Khan Academy

The Hessian

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:

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

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 $(x_0, y_0, \dots)$.

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

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

[The nature of higher order derivatives in a multivariable world]

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 $f(x, y) = x^3$

Solution: Ultimately we need all the second pa
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 matri
entries:

Math › Multivariable calculus › Applications of multivariable derivatives › Quadratic approximations
Quadratic approximations

Practice: The Hessian matrix

Next lesson
Optimizing multivariable f...

$$\mathbf{H}f(x,y) = \begin{bmatrix} f_{xx}(x,y) & f_{yx}(x,y) \\ f_{xy}(x,y) & f_{yy}(x,y) \end{bmatrix} = \begin{bmatrix} 6x & -2 \\ -2 & -30y^4 \end{bmatrix}$$

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

$$\mathbf{H}f(1,2) = \begin{bmatrix} 6(1) & -2 \\ -2 & -30(2)^4 \end{bmatrix} = \begin{bmatrix} 6 & -2 \\ -2 & -480 \end{bmatrix}$$

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:

$$\det\left(\begin{bmatrix} 6 & -2 \\ -2 & -480 \end{bmatrix}\right) = 6(-480) - (-2)(-2) = -2884$$

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.

will there be videos and exercises (mostly interested in the exercises) for these topics any time soon?

Reply • Comment (28 votes) Upvote Downvote Flag more

arkanmanva 5 years ago

more

Me too. Would love to see exercises in multivariable calculus, differential equations and linear algebra.

Comment (20 votes) Upvote Downvote Flag more

See 1 more reply

Marcel Brown 6 years ago

more

Should the determinant in the final step be: 180xy^4 - 4?

Reply • Comment (7 votes) Upvote Downvote Flag more

Kenz 6 years ago

more

I agree partially with Marcel Brown;
2x2 matrix by ad-bc, in this form bc
However, ab.coefficient = 6*-30 = -

Comment (12 votes)

See 1 more reply

Maria 6 years ago

more

Is the Hessian in any way related to the Jacobian matrix?

Reply • 1 comment (5 votes) Upvote Downvote Flag more

Jo Marino 5 years ago

more

More formally: **H** (f(**x**)) = **J** (∇ f(**x**))^T.
It also relates to the Laplacian as an operator: $\Delta f = \nabla^2 f = \text{trace}(\mathbf{H}(f))$.

Comment (4 votes) Upvote Downvote Flag more

gschex1112 6 years ago

more

Why is the last second partial derivative not -30y^4?

Reply • Comment (5 votes) Upvote Downvote Flag more

Emily H 6 years ago

more

Why is fyx= d/dy(3x^2-2y) and not d/dx(-2x-6y^5)? Wouldn't it be the partial derivative with respect to x of the first partial derivative with respect to y? I ask the same for fxy= d/dx(-2x-6y^5) not being d

Reply • Comment (3 votes)

Tejas 6 years ago

Are you a student or a teacher?

Student Teacher

It is both! Whether you derivate with respect to x first then y, or with respect to y first then x, you get the same answer. Notice here that fxy = fyx = -2. That is Clairaut's theorem.

Comment (3 votes) Upvote Downvote Flag more

yogesh kumar 6 years ago

more

why hessian makes sense only for a scalar valued function?

Reply • 1 comment (3 votes) Upvote Downvote Flag more

Aaron Hargrove 6 years ago

more

What are some of the practical applications of the determinant of a Hessian matrix?

Reply • Comment (2 votes) Upvote Downvote Flag more

Tejas 6 years ago

more

Evaluating it can tell you whether you are at a maximum, minimum, or a saddle point. It has all the same abilities as a second derivative in a uni-variate function.

Comment (2 votes)

Briahnna A 6 years ago

Around the last paragraph, the determinant is
But, I was wonder why, in my multivariate cal

hessian determinant and f_xx to determine local extrema. I was wondering what was the background or the basis for using the determinant of the Hessian Matrix to decided if critical points are local maximum or local minimums? Basically, what is important about the Hessian Determinant? Thank-You!

Reply • Comment (3 votes) Upvote Downvote Flag more

Tejas 6 years ago

more

We actually don't use the Hessian to determine whether the critical points are local maxima or local minima. We actually use the Hessian to determine whether they are local extrema or saddle points. As for using fxx, it doesn't have to be fxx. You could just as easily use fyy to determine whether the local extremum is a maximum or minimum.

If it is a local minimum, the gradient is pointing away from this point. If it is a local maximum, the gradient is always pointing toward this point. Of course, at all critical points, the gradient is 0. That should mean that the gradient of nearby points would be tangent to the change in the gradient. In other words, fxx and fyy would be high and fxy and fyx would be low.

On the other hand, if the point is a saddle point, then the gradient vectors will all be pointing around the critical point. Therefore at nearby points, the change in the gradient will be orthogonal to the gradient, not tangent. In other words, fxy and fyx would be high and fxx and fyy would be low, or fxx and fyy would Hessian determinant would be nega

Comment (0 votes)

disha.bandy 4 years ago

Since Hessian matrices are only used for trigonometric functions, what can be used to optimise a 3 variable trigonometric function?

Reply • Comment (1 vote) Upvote Downvote Flag more

muhammadahme2019 2 years ago

more

Where it says The Hessian matrix in this case is a 2 x 2 matrix with these functions as entries:
the fxx and fxy are flipped. Cuz under fxx it should be fxy. Then under fyx it should be fyy.

Reply • 1 comment (1 vote) Upvote Downvote Flag more

Show more...

‹ Vector form of multivariable quadratic approximation

Quadratic approximation ›