1 Differentiation in $\mathbb{R}^n$

Recall first the usual definition of differentiation of functions.

Definition 1 *A function $f : \mathbb{R} \to \mathbb{R}$ is differentiable at $x \in \mathbb{R}$ if there is a number $f'(x)$ such that*

$$\lim_{h \to 0} \frac{f(x+h) - f(x)}{h} = f'(x). \tag{1}$$

By rewriting 1 we obtain a new geometric interpretation,

$$f(x+h) - f(x) = f'(x)h + r(h), \tag{2}$$

where the remainder is small in the sense that

$$\lim_{h \to 0} \frac{r(h)}{h} = 0,$$

which we will write as $r(h) = o(h)$.

With this definition at hand, we can interpret the derivative in a slightly different manner. Notice that 2 expresses the behavior of the original function f around a point x (i.e. at points of the form $x+h$ with h small) as a sum of a linear function in h, $f'(x)h$ and the error term $r(h)$. In other words, the operation of taking derivatives is viewed as assigning linear functions that approximate the original function around a given point to all points in the domain of f.

The equation 1 certainly makes no sense in the general case of a function $f : \mathbb{R}^n \to \mathbb{R}^m$ but the view of the derivative as a local linear approximation to f around a point x certainly does.

Definition 2 *A function $f : \mathbb{R}^n \to \mathbb{R}^m$ is differentiable at $x \in \mathbb{R}^n$ if there is a linear function $\lambda : \mathbb{R}^n \to \mathbb{R}^m$ such that*

$$\lim_{h \to 0} \frac{\|f(x+h) - f(x) - \lambda h\|}{\|h\|} = 0.$$

The linear transformation λ is denoted $Df(x)$ and called **the derivative** (or differential or total derivative) of f at x. The matrix of $Df(x) : \mathbb{R}^n \to \mathbb{R}^m$ is a $m \times n$ matrix and is called the Jacobian matrix of f at x. If $f : \mathbb{R}^n \to \mathbb{R}$, then the acobian matrix is a row vector.

Proposition 1 *If a function $f : \mathbb{R}^n \to \mathbb{R}^m$ is differentiable at $x \in \mathbb{R}^n$ then there is a* ***unique linear function*** *$\lambda : \mathbb{R}^n \to \mathbb{R}^m$ such that*

$$\lim_{h \to 0} \frac{\|f(x+h) - f(x) - \lambda h\|}{\|h\|} = 0.$$

Proposition 2 ***(Chain Rule)*** *Suppose E is an open set in $\mathbb{R}^n$, f maps E into $\mathbb{R}^m$, f is differentiable at $x_0 \in E$, g maps an open set containing $f(E)$ into $\mathbb{R}^k$, and g is differentiable at $f(x_0)$. Then the mapping F of E into $\mathbb{R}^k$ defined by*

$$F(x) = g(f(x)),$$

is differentiable at x_0 and

$$DF(x_0) = Dg(f(x_0))\, Df(x_0).$$

Let f be a function, $f : \mathbb{R}^n \to \mathbb{R}^m$. The **components** of f are the real valued functions $f_1, ..., f_m$ defined by

$$f(x) = \sum_{i=1}^{m} f_i(x)\, e_i$$

or equivalently $f_i(x) = f(x)\, e_i$, where e_i is the unit vector in the $i - th$ direction. For $x \in \mathbb{R}^n$, $1 \leq i \leq m$, $1 \leq j \leq n$, we define

$$D_j f_i = \lim_{h \to 0} \frac{f_i(x + he_j) - f_i(x)}{h},$$

provided that the limit exists. Writing $f_i(x_1, ..., x_n)$ in place of $f_i(x)$ we see that $D_j f_i$ is the derivative of f_i with respect to x_j:

$$D_j f_i(x) = \frac{\partial f_i}{\partial x_j},$$

and is called a partial derivative.

Proposition 3 *If* $f : \mathbb{R}^n \to \mathbb{R}^m$ *is differentiable at* a*, then* $D_j f_i$ *exists for* $1 \le i \le m$*,* $1 \le j \le n$ *and* $Df(x)$ *is the* $m \times n$ *matrix* $D_j f_i(x)$*.*

Proposition 4 *If* $f : \mathbb{R}^n \to \mathbb{R}^m$*, then* $Df(x)$ *exists if all* $D_j f_i(x)$ *exist and are continuous in an open set containing* x*.*

A function that has continuous partial derivatives in an open set containing x is called continuously differentiable at x.

1.1 Partial Derivative, Directional Derivative and Gradient

In this section, we consider functions $f : \mathbb{R}^n \to \mathbb{R}$.

Definition 3 *The directional derivative of* f *at* x *in direction* $u \in \mathbb{R}^n$ *is denoted by the symbol* $f'(x; u)$ *and is defined by the equation*

$$f'(x; u) = \lim_{h \to 0} \frac{f(x + hu) - f(x)}{h},$$

whenever the limit on the right exists.

If $u = e_i$, the $i - th$ unit coordinate vector, then $f(x, e_i)$ is called the partial derivative and is denoted by $D_i f(x)$. If $f(x; u)$ exists in ecery direction, then in particular all the partial derivatives $D_1 f(x), ..., D_n f(x)$ exist. The converse is not true.

Example 1 *Consider the real-valued function* $f : \mathbb{R}^2 \to \mathbb{R}$ *given by*

$$f(x, y) = \begin{cases} x + y & \text{if } x = 0 \text{ or } y = 0, \\ 1 & \text{otherwise.} \end{cases}$$

A rather surprising fact is that a function can have a finite directional derivative $f(x; u)$ for every u but may fail to be even continuous and therefore at x.

Example 2 *Let*

$$f(x,y) = \begin{cases} \frac{xy^2}{(x^2+y^4)} & \text{if } x \neq 0, \\ 0 & \text{if } x = 0. \end{cases}$$

For this reason, directional derivatives as well as partial derivatives are a somewhat unsatisfactory extension of the one-dimesnional concept of derivative. The generalization that we defined earlier does not suffer from these defects..

Consider next a real valued function $f : \mathbb{R}^n \to \mathbb{R}$ Associated with each $x \in \mathbb{R}^n$ is a vector, the gradient of f at x, defined by

$$\nabla f(x) = Df(x)^\top .$$

The geometric interpretation of the gradient of a real-valued function of a vector variable is worth noting. Recall that

$$y - f(\hat{x}) = f'(\hat{x})(x - \hat{x})$$

is the equation in the hyperplane in $\mathbb{R}^{n+1}$, which is the tangent to $y = f(x)$ where $x = \hat{x}$.

$$\lim_{h \to 0} \frac{f(x+hu) - f(x)}{h} = \nabla f(x) \cdot u, \tag{3}$$

we define the directional derivative of f at x in the direction of the unit vector u, and is denoted by $(D_u f)(x)$. If f and x are fixed, but u varies, then 3 shows that $(D_u f)(x)$ attains its maximum when u is a positive scalar of $(\nabla f)(x)$. In other words, the gradient of a function gives the direction in which the function increases at its quickest rate.

Let $v \in \mathbb{R}^n$ be such that $\nabla f(x) \cdot v = 0$. Then the value of the function is locally constant around x in direction v. The set of points $\{x \in \mathbb{R}^n \,|f(x) = c\}$ is called the level curve of f at value c. This argument shows that the gradient is perpendicular to the tangent to the level curve of the function.

2 Taylor's Formula

The basic objective in differentiating a function is to obtain a linear approximation of the function under analysis. This allows for an easier analysis of the function locally around the point at which it is differentiated. For some questions, the linear approximation contains all the information that is really needed. For example, in the study of dynamical systems (which form an essential part of modern macroeconomics), it is often enough to linearize the system around a steady state of the system in order to analyze the local behavior of the system around that steady state. For other questions, the information contained in the linear approximation is insufficient. The primary instance of this is optimization theory. In this area, the basic task is to locate the points in the domain of a function where that function achieves its maximal (or minimal) value. Linear approximations are used in locating these points, but in order to tell apart the points where the maximum is obtained from the points where the function is minimized, more information is needed.

In this section, a function is approximated around a fixed point x by a polynomial. This approximation will obviously be at least as accurate as the linear approximation obtained in the theory of differentiation (since linear functions are also polynomials). The information contained in the polynomial approximation of a function is sufficient for the purposes of finding minima nad maxima.

We start by recalling some facts from eelementary calculus.

Proposition 5 ***(Rolle).*** *Assume f has a derivative at each point of an open interval (a, b), and assume that f is continuous at both endpoints a and b. If $f(a) = f(b)$ there is at least one interior point c at which $f'(c) = 0$.*

Proposition 6 ***(Generalized Mean Value Theorem)*** *Let f and g be two functions, each having a derivative at each point of an open interval (a, b)*

and each continuous at the endpoints a and b. Assume also that there is no interior point x at which both f'(x) and g'(x) are infinite. Then for some interior point c we have

$$f'(c)[g(b) - g(a)] = g'(c)[f(b) - f(a)].$$

These two results can be used to prove the famous approximation result for continuously differentiable real valued functions on the real line.

Proposition 7 ***(Taylor's theorem for*** $\mathbb{R}$***).*** *Let f be a function having finite n-th derivative $f^{(n)}$ everywhere in an open interval (a, b) and continuous $(n-1)$ st derivatives in the closed interval $[a, b]$. Assume that $c \in [a, b]$. Then, for every $x \in [a, b]$ and $x \neq c$, there exists a point x_1 interior to the interval joining x and c such that:*

$$f(x) = f(c) + \sum_{k=1}^{n-1} \frac{f^{(k)}(c)}{k!}(x-c)^k + \frac{f^{(n)}(x_1)}{n!}(x-c)^n.$$

Taylor's formula can be extended to real-valued functions f defined on subsets of $\mathbb{R}^n$. In order to state the general theorem in a form which resembles the one-dimensional case, we introduce special symbols

$$f''(x;t), \; f'''(x;t), ..., f^{(m)}(x;t),$$

for certain sums that arise in Taylor's formula. These play the role of higher order directional derivatives, and they are defined as follows. If x is a point in $\mathbb{R}^n$ where all second-order partial derivatives of f exist, and if $t = (t_1, ..., t_n)$ is an arbitrary point in $\mathbb{R}^n$, we write

$$f''(x;t) = \sum_{i=1}^{n} \sum_{j=1}^{n} D_{ij} f(x) t_j t_i.$$

We also define

$$f'''(x;t) = \sum_{i=1}^{n} \sum_{j=1}^{n} \sum_{k=1}^{n} D_{ijk} f(x) t_k t_j t_i$$

if all third-order partial derivatives exist at $\mathbf{x}$. The symbol $f^{(m)}(x;t)$ is similarly defined if all $m-th$ order partials exist.

Proposition 8 ***(Taylor's theorem for*** $\mathbb{R}^n$*). Assume that f and all its partial derivatives of order $< m$ are differentiable at each point of an open set S in $\mathbb{R}^n$. If a and b are two points of S such that $L(a,b) \subseteq S$, then there is a point z on the line segment $L(a,b)$ such that*

$$f(b) - f(a) = \sum_{k=1}^{m-1} \frac{f^{(k)}(a; b-a)}{k!} + \frac{f^{(m)}(z; b-a)}{m!}.$$

In the lectures on optimization, we will use the properties of $f''(x;t)$ to classify points where $f'(x) = 0$ as either local maxima or local minima of the function around that point.

3 The Implicit Function Theorem

The functions we analyzed so far were almost all explicit and of the form

$$y = f(x),$$

where the exogenous variable x appeared on the right hand side and the endogenous variable appeared on the left hand side. When endogenous and exogenous variables cannot be separated as in

$$f(x,y) = c, \tag{4}$$

then they define an implicit function. The problem is to decided whether this equation determines y as a function of x. If so, we have

$$y = g(x),$$

for some function g. We say that g is defined implicitly by (4). The problem assumes a more general form when we have a system of several equations involving several variables in terms of the remaining variables. An important special case is the familiar problem in algebra of solving n linear equations of the form

$$\sum_{j=1}^{n} a_{ij}y_j = t_i, \quad (i = 1, 2, ..., n), \tag{5}$$

Each equation (5) can be written in the form of (4) by setting

$$f_i(y, x) = \sum_{j=1}^{n} a_{ij}y_j - y_i = 0.$$

More generally, consider the linear function $f : \mathbb{R}^{n+k} \to \mathbb{R}^n$ given by the $n \times (n + k)$ matrix A. Denote the first n components of $\mathbb{R}^{n+k}$ by $y = (y_1, ..., y_n)$ and the last k components by $x = (x_1, ..., x_k)$. Suppose that we are interested in the behavior of the component $\mathbf{y}$ in the solutions to the equation

$$A(y, x) = 0.$$

Using the results from linear algebra, we may try to solve explicitly for the $y's$ in terms of the $x's$. For this purpose, write

$$A = (A_y, A_x)$$

where

$$A_y = \begin{pmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & \ddots & \vdots \\ a_{n1} & \cdots & a_{nn} \end{pmatrix} \text{ and } A_x = \begin{pmatrix} a_{1(n+1)} & \cdots & a_{1(n+k)} \\ \vdots & \ddots & \vdots \\ a_{n(n+1)} & \cdots & a_{n(n+k)} \end{pmatrix}.$$

Then the linear equation can be written as

$$A_y y + A_x x = 0,$$

and thus we can solve for y:

$$y = -A_y^{-1} A_x x.$$

Clearly, this can only be done if A_y is invertible, i.e. if $\det(A_y) \neq 0$.

In this section, our goal is to extend this analysis beyond linear functions. Unfortunately, it is in general impossible to find explicit solutions for n non-linear equations in n unknowns (the y above). Nevertheless, we can sometimes achieve a more modest goal by looking at the set of equations locally around a solution to the system. Recall that the basic method of local analysis was to linearize the function in question around a given point and analyze the linear approximation. The derivative of the function gives us the desired approximation, and we can use a similar line of argument to get local results as we did in thelinear example above.

More specifically, for a given implicit function $f(y, x)$ we want to know the answers to the following two questions:

1. does $f(y^0, x^0)$ determine y^0 as a continuous function of x for x near x^0?

2. how does a change in $\mathbf{x}$ affect the corresponding $\mathbf{y}'s$?

The answer to these questions is provided in the following theorem.

Theorem 9 ***(Implicit Function Theorem).*** *Let $f = (f_1, ..., f_n)$ be a vector-valued function defined on an open set S in $\mathbb{R}^{n+k}$ with values in $\mathbb{R}^n$. Suppose $f \in C^1$ on S. Let $(y_0; x_0)$ be a point in S for which $f(y^0; x^0) = 0$ and for which the $n \times n$ determinant $\det[D_y f(y^0; x^0)] \neq 0$. Then there exists a k-dimensional open set X^0 containing x^0 and a unique vector valued function g defined on X^0 and having values in $\mathbb{R}^n$, such that*

(a) $g \in C^1$ on T^0,

(b) $g(x^0) = y^0$,

(c) $f[g(x); x] = 0$, for every $x \in X^0$.

(*d*) $D_{\mathbf{x}}g(x^0) = -(D_y f(y^0, x^0))^{-1} D_x f(y^0, x^0)$, *where*

$$D_y f\left(y^0, x^0\right) = \begin{pmatrix} \frac{\partial f_1}{\partial y_1} & \cdots & \frac{\partial f_1}{\partial y_n} \\ \vdots & \ddots & \vdots \\ \frac{\partial f_n}{\partial y_1} & \cdots & \frac{\partial f_n}{\partial y_n} \end{pmatrix} \text{ and } D_x f\left(y^0, x^0\right) = \begin{pmatrix} \frac{\partial f_1}{\partial x_1} & \cdots & \frac{\partial f_1}{\partial x_k} \\ \vdots & \ddots & \vdots \\ \frac{\partial f_n}{\partial x_1} & \cdots & \frac{\partial f_n}{\partial x_k} \end{pmatrix}.$$

Another important result from multivariate calculus that uses the local linear approximation of the original function by its derivative is the following.

Proposition 10

Theorem 11 (Inverse Function Theorem). *Assume* $f = (f_1, ..., f_n) \in C^1$ *on an open set* S *in* $\mathbb{R}^n$, *and let* $T = f(S)$. *If the Jacobian determinant* $J_{\mathbf{f}}(a) \neq 0$ *for some point* $\mathbf{a} \in S$, *then there are two open sets* $X \subseteq S$ *and* $Y \subseteq T$ *and a uniquely determined function* g *such that:*

(*a*) $a \in X,\ f(y) \in \mathbf{Y}$,

(*b*) $Y = f(X)$,

(*c*) f *is one-to-one on* X,

(*d*) g *is defined on* Y, $g(Y) = X$, *and* $g[f(x)] = x$, *for every* $x \in X$,

(*e*) $g \in C^1$ *on* Y.

The Inverse Function Theorem is a powerful result. In general, it is nearly impossible to check directly that a given map is one-to-one and onto. By Theorem 11 we need only to show that its derivative is a nonsingular matrix - a much easier task.

Here are two examples of the use of Implicit Function Theorem that are typical in Economics.

Example 3 *Let* $f : \mathbb{R}^n \to \mathbb{R}$. *The level curve of* f *at value* c *is defined as the set* $\{x \in \mathbb{R}^n \,|f(x) = c\}$. *We may then ask when one of the endogenous variables, say* x_1 *is defined implicitly as a function of the other variables* $(x_2, ..., x_n) \in \mathbb{R}^{n-1}$ *around a point* x^* *on the level curve, i.e. at a point*

x^* satisfying $f(x^*) = c$. Implicit Function Theorem shows that whenever $\frac{\partial f(x^*)}{\partial x_1} \neq 0$, there is a function $g : \mathbb{R}^{n-1} \to \mathbb{R}$ and an $\varepsilon > 0$ such that

i) $x_1^* = g(x_2^*, ..., x_n^*)$,

ii) For all $(x_2, ..., x_n) \in B_\varepsilon((x_2^*, ..., x_n^*))$, $f(g(x_2, ..., x_n), (x_2, ..., x_n)) = c$,

iii) $\frac{\partial g((x_2^*,...,x_n^*))}{\partial x_i} = -\frac{\frac{\partial f(x^*)}{\partial x_i}}{\frac{\partial f(x^*)}{\partial x_1}}$.

Observe that the last conclusion follows from ii) by the chain rule. The interpretation of the result is the following. Suppose that one of the variables, x_i is increased by a small amount dx_i. The change in $x_1$needed in order to keep the value of the function f constant at c is then given by

$$-\frac{\frac{\partial f(x^*)}{\partial x_i}}{\frac{\partial f(x^*)}{\partial x_1}}.$$

In microeconomics, the problem of utility maximization results in level curves that are called indifference curves (since the buyer is indifferent between all those points as the utility index corresponding to them is constant). Expressions of the form $\frac{\frac{\partial f(x^*)}{\partial x_i}}{\frac{\partial f(x^*)}{\partial x_j}}$ are called the marginal rates of substitution between i and j.

Example 4 *Implicit function theorem is the primary tool of what is called the comparative statics analysis of economic models. In this analysis, economic considerations lead to behavioral equations linking the endogenous quantities (e.g. price and quantity of oil demanded and supplied) to exogenous variables (weather conditions, outbreak of a war etc.). The main task of the analysis is to predict how the endogenous variables will change in response to changes in the exogenous variables.*

Let y_1 denote the quantity of oil demanded, let y_2 denote the quantity of oil supplied and let y_3 denote the price of oil. Let also x_1 denote the average temperature (say that we are talking about the winter months) and x_2 denote a variable recording the military activity in the middle east. From *the consumers' optimal oil purchasing decisions, we can derive the demand function for oil:*

$$y_1 = M(y_3, x_1).$$

From the producers' optimal production decisions, we can derive the supply function for oil:

$$y_2 = S(y_3, x_2).$$

Finally, a market is in equilibrium if:

$$y_1 = y_2.$$

Using the notation from the implicit function theorem, we can write

$$f_1 = M(y_3, x_1) - y_1,$$

$$f_2 = S(y_3, x_2) - y_2,$$

$$f_3 = y_1 - y_2.$$

Then any point $y^0, x^0 = (y_1^0, y_2^0, y_3^0, x_1^0, x_2^0)$ *satisfying*

$$f\left(y^0, x^0\right) = \left(f_1\left(y^0, x^0\right), f_2\left(y^0, x^0\right), f_3\left(y^0, x^0\right)\right) = 0$$

gives an equilibrium of the economic model.

In order to see if implicit function theorem is applicable, we need to evaluate $D_y f(y^0, x^0)$. *But this is clearly:*

$$D_y f\left(y^0, x^0\right) = \begin{pmatrix} -1 & 0 & \frac{\partial M\left(\mathbf{y}^0, \mathbf{x}^0\right)}{\partial y_3} \\ 0 & -1 & \frac{\partial S\left(\mathbf{y}^0, \mathbf{x}^0\right)}{\partial y_3} \\ 1 & -1 & 0 \end{pmatrix}.$$

Hence $\det D_y f(y^0, x^0) = \frac{\partial M\left(\mathbf{y}^0, \mathbf{x}^0\right)}{\partial y_3} - \frac{\partial S\left(\mathbf{y}^0, \mathbf{x}^0\right)}{\partial y_3}$. *In any reasonable model of demand and supply,* $\frac{\partial M\left(\mathbf{y}^0, \mathbf{x}^0\right)}{\partial y_3} < 0$ *and* $\frac{\partial S\left(\mathbf{y}^0, \mathbf{x}^0\right)}{\partial y_3} > 0$, *i.e. demand is downward sloping and supply is upward sloping. Hence the deteminant is megative and* *thus nonzero. As a result, the implicit function theorem is applicable. Hence we may write around the initial point* (y^0, x^0) $y = g(x^0)$ *, where* $g : \mathbb{R}^2 \to \mathbb{R}^3$. *Since*

$$D_x f\left(y^0, x^0\right) = \begin{pmatrix} \frac{\partial M\left(\mathbf{y}^0,\mathbf{x}^0\right)}{\partial x_1} & 0 \\ 0 & \frac{\partial S\left(\mathbf{y}^0,\mathbf{x}^0\right)}{\partial x_2} \\ 0 & 0 \end{pmatrix},$$

$$D_x g\left(x^0\right) = -\frac{1}{\frac{\partial M(\mathbf{y}^0,\mathbf{x}^0)}{\partial y_3} - \frac{\partial S(\mathbf{y}^0,\mathbf{x}^0)}{\partial y_3}} \times$$
$$\begin{pmatrix} \frac{\partial S\left(\mathbf{y}^0,\mathbf{x}^0\right)}{\partial y_3} & -\frac{\partial M\left(\mathbf{y}^0,\mathbf{x}^0\right)}{\partial y_3} & \frac{\partial M\left(\mathbf{y}^0,\mathbf{x}^0\right)}{\partial y_3} \\ \frac{\partial S\left(\mathbf{y}^0,\mathbf{x}^0\right)}{\partial y_3} & -\frac{\partial M\left(\mathbf{y}^0,\mathbf{x}^0\right)}{\partial y_3} & \frac{\partial S\left(\mathbf{y}^0,\mathbf{x}^0\right)}{\partial y_3} \\ 1 & -1 & 1 \end{pmatrix} \begin{pmatrix} \frac{\partial M\left(\mathbf{y}^0,\mathbf{x}^0\right)}{\partial x_1} & 0 \\ 0 & \frac{\partial S\left(\mathbf{y}^0,\mathbf{x}^0\right)}{\partial x_2} \\ 0 & 0 \end{pmatrix}$$

It makes sense to assume that a decrease in mean temperature increases the demand for oil at all prices (during winter months). Thus $\frac{\partial M\left(\mathbf{y}^0,\mathbf{x}^0\right)}{\partial x_1} < 0$ *and we can calculate the effect of a change in temperature on the oil price (i.e.* $\frac{\partial y_3}{\partial x_1}$*) by evaluating the term* $\frac{\partial g_3\left(\mathbf{y}^0,\mathbf{x}^0\right)}{\partial x_1} = -\frac{\frac{\partial M\left(\mathbf{y}^0,\mathbf{x}^0\right)}{\partial x_1}}{\frac{\partial M(\mathbf{y}^0,\mathbf{x}^0)}{\partial y_3} - \frac{\partial S(\mathbf{y}^0,\mathbf{x}^0)}{\partial y_3}} < 0.$

4 Exercises

1. A function $\mathbf{f} : \mathbb{R}^2 \to \mathbb{R}^2$ is given. Verify that the inverse function theorem is applicable and find the inverse function $\mathbf{f}^{-1}$.

   (i) $y_1 = x_1,\ y_2 = x_1^2 + x_2,$

   (ii) $y_1 = 2x_1 - 3x_2,\ y_2 = x_1 + 2x_2.$

2. In each of the next two problems a function F and a point P are given. Verify that the implicit function theorem is applicable. Denoting the implicitly defined function by g, find the values of all the first partial derivatives of g at P.

$(i)\, F = (F^1, F^2)\,,\ P = (0,0,0,0)\,$, where $F^1 = 2x_1 - 3x_2 + y_1 - y_2,\ F^2 = x_1 + 2x_2 + y_1 - 2y_2,$

$(ii)\, F = (F^1, F^2)\,,\ P = (0,0,0,0)\,$, where $F^1 = 2x_1 - x_2 + 2y_1 - y_2,\ F^2 = 3x_1 + 2x_2 + y_1 + 2y_2.$

3. Consider the following system of equations:

$$f_1(y_1, y_2; \alpha, \beta, q) = y_1 + \alpha y_2^2 - q = 0$$
$$f_2(y_1, y_2; \alpha, \beta, q) = \beta y_1^2 + y_2 - q = 0$$

i)Can you solve (y_1, y_2) as functions of (α, β, q) in a neighborhood of the point $(y_1, y_2) = (1,1)\,,\ (\alpha, \beta, q) = (1,1,2)$?

ii) Find the matrix of partial derivatives:

$$\begin{pmatrix} \frac{\partial y_1}{\partial \alpha} & \frac{\partial y_1}{\partial \beta} & \frac{\partial y_1}{\partial q} \\ \frac{\partial y_2}{\partial \alpha} & \frac{\partial y_2}{\partial \beta} & \frac{\partial y_2}{\partial q} \end{pmatrix}.$$

4. The directional derivative of $f : \mathbb{R}^n \to \mathbb{R}$ is given by the formula

$$\lim_{h \to 0} \frac{f(x + hu) - f(x)}{h} = \nabla f(x) \cdot u.$$

Show that for all u such that $\|u\| = 1$,

$$\nabla f(x) \cdot u \leq \frac{\nabla f(x) \cdot \nabla f(x)}{\|\nabla f(x)\|},$$

thus verifying that the gradient gives the largest directional derivative for the function.

5. A function $f : \mathbb{R} \to \mathbb{R}$ is said to be of class C^∞ if it has continuous derivatives of all orders. f is said to be analytic, if it can be expressed as a limit of sums of power functions, i.e.

$$f(x) = \sum_{n=0}^{\infty} a_n x^n \qquad (6)$$

(i.e. the limit on the right hand side must be well defined). Obviously all polynomials are analytic functions.

Define the exponential function as follows:

$$e^x = \sum_{n=0}^{\infty} \frac{x^n}{n!}.$$

Show that the series on the right converges for all x and hence the exponential function is also analytic.

It is a well known theorem for analytic functions that the derivatives of f can be obtained by differentiating the series of functions term by term, i.e.

$$f'(x) = \sum_{n=0}^{\infty} n a_n x^{n-1}. \tag{7}$$

Use this fact to evaluate the derivative of the exponential function and show that the exponential function has a well defined inverse called the logarithmic function. Evaluate the derivative of the logarithmic function.

Deduce from 6 and 7 Taylor's theorem for the case where f is analytic:

$$f(x) = \sum_{n=0}^{\infty} \frac{f^{(n)}(a)}{n!} (x-a)^n. \tag{8}$$

Notice that all analytic functions are in class C^∞ since their derivatives exist for all orders. Finally, show that there are C^∞ functions that are not analytic. To achieve this, evaluate the derivatives of the following function

$$f(x,y) = \begin{cases} e^{-\frac{1}{x^2}} & \text{if } x \neq 0, \\ 0 & \text{if } x = 0. \end{cases}$$

Conclude that 8 fails for $a = 0$.

6. Let

$$f\left(x,y\right)=\begin{cases}\frac{xy^2}{(x^2+y^4)} & \text{if } x\neq 0,\\ 0 & \text{if } x=0.\end{cases}$$

Prove that f has well defined finite directional derivatives in all directions in $\mathbb{R}^2$ and prove also that f is not continuous at 0.