Gradient Descent

Recall: If $f(x)$ is a continuous function, then its derivative $f'(x) = \frac{d}{dx} f(x)$ is the rate of change of that function.

(1) Compute $f'(x) = \frac{d}{dx} f(x)$ the following derivatives:
(a) $f(x) = x^n$
(b) $f(x) = e^x$
(c) $f(x) = \ln(x)$
(d) $f(x) = \sin(x)$

(2) Let f and g be continuous functions. Compute
(a) $\dfrac{d}{dx}\left(f(x)g(x)\right)$
(b) $\dfrac{d}{dx}\left(\dfrac{f(x)}{g(x)}\right)$
(c) $\dfrac{d}{dx} f\left(g(x)\right)$

(3) What are each of the "rules" above called?

(4) What is the derivative of a continuous function $f(x)$ useful for?

Definition. We say $x = a$ is a **critical point** of a function f if $f'(a) = 0$ or $f'(a)$ is undefined.

(5) What is special about critical points?

(6) Find and classify the critical points of $g(x) = x^3 - 3x^2 + 3x$.

(7) Can we take derivatives of derivatives? If so, what is the benefit of this? Use $g(x) = x^3 - 3x^2 + 3x$ to give an example of your answer.

Recall: *Independent variables* are variables that do not depend on each other.

Definition. A **function** f **of two independent variables** x and y is a rule that assigns each ordered pair (x, y) in some set D to exactly one number $f(x, y)$

(8) Consider $f(x, y) = x^2 + xy$ with domain $D = \{(x, y) | x, y \in \mathbb{R}\}$.
(a) Compute $f(-1, 2)$.

(b) How do we usually denote the set D?
(c) Where does the graph of $f(x, y)$ live?
(d) Sketch the graph and compare your graph with an online graphing calculator like Desmos or GeoGebra.

(9) Functions with more than one variable! For example:

$$f(x, y, z) = x + y\cos(z) - 132,\ D = \mathbb{R}^3$$
$$g(x_1, , x_2, x_3, x_4, x_5) = x_1x_2 + x_3^2 + 5x_4 + x_5^3 + 1,\ D = \mathbb{R}^5$$

(a) Evaluate the two functions above at specific points in their domain.
(b) Where does the graph of the functions f and g live?

Definition. The **domain** of a function is the set of all inputs where the function is defined.

(10) Find the domain of the functions:
(a) $f(x, y) = \sqrt{x^2 - y}$
(b) $g(x, y) = \frac{1}{xy}$
(c) $h(x, y) = x^2 + y^2$

Definition. An x-**trace** is a curve $z = f(x, c)$ for some constant c. Similarly, a y-**trace** is a curve $z = f(d, y)$ for some constant d.

(11) Consider $z = f(x, y) = x^2 + y$, $D = \mathbb{R}^2$.
(a) Plot the x-traces $z = f(x, 0)$ and $z = f(x, 1)$ in the xz-plane and in $\mathbb{R}^3$.
(b) Plot the y-traces $z = f(0, y)$ and $z = f(2, y)$ in the yz-plane and in $\mathbb{R}^3$.
(c) What are the graphs of each?

Definition. A **level curve** or **contour** of a function $f(x, y)$ is a curve of the form $k = f(x, y)$ for some constant k.

(12) Again consider $f(x, y) = x^2 + y$, $D = \mathbb{R}^2$. Plot multiple contours of this function in the xy-plane. Compare your plots to the graph of f in $\mathbb{R}^3$. Save this drawing, you'll need it later!

(13) Think about your answers to exercises 4 and 7. Discuss how you might do similar things with the graphs of functions $z = f(x, y)$. Reference traces and contours in your answer.

Definition. The partial derivative of a differentiable function $f(x_1, x_2, \ldots, x_n)$ with respect to the variable x_i is

$$\lim_{h \to 0} \frac{f(x_1, \ldots, x_i + h, \ldots, x_n) - f(x_1, \ldots, x_i, \ldots, x_n)}{h}$$

Notation. We use $\frac{\partial f}{\partial x_i}$, f_{x_i}, and $\partial_{x_i} f$ to denote the x_i-partial derivative of f.

To compute a partial derivative, you follow the same rules as the usual differentiation that you are familiar with, except you treat all the variables that you are not differentiating as constants.

(14) Consider $f(x, y) = (x^2 + y^2)^3 - y$ and $g(x, y) = e^{xy} + 2$. Compute:
 (a) f_x
 (b) f_y
 (c) g_x
 (d) g_y

(15) We can also compute higher-order partial derivatives. For example $f_{xy} = \frac{\partial}{\partial y}\left(\frac{\partial f}{\partial x}\right)$. Let's try some examples with $f = x^2 - 5xy + y^3$. Compute:
 (a) f_{xy}
 (b) f_{xx}
 (c) f_{yy}
 (d) f_{xyy}

Chain Rule with Partial Derivatives Let us assume $z = f(x, y)$ and that both x and y are functions of other variables u, v, i.e $x = x(u, v), y = y(u, v)$. Then we can define $\frac{\partial f}{\partial u}$, $\frac{\partial f}{\partial v}$ using chain rule,

$$\frac{\partial f}{\partial u} = \frac{\partial f}{\partial x}\frac{\partial x}{\partial u} + \frac{\partial f}{\partial y}\frac{\partial y}{\partial u}, \quad \frac{\partial f}{\partial v} = \frac{\partial f}{\partial x}\frac{\partial x}{\partial v} + \frac{\partial f}{\partial y}\frac{\partial y}{\partial v}$$

(16) Let $z = f(x, y) = (x^2 + y^2)^3$, $x = \sin(u - v)$, $y = \cos(u + v)$. Compute $\frac{\partial z}{\partial u}$.

Tangent Plane. In 2D we can compute tangent lines. In 3D we can compute *tangent planes.*

The **tangent plane** of the graph of a function $z = f(x, y)$ at a point (a, b) is

$$y = f(a, b) + f_x(a, b)(x - a) + f_y(a, b)(y - b)$$

(17) Consider $f(x, y) = x^2 + y$. Compute the tangent plane of f at the point $(2, 0)$ and plot both f and this plane in an online graphing calculator.

Definition. A **vector-valued function** is a function whose outputs are vectors.

(18) Often we write vector valued functions with angular braces that look like $\langle\ \rangle$. Consider the function $\vec{F}(x, y) = \langle x^2 - y, \sin(x)\rangle$. Compute $\vec{F}$ at a few points in its domain $\mathbb{R}^2$.

Definition. The gradient vector of a differentiable function $f(x_1, x_2, \ldots, x_n)$ is the vector-valued function

$$\nabla f = \langle f_{x_1}, f_{x_2}, \ldots, f_{x_n}\rangle$$

Definition. The **critical points** of a multivariable function are the points where the gradient vector of the function is the 0-vector $\mathbf{0}$ or undefined.

Note. There are special rules for classifying critical points which involve partial derivatives. We won't mention them here because most of the functions we'll encounter will be so complicated that it would be nearly impossible to find and classify all of their critical points.

(19) Find the critical points of $f(x, y) = e^y(y^2 - x^2)$.

(20) The gradient vector is an incredibly powerful tool, even beyond finding critical points. For, $z = f(x, y) = x^2 + y$, $D = \mathbb{R}^2$,
- (a) compute ∇f,
- (b) determine the tangent plane at the point $(1, 2)$, and
- (c) plot the graph of the function f and the tangent plane of f at $(1, 2)$ in an online graphing calculator.
- (d) Go back to your contour plot in 12. Plot the gradient vector $\nabla f(1, 2)$ so that its starting point is the point $(1, 2)$. Also, sketch the line tangent to the contour $f(1, 2)$ at the point $(1, 2)$ What do you notice? What direction does the gradient vector point towards?
- (e) What do you notice about $-\nabla f(1, 2)$? What direction does it point?

Definition. Gradient descent is an algorithmic approach to searching for a minimum of a function. Let $f(x_1, x_2, \ldots, x_n)$ be a differentiable function and use vector notation to denote the point $\mathbf{x} = (x_1, x_2, \ldots, x_n)$. Then, a **gradient descent step** is

$$\mathbf{x}_{k+1} = \mathbf{x}_k - l\nabla f(\mathbf{x}_k)$$

where l is some constant which, in machine learning, we call the **learning rate**. Note here we get a sequence of points $\{\mathbf{x}_0, \mathbf{x}_1, \ldots\}$. For "nice" situations, with x_0 and l chosen efficiently, this sequence converges to a local minimum of f.

(21) Practice the gradient descent algorithm on the functions below. For each, use an online graphing calculator to visualize what is happening both in $\mathbb{R}^3$ and in a contour plot. Try different values of l.
- (a) $f(x, y) = x^2 + y^2$
- (b) $g(x, y) = \cos(x) + y^2 - \frac{1}{2}x$

(22) Write down some observations you made while working through the previous exercise.