Recall that the *gradient* $\nabla f(a)$ of f at $a \in \mathbb{R}^n$ is the column vector

$$\nabla f(a) = \begin{pmatrix} \frac{\partial f}{\partial x_1}(a) \\ \frac{\partial f}{\partial x_2}(a) \\ \vdots \\ \frac{\partial f}{\partial x_n}(a) \end{pmatrix},$$

and that

$$f'(a)(u) = \mathrm{D}f(a)(u) = \nabla f(a) \cdot u,$$

for any $u \in \mathbb{R}^n$ (where $\cdot$ means inner product). The above equation shows that *the direction of the gradient $\nabla f(a)$ is the direction of maximal increase of the function f at a* and that *$\|\nabla f(a)\|$ is the rate of change of f in its direction of maximal increase*. This is the reason why methods of "gradient descent" pick the direction *opposite* to the gradient (we are trying to minimize f).

The *Hessian matrix* $\nabla^2 f(a)$ of f at $a \in \mathbb{R}^n$ is the $n \times n$ symmetric matrix

$$\nabla^2 f(a) = \begin{pmatrix} \frac{\partial^2 f}{\partial x_1^2}(a) & \frac{\partial^2 f}{\partial x_1 \partial x_2}(a) & \dots & \frac{\partial^2 f}{\partial x_1 \partial x_n}(a) \\ \frac{\partial^2 f}{\partial x_1 \partial x_2}(a) & \frac{\partial^2 f}{\partial x_2^2}(a) & \dots & \frac{\partial^2 f}{\partial x_2 \partial x_n}(a) \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial^2 f}{\partial x_1 \partial x_n}(a) & \frac{\partial^2 f}{\partial x_2 \partial x_n}(a) & \dots & \frac{\partial^2 f}{\partial x_n^2}(a) \end{pmatrix},$$

and we have

$$\mathrm{D}^2 f(a)(u, v) = u^\top \nabla^2 f(a)\, v = u \cdot \nabla^2 f(a) v = \nabla^2 f(a) u \cdot v,$$

for all $u, v \in \mathbb{R}^n$. Then, we have the following three formulations of the formula of Taylor–Young of order 2:

$$\begin{aligned} f(a + h) &= f(a) + \mathrm{D}f(a)(h) + \frac{1}{2}\mathrm{D}^2 f(a)(h, h) + \|h\|^2 \, \epsilon(h) \\ f(a + h) &= f(a) + \nabla f(a) \cdot h + \frac{1}{2}(h \cdot \nabla^2 f(a) h) + (h \cdot h)\epsilon(h) \\ f(a + h) &= f(a) + (\nabla f(a))^\top h + \frac{1}{2}(h^\top \nabla^2 f(a)\, h) + (h^\top h)\epsilon(h). \end{aligned}$$

with $\lim_{h \mapsto 0} \epsilon(h) = 0$.

One should keep in mind that only the first formula is intrinsic (i.e., does not depend on the choice of a basis), whereas the other two depend on the basis and the inner product chosen