Why is the Hessian of a function well-defined only at its critical points?

Ivo Terek Couto

Defining $\mathrm{d}^2 f_p$:

Let M^n be a differentiable manifold, $f\colon M \to \mathbb{R}$ be a smooth function, and $p \in M$ be a critical point of f, that is, satisfying $\mathrm{d}f_p = 0$. This means that the partial derivatives of f with respect to any chart around p vanish when evaluated at p. This allows us to write the:

Definition. The *Hessian* of f at p is the bilinear form $\mathrm{d}^2 f_p\colon T_pM \times T_pM \to \mathbb{R}$ defined by

$$\mathrm{d}^2 f_p(v, w) \doteq \sum_{i,j=1}^n \frac{\partial^2 f}{\partial x_\alpha^i \partial x_\alpha^j}(p) v_\alpha^i w_\alpha^j,$$

where $(U_\alpha, \varphi_\alpha = (x_\alpha^1, \ldots, x_\alpha^n))$ is a chart around p for which we write

$$v = \sum_{i=1}^n v_\alpha^i \frac{\partial}{\partial x_\alpha^i}\bigg|_p \quad \text{and} \quad w = \sum_{i=1}^n w_\alpha^j \frac{\partial}{\partial x_\alpha^j}\bigg|_p.$$

To make this definition valid, we have to verify that the expression does not depend on the choice of chart around p. For this end, assume that we are given a second chart $(U_\beta, \varphi_\beta = (x_\beta^1, \ldots, x_\beta^n))$ around p. Then $U_\alpha \cap U_\beta$ is an *open* set around p (hence we are able to take derivatives), and we may assume without loss of generality (and to simplify the writing) that $\varphi_\alpha(p) = \varphi_\beta(p) = \mathbf{0} = (0, \ldots, 0) \in \mathbb{R}^n$. The relation between the coordinate vector fields along $U_\alpha \cap U_\beta$, evaluated at the correct points, is just

$$\frac{\partial}{\partial x_\beta^j} = \sum_{\ell=1}^n \frac{\partial x_\alpha^\ell}{\partial x_\beta^j} \frac{\partial}{\partial x_\alpha^\ell}, \qquad j = 1, 2, \ldots, n.$$

Seeing this as an equality between differential operators, it follows that

$$\frac{\partial f}{\partial x_\beta^j} = \sum_{\ell=1}^n \frac{\partial x_\alpha^\ell}{\partial x_\beta^j} \frac{\partial f}{\partial x_\alpha^\ell}, \qquad j = 1, 2, \ldots, n.$$

Applying $\partial/\partial x_\beta^i$ on both sides and applying the product rule, we get

$$\frac{\partial^2 f}{\partial x_\beta^i \partial x_\beta^j} = \sum_{\ell=1}^n \frac{\partial^2 x_\alpha^\ell}{\partial x_\beta^i \partial x_\beta^j} \frac{\partial f}{\partial x_\alpha^\ell} + \sum_{k,\ell=1}^n \frac{\partial x_\alpha^k}{\partial x_\beta^i} \frac{\partial x_\alpha^\ell}{\partial x_\beta^j} \frac{\partial^2 f}{\partial x_\alpha^k \partial x_\alpha^\ell}.$$

Evaluating the above at the point p kills the first sum in the right hand side, in view of the condition $\mathrm{d}f_p = 0$ (which implies that $(\partial f/\partial x^\ell_\alpha)(p) = 0$ for $\ell = 1, 2, \ldots, n$), resulting in

$$\frac{\partial^2 f}{\partial x^i_\beta \partial x^j_\beta}(p) = \sum_{k,\ell=1}^{n} \frac{\partial x^k_\alpha}{\partial x^i_\beta}(\mathbf{0}) \frac{\partial x^\ell_\alpha}{\partial x^j_\beta}(\mathbf{0}) \frac{\partial^2 f}{\partial x^k_\alpha \partial x^\ell_\alpha}(p).$$

Now, to compute the Hessian of f according to the chart (U_β, φ_β), we need to know the components of the tangent vectors v and w with respect to this new coordinate basis. Using self-evident notation, we have that

$$v^i_\beta = \sum_{r=1}^{n} \frac{\partial x^i_\beta}{\partial x^r_\alpha}(\mathbf{0}) v^r_\alpha \quad \text{and} \quad w^j_\beta = \sum_{s=1}^{n} \frac{\partial x^j_\beta}{\partial x^s_\alpha}(\mathbf{0}) w^s_\alpha, \qquad i, j = 1, 2, \ldots, n.$$

Putting everything together, we finally compute:

$$\begin{aligned}
\sum_{i,j=1}^{n} \frac{\partial^2 f}{\partial x^i_\beta \partial x^j_\beta}(p) v^i_\beta w^j_\beta &= \sum_{i,j,k,\ell,r,s=1}^{n} \frac{\partial x^k_\alpha}{\partial x^i_\beta}(\mathbf{0}) \frac{\partial x^\ell_\alpha}{\partial x^j_\beta}(\mathbf{0}) \frac{\partial^2 f}{\partial x^k_\alpha \partial x^\ell_\alpha}(p) \frac{\partial x^i_\beta}{\partial x^r_\alpha}(\mathbf{0}) v^r_\alpha \frac{\partial x^j_\beta}{\partial x^s_\alpha}(\mathbf{0}) w^s_\alpha \\
&= \sum_{k,\ell,r,s=1}^{n} \left(\sum_{i=1}^{n} \frac{\partial x^k_\alpha}{\partial x^i_\beta}(\mathbf{0}) \frac{\partial x^i_\beta}{\partial x^r_\alpha}(\mathbf{0}) \right) \left(\sum_{j=1}^{n} \frac{\partial x^\ell_\alpha}{\partial x^j_\beta}(\mathbf{0}) \frac{\partial x^j_\beta}{\partial x^s_\alpha}(\mathbf{0}) \right) \frac{\partial^2 f}{\partial x^k_\alpha \partial x^\ell_\alpha}(p) v^r_\alpha w^s_\alpha \\
&= \sum_{k,\ell,r,s=1}^{n} \delta^k_r \delta^\ell_s \frac{\partial^2 f}{\partial x^k_\alpha \partial x^\ell_\alpha}(p) v^r_\alpha w^s_\alpha \\
&= \sum_{k,\ell=1}^{n} \frac{\partial^2 f}{\partial x^k_\alpha \partial x^\ell_\alpha}(p) v^k_\alpha w^\ell_\alpha,
\end{aligned}$$

as wanted. This means that the Hessian is indeed well-defined if p is a critical point of the function f. Perhaps a more elegant approach for checking this last part, avoiding picking the tangent vectors v and w (but which obviously boils down to the same computation), is to write the transformation law for the differentials at the point p instead:

$$\mathrm{d}x^i_\beta\big|_p = \sum_{r=1}^{n} \frac{\partial x^i_\beta}{\partial x^r_\alpha}(\mathbf{0})\, \mathrm{d}x^r_\alpha\big|_p \qquad i = 1, 2, \ldots, n,$$

setting up

$$\sum_{i,j=1}^{n} \frac{\partial^2 f}{\partial x^i_\beta \partial x^j_\beta}(\mathbf{0})\, \mathrm{d}x^i_\beta\big|_p \otimes \mathrm{d}x^j_\beta\big|_p = \sum_{i,j,k,l=1}^{n} \frac{\partial x^k_\alpha}{\partial x^i_\beta}(\mathbf{0}) \frac{\partial x^\ell_\alpha}{\partial x^j_\beta}(\mathbf{0}) \frac{\partial^2 f}{\partial x^k_\alpha \partial x^\ell_\alpha}(p) \left(\sum_{r=1}^{n} \frac{\partial x^i_\beta}{\partial x^r_\alpha}(\mathbf{0})\, \mathrm{d}x^r_\alpha\big|_p \right) \otimes \left(\sum_{s=1}^{n} \frac{\partial x^j_\beta}{\partial x^s_\alpha}(\mathbf{0})\, \mathrm{d}x^s_\alpha\big|_p \right)$$

and recognizing δ^k_r and δ^ℓ_s to again obtain the same conclusion.

Generalizations:

Everything here uses in a crucial way the fact that $\mathrm{d}f_p = 0$. So this raises the natural question: is it possible to define such a Hessian for arbitrary points of the manifold M?

Without additional structure, the answer is *no*. If you do, however, have some extra structure to work with, here's what happens: let ∇ be a (Koszul) connection in the tangent bundle TM, and define the *covariant Hessian* of f with respect to ∇ at p as the map $\mathrm{Hess}^\nabla(f)_p\colon T_pM \times T_pM \to \mathbb{R}$ given by

$$\mathrm{Hess}^\nabla(f)_p(v,w) = v(\widetilde{\boldsymbol{w}}(f)) - \mathrm{d}f_p(\nabla_v \widetilde{\boldsymbol{w}}),$$

where $\widetilde{\boldsymbol{w}}$ is some extension of w to a neighborhood of p (i.e., a vector field defined in a neighborhood of p such that $\widetilde{\boldsymbol{w}}_p = w$). By the Leibniz rule for ∇ and its local character, we see that the right hand side above is actually independent of the choice of extension for w, and defines a bilinear form on T_pM. Note that if p happens to be a critical point of f, we recover $\mathrm{Hess}^\nabla(f)_p = \mathrm{d}^2 f_p$.

This actually induces a $\mathscr{C}^\infty(M)$-bilinear map $\mathrm{Hess}^\nabla(f)\colon \mathfrak{X}(M) \times \mathfrak{X}(M) \to \mathscr{C}^\infty(M)$, which is given in local coordinates $(U, (x^1, \ldots, x^n))$ by

$$\begin{aligned} \mathrm{Hess}^\nabla(f)(\partial_i, \partial_j) &= \partial_i\partial_j f - \mathrm{d}f(\nabla_{\partial_i}\partial_j) \\ &= \partial_i\partial_j f - \mathrm{d}f\left(\sum_{k=1}^n \Gamma_{ij}^k \partial_k\right) \\ &= \partial_i\partial_j f - \sum_{k=1}^n \Gamma_{ij}^k \partial_k f, \end{aligned}$$

where the n^3 functions Γ_{ij}^k are the connection components of ∇. Writing it in its full glory, we have

$$\mathrm{Hess}^\nabla(f) = \sum_{i,j=1}^n \left(\frac{\partial^2 f}{\partial x^i \partial x^j} - \sum_{k=1}^n \Gamma_{ij}^k \frac{\partial f}{\partial x^k}\right) \mathrm{d}x^i \otimes \mathrm{d}x^j.$$

One might also recognize the object $\mathrm{Hess}^\nabla(f)$ as the covariant differential $\nabla(\mathrm{d}f)$ of the $(0,1)$-tensor $\mathrm{d}f$, which is then a $(0,2)$-tensor. But despite all these ways of looking at the Hessian, we cannot expect it to necessarily have good properties, since the connection ∇ was so arbitrary. In fact, recall that the *torsion* of the connection ∇ is the $(0,2)$-tensor field $\tau^\nabla\colon \mathfrak{X}(M) \times \mathfrak{X}(M) \to \mathfrak{X}(M)$ given by

$$\tau^\nabla(\boldsymbol{X}, \boldsymbol{Y}) = \nabla_{\boldsymbol{X}}\boldsymbol{Y} - \nabla_{\boldsymbol{Y}}\boldsymbol{X} - [\boldsymbol{X}, \boldsymbol{Y}],$$

where $[\boldsymbol{X}, \boldsymbol{Y}]$ is the Lie bracket of $\boldsymbol{X}$ and $\boldsymbol{Y}$. The presence of $[\boldsymbol{X}, \boldsymbol{Y}]$ has the purpose of making the torsion τ^∇ $\mathscr{C}^\infty(M)$-bilinear. We'll conclude the discussion with the following characterization of this torsion:

Proposition. $\tau^\nabla = 0$ *if and only if* $\mathrm{Hess}^\nabla(f)$ *is a symmetric tensor, for every* $f \in \mathscr{C}^\infty(M)$.

Proof: Given vector fields $\boldsymbol{X}, \boldsymbol{Y} \in \mathfrak{X}(M)$, we compute directly that

$$\begin{aligned} \mathrm{Hess}^\nabla(f)(\boldsymbol{Y},\boldsymbol{X}) &= \boldsymbol{Y}(\boldsymbol{X}(f)) - \mathrm{d}f(\nabla_{\boldsymbol{Y}}\boldsymbol{X}) = \boldsymbol{X}(\boldsymbol{Y}(f)) - [\boldsymbol{X}, \boldsymbol{Y}](f) - \mathrm{d}f(\nabla_{\boldsymbol{Y}}\boldsymbol{X}) \\ &= \boldsymbol{X}(\boldsymbol{Y}(f)) - \mathrm{d}f(\nabla_{\boldsymbol{Y}}\boldsymbol{X} + [\boldsymbol{X}, \boldsymbol{Y}]) = \boldsymbol{X}(\boldsymbol{Y}(f)) - \mathrm{d}f(\nabla_{\boldsymbol{X}}\boldsymbol{Y} - \tau^\nabla(\boldsymbol{X}, \boldsymbol{Y})) \\ &= \mathrm{Hess}^\nabla(f)(\boldsymbol{X}, \boldsymbol{Y}) + \tau^\nabla(\boldsymbol{X}, \boldsymbol{Y})(f). \end{aligned}$$

The conclusion follows. □