-

- The new definition of differentiability from 2.1 is now the effective way to define differentiability. Now to prove that a function of several variables is differentiable we must show that a vector $\boldsymbol{c}$ and an error function $E(\boldsymbol{h})$ exist which satisfy the equality 2.16. See that in the proof of 2.19 a careful operation is performed to create such $\boldsymbol{c}$ and $E(\boldsymbol{h})$.

- On page 56, the definition of differentiability (2.16) is geometrically interpreted as: the difference between the actual graph of a function and the tangent line to it tends to zero faster than $\boldsymbol{h}$ goes to zero.

- Note that (in page 56) <u>if a function is differentiable</u> then the partial derivatives exist and the gradient vector will be the vector $\boldsymbol{c}$ (the gradient vector as oppose to the derivative vector.) But one must pay attention not to automatically replace the differentiability of a function with existence of its partial derivatives. There are examples where a function has all the partial derivative, hence the gradient vector exists but it is not satisfying the definition of differentiability as in 2.16. However if a function has continuous partial derivatives (a C^1 function) than it is differentiability. Read the discussion at the bottom of page 58: $C^1 \Longrightarrow$ differentiability $\Longrightarrow$ partial derivative exist.

- Read the discussion at the bottom of page 60 and on the top of page 61, in which the gradient (as a direction) is considered to be direction along which the function changes the fastest. That is, the directional derivative along the direction of gradient has the largest possible magnitude. In other words, if the variables (the input $(x_1, x-2, \ldots x_n)$ to the function) vary along the direction of the gradient then values of the function changes fastest.

- Pay attention to the use of mean value theorem for one variables in the middle of the proof of 2.19. This is a recurring theme.

- The differential is the linear approximation part of the definition 2.16, and it should not be mistaken with the gradient. This is traditionally the main idea of Calculus that we can replace the function with its linear approximation.

- Note also that the gradient appears as coefficients in the definition of the tangent plane to the surface defined by the function (in page 56). In the case of $n = 2$ we know that the gradient vector is located on the plane, and also one knows from linear algebra that the vector of coefficients in the equation of a plane form a vector perpendicular (or normal) to the plane. There may be a misunderstanding that the gradient vector may be both in the two dimensional (xy) plane and at the same time outside the (xy) plane and perpendicular to the surface defined by the function. However if we write the equation of tangent plane in a canonical way we see that it should be $\nabla f(\boldsymbol{a}) \cdot (\boldsymbol{x} - \boldsymbol{a}) - (z - f(\boldsymbol{a}) = 0$. in this case the vector perpendicular to the tangent plane is the same as the gradient plus a last component that is -1.

- Note that the differential $df(\boldsymbol{a}, \boldsymbol{h})$ in 2.22 iS presented as a function of two variables, the location $(\boldsymbol{a})$ and the increment $(\boldsymbol{h})$. Read the expression for du in the middle of page 59 and try to think of it as a kind of chain rule. Then again go back and think of it as the dot product as in 2.22.