# 3007-Final-Report

Sam Vredenburgh

November 2023

## 1 Introduction

**Seemingly disparate math techniques and tools can overlap to solve problems in really interesting ways. This final explores the connection between linear-algebra and solving a system of equations, and calculus using a multivariate function and partial derivatives to derive the same answer. Math is cool!**

**We are given a system of equations that can be expressed as either a matrix equation or a multivariate function. Using the method of gradient descent we can find the minimum of the multivariate function, using this to solve the system of equations.**

(i) If we are given the matrix multiplication in the first line below, we can rewrite that equation into a multivariate function with two inputs and two outputs. These two equations are functionally the same, and allow us to leverage different math tools between them to analyze or solve a problem. For example, we can then attempt to solve the system of equations created from the matrix multiplication, which is equivalent to finding when the output of the multivariate function is a vector $< 0, 0 >$.

$$f(x,y) = \begin{pmatrix} 3 & 1 \\ 2 & -1 \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} - \begin{pmatrix} -1 \\ 1 \end{pmatrix} = \begin{pmatrix} 3x + y + 1 \\ 2x - y - 1 \end{pmatrix}$$

$$f : \mathbb{R}^2 \to \mathbb{R}^2; (x,y) \mapsto \langle 3x + y + 1, 2x - y - 1 \rangle$$

(ii) As mentioned before, the output of the multivariate function is a vector, whose length we can calculate using the equation $|| < a, b > || = \sqrt{a^2 + b^2}$. We can then take the square of the output of this vector length and derive a new simpler function from the original, $f(x, y)$. This last function, $L(x, y) = a^2 + b^2$, can then be used to construct a gradient from, and through the method of gradient descent explored in the next steps we can find a minimum of $L(x, y)$ which in turn can tell us something about $f(x, y)$, and that in turn can help solve the system of the equations posed at the beginning.

(iii) Throughout this process the connections we need to keep in mind are below. $f(x, y)$ can be derived from the matrix equation. The length of the output vector from $f(x, y)$ can be computed with $|| < a, b > || = sqrt(a^2 + b^2)$. And finally the squared length, $L(x, y)$, can be derived from the previous steps and used to construct the method of gradient descent.

$$Multivariate function:$$

$$f(x,y) = \langle 3x + y + 1, 2x - y - 1 \rangle = A\vec{x} - \vec{b}$$

$$vectorlength:$$

$$||\langle a, b \rangle|| = \sqrt{a^2 + b^2}$$

$$f(x,y) \to L(x,y):$$

$$||f(x,y)||^2 =: L(x,y) = (3x + y + 1)^2 + (2x - y - 1)^2$$

## 2 Exploration of Gradient Descent

(i) We are provided with a link to desmos to analyze a multivariate function with different level sets of the output calculated to analyze. We are also given the calculation of the gradient with a moving point to understand the relationship between the gradient and level sets. When comparing the gradient to the level sets in the function $h(x, y) = x^2 + xy + y^2$ we see that the direction of the vector is perpendicular to the level set at any point along it. The same is true when testing any other level set and moving around to different points. An example can be seen in Figure 1 below.

(ii) We also notice along each level set the height does not change when moving perpendicular to the direction of the gradient vector. Each point along a single level set is the same surface height. Some examples can be seen in figure 2 below.

(iii) Given the two previous facts we can deduce that the direction the gradient is pointing, that is perpendicular to the continuous height of a level set, is in the direction in which the function $f(x, y)$ is increasing the fastest since the fastest way to move away from a straight line is to move perpendicular to it. If the surface is not changing in height, the quickest way to move away from the surface is straight up in a perpendicular direction. Figure 3 can help explain this point.

(iv) A gradient can be constructed from multivariate functions in the form of $h$ from $\mathbb{R}^n \to \mathbb{R}$, meaning multiple inputs and a single output. Since the function we are given is in the form of

$$f : \mathbb{R}^2 \to \mathbb{R}^2; (x, y) \mapsto \langle 3x + y + 1, 2x - y - 1\rangle$$

meaning it has two inputs and TWO outputs, we cannot construct the gradient on this function directly.

(v) However, we can create a multivariate function using the given function $f(x, y)$, and the length of the output vector $||\langle a, b\rangle|| = \sqrt{a^2 + b^2}$. For example given $f(3, 1)$ we can compute the length of the vector as $\sqrt{137}$.

$$f(x, y) = \langle 3x + y + 1, 2x - y - 1\rangle$$
$$f(3, 1) = \langle 11, 4\rangle$$
$$||\langle 11, 4\rangle|| = \sqrt{137}$$

(vi) We can use this previous calculation $||\langle a, b\rangle|| = \sqrt{a^2 + b^2}$ to compute the squared length of the vector, removing the messy square root from the result. If we create a new multivariate function using this result, $||f(x, y)||^2 =: L(x, y)$, we can now see a function in the form of $\mathbb{R}^2 \to \mathbb{R}$: $L(x, y) = (3x + y + 1)^2 + (2x - y - 1)^2$.

(vii) Since $L(x, y)$ is using the squared length of the vector from $f(x, y)$ , notice the output with the same input from before $(3, 1)$ will be the square, 137, of the previous output $\sqrt{137}$.

(viii) Relating back to the main purpose of this process, if both $a$ and $b$ in the output vector $\langle a, b\rangle$ are 0 the output vector will have zero length: $||\langle 0, 0\rangle|| = \sqrt{0^2 + 0^2} = 0$. We can find a pair of (x,y) input values to make this true, which will also solve the system of equations posed at the beginning!

(ix) Now that we have a multivariate function in the form of $\mathbb{R}^2 \to \mathbb{R}$, namely $L(x, y) = (3x + y + 1)^2 + (2x - y - 1)^2$, we can compute the gradient of this function with two inputs and ONE output using the partial derivatives $\nabla||f(x, y)||^2 = \nabla L(x, y) = \langle 26x + 2y + 2, 2x + 4y + 4\rangle$. Notice this is the gradient in the form of $\mathbb{R}^2 \to \mathbb{R}^2$ and outputs a vector.

$$\nabla||f(x, y)||^2 = \nabla L(x, y) = \langle 26x + 2y + 2, 2x + 4y + 4\rangle$$

(x) We can compare this gradient function to the matrix multiplication formula using the transpose of the matrix $A$, $A^T(A\vec{x} - \vec{b})$, resulting: $\begin{pmatrix} 13x + y + 1 \\ x + 2y + 2 \end{pmatrix}$

(xi) Here we see a clear relationship between the previous steps! This is exciting because we can use matrix multiplication to find minimum values in a multivariate function using a similar process in a way that makes it much easier to compute, for example creating simple Python code instead of partial derivative calculations.

$$\nabla||A\vec{x}-\vec{b}||^2 = 2(A^T(A\vec{x}-\vec{b}))$$

(xii) The general formula for the method of gradient descent to find a local minimum for a function $g(x,y)$ in the form of $g:\mathbb{R}^2 \to \mathbb{R}$ is given below:

$$\vec{a}_{n+1} = \vec{a}_n - \eta\nabla g(\vec{a}_n)$$

Given a starting point of $a_0$ we can generate a sequence of points by subtracting the gradient from each previous point at that specific point($a_n$ ) multiplied by a step size, $\eta$. This means we take a point and subtract a vector in the opposite direction of the steepest direction with a magnitude calculated from the partial derivatives. Since we are subtracting the vector the gradient produces we are moving in the opposite direction of the steepest increasing direction, thus moving down. As the points approach a minimum the partial derivatives will get smaller and smaller due to the fact they are calculated from the slope of the function at that given point and will eventually approach 0, meaning the gradient will reach a length of 0. Figure 4 below demonstrates a basic example of using the method of gradient descent on a simpler function approaching a minimum from both directions. If $\eta$, the step size, is too large the algorithm could diverge from a critical pair of values at a minimum by taking infinite steps without resolving with the gradient length being 0. This could be seen in code, an infinite loop could arise out of never meeting the condition to stop calculations, in particular when the gradient length is at or incredibly close to 0. With too small of a step size it could take much longer to calculate a local minimum. Figure 5 demonstrates issues that could arise from too large or too small of a value for $\eta$ on a simple function.

(xiii) As the provided Python code demonstrates, the input pairs $(x,y)$ converge towards a critical pair of values, $(0,-1)$, as the value of the squared length function $L(x,y)$ is decreasing towards 0.

(xiv) To reiterate why we can use this method of gradient descent on $L(x,y)$ and not on the given function $f(x,y)$, it's because $L(x,y)$ has a single output in the form of $\mathbb{R}^2 \to \mathbb{R}$ which we can use to determine the partial derivatives for the gradient from, whereas the function $f(x,y)$ outputs two values in the form of $\mathbb{R}^2 \to \mathbb{R}^2$. The gradient can only be constructed from a function with a single output. Figure 6 can help show this.

(xv) In summation of the previous few steps, the method of gradient descent will result in a local minimum, and in our case the gradient function is created using the squared length of the output vector from our original function, $f(x,y)$ as $L(x,y)$. The method of gradient descent creates a sequence of points which results in the output vector decreasing to 0 as you subtract the gradient from each point. When the minimum is reached through the method of gradient descent we also found that the output of $L(x,y)$ is approaching 0. This allows us to analyze the original function, $f(x,y)$, since $L(x,y)$ was created from the original function. Per step 8, the only way the squared length of the output vector, as defined by $L(x,y)$, is 0, will be when both $a$ and $b$ will equal 0. Since $L(x,y)$ was created from $f(x,y)$, these same values which make $L(x,y)=0$ can be used as the critical pair of values of $(x,y)$ that will make $f(x,y)$ $= \langle 0,0\rangle$.

(xvi) Looking at the provided results of the Python code we can estimate that at the point $(0,-1) f(x,y)$ will equal $\langle 0,0\rangle$. To demonstrate the previous point, of the relationship between $f(x,y)$ and $L(x,y)$, we can prove this same pair of $(x,y)$ values will make $L(x,y)=0$.

$$f(x,y) = \langle 3x+y+1, 2x-y-1\rangle$$

$$f(0,-1) = \langle 3(0)-1+1, 2(0)+1-1\rangle$$

$$f(0,-1) = \langle 0,0\rangle$$

$$L(x,y) = (3x+y+1)^2 + (2x-y-1)^2$$

$$L(0,-1) = (3(0) - 1 + 1)^2 + (2(0) + 1 - 1)^2$$
$$L(0,-1) = (0)^2 + (0)^2$$
$$L(0,-1) = 0$$

(xvii) It is worth noting if the given function, $f(x, y)$, did not have a solution for the system of equations, where the output vector is $\langle 0, 0 \rangle$, we could have taken similar steps, creating the square length of the output vector $L(x, y)$, determining the partial derivatives, starting at a given point and calculating a sequence of points using the gradient and found that the $(x, y)$ pair does not converge to a pair of critical values that would make $L(x, y) = 0$.

# 3 Part 2 - Hardest question of the year

**If you had to be a letter of the alphabet, what letter would you be? Cursive or print? Upper case or lower case?** I would have to choose a lowercase v, preferably in a shade of dark green. Yes, my last name does start with a V, thus a lifelong affinity for the letter. I like the lowercase look as it is more subdued, and in line with some of the experimental electronic music and idm artists and tracks I am a fan of.

# 4 Figures

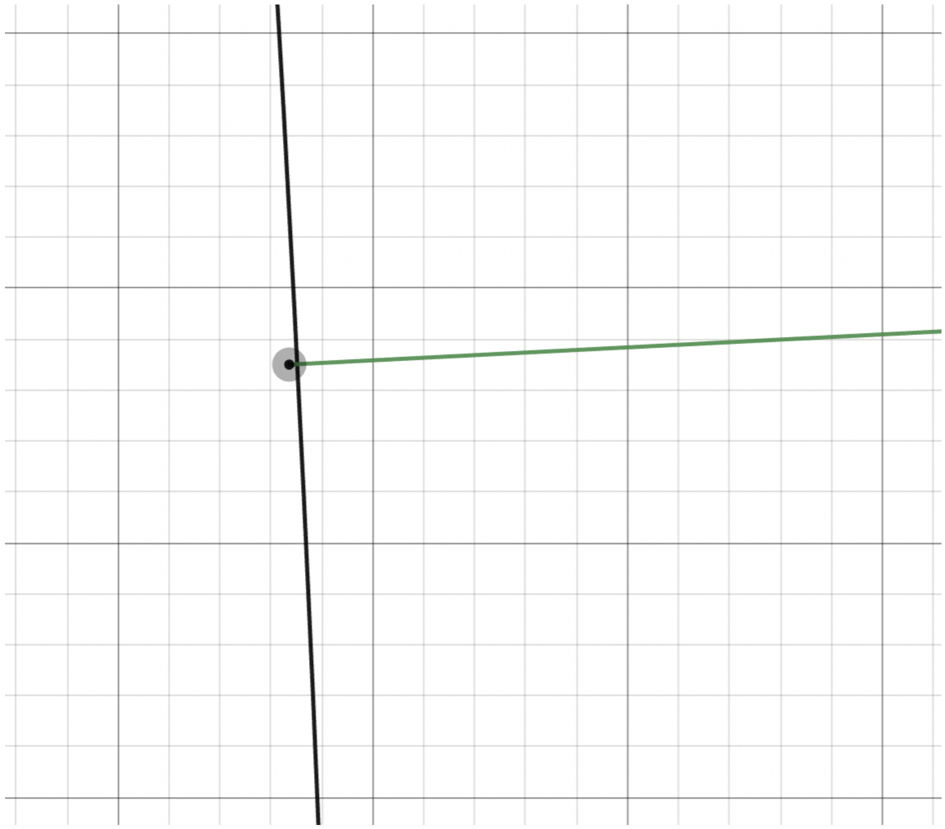

Figure 1: Gradient in green perpendicular to the level step in black

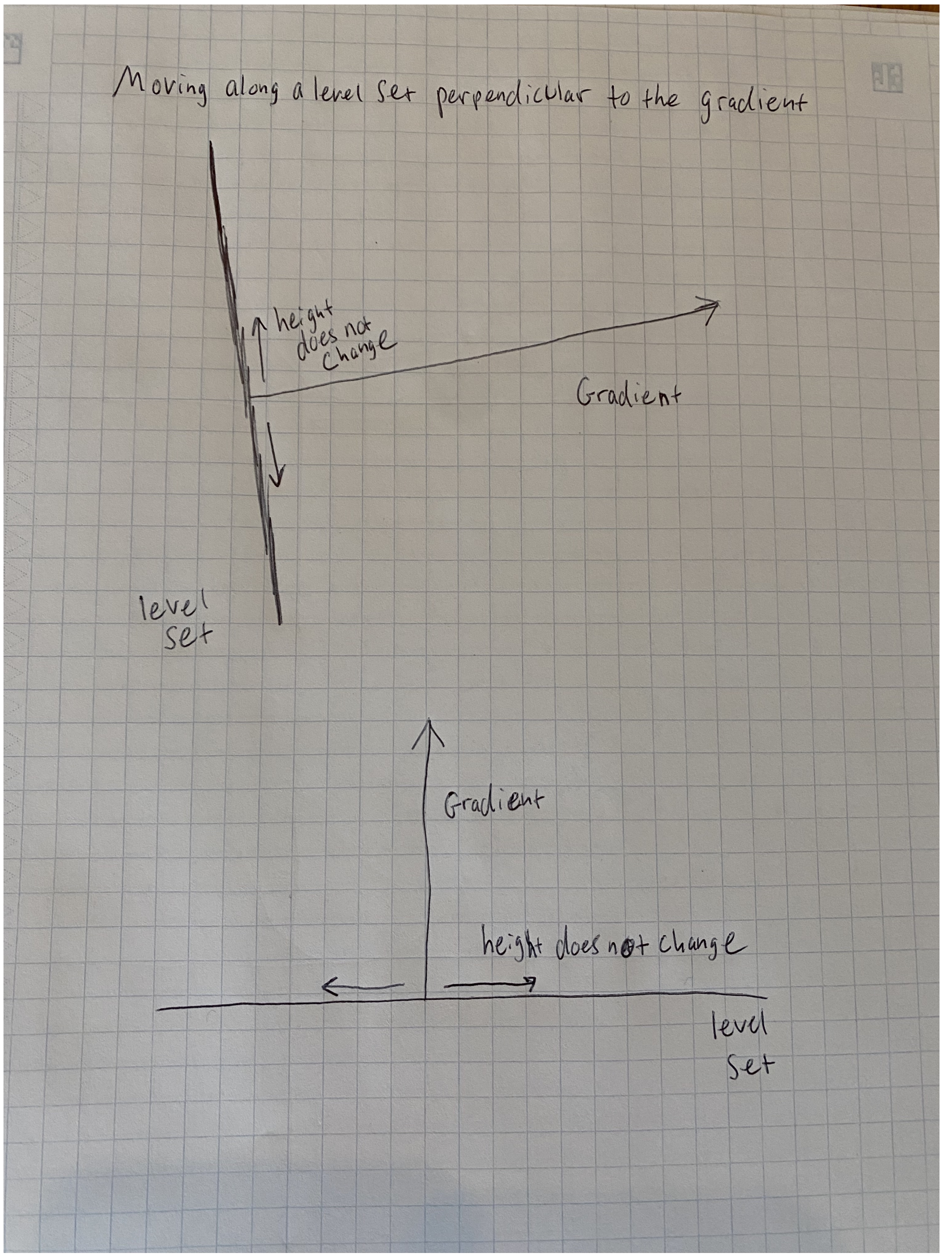


Figure 2: Examples of the level set height vs gradient.

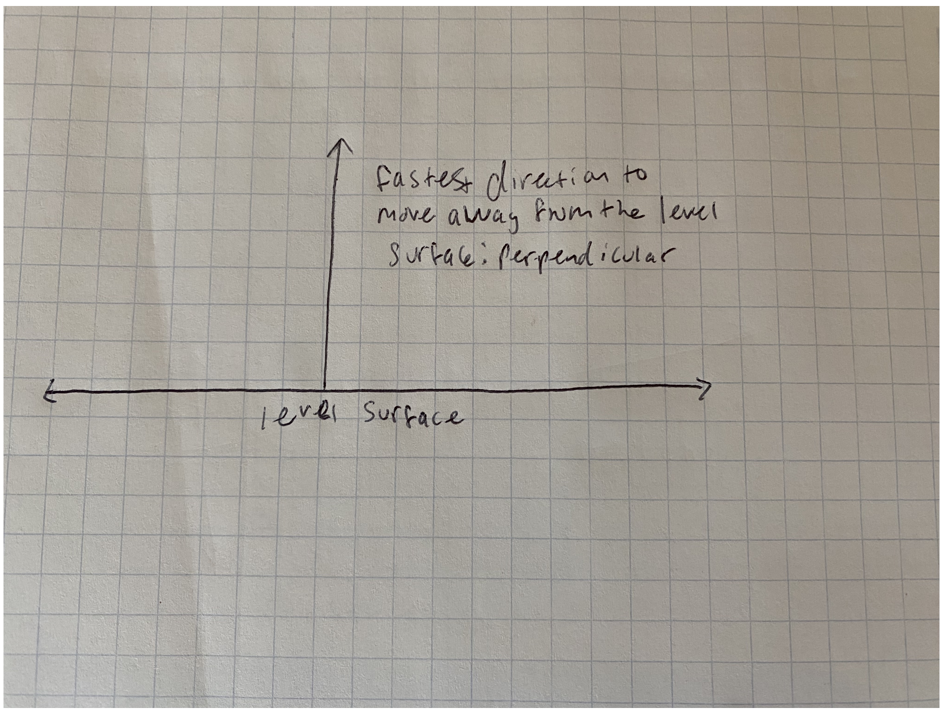


Figure 3: Fastest direction to move away from the level set is perpendicular.

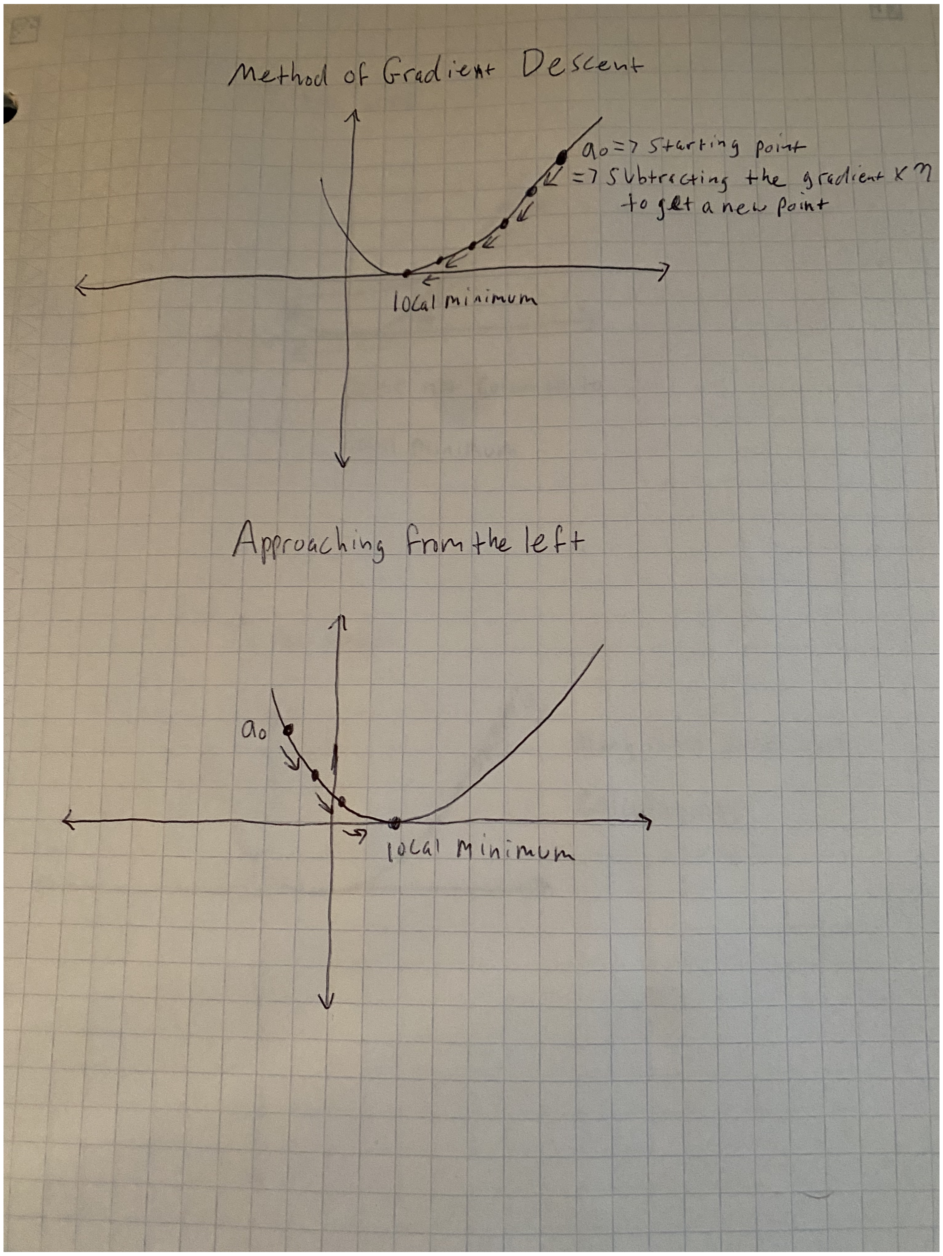


Figure 4: Example Method of Gradient Descent from both directions

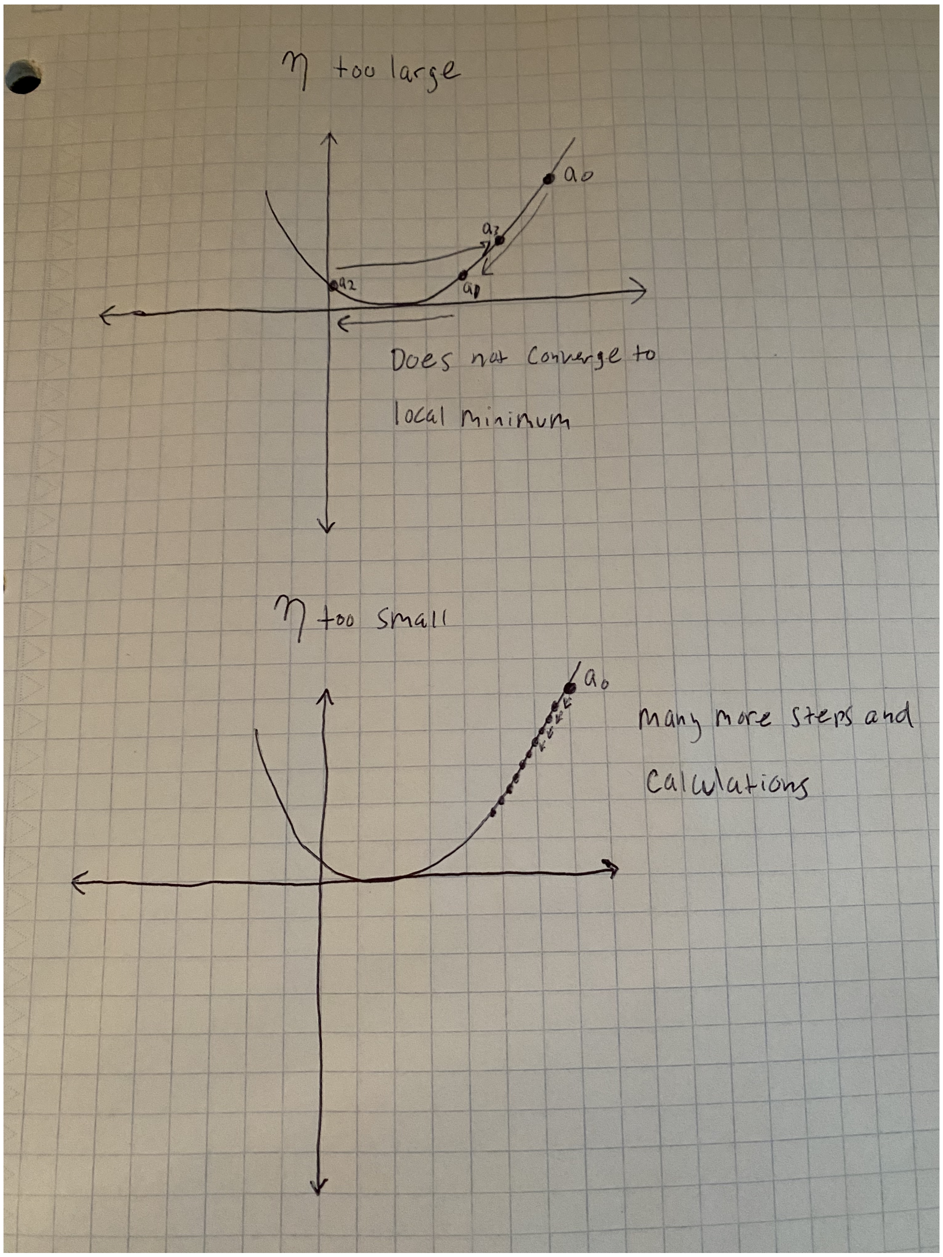


Figure 5: Examples of issues with eta too large and too small

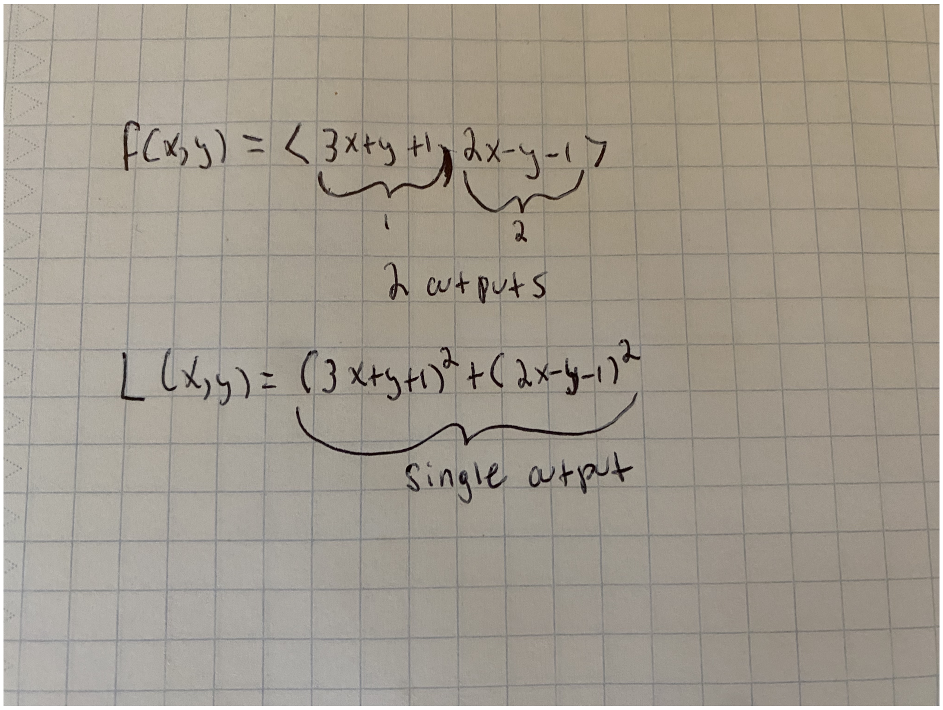


Figure 6: Why the gradient cannot be constructed from f(x,y)

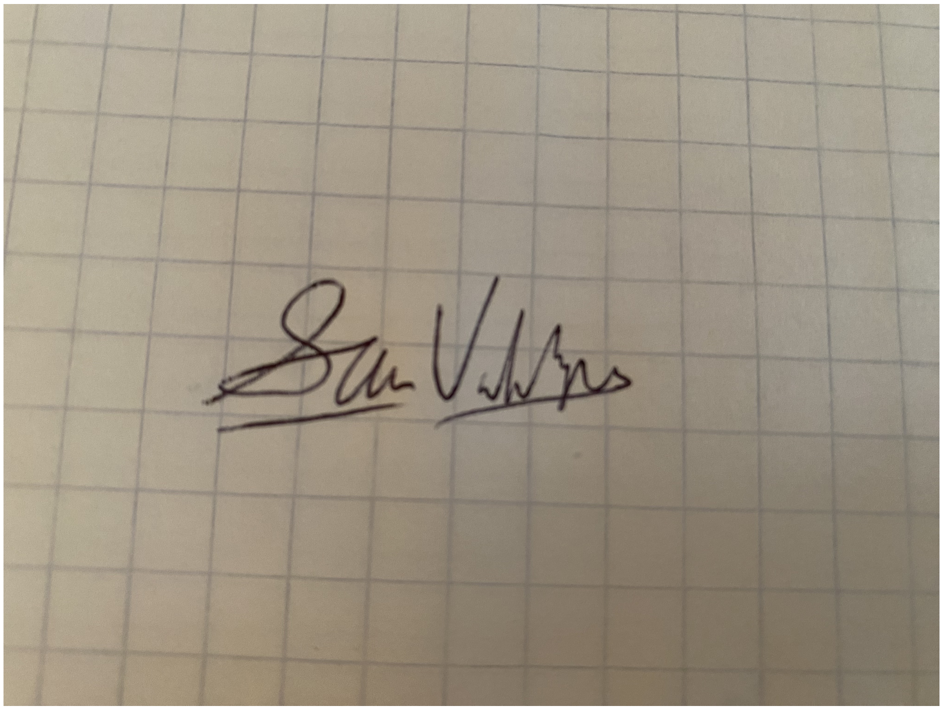

Figure 7: Acknowledgement of the Final terms