Prerequisites: Basic Differential Calculus

Text: *Crippen et al.*, 2008 [1] pp. 85–86, 94–96, 99–104

Main Concepts:

- *Limits*
- *Derivatives*
- *Chain Rule*
- *Taylor Series*

Sample Problem, Signal Processing: What is the instantaneous rate of change of a function with a scalar argument?

This is a brief summary of the main concepts that we will use from elementary differential calculus. We will not prove these results because a student is expected to know the prerequisite material. This tutorial is intended to help a student recall these basic ideas.

0.1 A Function with a Scalar Argument

We will deal with optimization objectives that have scalar values. Strictly speaking, we will use *functionals* but the optimization literature names them *functions*. The latter term is more general and we will adopt its use.

The simplest function in this course has a scalar argument that is a real number, so it is a member of the set $\mathbb{R}$. Such a function maps a real number to a real number. The *domain* of a function may be restricted to a subset of the real numbers, such as a finite closed interval; mostly, the domain will be all of $\mathbb{R}$. For brevity, we will refer to any member of a domain as a *point*, which we will write as t for the domain $\mathbb{R}$. We can write a function f as

$$f : \mathbb{R} \to \mathbb{R}$$

A simpler, more familiar notation is

$$f(t)$$

0.2 Limit of a Function that has a Scalar Argument

The modern definition of the *limit* of a function is the Bolzano-Weierstrass definition in terms of values ϵ and δ for a finite limit. We will write the limit of a function f, at a value a, as

$$\lim_{t \to a} f(t) = c$$

0.3 Continuous Function that has a Scalar Argument

Although continuity can be rigorously defined directly using the ϵ–δ formulation, we will take continuity at a value a to mean

$$\lim_{t \to a} f(t) = f(a)$$

This assumes that $f(a)$ exists, which is the usual case in optimization.

0.4 Derivative of a Function that has a Scalar Argument

We will use both the Leibniz notation and the Lagrange notation for a derivative. We will assume that the derivative is the same, regardless of the direction in which the limit is taken. The derivative of f at the point a is

$$\frac{df(a)}{dt} \quad \text{or} \quad \frac{d}{dt} f(a) \quad \text{or} \quad f'(a) \quad \stackrel{\text{def}}{=} \lim_{h \to 0} \frac{f(a+h) - f(a)}{h}$$

Figure 1 illustrates how a *chord*, which connects a point a to a point $a + h$ for a function $f(t)$, has a limit that is the derivative of the function.

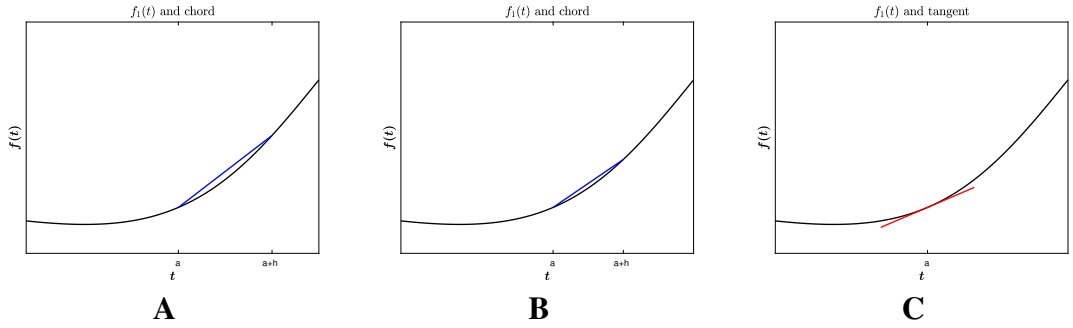

Figure 1: Graphically, the limit of a chord for $f(t)$ between $(a, f(a))$ and $(a + h, f(a + h))$ is the derivative of $f(t)$ at a. (A) A convex function $f(t)$, a scalar argument a, and a chord for a value h. (B) The same function and scalar argument, with a smaller value for h. (C) The derivative of $f(t)$ at a is tangent to the curve of the function.

The usual rules of finding derivatives will be used. In this course, the examples will mainly be polynomials. The derivatives for powers of a variable t, where $t^0 \stackrel{\text{def}}{=} 1$, are

$$\frac{d}{dt} t^a \quad = \quad a t^{a-1}$$

The derivative of a polynomial can be found using basic rules. We will also use trigonometric functions and exponential functions. The basic derivatives for these functions are:

$$\begin{aligned} & \frac{d}{dt}\sin(t) = \cos(t) \\ & \frac{d}{dt}\cos(t) = -\sin(t) \\ & \frac{d}{dt}e^t = e^t \\ (t>0) \rightarrow \quad & \frac{d}{dt}\ln(t) = \frac{1}{t} \end{aligned}$$

0.5 Constant, Sum, and Product Rules for Derivatives

Three simple rules we will use often are for two functions, f and g, which are assumed to be differentiable on the domain $\mathbb{V}$ that is specified. The Constant Rule is

$$\frac{d}{dt}(cf(t)) = c\frac{d}{dt}f(t) \tag{1}$$

The Sum Rule is

$$\frac{d}{dt}(f(t)+g(t)) = \frac{d}{dt}f(t) + \frac{d}{dt}g(t) \tag{2}$$

The Product Rule is also referred to as Leibniz's Rule:

$$\frac{d}{dt}(f(t)g(t)) = \left(\frac{d}{dt}f(t)\right)g(t) + f(t)\left(\frac{d}{dt}g(t)\right) \tag{3}$$

0.6 Chain Rule and Quotient Rule for Derivatives

The rule for the composition of functions is crucially important. For functions f and g that meet the appropriate conditions for existence, domains, and ranges, the composition of f with g is often written in either of two ways:

$$f(g(t)) \qquad \text{or} \qquad f \circ g(t)$$

Using Lagrange notation and composition notation, the Chain Rule is

$$(f \circ g)' = (f' \circ g)\, g' \tag{4}$$

A commonly encountered situation is that variables are defined as functions of other variables. For example, we might specify an independent variable t, and two dependent variables y and z, to be related as

$$z = f(y) \quad \text{and} \quad y = g(t)$$

Using these related variables, and Leibniz notation, the Chain Rule is

$$\frac{d}{dt} f(g(t)) = \frac{dz}{dt} = \frac{dz}{dy}\frac{dy}{dt} \tag{5}$$

Reciprocal Rule

An example application of the Chain Rule is a deduction of the Reciprocal Rule. Consider a function that is the reciprocal of another function; we can write this reciprocal as

$$h(t) = \frac{1}{g(t)} = f(g(t)) \qquad f(g) = g^{-1}$$

Using the Power Rule, we see that $f'(g) = -g^{-2}$. Applying the Chain Rule, the derivative of $h(t) = 1/g(t)$ is

$$\begin{aligned} h'(t) &= (f'(g))g' \\ &= -(g(t))^{-2} g' \\ &= -\frac{g'(t)}{(g(t))^2} \end{aligned} \tag{6}$$

Quotient Rule

Another example is a deduction of the Quotient Rule. Consider a function that composed of a numerator function $f(t)$ and a denominator function $g(t)$, which is

$$h(t) = \frac{f(t)}{g(t)} = f(t)\frac{1}{g(t)}$$

The Quotient Rule can be deduced by using the Reciprocal Rule and the Product Rule. To arrive at a common denominator, we observe that $1/g(t) = g(t)/(g(t))^2$ and thus have

$$\frac{d}{dt}\left(\frac{f(t)}{g(t)}\right) = \frac{\frac{df}{dt}g(t) - f(t)\frac{dg}{dt}}{(g(t))^2} \tag{7}$$

0.7 Taylor Series

An important relevant formula is the *Taylor series*. This is an infinite series for any "nice" function $f(t)$ or, formally, any analytic function. For any function that is analytic near a value $t_0 \in \mathbb{R}$, the Taylor series is defined as

$$\begin{aligned} f(t) &= \sum_{i=0}^{\infty} \frac{f^{(i)}(t_0)}{i!}(t-t_0)^i \\ &= f(t_0) + f'(t_0)(t-t_0) + \frac{f''(t_0)}{2}(t-t_0)^2 + \cdots \end{aligned} \tag{8}$$

We will use the Taylor series to approximate a function for a value of t that is "near" a value t_0. A linear approximation is

$$f(t) \approx f(t_0) + f'(t_0)(t-t_0) \tag{9}$$

and a quadratic approximation is

$$f(t) \approx f(t_0) + f'(t_0)(t-t_0) + \frac{f''(t_0)}{2}(t-t_0)^2 \tag{10}$$

There are many explicit form for the *remainder* term of Equation 8. We will sometimes use the Lagrange error term for a Taylor series, which is an expression in terms of a higher-order derivative. The interval of interest for $t > t_0$ is $[t_0\,,\,t]$ and the interval for $t < t_0$ is $[t\,,\,t_0]$. Assuming that the $(k+1)^{\text{th}}$ derivative is continuous on the closed interval, the Lagrange remainder guarantees that there exists a value ξ in the closed interval such that

$$f(t) = \sum_{i=0}^{k} \frac{f^{(i)}(t_0)}{i!}(t-t_0)^i + \frac{f^{(k+1)}(\xi)}{k!}(t-t_0)^{k+1} \tag{11}$$

In particular, the error of a local linear approximation is bounded by the value in the interval that has the largest second derivative, because

$$f(t) = f(t_0) + f'(t_0)(t-t_0) + \frac{1}{2}f''(\xi)(t-t_0)^2 \tag{12}$$

We will also use the Taylor series for a function that has a vector argument.

0.8 Partial Derivatives

Suppose that a function has two scalar arguments, which is a formal way of saying that it is a function of two variables. We will write each of the two arguments by using a subscript, so we will write the function as

$$f_2(w_1, w_2)$$

Figure 2 illustrates how such a function can be depicted graphically.

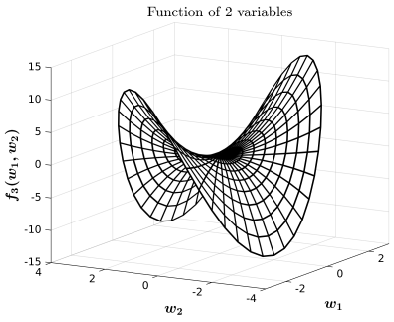

Figure 2: A non-convex function of two variables, w_1 and w_2, produces a surface with points $(w_1, w_2, f(w_1, w_2))$. In this example the function is continuous and is not convex.

Such a function of two variables does not have a single derivative. We will treat the derivatives of such functions in two ways. The first way is to find how the function changes with respect to one variable at a time, where we hold the other variable as locally constant. This is called the *partial derivative* of the function with respect to each variable.

Suppose that we have two scalar variables, w_1 and w_2, and are interested in how $f_2(w_1, w_2)$ behaves around the values $w_1 = a_1$ and $w_2 = a_2$ for constants a_1 and a_2. The partial derivatives of $f_2(w_1, w_2)$ at the values (a_1, a_2) are defined as

$$\begin{aligned} \frac{\partial f_2}{\partial w_1}(a_1, a_2) &\stackrel{\text{def}}{=} \lim_{h \to 0} \frac{f_2(a_1 + h, a_2) - f_2(a_1, a_2)}{h} \\ \frac{\partial f_2}{\partial w_2}(a_1, a_2) &\stackrel{\text{def}}{=} \lim_{h \to 0} \frac{f_2(a_1, a_2 + h) - f_2(a_1, a_2)}{h} \end{aligned} \tag{13}$$

Equation 13 implies that the partial derivative of f_2 with respect to the argument w_1 is the derivative of f_2 where we consider the value of w_2 to be some constant but arbitrary value and we only consider the argument w_1 to be varying. A sometimes useful analogy is to imagine that we

have created a first new function for which only the argument w_1 varies and the argument w_2 is "given". We might write such a function, having a single scalar argument t, as

$$g_2(t|w_2) \stackrel{\text{def}}{=} f_2(t, w_2) \tag{14}$$

Combining Equation 13 with Equation 14, we would have

$$\frac{\partial f_2}{\partial w_1}(a_1, a_2) \stackrel{\text{def}}{=} \frac{d}{dt} g_2(a_1|a_2) \tag{15}$$

The partial derivative of f_2 with respect to the argument w_2 is the derivative of f_2 where we consider the value of w_1 to be some constant but arbitrary value. We might create a second new function for which only the argument w_2 varies and the argument w_1 is "given". We might write such a function, having a single scalar argument t, as

$$h_2(t|w_1) \stackrel{\text{def}}{=} f_2(w_1, t) \tag{16}$$

Combining Equation 13 with Equation 16, we would have

$$\frac{\partial f_2}{\partial w_2}(a_1, a_2) \stackrel{\text{def}}{=} \frac{d}{dt} h_2(a_2|a_1) \tag{17}$$

Equation 13 easily generalizes to n variables.

References

[1] Crippin P, Donato R, Kirkpatrick C: Calculus and Vectors. Nelson Education Limited, 2008