Higher order derivatives of the inverse function

Andrej Liptaj*

Abstract

A general recursive and limit formula for higher order derivatives of the inverse function is presented. The formula is next used in couple of mathematical applications: expansion of the inverse function into Taylor series, solving equations, constructing random numbers with a given distribution from uniformly distributed random numbers and expanding a function in the neighborhood of a given point in an alternative way to the Taylor expansion.

Introductory note: This text was previously published on Scribd[1].

1 Introduction

The first derivative of the inverse function is given by the well-known formula

$$\frac{d}{dy}g(y)|_{y=f(x_0)} = \frac{1}{\frac{d}{dx}f(x)|_{x=x_0}},\ g = f^{-1}$$

and higher order derivatives are usually derived by a repetitive differentiation of the identity $f(g(x)) = x$. In this text I would like to present an alternative expression that has recursive and limit form and write down the explicit forms of the first ten derivatives of the inverse function. Even though these expressions are simple to derive, they are not commonly available in the literature and it may be useful to summarize them in one text. In addition I will show a couple of related examples and schemes that, I believe, have an application potential.

2 Higher order derivatives of the inverse function

2.1 Notations

For the purpose of this article it is convenient to introduce new notations. This is mostly motivated by the fact that higher order derivatives and higher powers will occur often in the text and thus would make formulas difficult to read. Let us consider a real function of a real variable $f(x)$ on the interval I. Let us suppose that f is bijective on I. Let x_0 be an internal point of I and $g(y)$ be the inverse function of f, $g(f(x_1)) = g(y_1) = x_1$ for any x_1 belonging to I. Let us define $y_0 = f(x_0)$. In this article I will use the following notations:

$$fn = \frac{d^n}{dx^n}f(x)|_{x=x_0}, \quad gn = \frac{d^n}{dy^n}g(y)|_{y=y_0}$$

Described objects are depicted on the Figure 1.

*Institute of Physics, Bratislava, Slovak Academy of Sciences, andrej.liptaj@savba.sk
I am willing to publish any of my ideas presented through free-publishing services in a journal, if someone (an editor) judges them interesting enough. Journals in the "*Current Contents*" database are strongly preferred.

[1] *https://www.scribd.com/document/113474210/Higher-order-derivatives-of-the-inverse-function-version-2*
http://www.scribd.com/doc/13699758/Higher-order-derivatives-of-the-inverse-function

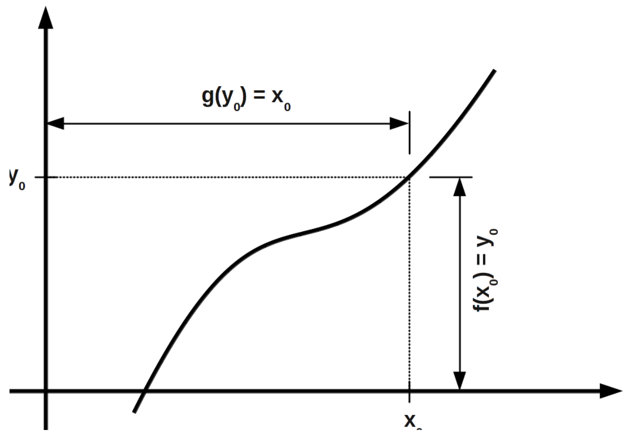

Figure 1: $f(x)$ and $g(y)$.

2.2 Formulas

The formula for higher order derivatives of the inverse function can be written in a recursive and limit form as follows

$$g1 = \frac{1}{f1}, \quad gn = \lim_{\Delta x \to 0} n! \frac{\Delta x - \Sigma_{i=1}^{n-1} \frac{1}{i!}(gi)\left[\Sigma_{j=1}^{n} \frac{1}{j!}(fj)(\Delta x)^j\right]^i}{\left[\Sigma_{i=1}^{n} \frac{1}{i!}(fi)(\Delta x)^i\right]^n} \equiv F[f1, \ldots, fn], \tag{1}$$

where the notation $F[f1, \ldots, fn]$ is introduced to refer in a simplified form to this recursive limit expression later in the text. This gives us for the first five derivatives

$$g1 = \frac{1}{f1}, \quad g2 = -\frac{f2}{(f1)^3}, \quad g3 = \frac{3(f2)^2 - (f1)(f3)}{(f1)^5}, \quad g4 = -\frac{15(f2)^3 - 10(f1)(f2)(f3) + (f1)^2(f4)}{(f1)^7},$$

$$g5 = \frac{105(f2)^4 - 105(f1)(f2)^2(f3) + 10(f1)^2(f3)^2 + 15(f1)^2(f2)(f4) - (f1)^3(f5)}{(f1)^9}. \tag{2}$$

The expressions for further derivatives are summarized in the appendix A.

2.3 Illustrative example

Let us define

$$f(x) = \ln(x), \ x_0 = 1.$$

First three derivatives are

$$f1 = \frac{d}{dx} f(x)|_{x=1} = 1, \quad f2 = \frac{d^2}{dx^2} f(x)|_{x=1} = -1, \quad f3 = \frac{d^3}{dx^3} f(x)|_{x=1} = 2.$$

The formulas 2 yield

$$g1 = \frac{1}{f1} = 1, \quad g2 = -\frac{f2}{(f1)^3} = 1, \quad g3 = \frac{3(f2)^2 - (f1)(f3)}{(f1)^5} = 1.$$

One recognizes the derivatives of the exponential function $g(y) = f^{-1} = e^y$ in the point $y_0 = \ln(1) = 0$.

2.4 Comments

- The formulas 2 contain only basic arithmetical operations and from this point of view they are suitable for computer calculations. The disadvantage is the rapid growth in the number of terms with increasing order of the derivative ($g10$ has already 30 additive terms).

- Since inverting a function twice leads to the initial function, one expects this property to be propagated into the formula 1 ("duality"). This needs to be true (if the formula is true), however it is still interesting to check it explicitly. Let h be the inverse function of g and, although identical, let us think of h and f separately. Then, for example, for the second derivative one has

$$h2 = -\frac{g2}{(g1)^3} = -\frac{-\frac{f2}{(f1)^3}}{(\frac{1}{f1})^3} = -\frac{-(f2)(f1)^3}{(f1)^3} = f2.$$

This property was checked on computer to by explicitly valid up to the order eight.

2.5 Derivation of the higher order derivatives formula for the inverse function

In this section the same notations are used as in the previous one. In the neighborhood of x_0 one can expand the functions f and g

$$f(x_0 + \Delta x) \approx f(x_0) + a_1\Delta x + a_2\Delta x^2 + \ldots + a_n\Delta x^n = f(x_0) + \sum_{j=1}^{n} a_j \Delta x^j,$$

$$g(y_0 + \Delta y) \approx g(y_0) + b_1\Delta y + b_2\Delta y^2 + \ldots + b_n\Delta y^n = g(y_0) + \sum_{i=1}^{n} b_i \Delta y^i.$$

Then using the inverse-function property in the point $x_0 + \Delta x$ one has for small Δx

$$\begin{aligned} x_0 + \Delta x &= g[f(x_0 + \Delta x)] \\ &\approx g[f(x_0) + \textstyle\sum_{j=1}^n a_j \Delta x^j] \quad \text{with} \quad \textstyle\sum_{j=1}^n a_j \Delta x^j = \Delta y \\ &\approx x_0 + \textstyle\sum_{i=1}^n b_i [\sum_{j=1}^n a_j \Delta x^j]^i \\ &\approx x_0 + \textstyle\sum_{i=1}^{n-1} b_i [\sum_{j=1}^n a_j \Delta x^j]^i + b_n [\sum_{j=1}^n a_j \Delta x^j]^n. \end{aligned}$$

Taking into account that our calculations are valid only for small Δx by applying the limit $\Delta x \to 0$ and using the relation between the polynomial coefficients and its derivatives $a_j = \frac{fj}{j!}$, $b_i = \frac{gi}{i!}$ one arrives to the expression

$$gn = \lim_{\Delta x \to 0} n! \frac{\Delta x - \sum_{i=1}^{n-1} \frac{gi}{i!} \left[\sum_{j=1}^n \frac{fj}{j!} \Delta x^j\right]^i}{[\sum_{j=1}^n \frac{fj}{j!} \Delta x^j]^n}$$

which is identical to the formula 1. A rigorous mathematical proof would require careful treatment of approximations and limits, which I am not doing here.

3 Applications

3.1 Inverse function Taylor series

The knowledge of the derivatives of the inverse function enables us to expand the inverse function into Taylor series. Knowing the initial function f, let in addition suppose that we know also the value of its inverse function g at some point y_0, $g(y_0) = x_0$. Calculating the numbers $g1, g2, g3\ldots$ from the formula 1, one can express the value of g at some different point y_1

$$\begin{aligned} g(y_1) &= g(y_0) + (g1)(y_1 - y_0) + \frac{1}{2!}(g2)(y_1 - y_0)^2 + \ldots \qquad (3) \\ &= x_0 + \sum_{i=1}^{\infty} \frac{1}{i!}(gi)(y_1 - y_0)^i. \end{aligned}$$

This of course holds only on condition that Taylor series for g converge to g at y_1.

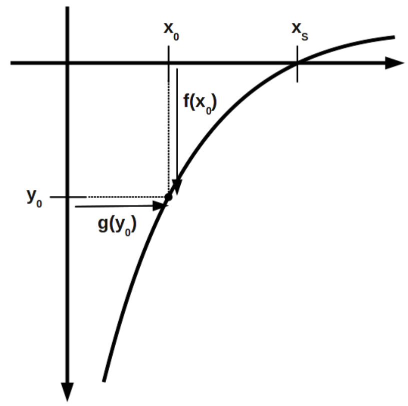

Figure 2: Solving equations

3.2 Solving equations

The knowledge of the inverse function (supposing it can be approximated by its Taylor series) allows for solving real equations of the real variable. Equation $f(x_s) = 0$ can be solved immediately doing $x_s = g[f(x_s)] = g(0)$. A complication comes from the fact that we *a priori* do not know the value of g at some given point. Consider the situation depicted in the Figure 2. The value of $g(0)$ can be expressed through $g(y_0)$

$$x_s = g(0) = g(y_0) + \sum_{i=1}^{\infty} \frac{1}{i!}(gi)(-y_0)^i.$$

However, one is not free to choose y_0 arbitrarily (otherwise one would directly choose $y_0 = 0$). A possible solution is to choose an arbitrary x_0 in some interval around x_s, where f is bijective. Then using $y_0 = f(x_0)$ one rewrites the previous equation in the form

$$x_s = g(0) = x_0 + \sum_{i=1}^{\infty} \frac{1}{i!}(gi)[-f(x_0)]^i, \tag{4}$$

or once more in the form

$$x_s = g(0) = x_0 + \sum_{i=1}^{\infty} \frac{1}{i!} F[f1, \ldots, fi][-f(x_0)]^i,$$

where it can explicitly be seen, that all information needed to solve the equation $f(x_s) = 0$ is contained in the (known) function f itself. This approach can be demonstrated on couple of simple examples that follow.

Solving transcendent equation

Consider the equation $x^x = 7$. We need to transform it into the form $f(x) = x^x - 7 = 0$. Next we need to have some reasonable guess for the solution. Let the guess be $x_0 = 2$ $[f(x_0) = -3]$. Having x_0, we calculate the derivatives $fn = \frac{d^n}{dx^n}(x^x - 7)|_{x=x_0=2}$ up to the desired order and then using the expression 1 calculate $g1$, $g2$, $g2, \ldots$ - the derivatives of the inverse function $g(y)$ at the corresponding point $y_0 = -3$. I have considered first 8 derivatives and calculations based on the formula 4 lead to $x_s \approx 2.309117\ldots$ giving $2.309117\ldots^{2.309117\ldots} = 6.90620\ldots$. The method thus gives nice results that can be improved if the whole procedure is used iteratively. If in the same situation one considers four derivatives only but repeats the whole procedure twice (taking the approximate solution from the first iteration as the starting point for the second iteration) one obtains numerically precise result $x_s \approx 2.31645\ldots$ giving $2.31645\ldots^{2.31645\ldots} = 6.999999205\ldots$.

Calculation of mathematical constants π and e

The previous approach can be used to calculate approximate values of important mathematical constants like π or e. For that purpose one needs to formulate equations that have these constants for solutions. In the following I will also consider two scenarios: a single iteration with 8 terms taken into account and two iterations with 4 derivatives considered.

For π a suitable equation might be $\cos\left(\frac{x}{4}\right) = \frac{\sqrt{2}}{2}$. One then defines $f(x) = \cos\left(\frac{x}{4}\right) - \frac{\sqrt{2}}{2}$ and applies the method of the inverse function to solve $f(x) = 0$. If we choose $x_0 = 3$, we get in the first scenario

$$\pi \approx 3.14159265357636280318084094735\ldots,$$

where 10 decimal places are correct. In the second case one obtains

$$\pi \approx 3.14159265358979323846269213732\ldots,$$

where 22 decimal digits are correct.

For e the equation $\ln(x) = 1$ might be consider, from which one defines $f(x) = \ln(x) - 1$ and solves $f(x) = 0$. Taking $x_0 = 2$ gives in the first scenario

$$e \approx 2.71828182832191473183292924148\ldots\ ,$$

where 9 decimal places are correct and in the second scenario

$$e \approx 2.71828182845904444095135772192\ldots\ ,$$

where 14 decimal digits are correct.

3.3 Generating random numbers

The aim of this section is to obtain random numbers which are distributed respecting a probability density function $\rho(x)$, given a random number generator which provides uniformly distributed random numbers from the interval $\langle 0, 1)$. I think this kind of problem might appear often in practice, since computers usually offer only uniformly distributed numbers from the interval $\langle 0, 1)$ and few other common distributions. The algorithm to solve the task is:

- construct the distribution function $f(x) = \int_{-\infty}^{x} \rho(y)\, dy$.
- construct the function g inverse to f: $g[f(x)] = f[g(x)] = x$.
- if the numbers $x_1, x_2, \ldots, x_n$ from the interval $\langle 0, 1)$ respect the uniform distribution then the numbers $y_1 = g(x_1), y_2 = g(x_2), \ldots, y_n = g(x_n)$ respect the distribution $\rho(x)$.

To arrive to this solution one can apply the methods presented in this article, however only on the condition that the value of the distribution function f is known at least in one point x_0 (this is often the case - for example for symmetric distributions $f(x_0 = 0) = \frac{1}{2}$). If so, one constructs the Taylor series for g using the formula 3.

The example I am going to present is $\rho(x) = \frac{1}{\sqrt{2\pi}} \exp(-\frac{x^2}{2})$. The corresponding distribution function $f(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} \exp(-\frac{y^2}{2})\, dy$ is not an elementary function, but its value and the value of all its derivatives fn at the point $x_0 = 0$ are known. Using the formula 1 the derivatives gn of the inverse function g at the point $y_0 = f(x_0 = 0) = \frac{1}{2}$ are calculated. Taking into account first 9 derivatives and using the expression 3 we get (x is used for the variable name rather than y)

$$g(x) \approx 0 + (2.506628)(x - 0.5) + 0 + \frac{1}{3!}(15.749609)(x - 0.5)^3 + \ldots$$

$$\ldots + \frac{1}{9!}(17068346.560783)(x - 0.5)^9.$$

The graph of this polynomial is plotted in the Figure 3. The result is not fully satisfactory because the generated numbers are always between approximately -2 and 2. The distribution coming out of our procedure is in the Figure 4 compared with the "true" normal distribution. One sees that a polynomial with higher number of terms would be needed to approximate the normal distribution with better accuracy, the presented example however well demonstrates one possible application of the formula 1 and the related procedure.

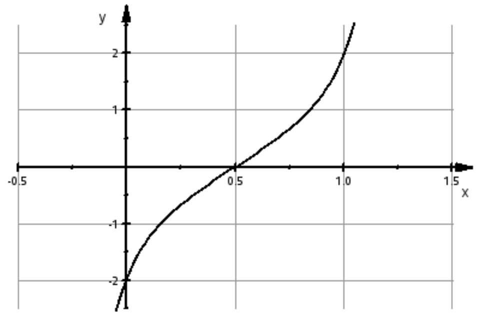

Figure 3: The approximation of the function g, $g = f^{-1}$, $f(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} \exp(-\frac{y^2}{2})\, dy$.

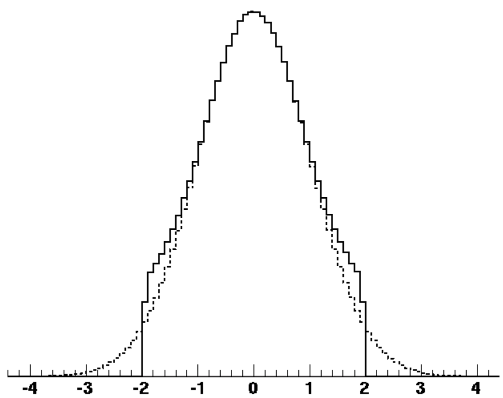

Figure 4: Plots produced with a high statistics of random numbers. Dashed line - normal distribution; Solid line - our approximation of the normal distribution

3.4 An alternative function expansion

The ideas presented in this section suppose that a numerical procedure for inverting a function (flip function graph around the axis $y = x$) is available. If so, one has an alternative method of approximating a function in a neighborhood of a point x_0: one expands the inverse function into Taylor series and then inverts numerically the result. In this way one gets a non-polynomial approximation (inverse of a polynomial is not, in general, a polynomial) of the initial function. In some cases this approximation has bigger interval of convergence than the Taylor series of the initial function and thus I consider it as a possible interesting application of the higher order derivatives of the inverse function.

I will demonstrate it on a simple example that was already presented in the section 2.3: $f(x) = \ln(x)$, $x_0 = 1$. The Taylor series for $\ln(x)$ at $x_0 = 1$ has the radius of convergence $R = 1$ and so one cannot use it to calculate function values beyond $x = 2$. If, however, one expands into Taylor series the inverse function, i.e. the function $\exp(y)$ at $y_0 = f(x_0 = 1) = 0$ one gets a series that converges for every $y \epsilon \mathbb{R}$. And so, numerically inverting the result, one ends up with a prediction for the value of $\ln(x)$ for every $x \epsilon \mathbb{R}$.

4 Conclusion

I presented the general formula that has a recursive and limit form for calculating higher order derivatives of the inverse function together with couple of examples and possible applications. More details on the calculations shown in this text can be found on the web page *http://www.scribd.com/doc/13699758/Higher-order-derivatives-of-the-inverse-function.*

A List of expressions

The list of the formulas for higher order derivatives of the inverse function up to the order ten.

$g1 = \frac{1}{f1}$

$g2 = -\frac{f2}{(f1)^3}$

$g3 = \frac{1}{(f1)^5}\,[\,3(f2)^2 - (f1)(f3)\,]$

$g4 = \frac{1}{(f1)^7}\,[\,-15(f2)^3 + 10(f1)(f2)(f3) - (f1)^2(f4)\,]$

$g5 = \frac{1}{(f1)^9}\,[\,105(f2)^4 - 105(f1)(f2)^2(f3) + 10(f1)^2(f3)^2 + 15(f1)^2(f2)(f4) - (f1)^3(f5)\,]$

$g6 = \frac{1}{(f1)^{11}}\,[\,-945(f2)^5 + 1260(f1)(f2)^3(f3) - 280(f1)^2(f2)(f3)^2 - 210(f1)^2(f2)^2(f4)$
$+ 35(f1)^3(f3)(f4) + 21(f1)^3(f2)(f5) - (f1)^4(f6)\,]$

$g7 = \frac{1}{(f1)^{13}}\,[\,10395(f2)^6 - 17325(f1)(f2)^4(f3) + 6300(f1)^2(f2)^2(f3)^2 - 280(f1)^3(f3)^3$
$+ 3150(f1)^2(f2)^3(f4) - 1260(f1)^3(f2)(f3)(f4) + 35(f1)^4(f4)^2 - 378(f1)^3(f2)^2(f5)$
$+ 56(f1)^4(f3)(f5) + 28(f1)^4(f2)(f6) - (f1)^5(f7)\,]$

$g8 = \frac{1}{(f1)^{15}}\,[\,-135135(f2)^7 + 270270(f1)(f2)^5(f3) - 138600(f1)^2(f2)^3(f3)^2 + 15400(f1)^3(f2)(f3)^3$
$- 51975(f1)^2(f2)^4(f4) + 34650(f1)^3(f2)^2(f3)(f4) - 2100(f1)^4(f3)^2(f4) - 1575(f1)^4(f2)(f4)^2$
$+ 6930(f1)^3(f2)^3(f5) - 2520(f1)^4(f2)(f3)(f5) + 126(f1)^5(f4)(f5) - 630(f1)^4(f2)^2(f6)$
$+ 84(f1)^5(f3)(f6) + 36(f1)^5(f2)(f7) - (f1)^6(f8)\,]$

$g9 = \frac{1}{(f1)^{17}}\,[\,2027025(f2)^8 - 4729725(f1)(f2)^6(f3) + 3153150(f1)^2(f2)^4(f3)^2 - 600600(f1)^3(f2)^2(f3)^3$
$+ 15400(f1)^4(f3)^4 + 945945(f1)^2(f2)^5(f4) - 900900(f1)^3(f2)^3(f3)(f4) + 138600(f1)^4(f2)(f3)^2(f4)$
$+ 51975(f1)^4(f2)^2(f4)^2 - 5775(f1)^5(f3)(f4)^2 - 135135(f1)^3(f2)^4(f5) + 83160(f1)^4(f2)^2(f3)(f5)$
$- 4620(f1)^5(f3)^2(f5) - 6930(f1)^5(f2)(f4)(f5) + 126(f1)^6(f5)^2 + 13860(f1)^4(f2)^3(f6)$
$- 4620(f1)^5(f2)(f3)(f6) + 210(f1)^6(f4)(f6) - 990(f1)^5(f2)^2(f7) + 120(f1)^6(f3)(f7)$
$+ 45(f1)^6(f2)(f8) - (f1)^7(f9)\,]$

$g10 = \frac{1}{(f1)^{19}}\,[\,-34459425(f2)^9 + 91891800(f1)(f2)^7(f3) - 75675600(f1)^2(f2)^5(f3)^2$
$+ 21021000(f1)^3(f2)^3(f3)^3 - 1401400(f1)^4(f2)(f3)^4 - 18918900(f1)^2(f2)^6(f4)$
$+ 23648625(f1)^3(f2)^4(f3)(f4) - 6306300(f1)^4(f2)^2(f3)^2(f4) + 200200(f1)^5(f3)^3(f4)$
$- 1576575(f1)^4(f2)^3(f4)^2 + 450450(f1)^5(f2)(f3)(f4)^2 - 5775(f1)^6(f4)^3 + 2837835(f1)^3(f2)^5(f5)$
$- 2522520(f1)^4(f2)^3(f3)(f5) + 360360(f1)^5(f2)(f3)^2(f5) + 270270(f1)^5(f2)^2(f4)(f5)$
$- 27720(f1)^6(f3)(f4)(f5) - 8316(f1)^6(f2)(f5)^2 - 315315(f1)^4(f2)^4(f6) + 180180(f1)^5(f2)^2(f3)(f6)$
$- 9240(f1)^6(f3)^2(f6) - 13860(f1)^6(f2)(f4)(f6) + 462(f1)^7(f5)(f6) + 25740(f1)^5(f2)^3(f7)$
$- 7920(f1)^6(f2)(f3)(f7) + 330(f1)^7(f4)(f7) - 1485(f1)^6(f2)^2(f8) + 165(f1)^7(f3)(f8)$
$+ 55(f1)^7(f2)(f9) - (f1)^8(f10)\,]$