The Hessian

Google Classroom

The Hessian is a matrix that organizes all the second partial derivatives of a function.

Background:

- Second partial derivatives

The Hessian matrix

The "**Hessian matrix**" of a multivariable function $f(x, y, z, \dots)$, which different authors write as $\mathbf{H}(f)$, $\mathbf{H}f$, or $\mathbf{H}_f$, organizes all second partial derivatives into a matrix:

$$\mathbf{H}f = \begin{bmatrix} \dfrac{\partial^2 f}{\partial x^2} & \dfrac{\partial^2 f}{\partial x \partial y} & \dfrac{\partial^2 f}{\partial x \partial z} & \cdots \\ \dfrac{\partial^2 f}{\partial y \partial x} & \dfrac{\partial^2 f}{\partial y^2} & \dfrac{\partial^2 f}{\partial y \partial z} & \cdots \\ \dfrac{\partial^2 f}{\partial z \partial x} & \dfrac{\partial^2 f}{\partial z \partial y} & \dfrac{\partial^2 f}{\partial z^2} & \cdots \\ \vdots & \vdots & \vdots & \ddots \end{bmatrix}$$

So, two things to notice here:

- This only makes sense for scalar-valued function.
- This object $\mathbf{H}f$ is no ordinary matrix; it is a matrix with *functions* as entries. In other words, it is meant to be evaluated at some point $(x_0, y_0, \dots)$.

$$\mathbf{H}f(x_0, y_0, \dots) = \begin{bmatrix} \dfrac{\partial^2 f}{\partial x^2}(x_0, y_0, \dots) & \dfrac{\partial^2 f}{\partial x \partial y}(x_0, y_0, \dots) & \cdots \\ \dfrac{\partial^2 f}{\partial y \partial x}(x_0, y_0, \dots) & \dfrac{\partial^2 f}{\partial y^2}(x_0, y_0, \dots) & \cdots \\ \vdots & \vdots & \ddots \end{bmatrix}$$

As such, you might call this object $\mathbf{H}f$ a "matrix-valued" function. Funky, right?

[Hide explanation]

> In the single variable world, things were so simple. The derivative, second derivative, etc. are all still single variable functions.
>
> Now, when a scalar-valued function f has multiple inputs, the analog of the first derivative is the gradient, ∇f, which is a vector-valued function, and the analog of the second derivative is this Hessian $\mathbf{H}f$, which is a matrix-valued function.
>
> You might wonder if there are analogous ideas for higher order derivatives, third, fourth, etc. There are! And there is even a nice extension of Taylor's formula using them all.
>
> "Really, what are they?", I hear you asking eagerly. Well, expressing them in their full glory involves something called "multilinear algebra". It is both fascinating and beautiful, but unfortunately a bit outside the scope of our current discussion. Patience.

One more important thing, **the word "Hessian" also sometimes refers to the determinant of this matrix,** instead of to the matrix itself.

Example: Computing a Hessian

Problem: Compute the Hessian of $f(x, y) = x^3 - 2xy - y^6$ at the point $(1, 2)$:

Solution: Ultimately we need all the second partial derivatives of f, so let's first compute both partial derivatives:

$$f_x(x, y) = \frac{\partial}{\partial x}(x^3 - 2xy - y^6) = 3x^2 - 2y$$

$$f_y(x, y) = \frac{\partial}{\partial y}(x^3 - 2xy - y^6) = -2x - 6y^5$$

With these, we compute all four second partial derivatives:

$$f_{xx}(x, y) = \frac{\partial}{\partial x}(3x^2 - 2y) = 6x$$

$$f_{xy}(x,y) = \frac{\partial}{\partial y}(3x^2 - 2y) = -2$$

$$f_{yx}(x,y) = \frac{\partial}{\partial x}(-2x - 6y^5) = -2$$

$$f_{yy}(x,y) = \frac{\partial}{\partial y}(-2x - 6y^5) = -30y^4$$

The Hessian matrix in this case is a 2×2 matrix with these functions as entries:

$$\mathbf{H}f(x,y) = \left[\begin{array}{cc} f_{xx}(x,y) & f_{yx}(x,y) \\ f_{xy}(x,y) & f_{yy}(x,y) \end{array} \right] = \left[\begin{array}{cc} 6x & -2 \\ -2 & -30y^4 \end{array} \right]$$

We were asked to evaluate this at the point $(x,y) = (1,2)$, so we plug in these values:

$$\mathbf{H}f(1,2) = \left[\begin{array}{cc} 6(1) & -2 \\ -2 & -30(2)^4 \end{array} \right] = \left[\begin{array}{cc} 6 & -2 \\ -2 & -480 \end{array} \right]$$

Now, the problem is ambiguous, since the "Hessian" can refer either to this matrix or to its determinant. What you want depends on context. For example, in optimizing multivariable functions, there is something called the "second partial derivative test" which uses the Hessian determinant. When the Hessian is used to approximate functions, you just use the matrix itself.

If it's the determinant we want, here's what we get:

$$\det\left(\left[\begin{array}{cc} 6 & -2 \\ -2 & -480 \end{array} \right] \right) = 6(-480) - (-2)(-2) = -2884$$

Uses

By capturing all the second-derivative information of a multivariable function, the Hessian matrix often plays a role analogous to the ordinary second derivative in single variable calculus. Most notably, it arises in these two cases:

- Quadratic approximations of multivariable functions, which is a bit like a second order Taylor expansion, but for multivariable functions.
- The second partial derivative test, which helps you find the maximum/minimum of a multivariable function.