1 Hessian

The Hessian of f at a point x, denoted $B = H(f)(x) = \nabla^2 f(x)$, is the second derivative of a function $f : \mathbb{R}^n \to \mathbb{R}$ at the point x and hence a bilinear map

$$B : \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}.$$

it is usually identified with the matrix

$$H_{ij} = \nabla^2 f(x)(e_i, e_j) = \frac{\partial^2 f}{\partial x_i \partial x_j}(x)$$

and with this identification we have

$$B(u, v) = u^T H v.$$

Let $f : \mathbb{R}^n \to \mathbb{R}$ be a C^2-function and $\overline{f}(u) = f(x_0 + Fu)$, where $x_0 \in \mathbb{R}^n$ and $F : \mathbb{R}^m \to \mathbb{R}^n$ is a linear map, i.e. $F \in Mat_{n\times m}$.

We want to compute the gradient and Hessian of h at any point u from those of f. To get these do a second order Taylor expansion of f about x:

$$f(x + h) = f(x) + h^T g + h^T Hh + o(\|h\|^2),$$

where $g = \nabla f(x)$ and $H = \nabla^2 f(x)$ are uniquely determined by the fact that the residual is $o(\|h\|^2)$. Applying this to the point $x = x_0 + Fu$ this implies that

$$\begin{aligned} \overline{f}(u + h) &= f(x_0 + Fu + Fh) \\ &= f(x_0 + Fu) + (Fh)^T g + (Fh)^T HFh + o(\|Fh\|^2). \end{aligned}$$

Since $o(\|Fh\|^2)$ is $o(\|h\|^2)$ we conclude from this that

$$\nabla \overline{f}(u) = F^T g \quad \text{and} \quad \nabla^2 \overline{f}(u) = F^T HF,$$

or, more explicitly

$$\nabla \overline{f}(u) = F^T \nabla f(x_0 + Fu) \quad \text{and} \quad \nabla^2 \overline{f}(u) = F^T \nabla^2 f(x_0 + Fu) F. \tag{1}$$

With a similar approach we can compute the Hessian of a composition $g(f(x))$, where here $g : \mathbb{R} \to \mathbb{R}$ is a scalar function of one variable (more general g are much harder to handle and we do not need them). Indeed, set

$$y = f(x), \quad \nabla = \nabla f(x) \quad H = \nabla^2 f(x) \text{and} \quad k = h^T \nabla + \tfrac{1}{2} h^T Hh$$

and use second order Taylorapproximations on f at the point x and g at the point $y = f(x)$ to obtain:

$$\begin{aligned} g(f(x + h))) &= g\left(f(x) + h^T \nabla + \tfrac{1}{2} h^T Hh\right) \\ &= g(y + k) = g(y) + g'(y) + \tfrac{1}{2} g''(y) k^2 + o(k^2) \\ &= g(y) + g'(y)\left(h^T \nabla + \tfrac{1}{2} h^T Hh\right) + \frac{1}{2} g''(y)\left(h^T \nabla + \tfrac{1}{2} h^T Hh\right)^2 + o(\|h\|^2). \end{aligned}$$

Here we have used that $o(k^2) = o(\|h\|^2)$. Collect terms of first and second order in h together and sticking all terms of higher order into the residual $o(\|h\|^2)$. Note that the squared term contributes no first order terms and only one second order term, this being the term

$$(h^T \nabla)^2 = (h^T \nabla)(h^T \nabla) = (h^T \nabla)(h^T \nabla)^T = h^T (\nabla \nabla^T) h.$$

We obtain

$$g(f(x+h))) = g(y) + h^T [g'(y) \nabla] + \frac{1}{2} h^T \left[g'(y) H + g''(y) \nabla \nabla^T \right] h + o(\|h\|^2).$$

and from this we can read off that

$$\nabla (g \circ f)(x) = g'(f(x)) \nabla f(x), \quad \text{and} \tag{2}$$
$$H(g \circ f)(x) = g'(f(x)) H + g''(f(x)) \nabla f(x) \nabla f(x)^T \tag{3}$$

Note that here $d = \nabla f(x)$ is viewed as a column vector and so $\nabla f(x) \nabla f(x)^T$ is the *outer product*

$$\nabla f(x) \nabla f(x)^T = dd^T = (d_i d_j)_{ij}.$$