

Functions of several variables and optimization with several variables

1 Find first partial derivatives

Find all of the first partial derivatives of each function.

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

Solution: The partial derivatives here involve straightforward power rule. As you do these partial derivatives, get used to seeing all other variables that you're not interested in as constants. It will become more natural over time.

$$\frac{\partial f}{\partial x} = 3 \quad \frac{\partial f}{\partial y} = -8y^3$$

b. $f(x, y) = x^5 + 3x^3y^2 + 3xy^4$

Solution: More power rule.

$$\frac{\partial f}{\partial x} = 5x^4 + 9x^2y^2 + 3y^4 \quad \frac{\partial f}{\partial y} = 6x^3y + 12xy^3$$

c. $g(x, y) = xe^{3y}$

Solution: Now, we have to start dealing with the chain rule. Note that even though there is technically a product involving x and y here, at no point does either variable show up in both parts of the product. As such, we don't need product rule.

The partial derivative with respect to x is straightforward. The function is linear with respect to x , so the partial derivative is e^{3y} .

The partial derivative with respect to y is a bit more complicated.

$$\frac{\partial g}{\partial y} = xe^{3y} \cdot \frac{\partial}{\partial y}(3y) = xe^{3y} \cdot 3 = 3xe^{3y}$$

d. $k(x, y) = \frac{x-y}{x+y}$

Solution: We require the use of quotient rule here. (Or as we showed in a previous homework, you can rewrite this as a product.) Recall that the quotient rule for some generic $h(x) = \frac{f(x)}{g(x)}$ is

$$h'(x) = \frac{f'(x)g(x) - f(x)g'(x)}{(g(x))^2}$$

So, we can apply this formula to derive the function with respect to each variable to get:

$$\begin{aligned}
\frac{\partial k}{\partial x} &= \frac{(\frac{\partial}{\partial x}(x-y))(x+y) - (x-y)(\frac{\partial}{\partial x}(x+y))}{(x+y)^2} \\
&= \frac{(1)(x+y) - (x-y)(1)}{(x+y)^2} \\
&= \frac{x+y-x+y}{(x+y)^2} \\
&= \frac{2y}{(x+y)^2} \\
\frac{\partial k}{\partial y} &= \frac{(\frac{\partial}{\partial y}(x-y))(x+y) - (x-y)(\frac{\partial}{\partial y}(x+y))}{(x+y)^2} \\
&= \frac{(-1)(x+y) - (x-y)(1)}{(x+y)^2} \\
&= \frac{-x-y-x+y}{(x+y)^2} \\
&= -\frac{2x}{(x+y)^2}
\end{aligned}$$

e. $f(x, y, z) = \log(x + 2y + 3z)$

Solution: This requires relatively straightforward chain rule.

$$\begin{aligned}
\frac{\partial f}{\partial x} &= \frac{1}{x + 2y + 3z} \cdot \frac{\partial}{\partial x}(x + 2y + 3z) = \frac{1}{x + 2y + 3z}(1) = \frac{1}{x + 2y + 3z} \\
\frac{\partial f}{\partial y} &= \frac{1}{x + 2y + 3z} \cdot \frac{\partial}{\partial y}(x + 2y + 3z) = \frac{1}{x + 2y + 3z}(2) = \frac{2}{x + 2y + 3z} \\
\frac{\partial f}{\partial z} &= \frac{1}{x + 2y + 3z} \cdot \frac{\partial}{\partial z}(x + 2y + 3z) = \frac{1}{x + 2y + 3z}(3) = \frac{3}{x + 2y + 3z}
\end{aligned}$$

f. $h(x, y, z) = x^2 e^{yz}$

Solution: This requires some chain rule, as well.

$$\begin{aligned}
\frac{\partial h}{\partial x} &= 2x^{(2-1)} e^{yz} = 2x e^{yz} \\
\frac{\partial h}{\partial y} &= x^2 e^{yz} \cdot \frac{\partial}{\partial y}(yz) = x^2 e^{yz} \cdot z = x^2 z e^{yz} \\
\frac{\partial h}{\partial z} &= x^2 e^{yz} \cdot \frac{\partial}{\partial z}(yz) = x^2 e^{yz} \cdot y = x^2 y e^{yz}
\end{aligned}$$

2 Find the gradient

Find the gradient ∇f of the following functions and evaluate them at the given points.

a. $f(x, y) = \sqrt{x^2 + y^2}$, $(x, y) = (3, 4)$

Solution: Let's first rewrite the function so that it is easier to differentiate.

$$f(x, y) = \sqrt{x^2 + y^2} = (x^2 + y^2)^{1/2}$$

Now, taking the partial derivatives with respect to x and y comes more naturally.

$$\begin{aligned}\frac{\partial f}{\partial x} &= \frac{1}{2}(x^2 + y^2)^{-1/2} \cdot \frac{\partial}{\partial x}(x^2 + y^2) \\ &= \frac{1}{2}(x^2 + y^2)^{-1/2} \cdot (2x) \\ &= x(x^2 + y^2)^{-1/2} \\ &= \frac{x}{\sqrt{x^2 + y^2}}\end{aligned}$$

$$\begin{aligned}\frac{\partial f}{\partial y} &= \frac{1}{2}(x^2 + y^2)^{-1/2} \cdot \frac{\partial}{\partial y}(x^2 + y^2) \\ &= \frac{1}{2}(x^2 + y^2)^{-1/2} \cdot (2y) \\ &= y(x^2 + y^2)^{-1/2} \\ &= \frac{y}{\sqrt{x^2 + y^2}}\end{aligned}$$

So, the gradient of this function is:

$$\nabla f(x, y) = \left(\frac{x}{\sqrt{x^2 + y^2}}, \frac{y}{\sqrt{x^2 + y^2}} \right)$$

And if we evaluate it at the given point, we get:

$$\nabla f(3, 4) = \left(\frac{3}{\sqrt{3^2 + 4^2}}, \frac{4}{\sqrt{3^2 + 4^2}} \right) = \left(\frac{3}{\sqrt{9 + 16}}, \frac{4}{\sqrt{9 + 16}} \right) = \left(\frac{3}{\sqrt{25}}, \frac{4}{\sqrt{25}} \right) = \left(\frac{3}{5}, \frac{4}{5} \right)$$

b. $f(x, y, z) = (x + z)e^{x-y}$, $(x, y, z) = (1, 1, 1)$

Solution: This question is slightly more involved; the partial derivative with respect to x will require product rule since x appears in both factors of the product. Recall that the product rule for a generic $h(x) = j(x)k(x)$ is $h'(x) = j'(x)g(x) + f(x)g'(x)$. The partial derivative with respect to y will require the chain rule.

$$\begin{aligned}\frac{\partial f}{\partial x} &= \left(\frac{\partial}{\partial x}(x + z) \right) e^{x-y} + (x + z) \left(\frac{\partial}{\partial x} e^{x-y} \right) \\ &= (1) \cdot e^{x-y} + (x + z) \cdot e^{x-y} \\ &= e^{x-y} + (x + z) \cdot e^{x-y} \\ &= e^{x-y}(x + z + 1)\end{aligned}$$

(The partial derivative above technically requires chain rule where you take the derivative of $x - y$ with respect to x , but that's just 1, so we omit that work here.)

$$\begin{aligned}
\frac{\partial f}{\partial y} &= (x+z) \cdot e^{x-y} \cdot \left(\frac{\partial}{\partial y}(x-y) \right) \\
&= (x+z) \cdot e^{x-y} \cdot (-1) \\
&= -e^{x-y}(x+z)
\end{aligned}$$

$$\begin{aligned}
\frac{\partial f}{\partial z} &= \frac{\partial}{\partial z}(xe^{x-y} + ze^{x-y}) \\
&= \frac{\partial}{\partial z}(xe^{x-y}) + \frac{\partial}{\partial z}(ze^{x-y}) \\
&= 0 + e^{x-y} \\
&= e^{x-y}
\end{aligned}$$

The gradient of this function, which we can also write as a vertical vector, is

$$\nabla f(x, y, z) = \begin{bmatrix} e^{x-y}(x+z+1) \\ -e^{x-y}(x+z) \\ e^{x-y} \end{bmatrix}$$

When we evaluate the gradient at the given value, we obtain

$$\nabla f(1, 1, 1) = \begin{bmatrix} e^{1-1}(1+1+1) \\ -e^{1-1}(1+1) \\ e^{1-1} \end{bmatrix} = \begin{bmatrix} e^0(3) \\ -e^0(2) \\ e^0 \end{bmatrix} = \begin{bmatrix} 1 \cdot 3 \\ -1 \cdot 2 \\ 1 \end{bmatrix} = \begin{bmatrix} 3 \\ -2 \\ 1 \end{bmatrix}$$

3 Find the Hessian

Find the Hessian H for the following functions.

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

Solution: 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 $f_{xy} = f_{yx}$. 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. $f(x, y, z) = xyz - x^2$

Solution: 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.

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

4 Find the critical points

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

Solution: 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. $k(x, y) = (1 + xy)(x + y)$

Solution: 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.

5 Least squares regression

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 just have one observation of each variable. Let Y be our single observation of the dependent variable (whether or not a person turns out to vote) and X be our single observation of the independent variable, (the number of years of education that same person has). We believe that the process used to generate our data takes the following form:

$$Y = \beta X + \epsilon$$

where ϵ is an error term. We include this error term because we think random occurrences in the world will mean our model produces estimates that are slightly wrong sometimes, but we believe that on average, this model accurately relates X to Y . We observe the values of X and Y , but what about β ? How do we know the value of β that best approximates this relationship, i.e., what's the slope of this line?

There are different criteria we could use, but a popular choice is the method of least squares. In this process, we solve for the value of β that minimizes the sum of squared errors, ϵ^2 , in our data. Using the tools of minimization we've been practicing, find the value of β that minimizes this quantity. (Hint: In this case there is only one observation, so the sum of squared errors is equal to the single error squared.)

Solution: First we must identify the quantity we are trying to minimize, the sum of squared errors. Using some simple algebra, we get:

$$\begin{aligned} Y &= \beta X + \epsilon \\ Y - \beta X &= \epsilon \\ (Y - \beta X)^2 &= \epsilon^2 \end{aligned}$$

If we have many data points, there would be a summation sign in front of both sides of this equation, but since there is only one data point here, we have the quantity of interest. We want to find the value of β that minimizes $(Y - \beta X)^2$, the squared error. Let's call this function $f(\beta)$.

The first step to minimizing a function is finding its first derivative with respect to the variable of interest, which is β in this case. For that we will need the chain rule, since this is a nested function. Define $f(\beta)$ as $h(g(\beta))$, where $h(u) = u^2$ and $g(u) = (Y - uX)$ for any u . So, by the chain rule,

$$\begin{aligned} f'(\beta) &= h'(g(\beta)) * g'(\beta) \\ &= 2(Y - \beta X) * -X \\ &= -2X(Y - \beta X) \end{aligned}$$

With our first derivative in hand, we need to solve for β^* , the value of β at which $f'(\beta) = 0$. Again, some algebra:

$$\begin{aligned} f'(\beta) &= -2X(Y - \beta^* X) = 0 \\ -2XY + 2X^2\beta^* &= 0 \\ -XY + X^2\beta^* &= 0 \\ X^2\beta^* &= XY \\ \beta^* &= \frac{XY}{X^2} \\ \beta^* &= \frac{Y}{X} \end{aligned}$$

Is this a minimum or a maximum? Remember, the first derivative will equal 0 in both cases. To find out, we perform the second derivative test.

$$\begin{aligned} f''(\beta) &= \frac{df'(\beta)}{d\beta} \\ &= \frac{d}{d\beta} - 2X(Y - \beta X) \\ &= \frac{d}{d\beta} - 2XY + 2X^2\beta \\ &= 2X^2 \end{aligned}$$

We know that the quantity $2X^2 > 0$, so the second derivative is positive, so we have found a minimum. We have set no bounds on the domain, and there is only one minimum (i.e. only one solution for β^*), so we are done. $\frac{Y}{X}$ is the value of β that minimizes the squared error. For any other value of β , the squared distance between the true outcome Y , and the predicted outcome, βX , would be larger, which means our model would not fit as well.

6 Least squares regression, refined

Following on the previous exercise, suppose we showed a colleague our model of voter turnout and she complained. “What a lame model”, our colleague said, “You definitely have to include an intercept term.” So in this problem we’ll follow our colleague’s advice and do just that.

Let Y be our single observation of the dependent variable (whether or not a person turned out to vote) and X_1 be our single observation of an independent variable, *education*, the number of years of schooling for this individual. Now though, 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_1 .

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.

$$Y = \beta_0 + \beta_1 X_1 + \epsilon$$

where ϵ is an error term.

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, find the values of β_0 and β_1 that minimize this quantity.

Solution:

$$\begin{aligned} Y &= \beta_0 + \beta_1 X_1 + \epsilon \\ Y - (\beta_0 + \beta_1 X_1) &= \epsilon \\ (Y - (\beta_0 + \beta_1 X_1))^2 &= \epsilon^2 \\ f(\beta_0, \beta_1 | x_i, y_i) &= (Y - \beta_0 - \beta_1 X_1)^2 \\ \frac{\partial f(\beta_0, \beta_1 | x_i, y_i)}{\partial \beta_0} &= -2(Y - \beta_0 - \beta_1 X_1) \\ &= -2Y + 2\beta_0 + 2\beta_1 X_1 \\ 0 &= 2Y - 2\beta_0 - 2\beta_1 X_1 \\ \hat{\beta}_0 &= Y - \beta_1 X_1 \\ \frac{\partial f(\beta_0, \beta_1 | x_i, y_i)}{\partial \beta_1} &= -2X_1(Y - \beta_0 - \beta_1 X_1) \\ &= -2YX_1 + 2\beta_0 X_1 + 2\beta_1 X_1^2 \\ 0 &= 2YX_1 - 2\beta_0 X_1 - 2\beta_1 X_1^2 \\ \beta_1 X_1^2 &= YX_1 - \beta_0 X_1 \end{aligned}$$

We then sub in $\hat{\beta}_0$ for β_0 and continue:

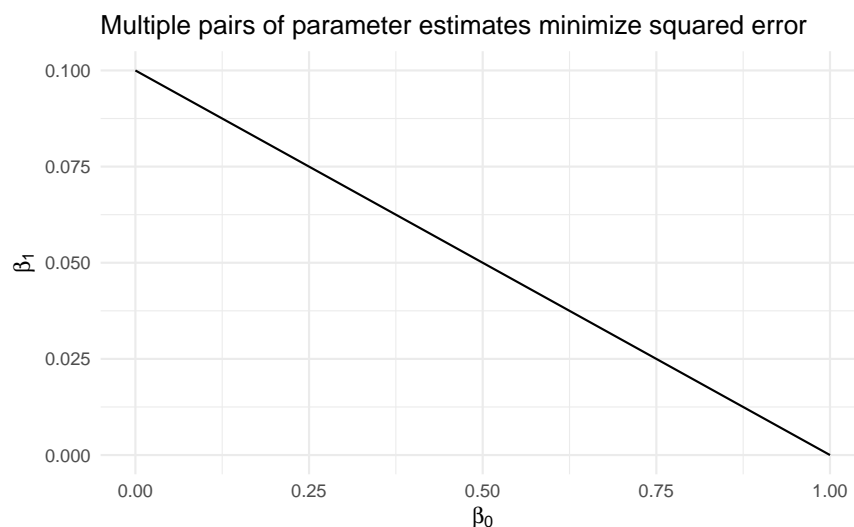
$$\begin{aligned} \beta_1 X_1^2 &= YX_1 - (Y - \beta_1 X_1)X_1 \\ \beta_1 X_1^2 &= YX_1 - YX_1 + \beta_1 X_1^2 \\ \beta_1 X_1^2 &= \beta_1 X_1^2 \\ \beta_1 X_1^2 - \beta_1 X_1^2 &= 0 \\ 0 &= 0 \end{aligned}$$

So, what’s going on? Why can’t we find a solution for $\hat{\beta}_1$?

The problem is that we are trying to use two parameters to describe a single observation of data. This is a case of a problem that occurs when solving systems of equations known as **underdetermination** (i.e., we have two unknown parameters, β_0 and β_1 , but only one equation with which to estimate them).

The intuition behind this problem is that, given our data, we don't have enough information to uniquely identify two parameters. With one observation we could minimize the squared error by *either*:

1. Fitting the point exactly with our intercept term, β_0
2. Fitting the point exactly with the parameter on education β_1 or
3. Fitting the point with some linear combination of values for β_0 and β_1 . Assuming this individual voted ($y=1$) and had 10 years of education ($x=10$), the line shows all the pairs of parameter estimates that would minimize the squared error with the two edge cases as end points (i.e., explain entirely with the intercept term or with the parameter attached to education) and a number of combinations of parameter values that also work in between (e.g., $\beta_0 = .5$ and $\beta_1 = .05$).



Since we have so many different combinations of parameter values that accomplish our goal, to minimize the squared error, equally well, no unique solution exists for this parameter vector. When this occurs we say our model is not identified.

If we start to get more observations than parameters we want to estimate (i.e., we have an overdetermined system of equations), this answer will change and we'll get unique solutions for each.

1. This problem illustrates, in a very simple manner, one symptom of underdetermination: perfect multicollinearity. Because we have a linearly dependent $\hat{\beta} = (X'X)^{-1}X'Y$.

$$\begin{aligned}
 X &= \begin{bmatrix} 1 & X_1 \end{bmatrix} \\
 X'X &= \begin{bmatrix} 1 \\ X_1 \end{bmatrix} \begin{bmatrix} 1 & X_1 \end{bmatrix} \\
 &= \begin{bmatrix} 1 & X_1 \\ X_1 & X_1^2 \end{bmatrix}
 \end{aligned}$$

Where we can show linear dependence by multiplying the first row in the matrix by X_1 to get an exact copy of the second row. We'll see that perfect multicollinearity can also occur even when we have more observations than variables, if we can write some variables as a linear combination of other variables.

2. If you got values for $\hat{\beta}_0$ and $\hat{\beta}_1$ on this problem you did something wrong (e.g., assumed more than 1 observation, made an algebra mistake).