M365C (Rusin) HW11 – comments

1. Let $g(x) = \log(f(x))$. Then g is continuous and $g(x+y) = g(x) + g(y)$ for any x and y, so this is essentially the same situation as the question on the last exam: $g(x) = cx$ where $c = g(1)$. Or you could recreate the spirit of that proof: let $c = \log(f(1))$ and then prove $f(x) = \exp(cx)$ first for $x \in \mathbf{N}$, then for $x \in \mathbf{Z}$, then for $x \in \mathbf{Q}$, then (using continuity) for $x \in \mathbf{R}$.

2. I wanted you to use the fact that a^b has been *defined* to mean $\exp(b\log(a))$, so we are computing the limit of $\exp(\log(1+x)/x)$. Since exp is continuous, this is the same as $\exp(\lim_{x\to 0}\log(1+x)/x)$. We may compute that inner limit using L'Hôpital's Rule; it is 1. (You could also use the *definition* of the logarithm function to show that for $x \ge 0$, $\log(1+x) = \int_1^{1+x}(1/t)\,dt \le \int_1^{1+x} 1\,dt = x$ and $\log(1+x) = \int_1^{1+x}(1/t)\,dt \ge \int_1^{1+x} 1/(1+x)\,dt = x/(1+x)$, so the inner limit is 1 by the squeeze theorem, at least as $x \to 0^+$. Similar inequalities show the other one-sided limit is also 1.

3. Just as we defined the logarithm as a certain integral, let us define a function $A : \mathbf{R} \to \mathbf{R}$ by

$$A(x) = \begin{cases} \int_{[0,x]} \frac{1}{1+t^2}\,dt & \text{if } x \ge 0 \\ -\int_{[x,0]} \frac{1}{1+t^2}\,dt & \text{if } x < 0 \end{cases}$$

(a) Of course $A(0) = 0$ because the interval $[0,0]$ has a width of zero! The fact that $A(-x) = -A(x)$ follows from the substitution $u = -t$.

(b) Continuity on $(0,\infty)$ and $(-\infty,0)$ follows from differentiability (see below). To get continuity at 0 we have to show that the two one-sided limits of $A(x)$ as $x \to 0$ are 0. But since the integrand $1/(1+t^2)$ is everywhere at most 1, it follows that $|A(x)| \le |x|$ for every x, which forces $\lim A(x) = 0$.

(c) Differentiability away from 0 follows from the Fundamental Theorem of Calculus; in fact $A'(x) = 1/(1+x^2)$ for these x. It's also true that $A'(0)$ exists and equals 1. To see this write $1/(1+t^2)$ as $1 - t^2/(1+t^2)$ and then integrate to get $A(x) = x - B(x)$ where $B(x) = \int_0^x t^2/(1+t^2)\,dt$. The point of this move is that $t^2/(1+t^2) \le t^2$ for all t, so $B(x) \le \int_0^x t^2\,dt = x^3/3$ (with a companion inequality for $x < 0$.) Then we are ready to compute $A'(0)$ straight from the definition of the derivative. Since $A(0) = 0$, it will be the limit at 0 of $A(x)/x = 1 - B(x)/x$, but $|B(x)/x| \le x^2/3$ so $B(x)/x \to 0$ and thus $A(x)/x \to 1$. That is, $A'(0)$ exists and equals 1.

(d) Since $A'(x) > 0$ for all x, A is everywhere increasing.

(e) For any $t > 0$ we have $1/(1+t^2) > 1/t^2$ and thus for any $b > a > 0$ we have $\int_a^b 1/(1+t^2)\,dt > \int_a^b 1/t^2\,dt = F(b) - F(a)$ where $F(t)$ is an antiderivative of $1/t^2$, such as $F(t) = -1/t$. So we have $\int_a^b (1/t)\,dt < 1/a - 1/b$ for all $b > a > 0$, and thus $A(x) = \int_0^a 1/(1+t^2)\,dt + \int_a^x 1/(1+t^2)\,dt < C - 1/x < C$ where $C = \int_0^a 1/(1+t^2)\,dt + 1/a$ for any choice of a, say $a = 1$. This shows the values of A are bounded, and hence have a least upper bound. (It's $\pi/2$.)

(f) An increasing function A has an inverse defined on the image of A, which in our case is $(-\pi/2, \pi/2)$. If the inverse is called T then T is differentiable at any point $c = A(b)$ where

$A'(b)$ exists and is nonzero; in fact $T'(c) = 1/A'(b) = 1/A'(T(c))$. All those statements apply in our case, and since $A'(x) = 1/(1+x^2)$ for all x, we have $T'(c) = 1 + (T(c))^2$. There is a function $S(x)$ defined on $(-\pi/2, \pi/2)$ to be $S(x) = \sqrt{1 + T(x)^2}$; then we have the formula $T'(x) = S(x)^2$ for all x in this interval.

Of course, these functions are typically given other names: the one I have called A is called the "arctangent" function; T is the tangent function (or more precisely the restriction of the tangent function to a single period), and likewise S is (the restriction of) the secant function. Other definitions of these functions exist; for example, after we define the cosinge function in class by means of a power series, we could define the secant to be the reciprocal of the cosine. One can show that the two definitions agree by, for example, using the definitions to compute the Taylor series for each of the two functions, and verifying that the Taylor series converge to the two functions (at least on $(-\pi/2, \pi/2)$); at that point one needs only observe that the two Taylor series are identical, so the functions must be identical too (at least on that interval).

4. This is a tricky topic. It's not hard to define a "functional square root" of the exponential function as shown in the exercises, but it's hard to arrange for it to be very smooth. One can show that there *cannot be* a complex-analytic such function f, for example, while Kneser proved in 1949 that there *is* a real-analytic function. Before addressing the questions asked in the homework, let me put the problem into a broader framework for you.

Somewhat atypically, in this problem we will *not* try to keep the domain and range as two separate copies of $\mathbf{R}$. Rather, the function $f : \mathbf{R} \to \mathbf{R}$ that we are trying to construct should be thought of as a transformation moving points within a single copy of the real line. The exponential function already does this: it moves 0 to 1, 1 to e, etc., and at the same time moves the whole interval $[0, 1]$ over to $[1, e]$. We will construct a function f to do something similar: we will describe a sequence of points $0 = a_0 < a_1 < a_2 < \ldots$ and arrange it so that f carries each a_n to the next point a_{n+1} in the sequence; our f will be increasing so it will then carry each interval $[a_{n-1}, a_n]$ onto the interval $[a_n, a_{n+1}]$. Then the composite $f \circ f$ can be viewed as simply "taking two steps" along the real line in this way. We simply want to arrange it so that that (Texas) two-step matches the motion of a single step of exp. (In particular, we will have for each $n \geq 0$ that $a_{n+2} = f(a_{n+1}) = f(f(a_n)) = \exp(a_n)$. Since for every $x > 0$ we surely have $\exp(x) > 1 + x$ — that's clear from the Taylor series for exp — it follows that each $a_{n+2} > a_n + 1$, and then by induction $a_{2k} > k$, which in particular means the set of points a_n is unbounded: our function f will be defined on all of $[0, \infty)$.)

With this picture in mind, let us begin to construct f.

We begin by selecting any $a \in (0, 1)$ to be our first point a_1; we have already decided we want a_{n+2} to be $\exp(a_n)$ so we will need $a_2 = 1$, and we want to have f carry $a_0 = 0$ to $a_1 = a$ and carry a_1 to $a_2 = 1$, and in fact carry the whole interval $[a_0, a_1]$ onto $[a_1, a_2]$. So we select any increasing function f_1 mapping $[0, a]$ onto $[a, 1]$. Note that f_1 has an inverse defined on $[a, 1]$.

The definition of f on the second interval $[a_1, a_2]$ is then forced on us by the requirement that $f \circ f = \exp$: each such point x is already of the form $f(y)$ for some y in the first interval $[a_0, a_1]$, so we must have $f(x) = f(f(y)) = \exp(y)$; that is, we are forced to have $f_2(x) = \exp(f^{-1}(x))$ for all x in the second interval. (Observe that this definition would imply $f_2(a_1) = \exp(f^{-1}(a_1)) = \exp(a_0) = \exp(0) = 1$, which agrees with the value of $f_1(a_1)$. That is, we have defined f separately on the first and second (closed) intervals, and the definitions given for f_1 and f_2 agree on the single point a_1 of intersection of those intervals.) We have now defined f on the interval $[0, a_2]$, and we have done so to ensure that for any $x \leq a_1$ we have $f(f(x)) = \exp(x)$.

At this point the extension to all of $[0, \infty)$ is clear. To define f on the third interval $[a_2, a_3]$, for example, we again use the formula $f_3(x) = \exp(f_2^{-1}(x))$. In words, we can walk forward one step (i.e. apply f) by walking backward one step (i.e. apply f^{-1}) and then walk forward two steps (which is supposed to be accomplished by applying exp). Alternatively, we may walk backward *two* steps (apply $\exp^{-1} = \log$), then forward one (apply f), then forward two more (apply exp); in formulas, $f_3(x) = \exp(f_1(\log(x)))$. This is possible because we have already defined the action of f (i.e. of f_1) on the first interval; and this formalism may be preferable because we do not have to invert f_2. (Of course we inverted $f_2 \circ f_2 = \exp$ instead, but you probably feel more comfortable with log than you do with f_2^{-1} !) In exactly the same way we define each function $f_n : [a_{n-1}, a_n] \to [a_n, a_{n+1}]$ by either $f_n(x) = \exp(f_{n-1}^{-1}(x))$ or $f_n(x) = \exp(f_{n-2}(\log(x)))$. As in the previous paragraph, the definitions of $f_{n-1}(a_{n-1})$ and $f_n(a_{n-1})$ agree (both send a_{n-1} to a_n).

We began with an invertible f_1 on the first interval. If f_1 is continuous there, then its inverse is continuous on the second interval ("a continuous, invertible function on a compact set has a continuous inverse"); since exp and log are continuous on their whole domains, it follows by induction that all the remaining functions f_n are continuous on their domains and, since the definitions are now of continuous functions on closed domains, we know that glueing them together gives a function that is continuous everywhere.

The fact that the resulting function $f : [0, \infty) \to \mathbf{R}$ then satisfies $f \circ f = \exp$ is now clear from the construction. Given any $x > 0$ choose the n such that $x \in [a_{n-1}, a_n]$; then $f(x) = f_n(x)$ will by construction lie in $[a_n, a_{n+1}]$, so we will have $f(f(x)) = f(f_n(x)) = f_{n+1}(f_n(x))$; but f_{n+1} was constructed precisely so that this composite would be $\exp(x)$

Now, for differentiability, we must assume at the outset that f_1 is differentiable on the first interval $[a_0, a_1]$. From the Inverse Function Theorem, we may conclude that f_1^{-1} is also differentiable everywhere on $[a_1, a_2]$ *assuming* that f_1' is never zero on $[0, a]$, a requirement which I forgot to include on the homework assignment. (It really is necessary to have $f'(x) > 0$ for all x: the condition $f \circ f = \exp$ would lead to a contradiction if $f'(c) = 0$ since the derivative of exp is nonzero at every point.) Once we know f_1^{-1} is differentiable, it follows that f_2 is differentiable (since exp is), using the Chain Rule. Then since exp and log are both differentiable, we prove that each f_n is differentiable on its domain since f_{n-2} is, again using the Chain Rule.

However, glueing two differentiable functions together does not necessarily give a function that is differentiable at the point where the domains overlap. We must know that the left- and right-derivatives agree (i.e. those limits with $h \to 0^+$ and with $h \to 0^-$ must agree). Specifically, differentiability at $a_1 = a$ requires that the right-hand derivative, which will be $\exp(a_0) f_1'(a_0) = 1/f_1'(0)$, must match the left-hand derivative, which is simply $f_1'(a)$. So we need $f_1'(0) \cdot f_1'(a) = 1$. Similarly differentiability at a_2 requires that $\exp(a_1)/f_2'(a_1) = f_2'(a_2)$, but that simplifies (eventually) to the same condition $f_1'(0) \cdot f_1'(a) = 1$. Around later endpoints a_n we may simply use the fact that $f = \exp \circ f \circ \log$ to deduce that if f is differentiable at a_{n-2} then it is also differentiable at a_n.

We can achieve greater amounts of smoothing of f by imposing more and more conditions about the values of f_1 and its derivatives at both a_0 and a_1.

Let me point out that in this problem I viewed the application of the function exp as a kind of discrete step each point takes to the right. We computed a "functional square root" f by cutting that step in half; then in fact we may view all of f, exp, and log as members of the family of all powers of f: $\exp = f^2, \log = f^{-2}$, and of course there are multiple steps f^3 and f^{-5}, etc. One can interpolate other functions in the family — functional cube roots of exp and so on — if we can refine our construction so that instead of having a single transformation $f : \mathbf{R} \to \mathbf{R}$ that makes every point jump far to the right, we instead could define an "evolution" $F_t(x)$ which prescribed where each point should move to at time t, starting with $F_0(x) = x$ and ending with $F_1(x) = \exp(x)$ (or more generally we could let time continue to move on; what we want is for $F_{t+1}(x) = \exp(F_t(x))$. The usual choice is to have the evolution be the same for all points, that is, if after an amount of time t we have watched point a move to point b and also seen point b move to point c, then we should expect that by time $2t$, point a will in turn have moved to that same point c. In symbols: $F_u(F_t(x)) = F_{u+t}(x)$. In that case, we really only need to watch where the point $x = 0$ goes at time t, i.e. we simply need a function $\phi(t) = F_t(0)$ which has the property that $\phi(t + 1) = exp(\phi(t))$. We would then be able to recover $F_t(x) = \phi(t + \phi^{-1}(x))$, and then could create our "functional square root" as $F_{1/2}(x)$. You may wish to read up on this *Abel functional equation*.