Definition: local minimizer of $f(\vec{w})$

For any $\vec{w} \in \mathbb{R}^n$, and any $f : \mathbb{R}^n \to \mathbb{R}$, a vector $\vec{t^*} \in \mathbb{R}^n$ is a *local minimizer* of f is defined as

$$\exists r \in \mathbb{R}_{++}(\|\vec{w} - \vec{t^*}\| < r) \rightarrow (f(\vec{t^*}) \leq f(\vec{w})) \tag{1.11}$$

Observation: In the definitions of minimizers, the function f can be discontinuous, e.g., can be defined over the domain of integers.

We will occasionally need to use a few basic concepts from point-set topology. The main concepts are interior points, boundary points, and open sets. Because our points are always members of vector spaces, our definitions are relatively straightforward.

An *interior point* of a set $\mathbb{V} \subseteq \mathbb{R}^n$ is a point that has a "ball" of points around it that are all in $\mathbb{V}$.

Definition: interior point of $\mathbb{V}$

For any $\mathbb{V} \subseteq \mathbb{R}^n$ and any $\vec{u} \in \mathbb{V}$, the vector $\vec{u}$ is an *interior point* of $\mathbb{V}$ is defined as

$$\exists r \in \mathbb{R}_{++}, \forall \vec{w} \in \mathbb{R}^n \left((\|\vec{w} - \vec{u}\| < r) \rightarrow (\vec{w} \in \mathbb{V})\right) \tag{1.12}$$

A *boundary point* is a point in a set that is not an interior point.

Definition: boundary point of $\mathbb{V}$

For any $\mathbb{V} \subseteq \mathbb{R}^n$ and any $\vec{u} \in \mathbb{V}$, the vector $\vec{u}$ is a *boundary point* of $\mathbb{V}$ is defined as

$$\neg\left(\exists r \in \mathbb{R}_{++}, \forall \vec{w} \in \mathbb{R}^n \left((\|\vec{w} - \vec{u}\| < r) \rightarrow (\vec{w} \in \mathbb{V})\right)\right) \tag{1.13}$$

A set $\mathbb{V} \subseteq \mathbb{R}^n$ is an *open set* if $\mathbb{V}$ contains only interior points.

Definition: open set

A set $\mathbb{V} \subseteq \mathbb{R}^n$ is an *open set* of $\mathbb{V}$ is defined as

$$\forall \vec{u} \in \mathbb{V}, \exists r \in \mathbb{R}_{++}, \forall \vec{w} \in \mathbb{R}^n \left((\|\vec{w} - \vec{u}\| < r) \rightarrow (\vec{w} \in \mathbb{V})\right) \tag{1.14}$$

References

[1] Beck A: Introduction to Nonlinear Optimization: Theory, Algorithms, and Applications with MATLAB. Siam Press, 2014

CISC 371 Class 2

Relevant Scalar Differential Calculus

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

Main Concepts:

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

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

This class 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.

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

$$f(t)$$

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

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

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

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

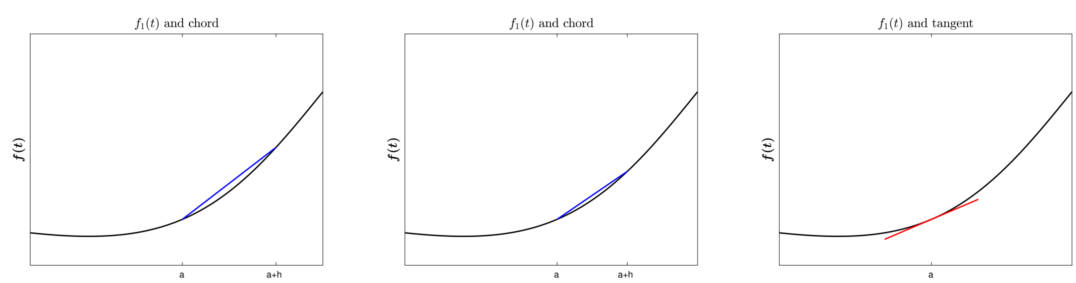
This assumes that $f(a)$ exists, which is the usual case in optimization.

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

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

Figure 2.1 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.



A **B** **C**

Figure 2.1: 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 $t^0 \overset{\text{def}}{=} 1$, are

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

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:

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

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

$$\frac{d}{dt}(cf(t)) = c\frac{d}{dt}f(t) \tag{2.1}$$

The Sum Rule is

$$\frac{d}{dt}(f(t) + g(t)) = \frac{d}{dt}f(t) + \frac{d}{dt}g(t) \tag{2.2}$$

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

$$\frac{d}{dt}(f(t)g(t)) = \left(\frac{d}{dt}f(t)\right) g(t) + f(t) \left(\frac{d}{dt}g(t)\right) \tag{2.3}$$

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

$$(f \circ g)' = (f' \circ g)\, g' \tag{2.4}$$

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

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

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

$$\frac{d}{dt} f(g(t)) = \frac{dz}{dt} = \frac{dz}{dy}\frac{dy}{dt} \tag{2.5}$$

$$\frac{d}{dt}\left(\frac{f(t)}{g(t)}\right) = \frac{\frac{df}{dt} g(t) - f(t)\frac{dg}{dt}}{g^2} \tag{2.6}$$

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

$$\begin{aligned} f(t) &= \sum_{i=0}^{\infty} \frac{f^{(i)}(t_0)}{i!}(t - t_0)^i \qquad (2.7)\\ &= f(t_0) + f'(t_0)(t - t_0) + \frac{f''(t_0)}{2}(t - t_0)^2 + \cdots \end{aligned}$$

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

$$f(t) \approx f(t_0) + f'(t_0)(t - t_0) \tag{2.8}$$

and a quadratic approximation is

$$f(t) \approx f(t_0) + f'(t_0)(t - t_0) + \frac{f''(t_0)}{2}(t - t_0)^2 \tag{2.9}$$

There are many explicit form for the *remainder* term of Equation 2.7. 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

$$f(t) = \sum_{i=0}^{k} \frac{f^{(i)}(t_0)}{i!}(t - t_0)^i + \frac{f^{(k+1)}(\xi)}{k!}(t - t_0)^{k+1} \tag{2.10}$$

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

$$f(t) = f(t_0) + f'(t_0)(t - t_0) + \frac{1}{2}f''(\xi)(t - t_0)^2 \tag{2.11}$$

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

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 2.2 illustrates how such a function can be depicted graphically.

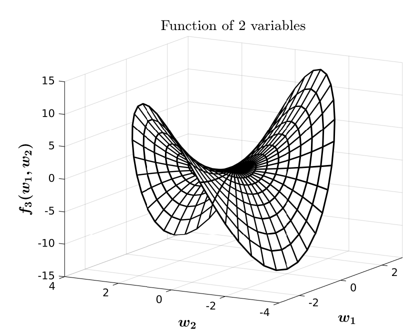


Figure 2.2: A non-convex function of two variables, w_1 and w_2, produces a surface with points $(w_1, w_2, f(w_1, w_2))$. In this example the function is continuous and is not convex.

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