix} f_{xx} & f_{xy} & f_{xz} \\ f_{yx} & f_{yy} & f_{yz} \\ f_{zx} & f_{zy} & f_{zz} \end{bmatrix} = \begin{bmatrix} -2 & z & y \\ z & 0 & x \\ y & x & 0 \end{bmatrix}$$

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

Solutions:

- (a) The first step, as the problem indicates, is to determine the gradient of the function. Taking the first derivatives here is quite straightforward.

$$\begin{aligned} \frac{\partial}{\partial x}(x^4 + y^4 - 4xy + 2) &= 4x^3 + 0 - 4y + 0 \\ &= 4x^3 - 4y \\ \frac{\partial}{\partial y}(x^4 + y^4 - 4xy + 2) &= 0 + 4y^3 - 4x + 0 \\ &= 4y^3 - 4x \\ \nabla f(x, y) &= (4x^3 - 4y, 4y^3 - 4x) \end{aligned}$$

Now we must set this gradient equal to the zero vector. So we know $4x^3 - 4y = 0$ and $4y^3 - 4x = 0$. The standard method for solving a system of equations is to solve one equation for one variable in terms of the other(s) and substitute that value into the other equations. In this case, let's choose to solve the second equation for x in terms of y . So we have $4y^3 - 4x = 0$. Dividing both sides by 4 gives $y^3 - x = 0$, and adding x to both sides gives $y^3 = x$. Now let's plug this into the other equation. So we have $4(y^3)^3 - 4y = 4y^9 - 4y = 0$. Dividing by 4 gives $y^9 - y = 0$. There are a few ways to go about looking at this equation to get a value for y , but let's take the most rigorous approach. First, let us try to simplify a bit; if y is not equal to 0, we can divide by y to get $y^8 - 1 = 0$, or $y^8 = 1$. So taking the square root of both sides gives $y^4 = \pm 1$, but it obviously cannot be -1, so $y^4 = 1$. Doing so again gives $y^2 = \pm 1$, but again, it cannot be that $y^2 = -1$, so $y^2 = 1$. Taking the square root one last time gives $y = \pm 1$. Both of these values are feasible. Since we know that $x = y^3$, we know that (1,1) and (-1, -1) are critical points of this function.

Are these the only ones, though? Well, not necessarily. Recall that in finding those critical points, we had to divide by y , which we cannot do when $y = 0$. In other words, we made an assumption that y was not 0 - which means we only found the non-zero values of y that produce critical points. Is there a critical point where $y = 0$? In this case, yes - if both y and x are zero, the gradient is zero as well. So (0,0) is also a critical point.

Now let's calculate the Hessian.

$$H(x, y) = \begin{bmatrix} f_{xx} & f_{xy} \\ f_{yx} & f_{yy} \end{bmatrix} = \begin{bmatrix} \frac{\partial}{\partial x}(4x^3 - 4y) & \frac{\partial}{\partial x}(4y^3 - 4x) \\ \frac{\partial}{\partial y}(4x^3 - 4y) & \frac{\partial}{\partial y}(4y^3 - 4x) \end{bmatrix} = \begin{bmatrix} 12x^2 & -4 \\ -4 & 12y^2 \end{bmatrix}$$

Now we simply have to plug in the x and y values at the critical points and apply the second derivative test.

$$H(1, 1) = H(-1, -1) = \begin{bmatrix} 12 & -4 \\ -4 & 12 \end{bmatrix}$$

So for the critical points (1,1) and (-1, -1), $AC - B^2 = 12 * 12 - (-4)^2 = 128 > 0$ and $A = 12 > 0$. So the Hessian at these points is positive definite, and the points (1,1) and (-1, -1) are local minima.

$$H(0,0) = \begin{bmatrix} 0 & -4 \\ -4 & 0 \end{bmatrix}$$

So for the critical point (0,0), $AC - B^2 = 0 * 0 - (-4)^2 = -16 < 0$, so the Hessian is indefinite and the point (0,0) is a saddle point.

- (b) As in the first problem, we begin by finding the gradient. To simplify the derivation process, note that $(1 + xy)(x + y) = x^2y + xy^2 + x + y$

$$\frac{\partial}{\partial x}(x^2y + xy^2 + x + y) = 2xy + y^2 + 1$$

$$\frac{\partial}{\partial y}(x^2y + xy^2 + x + y) = x^2 + 2xy + 1$$

$$\nabla f(x, y) = (y^2 + 2xy + 1, x^2 + 2xy + 1)$$

Now we need to set this gradient equal to the zero vector and solve the system of equations (sorry for the awful algebra here!). The standard method of solving one equation for one variable in terms of the other and plugging back into the other equation will work here (and I will write out the algebra below), but there is a much easier way of doing it. Notice that if $x^2 + 2xy + 1 = 0$ and $y^2 + 2xy + 1 = 0$, then it must be the case that $x^2 + 2xy + 1 = y^2 + 2xy + 1$. Now we can subtract $2xy + 1$ from both sides to get $x^2 = y^2$, or $x = \pm y$. This gives us two cases to consider: either $x = y$ or $x = -y$. Let's start with $x = y$. Plugging this value of x into the first equation gives us $y^2 + 2yy + 1 = 3y^2 + 1 = 0$. But then $3y^2 = -1$, which is impossible for any real value of y . So now let's look at $x = -y$. Then the first equation gives us $y^2 + 2(-y)y + 1 = -y^2 + 1 = 0$, which means that $y^2 = 1$, or $y = \pm 1$. If $y = 1$, we know $x = -y = -1$, and if $y = -1$, $x = -y = 1$. So (-1, 1) and (1, -1) are critical points of this function (you can confirm for yourselves that these points are solutions to the other equation as well).

The longer way to do it is as follows. First, let's solve the first equation for x in terms of y . So we begin with $y^2 + 2xy + 1 = 0$. Subtract $2xy$ to get $y^2 + 1 = -2xy$. Clearly, we will want to divide by y , but first, let's check what happens when $y = 0$. If $y = 0$, then $y^2 + 1 = 1$ and $-2xy = 0$ for all x - so there are no solutions. So now, let's divide both sides of $y^2 + 1 = -2xy$ by $-2y$. Then we have $x = -\frac{y^2+1}{2y}$. Plugging this value into the second equation gives us $(-\frac{y^2+1}{2y})^2 + 2y(-\frac{y^2+1}{2y}) + 1 = 0$. Distributing out the first term and cancelling out the $2y$ in the second leaves us with $\frac{y^4+2y^2+1}{4y^2} - y^2 = 0$. Now let's multiply both sides by $4y^2$. Then we have $y^4 + 2y^2 + 1 - 4y^4 = -3y^4 + 2y^2 + 1 = 0$. We can then apply the quadratic formula to get

$$y^2 = \frac{-2 \pm \sqrt{2^2 - 4(-3)(1)}}{2(-3)} = \frac{-2 \pm \sqrt{16}}{-6} = \frac{-2 \pm 4}{-6}$$

Clearly, if $y^2 = \frac{-2+4}{-6} = -1/3$, y has no real values. So we must have $y^2 = \frac{-2-4}{-6} = 1$, or $y = \pm 1$, as before, with the same x values following.

Now that we have established (-1,1) and (1, -1) as critical points, we must find the Hessian.

$$H(x, y) = \begin{bmatrix} f_{xx} & f_{xy} \\ f_{yx} & f_{yy} \end{bmatrix} = \begin{bmatrix} \frac{\partial}{\partial x}(y^2 + 2xy + 1) & \frac{\partial}{\partial x}(x^2 + 2xy + 1) \\ \frac{\partial}{\partial y}(y^2 + 2xy + 1) & \frac{\partial}{\partial y}(x^2 + 2xy + 1) \end{bmatrix} = \begin{bmatrix} 2y & 2x + 2y \\ 2y + 2x & 2x \end{bmatrix}$$

Then we have

$$H(-1, 1) = \begin{bmatrix} 2 & 0 \\ 0 & -2 \end{bmatrix}$$

and

$$H(1, -1) = \begin{bmatrix} -2 & 0 \\ 0 & 2 \end{bmatrix}$$

In both cases, $AC - B^2 = 2(-2) - 0^2 = -4 < 0$, so the Hessian is indefinite and these are saddle points. So there are no local minima or maxima.

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

Solutions

(a)

$$\begin{aligned} \int_0^1 \int_2^3 x^2 y^3 dx dy &= \int_0^1 \left[\frac{1}{3} x^3 y^3 \right]_{x=2}^{x=3} dy \\ &= \int_0^1 \left[\frac{1}{3} 3^3 y^3 - \frac{1}{4} 2^3 y^3 \right] dy = \int_0^1 \frac{19}{3} y^3 dy \\ &= \frac{19}{12} y^4 \Big|_{y=0}^{y=1} = \frac{19}{12} \end{aligned}$$

(b)

$$\begin{aligned} \int_2^3 \int_0^1 x^2 y^3 dy dx &= \int_2^3 \left[\frac{1}{4} x^2 y^4 \right]_{y=0}^{y=1} dx \\ &= \int_2^3 \left[\frac{1}{4} 1^4 x^2 - \frac{1}{4} 0^4 x^2 \right] dx = \int_2^3 \frac{1}{4} x^2 dx \\ &= \frac{1}{12} x^3 \Big|_{x=2}^{x=3} = \frac{19}{12} \end{aligned}$$

A and B demonstrate the order of integration doesn't matter, you should get the same answer either way.

(c)

$$\begin{aligned} \int_0^1 \int_0^{\sqrt{1-x^2}} 2x^3 y dy dx &= \int_0^1 x^3 y^2 \Big|_{y=0}^{y=\sqrt{1-x^2}} dx \\ &= \int_0^1 x^3 (1 - x^2) dx = \int_0^1 x^3 - x^5 dx \\ &= \frac{1}{4} x^4 - \frac{1}{6} x^6 \Big|_{x=0}^{x=1} = \frac{1}{12} \end{aligned}$$

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

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

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

Answers:

$$\begin{aligned}
3x^2 - 9y^2 &= 12 \\
\frac{\partial}{\partial x} [3x^2 - 9y^2] &= \frac{\partial}{\partial x} [12] \\
6x - \frac{\partial}{\partial x} [9y^2] &= 0 \\
6x - \frac{\partial}{\partial x} [9(y(x))^2] &= 0
\end{aligned}$$

Viewing y as a function of x , we can apply the chain rule.

$$\begin{aligned}
6x - \frac{\partial}{\partial x} [9(y(x))^2] &= 0 \\
6x - 18y * \frac{\partial y}{\partial x} &= -6x \\
\frac{\partial y}{\partial x} &= \frac{-6x}{-18y} \\
\frac{\partial y}{\partial x} &= \frac{x}{3y}
\end{aligned}$$

$$\begin{aligned}
e^x + e^y &= x^5 \\
\frac{\partial}{\partial x} [e^x + e^y] &= \frac{\partial}{\partial x} [x^5] \\
e^x + \frac{\partial}{\partial x} [e^{y(x)}] &= 5x^4 \\
e^x + e^y \frac{dy}{dx} &= 5x^4 \\
\frac{dy}{dx} &= \frac{5x^4 - e^x}{e^y}
\end{aligned}$$

Question 7. Suppose we were interested in learning about how years of schooling affect the probability that a person turns out to vote. To simplify things, let's say we're trying to predict whether one individual voted. 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: Use the method of least squares to solve for the values of β_0, β_1 that minimizes the sum of squared errors in the our data. Using the tools of multivariate minimization we've been practicing, set the partial derivatives of β_0 and β_1 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: Use what we learned about multivariate optimization to find the 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.

Useful Algebra For Working With Summations:

- (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., $\sum_{i=1}^n (X_i + Y_i) = \sum_{i=1}^n X_i + \sum_{i=1}^n Y_i$)

Solutions

$$\begin{aligned} f(\beta_0, \beta_1 | x_i, y_i) &= \sum_{i=1}^n (Y_i - \beta_0 - \beta_1 X_i)^2 \\ \frac{\partial f(\beta_0, \beta_1 | x_i, y_i)}{\partial \beta_0} &= -2 \left(\sum_{i=1}^n (Y_i - \beta_0 - \beta_1 X_i) \right) \\ &= \sum_{i=1}^n -2Y_i + 2\beta_0 + 2\beta_1 X_i \\ 0 &= \sum_{i=1}^n -2Y_i + 2\beta_0 + 2\beta_1 X_i \\ 0 &= -2 \sum_{i=1}^n Y_i + 2 \sum_{i=1}^n \beta_0 + 2\beta_1 \sum_{i=1}^n X_i \\ 0 &= -2 \sum_{i=1}^n Y_i + (n \times 2\beta_0) + 2\beta_1 \sum_{i=1}^n X_i \\ n \times 2\beta_0 &= 2 \sum_{i=1}^n Y_i - 2\beta_1 \sum_{i=1}^n X_i \\ \hat{\beta}_0 &= \frac{2 \sum_{i=1}^n Y_i}{2n} - \frac{2\beta_1 \sum_{i=1}^n X_i}{2n} \\ &= \frac{\sum_{i=1}^n Y_i}{n} - \beta_1 \frac{\sum_{i=1}^n X_i}{n} \\ \hat{\beta}_0 &= \bar{Y} - \beta_1 \bar{X} \end{aligned}$$

There are two key moves in the above steps. First, when we take the sum of a constant (e.g., $2 \sum_{i=1}^n \beta_0$) we can pass it through the summation sign and multiply the constant by n . For intuition think about taking the sum of 1 many times (e.g., $\sum_{i=1}^n 1 = n \times 1 = n$). Second, we can write a term like $\frac{\sum_{i=1}^n Y_i}{n}$ as a sample mean, \bar{Y} .

Ok now let's move on to β_1 . Note: In the fourth step below, I sub in $\hat{\beta}_0$, what we just got in the above problem, for β_0 .

$$\begin{aligned}
 \frac{\partial f(\beta_0, \beta_1 | x_i, y_i)}{\partial \beta_1} &= \sum_{i=1}^n -2X_i(Y_i - \beta_0 - \beta_1 X_i) \\
 &= \sum_{i=1}^n -2Y_i X_i + 2\beta_0 X_i + 2\beta_1 X_i^2 \\
 0 &= \sum_{i=1}^n -2Y_i X_i + 2\beta_0 \sum_{i=1}^n X_i + 2\beta_1 \sum_{i=1}^n X_i^2 \\
 &= \sum_{i=1}^n -2Y_i X_i + 2(\bar{Y} - \beta_1 \bar{X}) \sum_{i=1}^n X_i + 2\beta_1 \sum_{i=1}^n X_i^2 \\
 &= \sum_{i=1}^n -2Y_i X_i + 2\bar{Y} \sum_{i=1}^n X_i - 2\beta_1 \bar{X} \sum_{i=1}^n X_i + 2\beta_1 \sum_{i=1}^n X_i^2 \\
 2\beta_1 \sum_{i=1}^n X_i^2 - 2\beta_1 \bar{X} \sum_{i=1}^n X_i &= \sum_{i=1}^n 2Y_i X_i - 2\bar{Y} \sum_{i=1}^n X_i \\
 \beta_1 \left(\sum_{i=1}^n X_i^2 - \bar{X} \sum_{i=1}^n X_i \right) &= \sum_{i=1}^n Y_i X_i - \bar{Y} \sum_{i=1}^n X_i \\
 \hat{\beta}_1 &= \frac{\sum_{i=1}^n Y_i X_i - \bar{Y} \sum_{i=1}^n X_i}{\sum_{i=1}^n X_i^2 - \bar{X} \sum_{i=1}^n X_i}
 \end{aligned}$$

Later in the course we'll discuss the variance and covariance of variables. Using this knowledge you will then be able to simplify $\hat{\beta}_1$ to $\frac{\text{Covariance}(X, Y)}{\text{Variance}(X)}$. This will be discussed in greater depth in 350A so don't worry if the estimator for β_1 doesn't make a ton of sense right now.

Challenge Portion

As a final check that these values minimize the sum of squared errors, we need to assess the hessian.

$$\begin{aligned}
 f(\beta_0, \beta_1 | x_i, y_i) &= \sum_{i=1}^n (Y_i - \beta_0 - \beta_1 X_i)^2 \\
 \frac{\partial f(\beta_0, \beta_1 | x_i, y_i)}{\partial \beta_0} &= \sum_{i=1}^n -2Y_i + 2\beta_0 + 2\beta_1 X_i \\
 \frac{\partial f(\beta_0, \beta_1 | x_i, y_i)}{\partial \beta_0 \beta_0} &= \sum_{i=1}^n 2 \\
 &= 2n \\
 \frac{\partial f(\beta_0, \beta_1 | x_i, y_i)}{\partial \beta_0 \beta_1} &= \sum_{i=1}^n 2X_i \\
 \frac{\partial f(\beta_0, \beta_1 | x_i, y_i)}{\partial \beta_1} &= \sum_{i=1}^n -2Y_i X_i + 2\beta_0 X_i + 2\beta_1 X_i^2 \\
 \frac{\partial f(\beta_0, \beta_1 | x_i, y_i)}{\partial \beta_1 \beta_1} &= \sum_{i=1}^n 2X_i^2
 \end{aligned}$$

This produces the following hessian:

$$\begin{aligned}
 H(f(\beta_0, \beta_1 | x_i, y_i)) &= \begin{bmatrix} \frac{\partial f(\beta_0, \beta_1 | x_i, y_i)}{\partial \beta_0 \beta_0} & \frac{\partial f(\beta_0, \beta_1 | x_i, y_i)}{\partial \beta_0 \beta_1} \\ \frac{\partial f(\beta_0, \beta_1 | x_i, y_i)}{\partial \beta_0 \beta_1} & \frac{\partial f(\beta_0, \beta_1 | x_i, y_i)}{\partial \beta_1 \beta_1} \end{bmatrix} \\
 &= \begin{bmatrix} 2n & \sum_{i=1}^n 2X_i \\ \sum_{i=1}^n 2X_i & \sum_{i=1}^n 2X_i^2 \end{bmatrix}
 \end{aligned}$$

Let's use what we learned about Hessians yesterday in class to determine whether we are at a minimum. If what we've found minimizes the sum of squares, this Hessian must be positive definite.

First, to start with $A = 2n > 0$. Since n can only take positive values, this means the first condition for a positive definite Hessian is met.

Second, we need to show that $AD - B^2 > 0$:

$$\begin{aligned}
 AD &= 2n \times \sum_{i=1}^n 2X_i^2 \\
 B^2 &= (2 \sum_{i=1}^n X_i)^2 \\
 AD - B^2 &= 2n \times \sum_{i=1}^n 2X_i^2 - (2 \sum_{i=1}^n X_i)^2 \\
 &= 4(n \sum_{i=1}^n X_i^2 - (\sum_{i=1}^n X_i)^2) \\
 &= 4n^2 \left(\frac{\sum_{i=1}^n X_i^2}{n} - \frac{(\sum_{i=1}^n X_i)^2}{n^2} \right) \\
 &= 4n^2 \left(\frac{\sum_{i=1}^n X_i^2}{n} - \left(\frac{\sum_{i=1}^n X_i}{n} \right)^2 \right) \\
 &= 4n^2 \left(\frac{\sum_{i=1}^n X_i^2}{n} - (\bar{X})^2 \right) \\
 &= 4n^2 \times \frac{1}{n} \sum_{i=1}^n (X_i^2 - \bar{X})^2 \\
 &= 4n \sum_{i=1}^n (X_i^2 - \bar{X})^2 \\
 4n \sum_{i=1}^n (X_i^2 - \bar{X})^2 &> 0 \\
 AD - B^2 &> 0
 \end{aligned}$$

The key step is restating $\frac{\sum_{i=1}^n X_i^2}{n} - (\bar{X})^2$ as $\frac{1}{n} \sum_{i=1}^n (X_i^2 - \bar{X})^2$. If you're unclear on how we get from this first statement to the second, expand $\frac{1}{n} \sum_{i=1}^n (X_i^2 - \bar{X})^2$ and try and make it look like the first statement.

Once we've stated the relationship this way, it's clear that $AD - B^2$ will always be greater than 0 (barring the pathological case where every X_i equals 0). n is always positive and, because we square the difference of $X_i^2 - \bar{X}$, the second part of the statement will always be positive as well. As a result, the $\hat{\beta}_0$ and $\hat{\beta}_1$ we identified earlier will minimize the sum of squared errors.

Optional Challenge Problem

Question 8: 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

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

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`

Solution:

This problem required some painful algebra. We all hate fighting through algebra... but it's a useful skill to have, sometimes.

The approach we'll take for this problem follows that of problem 4. First, let's multiply it out to make the derivatives easier to find.

$$\begin{aligned} -(x^2 - 1)^2 - (x^2y - x - 1)^2 &= -x^4 + 2x^2 - 1 - x^4y^2 + 2x^3y + 2x^2y - x^2 - 2x - 1 \\ &= -x^4 + x^2 - 2 - x^4y^2 + 2x^3y + 2x^2y - 2x \end{aligned}$$

So the first derivative with respect to x is $f_x = -4x^3 + 2x - 4x^3y^2 + 6x^2y + 4xy - 2$ (yes, there is no question that this is ugly and horrible) and the first derivative with respect to y is $f_y = -2x^4y + 2x^3 + 2x^2$. Let's set $f_y = 0$ and solve for y . Note that when $x = 0$, $f_x = 0$ has no solutions, and so we can divide by x for the purpose of finding critical points. When $f_y = 0$, we have $-2x^4y + 2x^3 + 2x^2 = 0$. Adding $2x^4y$ to both sides and then dividing by $2x^2$ gives $x + 1 = x^2y$. Dividing by x^2 then gives $y = \frac{x+1}{x^2}$. Now, painful though it is, we can plug this value of y into $f_x = 0$. So we have

$$\begin{aligned} -4x^3 + 2x - 4x^3y^2 + 6x^2y + 4xy - 2 &= 0 \\ -4x^3 + 2x - 4x^3\left(\frac{x+1}{x^2}\right)^2 + 6x^2\left(\frac{x+1}{x^2}\right) + 4x\left(\frac{x+1}{x^2}\right) - 2 &= 0 \\ -4x^3 + 2x - 4x^3\left(\frac{x^2 + 2x + 1}{x^4}\right) + 6(x+1) + 4x\left(\frac{x+1}{x^2}\right) - 2 &= 0 \\ -4x^3 + 2x - 4\left(\frac{x^2 + 2x + 1}{x}\right) + 6x + 6 + 4x\left(\frac{x+1}{x^2}\right) - 2 &= 0 \\ -4x^3 + 2x - \frac{4x^2}{x} - \frac{8x}{x} - \frac{4}{x} + 6x + 4 + \frac{4x^2}{x^2} + \frac{4x}{x^2} &= 0 \\ -4x^3 + 2x - 4x - 8 - \frac{4}{x} + 6x + 4 + 4 + \frac{4}{x} &= 0 \\ -4x^3 + (2x - 4x + 6x) + (4 + 4 - 8) + \left(\frac{4}{x} - \frac{4}{x}\right) &= 0 \\ -4x^3 + 4x &= 0 \\ x^2 - 1 &= 0 \\ x^2 &= 1 \\ x &= \pm 1 \end{aligned}$$

Though the algebra was awful, we ended up with something nice and simple. Remembering that $y = \frac{x+1}{x^2}$, we know that when $x = 1$, $y = \frac{1+1}{1^2} = 2$, and when $x = -1$, $y = \frac{1-1}{(-1)^2} = 0$. So our critical points are $(1, 2)$ and $(-1, 0)$. Now we must find the Hessian.

$$\begin{aligned} H(x, y) &= \begin{bmatrix} f_{xx} & f_{xy} \\ f_{yx} & f_{yy} \end{bmatrix} \\ &= \begin{bmatrix} \frac{\partial}{\partial x}(-4x^3 + 2x - 4x^3y^2 + 6x^2y + 4xy - 2) & \frac{\partial}{\partial x}(-2x^4y + 2x^3 + 2x^2) \\ \frac{\partial}{\partial y}(-4x^3 + 2x - 4x^3y^2 + 6x^2y + 4xy - 2) & \frac{\partial}{\partial y}(-2x^4y + 2x^3 + 2x^2) \end{bmatrix} \\ &= \begin{bmatrix} (-12x^2 + 2 - 12x^2y^2 + 12xy + 4y) & (-8x^3y + 6x^2 + 4x) \\ (-8x^3y + 6x^2 + 4x) & (-2x^4) \end{bmatrix} \end{aligned}$$

So now we must simply plug in the critical values of (x, y)

$$\begin{aligned} H(1, 2) &= \begin{bmatrix} (-12(1^2) + 2 - 12(1^2)(2^2) + 12(1)(2) + 4(2)) & (-8(1^3)(2) + 6(1^2) + 4(1)) \\ (-8(1^3)(2) + 6(1^2) + 4(1)) & (-2(1^4)) \end{bmatrix} \\ &= \begin{bmatrix} (-12 + 2 - 48 + 24 + 8) & (-16 + 6 + 4) \\ (-16 + 6 + 4) & (-2) \end{bmatrix} \\ &= \begin{bmatrix} -26 & -6 \\ -6 & -2 \end{bmatrix} \end{aligned}$$

So at $(1,2)$, $AC - B^2 = (-26)(-2) - (-6)^2 = 52 - 36 = 18 > 0$, and $A = -26 < 0$, so the Hessian is negative definite and $(1,2)$ is a local maximum.

$$\begin{aligned} H(-1, 0) &= \begin{bmatrix} (-12(-1^2) + 2 - 12(-1^2)(0^2) + 12(-1)(0) + 4(0)) & (-8(-1^3)(0) + 6(-1^2) + 4(-1)) \\ (-8(-1^3)(0) + 6(-1^2) + 4(-1)) & (-2(-1^4)) \end{bmatrix} \\ &= \begin{bmatrix} (-12 + 2) & (6 - 4) \\ (6 - 4) & (-2) \end{bmatrix} \\ &= \begin{bmatrix} -10 & 2 \\ 2 & -2 \end{bmatrix} \end{aligned}$$

So at $(-1, 0)$, $AC - B^2 = (-10)(-2) - 2^2 = 20 - 4 = 16 > 0$, and $A = -10 < 0$, so the Hessian is negative definite and $(-1, 0)$ is a local maximum as well. We therefore have only two critical points, and they are both local maxima.