Here is the traditional proof of the chain rule which appeared in calculus books at the turn of the century before being "lost".
Trivial lemma: if the domain of a function f is the union of two sets and the restriction of f to each of those sets converges to 0 as x approaches a, then f itself converges to 0 as x approaches a.
Now assume $z(y(x))$ is a composite of two differentiable functions and that on every deleted neighborhood of a, $\Delta y = 0$ somewhere. Then clearly $dy/dx = 0$ at a. Hence to prove the chain rule there, means to show that $\Delta z/\Delta x$ approaches 0. On the set where $\Delta y = 0$, then $\Delta z/\Delta x$ also equals 0, so this set poses no problem. On the set where $\Delta y \neq 0$, we have $\Delta z/\Delta x = (\Delta z/\Delta y)(\Delta y/\Delta x)$ so the result follows by the product rule for limits.

From this point of view, the so called "problem set" is the easier one to deal with.

This result was traditionally proved correctly in turn of the century English language books, such as Pierpont's Theory of functions of a real variable, and in 19th century European books such as that of Tannery [see the article by Carslaw, in vol XXIX of B.A.M.S.], but unfortunately not in the first three editions of the influential book Pure Mathematics, by G.H.Hardy. Although Hardy reinstated the classical proof in later editions, modern books usually deal with the problem by giving the slightly more sophisticated linear approximation proof, or making what to me are somewhat artificial constructions. The classical proof seems to have merit, so I recall it here.
The point is simply that in proving a function has limit L, one only needs to prove it at points where the function does not already have value L. Thus to someone who says that the usual argument for the chain rule for $y(u(x))$, does not work for x's where $\Delta u = 0$, one can simply reply that these points are irrelevant.

Assume f is differentiable at $g(a)$, g is differentiable at a, and on every neighborhood of a there are points x where $g(x) = g(a)$. We claim the derivative of $f(g(x))$ at a equals $f'(g(a))(g'(a))$.
Proof:
1) Clearly under these hypotheses, $g'(a) = 0$.
Consequently,
2) the chain rule holds at a if and only if $\lim \Delta f/\Delta x = 0$ as x approaches a.
3) Note that $\Delta f = \Delta f/\Delta x = 0$ at all x such that $g(x) = g(a)$.
4) In general, to prove that $\lim h(x) = L$, as x approaches a, it suffices to prove it for the restriction of h to those x such that $h(x) \neq L$.
5) Thus in arguing that $\Delta f/\Delta x$ approaches 0, we may restrict to x such that $g(x) \neq g(a)$, where the usual argument applies.