Second Partial Derivatives and the Hessian

Last modified: 2026-11-24 09:19:45 (PST)

Second partial derivatives and the Hessian

Definition 0.1 (Open ball)**.** Let $\underset{\sim}{x} \in \mathbb{R}^p$ and $r > 0$. The **open ball** [1] of radius r around $\underset{\sim}{x}$ is the set $\{\underset{\sim}{y} \in \mathbb{R}^p : \|\underset{\sim}{y} - \underset{\sim}{x}\| < r\}$, of points whose distance from $\underset{\sim}{x}$ (the Euclidean norm of the difference) is less than r.

Definition 0.2 (Closed ball)**.** For $\underset{\sim}{x}$ and r as in Definition 0.1, the **closed ball** of radius r around $\underset{\sim}{x}$ is $\{\underset{\sim}{y} \in \mathbb{R}^p : \|\underset{\sim}{y} - \underset{\sim}{x}\| \leq r\}$.

Definition 0.3 (Unit ball)**.** The open balls and closed balls of radius 1 around $\underset{\sim}{0}$ are the open and closed **unit balls**.

Example 0.1 (Points in the unit balls of $\mathbb{R}^2$)**.**

- $(0.5, 0.5)^\top$ has

$$\begin{aligned} \left\|(0.5, 0.5)^\top\right\|^2 &= 0.25 + 0.25 \\ &= 0.5 \\ &< 1, \end{aligned}$$

  so it is in both the open and the closed unit ball.

- $(0.6, 0.8)^\top$ has

$$\begin{aligned} \left\|(0.6, 0.8)^\top\right\|^2 &= 0.36 + 0.64 \\ &= 1, \end{aligned}$$

  so it is in the closed unit ball but not the open one.

- For $p = 1$, the open ball of radius r around c is the open interval $(c - r, c + r)$.

Definition 0.4 (Second partial derivative)**.** Let f be a scalar-valued function of a $p \times 1$ vector $\underset{\sim}{x}$, and let $i, j \in \{1, \ldots, p\}$. Suppose the partial derivative $\frac{\partial}{\partial x_j} f$ (Definition 1 in Partial Derivatives and Gradients) exists at every point of an open ball around $\underset{\sim}{x}$ (Definition 0.1). The **second partial derivative** of f at $\underset{\sim}{x}$, first with respect to x_j and then with respect to x_i, is

$$\frac{\partial}{\partial x_i}\left(\frac{\partial}{\partial x_j} f(\underset{\sim}{x})\right),$$

the partial derivative with respect to x_i of the function $\frac{\partial}{\partial x_j} f$, when it exists.

Example 0.2 (Second partial derivatives of $x_1^2 x_2 + 3x_2$)**.** For $f(\underset{\sim}{x}) = x_1^2 x_2 + 3x_2$ of Example 1 in Partial Derivatives and Gradients, $\frac{\partial}{\partial x_1} f(\underset{\sim}{x}) = 2x_1 x_2$ and $\frac{\partial}{\partial x_2} f(\underset{\sim}{x}) = x_1^2 + 3$. Differentiating each of these again:

- $$\begin{aligned}\frac{\partial}{\partial x_1}\left(\frac{\partial}{\partial x_1} f(\underset{\sim}{x})\right) &= \frac{\partial}{\partial x_1}(2x_1 x_2) \\ &= 2x_2;\end{aligned}$$
- $$\begin{aligned}\frac{\partial}{\partial x_2}\left(\frac{\partial}{\partial x_1} f(\underset{\sim}{x})\right) &= \frac{\partial}{\partial x_2}(2x_1 x_2) \\ &= 2x_1;\end{aligned}$$
- $$\begin{aligned}\frac{\partial}{\partial x_1}\left(\frac{\partial}{\partial x_2} f(\underset{\sim}{x})\right) &= \frac{\partial}{\partial x_1}(x_1^2 + 3) \\ &= 2x_1;\end{aligned}$$
- $$\begin{aligned}\frac{\partial}{\partial x_2}\left(\frac{\partial}{\partial x_2} f(\underset{\sim}{x})\right) &= \frac{\partial}{\partial x_2}(x_1^2 + 3) \\ &= 0.\end{aligned}$$

The two mixed partial derivatives are both $2x_1$.

Definition 0.5 (Mixed partial derivative)**.** When $i \neq j$, the second partial derivative of Definition 0.4 is a **mixed partial derivative**.

Definition 0.6 (Hessian matrix)**.** Let f be a scalar-valued function of a $p \times 1$ vector $\underset{\sim}{x}$ whose first partial derivatives exist on an open ball around $\underset{\sim}{x}$ (Definition 0.1) and whose second partial derivatives (Definition 0.4) exist at $\underset{\sim}{x}$. The **Hessian matrix** [2] of f at $\underset{\sim}{x}$ is the derivative (Definition 2 in Jacobians and Derivative Rules) of the transposed gradient $\left(\frac{\partial}{\partial \underset{\sim}{x}} f(\underset{\sim}{x})\right)^\top$ (Definition 2 in Partial Derivatives and Gradients):

$$\underbrace{\mathbf{H}_f(\underset{\sim}{x})}_{p \times p} \stackrel{\text{def}}{=} \frac{\partial}{\partial \underset{\sim}{x}} \left(\frac{\partial}{\partial \underset{\sim}{x}} f(\underset{\sim}{x}) \right)^\top. \tag{1}$$

Example 0.3 (A Hessian matrix)**.** Let $f(\underset{\sim}{x}) = e^{2x_1 + x_2} - x_1$ for $\underset{\sim}{x} = (x_1, x_2)^\top$, where e is Euler's number, and write $u = 2x_1 + x_2$. By the chain rule, $\frac{\partial}{\partial x_1} e^u = 2e^u$ and $\frac{\partial}{\partial x_2} e^u = e^u$, so

$$\frac{\partial}{\partial \underset{\sim}{x}} f(\underset{\sim}{x}) = \begin{bmatrix} 2e^u - 1 \\ e^u \end{bmatrix}.$$

Differentiating each entry again,

$$\begin{aligned} \mathbf{H}_f(\underset{\sim}{x}) &= \begin{bmatrix} \frac{\partial}{\partial x_1}(2e^u - 1) & \frac{\partial}{\partial x_1} e^u \\ \frac{\partial}{\partial x_2}(2e^u - 1) & \frac{\partial}{\partial x_2} e^u \end{bmatrix} \\ &= \begin{bmatrix} 4e^u & 2e^u \\ 2e^u & e^u \end{bmatrix}, \end{aligned}$$

and at $\underset{\sim}{x} = \underset{\sim}{0}$, where $u = 0$, $\mathbf{H}_f(\underset{\sim}{0}) = \begin{bmatrix} 4 & 2 \\ 2 & 1 \end{bmatrix}$.

Example 0.4 (A function with a gradient but no Hessian at a point)**.** Let $f(\underset{\sim}{x}) = x_1 |x_1|$ for $\underset{\sim}{x} = (x_1, x_2)^\top$, where $|\cdot|$ is the absolute value. For $x_1 > 0$, $f = x_1^2$ and $\frac{\partial}{\partial x_1} f = 2x_1$; for $x_1 < 0$, $f = -x_1^2$ and $\frac{\partial}{\partial x_1} f = -2x_1$; and at $x_1 = 0$ the difference quotient is $h|h|/h = |h| \to 0$. So the gradient exists everywhere: it is $(2|x_1|,\ 0)^\top$. But $2|x_1|$ has no derivative in x_1 at $x_1 = 0$ (its difference quotient $2|h|/h$ is 2 for $h > 0$ and -2 for $h < 0$), so $\left[\mathbf{H}_f(\underset{\sim}{x})\right]_{11}$, and with it the Hessian, does not exist at any $\underset{\sim}{x}$ with $x_1 = 0$.

Theorem 0.1 (Other ways to write the Hessian, and its entries)**.** *Let f and $\underset{\sim}{x}$ be as in Definition 0.6.*

1. *The Hessian can be written with the row-vector derivative (Definition 1 in Jacobians and Derivative Rules) of f:*

$$\begin{aligned}\mathbf{H}_f(\underset{\sim}{x}) &= \frac{\partial}{\partial \underset{\sim}{x}}\left(\frac{\partial f(\underset{\sim}{x})}{\partial \underset{\sim}{x}^\top}\right) \\ &= \frac{\partial^2 f(\underset{\sim}{x})}{\partial \underset{\sim}{x}\partial \underset{\sim}{x}^\top}.\end{aligned} \tag{2}$$

2. The (i, j) entry of the Hessian is

$$\left[\mathbf{H}_f(\underset{\sim}{x})\right]_{ij} = \frac{\partial}{\partial x_i}\left(\frac{\partial}{\partial x_j} f(\underset{\sim}{x})\right). \tag{3}$$

Proof

Proof. **Part 1.** By Theorem 1 in Jacobians and Derivative Rules, the transposed gradient $\left(\frac{\partial}{\partial \underset{\sim}{x}} f(\underset{\sim}{x})\right)^\top$ equals the row-vector derivative $\frac{\partial f(\underset{\sim}{x})}{\partial \underset{\sim}{x}^\top}$. Substituting it into Equation 1 gives the first line. The second line is the same expression written as a second derivative. It follows the shape rule of matrix multiplication (Remark 2 in Jacobians and Derivative Rules): the $p \times 1$ column-vector operator $\frac{\partial}{\partial \underset{\sim}{x}}$ stands on the left of a $1 \times p$ row vector, giving a $p \times p$ matrix.

Part 2. Let $y_j \overset{\text{def}}{=} \frac{\partial}{\partial x_j} f(\underset{\sim}{x})$ be entry j of the gradient. By Definition 2 in Jacobians and Derivative Rules, the (i, j) entry of the derivative of the vector $\underset{\sim}{y} = (y_1, \ldots, y_p)$ with respect to $\underset{\sim}{x}$ is

$$\begin{aligned}\left[\mathbf{H}_f(\underset{\sim}{x})\right]_{ij} &= \left[\frac{\partial}{\partial \underset{\sim}{x}} \underset{\sim}{y}^\top\right]_{ij} && \text{(Equation 1, with } \underset{\sim}{y} \text{ the gradient)} \\ &= \frac{\partial}{\partial x_i} y_j && \text{(Definition 2 in Jacobians and Derivative Rules)} \\ &= \frac{\partial}{\partial x_i}\left(\frac{\partial}{\partial x_j} f(\underset{\sim}{x})\right). && \left(y_j = \frac{\partial}{\partial x_j} f(\underset{\sim}{x})\right)\end{aligned}$$

□

Example 0.5 (Computing the Hessian entry by entry, and checking it by finite differences)**.** Take $f(\underset{\sim}{x}) = e^{2x_1 + x_2} - x_1$ from Example 0.3. By Equation 3, entry (i, j) of the Hessian is the derivative with respect to x_i of the derivative with respect to x_j. The code differentiates `f` symbolically in that order with `D()`, evaluates the four entries at $\underset{\sim}{x} = (0.5, -1)$, and compares them with central differences of the gradient.

```r
f_expr <- quote(exp(2 * x1 + x2) - x1)
vars <- c("x1", "x2")
point <- c(x1 = 0.5, x2 = -1)

grad_expr <- lapply(vars, function(v) D(f_expr, v))
hess_expr <- lapply(vars, function(i) {
  lapply(grad_expr, function(g_j) D(g_j, i))
})
H <- sapply(1:2, function(j) {
  sapply(1:2, function(i) eval(hess_expr[[i]][[j]], as.list(point)))
})

grad_at <- function(x) {
  sapply(grad_expr, function(g) eval(g, as.list(setNames(x, vars))))
}
h <- 1e-5
H_fd <- sapply(1:2, function(j) {
  sapply(1:2, function(i) {
    step <- replace(c(0, 0), i, h)
    (grad_at(point + step)[j] - grad_at(point - step)[j]) / (2 * h)
  })
})

H
#>      [,1] [,2]
#> [1,]    4    2
#> [2,]    2    1
H_fd
#>      [,1] [,2]
#> [1,]    4    2
#> [2,]    2    1
```

At this point $u = 2x_1 + x_2 = 0$, so the closed form in Example 0.3 gives $4e^u = 4$ for entry $(1, 1)$, and the code gives 4. The off-diagonal entries are 2 and 2, which are equal. The finite differences agree with the symbolic entries to within 0.00000000026.

Videos

Jon Krohn's "Calculus for Machine Learning" YouTube playlist has videos on second and higher partial derivatives:

- Higher-Order Partial Derivatives
- Exercise on Higher-Order Partial Derivatives

Definition 0.7 (Continuity at a point in several variables)**.** A function f from $\mathbb{R}^p$ to $\mathbb{R}^q$ is **continuous at** $\underset{\sim}{x}$ if for every $\varepsilon > 0$ there is a $\delta > 0$ such that $\|f(\underset{\sim}{y}) - f(\underset{\sim}{x})\| < \varepsilon$ whenever $\|\underset{\sim}{y} - \underset{\sim}{x}\| < \delta$. For $q = 1$, $\|f(\underset{\sim}{y}) - f(\underset{\sim}{x})\| = |f(\underset{\sim}{y}) - f(\underset{\sim}{x})|$.

Definition 0.8 (Continuity on a set in several variables)**.** A function f from $\mathbb{R}^p$ to $\mathbb{R}^q$ is **continuous on** a set if it is continuous at every point of the set (Definition 0.7).

Example 0.6 (A continuous function, and a discontinuous one)**.**

- $f(\underset{\sim}{x}) = x_1 + x_2$ is continuous at every $\underset{\sim}{x}$. Each $|y_i - x_i| \le \|\underset{\sim}{y} - \underset{\sim}{x}\|$, since the squared length is a sum of nonnegative squares, so

$$\begin{aligned} |f(\underset{\sim}{y}) - f(\underset{\sim}{x})| &= |(y_1 - x_1) + (y_2 - x_2)| && \text{(subtract)} \\ &\le |y_1 - x_1| + |y_2 - x_2| && \text{(triangle inequality for numbers)} \\ &\le 2\,\|\underset{\sim}{y} - \underset{\sim}{x}\|, && \text{(each term is at most } \|\underset{\sim}{y} - \underset{\sim}{x}\|) \end{aligned}$$

  which is less than ε whenever $\|\underset{\sim}{y} - \underset{\sim}{x}\| < \delta = \varepsilon/2$.

- $f(\underset{\sim}{x}) = 1$ if $x_1 > 0$ and $f(\underset{\sim}{x}) = 0$ otherwise is not continuous at $\underset{\sim}{0}$: for $\varepsilon = \frac{1}{2}$ and any $\delta > 0$, the point $\underset{\sim}{y} = (\delta/2, 0)^\top$ has $\|\underset{\sim}{y} - \underset{\sim}{0}\| = \delta/2 < \delta$ but $|f(\underset{\sim}{y}) - f(\underset{\sim}{0})| = 1 > \frac{1}{2}$.

Theorem 0.2 (Continuity in one variable, and sums, multiples and compositions)**.**

1. *For $p = q = 1$, a function f is continuous at x in the sense of Definition 0.7 exactly when $\lim_{y \to x} f(y) = f(x)$ (limit), which is the usual continuity at x.*
2. *If $f : \mathbb{R}^p \to \mathbb{R}^q$ is continuous at $\underset{\sim}{x}$ and $g : \mathbb{R}^q \to \mathbb{R}^r$ is continuous at $f(\underset{\sim}{x})$, then the composition $g \circ f$ is continuous at $\underset{\sim}{x}$.*
3. *If $f, g : \mathbb{R}^p \to \mathbb{R}$ are continuous at $\underset{\sim}{x}$, then $f + g$ is continuous at $\underset{\sim}{x}$.*
4. *If $f : \mathbb{R}^p \to \mathbb{R}$ is continuous at $\underset{\sim}{x}$ and $c \in \mathbb{R}$, then cf is continuous at $\underset{\sim}{x}$.*

If the hypotheses of part 2, 3 or 4 hold at every point of a set, the conclusion holds at every point of that set.

Proof

Proof. Let $\varepsilon > 0$ be given.
Part 1. For $p = q = 1$, $\|f(y) - f(x)\| = |f(y) - f(x)|$ and $\|y - x\| = |y - x|$. Definition 0.7 then asks for a $\delta > 0$ such that $|f(y) - f(x)| < \varepsilon$ whenever $|y - x| < \delta$. The limit definition with $L = f(x)$ asks for the same inequality, but only whenever $0 < |y - x| < \delta$. The two agree, because at $y = x$ the inequality $|f(x) - f(x)| = 0 < \varepsilon$ holds for every δ.
Part 2. Because g is continuous at $f(\underset{\sim}{x})$, there is a $\delta_g > 0$ such that $\|g(\underset{\sim}{z}) - g(f(\underset{\sim}{x}))\| < \varepsilon$ whenever $\|\underset{\sim}{z} - f(\underset{\sim}{x})\| < \delta_g$. Because f is continuous at $\underset{\sim}{x}$, applied with δ_g in the role of ε, there is a $\delta > 0$ such that $\|f(\underset{\sim}{y}) - f(\underset{\sim}{x})\| < \delta_g$ whenever $\|\underset{\sim}{y} - \underset{\sim}{x}\| < \delta$. So if $\|\underset{\sim}{y} - \underset{\sim}{x}\| < \delta$, then $\underset{\sim}{z} = f(\underset{\sim}{y})$ has $\|\underset{\sim}{z} - f(\underset{\sim}{x})\| < \delta_g$, and therefore $\|g(f(\underset{\sim}{y})) - g(f(\underset{\sim}{x}))\| < \varepsilon$.
Part 3. Because f and g are continuous at $\underset{\sim}{x}$, there are $\delta_f > 0$ and $\delta_g > 0$ such that $|f(\underset{\sim}{y}) - f(\underset{\sim}{x})| < \varepsilon/2$ whenever $\|\underset{\sim}{y} - \underset{\sim}{x}\| < \delta_f$, and $|g(\underset{\sim}{y}) - g(\underset{\sim}{x})| < \varepsilon/2$ whenever $\|\underset{\sim}{y} - \underset{\sim}{x}\| < \delta_g$. Let $\delta \stackrel{\text{def}}{=} \min\{\delta_f, \delta_g\}$. If $\|\underset{\sim}{y} - \underset{\sim}{x}\| < \delta$, then

$$\begin{aligned}
|(f+g)(\underset{\sim}{y}) - (f+g)(\underset{\sim}{x})| &= |(f(\underset{\sim}{y}) - f(\underset{\sim}{x})) + (g(\underset{\sim}{y}) - g(\underset{\sim}{x}))| && \text{(regroup)} \\
&\le |f(\underset{\sim}{y}) - f(\underset{\sim}{x})| + |g(\underset{\sim}{y}) - g(\underset{\sim}{x})| && \text{(triangle inequality for numbers)} \\
&< \frac{\varepsilon}{2} + \frac{\varepsilon}{2} && \text{(choice of } \delta) \\
&= \varepsilon. && \text{(add)}
\end{aligned}$$

Part 4. Because f is continuous at $\underset{\sim}{x}$, there is a $\delta > 0$ such that $|f(\underset{\sim}{y}) - f(\underset{\sim}{x})| < \varepsilon/(|c|+1)$ whenever $\|\underset{\sim}{y} - \underset{\sim}{x}\| < \delta$. For such $\underset{\sim}{y}$,

$$\begin{aligned}
|cf(\underset{\sim}{y}) - cf(\underset{\sim}{x})| &= |c|\,|f(\underset{\sim}{y}) - f(\underset{\sim}{x})| && \text{(factor out } c) \\
&\le |c|\,\frac{\varepsilon}{|c|+1} && \text{(choice of } \delta) \\
&< \varepsilon. && (|c| < |c|+1)
\end{aligned}$$

□

Example 0.7 (Choosing δ for a sum, a multiple and a composition)**.** Let $f(\underset{\sim}{x}) = x_1$ and $g(\underset{\sim}{x}) = x_2$ on $\mathbb{R}^2$, and $h(t) = 3t$ on $\mathbb{R}$. Since $|y_i - x_i| \le \|\underset{\sim}{y} - \underset{\sim}{x}\|$, the functions f and g are continuous with $\delta = \varepsilon$, and h is continuous with $\delta = \varepsilon/3$, because $|h(s) - h(t)| = 3\,|s - t|$. Take $\underset{\sim}{x} = (1, 2)$ and $\varepsilon = 0.1$. The proof of Theorem 0.2 gives these values of δ:

- for $f + g$ (part 3), $\delta = \min\{\varepsilon/2, \varepsilon/2\}$;
- for $3f$ (part 4), $\delta = \varepsilon/(3+1)$;
- for $h \circ (f + g)$ (part 2), the outer value $\delta_h = \varepsilon/3$ plays the role of ε for $f + g$, so $\delta = \delta_h/2$.

The code draws many points $\underset{\sim}{y}$ at distance less than δ from $\underset{\sim}{x}$ and finds the largest change

in each function.

```
set.seed(571)
eps <- 0.1
x <- c(1, 2)
f <- function(x) x[1]
g <- function(x) x[2]
h <- function(t) 3 * t

fun_sum <- function(x) f(x) + g(x)
fun_multiple <- function(x) 3 * f(x)
fun_composition <- function(x) h(fun_sum(x))

deltas <- c(
  sum = eps / 2,
  multiple = eps / (3 + 1),
  composition = (eps / 3) / 2
)
funs <- list(
  sum = fun_sum, multiple = fun_multiple, composition = fun_composition
)

n <- 10000L
largest_change <- sapply(names(funs), function(nm) {
  radius <- deltas[[nm]] * sqrt(runif(n))
  angle <- runif(n, 0, 2 * pi)
  y <- cbind(x[1] + radius * cos(angle), x[2] + radius * sin(angle))
  max(abs(apply(y, 1, funs[[nm]]) - funs[[nm]](x)))
})
data.frame(delta = deltas, largest_change = largest_change, eps = eps)
#>                  delta largest_change eps
#> sum         0.05000000     0.07063484 0.1
#> multiple    0.02500000     0.07494539 0.1
#> composition 0.01666667     0.07063336 0.1
```

For 3 of the 3 functions, the largest change among the 10000 points is below $\varepsilon = 0.1$, as Theorem 0.2 guarantees.

Example 0.8 (Partial derivatives can exist where a function is not continuous)**.** Let $f : \mathbb{R}^2 \to \mathbb{R}$ be

$$f(\underset{\sim}{x}) = \begin{cases} \dfrac{x_1 x_2}{x_1^2 + x_2^2} & \text{if } \underset{\sim}{x} \neq \underset{\sim}{0}, \\ 0 & \text{if } \underset{\sim}{x} = \underset{\sim}{0}. \end{cases}$$

Both partial derivatives exist at $\underset{\sim}{0}$. For $h \neq 0$,

$$\begin{aligned} f(h, 0) &= \frac{h \cdot 0}{h^2 + 0} \\ &= 0, \end{aligned}$$

so

$$\begin{aligned} \frac{f(0+h, 0) - f(0,0)}{h} &= \frac{0 - 0}{h} && (f(h,0) = 0 \text{ and } f(0,0) = 0) \\ &= 0, && (0/h = 0 \text{ for } h \neq 0) \end{aligned}$$

and the limit as $h \to 0$ is 0: $\frac{\partial}{\partial x_1} f(\underset{\sim}{0}) = 0$ (Definition 1 in Partial Derivatives and Gradients). The same steps with $f(0, h) = 0$ give $\frac{\partial}{\partial x_2} f(\underset{\sim}{0}) = 0$.

But f is not continuous at $\underset{\sim}{0}$ (Definition 0.7). Take $\varepsilon = \frac{1}{4}$ and any $\delta > 0$, and let $t = \delta/2$ and $\underset{\sim}{y} = (t, t)^\top$. Then

$$\begin{aligned} \|\underset{\sim}{y} - \underset{\sim}{0}\| &= \sqrt{t^2 + t^2} \\ &= t\sqrt{2} \\ &= \delta/\sqrt{2} \\ &< \delta, \end{aligned}$$

but

$$\begin{aligned} |f(\underset{\sim}{y}) - f(\underset{\sim}{0})| &= \left| \frac{t \cdot t}{t^2 + t^2} - 0 \right| && (\underset{\sim}{y} \neq \underset{\sim}{0}, \text{ since } t > 0) \\ &= \frac{t^2}{2t^2} && (t^2 + t^2 = 2t^2) \\ &= \tfrac{1}{2}, && (\text{cancel } t^2 > 0) \end{aligned}$$

which is more than $\varepsilon = \frac{1}{4}$.

The partial derivatives only look along the two coordinate axes, where f is 0; along the line $x_1 = x_2$, f is $\frac{1}{2}$ everywhere except at $\underset{\sim}{0}$. So partial derivatives existing at a point says little about how f behaves near it.

Theorem 0.3 (Symmetry of the Hessian)**.** *If the second partial derivatives of f exist and are continuous (Definition 0.7) on an open ball around $\underset{\sim}{x}$ (Definition 0.1), then $\frac{\partial}{\partial x_i}\left(\frac{\partial}{\partial x_j} f(\underset{\sim}{x})\right) = \frac{\partial}{\partial x_j}\left(\frac{\partial}{\partial x_i} f(\underset{\sim}{x})\right)$ for all i, j [3], so $\mathbf{H}_f(\underset{\sim}{x})$ (Definition 0.6) is symmetric*

(symmetric matrix).

The proof applies the one-variable mean value theorem twice, which these notes do not develop; see [4], Theorem 9.41, which is stated for two variables: apply it to f as a function of x_i and x_j, with the other coordinates held fixed.

Example 0.9 (Mixed partial derivatives agree)**.** In Example 0.3, the $(1,2)$ and $(2,1)$ entries of $\mathbf{H}_f(\underset{\sim}{x})$ are both $2e^{2x_1+x_2}$.

Example 0.10 (Without continuity, the mixed partials can differ)**.** Let $f(x_1,x_2) = \dfrac{x_1x_2(x_1^2-x_2^2)}{x_1^2+x_2^2}$ for $\underset{\sim}{x} \neq \underset{\sim}{0}$, and $f(\underset{\sim}{0}) = 0$. For $x_2 \neq 0$, the partial derivative (Definition 1 in Partial Derivatives and Gradients) in x_1 at $(0,x_2)$ is

$$\begin{aligned}
\frac{\partial}{\partial x_1} f(0,x_2) &= \lim_{h\to 0} \frac{f(h,x_2)-f(0,x_2)}{h} && \text{(definition of the partial derivative, Definition 1 in Partial Derivative} \\
&= \lim_{h\to 0} \frac{x_2(h^2-x_2^2)}{h^2+x_2^2} && (f(0,x_2)=0\text{; cancel } h) \\
&= \frac{x_2\cdot(-x_2^2)}{x_2^2} \\
&= -x_2, && \text{(the quotient is continuous at } h=0)
\end{aligned}$$

and

$$\begin{aligned}
\frac{\partial}{\partial x_1} f(\underset{\sim}{0}) &= \lim_{h\to 0}(0-0)/h \\
&= 0,
\end{aligned}$$

so $\frac{\partial}{\partial x_1} f(0,x_2) = -x_2$ holds at $x_2 = 0$ too. In the same way, with the roles of x_1 and x_2 swapped, $f(x_1,k)/k = x_1(x_1^2-k^2)/(x_1^2+k^2) \to x_1$, so $\frac{\partial}{\partial x_2} f(x_1,0) = x_1$ for every x_1. So at $\underset{\sim}{0}$ (Definition 0.6)

$$
\begin{aligned}
\left[\mathbf{H}_f(\underset{\sim}{0})\right]_{21} &= \frac{\partial}{\partial x_2}\left(\frac{\partial}{\partial x_1} f\right) \\
&= \frac{d}{dx_2}(-x_2) \\
&= -1, \\
\left[\mathbf{H}_f(\underset{\sim}{0})\right]_{12} &= \frac{\partial}{\partial x_1}\left(\frac{\partial}{\partial x_2} f\right) \\
&= \frac{d}{dx_1} x_1 \\
&= 1 :
\end{aligned}
$$

the Hessian at $\underset{\sim}{0}$ is not symmetric. By Theorem 0.3, then, the second partial derivatives of this f cannot all be continuous near $\underset{\sim}{0}$.

Theorem 0.4 (Hessian of a quadratic form)**.** *If* $\mathbf{S}$ *is a symmetric* $p \times p$ *matrix that is constant with respect to* $\underset{\sim}{x}$*, then* $f(\underset{\sim}{x}) = \underset{\sim}{x}^\top \mathbf{S} \underset{\sim}{x}$ *has* $\mathbf{H}_f(\underset{\sim}{x}) = 2\mathbf{S}$ *for every* $\underset{\sim}{x}$*.*

Proof

Proof.

$$
\begin{aligned}
\mathbf{H}_f(\underset{\sim}{x}) &= \frac{\partial}{\partial \underset{\sim}{x}}\left(\frac{\partial}{\partial \underset{\sim}{x}}(\underset{\sim}{x}^\top \mathbf{S} \underset{\sim}{x})\right)^\top && \text{(Definition 0.6)} \\
&= \frac{\partial}{\partial \underset{\sim}{x}}(2\mathbf{S}\underset{\sim}{x})^\top && \text{(Corollary 2 in Jacobians and Derivative Rules)} \\
&= \left(\frac{\partial}{\partial \underset{\sim}{x}} \underset{\sim}{x}\right)(2\mathbf{S})^\top && \text{(Theorem 5 in Jacobians and Derivative Rules, with } \mathbf{A} = 2\mathbf{S} \text{ and } \underset{\sim}{v} = \underset{\sim}{x};\ \frac{\partial}{\partial \underset{\sim}{x}}(2\mathbf{S} \\
&= \mathbf{I}_p\,(2\mathbf{S})^\top && \left(\frac{\partial}{\partial \underset{\sim}{x}} \underset{\sim}{x} = \mathbf{I}_p, \text{ Remark 4 in Jacobians and Derivative Rules}\right) \\
&= 2\mathbf{S}. && (\mathbf{S} \text{ is symmetric})
\end{aligned}
$$

□

Example 0.11 (The Hessian of a 2×2 quadratic form)**.** For $\mathbf{S} = \begin{bmatrix} 3 & 1 \\ 1 & 2 \end{bmatrix}$ as in Example 12 in Jacobians and Derivative Rules, Theorem 0.4 gives $\mathbf{H}_f(\underset{\sim}{x}) = \begin{bmatrix} 6 & 2 \\ 2 & 4 \end{bmatrix}$. Directly, the gradient found there, with β_i renamed x_i, is $(6x_1 + 2x_2,\ 2x_1 + 4x_2)^\top$, and differentiating its entries by x_1 and by x_2 (Definition 0.6) gives

$$\begin{aligned}\mathbf{H}_f(\underset{\sim}{x}) &= \begin{bmatrix} \frac{\partial}{\partial x_1}(6x_1 + 2x_2) & \frac{\partial}{\partial x_1}(2x_1 + 4x_2) \\ \frac{\partial}{\partial x_2}(6x_1 + 2x_2) & \frac{\partial}{\partial x_2}(2x_1 + 4x_2) \end{bmatrix} \\ &= \begin{bmatrix} 6 & 2 \\ 2 & 4 \end{bmatrix}.\end{aligned}$$

[1] Wikipedia contributors, "Ball (mathematics) — Wikipedia, the free encyclopedia." 2026. Available: https://en.wikipedia.org/wiki/Ball_(mathematics)

[2] Wikipedia contributors, "Hessian matrix — Wikipedia, the free encyclopedia." 2026. Available: https://en.wikipedia.org/wiki/Hessian_matrix

[3] Wikipedia contributors, "Symmetry of second derivatives — Wikipedia, the free encyclopedia." 2026. Available: https://en.wikipedia.org/wiki/Symmetry_of_second_derivatives

[4] W. Rudin, *Principles of mathematical analysis*, 3rd ed. in International series in pure and applied mathematics. New York: McGraw-Hill, 1976.