

Taylor Expansions and (log)linearizing

Stéphane Dupraz

1 Mean Value Thoerem

Theorem 1.1. Mean Value Theorem in \mathbb{R}

Let $f : [a, b] \rightarrow \mathbb{R}$.

If f is continuous on $[a, b]$ and differentiable on (a, b) , then there exists $z \in (a, b)$ such that:

$$f(b) - f(a) = f'(z)(b - a)$$

Proof. The proof is in two steps. First we prove the result—called **Rolle's theorem**—in the particular case when $f(b) = f(a)$. In this case we want to find z such that $f'(z) = 0$: a critical point. Since f is continuous on $[a, b]$ compact, f has a maximum and a minimum on $[a, b]$ by Weierstrass theorem. To conclude, we just need to make sure that there is one maximum or minimum on the interior (a, b) . There are three cases. First, if f is constant over $[a, b]$ then $f' = 0$ and any z will do. Second, if there exists some x such that $f(x) > f(a)$, then there is an interior maximum. Finally, if there exists some x such that $f(x) < f(a)$, then there is an interior minimum.

For the general case, the trick is to define the function $g(t) = (f(b) - f(a))t - (b - a)f(t)$. It satisfies $g(b) = g(a)$, so we can apply Rolle's theorem to g : there exists $z \in (a, b)$ such that $g'(z) = f(b) - f(a) - (b - a)f'(z) = 0$. QED. □

The mean value theorem has a straightforward extension to function from \mathbb{R}^n to \mathbb{R} .

Theorem 1.2. Mean Value Theorem in \mathbb{R}^n

Let $f : S \rightarrow \mathbb{R}$, S an open subset of \mathbb{R}^n .

If f is differentiable, then for any $x, y \in S$ such that $[x, y] \in S$, there exists $z \in (x, y)$ (that is $z =$

$\lambda x + (1 - \lambda)y$ for some $\lambda \in (0, 1)$ such that:

$$f(x) - f(y) = f'(z)(x - y)$$

Proof. Just define $g(\lambda) = f(\lambda x + (1 - \lambda)y)$ for $\lambda \in [0, 1]$ and apply the mean value theorem for function from \mathbb{R} to \mathbb{R} . □

2 Taylor Expansions

We start with k^{th} -order Taylor expansions for functions from \mathbb{R} to \mathbb{R} , then consider first and second-order Taylor expansions for functions from \mathbb{R}^n to \mathbb{R} .

2.1 Taylor formula for functions from \mathbb{R} to \mathbb{R}

We have seen that we can look at the derivative of a function f at a point x as providing the best affine approximation of f around x . We say that we **linearize** the function f around x . A Taylor expansion of order k of f at x generalizes this notion by looking at the best linear approximation of f around x by a polynomial of order k . There exist variants of the Taylor theorem, depending on the precise meaning of “best approximation” (some require f to be $k + 1$ times differentiable, other only k times differentiable). Here we state the Taylor-Lagrange theorem, whose proof relies on the mean value theorem.

Theorem 2.1. *Taylor-Lagrange theorem for functions from \mathbb{R} to \mathbb{R}*

Let $f : [a, b] \rightarrow \mathbb{R}$.

If f is $k + 1$ times differentiable on $[a, b]$, then for all $x, y \in [a, b]$, there exists $z \in (x, y)$ such that:

$$f(y) = \left(f(x) + \sum_{i=1}^k \frac{f^{(i)}(x)}{i!} (y-x)^i \right) + \frac{f^{(k+1)}(z)}{(k+1)!} (y-x)^{k+1}$$

We call the term $f(x) + \sum_{i=1}^k \frac{f^{(i)}(x)}{i!} (y-x)^i$ the Taylor expansion of order k of f at x .

We often note the result less rigorously as:

$$f(y) \simeq f(x) + \sum_{i=1}^k \frac{f^{(i)}(x)}{i!} (y-x)^i$$

Proof. The proof is by induction on k . The base case $k = 0$ is the mean value theorem. The induction step relies on the mean-value theorem. We do not go into the details here. \square

2.2 Second-order Taylor approximations for functions from \mathbb{R}^n to \mathbb{R}

For functions of n variables, we restrict to first and second-order Taylor expansions. We have already seen first-order Taylor expansions—differentiation/linearization. Stated not too rigorously, the second-order Taylor

expansion of a twice differentiable function from \mathbb{R}^n to \mathbb{R} is

$$\begin{aligned} f(y) &\simeq f(x) + f'(x)(y - x) + \frac{1}{2}(y - x)' f''(x)(y - x) \\ &\simeq f(x) + \sum_{i=1}^n f'_i(x)(y_i - x_i) + \frac{1}{2} \sum_{i=1}^n \sum_{j=1}^n f''_{ij}(x)(y_i - x_i)(y_j - x_j) \end{aligned}$$

The main application of first and second-order Taylor approximations you will see will be in solving **Dynamic Stochastic General Equilibrium (DSGE)** models in macroeconomics. The **perturbations method** solves numerically for the equilibrium of an economy defined by a set of difference equations by first taking a first-order or second-order approximation to the equations around the steady-state of the system.

3 Linearizing

We now take a more applied turn on first-order approximations: linearization.

3.1 Linearizing a function

As we have seen, to get the first-order Taylor expansion of a function f at a point x^* —to **linearize the function** f at x^* —we simply need to calculate its derivative at x^* , $f'(x^*)$. Then the approximation is $f(x) = f(x^*) + f'(x^*)(x - x^*)$. Defining $dx \equiv x - x^*$ and $df(x) \equiv f(x) - f(x^*)$, we can note this as:

$$df(x) = f'(x^*)dx = f'_1(x^*)dx_1 + f'_2(x^*)dx_2 + \dots + f'_n(x^*)dx_n$$

and see it as relating the deviation of $f(x)$ from $f(x^*)$ to the deviation of x from x^* .

This is most useful to **linearize equations**. Consider an equation of the form:

$$f(x_1, \dots, x_n) = 0$$

We can differentiate both sides at x^* to get:

$$f'(x^*)dx = f'_1(x^*)dx_1 + f'_2(x^*)dx_2 + \dots + f'_n(x^*)dx_n = 0$$

This turns the equation that relates the variables x_1, \dots, x_n into an linear equation that relates the deviations of each variables dx_1, \dots, dx_n .

3.2 Practical rules

The rules of differentiation can be easily expressed in those terms. The linearity of the differentiation operator and the chain rule translate into:

1. $d(x + y) = dx + dy$.
2. $d(\lambda x) = \lambda dx$.
3. $df(g(x)) = f'(g(x^*))dg(x) = f'(g(x^*))g'(x^*)dx$.

For example, to linearize a resource constraint:

$$f(K_t) = C_t + K_{t+1} - (1 - \delta)K_t$$

around $C_t = C^*$ and $K_{t+1} = K_t = K^*$:

$$f'(K^*)dK_t = dC_t + dK_{t+1} - (1 - \delta)dK_t$$

4 Loglinearizing

In economics, it is often more relevant to focus on **relative deviations** (or percentage deviations) rather than on **absolute deviations** (or level deviations). For instance, knowing that “consumption is 5% below its steady-state value” is more meaningful than knowing that “consumption is \$100 below its steady-state value”, the meaning of which depends on how high steady-state consumption is. We are more interested in the relative deviation variables, that we will denote with a hat:

$$\hat{x} = \frac{dx}{x^*} = \frac{x - x^*}{x^*}$$

But it turns out that at first order¹:

$$\hat{x} = \frac{dx}{x^*} = dx \times \ln'(x^*) = d \ln(x)$$

where the last term means (as for any function): $d \ln(x) = \ln(x) - \ln(x^*)$. For this reason, turning a function—a relationship between inputs x_1, \dots, x_n and output $f(x)$ —into a linear function of the relative variation in x , $\hat{x}_1, \dots, \hat{x}_n$, and the relative variation in y , \hat{y} , is called **loglinearizing the function**. (Turning an equation—a relationship between variables x_1, \dots, x_n —into a linear equation that relates the relative variations of each variables $\hat{x}_1, \dots, \hat{x}_n$, is called **loglinearizing the equation**).

4.1 Formally

Formally, to loglinearize a function $f(x)$ instead of linearizing it, we first do some changes of variables.

- Instead of the variation dx_i , we want the variation $d \ln(x_i)$, so we define $y_i = \ln(x_i)$.
- Instead of the variation $df(x)$, we want the variation $d \ln(f(x))$, so we consider the function $\ln(f(x))$.

This way, we consider the function:

$$g(y) = \ln(f(e^{y_1}, \dots, e^{y_n}))$$

¹You can also see it through : $d \ln(x) = \ln(x) - \ln(x^*) = \ln\left(\frac{x}{x^*}\right) = \ln\left(\frac{x-x^*}{x^*} + 1\right) \simeq \frac{x-x^*}{x^*} \equiv \hat{x}$.

Linearizing:

$$dg(y) = \sum_{i=1}^n \frac{f'_i(e^{y_1^*}, \dots, e^{y_n^*})}{f(e^{y_1^*}, \dots, e^{y_n^*})} e^{y_i^*} dy_i$$

$$d \ln(f(x)) = \sum_{i=1}^n \frac{f'_i(x^*)}{f(x^*)} x_i^* d \ln(x_i)$$

Definition 4.1. The *elasticity* of a function f at x^* with respect to x_i is:

$$\varepsilon_f^i(x^*) = \frac{f'_i(x^*)}{f(x^*)} x_i^* = \frac{\partial \ln(f(x^*))}{\partial \ln(x_i)}$$

It expresses the percentage change of $f(x)$ for a percentage increase in x_i of 1% (at x^*).

We finally get:

$$\widehat{f(x)} = \varepsilon_1(x^*) \hat{x}_1 + \dots + \varepsilon_n(x^*) \hat{x}_n$$

In practice, we can rely on two approaches to loglinearize a function or an equation.

4.2 First Approach and Practical Rules

In practice, we can use the following method:

Approach 1. To loglinearize a function f :

1. Take the log of f .
2. Linearize $\ln(f)$, remembering $d \ln(x) = \hat{x}$.

For instance let us loglinearize a Cobb-Douglas production function $Y = AK^\alpha L^{1-\alpha}$ at any point:

$$\ln(Y) = \ln(A) + \alpha \ln(K) + (1 - \alpha) \ln(L)$$

$$d \ln(Y) = d \ln(A) + \alpha d \ln(K) + (1 - \alpha) d \ln(L)$$

$$\hat{Y} = \hat{A} + \alpha \hat{K} + (1 - \alpha) \hat{L}$$

This example is straightforward because of the following practical rules for loglinearization, which are simply the translation of the corresponding rules for linearization:

1. $\widehat{xy} = \widehat{x} + \widehat{y}$
2. $\widehat{x^\alpha} = \alpha \widehat{x}$
3. $\widehat{f(g(x))} = \varepsilon_f(g(x^*)) \varepsilon_g(x^*) \widehat{x}$.

Power functions have constant elasticities; they are to loglinearization what linear functions, which have a constant derivative, are to linearization.

4.3 Second approach to deal with sums

The previous rules imply that it is straightforward to loglinearize a function that contains only products and power functions using the first approach. However, with the first approach it is not so easy to deal with sums. In contrast, it was easy to deal with sums when linearizing. Hence, when a function f includes sums, it is easier to loglinearize it using the alternative method:

Approach 2. *To loglinearize a function f :*

1. *Linearize f .*
2. *Replace absolute deviations dx by $x^* \widehat{x}$.*

For instance, to loglinearize the resources constraint above:

$$\begin{aligned} f'(K^*)dK_t &= dC_t + dK_{t+1} - (1 - \delta)dK_t \\ f'(K^*)K^* \widehat{K}_t &= C^* \widehat{C}_t + K^* \widehat{K}_{t+1} - (1 - \delta)K^* \widehat{K}_t \\ f'(K^*) \widehat{K}_t &= \frac{C^*}{K^*} \widehat{C}_t + \widehat{K}_{t+1} - (1 - \delta) \widehat{K}_t \end{aligned}$$

4.4 Mixing the two approaches

Which of the two approaches to use in practice? We have seen that when a function involves products and power functions (for instance the Cobb-Douglas production function), taking the log and linearizing is the quickest way to go. But when the function is a sum (for instance a resource or budget constraint), then it is easier to linearize and replace $dx = x^* \widehat{x}$. Now most functions are sums of products or products of sums. Consider for instance the following resource constraint, where the production function is Cobb-Douglas:

$$AK^\alpha L^{1-\alpha} = C + I$$

It is a sum, but with one of its terms (on the LHS) being itself a product. Start by seeing it as a sum, as it is a sum, so linearize, remembering $dx = x^* \hat{x}$:

$$(A^* K^{*\alpha} L^{*1-\alpha}) \widehat{AK^\alpha L^{1-\alpha}} = C^* \widehat{C} + I^* \widehat{I}$$

Then treat the product $A_t \widehat{K_t^\alpha L_t^{1-\alpha}}$ with the rule on products and power functions to get:

$$(A^* K^{*\alpha} L^{*1-\alpha})(\widehat{A} + \alpha \widehat{K} + (1-\alpha)\widehat{L}) = C^* \widehat{C} + I^* \widehat{I}$$

$$\widehat{A} + \alpha \widehat{K} + (1-\alpha)\widehat{L} = \left(\frac{C^*}{A^* K^{*\alpha} L^{*1-\alpha}} \right) \widehat{C} + \left(\frac{I^*}{A^* K^{*\alpha} L^{*1-\alpha}} \right) \widehat{I}$$

Alternatively, you can face a product some of the factors of which are sums. Then treat it as a product (take the log and linearize), and then use the rules on sums to treat each factor. Finally, if you are running into a function whose functional form is not made explicit, use the general formula with elasticities.

4.5 When loglinearizing equations, reduce the number of constants

When we are loglinearizing a function $f(x_1, \dots, x_n)$ at x^* , constants may appear that depend on x_1^*, \dots, x_n^* . If we are loglinearizing an equation $f(x_1, \dots, x_n) = 0$, those constants are linked through the equation at x^* , $f(x_1^*, \dots, x_n^*) = 0$. For instance, in the previous example, we have two constant that depend on x^* . However, they are linked through:

$$(A^* K^{*\alpha} L^{*1-\alpha}) = C^* + I^*$$

$$1 = \frac{C^*}{A^* K^{*\alpha} L^{*1-\alpha}} + \frac{I^*}{A^* K^{*\alpha} L^{*1-\alpha}}$$

Noting $s_C = \frac{C^*}{A^* K^{*\alpha} L^{*1-\alpha}}$ the share of consumption in GDP, the loglinearized equation becomes:

$$\widehat{A} + \alpha \widehat{K} + (1-\alpha)\widehat{L} = s_C \widehat{C} + (1-s_C)\widehat{I}$$

There is some arbitrariness in the choice of the constants to keep.

4.6 A remark on interest, inflation, ... rates

Just a remark to finish: many authors use a different definition of the hat variable for interest and inflation rates. What they note \widehat{r} , which could be taken as the relative deviation of the net interest rate, is actually the

deviation from trend of the gross interest rate, $\widehat{1+r}$. The connection between the two is :

$$\widehat{1+r} = \frac{d(1+r)}{1+r^*} = \frac{dr}{1+r^*} = \frac{r^*}{1+r^*} \widehat{r}$$

The advantage of doing so is that it allows to consider the case of a steady-state value of the net rate equal to zero, which might be of particular interest for the nominal interest rate or the inflation rate. Indeed, we cannot loglinearize the net interest rate around zero if its steady-state value is zero, but we can loglinearize the gross interest rate around 1. Note that in this case, the relative deviation of the net interest rate is the net interest rate itself:

$$\widehat{1+r_t} = dr_t = r_t$$