On the Properties of the Hessian Tensor for vector functions

Eduardo Magalhães

May 3, 2020

Abstract

In this paper some properties and the chain rule for the hessian tensor for combined vector functions are derived. We will derive expressions for $H(T+L)$, $H(aT)$, and $H(T \circ L)$ (chain rule for hessian tensors) and show some specific examples of the chain rule in certain types of composite maps.

1 Introduction

Let $f : \mathbb{R}^n \longrightarrow \mathbb{R}$. This function takes as input a vector $\mathbf{x} \in \mathbb{R}^n$ and outputs a scalar $f(\mathbf{x}) \in \mathbb{R}$. let:

$$\mathbf{x} = \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix}$$

Then the hessian matrix of this function is defined as:

$$\mathbf{H}(f) = \begin{pmatrix} \partial_{x_1 x_1} f & \cdots & \partial_{x_1 x_n} f \\ \vdots & \ddots & \vdots \\ \partial_{x_n x_1} f & \cdots & \partial_{x_n x_n} f \end{pmatrix}$$

We can find the value of each entry of the matrix by the following formula:

$$\mathbf{H}_{ij} = \partial_{x_i x_j} f$$

We can generalize this concept for any map between two vector spaces:

Let V and W be two vector spaces, and let $T : V \longrightarrow W$ be a function between them. Let:

$$T(\mathbf{x}) = \begin{pmatrix} g_1(\mathbf{x}) \\ \vdots \\ g_m(\mathbf{x}) \end{pmatrix}$$

Then we define the hessian of this function as:

$$\mathbf{H} = \begin{pmatrix} \mathbf{H}(g_1) & \cdots & \mathbf{H}(g_m) \end{pmatrix}$$

So this is a third-order tensor. We can denote each component of the tensor as:

$$H_{\gamma ij} = \partial_{x_i x_j} g_\gamma$$

We will explore some properties of this tensor when there is composition of vector functions.

2 Content

2.1 Linearity

Proposition 1. Let V and W be two vector spaces over a field $\mathbb{F}$, and let $T : V \longrightarrow W$,$L : V \longrightarrow W$ be 2 functions between those 2 vector spaces. Then we have that:

$$\mathbf{H}(T + L) = \mathbf{H}(T) + \mathbf{H}(L) \tag{1}$$

$$\mathbf{H}(\alpha T) = \alpha \mathbf{H}(T), \quad \alpha \in \mathbb{F} \tag{2}$$

Proof 1.

(1) Let V and W be two vector spaces over a field $\mathbb{F}$, and let $T : V \longrightarrow W$,$L : V \longrightarrow W$ be 2 functions between those 2 vector spaces. Let's say that:

$$T(\mathbf{x}) = \begin{pmatrix} t_1(\mathbf{x}) \\ \vdots \\ t_m(\mathbf{x}) \end{pmatrix} \quad \text{and} \quad L(\mathbf{x}) = \begin{pmatrix} l_1(\mathbf{x}) \\ \vdots \\ l_m(\mathbf{x}) \end{pmatrix}$$

we have that:

$$(T + L)(\mathbf{x}) = \begin{pmatrix} t_1(\mathbf{x}) + l_1(\mathbf{x}) \\ \vdots \\ t_m(\mathbf{x}) + l_m(\mathbf{x}) \end{pmatrix}$$

So the components of the hessian tensor for the sum of those functions will be:

$$H_{\gamma ij}(T + L) = \partial_{x_i x_j}(t_\gamma + l_\gamma) =$$

$$\partial_{x_i x_j} t_\gamma + \partial_{x_i x_j} l_\gamma = H_{\gamma ij}(T) + H_{\gamma ij}(L)$$

Thus, $\mathbf{H}(T + L) = \mathbf{H}(T) + \mathbf{H}(L)$.

(2) Let V and W be two vector spaces over a field $\mathbb{F}$, $\alpha \in \mathbb{F}$, and let $T : V \longrightarrow W$ be a function between those 2 vector spaces with:

$$T(\mathbf{x}) = \begin{pmatrix} t_1(\mathbf{x}) \\ \vdots \\ t_m(\mathbf{x}) \end{pmatrix}$$

then, we have that:

$$(\alpha T)(\mathbf{x}) = \begin{pmatrix} \alpha t_1(\mathbf{x}) \\ \vdots \\ \alpha t_m(\mathbf{x}) \end{pmatrix}$$

The components of the hessian tensor will be:

$$H_{\gamma ij}(\alpha T) = \partial_{x_i x_j}(\alpha t_\gamma) =$$

$$\alpha \partial_{x_i x_j}(t_\gamma) = \alpha H_{\gamma ij}(T)$$

Thus, $\mathbf{H}(\alpha T) = \alpha \mathbf{H}(T)$.

2.2 Composition of functions

We will now deduce a formula for the Hessian tensor of composite functions.

Let V, W, K be vector spaces such that $\dim V = v, \dim W = w$ and $\dim K = k$. And let T and L be two functions $L : V \longrightarrow W$ and $T : W \longrightarrow K$, Such that:

$$T(\mathbf{x}) = \begin{pmatrix} t_1(\mathbf{x}) \\ \vdots \\ t_k(\mathbf{x}) \end{pmatrix}, \quad \mathbf{x} \in W$$

$$L(\mathbf{x}) = \begin{pmatrix} t_1(\mathbf{x}) \\ \vdots \\ t_w(\mathbf{x}) \end{pmatrix}, \quad \mathbf{x} \in V$$

Then we have that $T \circ L : V \longrightarrow K$ such that:

$$(T \circ L)(\mathbf{x}) = \begin{pmatrix} t_1(L(\mathbf{x})) \\ \vdots \\ t_k(L(\mathbf{x})) \end{pmatrix}, \quad \mathbf{x} \in V$$

because we have that $L(\mathbf{x}) = \begin{pmatrix} l_1(\mathbf{x}) \\ \vdots \\ l_w(\mathbf{x}) \end{pmatrix}$ we can write an expression for every component of the vector $(T \circ L)(\mathbf{x})$:

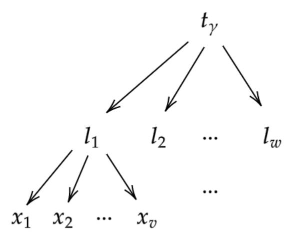
$$[(T\circ L)(\mathbf{x})]_\gamma = t_\gamma \begin{pmatrix} l_1(\mathbf{x}) \\ \vdots \\ l_w(\mathbf{x}) \end{pmatrix}$$

Now we can compute the components for the Hessian tensor:

$$H_{\gamma ij}(T\circ L) = \partial_{x_i x_j} t_\gamma \begin{pmatrix} l_1(\mathbf{x}) \\ \vdots \\ l_w(\mathbf{x}) \end{pmatrix} =$$

$$\partial_{x_j}\left[\partial_{x_i} t_\gamma \begin{pmatrix} l_1(\mathbf{x}) \\ \vdots \\ l_w(\mathbf{x}) \end{pmatrix}\right]$$

Let's first evaluate the derivative with respect to x_i. We can use the chain rule to do that. We have:



So we have that:

$$\partial_{x_i} t_\gamma \begin{pmatrix} l_1(\mathbf{x}) \\ \vdots \\ l_w(\mathbf{x}) \end{pmatrix} = \sum_w \partial_{l_w} t_\gamma(L)\partial_{x_i} l_w(\mathbf{x})$$

This gives us:

$$\partial_{x_j}\left[\partial_{x_i} t_\gamma \begin{pmatrix} l_1(\mathbf{x}) \\ \vdots \\ l_w(\mathbf{x}) \end{pmatrix}\right] = \partial_{x_j}\left[\sum_w \partial_{l_w} t_\gamma(L)\partial_{x_i} l_w(\mathbf{x})\right]$$

We can use the fact the the partial derivative is linear:

$$\partial_{x_j}\left[\sum_w \partial_{l_w} t_\gamma(L)\partial_{x_i} l_w(\mathbf{x})\right] = \sum_w \partial_{x_j}\left(\partial_{l_w} t_\gamma(L)\partial_{x_i} l_w(\mathbf{x})\right)$$

Using the product rule for partial derivatives we have:

$$\sum_w \partial_{x_j} \left(\partial_{l_w} t_\gamma(L) \partial_{x_i} l_w(\mathbf{x})\right) = \sum_w \left(\partial_{x_j} \left[\partial_{l_w} t_\gamma(L)\right] \partial_{x_i} l_w(\mathbf{x}) + \partial_{l_w} t_\gamma(L) \partial_{x_i x_j} l_w(\mathbf{x})\right) =$$

$$\sum_w \partial_{x_j} \left[\partial_{l_w} t_\gamma(L)\right] \partial_{x_i} l_w(\mathbf{x}) + \sum_w \partial_{l_w} t_\gamma(L) \partial_{x_i x_j} l_w(\mathbf{x})$$

note that $\partial_{x_i x_j} l_w$ is the component H_{wij} of the hessian tensor of L. So we can rewrite this as:

$$\sum_w \partial_{x_j} \left[\partial_{l_w} t_\gamma(L)\right] \partial_{x_i} l_w(\mathbf{x}) + \sum_w \partial_{l_w} t_\gamma(L) H_{wij}(L)$$

So we get:

$$H_{\gamma ij}(T \circ L) = \sum_w \partial_{x_j} \left[\partial_{l_w} t_\gamma(L)\right] \partial_{x_i} l_w(\mathbf{x}) + \sum_w \partial_{l_w} t_\gamma(L) H_{wij}(L) \tag{3}$$

2.2.1 Specific cases

Let's now look at some specific cases and see how formula (3) transform under those certain specific circumstances.

(1) Let's assume the same things we assumed in 2.2, but this time let's assume that $L : V \longrightarrow W$ is linear. He have that:

$$H_{\gamma ij}(T \circ L) = \sum_w \partial_{x_j} \left[\partial_{l_w} t_\gamma(L)\right] \partial_{x_i} l_w(\mathbf{x}) + \sum_w \partial_{l_w} t_\gamma(L) H_{wij}(L)$$

Because L is linear we have that $\mathbf{H}(L) = \mathbf{0}$ [1], so $H_{\gamma ij}(L) = 0$. Because of this $\sum_w \partial_{l_w} t_\gamma(L) H_{wij}(L) = 0$.

$$H_{\gamma ij}(T \circ L) = \sum_w \partial_{x_j} \left[\partial_{l_w} t_\gamma(L)\right] \partial_{x_i} l_w(\mathbf{x})$$

L can also be written in terms of a matrix because it is linear:

$$L(\mathbf{x}) = \begin{pmatrix} A_{11} & \cdots & A_{1v} \\ \vdots & \ddots & \vdots \\ A_{w1} & \cdots & A_{wv} \end{pmatrix} \begin{pmatrix} x_1 \\ \vdots \\ x_v \end{pmatrix} =$$

$$\begin{pmatrix} \sum_v A_{1v} x_v \\ \vdots \\ \sum_v A_{wv} x_v \end{pmatrix}$$

so, for any γ:

$$l_\gamma(\mathbf{x}) = \sum_v A_{\gamma v} x_v$$

If we plug this in the previous equation we get:

$$H_{\gamma ij}(T \circ L) = \sum_w \partial_{x_j} \left[\partial_{l_w} t_\gamma(L)\right] \partial_{x_i} \sum_v A_{wv} x_v =$$

$$\sum_w \partial_{x_j} \left[\partial_{l_w} t_\gamma(L)\right] \sum_v A_{wv} \partial_{x_i} x_v =$$

$$\sum_w \partial_{x_j} \left[\partial_{l_w} t_\gamma(L)\right] \sum_v A_{wv} \delta_{iv}$$

Where $\delta_{iv}$is the Kronecker delta.

$$\sum_v A_{wv}\delta_{iv} = A_{wi}\delta_{ii} + \sum_{v \neq i} A_{wv}\delta_{iv} = A_{wi}$$

We can now plug this back in your equation giving us:

$$\sum_w \partial_{x_j} \left[\partial_{l_w} t_\gamma(L)\right] \sum_v A_{wv}\delta_{iv} = \sum_w A_{wi} \partial_{x_j} \left[\partial_{l_w} t_\gamma(L)\right]$$

So, if L is a linear map, then:

$$H_{\gamma ij}(T \circ L) = \sum_w A_{wi} \partial_{x_j} \left[\partial_{l_w} t_\gamma(L)\right] \tag{4}$$

(2) Now let's show, unsing formula (4), that $\mathbf{H}(T \circ Id) = \mathbf{H}(T)$. Let V be a vector space such that $\dim V = v$. Let $T : V \longrightarrow V$ and $Id : V \longrightarrow V$. Because Id is linear we have:

$$H_{\gamma ij}(T \circ Id) = \sum_v I_{vi} \partial_{x_j} \left[\partial_{l_v} t_\gamma(Id)\right]$$

Where I is the $v \times v$ identity matrix, and where $Id(\mathbf{x}) = \begin{pmatrix} l_1(\mathbf{x}) \\ \vdots \\ l_v(\mathbf{x}) \end{pmatrix}$. Because the defining property of Id is that $Id(\mathbf{x}) = \mathbf{x}$ then $l_\gamma = x_\gamma$, If we make this substitution on the equation we get:

$$H_{\gamma ij}(T \circ Id) = \sum_v I_{vi} \partial_{x_j} \left[\partial_{x_v} t_\gamma(\mathbf{x})\right] =$$

$$\sum_v \delta_{vi} \partial_{x_j x_v} t_\gamma$$

Where δ_{vi} is the Kronecker delta.

$$\sum_{v} \delta_{vi} \partial_{x_j x_v} t_\gamma = \delta_{ii} \partial_{x_j x_i} t_\gamma + \sum_{v \neq i} \delta_{vi} \partial_{x_j x_v} t_\gamma = \partial_{x_i x_j} t_\gamma$$

This gives us:

$$H_{\gamma ij}(T \circ Id) = \partial_{x_i x_j} t_\gamma = H_{\gamma ij}(T)$$

(3) If we let $f, g : \mathbb{R} \longrightarrow \mathbb{R}$, then, because the Hessian matrix and the hessian tensor is a generalization of the second derivative, the formula (3) used to calculate $\mathbf{H}(f \circ g)$ will give us an expression for the second derivative of $(f \circ g)$.

Formula (3) gives us:

$$H_{\gamma ij}(T \circ L) = \sum_{w} \partial_{x_j} \left[\partial_{l_w} t_\gamma(L)\right] \partial_{x_i} l_w(\mathbf{x}) + \sum_{w} \partial_{l_w} t_\gamma(L) H_{wij}(L)$$

If $f, g\colon \mathbb{R} \longrightarrow \mathbb{R}$,then the hessian tensor will have conly one component, making it a constant. so we can get rid of all those indices relative to the specific component of the tensor we are calculating. The equation will simplify to:

$$H(f \circ g) = \sum_{w} \partial_x \left[\partial_{g_w} f(g)\right] \partial_x g_w(x) + \sum_{w} \partial_{g_w} f(g) H(g)$$

Because this are functions are single variable functions we can change the partial derivatives to normal ones, and we can get rid of the slums because $w \in \{1\}$. The equation simplifies further to:

$$H(f \circ g) = \frac{d}{dx}\left[\frac{d}{dg} f(g)\right] \frac{d}{dx} g(x) + \frac{d}{dg} f(g) H(g)$$

The Hessian of g is simply the second derivative of g:

$$H(f \circ g) = \frac{d}{dx}\left[\frac{d}{dg} f(g)\right] \frac{d}{dx} g(x) + \frac{d}{dg} f(g) g'' =$$

$$\frac{d}{dx}\left[f'(g)\right] g' + g'' f'(g) = \left[g'\right]^2 f''(g) + g'' f'(g)$$

Thus giving us: $H(f \circ g) = \left[g'\right]^2 f''(g) + g'' f'(g)$. Because the hessian of a single variable function is the second derivative of that function we get:

$$(f \circ g)'' = \left[g'\right]^2 f''(g) + g'' f'(g)$$

It's easy to show that this is true using the chain rule for single variable functions.

References

[1] Eduardo Magalhães (2020) *Geometric Definition of Linear Transformations* `https://vixra.org/abs/2005.0018`