THE INVERSE FUNCTION THEOREM AND THE LEGENDRE TRANSFORM

TOBY LAM

If I give you a monotone, differentiable function $f(x)$ that has an inverse $f^{-1}(x)$, could you tell me what $\frac{d}{dx}f^{-1}(x)$ and $\int f^{-1}(x)\,dx$ would be? The first question is easy! We all know about the inverse function theorem. 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 what $f^{-1}(x)$ and $\int f(x)dx$ is to answer the second question as well. This is similar to how you only need to know what $f^{-1}(x)$ and $f'(x)$ is to figure out $\frac{d}{dx}f^{-1}(x)$.

We shall begin by proving the inverse function theorem with a geometric flavour. We would then show how the Legendre transform provides the answer to the second question. Finally, we would show how they are both connected to each other.

A Geometric Proof of the Inverse Function Theorem

As we know,

Theorem 0.1 (Inverse Function Theorem)**.** *Given a continuously differentiable function $f : \mathbb{R} \to \mathbb{R}$ with $f'(a) \neq 0$ at some point, there exists some interval I with $a \in I$ such that there exists a continuously differentiable inverse f^{-1} defined on $f(I)$ such that*

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

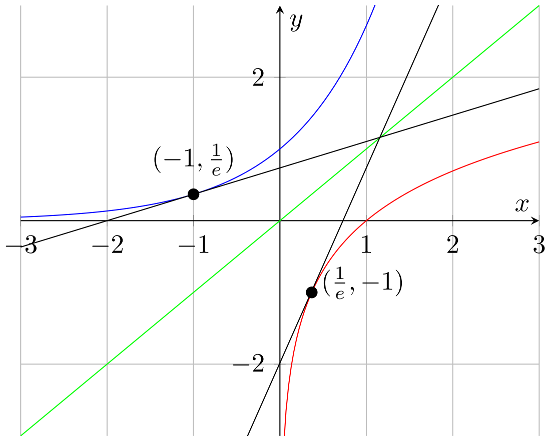
In high school, we would prove the above algebraically by differentiating both sides of $f^{-1}(f(x)) = x$. In a course in analysis, we would be even more rigorous.

However geometrically, we could do the following instead.

(1) Plot the function $y = f(x)$ on the Cartesian plane
(2) Reflect the graph over the diagonal $y = x$

The graph we get out of this would be the plot of the inverse function! As every tangent line of $f(x)$ would be reflected over the diagonal, all their slopes would also be multiplicatively inversed. This is in essence already nets us the inverse function theorem, although we need to be careful tracking the points.

Example 0.2. Below is the plot of $y = e^x$ in blue and the plot of $y = \ln x$ in red. We see how we they're related by a reflection across the green diagonal line. 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 $\frac{d}{dx}|_{x=1/e} \ln x = 1/\frac{d}{dx}|_{x=-1}\, e^x = e$



This suggests another way to attack the problem. Recall that to plot a function $f(x)$, we're essentially parametrizing a curve on $\mathbb{R}^2$ by sending x to $(x, f(x))$. The slope of this curve at x, $\frac{df}{dx}/\frac{dx}{dx}$ in turn aligns with the derivative of $f(x)$.

Now what if we parametrized the plot of the inverse function by sending x to $\phi((x, f(x))) = (f(x), x)$? The slope of the curve at $f(x)$ would now be $\frac{dx}{dx}/\frac{df}{dx} = 1/\frac{df}{dx}$, which gives us the inverse function theorem!

The Legendre Transform

Finally, let's try to work out a formula for the antiderivative of $f^{-1}(x)$. Allow me to pull something out of the hat. Let $F(x)$ be the integral of $f(x)$. Then we can observe

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

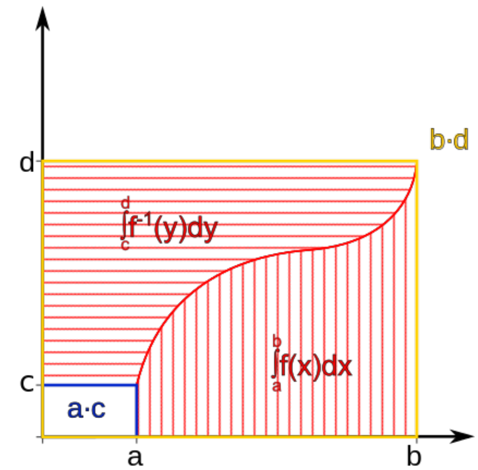
For example, you could work out that $\int \arctan(x)\,dx = x\arctan(x) - \ln|\sec(\arctan x)| + C$ for some constant C. Unfortunately, it's not at all obvious where the formula cames from. One could reverse the logic by performing an integration by parts, although I'm not so sure if it helps that much!

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

Perhaps the best proof is once again a geometric one. I claim that

$$\int_c^d f^{-1}(x)dx + \int_a^b f(x)dx = bd - ac$$

This is a result by Laisant in 1905. Here's a proof without words.



By Jonathan Steinbuch - Own work, CC BY-SA 3.0

Connecting this to the algebraic formula is an exercise left to the reader!

In fact, the transformation of $F(x)$ to $xf^{-1}(x) - F(f^{-1}(x))$ is called the Legendre transformation, 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!

Summary

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

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 in a sense, 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 derivative then take the inverse.

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