Context: taking an online Calculus course, 18.01.2 MIT Online Learning Library, and on a video about convergence of Newton's method where they analyze the error rate and derive a nifty recurrence relation that I want to show in a way understandable to me: I want to show $e_n = O(e_{n-1}^2)$. And there is one thing I will simply take for granted and accepted: that a linear approximation of a function, $f(x)$, given a shift Δx has an error term, the difference of the approximation from the true value, of $O\left((\Delta x)^2\right)$. If I accept that f can be represented by a Taylor series, then I am willing to accept this property of linear approximations.

First let me recap Newton's Method So I am using Newton's method to find a root of $f(x)$ near my initial guess, x_0. I can compute the height, $f(x_0)$ which gives me a sense of how far off my guess is to being a root (the closer $f(x_0)$ is to 0 the more accurate the guess). Then I need to consult the derivative, $f'(x_0)$ to decide two things: which direction to move x, forwards or backwards based on the signs of $f(x_0)$ and $f'(x_0)$ and what magnitude, like how much to shift x_0 to get my new guess x_1. To this end, I use a linear approximation, pretending that on average f behaves like its tangent line (average rate of change is $f'(x_0)$), to compute how much to offset x_0 by, $x_1 = x_0 - \frac{f(x_0)}{f'(x_0)}$. Why? Because moving x_0 k units to the right would result in the linear approximation $f(x_0 + k) \approx f(x_0) + f'(x_0)k$. That is, the change k induces is $f'(x_0)k$. Now, I want to move k units such that the change is $-f(x_0)$ in order for $f(x_0 + k) \approx 0$. So $k = -\frac{f(x_0)}{f'(x_0)}$ and that is how much I increment x_0 by to get my next guess of the root, x_1.

Now time to introduce error terms, let x^* be the true root of f near x_0 I'm trying to find, and let $e_0 = x^* - x_0$ and $e_1 = x^* - x_1$. Without loss of generality, let me assume that $x_0 < x^*$. Then there are 3 possibilities: $x^* = x_1$ and root found, I am done or either $x^* < x_1$ or $x^* > x_1$. The first case is not interesting at all, as $e_1 = 0$ then so let me not consider it at all. Recall that $x_1 - x_0 = k = -\frac{f(x_0)}{f'(x_0)}$. A key idea is to represent k as $k = e_0 - e_1$. That is, I view k as the both the path from $x_0 \to x_1$ and as a combined path of $x_0 \to x^* \to x_1$. In the latter combined path formulation, $x_0 \to x^*$ means go $e_0 = x^* - x_0$ units right and similarly $x^* \to x_1$ means then go $-e_1$ units right. (I trust readers can understand working with signs and directions). So again, by going k units from x_0, this induces a change in the linear approxmiation of f of $-f(x_0)$, so from $f(x_0)$ to approximately 0. But as $x^* \neq x_1$, consider what happens when I go e_0 units right. I won't vertically traverse $-f(x_0)$ units, I would either undershoot or overshoot it and the linear approximation $f(x_0 + e_0) = f(x_0) + f'(x_0)e_0 \neq 0$. I traversed $f'(x_0)e_0$ units but I'm not quite at the desired $-f(x_0)$ units. Now if I were to then go $-e_1$ units (note the negative sign), this would be the same by linearity of going k units and net change of $-f(x_0)$ vertically thus reaching 0 from $f(x_0$. Going $-e_1$ units induces $f'(x_0) - e_1 = f'(x_0)(k - e_0) = f'(x_0)k - f'(x_0)e_0 = -f(x_0) - f'(x_0)e_0$ vertical unit change Maybe also convince yourself visually that this holds for both cases where x^* is greater and less than x_1. TODO show diagrams for these two cases as pictures are quite helpful if not essential

The point is, after taking the first e_0 step, I'm not quite at 0 vertically from $f(x_0)$ start point. Where I am, particularly, is $f(x_0) + f'(x_0)e_0$ so its negation is how much offset remaining I have left: offset $= -f(x_0) + -f'(x_0)e_0$ And I need to take $-e_1$ steps to then cover the offset, offset $= f'(x_0) - e_1$. Again, I have already shown all this in the prior paragraph. The key point is that by virtue of the first e_0 step being a linear approximation, the offset from 0 (from 0 because $f(x_0 + e_0) = f(x^*) = 0$) is quadratically bounded in terms of the e_0 step size: offset $= O(e_0^2)$. And I'm basically done, combine the offset equations: $f'(x_0) - e_1 = O(e_0^2)$, and as $f'(x_0)$ is a scalar, $e_1 = O(e_0^2)$.

Parting Remark + Summary: while exploiting quadratic error big O term of linear approximation is cool, to me the highlight is simple linearity and splitting up the horizontal displacement from x_0 to x_1 into 2 sub-displaments of x_0 to x^* and then necessarily from x^* to x_1 as that's what's left. This split means I take steps e_0 and then $-e_1$ to span $-f(x_0)$ change: $-f(x_0) = f'(x_0)(e_0 + -e_1) = f'(x_0)e_0 + f'(x_0) - e_1$. And the point of this split is then to leverage the property of linear approximation: by linear approxmation I also have $-f(x_0) = f'(x_0)e_0 + O(e_0^2)$. Combining these equations, these 2 ways to cover $-f(x_0)$, I have the desired result that $e_1 = \frac{1}{f'(x_0)} O(e_0^2) = O(e_0^2)$.