Optimization. A first course of mathematics for economists

Xavier Martinez-Giralt

Universitat Autònoma de Barcelona

xavier.martinez.giralt@uab.eu

I.3.- Differentiability

Differentiation and derivative

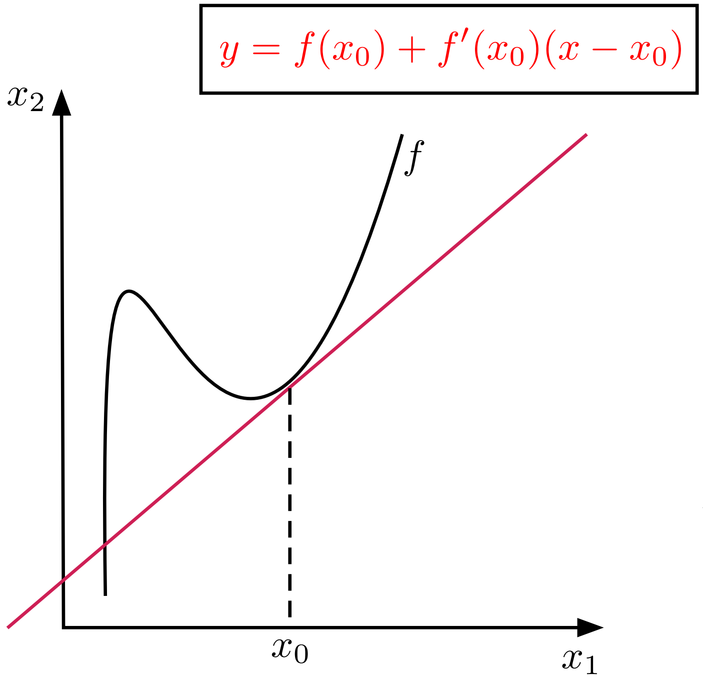
Definitions

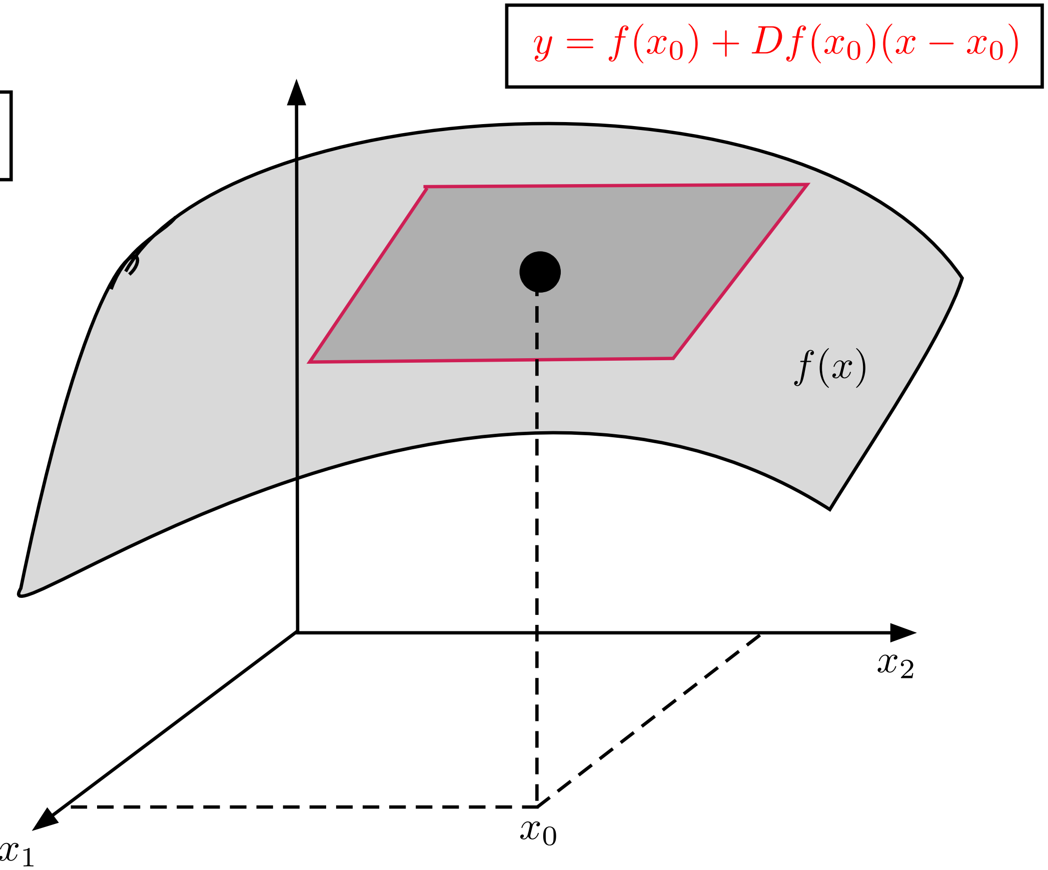
- A function $f : A \subset \mathbf{R} \rightarrow \mathbf{R}$ is differentiable at a point $x_0 \in A$ if the limit
$f'(x_0) = \lim_{h\rightarrow 0} \frac{f(x_o+h)-f(x_0)}{h}$ exists. Equivalently,
$\lim_{h\rightarrow 0} \frac{f(x_0+h)-f(x_0)-f'(x_0)h}{h} = 0$ or
$\lim_{x\rightarrow x_0} \frac{|f(x)-f(x_0)-f'(x_0)(x-x_0)|}{|x-x_0|} = 0.$
- A function $f : A \subset \mathbf{R}^n \rightarrow \mathbf{R}^m$ is differentiable at a point $x_0 \in A$ if we can find a linear function $Df(x_0) : \mathbf{R}^n \rightarrow \mathbf{R}^m$ (that we refer to as the derivative of f at x_0) such that

$$\lim_{x\rightarrow x_0} \frac{\|f(x) - [f(x_0) + Df(x_0)(x - x_0)]\|}{\|x - x_0\|} = 0$$

- If f is differentiable $\forall x \in A$, we say that f is differentiable in$A$.
- Derivative is the slope of the linear approximation of f at x_0.

Differentiability- Illustration





Differentiation and derivative (2)

Some theorems

- Theorem 1: Let $f : A \to \mathbf{R}^m$ be differentiable at $x_0 \in A$. Assume $A \subset \mathbf{R}^n$ is an open set. Then, there is a unique linear approximation $Df(x_0)$ to f at x_0.

Recall some one-dimensional results

- Theorem 2 (Fermat): Let $f : (a, b) \to \mathbf{R}$ be differentiable at $c \in (a, b)$. If c is an extreme point of f then, $f'(c) = 0$.
- Theorem 3 (Rolle): Let $f : [a, b] \to \mathbf{R}$ be continuous. Assume f is differentiable in (a, b). Assume also $f(a) = f(b) = 0$. Then, $\exists c \in (a, b)$ such that $f'(c) = 0$.
- Theorem 4 (Mean-Value): Let $f : [a, b] \to \mathbf{R}$ be continuous. Assume f is differentiable in (a, b). Then, $\exists c \in (a, b)$ such that $f(b) - f(a) = f'(c)(b - a)$.
- Corollary: If, in addition, $f' = 0$ on (a, b), then f is constant.

Differentiation and derivative (3)

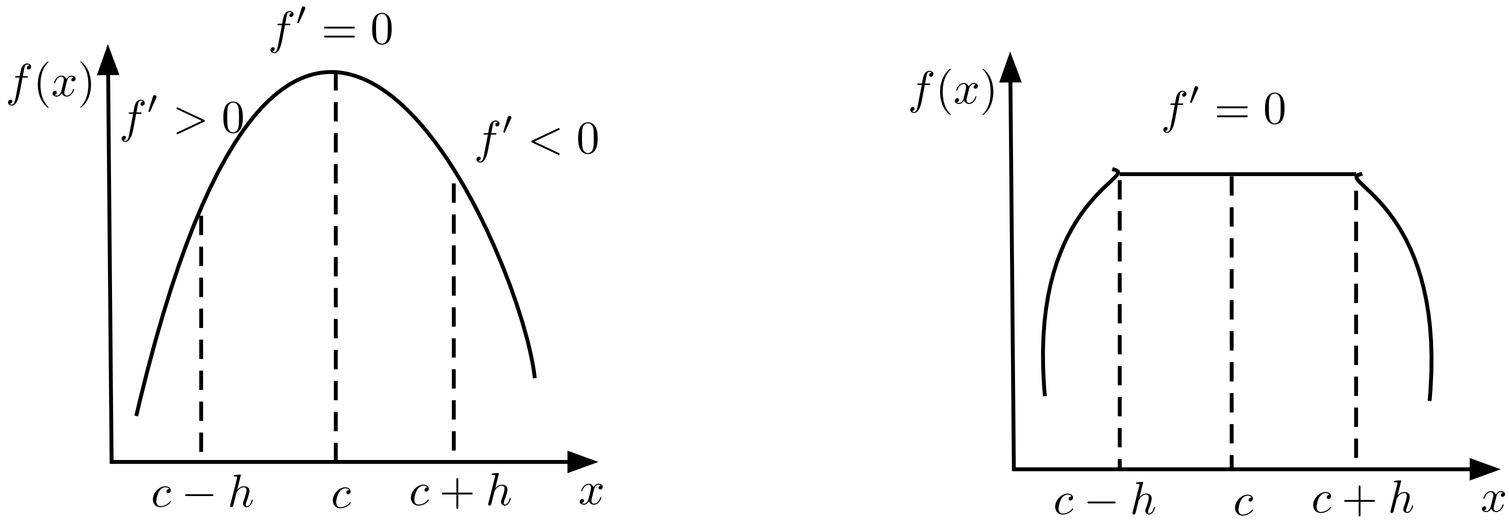
Proof of theorem 2

- Let f have a maximum at c. Then, for $h \geq 0,\ [f(c+h) - f(c)]/h \leq 0$. Letting $h \to 0, h \geq 0$ we get $f'(c) \leq 0$. Similarly, for $h \leq 0$, it follows that $f'(c) \geq 0$. Hence, $f'(c) = 0$.
- A parallel argument holds when f has a minimum at c.

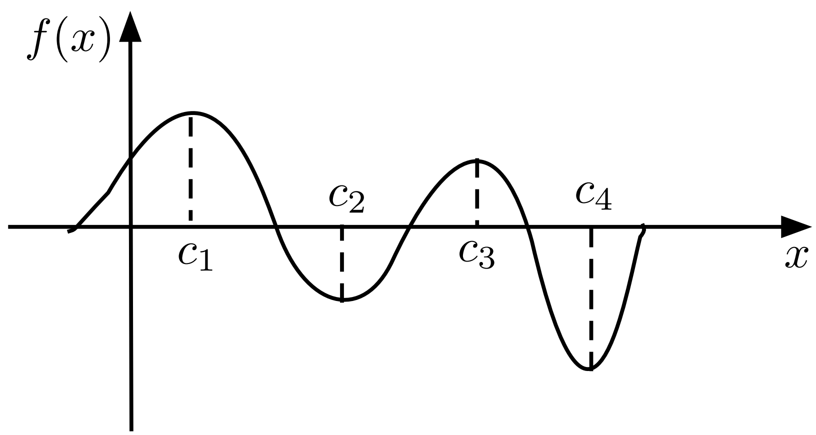
Proof of theorem 3

- If $f(x) = 0, \forall x \in [a, b]$, we can choose any c.
- If $f \neq 0$, applying the boundedness theorem $\exists c_1$ where f reaches a maximum and $\exists c_2$ where f reaches a minimum.
- Since $f(a) = f(b) = 0$, at least one of c_1, c_2 lies in (a, b).
- Assume $c_1 \in (a, b)$. Then, applying theorem 1 $f'(c_1) = 0$. Mutatis mutandis for c_2

One-dimension theorems - Illustration



Theorem 2



Theorem 3

Differentiation and derivative (4)

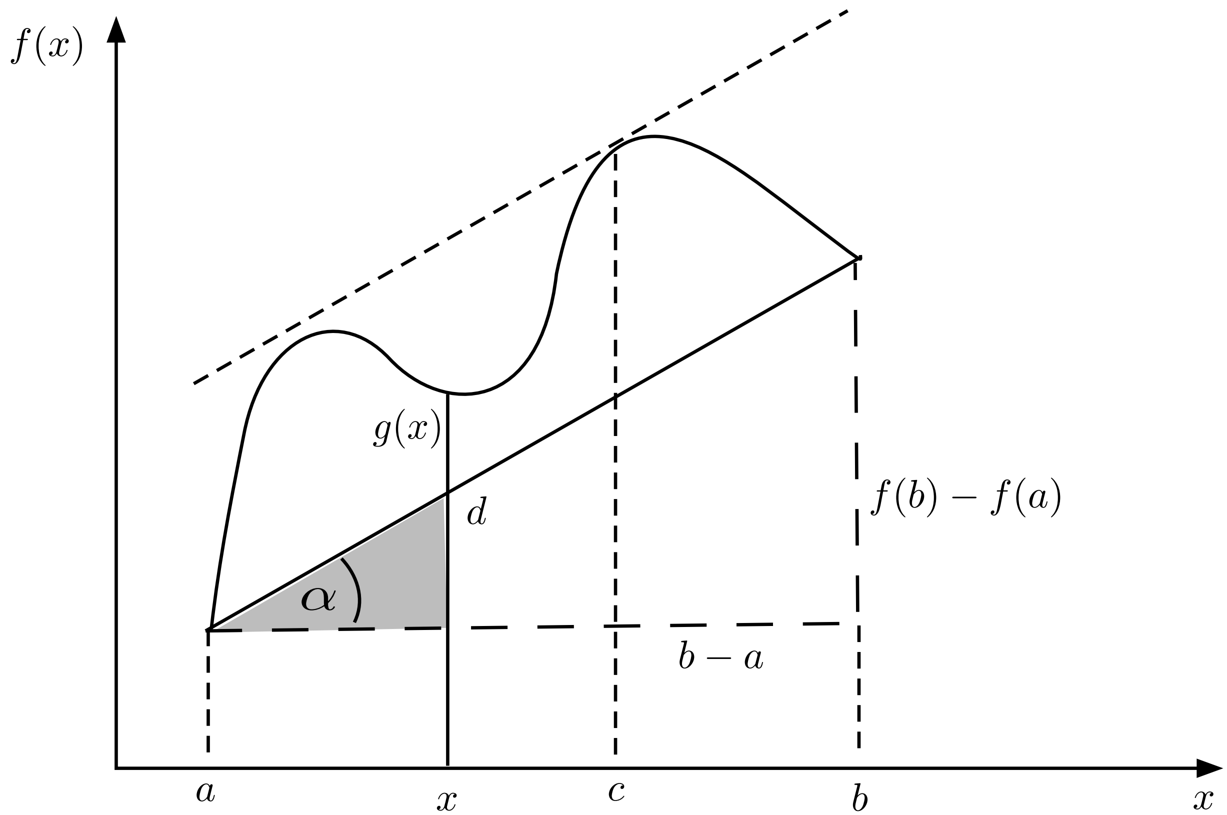
Proof of theorem 4

- Define an auxiliary function $g(x) = f(x) - f(a) - (x-a)\frac{f(b)-f(a)}{b-a}$
- $g(x)$ is continuous in $[a,b]$ and differentiable in (a,b).
- $g'(x) = f'(x) - \frac{f(b)-f(a)}{b-a}$. Also, $g(a) = g(b) = 0$.
- Applying Rolle's theorem, $g'(c) = 0$, Hence, $f(b) - f(a) = f'(c)(b-a)$.

Proof of the Corollary

- Apply Theorem 4 to f on $[a,x]$.
- Then, $f(x) - f(a) = f'(c)(x-a) = 0$.
- Thus, $f(x) = f(a)\ \forall x \in [a,b]$
- Therefore, f is constant.

One-dimension theorems - Illustration



$$tg(\alpha) = \frac{f(b)-f(a)}{b-a} = \frac{d-f(a)}{x-a}$$

$$g(x) = f(x) - [f(a) + tg(\alpha)(x-a)] = f(x) - d$$

Theorem 4

The general Mean-Value theorem

- We say that c is on the line segment joning x and y if it can be written as a convex combination of x and y. Namely, $c = (1-\lambda)x + \lambda y,\ \lambda \geq 0,\ \lambda \in [0,1]$

The general Mean-Value theorem

(i)
- Let $f : A \subset \mathbf{R}^n \to \mathbf{R}$ be differentiable on A (an open set).
- Consider $(x, y) \in A$ s.t. the segment defined by their convex combination lies in A.
- Then $\exists c$ in that segment such that $f(y) - f(x) = Df(c)(y - x)$

(ii)
- Let $f : A \subset \mathbf{R}^n \to \mathbf{R}^m$ be differentiable on A (open set).
- Let $f = (f_1, f_2, \dots, f_m)$. Consider $(x, y) \in A$ s.t. the segment defined by their convex combination lies in A.
- Then $\exists (c_1, c_2, \dots, c_m)$ on that segment such that $f_i(y) - f_i(x) = Df_i(c_i)(y - x),\ i = 1, 2, \dots, n$

The Jacobian matrix

- Let $f : A \subset \mathbf{R}^n \to \mathbf{R}^m$: $f(x_1, \ldots, x_n) = [f_1(x_1, \ldots, x_n), f_2(x_1, \ldots, x_n), \ldots, f_m(x_1, \ldots, x_n)]$
- Compute the partial derivative of f_j wrt x_i, i.e. $\frac{\partial f_j}{\partial x_i}$, $j = 1, \ldots, m;\ i = 1, \ldots, n$
- Partial derivative. Definition $\frac{\partial f_j}{\partial x_i}(x_1, \ldots, x_n) = \lim_{h \to 0} \frac{f_j(x_1, \ldots, x_i + h, \ldots, x_n) - f_j(x_1, \ldots, x_n)}{h}$
- Theorem Let $f : A \to \mathbf{R}^m$ be differentiable. Assume $A \subset \mathbf{R}^n$ is an open set. Then, the partial derivatives $\partial f_j / \partial x_i$ exist and the matrix $Df(x)$ is given by

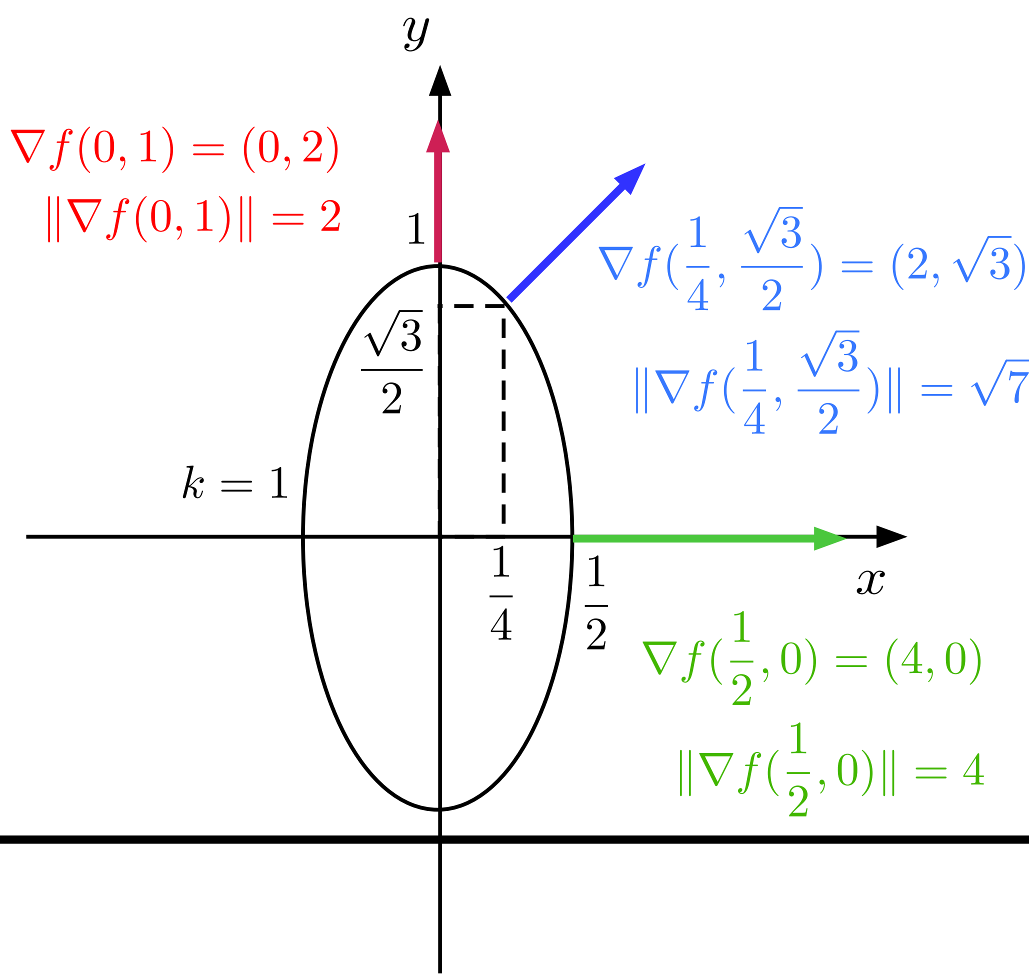
$$\begin{pmatrix} \frac{\partial f_1}{\partial x_1}(x) & \frac{\partial f_1}{\partial x_2}(x) & \ldots & \frac{\partial f_1}{\partial x_n}(x) \\ \frac{\partial f_2}{\partial x_1}(x) & \frac{\partial f_2}{\partial x_2}(x) & \ldots & \frac{\partial f_2}{\partial x_n}(x) \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial f_m}{\partial x_1}(x) & \frac{\partial f_m}{\partial x_2}(x) & \ldots & \frac{\partial f_m}{\partial x_n}(x) \end{pmatrix} \quad \text{(Jacobian matrix of } f\text{)}$$

The gradient of f

- The gradient is the generalization of the concept of derivative of a function in **R** to a function in $\mathbf{R}^m$.
- The gradient is the vector of the n partial derivatives of f.
- Gradient of f. Definition. Let $f : A \subset \mathbf{R}^n \to \mathbf{R}$ be differentiable. The gradient of f is the vector whose components are the elements of $Df(x)$. That is,
- $\nabla f = \left(\frac{\partial f}{\partial x_1}(x), \frac{\partial f}{\partial x_2}(x), \ldots, \frac{\partial f}{\partial x_n}(x) \right)$
- The gradient points towards the direction of greatest rate of increase of the function f. (See below, p. 19)
- Recall: Let $f : A \to \mathbf{R}$. If f is differentiable at x_0 then it is continuous at that point.

Gradient - Illustration

- Consider $f(x,y) = 4x^2 + y^2$. Then, $\nabla f = (8x, 2y)$. The figure represents the gradient at three points for the level set $k = 1$. They are orthogonal to the level surfice as they show the direction of greatest increase of f.



Continuity and differentiability

- Theorem: Let $f : A \to \mathbf{R}^m$ be differentiable in A. Assume $A \subset \mathbf{R}^n$ is an open set. Then, f is continuous.
- Proof:
  - f differentiable at x_0 means $f'(x_0) = \lim_{x \to x_0} \frac{f(x) - f(x_0)}{x - x_0}$
  - We need to prove that $\lim_{x \to x_0} f(x) = f(x_0)$

$$\lim_{x \to x_0} [f(x) - f(x_0)] = \lim_{x \to x_0} (x - x_0) \frac{f(x) - f(x_0)}{x - x_0}$$

$$= \lim_{x \to x_0} (x - x_0) \lim_{x \to x_0} \frac{f(x) - f(x_0)}{x - x_0} = 0 \cdot f'(x_0) = 0$$

  - Therefore, $\lim_{x \to x_0} f(x) = f(x_0)$ proving that f is continuous.

Continuity and differentiability (2)

- Theorem (Lipschitz property): Let $f : A \to \mathbf{R}^m$ be differentiable in A. Assume $A \subset \mathbf{R}^n$ is an open set. Then, f is continuous. More precisely, $\forall x_0 \in A, \exists M > 0$ and $\delta_0 > 0$ such that $\|x - x_0\| < \delta_0$ implies $\|f(x) - f(x_0\| \leq M\|x - x_0\|$.
- The Lipschitz property defines a stronger notion of continuity where, the number M (called the “Lipschitz constant") represents the bound of the slope of the function at x_0. A particular case of Lipschitz continuity is the property of a function being a *contraction*, when $M < 1$ (useful for fixed-point theorems, and stability of equilibria).
- Theorem: Consider $f : A \to \mathbf{R}^m$. Assume $A \subset \mathbf{R}^n$ is an open set. Assume $f = (f_1, f_2, \ldots, f_m)$. If each of the partial derivatives $\partial f_j / \partial x_i$ exists, and is a continuous function in A, then f is differentiable in A.

Directional derivatives

- Intuition: Consider a function defined in a n-dimensional space. The directional derivative is the rate of change of the function f in a particular direction e.
- Definition: Let $f : \mathbf{R}^n \to \mathbf{R}^m$. Assume f is defined in a neighborhood of $x_0 \in \mathbf{R}^n$. Let $e \in \mathbf{R}^n$ be a unit vector. Then, the directional derivative of f at x_0 in the direction e is defined as
$D_u f(x_0) \equiv \frac{d}{dh} f(x_0 + he)|_{h=0} = \lim_{h \to 0} \frac{f(x_0+he)-f(x_0)}{h}$
- This is very similar to the definition of a partial derivative. However, this limit may be difficult to compute. An equivalent formula can be derived using the gradient of f.

Directional derivatives - using the gradient

Introduction

- For illustrative purposes, the argument is developed in $\mathbf{R}^2$, but straightforward generalization
- Consider $f(x, y)$ and a unit vector $e = (e_1, e_2)$.
  Define $g(z) \equiv f(x, y)$ with $x = \tilde{x} + e_1 z$ and $y = \tilde{y} + e_2 z$.

Step 1

- Compute $g'(z) = \lim_{h \to 0} \frac{g(z+h) - g(z)}{h}$
- Evaluate at $z = 0: \;\; g'(0) = \lim_{h \to 0} \frac{g(h) - g(0)}{h}$
- Substitute in $g(\cdot)$ for $f(\cdot)$ to obtain
  $g'(0) = \lim_{h \to 0} \frac{f(\tilde{x}+e_1 h, \tilde{y}+e_2 h) - f(\tilde{x}, \tilde{y})}{h}$
- Note that this limit is precisely the directional derivative of f at $(\tilde{x}, \tilde{y})$, i.e. $g'(0) = D_e f(\tilde{x}, \tilde{y})$

Directional derivatives - using the gradient (2)

Step 2

- Compute $g'(z)$ using the Chain rule:
- $g'(z) = \frac{dg}{dz} = \frac{\partial f}{\partial x}\frac{\partial x}{\partial z} + \frac{\partial f}{\partial y}\frac{\partial y}{\partial z} = \frac{\partial f}{\partial x}e_1 + \frac{\partial f}{\partial y}e_2$, i.e.
- $g'(z) = \frac{\partial f}{\partial x}e_1 + \frac{\partial f}{\partial y}e_2$
- Evaluating at $z = 0$, $g'(0) = \frac{\partial f}{\partial x}(\tilde{x})e_1 + \frac{\partial f}{\partial y}(\tilde{y})e_2$

Step 3

- Combining the two expressions obtained for $g'(0)$ in the two previous steps, it follows that

$$D_e f(\tilde{x}, \tilde{y}) = \frac{\partial f}{\partial x}(\tilde{x})e_1 + \frac{\partial f}{\partial y}(\tilde{y})e_2 = \nabla f \cdot e$$

Directional derivatives - using the gradient (3)

- Let $e = (1, 0, 0, \ldots, 0)$ This is a unit vector in the direction x_1. Accordingly, the directional derivative coincides with the partial derivative $\partial f / \partial x_1$.
- Thus, for a general direction $e = (e_1, \ldots, e_n)$, the directional derivative is a combination of all the partial derivatives with weights $e = (e_1, \ldots, e_n)$ for each of the n directions respectively.
- Operative definition of directional derivative: Consider a function $f : \mathbf{R}^n \to \mathbf{R}$. Let $e = (e_1, \ldots, e_n)$ be a unit vector (i.e. a vector of lenght one). Then, the directional derivative is the dot product of the gradient and the unit vector:
  $D_u f = \nabla f \cdot e = \sum_{i=1}^{n} \frac{\partial f}{\partial x_i} e_i$

Directional derivatives - using the gradient (4)

- As the directional derivative is the dot product of two vectors, it can be written as $D_u f = \nabla f \cdot e = \|\nabla f\| \|e\| \cos\theta$ where θ is tha angle between the gradient vector and the unit vector.
- Note that $D_u f$ is decreasing in $\cos\theta$. That is, the greatest positive value of the directional derivative occurs at $\theta = 0$. Hence, the direction of greatest increase of f is the same direction of the gradient vector.
- Also, the greatest negative value of the directional derivative occurs at $\theta = \pi$. Hence, the direction of greatest decrease of f is the direction opposite to the gradient vector.
- Thus if two vectors a and b are orthogonal (i.e. $\theta = \pi/2$), $\cos\theta = 0$ and thus $a \cdot b = 0$.
- Similarly, two vectors a and b are parallel (i.e. $\theta = \{0, \pi\}$), $\cos\theta = \pm 1$ if $a \cdot b = \|a\| \|b\|$.

Directional derivatives - using the gradient (5)

- Example:
- Let $f(x, y) = 4x^2 + y^2$. Find the directional derivative in the direction $u = (2, 1)$ at the point $(x, y) = (1, 1)$.
- Compute the gradient: $\nabla f = (8x, 2y)$
- Evaluate the gradient at the point (1,1): $\nabla f(1, 1) = (8, 2)$
- Compute the unit vector $e = (e_1, e_2)$:
  - Given the direction $u = (2, 1)$, the length of this vector is $\|u\| = \sqrt{2^2 + 1^2} = \sqrt{5}$.
  - Then $e = (e_1, e_2) = \frac{u}{\|u\|} = (\frac{2}{\sqrt{5}}, \frac{1}{\sqrt{5}})$.
  - so that $\|e\| = \sqrt{(\frac{2}{\sqrt{5}})^2 + (\frac{1}{\sqrt{5}})^2} = 1$
- The directional derivative requested is $\nabla f(1, 1) \cdot (e_1, e_2)^T = (8, 2) \cdot (\frac{2}{\sqrt{5}}, \frac{1}{\sqrt{5}})^T = \frac{18}{\sqrt{5}}$

The Chain rule - Differentiating composite functions

- Many economic applications involve composite functions.
- Directional derivatives is an application of the chain rule.
- Set-up (in $\mathbf{R}^2$)
  - Let $z = f(x_1, x_2), x_1 = g(t), x_2 = h(t)$, be differentiable.
  - Write $z = f(g(t), h(t)) = \phi(t)$ Question: Value of $d\phi/dt$?
  - Answer (theorem): $\frac{d\phi}{dt} = \frac{\partial f}{\partial x_1}\frac{dx_1}{dt} + \frac{\partial f}{\partial x_2}\frac{dx_2}{dt}$
- A more general set-up
  - $z = f(x_1, x_2),\ x_i = g_i(t_1, t_2, t_3),\ (i = 1, 2),\ z = \phi(t_1, t_2, t_3)$
  - Then, $\frac{\partial \phi}{\partial t_j} = \frac{\partial f}{\partial x_1}\frac{\partial x_1}{\partial t_j} + \frac{\partial f}{\partial x_2}\frac{\partial x_2}{\partial t_j},\ (j = 1, 2, 3)$
- General set-up
  - $z = f(x_1, \ldots, x_n),\ x_i = g_i(t_1, \ldots, t_m),\ z = \phi(t_1, \ldots, t_m)$
  - Then, $\frac{\partial \phi}{\partial t_j} = \sum_{i=1}^{n} \frac{\partial f}{\partial x_i}\frac{\partial x_i}{\partial t_j},\ (j = 1, \ldots, m)$

The Chain rule - Proof

- Use definition of derivative: $\frac{d\phi}{dt} = \lim_{\Delta t \to 0} \frac{\phi(t+\Delta t)-\phi(t)}{\Delta t} = \lim_{\Delta t \to 0} \frac{f(g(t+\Delta t),h(t+\Delta t))-f(g(t),h(t))}{\Delta t}$
- Define $\Delta x_1 = g(t+\Delta t) - g(t),\ \Delta x_2 = h(t+\Delta t) - h(t)$
- Substitute:
  $\frac{d\phi}{dt} = \lim_{\Delta t \to 0} \frac{\phi(t+\Delta t)-\phi(t)}{\Delta t} = \lim_{\Delta t \to 0} \frac{f(x_1+\Delta x_1,x_2+\Delta x_2)-f(x_1,x_2)}{\Delta t}$
- Add and substract $f(x_1, x_2 + \Delta x_2)$

$$= \lim_{\Delta t \to 0} \frac{f(x_1+\Delta x_1,x_2+\Delta x_2)-f(x_1,x_2)+f(x_1,x_2+\Delta x_2)-f(x_1,x_2+\Delta x_2)}{\Delta t} =$$

$$= \lim_{\Delta t \to 0} \left(\frac{f(x_1+\Delta x_1,x_2+\Delta x_2)-f(x_1,x_2+\Delta x_2)}{\Delta t} \frac{\Delta x_1}{\Delta x_1} + \frac{f(x_1,x_2+\Delta x_2)-f(x_1,x_2)}{\Delta t} \frac{\Delta x_2}{\Delta x_2} \right)$$

$$= \lim_{\Delta t \to 0} \left(\frac{f(x_1+\Delta x_1,x_2+\Delta x_2)-f(x_1,x_2+\Delta x_2)}{\Delta x_1} \frac{\Delta x_1}{\Delta t} + \frac{f(x_1,x_2+\Delta x_2)-f(x_1,x_2)}{\Delta x_2} \frac{\Delta x_2}{\Delta t} \right) =$$

The Chain rule - Proof (cont'd)

- $= \frac{f(x_1+\Delta x_1, x_2+\Delta x_2) - f(x_1, x_2+\Delta x_2)}{\Delta x_1} \lim_{\Delta t \to 0} \frac{\Delta x_1}{\Delta t} + \frac{f(x_1, x_2+\Delta x_2) - f(x_1, x_2)}{\Delta x_2} \lim_{\Delta t \to 0} \frac{\Delta x_2}{\Delta t} = $
  $= \frac{f(x_1+\Delta x_1, x_2+\Delta x_2) - f(x_1, x_2+\Delta x_2)}{\Delta x_1} \frac{dx_1}{dt} + \frac{f(x_1, x_2+\Delta x_2) - f(x_1, x_2)}{\Delta x_2} \frac{dx_2}{dt}$
- Note that when $\Delta t \to 0$ it follows that $\Delta x_1 \to 0$ and $\Delta x_2 \to 0$
- Note that $\lim_{\Delta x_1 \to 0} \frac{f(x_1+\Delta x_1, x_2+\Delta x_2) - f(x_1, x_2+\Delta x_2)}{\Delta x_1} = \frac{\partial f}{\partial x_1}$
- and $\lim_{\Delta x_2 \to 0} \frac{f(x_1, x_2+\Delta x_2) - f(x_1, x_2)}{\Delta x_2} = \frac{\partial f}{\partial x_2}$
- Hence, we conclude $\frac{d\phi}{dt} = \frac{\partial f}{\partial x_1} \frac{dx_1}{dt} + \frac{\partial f}{\partial x_2} \frac{dx_2}{dt}$.

The Chain rule and directional derivatives

- Consider $f(x, y)$ and a point (x_0, y_0) in the domain of f.
- Consider any vector $(h, k) \neq 0$. It gives a direction to move away from (x_0, y_0) in a straight line towards points $(x, y) = (x(t), y(t)) = (x_0 + th, y_0 + tk)$
- Given (x_0, y_0) and (h, k), define the directional function $g(t) = f(x_0 + th, y_0 + tk)$.
- Question dg/dt?
- Apply Chain-rule: $\frac{dg}{dt} = \frac{\partial f}{\partial x}\frac{dx}{dt} + \frac{\partial f}{\partial y}\frac{dy}{dt} = \frac{\partial f}{\partial x}h + \frac{\partial f}{\partial y}k$.
- Let $t = 0$. Then, $\frac{dg}{dt}|_{t=0} = \frac{\partial f}{\partial x}|_{(x_0,y_0)}h + \frac{\partial f}{\partial y}|_{(x_0,y_0)}k = \nabla f \cdot (h, k)$.
- When (h, k) is the unit vector (i.e. $h^2 + k^2 = 1$), the derivative of f in the direction (h, k) is the *directional derivative* of f at (x_0, y_0).

The Implicit function theorem

Motivation

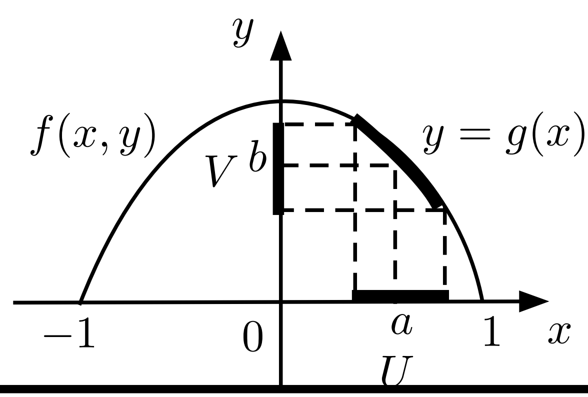
- Consider $f(x,y) = x^2 + y^2 - 1$.
- 1.- Can we find a function $y = g(x)$ for (x,y) s.t. $f(x,y) = 0$?
- i.e., can we write $f(x,g(x)) = 0$ for all x in the domain of g?
- 2.- how changes in x affect y?

Some examples

- Example 1
  - Let $f(x,y) = ay - bx - c$
  - values that satisfy $f(x,y) = 0$ are $ay - bx - c = 0$
  - **Suppose** $a \neq 0$
  - Then, $y(x) = (b/a)x + c/a$
  - $y(x)$ contiunuous $\forall x$; $y(x)$ differentiable, $dy/dx = b/a$
  - Note $\frac{\partial f}{\partial y} = a$. Hence, $f(x)$ exists and differentiable iff $a \neq 0$

The Implicit function theorem - example 2

- Consider $f(x, y) = x^2 + y^2 - 1$.
- Let $x \in [-1, 1]$ and $y \geq 0$
- Consider points (a, b) such that $f(x, y) = 0$
- $Jf = \left(\frac{\partial f}{\partial x}(a, b), \frac{\partial f}{\partial y}(a, b)\right) = (2a, 2b)$
- $\frac{\partial f}{\partial y}(a, b) = 2b \neq 0$ if $b \neq 0$.
- Then $y = g(x) = \sqrt{1 - x^2}$ and
  $f(x, g(x)) = x^2 + (\sqrt{1 - x^2})^2 - 1 = x^2 + 1 - x^2 - 1 = 0$



The Implicit function theorem - example 3

- Consider the functions

$$f_1 : \mathbf{R}^2 \times \mathbf{R}^2 \to \mathbf{R} : (x, y, z, w) \to x^2 + y^2 + z^2 + w^2 - 2$$
$$f_2 : \mathbf{R}^2 \times \mathbf{R}^2 \to \mathbf{R} : (x, y, z, w) \to x^2 - y^2 + z^2 - w^2$$

- Suppose $\exists (x_0, y_0, z_0, w_0)$ with $z_0 > 0, w_0 > 0$ satisfying $f_1(x_0, y_0, z_0, w_0) = 0, f_2(x_0, y_0, z_0, w_0) = 0$
- Note that

$$\Delta = \left| \begin{matrix} \frac{\partial f_1}{\partial z} & \frac{\partial f_1}{\partial w} \\ \frac{\partial f_2}{\partial z} & \frac{\partial f_2}{\partial w} \end{matrix} \right|_{(z_0, w_0)} = \left| \begin{matrix} 2z_0 & 2w_0 \\ 2z_0 & -2w_0 \end{matrix} \right| = -8z_0 w_0 \neq 0$$

- Then, it is easy to verify that the functions $z = g_1(x, y) = \sqrt{1 - x^2}$ and $w = g_2(x, y) = \sqrt{1 - y^2}$ satisfy $f_1(x, y, g_1(x, y), g_2(x, y)) = 0$ and $f_2(x, y, g_1(x, y), g_2(x, y)) = 0$

The Implicit function theorem (2)

The general question

- Consider a function $f : \mathbf{R}^n \times \mathbf{R}^m \to \mathbf{R}^m$. Consider $f(x, y) = 0$:

$$
\begin{aligned}
f_1(x_1, \ldots, x_n; y_1, \ldots, y_m) &= 0 \\
\vdots \qquad\qquad & \vdots \\
f_m(x_1, \ldots, x_n; y_1, \ldots, y_m) &= 0
\end{aligned}
$$

- We aim at solving for the m unknowns $(y_1, \ldots, y_m)$ from the m equations in terms of $(x_1, \ldots, x_n)$.

The Implicit function theorem (3)

- Let $i = 1, 2, \ldots, m$
- Suppose $f_i : \mathbf{R}^n \times \mathbf{R}^m \to \mathbf{R}$ has continuous partial derivatives.
- Consider $(x_0, y_0) \in \mathbf{R}^n \times \mathbf{R}^m$ with $f_i(x_0, y_0) = 0, \forall i$.
- Assume the determinant Δ evaluated at (x_0, y_0) is not zero.

$$\Delta = \left| \begin{matrix} \frac{\partial f_1}{\partial y_1} & \cdots & \frac{\partial f_1}{\partial y_m} \\ \vdots & \vdots & \vdots \\ \frac{\partial f_m}{\partial y_1} & \cdots & \frac{\partial f_m}{\partial y_m} \end{matrix} \right|_{(x_0, y_0)} \neq 0$$

- Then, $\exists U = B(x_0, r) \subset \mathbf{R}^n$ and $V = B(y_0, s) \subset \mathbf{R}^m$ and a unique functions $g_i : U \to V,\ \forall x \in U,\ \forall y \in V$ such that $f_i(x, g_1(x), \ldots, g_m(x)) = 0, \forall i$.
- This is an essential result for the comparative statics analysis

The Implicit function theorem - A particular case

- Suppose $f : \mathbf{R}^n \times \mathbf{R} \to \mathbf{R}$ has continuous partial derivatives.
- Suppose $\exists (x_0, y_0) \in \mathbf{R}^n \times \mathbf{R}$ s.t. $f(x_0, y_0) = 0,\ \frac{\partial f}{\partial y}|_{(x_0,y_0)} \neq 0$
- Then, $\exists U = B(x_0, r) \subset \mathbf{R}^n$ and $V = B(y_0, s) \subset \mathbf{R}$ such that there is a unique function
  $y = g(x) = g(x_1, \ldots, x_n)$
  defined for $x \in U$ and $y \in V$, satisfying
  $f(x, g(x)) = 0$

Proof for $n = 2$

The Implicit function theorem - Proof $(n = 2, m = 1)$

- Notation: $(\mathbf{x}, z) = (x, y, z)$, $(\mathbf{x}_0, z_0) = (x_0, y_0, z_0)$
- Let $f : \mathbf{R}^2 \times \mathbf{R} \to \mathbf{R}$ with $f(\mathbf{x}_0, z_0) = 0$ and $\frac{\partial f}{\partial z}|_{(\mathbf{x}_0, z_0)} \neq 0$
- Suppose (wlog) $\frac{\partial f}{\partial z}|_{(\mathbf{x}_0, z_0)} > 0$ (otherwise, consider $-f$)
- Because $\frac{\partial f}{\partial z}$ is continuously differentiable, $\exists a > 0$ and $b > 0$ such that for $\|\mathbf{x} - \mathbf{x_0}\| < a$ and $|z - z_0| < a$, $\frac{\partial f}{\partial z} > b$.
- Also, we may assume $\exists M > 0$ such that $|\frac{\partial f}{\partial x}| < M$ and $|\frac{\partial f}{\partial y}| < M$ in the same region.
- Since $f(\mathbf{x}_0, z_0) = 0$, we can rewrite it as

$$f(\mathbf{x}, z) = [f(\mathbf{x}, z) - f(\mathbf{x}_0, z)] + [f(\mathbf{x}_0, z) - f(\mathbf{x}_0, z_0)]$$

The IFT - Proof (2)

- Consider the term $[f(\mathbf{x}, z) - f(\mathbf{x}_0, z)]$
- The line segment in $\mathbf{R}^3$ linking $(\mathbf{x}, z)$ to $(\mathbf{x}_0, z)$ is:

$$
\begin{aligned}
L : [0,1] \to \mathbf{R}^3 : t \to (t\mathbf{x} + (1-t)\mathbf{x}_0, z) \\
= (tx + (1-t)x_0, ty + (1-t)y_0, z)
\end{aligned}
$$

- Next, define $h = f \circ L : [0,1] \to \mathbf{R}$. Then, for some $\theta \in (0,1)$, applying the Mean value theorem it follows

$$f(\mathbf{x}, z) - f(\mathbf{x}_0, z) = h(1) - h(0) = h'(\theta)$$

- Applying the chain rule to compute $h'(\theta)$:

The IFT - Proof (3)

$$h'(\theta) = \begin{pmatrix} \frac{\partial f}{\partial x}|_{L(\theta)} & \frac{\partial f}{\partial y}|_{L(\theta)} & \frac{\partial f}{\partial z}|_{L(\theta)} \end{pmatrix} \begin{pmatrix} x - x_0 \\ y - y_0 \\ 0 \end{pmatrix}$$

$$= \left(\frac{\partial f}{\partial x}|_{(\theta\mathbf{x}+(1-\theta)\mathbf{x}_0,z)}\right)(x - x_0) + \left(\frac{\partial f}{\partial y}|_{(\theta\mathbf{x}+(1-\theta)\mathbf{x}_0,z)}\right)(y - y_0). \qquad (1)$$

The IFT - Proof (4)

- Consider the term $[f(\mathbf{x}_0, z) - f(\mathbf{x}_0, z_0)]$
- The line segment in **R**^{3} linking $(\mathbf{x}_0, z)$ to $(\mathbf{x}_0, z_0)$ is:

$$L : [0,1] \to \mathbf{R}^3 : t \to (\mathbf{x}_0, tz + (1-t)z_0)$$
$$= (x_0, y_0, tz + (1-t)z_0)$$

- Next, define $h = f \circ L : [0,1] \to \mathbf{R}$. Then, for some $\phi \in (0,1)$, applying the Mean value theorem it follows

$$f(\mathbf{x}_0, z) - f(\mathbf{x}_0, z_0) = h(1) - h(0) = h'(\phi)$$

- Applying the chain rule to compute $h'(\phi)$:

The IFT - Proof (5)

$$
\begin{aligned}
h'(\phi) &= \begin{pmatrix} \frac{\partial f}{\partial x}|_{L(\phi)} & \frac{\partial f}{\partial y}|_{L(\phi)} & \frac{\partial f}{\partial z}|_{L(\phi)} \end{pmatrix} \begin{pmatrix} 0 \\ 0 \\ z - z_0 \end{pmatrix} \\
&= \left(\frac{\partial f}{\partial z}|_{(\mathbf{x}_0, \phi z + (1-\phi) z_0)} \right) (z - z_0). \qquad (2)
\end{aligned}
$$

The IFT - Proof (6)

- From (1) and (2) we can write

$$f(\mathbf{x}, z) = \left(\frac{\partial f}{\partial x}|_{(\theta\mathbf{x}+(1-\theta)\mathbf{x}_0, z)}\right)(x - x_0) + \left(\frac{\partial f}{\partial y}|_{(\theta\mathbf{x}+(1-\theta)\mathbf{x}_0, z)}\right)(y - y_0) + \left(\frac{\partial f}{\partial z}|_{(\mathbf{x}_0, \phi z+(1-\phi)z_0)}\right)(z - z_0). \quad (3)$$

for some $\theta, \phi \in (0, 1)$.

- Now choose

$$a_0 \in (0, a), \quad \text{and} \quad \delta < \min\{a_0, \frac{ba_0}{2M}\}$$

The IFT - Proof (7)

- If $\|\mathbf{x} - \mathbf{x}_0\| < \delta$, then it is easy to see that

$$|\Big(\frac{\partial f}{\partial x}|_{(\theta\mathbf{x}+(1-\theta)\mathbf{x}_0,z)}\Big)(x-x_0)+\Big(\frac{\partial f}{\partial y}|_{(\theta\mathbf{x}+(1-\theta)\mathbf{x}_0,z)}\Big)(y-y_0)| < ba_0$$

so that

$$f(\mathbf{x}, z_0 + a_0) > 0 \quad \text{and} \quad f(\mathbf{x}, z_0 - a_0) < 0.$$

- Applying the intermediate value theorem,

$$\exists z \in (z_0 - a_0, z_0 + a_0) \quad \text{s.t.} \quad f(\mathbf{x}, z) = 0$$

- Also that value is unique, because since $\frac{\partial f}{\partial z} > 0$ it may have at most one root.

The IFT - Proof (8)

- In other words, take

$$U = B(\mathbf{x}_0, \delta) \quad \text{and} \quad V = (z_0 - a_0, z_0 + a_0)$$

  for each $\mathbf{x} \in U$ there is a unique $z \in V$ such that $f(\mathbf{x}, z) = 0$.
- Thus, we can write $z = g(x, y)$.

Differentiation of an implicit function

- Let $f(x_1, x_2) = k,\ k \in$ **R** be (continuously) differentiable.
- This is a level set of function $f(x_1, x_2)$.
- Assume this function allows to define $x_2 = g(x_1),\ \forall x_1 \in I \subset$ **R**
- Hence, $f(x_1, x_2) = f(x_1, g(x_1)) = \phi(x_1)$ and thus $\phi(x_1) = k$
- Question: Value of dx_2/dx_1 at a point p?
- Answer: slope of the tangent to $f(x_1, x_2) = k$
- How to compute that slope?
- Applying the chain rule, $\frac{d\phi(x_1)}{dx_1} = \frac{\partial f}{\partial x_1} + \frac{\partial f}{\partial x_2}\frac{dx_2}{dx_1}$
- Since $\phi(x_1) = k, \forall x_1 \in I$, it follows $\frac{d\phi}{dx_1} = 0$.
- Thus, $\frac{d\phi(x_1)}{dx_1} = \frac{\partial f}{\partial x_1} + \frac{\partial f}{\partial x_2}\frac{dx_2}{dx_1} = 0$ or
- $\frac{dx_2}{dx_1}|_p = -\frac{\partial f}{\partial x_1}|_p / \frac{\partial f}{\partial x_2}|_p$, with $\frac{\partial f}{\partial x_2} \neq 0$.

Differentiation of an implicit function - Illustration

- Consider $f(x,y) = y^3 + x^2 - 3xy - 7 = 0$ around $(x,y) = (4,3)$
- Suppose $y(x)$ exists solving $f(x,y) = 0$ around $(4,3)$
- Substitute $y(x)$ into $f(x,y)$:
  $[y(x)]^3 + x^2 - 3x[y(x)] - 7 = 0$
- Differentiate wrt x (use Chain rule):

$$3[y(x)]^2 \frac{\partial dy}{\partial dx} + 2x - 3y(x) - 3x \frac{\partial dy}{\partial dx} = 0$$

$$\frac{\partial dy}{\partial dx} = \frac{3y(x) - 2x}{3[y(x)]^2 - 3x}$$

- Then, $\frac{dy}{dx}|_{(4,3)} = \frac{1}{15}$
- Remark: $\frac{dy}{dx}$ exists if $3[y(x)]^2 - 3x \neq 0$.
- Again $\frac{df}{dy} \neq 0$ requiered.

The Implicit function theorem - an economic application

A macro model of income determination

- Notation
  - Y: national income = GDP
  - T: taxes (lump sum)
  - Y_d: disposable income, $Y_d = Y - T$
  - C: consumption. $C(Y_d),\ dC/dY_d \in (0,1)$
  - I: investment
  - G: government expenditure
- suppose macro equilibrium: aggr supply = aggr demand
  $Y = C(Y - T) + I + G$

Questions

- Can we express Y as a function of I, G, T?
- How variations in I, G, T affect Y?

The Implicit function theorem - an economic application (2)

Question (a)

- Define $F(Y, I, G, T) = Y - C(Y - T) - I - G$
- The Implicit function theorem tells us that $Y^*(I, G, T)$ exists in a neighborhood of (I, G, T) if $\frac{\partial F}{\partial Y} \neq 0$.
- Let us verify it:

$$\frac{\partial F}{\partial Y} = 1 - \frac{\partial C}{\partial Y_d}\frac{\partial Y_d}{\partial Y} = 1 - \frac{\partial C}{\partial Y_d} > 0$$

- therefore such a function exists.

The Implicit function theorem - an economic application (3)

Question (b) - Comparative statics

- The Implicit function theorem tells us that

$$\frac{\partial Y^*}{\partial I} = -\frac{\partial F/\partial I}{\partial F/\partial Y} = -\frac{-1}{1-\frac{\partial C}{\partial Y_d}} = \frac{1}{1-\frac{\partial C}{\partial Y_d}} > 0$$

$$\frac{\partial Y^*}{\partial G} = -\frac{\partial F/\partial G}{\partial F/\partial Y} = -\frac{-1}{1-\frac{\partial C}{\partial Y_d}} = \frac{1}{1-\frac{\partial C}{\partial Y_d}} > 0$$

$$\frac{\partial Y^*}{\partial T} = -\frac{\partial F/\partial T}{\partial F/\partial Y} = -\frac{-\frac{\partial C}{\partial Y_d}\frac{\partial Y_d}{\partial T}}{1-\frac{\partial C}{\partial Y_d}} = -\frac{\frac{\partial C}{\partial Y_d}}{1-\frac{\partial C}{\partial Y_d}} < 0$$

More on comparative statics

- A firm produces y using an input x; $f(x) = x^{\alpha}, \ \alpha \in (0,1)$
- Competitive markets for output and input
- Market prices: p and w
- Profit function: $\pi(x) = px^{\alpha} - wx$
- Questions: (a) x max π; (b) dx/dw
- (a) x max π
  - $d\pi/dx = 0 \rightarrow p\alpha x^{(\alpha-1)} - w = 0$
- (b) Assess dx/dw
  - Define $F(x,w) = p\alpha x^{(\alpha-1)} - w$
  - Since $dF/dx = (\alpha-1)\alpha p x^{(\alpha-2)} < 0$, (i.e. $\neq 0$), apply IFT
  - $$\frac{dx}{dw} = -\frac{\partial F/\partial w}{\partial F/\partial x} = -\frac{-1}{(\alpha-1)\alpha p x^{(\alpha-2)}} < 0$$

A more general set-up

- Let $f(x_1, x_2, x_3) = k$ be (continuously) differentiable.
- Assume this function allows to define $x_3 = g(x_1, x_2),\ \forall (x_1, x_2) \in I$
- The IFT guarantees g is also continuously differentiable.
- Rewrite $f(x_1, x_2, x_3) = f(x_1, x_2, g(x_1, x_2)) = \phi(x_1, x_2)$
- Since $\phi(x_1, x_2) = k, \forall x_1 \in I$, it follows $\frac{d\phi}{dx_1} = \frac{d\phi}{dx_2} = 0$.
- Apply chain rule to compute $\frac{d\phi}{dx_1}$ and $\frac{d\phi}{dx_2}$:
- $\frac{d\phi}{dx_1} = 0 = \frac{\partial f}{\partial x_1} + \frac{\partial f}{\partial x_3}\frac{\partial x_3}{\partial x_1}$. Therefore, $\frac{\partial x_3}{\partial x_1} = -\frac{\partial f}{\partial x_1} / \frac{\partial f}{\partial x_3}$
- $\frac{d\phi}{dx_2} = 0 = \frac{\partial f}{\partial x_2} + \frac{\partial f}{\partial x_3}\frac{\partial x_3}{\partial x_2}$. Therefore, $\frac{\partial x_3}{\partial x_2} = -\frac{\partial f}{\partial x_2} / \frac{\partial f}{\partial x_3}$
- with $\frac{\partial f}{\partial x_3} \neq 0$.

The general case

- Let $f(x_1, \ldots, x_n) = k$ be differentiable.
- Assume this function allows to define
  $x_n = g(x_1, \ldots, x_{n-1}),\ \forall (x_1, \ldots, x_{n-1}) \in I$
  where I is defined as the set of points $(x_1, \ldots, x_{n-1})$ satisfying $f(x_1, \ldots, x_n) = k$.
- The IFT guarantees g is also continuously differentiable.
- Rewrite $f(x_1, \ldots, x_n) = \phi(x_1, \ldots, x_{n-1})$
- Since $\phi(x_1, \ldots, x_{n-1}) = k$, it follows $\frac{d\phi}{dx_i} = 0, i = 1, \ldots, n-1$.
- Apply chain rule to compute $\frac{d\phi}{dx_i}$:
- $\frac{d\phi}{dx_i} = 0 = \frac{\partial f}{\partial x_i} + \frac{\partial f}{\partial x_n}\frac{\partial x_n}{\partial x_i},\ \ i = 1, \ldots, n-1.$
- Thus, $\frac{\partial z}{\partial x_i} = -\frac{\partial f}{\partial x_i} / \frac{\partial f}{\partial x_n},\ \ i = 1, \ldots, n-1$, with $\frac{\partial f}{\partial x_n} \neq 0$.

The inverse function

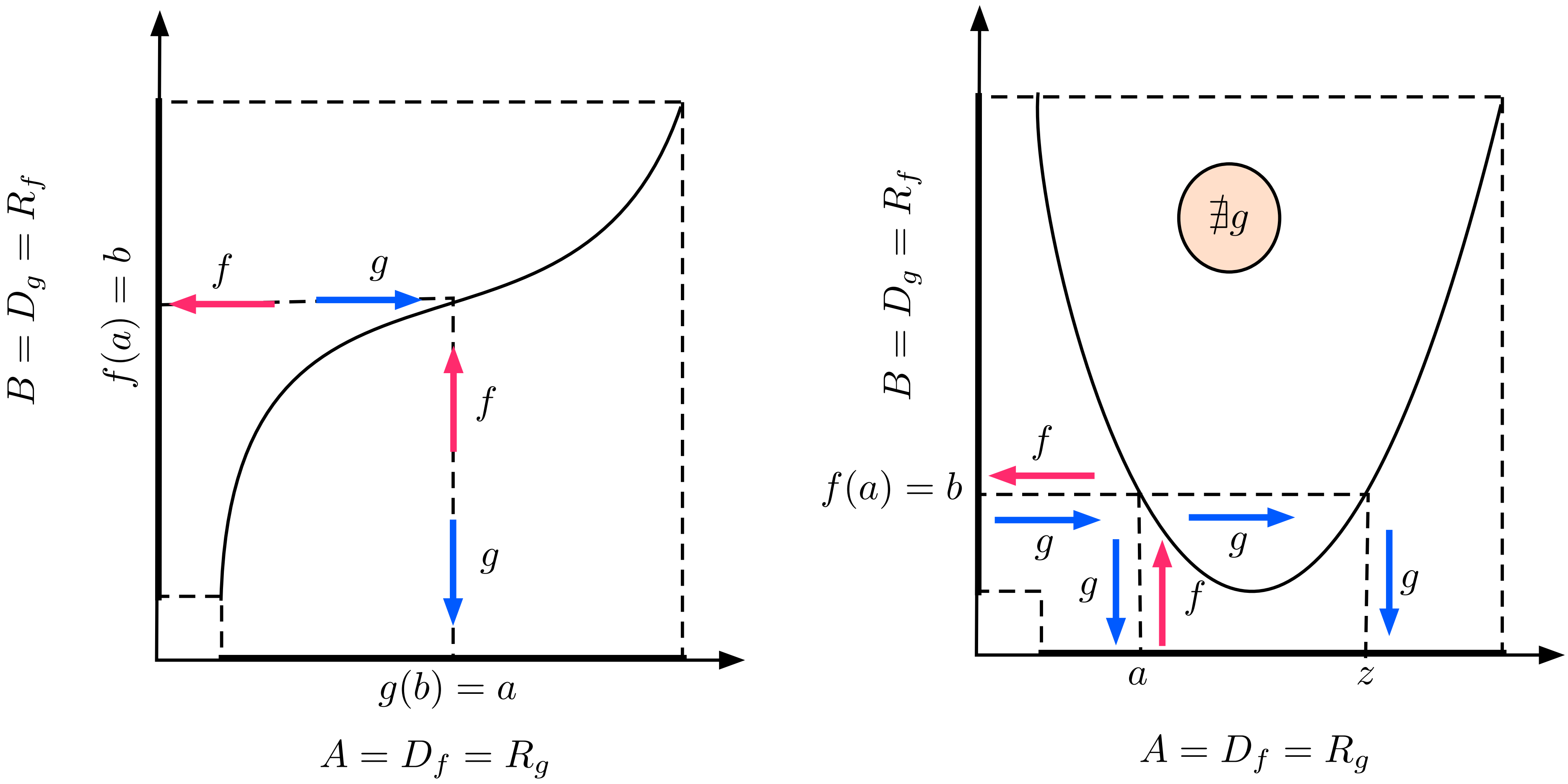
Definition

- Let A and B be two sets
- Let $f : A \to B$ be one-to-one.
- An inverse function for f (often denoted as f^{-1}) is a function $g : B \to A$ satisfying

$$f(g(b)) = b, \ \forall b \in B, \text{ and}$$
$$f(g(a)) = a, \ \forall a \in A$$

- Remark 1: If g is inverse for f, then f is inverse for g
- Remark 2: $[D_f = R_g; \ D_g = R_f] \to f$ and g defined on the same dimensional space.

Inverse function - Illustration



The inverse function (2)

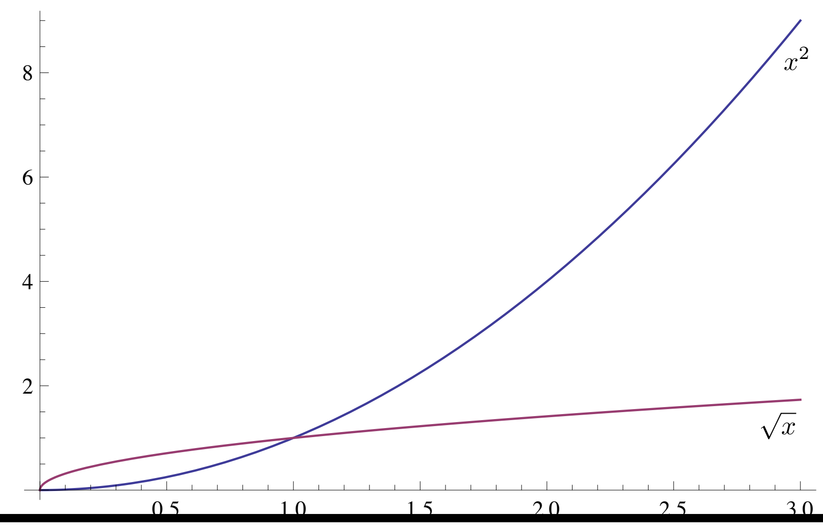
Example

$$f(x) = x^2,\ x \in [0, \infty) \ \leftrightarrow\ g(x) = \sqrt{x}, \forall x \in [0, \infty)$$

Alternative notations

$$f(x) = x^2,\ x \in [0, \infty) \ \leftrightarrow\ f^{-1}(x) = \sqrt{x}, \forall x \in [0, \infty)$$

$$y = x^2,\ x \in [0, \infty) \ \leftrightarrow\ x = \sqrt{y}, \forall y \in [0, \infty)$$



Inverse function - Theorem

One variable

- If $f(x) : \mathbf{R} \to \mathbf{R}$ is continuously differentiable (C^1) at x_0 with $f'(x_0) \neq 0$, then
  a) f is invertible in $B(x_0, r)$, and
  b) f^{-1} is also C^1 and $\left(f^{-1}\right)'(f(x_0)) = \frac{1}{f'(x_0)}$

Several variables

- If $f(x_1, \ldots, x_n) : \mathbf{R}^n \to \mathbf{R}^n$ is continuously differentiable (C^1) at $p = (\tilde{x}_1, \ldots, \tilde{x}_n)$ with $J_f(p) \neq 0$, then
  a) f is invertible around p, and
  b) f^{-1} is also C^1 and $J_{f^{-1}}(f(p)) = \frac{1}{J_f(p)}$

Inverse function - Example

Compute inverse function

$$f(x) = x^3 - 4 \longleftrightarrow f^{-1}(x) = \Big(f(x) + 4\Big)^{1/3}$$

Compute $(f^{-1})'(212)$

Step 1

$$\left.\begin{aligned} f(x) &= 212 \\ x = (216)^{1/3} &= 6 \end{aligned}\right\} \longrightarrow f(6) = 212$$

Step 2 $f'(x) = 3x^2; \quad f'(6) = 108$

Step 3 $(f^{-1})'(212) = \frac{1}{f'(6)} = 1/108.$

The inverse function theorem - example

- Let $f = (f_1, f_2)$ with $u(x, y) \equiv f_1(x, y) = \frac{x^4+y^4}{x}$ and $v(x, y) \equiv f_2(x, y) = \sin x + \cos y$.
- 1.- Find the points (x_0, y_0) around which we can solve for (x, y) in terms of (u, v)
  - Compute the Jacobian matrix:
    $$Jf = \begin{pmatrix} \frac{\partial f_1}{\partial x} & \frac{\partial f_1}{\partial y} \\ \frac{\partial f_2}{\partial x} & \frac{\partial f_2}{\partial y} \end{pmatrix} = \begin{pmatrix} \frac{3x^4-y^4}{x^2} & \frac{4y^3}{x} \\ \cos x & -\sin y \end{pmatrix}$$
  - Compute $detJf = -\frac{(3x^4-y^4)\sin y}{x^2} - \frac{4y^3 \cos x}{x}$
  - We are looking for (x, y) s.t. $detJf \neq 0$. In general this cannot be solved explicitly.
  - In this example one such points is $(x_0, y_0) = (\frac{\pi}{2}, \frac{\pi}{2})$.
  - Around that point we can obtain $x = g_1(u, v)$ and $y = g_2(u, v)$.

The inverse function theorem - example (cont'd)

- 2.- Compute $\frac{\partial x}{\partial u}, \frac{\partial x}{\partial v}, \frac{\partial y}{\partial u}, \frac{\partial y}{\partial v}$
  - The theorem tells us to invert the Jacobian matrix:

$$[Jf]^{-1} = \frac{1}{detJf} \begin{pmatrix} \frac{\partial v}{\partial y} & -\frac{\partial u}{\partial y} \\ -\frac{\partial v}{\partial x} & \frac{\partial u}{\partial x} \end{pmatrix} = \begin{pmatrix} \frac{\partial x}{\partial u} & \frac{\partial x}{\partial v} \\ \frac{\partial y}{\partial v} & \frac{\partial y}{\partial v} \end{pmatrix}$$

  - That is

$$\frac{\partial x}{\partial u} = \left(\frac{1}{-\frac{(3x^4-y^4)\sin y}{x^2} - \frac{4y^3\cos x}{x}} \right) \sin y = \frac{-x^2 \sin y}{\sin y(y^4-3x^4)-4y^3 x \cos x}$$

etc.

The inverse function - An economic illustration

- Assume the demand of a certain consumption good depends on its price according to $q_d = f(p)$
- From the perspective of the consumer this is often the way to look at the demand decision. Consumers observe the price and decide the amount to buy.
- From the perspective of an oligopolistic firm the decision of the choice variable hinges on the type of competition:
  - Cournot competition: firms choose the (optimal) volume of production and market forces determine the price. These firms defines their profit functions using the inverse demand function $p = f^{-1}(q_d)$
  - Bertrand competition: firms choose the (optimal) price and market forces determine the quantity bought. These firms defines their profit functions using the (direct) demand function $q_d = f(p)$

Homogeneous functions

Definition

- A function $f(x_1, \ldots, x_n)$ defined on a domain D is homogeneous of degree k if $\forall (x_1, \ldots, x_n) \in D,\ f(tx_1, \ldots, tx_n) = t^k f(x_1, \ldots, x_n)$

Example: Cobb-Douglas function

- $f(x, y) = x^a y^b$ is homogeneous of degree $a + b$ (H^{a+b})
- $f(tx, ty) = (tx)^a (ty)^b = t^{(a+b)} x^a y^b = t^{(a+b)} f(x, y)$
- Let f be a production function
- If $(a + b) = 1, f(tx, ty) = tf(x, y)$ Constant returns to scale
- If $(a + b) > 1, f(tx, ty) = t^{(a+b)} f(x, y) > tf(x, y)$ IRS
- If $(a + b) < 1, f(tx, ty) = t^{(a+b)} f(x, y) < tf(x, y)$ DRS

Homogeneous functions (2)

Theorem (Euler)

- Let $f(x_1, \ldots, x_n)$ be continuously differentiable in an open domain D.
- Let $t > 0$ such that $(x_1, \ldots, x_n) \in D$ implies $(tx_1, \ldots, tx_n) \in D$.
- Then, f is homogeneous of degree k *iff* $\sum_{i=1}^{n} x_i \frac{\partial f(x_1,\ldots,x_n)}{\partial x_i} = kf(x_1, \ldots, x_n), \forall (x_1, \ldots, x_n) \in D$.
- i.e. $f(tx_1, \ldots, tx_n) = t^k f(x_1, \ldots, x_n) \Longleftrightarrow \sum_{i=1}^{n} x_i \frac{\partial f(x_1,\ldots,x_n)}{\partial x_i} = kf(x_1, \ldots, x_n)$

Homogeneous functions (3)

Proof of Euler's theorem

- Step 1 ($\Rightarrow$)
  - Suppose f is homogeneous if degree k. Then, $f(tx_1,\ldots,tx_n) = t^k f(x_1,\ldots,x_n)$
  - Differentiating wrt t we obtain $\sum_{i=1}^n x_i \frac{\partial f(tx_1,\ldots,tx_n)}{\partial x_i} = kt^{k-1} f(x_1,\ldots,x_n)$
  - Set $t=1$ so that $\sum_{i=1}^n x_i \frac{\partial f(x_1,\ldots,x_n)}{\partial x_i} = kf(x_1,\ldots,x_n)$
- Step 2 ($\Leftarrow$)
  - Assume $\sum_{i=1}^n x_i \frac{\partial f(x_1,\ldots,x_n)}{\partial x_i} = kf(x_1,\ldots,x_n) \quad [\alpha]$
  - Fix $(x_1,\ldots,x_n)$ and define $\forall t > 0$, $g(t) = t^{-k} f(tx_1,\ldots,tx_n) - f(x_1,\ldots,x_n) \quad [\beta]$

Homogeneous functions (4)

Proof of Euler's theorem (cont'd)

- Step 2 ($\Leftarrow$)
  - Differentiate $g(t)$ to obtain
    $g'(t) =$
    $-kt^{-k-1}f(tx_1,\ldots,tx_n) + t^{-k}\sum_{i=1}^{n} x_i \frac{\partial f(tx_1,\ldots,tx_n)}{\partial x_i} \quad [\gamma]$
  - Given that $(tx_1,\ldots,tx_n) \in D$, $[\alpha]$ must also hold when replacing x_i by tx_i. Therefore,
    $\sum_{i=1}^{n} tx_i \frac{\partial f(tx_1,\ldots,tx_n)}{\partial tx_i} = kf(tx_1,\ldots,tx_n) \quad [\delta]$
  - Substitute $[\delta]$ in $[\beta]$ to obtain $\forall t > 0$
    $g'(t) = -kt^{-k-1}f(tx_1,\ldots,tx_n) + kt^{-k-1}f(tx_1,\ldots,tx_n)$
    i.e. $g'(t) = 0$
  - Accordingly $g(t)$ must be a constant function. To identify that constant, just note that from $[\beta]$ we obtain $g(1) = 0$. Therefore $g(t) = 0$.

Homogeneous functions (5)

Homogeneous functions - Proof of Euler's theorem (cont'd)

- Step 2 ($\Leftarrow$)(cont'd)
  - Applying $g(t) = 0$ in $[\beta]$ yields
    $t^{-k} f(tx_1, \dots, tx_n) = f(x_1, \dots, x_n)$ or
    $f(tx_1, \dots, tx_n) = t^k f(x_1, \dots, x_n)$
    meaning that f is homogeneous of degree k.

Homothetic functions

Definition

- A function $f : \mathbf{R}^n \to \mathbf{R}$ is homothetic if it can be obtained as the composition of a homogeneous function $h : \mathbf{R}^n \to \mathbf{R}$ and a monotonic increasing function $g : \mathbf{R} \to \mathbf{R}$.
  That is, $f = g(h(x_1, \ldots, x_n))$ or equivalently, f is a monotonic transformation of a homogeneous function.

Two properties

- Theorem 1: The level sets of a homothetic function are radial expansions of one another, that is
  $f(x_1, \ldots, x_n) = f(y_1, \ldots, y_n)$ implies
  $f(tx_1, \ldots, tx_n) = f(ty_1, \ldots, ty_n), \; t > 0.$
- Theorem 2: the slopes of the level sets of a homothetic function along a ray from the origin are constant, that is,
  $$-\frac{\frac{\partial f(tx_1,\ldots,tx_n)}{\partial x_i}}{\frac{\partial f(tx_1,\ldots,tx_n)}{\partial x_j}} = -\frac{\frac{\partial f(x_1,\ldots,x_n)}{\partial x_i}}{\frac{\partial f(x_1,\ldots,x_n)}{\partial x_j}}, \; \forall i, j, t > 0.$$

Homothetic functions (2)

Proof of theorem 1

- Because f is homothetic, $f(tx_1, \ldots, tx_n) = g(h(tx_1, \ldots, tx_n))$
- Because $h(x_1, \ldots, x_n)$ is homogeneous, $h(tx_1, \ldots, tx_n) = t^k h(x_1, \ldots, x_n)$
- Because we deal with level sets, $h(x_1, \ldots, x_n) = h(y_1, \ldots, y_n)$
- Combining altoghether, $f(tx_1, \ldots, tx_n) = g(h(tx_1, \ldots, tx_n)) = g(t^k h(x_1, \ldots, x_n)) = g(t^k h(y_1, \ldots, y_n)) = g(h(ty_1, \ldots, y_n)) = f(ty_1, \ldots, ty_n)$

Homothetic functions (3)

Proof of theorem 2

- In consumer theory, the theorem would say that the MRS for a homothetic function is homogeneous of degree zero.
- Because $f(tx_1, \ldots, tx_n)$ is homogeneous, $\frac{\partial f(tx_1,\ldots,tx_n)}{\partial x_i} = \frac{\partial g(h(tx_1,\ldots,tx_n))}{\partial x_i}$.
- Computing the derivative, $\frac{\partial g(h(tx_1,\ldots,tx_n))}{\partial x_i} = g'(h(tx_1, \ldots, tx_n))\frac{\partial h(tx_1,\ldots,tx_n)}{\partial x_i}$
- Combining these expressions,

$$\frac{\frac{\partial f(tx_1,\ldots,tx_n)}{\partial x_i}}{\frac{\partial f(tx_1,\ldots,tx_n)}{\partial x_j}} = \frac{\frac{\partial g(h(tx_1,\ldots,tx_n))}{\partial x_i}}{\frac{\partial g(h(tx_1,\ldots,tx_n))}{\partial x_j}} =$$

$$\frac{g'(h(tx_1, \ldots, tx_n))\frac{\partial h(tx_1,\ldots,tx_n)}{\partial x_i}}{g'(h(tx_1, \ldots, tx_n))\frac{\partial h(tx_1,\ldots,tx_n)}{\partial x_j}} = \frac{\frac{\partial h(tx_1,\ldots,tx_n)}{\partial x_i}}{\frac{\partial h(tx_1,\ldots,tx_n)}{\partial x_j}}$$

Homothetic functions (4)

Proof of theorem 2 (cont'd)

- Because h is homogeneous,

$$\frac{\frac{\partial h(tx_1,\ldots,tx_n)}{\partial x_i}}{\frac{\partial h(tx_1,\ldots,tx_n)}{\partial x_j}} = \frac{\frac{\partial t^k h(x_1,\ldots,x_n)}{\partial x_i}}{\frac{\partial t^k h(x_1,\ldots,x_n)}{\partial x_j}} = \frac{t^k \frac{\partial h(x_1,\ldots,x_n)}{\partial x_i}}{t^k \frac{\partial h(x_1,\ldots,x_n)}{\partial x_j}} = \frac{\frac{\partial h(x_1,\ldots,x_n)}{\partial x_i}}{\frac{\partial h(x_1,\ldots,x_n)}{\partial x_j}}$$

- Summarizing we have obtained

$$\frac{\frac{\partial f(tx_1,\ldots,tx_n)}{\partial x_i}}{\frac{\partial f(tx_1,\ldots,tx_n)}{\partial x_j}} = \frac{\frac{\partial h(x_1,\ldots,x_n)}{\partial x_i}}{\frac{\partial h(x_1,\ldots,x_n)}{\partial x_j}} \qquad [\alpha]$$

- For $t = 1$, $[\alpha]$ becomes

$$\frac{\frac{\partial f(x_1,\ldots,x_n)}{\partial x_i}}{\frac{\partial f(x_1,\ldots,x_n)}{\partial x_j}} = \frac{\frac{\partial h(x_1,\ldots,x_n)}{\partial x_i}}{\frac{\partial h(x_1,\ldots,x_n)}{\partial x_j}} \qquad [\beta]$$

- Combining $[\alpha]$ and $[\beta]$ completes the proof.

Homogeneous vs homothetic functions

- A homogeneous function of degree k is homothetic
  - Let $f(x)$ be a homogeneous function of degree k
  - Let H be a strictly increasing function
  - Define $F(x) = H(f(x))$. Then F is homothetic. To see why, take (x, y) such that
  - $F(x) = F(y)$ and show that $F(tx) = F(ty)$.
  - If $F(x) = F(y)$ then $H(f(x)) = H(f(y))$.
  - because $H' > 0$ it follows $f(x) = f(y)$
  - because f is homogeneous of degree k, for $t > 0$ we have $F(tx) = H(f(tx)) = H(t^k f(x)) = H(t^k F(y)) = H(f(ty)) = F(ty)$
  - thus proving that F is homothetic.
- the converse does not hold.

Homogeneous vs homothetic functions (2)

- Not all homothetic functions are homogeneous.
  - Let $F(x, y) = a\log(x) + b\log(y) = \log(x^a y^b)$ for all $x > 0, y > 0$ with $a > 0, b > 0$
  - the *log* function is strictly increasing
  - the function $x^a y^b$ is homogeneous of degree $a + b$.
  - Thus, $F(x, y)$ is a stricly increasing function of a homogeneous function. But it is not homothetic. Let's see why:
  - $F(tx, ty) = \log((tx)^a (ty)^b) = \log(t^{a+b} x^a y^b) = (a + b)\log(t) + \log(x^a y^b)$
  - which cannot be written as $t^k \log(x^a y^b)$ for any value of k.

Homogeneous vs homothetic functions (3)

- Economic applications
  - Consumer theory: Homogeneous preferences and implications for properties of the demand functions
  - Producer theory: Production functions (and their dual cost functions) and their implications for properties os supply functions
  - Implications of homogeneous/homothetic functions on the properties of market equilibrium.

Approximation of functions

Motivation

- f may be extremely complex
- often interest of analysis only around some point (e.g. equilibrium point), or subdomain
- obtaining information about $f(x)$ for $x \in B(x_0, r)$ is often sufficient
- approximating $f(x)$ for $x \in B(x_0, r)$ by means of an auxiliary (polynomial) function
- trade-off between simplicity of approximation and its accuracy

Approximation

- linear, quadratic, cubic, ...
- the higher the order of the polynomial the higher the accuracy of the approximation

Linear approximations

Definition

- Let $f(x)$ be differentiable.
- Let x_0 be a point in D_f
- A linear approximation to the value of $f(x)$ around x_0, is the tangent line to $f(x)$ at x_0
- The tangent line to $f(x)$ at x_0 has the equation: $P(x) = A_0 + A_1(x - x_0)$
- Question: How to determine A_0 and A_1?
- $P(x)$ has to satisfy 2 conditions $P(x_0) = f(x_0)$ and $P'(x_0) = f'(x_0)$
- where $P(x_0) = A_0$ and $P'(x_0) = A_1$
- then $P(x) = f(x_0) + f'(x_0)(x - x_0)$ and
- $f(x) \approx P(x)$ for $x \in B(x_0, r)$

Linear approximations (2)

Example

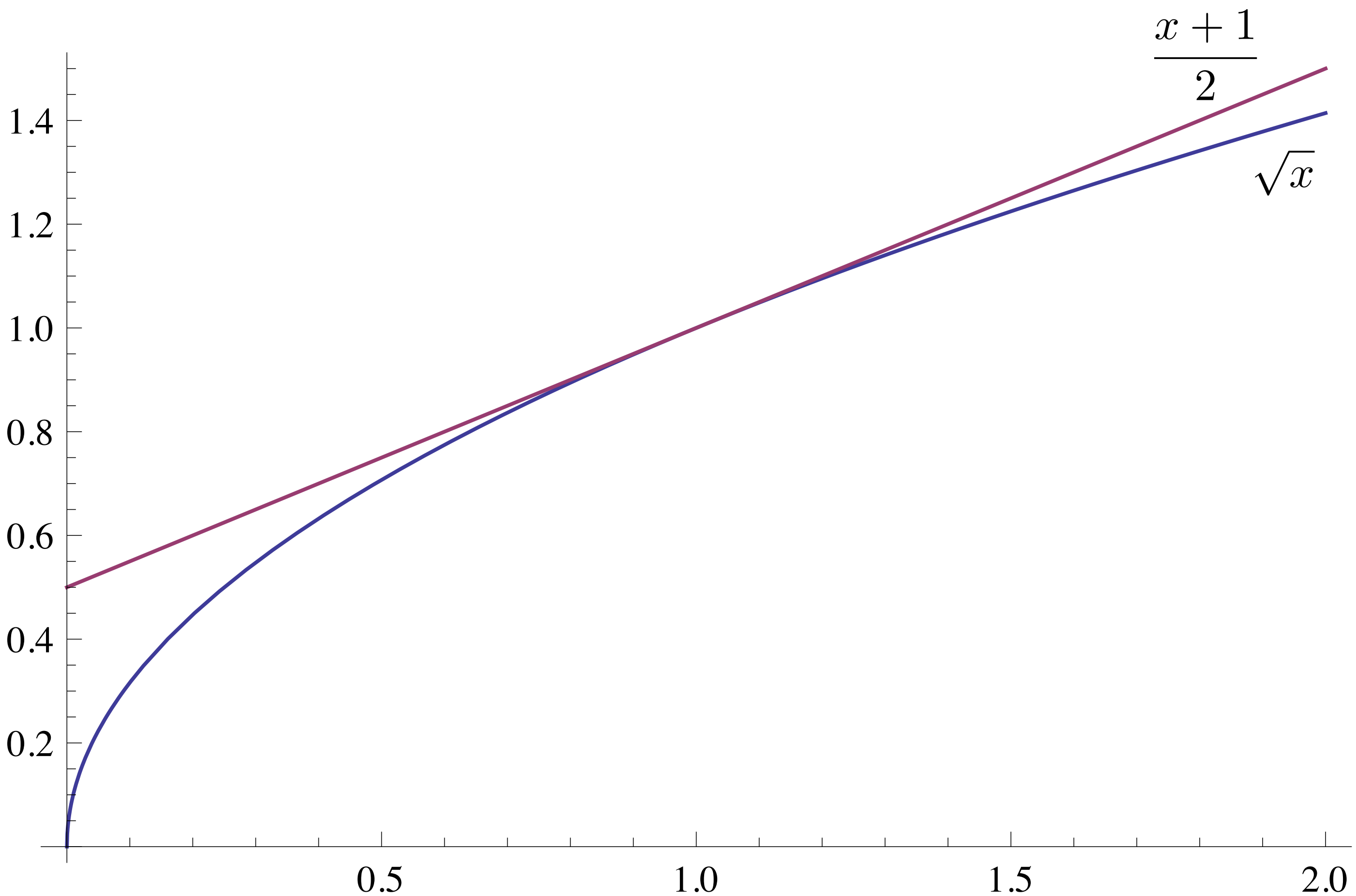
- Let $f(x) = \sqrt{x}$
- Find a linear approximation to $f(x)$ around $x_0 = 1$
- Near $x_0 = 1$ we have

$$P(x) = f(1) + f'(1)(x-1)$$

$$P(x) = 1 + \frac{1}{2}(x-1)$$

- A linear approximation to $f(x) = \sqrt{x}$ around $x = 1$ is given by $P(x) = 1 + \frac{1}{2}(x-1) = \frac{x+1}{2}$

Linear approximations (3)

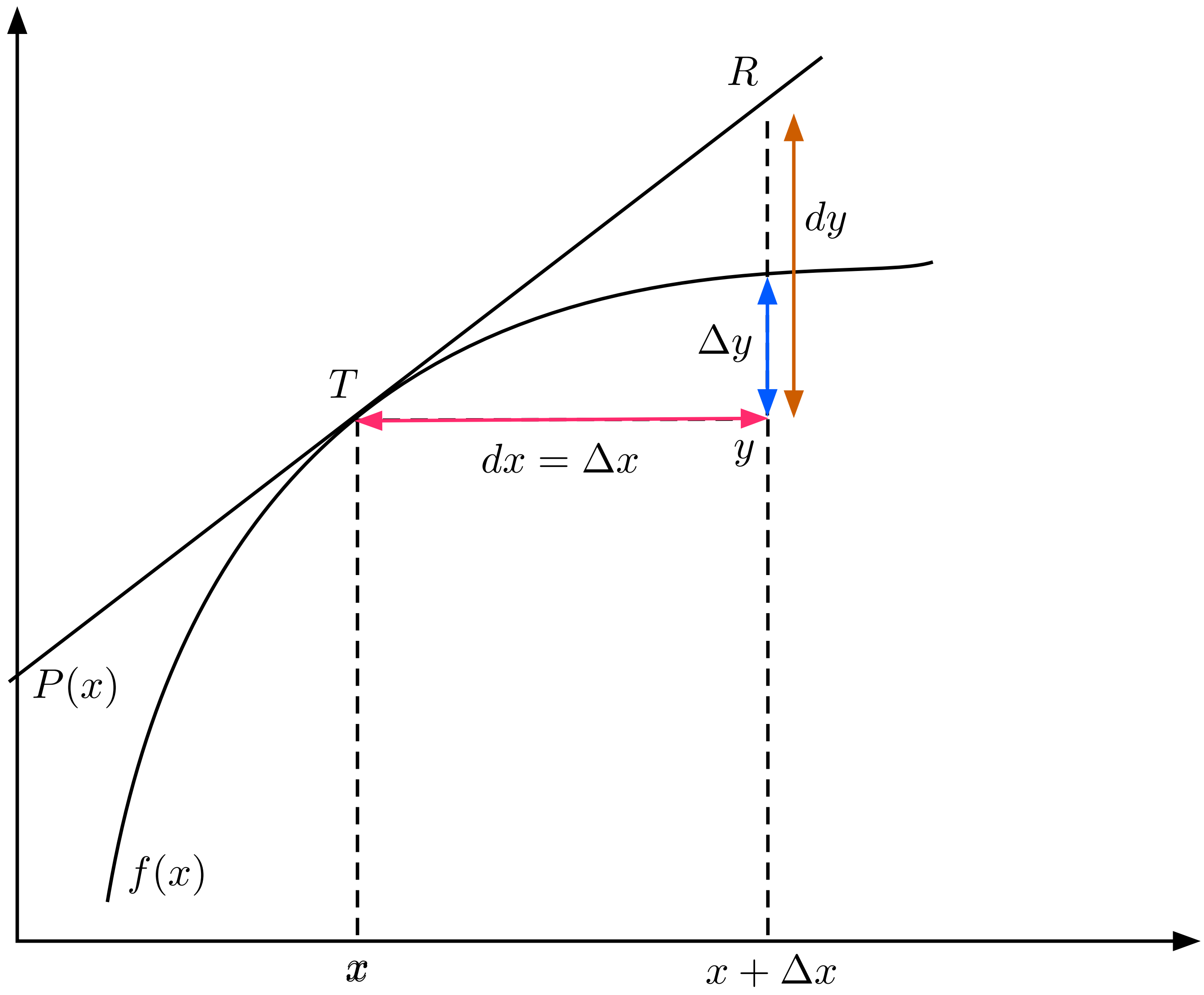


Linear approximation and differential of f

Differential of f

- Let $y = f(x)$ be differentiable.
- The differential dy is defined as
  $dy = f'(x)dx$
- where differential dx is an arbitrary variation of in the value of x.
- Remark: dy is proportional to dx with $f'(x)$ being the factor of proportionality $\rightarrow$
- dy represents NOT the change value of f(x) when x changes to $x + \Delta x$ (i.e. along $f(x)$) BUT the change in the value of y along the straight line with slope $f'(x)$ (i.e. along $P(x)$)
- Consider a movement from t to R (see figure).
  - moving along $f(x)$ yields $\Delta y = f(x + \Delta x) - f(x)$
  - moving along $P(x)$ yields $dy = P(x + \Delta x) - P(x)$

Linear approximation and differential of f (2)



Linear approximation and differential of f (2)

- for small $\Delta x, P(x)$ represents a linear approximation to $f(x)$
- Therefore $|dy - \Delta y|$ gives a measure of the error incurred when following the linear approximation instead of the function
- similar arguments in $\mathbf{R}^3$ and higher dimensions

Linear approximations in $\mathbb{R}^3$

Definition

- Let $z = f(x, y)$ be differentiable.
- Let $P = (a, b, c)$ be a point with $c = f(a, b)$.
- The tangent plane to $f(x, y)$ at P has the equation:
  $z - c = \frac{\partial f(a,b)}{\partial x}(x - a) + \frac{\partial f(a,b)}{\partial y}(y - b)$
- The tangent plane to $f(x, y)$ at P is a linear approximation to the value of $f(x, y)$ around P, i.e.
  $f(x, y) \approx f(a, b) + \frac{\partial f(a,b)}{\partial x}(x - a) + \frac{\partial f(a,b)}{\partial y}(y - b)$

Differential of a function in $\mathbb{R}^3$

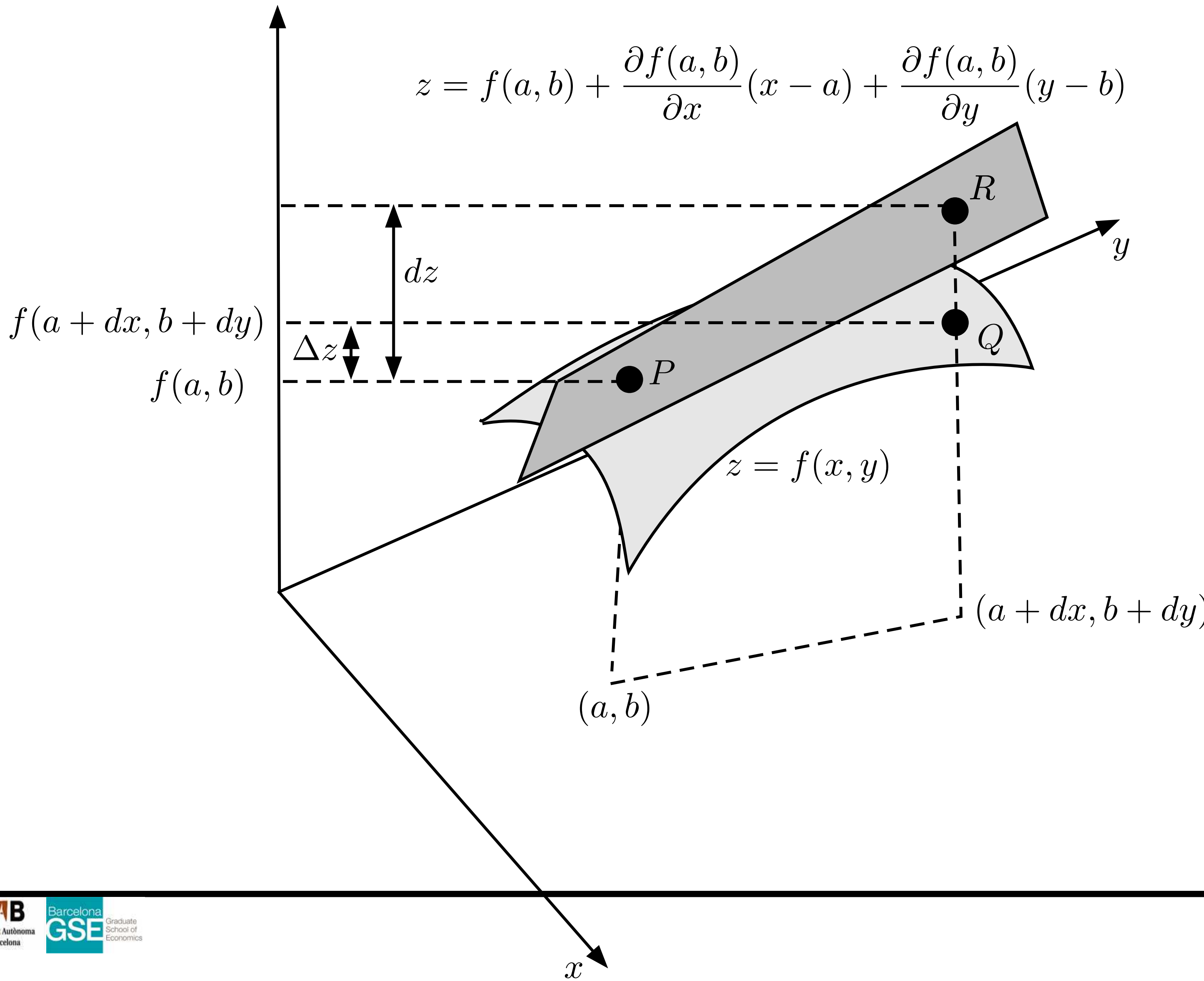
Definition

- Let $z = f(x, y)$ be differentiable.
- Let dx and dy be arbitrary real numbers (small or not)
- The differential of $z = f(x, y)$ at (a, b), denoted by dz (or df) is defined as $dz = \frac{\partial f(a,b)}{\partial x} dx + \frac{\partial f(a,b)}{\partial y} dy$
- In general, for $z = f(x_1, \ldots, x_n)$, $dz = \sum_{i=1}^{n} \frac{\partial f}{\partial x_i} dx_i$.

Measurement error

- Assume (a, b) varies to $(a + dx, b + dy)$.
- The variation in the value of f is $\Delta z = f(a + dx, b + dy) - f(a, b)$
- If dx and dy are small, then, $\Delta z \approx dz$.
- The difference $dz - \Delta z$ results from following the tangent plane instead of the surface.

Differential and tangent plane - Illustration



Higher order approximations and Taylor's theorem

Introduction

- Linear approximation $\to$ measurement error.
- Two questions:
  a) how to improve the accuracy of the approximation.
  b) how to evaluate the measurement error.
- Answers:
  a) Taylor's polynomial of degree n.
  b) Taylor's theorem and (extended) mean-value theorem.

Improving accuracy

- Let $f : [a, b] \to \mathbb{R}$ be continuously differentiable at $c \in (a, b)$.
- Linear approximation: fits slope around c:
  $f(x) \approx f(c) + f'(c)(x - c)$
- Quadratic approximation: fits slope and approximates curvature around c:
  $f(x) \approx f(c) + f'(c)(x - c) + \frac{1}{2!}f''(c)(x - c)^2$
- Approximations with polynomials of degrees $3, 4, \ldots$ allow to capture better and better the properties of $f(x)$ around c.
- Taylor's polynomial of degree n, $P_n(x)$:
  $P_n(x) =$
  $f(c) + f'(c)(x - c) + \frac{1}{2!}f''(c)(x - c)^2 + \cdots + \frac{1}{n!}f^{(n)}(c)(x - c)^n$
- Still measurement error: $E_n(x) = f(x) - P_n(x)$

Improving accuracy - Quadratic approximation

- A quadratic approximation to $f(x)$ around $x = x_0$ is a quadratic function tangent to $f(x)$ at x_0.
- The tangent quadratic function has the equation $P(x) = A_0 + A_1(x - x_0) + A_2(x - x_0)^2$
- Question: Determine A_0, A_1, A_2?
- $P(x)$ has to satisfy three conditions $P(x_0) = f(x_0), P'(x_0) = f'(x_0)$ and $P''(x) = f''(x)$
- As before, $A_0 = f(x_0), A_1 = f'(x_0)$
- $P''(x_0) = 2A_2$ so that $A_2 = \frac{1}{2}f''(x_0) = \frac{1}{2!}f''(x_0)$
- then $P(x) = f(x_0) + f'(x_0)(x - x_0) + \frac{1}{2!}f''(x_0)(x - x_0)^2$
- and $f(x) \approx f(x_0) + f'(x_0)(x - x_0) + \frac{1}{2!}f''(x_0)(x - x_0)^2$

Improving accuracy - Quadratic approximation (2)

Example

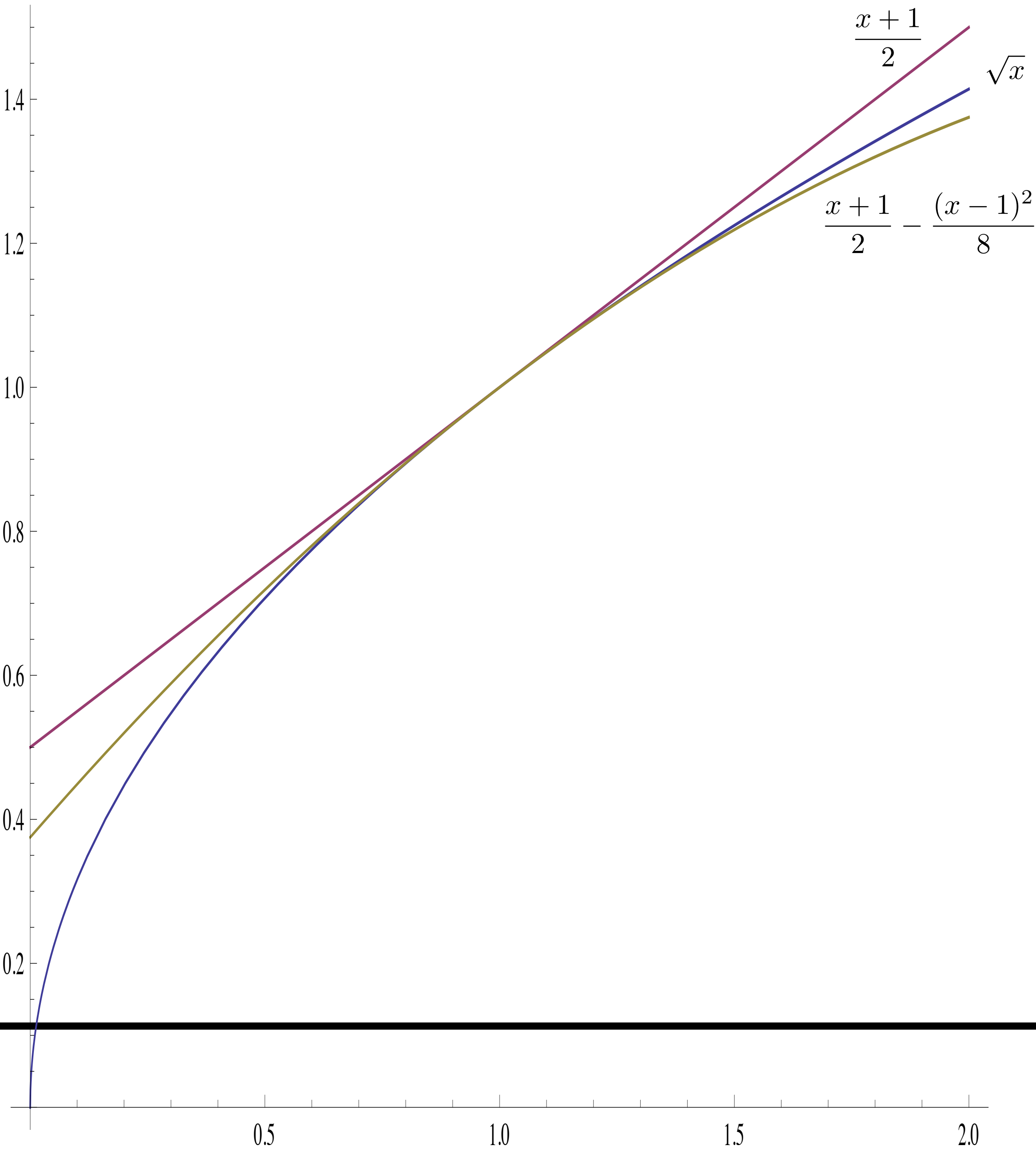
- Let $f(x) = \sqrt{x}$
- Find a quadratic approximation to $f(x)$ around $x_0 = 1$
- Near $x_0 = 1$ we have

$$P(x) = f(1) + f'(1)(x-1) + \frac{1}{2!}f''(1)(x-1)^2$$
$$P(x) = 1 + \frac{1}{2}(x-1) + \frac{1}{2!}\frac{-1}{4}(x-1)^2$$

- A quadratic approximation to $f(x) = \sqrt{x}$ around $x = 1$ is given by

$$P(x) = \frac{x+1}{2} - \frac{(x-1)^2}{8}$$

Improving accuracy - Quadratic approximation (3)



Improving accuracy (3)

- Generalization to function of multiple variables
- Let $f : \mathbf{R}^n \to \mathbf{R}$ be continuously differentiable at $c = (c_1, \ldots, c_n)$.
- Linear approximation: fits slope around c:
  $f(x) \approx f(c) + Df(c)(x - c)$ where $Df(c)$ is Jacobian matrix.
- Quadratic approximation: fits slope and approximates curvature around c:
  $f(x) \approx f(c) + f'(c)(x - c) + \frac{1}{2!}(x - c)^T Hf(c)(x - c)$ where $Hf(c)$ is Hessian matrix

Measuring the error

Recall

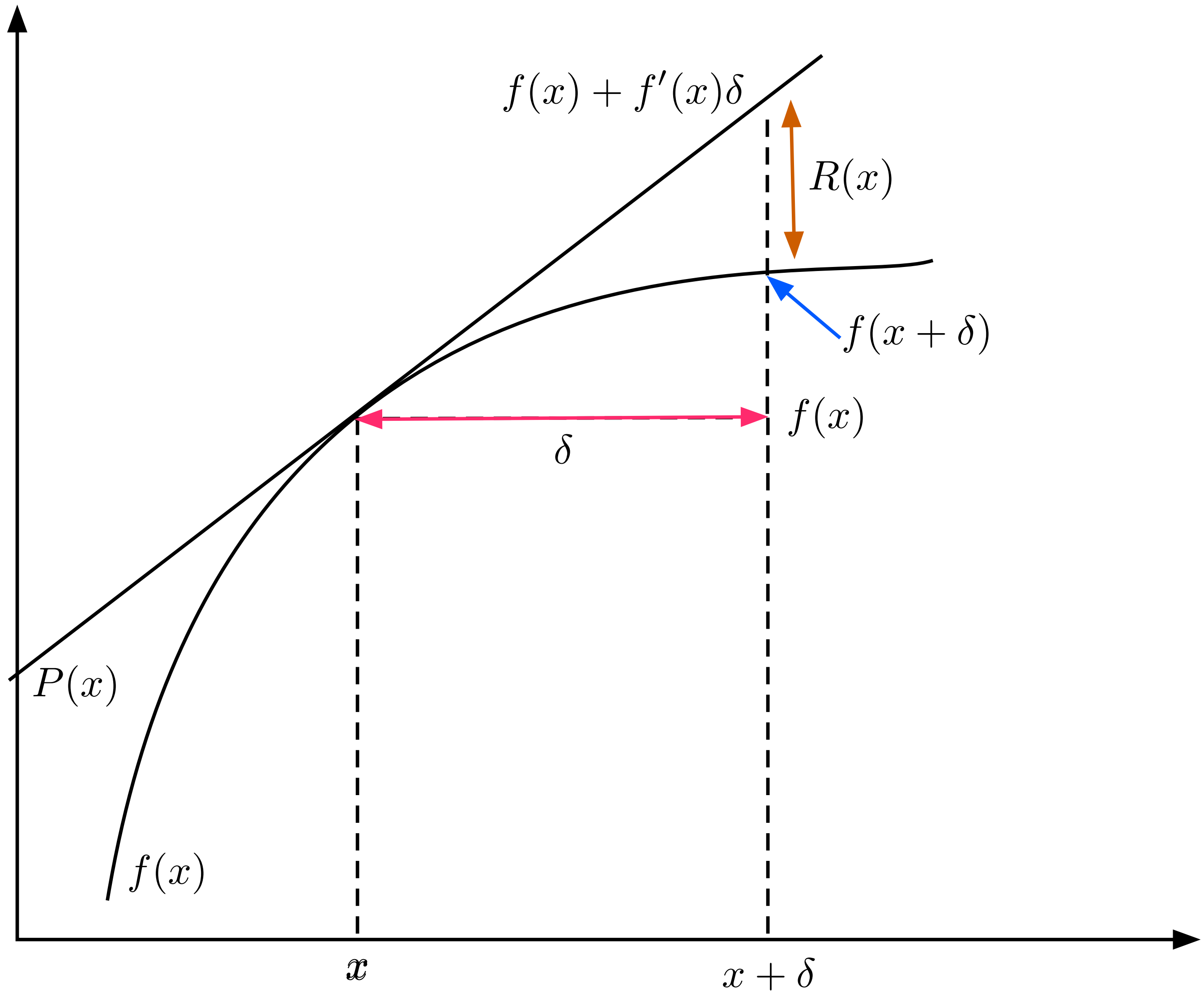
- Mean-value theorem
  - Let $f : [a, b] \to \mathbf{R}$ be continuous in $[a, b]$ and differentiable in (a, b). Then, $\exists c \in (a, b)$ such that $f'(c) = \frac{f(b)-f(a)}{b-a}$,
  - or equivalently, $f(b) = f(a) + f'(c)(b - a)$.
- Extended mean-value theorem
  - Let $f : [a, b] \to \mathbf{R}$. If f and f' are continuous in $[a, b]$ and differentiable in (a, b). Then, $\exists c \in (a, b)$ such that $f(b) = f(a) + f'(c)(b - a) + \frac{1}{2} f''(c)(b - a)^2$.
- Rolle's theorem
  - Let $f : [a, b] \to \mathbf{R}$ be continuous in $[a, b]$ and differentiable in (a, b). Suppose $f(a) = f(b)$. Then, $\exists c \in (a, b)$ such that $f'(c) = 0$.

Measuring the error (2)

Taylor's theorem

- The measurement error associated to the Taylor's polynomial of degree n is: $E_n(x) = f(x) - P_n(x)$
- Taylor's theorem provides an estimation for this error function $E_n(x)$.
- The basic content of the theorem is that the error is determined by the distance between x and c and by the $(n+1)^{st}$ derivative of f. Formally,
- Let f be $(n+1)$-times differentiable.
- Let $P_n(x)$ be the Taylor polynomial of degree n of f around c.
- Then for any value $x \neq c$, $\exists b \in (c, x)$ such that $f(x) = P_n(x) + \frac{1}{(n+1)!} f^{(n+1)}(b)(x-c)^{n+1}$ where the last term is called the error term of the approximation, $R_{n+1}(x)$.

Linear approximation and differential of f (2)



Measuring the error (3)

Taylor's theorem (cont'd)

- Remark 1: For $n = 0$ the theorem reduces to the Mean-value theorem.
- Remark 2: an equivalent way of stating the theorem is:
  - Let $M \leq |f^{(n+1)}(x)|$ on a neighborhood of c.
  - Then, for any x, the error of the Taylor approximation is bounded as $|f(x) - P_n(x)| \leq \frac{1}{(n+1)!} M |x - c|^{n+1}$
- Remark 3: If $f^{(n+1)}(x) = 0$, then $R_{n+1}(x) = 0$. It means that f is a polynomial of degree n. Therefore, the Taylor approximation of degree n is exact.

Taylor's theorem - Proof

Step 1. A Lemma

- Let f be $(n+1)$-times differentiable.
- Suppose that $f(c) = f'(c) = f''(c) = \cdots = f^{(n)}(c) = 0$
- Suppose that $\exists x \neq c$ such that $f(x) = 0$.
- Then, $\exists b \in (c, x)$ such that $f^{(n+1)}(b) = 0$.

Proof

- As $f(c) = 0, f(x) = 0$ Rolle's thm $\exists b_1 \in (c, x)$ s.t. $f'(b_1) = 0$
- As $f'(c) = 0, f'(b_1) = 0$ Rolle's thm $\exists b_2 \in (c, b_1)$ s.t. $f''(b_2) = 0$
- Iterate argument to generate sequence $b_1, b_2, \ldots, b_n$
- Eventually, we will find $b_n \in (c, x)$ s.t. $f^{(n+1)} = 0$
- Select $b_n = b$ as the desired value of b.

Taylor's theorem - Proof (2)

Step 2

- Let $P_n(x)$ be the degree n Taylor approx at c.
- Define $g(x) = f(x) - P_n(x)$ (error at $x \neq c$).
- Then, $g(c) = g'(c) = g''(c) = \ldots, g^{(n)}(c) = 0$
- Define $k = -\frac{g(x)}{(x-c)^{(n+1)}}$ or $g(x) = -k(x-c)^{(n+1)}$ $\quad [\alpha]$
- Define $h(x) = g(x) + k(x-c)^{(n+1)}$.
- Then, $h(c) = h'(c) = h''(c) = \ldots, h^{(n)}(c) = 0$ and $h(x) = 0$.
- Lemma$\rightarrow$ $\exists b \in (c, y)$ s.t. $h^{(n+1)}(b) = 0$
- Observe that $h^{(n+1)}(x) = g^{(n+1)}(x) + k(n+1)!$
- Also, $g^{(n+1)}(x) = f^{(n+1)}(x)$ (as $P_n(x)$ has degree n)
- Thus, $h^{(n+1)}(x) = f^{(n+1)}(x) + k(n+1)!$

Taylor's theorem - Proof (3)

Step 2 (cont'd)

- At $x = b$, using lemma $h^{(n+1)}(b) = f^{(n+1)}(b) + k(n+1)! = 0$
- Thus, $k = -\frac{f^{(n+1)}(b)}{(n+1)!} \quad [\beta]$
- Combining $[\alpha]$ and $[\beta]$ it follows

  $\frac{g(x)}{(x-c)^{(n+1)}} = \frac{f^{(n+1)}(b)}{(n+1)!}$

  $g(x) = \frac{f^{(n+1)}(b)}{(n+1)!}(x-c)^{(n+1)}$

  $f(x) - P_n(x) = \frac{f^{(n+1)}(b)}{(n+1)!}(x-c)^{(n+1)}$

  $f(x) = P_n(x) + \frac{f^{(n+1)}(b)}{(n+1)!}(x-c)^{(n+1)}$

  and this is Taylor's theorem.

Linear approximation and inverse function

Intuition

- Let $A \subset \mathbf{R}^n$ be an open set.
- Consider $x_0 \in A$ and $f : A \to \mathbf{R}^n$ be of class C^1.
- A linear approximation to f around x_0 is defined as the sum of $f(x_0)$ and a linear function $Jf(x_0)$.
- If $Jf(x_0)$ is invertible (i.e. $detJf(x_0) \neq 0$), then we may hope that f will be invertible as well around x_0.
- Note that f being invertible is a local property defined around a point $x_0 \in A$.
- The inverse function theorem is useful because it asserts whether there are solutions to equations and explains how to differentiate the solutions, although it may be impossible to solve the equations explicitly.

Linear approximation and inverse function (2)

Theorem

- Let $A \subset \mathbf{R}^n$ be an open set.
- Consider $x_0 \in A$ and $f : A \to \mathbf{R}^n$ be of class C^1.
- Suppose $det Jf(x_0) \neq 0$.
- Then, $\exists U = B(x_0, r) \subset A$ and $\exists V = B(f(x_0), s)$ open, such that $f(U) = V$ and f has a C^1 inverse $f^{-1} : V \to U$.
- Moreover, for $y \in V, x = f^{-1}(y)$, we have $Jf^{-1}(y) = [Jf(x)]^{-1}$.
- If f is of class C^p, so is f^{-1}.