THE INVERSE FUNCTION THEOREM AND THE LEGENDRE TRANSFORM

TOBY LAM

If I give you a monotone, differentiable function $f(x)$ that has an inverse $f^{-1}(x)$, can you tell me the derivative and the antiderivative of the inverse function?

The first question is easy! We all know about the inverse function theorem. We just need to know $f^{-1}(x)$ and $f'(x)$.

The second question, however, might bring up bad memories of your past self trying to remember calculus identities. It turns out that you only need to know $f^{-1}(x)$ and $\int f(x)\,dx$.

We will try to answer these two questions geometrically and show how they are related to each other. We will begin by proving the inverse function theorem with a geometric flavour.

A geometric proof of the inverse function theorem

To remind ourselves, we start off with a formal statement of the inverse function theorem.

Theorem (Inverse function theorem)**.** *Given a continuously differentiable function $f:\mathbb{R}\to\mathbb{R}$ with $f'(a)\neq 0$ at some point a, there exists some interval I with a in its interior on which f has a continuously differentiable inverse f^{-1}, defined on $f(I)$, and with derivative*

$$\frac{d}{dx}f^{-1}(x)=\frac{1}{f'(f^{-1}(x))}\ \forall x\in I.$$

In high school, we proved the above algebraically by differentiating both sides of $f^{-1}(f(x))=x$. In a course in analysis, we would be even more rigorous.

We give the following geometric argument instead. We can first plot the function $y=f(x)$ on the Cartesian plane. Then we reflect the plot across the diagonal $y=x$. The reflected plot is exactly the plot of the inverse function. As all tangent lines of $f(x)$ are reflected over the diagonal, the slope of each tangent line is be multiplicatively inversed. This in essence the inverse function theorem, although we need to be careful keeping track of where each point gets reflected to.

Example. *See figure 1. It lies the plot of $y=e^x$ in red and the plot of $y=\ln x$ in blue. We see how we they are related to each other by a reflection across the green diagonal line, $y=x$. The tangent line of e^x at $(-1,1/e)$ is reflected across the diagonal to give the tangent line of $\ln x$ at $(1/e,-1)$. As such $\frac{d}{dx}\big|_{x=1/e}\ln x=1/\frac{d}{dx}\big|_{x=-1}e^x=e$.*

This suggests another way of thinking about the inverse function theorem. Recall that to plot a function $f(x)$, we are essentially parametrizing a curve on $\mathbb{R}^2$ by sending x to $(x,f(x))$. We can find the slope of the curve at $(x,f(x))$ by finding the slope of the velocity vector $(1,f'(x))$ to the curve, which is exactly $f'(x)$.

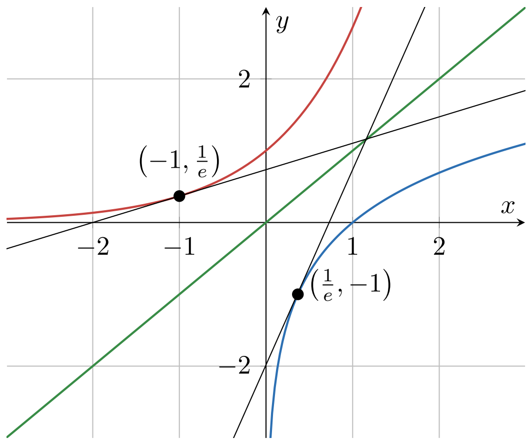

Figure 1.

Now, what if we parametrized the plot of the inverse function by sending x to $(f(x), x)$? The slope of the curve at $(f(x), x)$ will be the slope of the velocity vector $(f'(x), 1)$, which is $1/f'(x)$. This is exactly what the inverse function theorem says.

The Legendre transform

Now we try to answer the second question, what is the antiderivative of the inverse function? Similar to the inverse function theorem, we can do so algebraically and geometrically. While the algebraic method may not give us much insight, it is still worthwhile to mention. Almost all of the results in this section come from a 1905 paper by Laisant.

Allow me to pull something out of the hat. Let $F(x)$ be some antiderivative of $f(x)$. We can check that $xf^{-1}(x) - F(f^{-1}(x))$ is an antiderivative of $f^{-1}(x)$: By the product rule we have

$$\frac{d}{dx} x f^{-1}(x) = f^{-1}(x) + \frac{x}{f'(f^{-1}(x))}.$$

By the chain rule we have

$$\begin{aligned} \frac{d}{dx} F(f^{-1}(x)) &= \frac{f(f^{-1}(x))}{f'(f^{-1}(x))} \\ &= \frac{x}{f'(f^{-1}(x))}. \end{aligned}$$

Combining the two, we get

$$\frac{d}{dx}\left[x f^{-1}(x) - F(f^{-1}(x))\right] = f^{-1}(x).$$

For example, you can work out that $d/dx\,[x \ln(x) - x + C] = \ln x$ using the above. Unfortunately, it is not at all obvious where the formula comes from. One could reverse the logic by performing an integration by parts, although I am not so sure

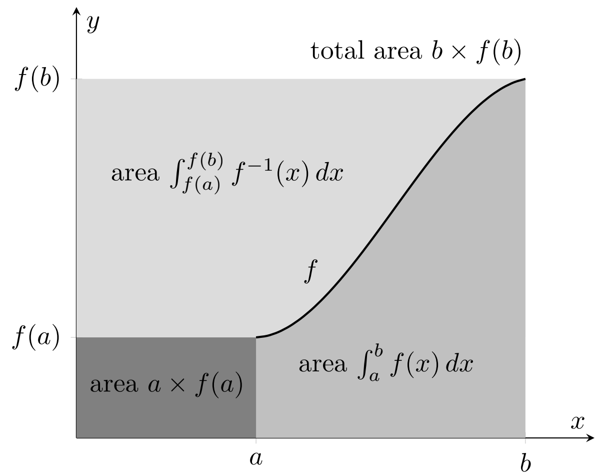

FIGURE 2.

if it helps that much!

$$\begin{aligned}
&\int f^{-1}(x)\,dx \\
&= xf^{-1}(x) - \int x \frac{d}{dx} f^{-1}(x)\,dx && \text{by parts} \\
&= xf^{-1}(x) - \int f(f^{-1}(x)) \frac{d}{dx} f^{-1}(x)\,dx \\
&= xf^{-1}(x) - F(f^{-1}(x)) + C && \text{by FTC}
\end{aligned}$$

for some constant C.

Perhaps the best proof is once again a geometric one. Laisant's formula tells us that if f is continuous and strictly increasing, then

$$\int_{f(a)}^{f(b)} f^{-1}(x)dx + \int_a^b f(x)dx = bf(b) - af(a).$$

A proof without words is presented in figure 2.

We can connect Laisant's formula to the algebraic approach. Further assume that f is differentiable. Informally, consider the points $(a, f(a))$ and $(a+\epsilon, f(a+\epsilon))$ for some $a \in \mathbb{R}$ and some small $\epsilon > 0$. Laisant's formula tells us that

$$\int_{f(a)}^{f(a+\epsilon)} f^{-1}(x)\,dx + \int_a^{a+\epsilon} f(x)\,dx = (a+\epsilon)f(a+\epsilon) - af(a)$$

so

$$G(f(a+\epsilon)) - G(f(a)) + F(a+\epsilon) - F(a) = (a+\epsilon)f(a+\epsilon) - af(a)$$

for some antiderivative $G(x)$ of $f^{-1}(x)$. Dividing both sides by ϵ and taking the limit $\epsilon \to 0$ we get

$$\left.\frac{d}{dx}\right|_{x=a} [G(f(x)) + F(x) - xf(x)] = 0.$$

As this is true for all $a \in \mathbb{R}$, we see that there is some constant C such that

$$G(f(x)) + F(x) - xf(x) = C.$$

Letting $\tilde{x} = f(x)$ and rearranging terms we see that

$$\tilde{G}(\tilde{x}) := \tilde{x}f^{-1}(\tilde{x}) - F(f^{-1}(\tilde{x}))$$

is an antiderivative of $f^{-1}(\tilde{x})$, which is what we expect.

Summary

In fact, taking $F(x)$ to $xf^{-1}(x) - F(f^{-1}(x))$ is called the Legendre transform, which is of immense significance in Lagrangian / Hamiltonian mechanics. Unfortunately, explaining this is beyond the scope of the article (and my knowledge). I encourage the reader to further ponder on this connection!

To summarise, we draw a commutative diagram

$$\begin{tikzcd}
F(x) \arrow[r, "\text{Legendre Transform}"] \arrow[d, "\frac{d}{dx}"] & G(y) \arrow[d, "\frac{d}{dy}"] \\
f(x) \arrow[r, "\text{Taking the inverse}"] \arrow[d, "\frac{d}{dx}"] & g(y) := f^{-1}(y) \arrow[d, "\frac{d}{dy}"] \\
f'(x) \arrow[r, "\text{Inverse Function Theorem}"] & g'(y)
\end{tikzcd}$$

with the following relations

$$\begin{aligned}
\text{IFT:} \quad & g'(y) = \frac{1}{f'(g(y))} \\
\text{Legendre:} \quad & G(y) = y \times g(y) - F(g(y)) + C
\end{aligned}$$

As such, the inverse function theorem is about what happens if you take the inverse then take the derivative. The Legendre transform is about what happens if you take the inverse then take the antiderivative.

References

Here is Laisant's original paper.

- C.-A. Laisant. "Intégration des fonctions inverses". In: *Nouvelles annales de mathématiques : journal des candidats aux écoles polytechnique et normale* 5 (1905), pp. 253-257. URL: `http://www.numdam.org/item/NAM_1905_4_5__253_0/`

Laisant's formula (and its curious connections with Young's inequality) is also exposed in Spivak's Calculus.

- Michael Spivak. *Calculus*. 4th ed. Publish or Perish, Inc., 2008. ISBN: 978-0-914098-91-1, Chapter 13, Problems 21 & 22, p. 276

Balliol College, University of Oxford
Email address: `toby.lam@balliol.ox.ac.uk`