On Hessians of Composite Functions

Čestmír Bárta, Martin Kolář and Jan Šustek

Department of Mathematics, Faculty of Science
The University of Ostrava, 701 03 Ostrava
Czech Republic

Copyright © 2014 Čestmír Bárta, Martin Kolář and Jan Šustek. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Abstract

In this paper we will study the Hessian matrices of composite functions $F$ which are in the form $F(x) = f(g(x))$ with $f: \mathbb{R} \to \mathbb{R}$ and $g: \mathbb{R}^n \to \mathbb{R}$. We present an explicit formulas for the Hessian of $F$ and for the inverse of the Hessian matrix.

Mathematics Subject Classification: 26B05

Keywords: Hessian, composite function

1 Introduction

For a function $f: \mathbb{R}^n \to \mathbb{R}$ the Hessian matrix $Hf(x)$ is defined as the $n \times n$ matrix of the second-order partial derivatives,

$$Hf(x) = \begin{pmatrix}
\frac{\partial^2 f}{\partial x_1^2} & \cdots & \frac{\partial^2 f}{\partial x_1 \partial x_n} \\
\vdots & \ddots & \vdots \\
\frac{\partial^2 f}{\partial x_n \partial x_1} & \cdots & \frac{\partial^2 f}{\partial x_n^2}
\end{pmatrix}.$$ 

The determinant of the above matrix is called Hessian. The Hessian matrix plays an important role in many areas of mathematics. It is important in neural computing, for instance several non-linear optimization algorithms used for training neural networks are based on considerations of the second-order properties of the error surface, which are controlled by the Hessian matrix [1].
The inverse of the Hessian matrix has been used to identify the least significant weights in a network as a part of network ‘pruning’ algorithms, or it can also be used to assign error bars to the predictions made by a trained network [1]. Hessian matrices are also applied in Morse theory [4].

Throughout this paper we will study the Hessian matrices of composite functions \( F \) which are in the form \( F(x) = f(g(x)) \) with \( f : \mathbb{R} \rightarrow \mathbb{R} \) and \( g : \mathbb{R}^n \rightarrow \mathbb{R} \). We present an explicit formulas for the Hessian of \( F \) and for the inverse of the Hessian matrix. In previous years, several papers were published with such formulas, but only for special functions \( F \), see e.g. [2] or [6]. These results follow easily from our general results.

We will use the following notation for simple work with vectors and matrices

\[
\begin{align*}
col a(i) &= \begin{pmatrix} a(1) \\ \vdots \\ a(n) \end{pmatrix}, \\
row a(i) &= (a(1) \ldots a(n)), \\
mat a(i,j) &= \begin{pmatrix} a(1,1) \ldots a(1,n) \\ \vdots \\ a(n,1) \ldots a(n,n) \end{pmatrix}, \\
diag a(i) &= \begin{pmatrix} a(1) & 0 \\ \vdots & \ddots \\ 0 & a(n) \end{pmatrix}.
\end{align*}
\]

Using this notation we have

\[
\nabla g(x) = \col \frac{\partial g}{\partial x_i} \quad \text{and} \quad I = \text{diag} 1 = \mat \delta_{ij},
\]

where \( \delta_{ij} \) is the Kronecker symbol, i.e. \( \delta_{ii} = 1 \) and \( \delta_{ij} = 0 \) for \( i \neq j \).

2 Results

Our first result concerns the determinant of the Hessian matrix.

**Theorem 1** Let \( f : \mathbb{R} \rightarrow \mathbb{R} \) and \( g : \mathbb{R}^n \rightarrow \mathbb{R} \) be real functions and let \( F(x) = f(g(x)) \) be their composition. Then the Hessian of the function \( F \) equals

\[
\det H F(x) = \left( 1 + \frac{f''(g(x))}{f'(g(x))} \nabla g(x)^T \cdot H g(x)^{-1} \cdot \nabla g(x) \right) f'(g(x))^n \det H g(x).
\]

The following result concerns the inverse of the Hessian matrix.

**Theorem 2** Let \( f : \mathbb{R} \rightarrow \mathbb{R} \) and \( g : \mathbb{R}^n \rightarrow \mathbb{R} \) be real functions and let \( F(x) = f(g(x)) \) be their composition. Then the inverse of the Hessian matrix of the function \( F \) equals

\[
H F(x)^{-1} = \left( I - \frac{H g(x)^{-1} \cdot \nabla g(x) \cdot \nabla g(x)^T}{f''(g(x))} \right) \frac{H g(x)^{-1} \cdot \nabla g(x) \cdot \nabla g(x)^T}{f'(g(x))}.
\]
Theorem 3 is a consequence of previous theorems for a particular class of functions $g$.

**Theorem 3** Let $\alpha_i \in \mathbb{R}$ for $i = 1, \ldots, n$ and let $f: \mathbb{R} \to \mathbb{R}$ be a real function. Put $g(x) = \prod_{k=1}^{n} x_k^{\alpha_k}$ and $F(x) = f(g(x))$. Then the Hessian of $F$ equals

$$
\det H F(x) = \left(1 + \frac{g(x) f''(g(x))}{f'(g(x))} \frac{|\alpha|}{|\alpha| - 1} \right) f'(g(x))^{n-1} (-1)^{n+1} (|\alpha| - 1) \prod_{k=1}^{n} \alpha_k \cdot x_k^{n\alpha_k - 2},
$$

where $|\alpha| = \sum_{k=1}^{n} \alpha_k$. The inverse of the Hessian matrix is

$$
HF(x)^{-1} = \frac{1}{g(x)f'(g(x))} \text{mat} \left( \frac{1}{|\alpha| - 1} - \frac{f'(g(x))}{g(x)f''(g(x))} (|\alpha| - 1)^2 + |\alpha| (|\alpha| - 1) \frac{\delta_{ij}}{\alpha_i} \right) x_i x_j.
$$

**Example 1** Chen [2] considered function of the form $F(x) = f(\sum_{k=1}^{n} h_k(x_k))$ and computed inductively its Hessian. We obtain the same result directly using Theorem 1. We have $g(x) = \sum_{k=1}^{n} h_k(x_k)$. Hence

$$
\nabla g(x) = \text{col} h'_k(x_k), \quad \text{H}_g(x) = \text{diag} h''_k(x_k),
$$

$$
\det H g(x) = \prod_{k=1}^{n} h''_k(x_k), \quad \text{H}_g(x)^{-1} = \text{diag} \frac{1}{h''_k(x_k)}.
$$

Then Theorem 1 implies that

$$
\det H F(x) = \left(1 + \frac{f''(g(x))}{f'(g(x))} \sum_{k=1}^{n} \frac{h_k(x_k)^2}{h''_k(x_k)} \right) f'(g(x))^n \prod_{k=1}^{n} h''_k(x_k).
$$

Moreover, using Theorem 2, we obtain the inverse of the Hessian matrix.

$$
HF(x)^{-1} = \left(\mathbf{I} - \frac{\text{diag} \frac{1}{h''_k(x_k)} \cdot \text{row} h'_k(x_k) \cdot \text{col} h'_k(x_k)}{\frac{f'(g(x))}{f''(g(x))} + \text{row} h'_k(x_k) \cdot \text{diag} \frac{1}{h''_k(x_k)} \cdot \text{col} h'_k(x_k)} \right) f'(g(x)) \frac{1}{h''_k(x_k)}
$$

$$
= \text{mat} \left( \delta_{ij} - \frac{h'_k(x_k) h'_l(x_l)}{h''_k(x_k) h''_l(x_l)} \right) \cdot \text{diag} \frac{1}{f'(g(x)) h''_k(x_k)}
$$

$$
= \text{mat} \left( \delta_{ij} - \frac{h'_k(x_k) h'_l(x_l)}{f'(g(x)) h''_k(x_k) + \sum_{k=1}^{n} \frac{h_k^2(x_k)}{h''_k(x_k)}} \right) \frac{1}{f'(g(x)) h''_k(x_k) h''_l(x_l)}
$$
Example 2 Trojovský and Hladíková [6] considered function \( F(x) = \exp \prod_{k=1}^{n} x_{k} \) and computed its Hessian as the sum of \( 2^n \) simpler determinants. We obtain the same result directly using Theorem 3. We have \( f(t) = e^t \) and \( \alpha_{i} = 1 \) for every \( i \). Hence \( f'(t) = f''(t) = e^t \). Then Theorem 3 implies that

\[
\det H F(x) = \left(1 + g(x) \frac{n}{n-1}\right) e^{ng(x)} (-1)^{n+1} (n-1) g(x)^{n-2}
\]

\[
= (-1)^{n+1} \left(n - 1 + n \prod_{k=1}^{n} x_{k}\right) e^{n \prod_{k=1}^{n} x_{k}} x_{n-2}.
\]

Moreover we obtain the inverse of the Hessian matrix.

\[
HF(x)^{-1} = \frac{1}{g(x) e^{g(x)}} \text{mat} \left( \frac{1}{n-1} - \frac{1}{g(x)} + \frac{n}{n-1} - \delta_{ij} \right) x_{i} x_{j}
\]

\[
= \text{mat} \frac{1}{\prod_{k=1}^{n} x_{k}} \left( \frac{1}{n-1} - \frac{\prod_{k=1}^{n} x_{k}}{(n-1)^2 + n(n-1) \prod_{k=1}^{n} x_{k}} - \delta_{ij} \right) x_{i} x_{j}
\]

Example 3 Consider the function \( F(x) = \sqrt[n]{\prod_{k=1}^{n} x_{k}} \), i.e. the geometric mean.

Using the notation of Theorem 3, we have \( f(t) = \sqrt[t]{t} \) and \( \alpha_{i} = 1 \) for every \( i \). Hence \( f'(t) = \frac{1}{n} t^{\frac{1}{n} - 1} \) and \( f''(t) = \frac{1}{n} \left( \frac{1}{n} - 1 \right) t^{\frac{1}{n} - 2} \). Then Theorem 3 implies that

\[
\det H F(x) = \left(1 + g(x) \frac{1}{n} \left( \frac{1}{n} - 1 \right) g(x)^{\frac{1}{n} - 2} \right) \frac{1}{n^n} g(x)^{1-n} (-1)^{n+1} (n-1) g(x)^{n-2} = 0.
\]

3 Proofs

In our proofs we will need the following two lemmas. We present them also with their short proofs.

**Lemma 1** (Matrix determinant lemma) ([3], Lemma 1.1) Let \( A \) be an invertible \( n \times n \) matrix and let \( u, v \in \mathbb{R}^n \) be column vectors. Then

\[
\det(A + uv^T) = (1 + v^T A^{-1} u) \det A.
\]

**Proof.** By a direct computation we easily obtain the identity

\[
\begin{pmatrix}
A & 0 \\
v^T & 1
\end{pmatrix}
\begin{pmatrix}
I + A^{-1} uv^T & A^{-1} u \\
0^T & 1
\end{pmatrix}
\begin{pmatrix}
I & 0 \\
v^T & 1
\end{pmatrix}
= \begin{pmatrix}
A & u \\
0^T & 1 + v^T A^{-1} u
\end{pmatrix}.
\]
Taking determinants of both sides we immediately obtain
\[
\det(A + uv^T) = \det A \cdot \det(I + A^{-1}uv^T) = \det A \cdot (1 + v^T A^{-1}u). \quad \square
\]

**Lemma 2** (Sherman-Morrison formula) ([5], Section 2.7.1) Let \( A \) be an invertible \( n \times n \) matrix and let \( u, v \in \mathbb{R}^n \) be column vectors. Suppose that the matrix \( A + uv^T \) is invertible. Then

\[
(A + uv^T)^{-1} = \left( I - \frac{A^{-1}uv^T}{1 + v^T A^{-1}u} \right) A^{-1}.
\]

**Proof.** From the assumption and Lemma 1 it follows that \( 1 + v^T A^{-1}u \neq 0 \). Then

\[
I = I + A^{-1}uv^T - \frac{A^{-1}uv^T + (v^T A^{-1}u)A^{-1}uv^T}{1 + v^T A^{-1}u}
= I + A^{-1}uv^T - \frac{A^{-1}uv^T + A^{-1}u(v^T A^{-1}u)v^T}{1 + v^T A^{-1}u}
= A^{-1}(A + uv^T) - \frac{A^{-1}uv^T A^{-1}}{1 + v^T A^{-1}u} (A + uv^T)
= \left( A^{-1} - \frac{A^{-1}uv^T A^{-1}}{1 + v^T A^{-1}u} \right) (A + uv^T)
\]

We obtain the result by multiplying by \((A + uv^T)^{-1}\). \quad \square

Using these lemmas we now prove our theorems.

**Proof of Theorem 1.** The second-order partial derivatives of \( F \) are

\[
\frac{\partial^2 F}{\partial x_i \partial x_j} = f''(g(x)) \frac{\partial g}{\partial x_i}(x) \frac{\partial g}{\partial x_j}(x) + f'(g(x)) \frac{\partial^2 g}{\partial x_i \partial x_j}(x). \tag{1}
\]

Hence the Hessian matrix can be written as \( HF(x) = A + uv^T \), where

\[
A = f'(g(x))Hg(x), \quad u = f''(g(x)) \nabla g(x), \quad v = \nabla g(x). \tag{2}
\]

From Lemma 1 we obtain

\[
\det HF(x) = \left( 1 + \nabla g(x)^T \left( f'(g(x))Hg(x) \right)^{-1} \cdot f''(g(x))\nabla g(x) \right) \det \left( f'(g(x))Hg(x) \right)
= \left( 1 + \frac{f''(g(x))}{f'(g(x))} \nabla g(x)^T \cdot Hg(x)^{-1} \cdot \nabla g(x) \right) f'(g(x))^n \det Hg(x). \quad \square
\]
**Proof of Theorem 2.** As in the proof of Theorem 1 we obtain (1) and (2). Then Lemma 2 implies
\[
Hf(x)^{-1} = \left( I - \frac{(f'(g(x)))Hg(x)^{-1} \cdot (f''(g(x))\nabla g(x))^T \cdot (f'(g(x)))^T}{1 + (\nabla g(x))^T \cdot (f'(g(x)))Hg(x)^{-1} \cdot (f''(g(x))\nabla g(x))} \right) \cdot (f'(g(x)))^T \cdot (f'(g(x)))^{-1}
\]
\[
= \left( I - \frac{Hg(x)^{-1} \cdot \nabla g(x) \cdot \nabla g(x)^T}{f'(g(x))} \right) \cdot (f'(g(x)))^{-1}.
\]
\[\square\]

**Proof of Theorem 3.** The derivatives of the function \( g \) are
\[
\frac{\partial g}{\partial x_i} = \alpha_i^x \frac{g(x)}{x_i}, \quad \frac{\partial^2 g}{\partial x_i^2} = \alpha_i(\alpha_i - 1) x_i^2 g(x), \quad \frac{\partial^2 g}{\partial x_i \partial x_j} = \alpha_i\alpha_j x_i x_j g(x).
\]
Therefore its gradient is equal to
\[
\nabla g(x) = g(x) \cdot \frac{\alpha_i}{x_i}
\]
and its Hessian matrix is
\[
Hg(x) = \begin{pmatrix}
\text{mat} \frac{\alpha_i\alpha_j}{x_i x_j} - \text{diag} \frac{\alpha_i}{x_i^2}
\end{pmatrix} g(x) = -g(x)(A + uv^T),
\]
where
\[
A = \text{diag} \frac{\alpha_i}{x_i^2}, \quad u = -\text{col} \frac{\alpha_i}{x_i}, \quad v = \text{col} \frac{\alpha_i}{x_i}.
\]
Then Lemma 1 implies that
\[
\det Hg(x) = (-g(x))^n \left( 1 - \text{row} \frac{\alpha_i}{x_i} \cdot \text{diag} \frac{x_i^2}{\alpha_i} \cdot \text{col} \frac{\alpha_i}{x_i} \right) \prod_{k=1}^{n} \frac{\alpha_k}{x_k^2}
\]
\[
= (-g(x))^n (1 - |\alpha|) \prod_{k=1}^{n} \frac{\alpha_k}{x_k^2}
\]
\[
= (-1)^n (1 - |\alpha|) \prod_{k=1}^{n} \alpha_k x_k^{|\alpha| - 2}.
\]
\[\text{(3)}\]

Lemma 2 implies that
\[
Hg(x)^{-1} = -\frac{1}{g(x)} \begin{pmatrix}
\text{diag} \frac{x_i^2}{\alpha_i} + \text{diag} \frac{x_i^2}{\alpha_i} \cdot \text{col} \frac{\alpha_i}{x_i} \cdot \text{row} \frac{\alpha_i}{x_i} & \text{diag} \frac{x_i^2}{\alpha_i} \\
1 - \text{row} \frac{\alpha_i}{x_i} \cdot \text{diag} \frac{x_i^2}{\alpha_i} \cdot \text{col} \frac{\alpha_i}{x_i}
\end{pmatrix}
\]
\[
= \frac{-1}{g(x)} \begin{pmatrix}
\text{diag} \frac{x_i^2}{\alpha_i} + \text{mat} x_i x_j \\
1 - |\alpha|
\end{pmatrix}.
\]
\[\text{(4)}\]
We will use the following identity
\[
\text{row } \frac{\alpha_i}{x_i} \cdot \text{diag } \frac{x_i^2}{\alpha_i} \cdot \text{col } \frac{\alpha_i}{x_i} + \text{row } \frac{\alpha_i}{x_i} \cdot \text{mat } \frac{x_i x_j}{1 - |\alpha|} \cdot \text{col } \frac{\alpha_i}{x_i} = |\alpha| + \frac{|\alpha|^2}{1 - |\alpha|} = \frac{|\alpha|}{1 - |\alpha|}.
\]

Then (3), (4), (5) and Theorem 1 imply that
\[
\det H F(x) = \left(1 + \frac{f''(g(x))}{f'(g(x))} g(x) \text{ row } \frac{\alpha_i}{x_i} \cdot \frac{-1}{g(x)} \left(\text{diag } \frac{x_i^2}{\alpha_i} + \text{mat } \frac{x_i x_j}{1 - |\alpha|} \right) \cdot g(x) \text{ col } \frac{\alpha_i}{x_i}\right)
\times f'(g(x))^n (-1)^n (1 - |\alpha|) \prod_{k=1}^{n} \alpha_k x_k^{n \alpha_k - 2}
\]
\[
= \left(1 - \frac{g(x) f''(g(x))}{f'(g(x))} \left(\text{row } \frac{\alpha_i}{x_i} \cdot \text{diag } \frac{x_i^2}{\alpha_i} \cdot \text{col } \frac{\alpha_i}{x_i} + \text{row } \frac{\alpha_i}{x_i} \cdot \text{mat } \frac{x_i x_j}{1 - |\alpha|} \cdot \text{col } \frac{\alpha_i}{x_i}\right)\right)
\times f'(g(x))^n (-1)^{n+1} (|\alpha| - 1) \prod_{k=1}^{n} \alpha_k x_k^{n \alpha_k - 2}
\]
\[
= \left(1 + \frac{g(x) f''(g(x))}{f'(g(x))} \frac{|\alpha|}{|\alpha| - 1}\right) f'(g(x))^n (-1)^{n+1} (|\alpha| - 1) \prod_{k=1}^{n} \alpha_k x_k^{n \alpha_k - 2}.
\]

By a direct computation, using (5), we obtain
\[
\nabla g(x)^T \cdot Hg(x)^{-1} \cdot \nabla g(x) = g(x) \text{ row } \frac{\alpha_i}{x_i} \cdot \frac{-1}{g(x)} \left(\text{diag } \frac{x_i^2}{\alpha_i} + \text{mat } \frac{x_i x_j}{1 - |\alpha|} \right) \cdot g(x) \text{ col } \frac{\alpha_i}{x_i}
\]
\[
= -g(x) \left(\text{row } \frac{\alpha_i}{x_i} \cdot \text{diag } \frac{x_i^2}{\alpha_i} \cdot \text{col } \frac{\alpha_i}{x_i} + \text{row } \frac{\alpha_i}{x_i} \cdot \text{mat } \frac{x_i x_j}{1 - |\alpha|} \cdot \text{col } \frac{\alpha_i}{x_i}\right)
\]
\[
= g(x) \frac{|\alpha|}{|\alpha| - 1}
\]
\]
and
\[
Hg(x)^{-1} \cdot \nabla g(x) \cdot \nabla g(x)^T = -\frac{1}{g(x)} \left(\text{diag } \frac{x_i^2}{\alpha_i} + \text{mat } \frac{x_i x_j}{1 - |\alpha|} \right) \cdot g(x) \text{ col } \frac{\alpha_i}{x_i} \cdot g(x) \text{ row } \frac{\alpha_i}{x_i}
\]
\[
= -g(x) \left(\text{diag } \frac{x_i^2}{\alpha_i} \cdot \text{col } \frac{\alpha_i}{x_i} \cdot \text{row } \frac{\alpha_i}{x_i} + \frac{1}{|\alpha|} \text{ mat } x_i x_j \cdot \text{col } \frac{\alpha_i}{x_i} \cdot \text{row } \frac{\alpha_i}{x_i}\right)
\]
\[
= g(x) \frac{|\alpha|}{|\alpha| - 1} \text{ mat } \frac{\alpha_j x_i}{x_j}. \]
Then (6), (7) and Theorem 2 imply

\[
\mathbf{H} F(x)^{-1} = \left( \mathbf{I} - \frac{\mathbf{H} g(x)^{-1} \cdot \nabla g(x) \cdot \nabla g(x)^T}{f'(g(x))} \right) \cdot \frac{\mathbf{H} g(x)^{-1}}{f'(g(x))} \\
= \left( \mathbf{I} - \frac{g(x)}{|\alpha| - 1} \text{mat} \frac{\alpha_i x_i}{x_j} \right) \cdot \frac{-1}{g(x)f'(g(x))} \left( \text{diag} \frac{x_i^2}{\alpha_i} + \text{mat} x_i x_j \right),
\]

where

\[
\beta = \frac{g(x)}{|\alpha| - 1} \frac{f'(g(x))}{f'(g(x)) + g(x)}.
\]

Now we have

\[
\beta \text{mat} \frac{\alpha_j x_i}{x_j} \cdot \text{diag} \frac{x_i^2}{\alpha_i} = \beta \text{mat} x_i x_j,
\]

\[
\beta \text{mat} \frac{\alpha_j x_i}{x_j} \cdot \text{mat} x_i x_j \frac{1}{1 - |\alpha|} = \frac{\beta|\alpha|}{1 - |\alpha|} \text{mat} x_i x_j,
\]

\[
-\mathbf{I} \cdot \text{diag} \frac{x_i^2}{\alpha_i} = -\text{mat} \frac{\delta_{ij}}{\alpha_i} x_i x_j,
\]

\[
-\mathbf{I} \cdot \text{mat} \frac{x_i x_j}{1 - |\alpha|} = \frac{1}{|\alpha| - 1} \text{mat} x_i x_j.
\]

From the identity \( \beta + \frac{\beta|\alpha|}{1 - |\alpha|} = \frac{\beta}{1 - |\alpha|} \) we obtain that

\[
(9) + (10) = \frac{-1}{\frac{f'(g(x))}{g(x)f'(g(x))}(|\alpha| - 1)^2 + |\alpha|(|\alpha| - 1)} \text{mat} x_i x_j.
\]

Finally, (8), (11), (12) and (13) imply that

\[
\mathbf{H} F(x)^{-1} = \frac{1}{g(x)f'(g(x))} \text{mat} \left( \frac{1}{|\alpha| - 1} - \frac{f'(g(x))}{g(x)f'(g(x))}(|\alpha| - 1)^2 + |\alpha|(|\alpha| - 1) \frac{\delta_{ij}}{\alpha_i} \right) x_i x_j.
\]

**Acknowledgement**

The authors were supported by grant GAČR P201/12/2351 of the Czech Science Foundation and by grant SGS08/PrF/2014 of the University of Ostrava.
References


Received: October 19, 2014; Published: December 1, 2014