Why/How does the determinant of the Hessian matrix, combined with the 2nd derivatives, tell us max., min., saddle points? Reasoning behind it?

Ask Question

In my classes, we are taught the following.

If the determinant of the Hessian matrix at the critical point $\det(D^2 f(c)) > 0$ and $f_{xx}(c) > 0$, the function f at c is concave up.

If the determinant of the Hessian matrix at the critical point $\det(D^2 f(c)) > 0$ and $f_{xx}(c) < 0$, the function f at c is concave down.

If the determinant of the Hessian matrix at the critical point $\det(D^2 f(c)) < 0$, the function f at c is a saddle point.

However, the reasoning behind this is never explained. We are never taught WHY or HOW.

I would like to know **why** the determinant of the Hessian matrix, combined with the second derivative at the critical point, contains this information about max., min., and saddle points. I would also like to know **how** this is derived, as I think this would likely go hand-in-hand with why.

Please give clear reasoning behind each step - not just 'this is what it is' without any reasoning.

When researching this topic, I recall reading mentions regarding eigenvectors or eigenvalues, but I honestly cannot remember.

Thank you.

- Do you know what eigenvalues and eigenvectors actually are, or have you merely seen them mentioned? – Ian Oct 26 '16 at 12:36
- @Ian I have just begun studying them. I previously saw them mentioned as being connected to this theory behind the Hessian matrix. – The Pointer Oct 26 '16 at 12:38
- 1
  OK. They're central to the way that this subject actually works. – Ian Oct 26 '16 at 12:39

add a comment |

3 Answers

active oldest votes

Given a smooth function $f : \mathbb{R}^n \to \mathbb{R}$, we can write a second order Taylor expansion in the form:

$$f(x + \Delta x) = f(x) + \nabla f(x)\Delta x + \frac{1}{2}(\Delta x)^t Hf(x)\Delta x + O(|\Delta x|^3)$$

where $\nabla f(x)$ is the gradient of f at x (written as a row vector), $Hf(x)$ is the Hessian matrix of f at x (which is symmetric, of course), and Δx is some small displacement.

Suppose you have have a critical point at $x = a$, then $\nabla f(a) = 0$. Then your Taylor expansion looks like:

$$f(a+\Delta x) = f(a) + \frac{1}{2}(\Delta x)^t Hf(a)\Delta x + O(|\Delta x|^3).$$

Thus, for small displacements Δx, the Hessian tells us how the function behaves around the critical point.

- The Hessian $Hf(a)$ is **positive definite** if and only if $(\Delta x)^t Hf(a)\Delta x > 0$ for $\Delta x \neq 0$. Equivalently, this is true if and only if all the eigenvalues of $Hf(a)$ are positive. Then no matter which direction you move away from the critical point, the value of $f(a+\Delta x)$ grows (for small $|\Delta x|$), so a is a **local minimum**.

- Likewise, the Hessian $Hf(a)$ is **negative definite** if and only if $(\Delta x)^t Hf(a)\Delta x < 0$ for $\Delta x \neq 0$. Equivalently, this is true if and only if all the eigenvalues of $Hf(a)$ are negative. Then no matter which direction you move away from the critical point, the value of $f(a+\Delta x)$ decreases (for small $|\Delta x|$), so a is a **local maximum**.

- Now suppose that the Hessian $Hf(a)$ is **indefinite**, but $(\Delta x)^t Hf(a)\Delta x \neq 0$ for $\Delta x \neq 0$. Equivalently, this is the same thing as saying that $Hf(a)$ has mixed positive and negative (but all nonzero) eigenvalues. Then (for small $|\Delta x|$) the value of $f(a+\Delta x)$ decreases or increases as you move away from the critical point, depending on which direction you take, so a is a **saddle point**.

- Lastly, suppose that there exists some $\Delta x \neq 0$ such that $(\Delta x)^t Hf(a)\Delta x = 0$. This is true if and only if $Hf(a)$ has a 0 eigenvalue. In this case the test fails: along this direction we aren't really sure whether the function f is increasing or decreasing as we move away from a; our second order approximation isn't good enough and we need higher order data to decide.

What I've described for you here is the intuition for the general situation on $\mathbb{R}^n$, but since it seems like you're working in $\mathbb{R}^2$, the test becomes a bit simpler. In $\mathbb{R}^2$ we can only have two (possibly identical) eigenvalues λ_1 and λ_2 for $Hf(a)$, since it is a 2×2 matrix. We can take advantage of the fact that the determinant of a matrix is the product of the eigenvalues, and the trace is their sum: $\det(Hf(a)) = \lambda_1\lambda_2$ and $\mathrm{tr}(Hf(a)) = \lambda_1 + \lambda_2$.

In this situation:

1. $\det(Hf(a)) = 0$ means that there is a zero eigenvalue and so the test fails.

2. $\det(Hf(a)) < 0$ means that both eigenvalues have different sign, so we have a saddle point at a.

3. $\det(Hf(a)) > 0$ means that both eigenvalues have the same sign: either both positive or both negative, and we must use the trace to decide which it is. In fact, rather than use the trace, it actually suffices to just use the top left entry $\frac{\partial^2 f}{\partial x^2}(a)$ of $Hf(a)$ to decide, by Sylvester's criterion. In other words, $\frac{\partial^2 f}{\partial x^2}(a) > 0$ means both eigenvalues are positive (local min at a a), whereas $\frac{\partial^2 f}{\partial x^2}(a) < 0$ means both eigenvalues are negative (local max at a).

share|cite|improve this answer
edited Oct 26 '16 at 12:50
answered Oct 26 '16 at 12:16

ಠ ಠ

5,41721242

- Thanks. What is the $O(|\Delta x|^3)$ term? I'm not sure what the O is supposed to represent. – The Pointer Oct 26 '16 at 17:04
- 1

It's big O notation. It's not a great notation, in my opinion, but it tells you rough size of the error in the second order Taylor approximation. – ಠ ಠ Oct 26 '16 at 20:27

- At a saddle point, there are directions where f neither increases nor decreases to second order. Consider xy on the axes. Please check your third and fourth bullet points. A zero eigenvalue is not the same thing as having v^t orthogonal to Hv. The latter is guaranteed at saddle points. – Douglas Zare Dec 10 '17 at 6:15

add a comment |

It rests on on the multivariable Taylor's formula:

$$f(x) = f(c) + Df(c)\cdot(x-c) + \frac{1}{2!}(D^2f)_{x=c}(x-c) + o(\|x-c\|^2)$$

For a critical pont, $Df(c) = 0$, and the formula can be written as

$$f(x) - f(c) = \frac{1}{2!}(D^2f)_{x=c}(x-c) + o(\|x-c\|^2)$$

where $(D^2f)_{x=c}$ denotes the quadratic form associated with the hessian matrix at $x = c$.

When $\|x-c\|$ is small enough and the quadratic form is not 0, the sign of this expression is the sign of the quadratic form. Hence if the quadratic form is positive, $f(x) - f(c) > 0$, corresponding to $f(c)$ being a local minimum, if it is negative, it corresponds to a local maximum.

For two variables, setting $x - c = (h, k)$, the quadratic form is, explicitly:

$$\frac{\partial^2 f}{\partial x^2}(c)h^2 + 2\frac{\partial^2 f}{\partial x \partial y}(c)hk + \frac{\partial^2 f}{\partial y^2}(c)k^2$$

This is a homogeneous quadratic form in two variables, and id its sign is the same as the sign of the dehomogenised quadratic form in one variable (setting $t = h/k$):

$$\frac{\partial^2 f}{\partial x^2}(c)t^2 + 2\frac{\partial^2 f}{\partial x \partial y}(c)t + \frac{\partial^2 f}{\partial y^2}(c)$$

Now a non-zero quadratic polynomial has constant sign if and only if its (reduced) discriminant is `negative`, whence the condition for an extremum:

$$\left[\left(\frac{\partial^2 f}{\partial x^2}\right)^2 - \frac{\partial^2 f}{\partial x \partial y}\right](c) < 0.$$

Furthermore, under these circumstances, the sign of the quadratic polynomial is the sign of its leading coefficient.Hence it is

- positive (*local minimum*) if $\frac{\partial^2 f}{\partial x^2}(c) < 0$,
- negative (*local maximum*) if $\frac{\partial^2 f}{\partial x^2}(c) > 0$.

share|cite|improve this answer
edited Oct 26 '16 at 12:50
answered Oct 26 '16 at 12:27

Bernard

118k639112

- $D^2f(x-c)^2$ is kind of a weird way to write it, since we don't multiply vectors. I'd prefer to write either $(x-c)^T D^2 f(x-c)$ (thinking of $D^2 f$ as a matrix) or $D^2 f(x-c, x-c)$ (thinking of $D^2 f$ as a tensor, i.e. a multilinear scalar-valued function). – Ian Oct 26 '16 at 12:38
- @Ian: I agree it's weird, but I specified it's a quadratic form. I tried to use a compact notation, but the problem is the `(c)`. Finally I changed it to $(D^2 f)_{x=c}(x-c)$, which is mathematically correct. – Bernard Oct 26 '16 at 12:51
- I think you meant to write $[\frac{\partial^2 f}{\partial x \partial y} - \frac{\partial^2 f}{\partial x^2}\frac{\partial^2 f}{\partial y^2}](c) < 0$, no? Since that's the reduced discriminant of the polynomial of variable t? Amazing answer, btw! – Striker Aug 19 '18 at 20:18

add a comment |

You can see it in this way. Determinant is the product of all eigenvalues of the Hessian matrix (2 eigenvalues, in the case of two variables). Then checking the sign of determinant is sufficient to tell the sign of eigenvalues, which is a more general way to test the min/max points.

FYI: wiki.

share|cite|improve this answer
answered Apr 22 '18 at 9:44

Jan Fan
111

add a comment |