Notes for Section 3.1

Start by understanding what a local minimum/maximum is and what an absolute/global minimum/maximum is, and how these two are different. Figure 2 may help. Finding the minima and maxima of functions is a common problem in the sciences and social sciences and is one of the most important practical applications of calculus. It is not only absolute minima and maxima that matter in practice, local minima and maxima also play significant roles in the quantitative sciences. For example, in chemistry, matter typically wants to move from high energy states to low energy states. In other words, it tends to move toward the absolute minimum of some kind of energy function that describes its state. But a chemical that is at a local minimum of its energy function may be in a stable state even if it is not at the absolute minimum of the energy function. For example, if you mix natural gas (methane) with oxygen at room temperature, nothing much happens. The mixture is stable, even though it is not at the absolute minimum of its possible energy states. It does in fact have a lower energy state: if the gas burns, it releases energy and thus moves to a lower energy state as water and carbon-dioxide. But a methane-oxygen mixture can exist for a long time without any such chemical reaction taking place, because it is at a local minimum of its possible energy states.

A function can have one or more local maxima, or none at all. And it can have one or more absolute maxima (because it can attain the same absolute maximum value at several different inputs), or none at all. The same goes for local and absolute minima. Look at Figures 3-5 to see how such things can happen.

Here are a few words on correct terminology. Minimum and maximum are the singular terms, minima and maxima are the plural. There is no such thing as minimums or maximas, or a minima. Minima and maxima are collectively referred to as extrema (the singular is extremum), or extreme values. So there are local extrema and absolute or global extrema.

It is also an important practical problem to find the extrema of a function whose input variable is subject to some kind of a constraint, for example it must be in some interval. While a function need not have absolute extrema, there is an important special case which guarantees the existence of absolute extrema: a continuous function $f(x)$ over a closed interval $x \in [a, b]$ always has both an absolute minimum and an absolute maximum. Of course, there could be more than one value of x where f attains the same minimum or maximum value. This is what the Extreme Value Theorem says. We will not prove the Extreme Value Theorem because its proof requires ideas that are beyond the level of a course like MCS 119. But if you think about this a little bit, the theorem makes good intuitive sense. Since the graph of f is a continuous curve between the points $(a, f(a))$ and $(b, f(b))$, it must have a highest point and a lowest point. Figures 8 and 9 show you how the conclusion of the Extreme Value Theorem can fail to hold if its conditions are not satisfied. Figure 8 shows a discontinuous function f on the closed interval $[0, 2]$, which has no absolute maximum. Figure 9 shows a continuous function g on the open interval $(0, 2)$, wich has no absolute maximum because its values can get arbitrarily large. Note that g has no absolute minimum either because its vaues include every real numbers $y > 1$, but and there is no smallest among such real numbers.

The next question is how we can find the extrema of a function. Local extrema are easier to deal with, so we will start with those. Fermat's Theorem says that if f has a local minimum or maximum at a point c, then the derivative $f'(c)$ must either be 0 or must not exist. If you think about this a bit, you may find that it also makes good intuitive sense. Let us say f has a local maximum at c. Then f has a local peak at c, so it must be an increasing function just before c and must be a decreasing function just past c. So $f'(x)$ must be positive just before c and $f'(x)$ must be negative just past c. Therefore f' must change its sign as the value of x passes through c. One way to do it is to pass through 0, that is $f'(c) = 0$. You can also think about this in terms of the tangent line. The tangent line must be sloping up just before x gets to c and must start sloping down once x gets past c. So it makes sense that the tangent line may be flat right at c. In fact, it may be hard at first to imagine what else could happen. Why could the derivative $f'(c)$ fail to exist? That is because the local peak at c could look like one of the peaks in the Himalayas: it could be quite pointy. And we know that when the graph of f has a sharp corner (cusp) at a point, f is not differentiable there. In this case, f' is a discontinuous function at c: its values are positive to the left of c and negative to the right of c, but it has not value at c. In other words, it goes from positive to negative by jumping over 0. You can think about the behavior of f at a local minimum in very similar terms.

Read and understand the proof of Fermat's Theorem. It uses limits, in fact one-sided limits, but it is not particulary difficult. The proof given in the book is a perfectly good one, so I will not spell it out here.

It gets tiresome to say "c is a point where $f'(c) = 0$ or $f'(c)$ does not exist," so we refer instead to such a point as a *critical number* or *critical point*. So Fermat's Theorem says that if c is a local extremum of f, then c is a critical point.

It is important to understand that the converse of Fermat's Theorem is not true: a critical point c of f need not be a local extremum. A good example of this is $f(x) = x^3$ at $x = 0$. The derivative $f'(x) = 3x^2$ is 0 at $x = 0$. But f does not have a local extremum at 0. It is not hard to understand what happens at this critical point. To the left of 0, f is increasing, so f' is positive, but the rate of increase is really slowing down as x gets closer to 0. Then at 0, the tangent line become horizontal, for just a brief moment really, before f starts increasing again, and so f' is positive again. Graph $f(x) = x^3$ in one of the Desmos apps that allows you to move the tangent line along the graph.

Another example of a critical point that is not a local extremum is the critical point of $f(x) = \sqrt[3]{x}$ at $x = 0$. We saw in Section 2.2 that f is not differentiable at 0 because its tangent line there is vertical and the slope of a vertical line is not defined. Therefore $f'(0)$ does not exist, so $x = 0$ is a critical point. But it is clear from the graph the that is is neither a local minimum nor a local maximum. Graph this functin in one of the Desmos apps that allows you to move the tangent line along the graph to visualize all this.

Here is why this is important. It is typically not difficult to find the critical points of a function f. You can find the derivative f' and then think about when $f'(c)$ does not exist or when $f'(c) = 0$. Example 5 illustrates how this may work in practice. Now, you know that any local extrema f may have must be among the critical points. But not all critical points have to be local extrema. Yet finding the critical points already narrows down the potential extrema typically to a few possibilities. If you really want to identify which of these are the local extrema, you need a good way to distinguish those critical points that are local extrema from those that are not. We will learn good ways to do this in Section 3.3.

Let us get back to continuous functions on a closed interval $[a, b]$. If f is such a function, the Extreme Value Theorem guarantees that it has an absolute minimum and an absolute maximum. But how can we find those? It is actually not hard to do. Let's think about this a little. Suppose $c \in [a, b]$ is an absolute maximum of f. Then $f(c) \geq f(x)$ for any other $x \in [a, b]$. If $a < c < b$, then c must also be a local maximum because if you look at a neighborhood of c that is small enough to be in $[a, b]$, then it must also be true for every x in that neighborhood that $f(c) \geq f(x)$. So either c is a local maximum or $c = a$ or $c = b$. If c is a local maximum then it must also be a critical point by Fermat's Theorem. A similar argument shows that if d is an absolute minimum then d is either a critical point or is one of the endpoints of the interval. So if we want to find the absolute extrema of f on $[a, b]$, we can proceed to find all of the critical points of f, and then we can compare the values of f at these critical points with each other and with $f(a)$ and $f(b)$. The largest of these values is (or are) the absolute maximum (or maxima) and the smallest is (or are) the absolute minimum (or minima). Typically, there are only a small number of critical points and therefore it is easy enough to calculate the values of f at those and at the two endpoints of the interval. Look at Example 6 for how this can be done.