Taylor's Theorem in several variables

One version of Taylor's Theorem in one-variable Calculus is

$$f(a+h) = \sum_{n=0}^{k} \frac{1}{n!} f^{(n)}(a) h^n + \frac{1}{k!} \int_{t=0}^{1} (1-t)^k f^{(k+1)}(a+th) h^{k+1} \, \mathrm{d}t.$$

To be more explicit, here is the statement for the first few values of k:

$$\begin{aligned}
f(a+h) &= f(a) + \int_{t=0}^{1} f'(a+th) h \, \mathrm{d}t \\
&= f(a) + f'(a)h + \int_{t=0}^{1} (1-t) f''(a+th) h^2 \, \mathrm{d}t \\
&= f(a) + f'(a)h + \frac{1}{2} f''(a) h^2 + \frac{1}{2} \int_{t=0}^{1} (1-t)^2 f'''(a+th) h^3 \, \mathrm{d}t \\
&\;\;\vdots
\end{aligned}$$

Here, a and h are real numbers, k is a whole number, and f is a function that is continuously differentiable $k+1$ times (at least) between a and $a+h$. These statements may be proved by repeated application of integration by parts (and the Fundamental Theorem of Calculus, which is why $f^{(k+1)}$ must not only exist but also be continuous).

To write down the general statement in several variables requires more advanced notation than we use in this class, but I will write down the first few statements when f is a function of 2 variables:

$$\begin{aligned}
f(a+h, b+i) &= f(a,b) + \int_{t=0}^{1} D_1 f(a+th, b+ti) h \, \mathrm{d}t + \int_{t=0}^{1} D_2 f(a+th, b+ti) i \, \mathrm{d}t \\
&= f(a,b) + D_1 f(a,b) h + D_2 f(a,b) i \\
&\quad + \int_{t=0}^{1} (1-t) D_{1,1} f(a+th, b+ti) h^2 \, \mathrm{d}t + \int_{t=0}^{1} (1-t) D_{1,2} f(a+th, b+ti) hi \, \mathrm{d}t \\
&\quad + \int_{t=0}^{1} (1-t) D_{2,1} f(a+th, b+ti) ih \, \mathrm{d}t + \int_{t=0}^{1} (1-t) D_{2,2} f(a+th, b+ti) i^2 \, \mathrm{d}t \\
&= f(a,b) + D_1 f(a,b) h + D_2 f(a,b) i \\
&\quad + \frac{1}{2} D_{1,1} f(a,b) h^2 + \frac{1}{2} D_{1,2} f(a,b) hi + \frac{1}{2} D_{2,1} f(a,b) ih + \frac{1}{2} D_{2,2} f(a,b) i^2 \\
&\quad + \frac{1}{2} \int_{t=0}^{1} (1-t)^2 D_{1,1,1} f(a+th, b+ti) h^3 \, \mathrm{d}t + \frac{1}{2} \int_{t=0}^{1} (1-t)^2 D_{1,1,2} f(a+th, b+ti) h^2 i \, \mathrm{d}t \\
&\quad + \frac{1}{2} \int_{t=0}^{1} (1-t)^2 D_{1,2,1} f(a+th, b+ti) hih \, \mathrm{d}t + \frac{1}{2} \int_{t=0}^{1} (1-t)^2 D_{1,2,2} f(a+th, b+ti) hi^2 \, \mathrm{d}t \\
&\quad + \frac{1}{2} \int_{t=0}^{1} (1-t)^2 D_{2,1,1} f(a+th, b+ti) ih^2 \, \mathrm{d}t + \frac{1}{2} \int_{t=0}^{1} (1-t)^2 D_{2,1,2} f(a+th, b+ti) ihi \, \mathrm{d}t \\
&\quad + \frac{1}{2} \int_{t=0}^{1} (1-t)^2 D_{2,2,1} f(a+th, b+ti) i^2 h \, \mathrm{d}t + \frac{1}{2} \int_{t=0}^{1} (1-t)^2 D_{2,2,2} f(a+th, b+ti) i^3 \, \mathrm{d}t \\
&\;\;\vdots
\end{aligned}$$

These may again be proved by using integration by parts. In fact, by doing the integration by parts in slightly different ways, we can rearrange the order of the mixed partial derivatives (such as $D_{1,2} f$ and

$D_{2,1}f$); this both proves the theorem that the mixed partial derivatives are the same in either order (when they are continuous) but also allows us to simplify the formulas slightly:

$$\begin{aligned}
f(a+h,b+i) &= f(a,b) + \int_{t=0}^{1} D_1 f(a+th,b+ti)h\,\mathrm{d}t + \int_{t=0}^{1} D_2 f(a+th,b+ti)i\,\mathrm{d}t \\
&= f(a,b) + D_1 f(a,b)h + D_2 f(a,b)i + \int_{t=0}^{1} (1-t) D_{1,1} f(a+th,b+ti)h^2\,\mathrm{d}t \\
&\quad + 2\int_{t=0}^{1} (1-t) D_{1,2} f(a+th,b+ti)hi\,\mathrm{d}t + \int_{t=0}^{1} (1-t) D_{2,2} f(a+th,b+ti)i^2\,\mathrm{d}t \\
&= f(a,b) + D_1 f(a,b)h + D_2 f(a,b)i + \frac{1}{2} D_{1,1} f(a,b)h^2 + D_{1,2} f(a,b)hi + \frac{1}{2} D_{2,2} f(a,b)i^2 \\
&\quad + \frac{1}{2}\int_{t=0}^{1} (1-t)^2 D_{1,1,1} f(a+th,b+ti)h^3\,\mathrm{d}t + \frac{3}{2}\int_{t=0}^{1} (1-t)^2 D_{1,1,2} f(a+th,b+ti)h^2 i\,\mathrm{d}t \\
&\quad + \frac{3}{2}\int_{t=0}^{1} (1-t)^2 D_{1,2,2} f(a+th,b+ti)hi^2\,\mathrm{d}t + \frac{1}{2}\int_{t=0}^{1} (1-t)^2 D_{2,2,2} f(a+th,b+ti)i^3\,\mathrm{d}t \\
&\ \ \vdots
\end{aligned}$$

However, in my opinion, the pattern is not so clear when it's put this way.

For purposes of approximation, it's useless to actually work out the integrals that appear here; if you knew the exact value of the derivatives of f at all the points between (a,b) and $(a+h,b+i)$, then you could probably just evaluate f at $(a+h,b+i)$ directly. However, if there is a value M such that you know that none of the derivatives of f of order $k+1$ have an absolute value greater than M at any point between (a,b) and $(a+h,b+i)$, then you can leave off the integrals to get an approximation of $f(a+h,b+i)$ and then use M to get an estimate of the error of this approximation:

$$\begin{aligned}
f(a+h,b+i) &\approx f(a,b), \text{ a constant approximation, if } f \text{ is continuous;} \\
f(a+h,b+i) &\approx f(a,b) + D_1 f(a,b)h + D_2 f(a,b)i, \text{ a linear approximation, if } f \text{ is differentiable;} \\
f(a+h,b+i) &\approx f(a,b) + D_1 f(a,b)h + D_2 f(a,b)i + \frac{1}{2} D_{1,1} f(a,b)h^2 + D_{1,2} f(a,b)hi + \frac{1}{2} D_{2,2} f(a,b)i^2, \\
&\qquad \text{a quadratic approximation, if } f \text{ is twice differentiable;} \\
&\ \ \vdots
\end{aligned}$$

with

$$|f(a+h,b+i) - f(a,b)| \leq M_1(|h|+|i|)$$

if $|D_1 f|$ and $|D_2 f|$ are never greater than M_1 between (a,b) and $(a+h,b+i)$,

$$\Big|f(a+h,b+i) - \big(f(a,b) + D_1 f(a,b)h + D_2 f(a,b)i\big)\Big| \leq \frac{1}{2} M_2(|h|+|i|)^2$$

if $|D_{1,1}f|$, $|D_{1,2}f|$, and $|D_{2,2}f|$ are never greater than M_2 between (a,b) and $(a+h,b+i)$,

$$\begin{aligned}
&\Big|f(a+h,b+i) - \Big(f(a,b) + D_1 f(a,b)h + D_2 f(a,b)i + \frac{1}{2} D_{1,1} f(a,b)h^2 + D_{1,2} f(a,b)hi + \frac{1}{2} D_{2,2} f(a,b)i^2\Big)\Big| \\
&\qquad \leq \frac{1}{6} M_3(|h|+|i|)^3
\end{aligned}$$

if $|D_{1,1,1}f|$, $|D_{1,1,2}f|$, $|D_{1,2,2}f|$, and $|D_{2,2,2}f|$ are never greater than M_3 between (a,b) and $(a+h,b+i)$, etc.

Using vectors, we can write the first approximation and its error in any number of variables:

$$f(P+\mathbf{v}) \approx f(P),$$
$$|f(P+\mathbf{v}) - f(P)| \leq M_1 \, |\mathbf{v}|_1,$$

where $|\mathbf{v}|_1$ is the so-called 1-norm of $\mathbf{v}$, found by adding up the absolute values of its components. (The usual magnitude is then called the 2-norm, because these absolute values are raised to the power of 2 before they are added and then the principal root of index 2 is extracted; in general, you can consider the p-norm $|\mathbf{v}|_p$ for any positive real number p, or even other values of p if you're sufficiently clever.) We can also write the second approximation and its error using vectors:

$$f(P+\mathbf{v}) \approx f(P) + \nabla f(P) \cdot \mathbf{v},$$
$$\Big| f(P+\mathbf{v}) - \big(f(P) + \nabla f(P) \cdot \mathbf{v}\big) \Big| \leq \frac{1}{2} M_2 \, |\mathbf{v}|_1^2$$

for the error estimate. The next approximation, however, requires dyads to write down, which are more complicated than vectors; to write down the general case to any order involves a massive generalization of vectors called tensors. However, you can always write it down in any specific dimension by writing a lot of terms according to the appropriate pattern, as I did on the first page; there is also a technique, called multi-index notation, to encode these patterns, which you can see (for example) on the English Wikipedia article on Taylor's Theorem.

It's handy to describe these approximations in terms of differentials and differences. While a differential represents an infinitesimal (infinitely small) change, a **difference** represents an appreciable or finitesimal (not infinitely small) change. As $R = (x, y)$ (or (x, y, z) etc) changes from P to $P + \mathbf{v}$, we say that the difference in R is

$$\Delta R = (P + \mathbf{v}) - P = \mathbf{v}.$$

Meanwhile, if $u = f(R)$, then the difference in u is

$$\Delta u|_{\substack{R=P,\\ \Delta R=\mathbf{v}}} = f(P+\mathbf{v}) - f(P).$$

Then the constant approximation says

$$\Delta u|_{\substack{R=P,\\ \Delta R=\mathbf{v}}} \approx 0,$$

while the linear approximation says (more precisely)

$$\Delta u|_{\substack{R=P,\\ \Delta R=\mathbf{v}}} \approx \mathrm{d}u|_{\substack{R=P,\\ \mathrm{d}R=\mathbf{v}}}.$$

So in the end, the linear approximation replaces differences with differentials. The next (quadratic) approximation can be written using the second differential d^2u, and so on, but we won't cover that in this class. The error estimates are

$$\Big| \Delta u|_{\substack{R=P,\\ \Delta R=\mathbf{v}}} \Big| \leq M_1 \, |\mathbf{v}|_1$$

and

$$\Big| \Delta u|_{\substack{R=P,\\ \Delta R=\mathbf{v}}} - \mathrm{d}u|_{\substack{R=P,\\ \mathrm{d}R=\mathbf{v}}} \Big| \leq \frac{1}{2} M_2 \, |\mathbf{v}|_1^2.$$