

**UNIVERSIDAD DEL CEMA  
Buenos Aires  
Argentina**

Serie  
**DOCUMENTOS DE TRABAJO**

**Área: Matemática y Economía**

**Intuitive Mathematical Economics Series.  
Chain Rule and Derivatives of Functions Defined Implicitly**

**Sergio A. Pernice**

**Diciembre 2018  
Nro. 679**

**[www.cema.edu.ar/publicaciones/doc\\_trabajo.html](http://www.cema.edu.ar/publicaciones/doc_trabajo.html)  
UCEMA: Av. Córdoba 374, C1054AAP Buenos Aires, Argentina  
ISSN 1668-4575 (impreso), ISSN 1668-4583 (en línea)  
Editor: Jorge M. Streb; asistente editorial: Valeria Dowding [jae@cema.edu.ar](mailto:jae@cema.edu.ar)**



# Intuitive Mathematical Economics Series

## Chain Rule and Derivatives of Functions Defined Implicitly

SERGIO A. PERNICE<sup>1</sup>

Universidad del CEMA  
Av. Córdoba 374, Buenos Aires, 1054, Argentina

December 26, 2018

### Abstract

In this paper we present some elements of calculus for economics: the chain rule and extended chain rule for calculation of derivatives of composite functions, and differentiation of functions defined implicitly. The emphasis, as always in this series, is in providing a pedagogical, intuitive presentation to these topics.

*Keywords:* Calculus, chain rule, implicit derivatives.

## 1 Introduction

This paper belongs to the “Series on Intuitive Mathematical Economics”, whose objective is to present mathematical economics in a more intuitive way than the one usually encountered in the standard textbooks on the subject. In this case we cover the concepts of differentiation of composite functions (or the chain rule), and differentiation of functions defined implicitly.

The most important notion of single-variable calculus is that you can approximate any sufficiently smooth function  $f(x)$  near a value  $x = a$  by a straight line.

$$f(x) \approx \alpha x + \beta \tag{1}$$

where  $\alpha = df(x)/dx|_{x=a}$  is the derivative of  $f(x)$  evaluated at  $x = a$ , see Fig. 1.

It cannot be overestimated the importance of this simple equation, that also extends to many dimensions, in the sense that any smooth vector function  $\mathbf{f}(x_1, x_2, \dots, x_n)$  of  $n$  variables can be approximated near the point  $(a_1, a_2, \dots, a_n)$  by the linear function

$$\mathbf{f}(\mathbf{x}) \approx \mathbf{A}\mathbf{x} + \mathbf{b} \tag{2}$$

---

<sup>1</sup>sp@ucema.edu.ar

The points of view of the author do not necessarily represent the position of Universidad del CEMA.

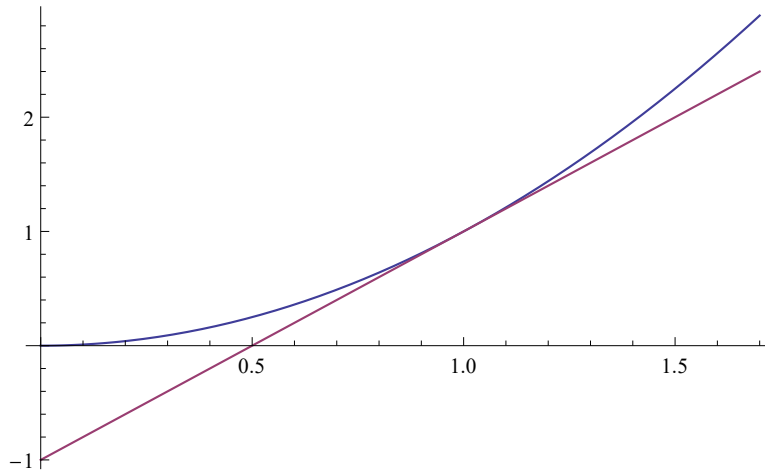


Figure 1: The function  $f(x) = x^2$  (blue) is approximated by the straight line  $y = 2x - 1$  (purple) at  $x = 1$ . The straight line is tangent at that point.

where the small, bold letters, now represent vectors, and  $A$  is a matrix. If you do not fully grasp equation (2) yet, don't worry, the extension to many variables will be covered in different documents. But in this paper, equation (1) is assumed to be known.

Another way of writing the straight line (1) is this:

$$f(x) = f(a + \Delta x) \approx f(a) + f'(a)\Delta x \quad (3)$$

where  $\Delta x = x - a$  is assumed small. Comparing (3) with (1), it is clear that  $\beta = f(a) - f'(a)a$ . So, the equation of the straight line tangent to  $f(x)$  at  $x = a$  is:

$$y = f'(a)x + f(a) - f'(a)a \quad (4)$$

where it is explicit that at  $x = a$ ,  $y = f(a)$  (see problem 1; 1).

Yet another way of writing (1) or (3) is:

$$\Delta f \equiv f(a + \Delta x) - f(a) \approx f'(a)\Delta x \quad (5)$$

Although the notation  $\Delta x$  ( $\Delta f$ ) is normally used for finite differences, and  $dx$  ( $df$ ) for differentials, we will not distinguish in this work between these notations<sup>2</sup>.

<sup>2</sup>Another way of writing all this is in vector notation, whose importance for an intuitive approach to Mathematical Economics cannot be overemphasized and will be addressed in a different work, the meaning of the linear approximation presented in (1), (3) and (5), is that the vector

$$\vec{v} = \hat{x} + f'(a)\hat{f} \quad (6)$$

is tangent to the function  $f(x)$  at  $x = a$ , where  $\hat{x}$  and  $\hat{f}$  are, respectively, the unit vectors in the horizontal ( $x$ ) and vertical ( $f$ ) directions in the  $(x, f)$  plane.

**HW:** "digest" expression (6). As vectors, what is the relation between vector  $\vec{v}$  and  $\vec{r} = \Delta x \vec{v} = f'(a)\Delta x \hat{f} + \Delta x \hat{x}$ , for some value of  $\Delta x$ ?

**HW:** Prove that, in vector notation, the straight line (6) is:  $\vec{w}(\Delta x) = \Delta x \vec{v} + f(a)\hat{f} + a\hat{x}$ , for arbitrary  $\Delta x$ .

In the rest of the paper we study the form these formulas take for, composite functions in section 2, and for functions defined implicitly in section 3.

## 2 Differentiating composite functions: the chain rule

If the function  $h(x)$  can be written in the composite form  $h(x) = f(g(x))$ , the “chain rule” states that its derivative can be computed as

$$\left. \frac{dh}{dx} \right|_a = \left. \frac{df}{dg} \right|_{g(a)} \left. \frac{dg}{dx} \right|_a \quad (7)$$

In words, if  $h(x) = f(g(x))$ , the derivative of  $h$  with respect to  $x$ , evaluated at  $x = a$ , equals the derivative of  $f$  with respect to  $g$ , evaluated at  $g(a)$ , times the derivative of  $g$  with respect to  $x$ , evaluated at  $x = a$ .

For example, if  $h(x) = x^n$ ,  $dh/dx = nx^{n-1}$ , but we can write  $x^n = (x^r)^{n/r}$ , so, if  $g(x) = x^r$ , and  $f(g) = g^{n/r}$ , we have  $h(x) = f(g(x))$ , so let us see the chain rule in action:

$$\begin{aligned} \left. \frac{dh}{dx} \right|_a &= na^{n-1} & (8) \\ &= \left. \frac{df}{dg} \right|_{g(a)} \left. \frac{dg}{dx} \right|_a \\ &= \left. \frac{dg^{\frac{n}{r}}}{dg} \right|_{g(a)} \left. \frac{dx^r}{dx} \right|_a \\ &= \left. \frac{n}{r} g^{\frac{n}{r}-1} \right|_{g(a)} \left. rx^{r-1} \right|_a \\ &= n (a^r)^{\frac{n-r}{r}} a^{r-1} \\ &= n a^{n-r} a^{r-1} \\ &= n a^{n-1} \end{aligned}$$

in this derivation we used  $g(a) = a^r$ , according to the definition of  $g(x) = x^r$ . Comparing the first and last line we see that, at least in this case, the chain rule works. The objective of this section is to try to provide the intuition behind the general validity of this rule for any sufficiently smooth function.

What is the meaning of  $h(a + \Delta x)$  when  $h$  is a function of the form  $h(x) = f(g(x))$ ? It means  $h(a + \Delta x) = f(g(a + \Delta x))$ .

So, on the one hand,

$$h(a + \Delta x) \approx h(a) + \left. \frac{dh}{dx} \right|_a \Delta x \quad (9)$$

$$= f(g(a)) + \left. \frac{df}{dg} \right|_{g(a)} \Delta x \quad (10)$$

where (9) applies for any smooth function and in (10) we used  $h(a) = f(g(a))$ .

On the other

$$h(a + \Delta x) = f(g(a + \Delta x)) \quad (11)$$

$$\approx f\left(g(a) + \left.\frac{dg}{dx}\right|_a \Delta x\right) \quad (12)$$

$$\approx f(g(a)) + \left.\frac{df}{dg}\right|_{g(a)} \left.\frac{dg}{dx}\right|_a \Delta x \quad (13)$$

Going from (11) to (12) we used the linear approximation (3) for the function  $g$  (assumed sufficiently smooth), and from (12) to (13) we used the same linear approximation (3) for  $f$ , evaluated at the point  $g(a) + \left.\frac{dg}{dx}\right|_a \Delta x$ , which is assumed sufficiently close to the point  $g(a)$  for the linear approximation to apply. This will happen if  $\left.\frac{dg}{dx}\right|_a$  is finite and  $\Delta x$  is itself “sufficiently small”.

So, comparing (10) and (13), we finally arrive at:

$$\left.\frac{dh}{dx}\right|_a = \left.\frac{df}{dg}\right|_{g(a)} \left.\frac{dg}{dx}\right|_a \quad (14)$$

which is the chain rule we wanted to prove.

One point that confuses many students is that in going from (11) to (13) we make first an approximation and then another. Will the composition of the two approximations still allow the linear approximation to work? How small  $\Delta x$  has to be to ensure that  $\left.\frac{dg}{dx}\right|_a \Delta x$  is “sufficiently small”?

Without given a rigorous proof, since the emphasis here is on intuition and a rigorous proof can be found in many excellent textbooks, we will help to construct the intuition of why, if the linear approximation works for  $f(g)$  and for  $g(x)$ , then it should work for  $f(g(x))$ .

The idea is very simple, the fact that the linear approximation works for  $f$  at the value of the independent variable  $g(a)$  means that for sufficiently small values of  $\Delta g$ ,  $f(g(a) + \Delta g) \approx f(g(a)) + \left.\frac{df}{dg}\right|_{g(a)} \Delta g$ . Looking at (12) the question then is, can we make  $\Delta g \equiv \left.\frac{dg}{dx}\right|_a \Delta x$  sufficiently small? The answer is yes, as you can show in problem 2: 1. This is the intuition behind the chain rule.

## 2.1 Extensions of the chain rule

There is an extension of the chain rule, that strictly speaking belongs to multiple variable calculus, but we will present here and “justify” it by simply showing how it works in the example (8).

Suppose that the function  $h$  can be written in the form  $h(x) = f(g(x), x)$ . Note that  $f$  is now a *two* variable function  $f(g, x)$ . This means that part of the dependence on the variable  $x$  is through the

variable  $g$ , that itself is a function of  $x$ , and part is the explicit dependence of  $f$  on  $x$ . Then the derivative of  $h$  with respect to  $x$  can be computed as

$$\left. \frac{dh}{dx} \right|_a = \left. \frac{\partial f}{\partial g} \right|_{g=g(a), x=a} \left. \frac{dg}{dx} \right|_{x=a} + \left. \frac{\partial f}{\partial x} \right|_{g=g(a), x=a} \quad (15)$$

in words, the derivative of  $h(x)$  with respect to  $x$ , evaluated at  $x = a$ , where  $h$  can be written as a function of two variables  $f(g, x)$ , and  $g$  itself is a function of  $x$ , equals to the sum of two terms. The first term is the *partial* derivative of  $f(g, x)$  with respect to  $g$ , evaluated at  $g = g(a)$  and  $x = a$ , times the derivative of  $g$  with respect to  $x$ , evaluated at  $x = a$ . And the second term is the *partial* derivative of  $f(g, x)$  with respect to  $x$ , evaluated at  $g = g(a)$  and  $x = a$ . This is an extension of the chain rule (7).

To see this in action, consider again the example (8),  $h(x) = x^n$ , but now written as

$$h(x) = x^n = x^{n-2} x^2 = (x^r)^{(n-2)/r} x^2 = g^{(n-2)/r} x^2, \quad \text{where } g = g(x) = x^r \quad (16)$$

so, the function  $f$ , of *two variables*, is  $f(g, x) = g^{(n-2)/r} x^2$ .

Now, on the one hand, as in (8),  $dh/dx|_{x=a} = na^{n-1}$ . On the other, applying (15),

$$\begin{aligned} \left. \frac{dh}{dx} \right|_a &= \left. \frac{n-2}{r} g^{(n-2-r)/r} x^2 \right|_{g=a^r, x=a} \left. r x^{r-1} \right|_{x=a} + \left. g^{(n-2)/r} 2x \right|_{g=a^r, x=a} \quad (17) \\ &= (n-2) (a^r)^{(n-2-r)/r} a^2 a^{r-1} + (a^r)^{(n-2)/r} 2a \\ &= (n-2) a^{n-1} + 2a^{n-1} \\ &= na^{n-1} \end{aligned}$$

This shows that (15) works for our particular case. We leave the general proof, that involves multi-variable calculus, for another work.

For an even greater extension of the chain rule, suppose that the function  $h$  can be written in the form

$$h(x) = f(g_1(x), g_2(x), g_3(x), \dots, g_k(x), x) \quad (18)$$

Now  $f$  is a function of  $k+1$  variables,  $f(g_1, g_2, g_3, \dots, g_k, x)$ . This means that part of the dependence on the variable  $x$  is through the variables  $g_i$ ,  $i = 1, \dots, k$ , that themselves are functions of  $x$ , and part is the explicit dependence of  $f$  on  $x$ . Then the derivative of  $h$  with respect to  $x$  can be computed as

$$\left. \frac{dh}{dx} \right|_a = \sum_{i=1}^k \left. \frac{\partial f}{\partial g_i} \right|_{g_1=g_1(a), \dots, g_k=g_k(a), x=a} \left. \frac{dg_i}{dx} \right|_{x=a} + \left. \frac{\partial f}{\partial x} \right|_{g_1=g_1(a), \dots, g_k=g_k(a), x=a} \quad (19)$$

$$= \sum_{i=1}^k \frac{\partial f}{\partial g_i} \frac{dg_i}{dx} + \frac{\partial f}{\partial x} \quad (20)$$

(19) is the extended chain rule, and (20) is meant to be interpreted as in (19), but in a shorter notation where the points of evaluation of the derivatives are assumed understood (see problem 2.1: I).

### 3 Differentiation of functions defined implicitly

Consider in Fig. 2 the circle given by the equation

$$x^2 + g^2 = 1 \tag{21}$$

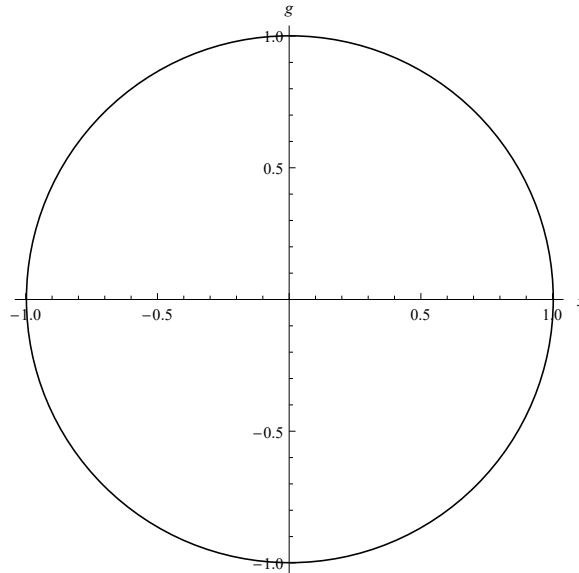


Figure 2: Circle  $x^2 + g^2 = 1$ .

In the plane  $(x, g)$ , (21) implicitly defines two functions:  $g_+(x) = \sqrt{1 - x^2}$  (the upper half circle) and  $g_-(x) = -\sqrt{1 - x^2}$  (the lower half circle). Note that these two functions can be easily found from their defining equation (21). Consider for example  $g_+(x)$ , omitting for simplicity the subindex “+”:  $g(x) = (1 - x^2)^{1/2}$ . Its derivative with respect to  $x$  is

$$\frac{dg}{dx} = -\frac{x}{\sqrt{1 - x^2}} \tag{22}$$

There is another, more general way, to view the same problem, that works even in cases in which one can not disentangle from the defining equation the explicit function  $g(x)$ . Consider in Fig. 3 the two variable function  $f(x, g) = x^2 + g^2$ . A “level curve” of a function of two variables  $f(x, g)$ , is the set of points defined by the equation  $f(x, g) = c$ . In Fig. 3, one can see the level curves corresponding to  $c = 0.3, 1$  and  $2$ .

Each of these level curves defines implicitly a function  $g(x)$  (see the projection of the level curves in the  $(x, g)$  plane in Fig. 3). So, for each level curve, we have a situation similar to the one in section 2.1: there must be a function  $h(x)$  whose  $x$  dependence can be written as  $h(x) = f(x, g)$ ,



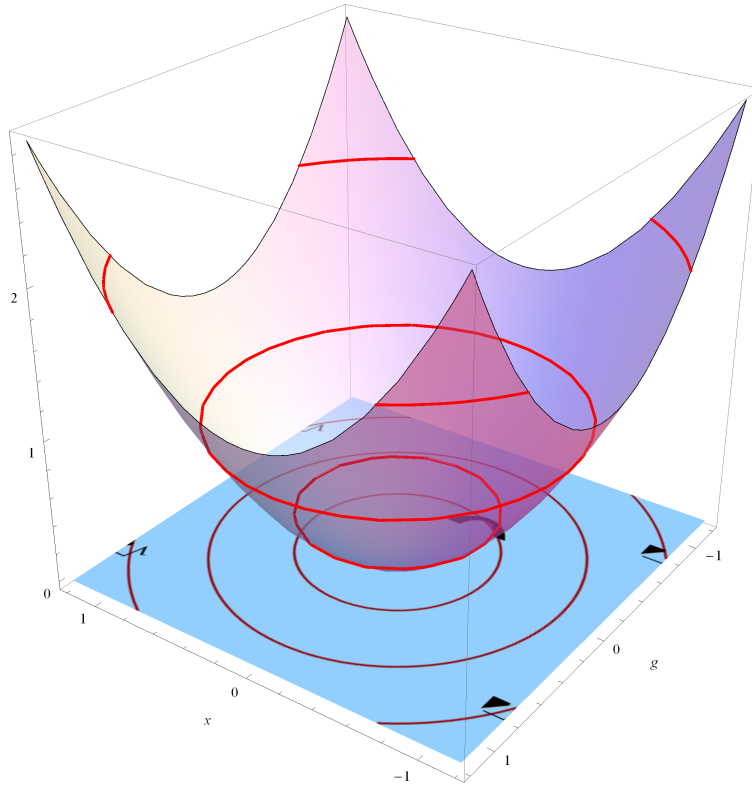


Figure 3: Two variable function  $f(x, g) = x^2 + g^2$  and its level curves  $f = 0.3, 1, 2$ .

with  $g$  itself a function of  $x$ ,  $g(x)$ . But equation  $f(x, g) = c$  tells us that this function is a constant  $c$ . So, on the one hand we have

$$\frac{dh}{dx} = 0 \quad (23)$$

and, on the other, applying the extended chain rule (15), we have

$$\frac{dh}{dx} = \frac{\partial f}{\partial g} \frac{dg}{dx} + \frac{\partial f}{\partial x} \quad (24)$$

from (23) and (24),  $\frac{\partial f}{\partial g} \frac{dg}{dx} + \frac{\partial f}{\partial x} = 0$ , from where we derive

$$\frac{dg}{dx} = -\frac{\frac{\partial f}{\partial x}}{\frac{\partial f}{\partial g}} \quad (25)$$

This is the general form of the derivative of a function given implicitly in the form  $f(x, g) = c$ , i.e., a level line of  $f$ . Note that this formula does not require knowing explicitly  $g(x)$ , as was the case in (22).

Let us return to our circle,  $f(x, g) = x^2 + g^2$ .  $\partial f/\partial x = 2x$  and  $\partial f/\partial g = 2g$ , so (25) implies

$$\frac{dg}{dx} = -\frac{x}{g} \quad (26)$$

but since in this case we know explicitly that  $g(x) = \sqrt{1 - x^2}$ ,

$$\frac{dg}{dx} = -\frac{x}{\sqrt{1 - x^2}} \quad (27)$$

as in (22).

—x—

There is yet another, slightly different perspective, to view the same problem. Consider the function  $f(x, g)$ , whose level curve  $f(x, g) = c$  implicitly defines  $g(x)$ . The generalization of (5) to two variables is

$$\Delta f \equiv f(a + \Delta x, b + \Delta g) - f(a, b) \approx \left. \frac{\partial f}{\partial x} \right|_{x=a, g=b} \Delta x + \left. \frac{\partial f}{\partial g} \right|_{x=a, g=b} \Delta g \quad (28)$$

where  $\Delta x = x - a$  and  $\Delta g = g - b$  are supposed to be small. The same relation can be written in the form (3),

$$f(x, g) = f(a + \Delta x, b + \Delta g) \approx f(a, b) + \left. \frac{\partial f}{\partial x} \right|_{x=a, g=b} \Delta x + \left. \frac{\partial f}{\partial g} \right|_{x=a, g=b} \Delta g \quad (29)$$

again, valid for  $(x, g)$  sufficiently close to  $(a, b)$ .

In the 3-D space  $(x, g, z)$ , the plane

$$z = f(a, b) + \left. \frac{\partial f}{\partial x} \right|_{x=a, g=b} (x - a) + \left. \frac{\partial f}{\partial g} \right|_{x=a, g=b} (g - b) \quad (30)$$

is tangent to the surface  $z = f(x, g)$  at  $x = a$  and  $g = b$ . And equation (29) says that for points in the  $(x, g, z)$  space with  $x$  and  $g$  close enough to  $a$  and  $b$ , the function is very well approximated by the tangent plane, see Fig. 4.

Now, let us refocus on the level curve  $f(x, g) = c$  of a function. In equations (28-29),  $\Delta x$  and  $\Delta g$  vary independently, and, as Fig. 4 shows for the particular case  $f = x^2 + g^2$ , as they independently vary, these equations give us an approximation to the change in  $f$ . But if we want  $f(x, g)$  to be fixed at the value  $c$ , as Fig. 3 shows, we are forced to move in the  $(x, g)$  plane only in those points that make  $f(x, g) = c$ . We need  $\Delta f = 0$ , and equation (28) therefore indicates that  $\Delta x$  and  $\Delta g$  cannot vary independently. To stay in the level curves, small changes  $\Delta g$  of the variable  $g$ , have to depend on small changes  $\Delta x$  of the variable  $x$ , so that

$$\frac{\partial f}{\partial x} \Delta x + \frac{\partial f}{\partial g} \Delta g = 0 \quad (31)$$

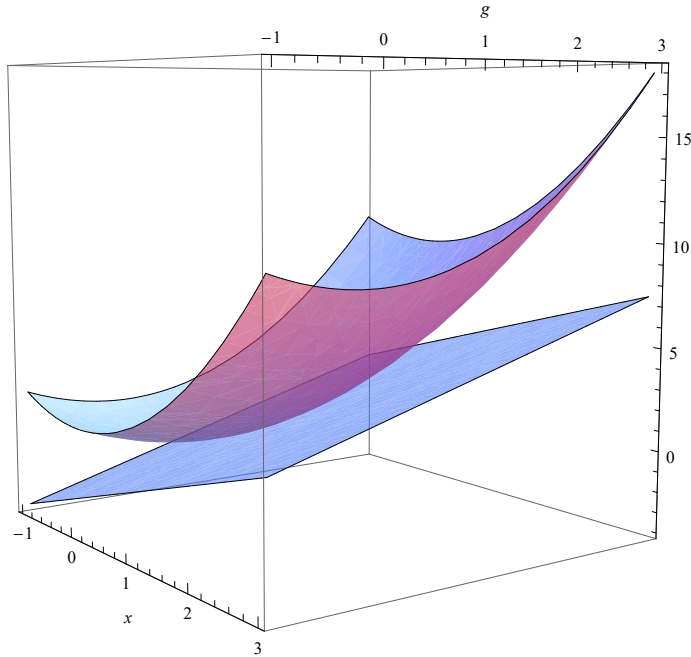


Figure 4: The two variable function  $f(x, g) = x^2 + g^2$  and its tangent plane at  $x = g = \frac{1}{\sqrt{2}}$ .

In the  $(x, g)$  plane this implies that the implicit function  $g(x)$  defined by  $f(x, g) = c$  is such that small variations in  $x$  and small variations in  $g$  should be related so that

$$\frac{\Delta g}{\Delta x} = -\frac{\frac{\partial f}{\partial x}}{\frac{\partial f}{\partial g}} \quad (32)$$

But this, in the limit, is precisely the definition of the derivative of  $g$  with respect to  $x$ , as we had already seen in (25).

## 4 Conclusions

We have provided various intuitive perspectives for understanding the chain rule and the derivative of functions given implicitly by level curves of a function  $f(x, g)$ . In a different work we will show economic applications of these technics.

## Problems:

### Section 1:

1. Prove (4).

### Section 2:

1. Show, first graphically and then algebraically, that if one wants  $|g'(a)\Delta x|$  to be smaller than a given number  $\epsilon$  so as to make sure that the linear approximation of  $f$  applies in (12), then, if  $g'(a)$  is finite, one can find a number  $\delta$  such that if  $|\Delta x| < \delta$ , then  $|g'(a)\Delta x| < \epsilon$ .

### Section 2.1:

1. Consider the function  $h(x) = x^n$ , written as

$$\begin{aligned}h(x) &= x^n = x^r x^s x^2 = g_1 g_2 x^2 = f(g_1, g_2, x) \\f(g_1, g_2, x) &= g_1 g_2 x^2 \\g_1 &= g_1(x) = x^r \\g_2 &= g_2(x) = x^s \\n &= r + s + 2\end{aligned}$$

Prove that the extended chain rule (19-20) works in this example.