

## Test for Local Minimum/Maximum (Elaborating on Strang 6.5)

### 1. Approximation of a function<sup>1</sup>

- (a) First order (linear) approximation

The linearization of a function of one variable  $f(x)$  can be written as

$$L(x) = f(a) + f'(a)(x - a) \quad (1)$$

The graph of  $L$  is the tangent line to the curve  $z = f(x)$  at  $(a, f(a))$ , and  $f(x) \approx L(x)$  around  $x = a$ . This is called the first degree Taylor polynomial of  $f$  at  $x = a$ . This can be extended to a function  $g$  of two variables at a point  $(a, b)$ :

$$L(x, y) = g(a, b) + f_x(a, b)(x - a) + f_y(a, b)(y - a) \quad (2)$$

The graph of  $L$  is the tangent plane to the surface  $z = g(x, y)$  at  $(a, b, g(a, b))$ , and  $g(x, y) \approx L(x, y)$  around  $(a, b)$ .

- (b) Second order (quadratic) approximation

If  $g$  has continuous second-order partial derivatives at  $(a, b)$ , then the second-degree Taylor polynomial of  $g$  at  $(a, b)$  is

$$Q(x, y) = g(a, b) + g_x(a, b)(x - a) + g_y(a, b)(y - a) \quad (3)$$

$$+ \frac{1}{2}g_{xx}(a, b)(x - a)^2 + g_{xy}(a, b)(x - a)(y - b) + \frac{1}{2}g_{yy}(a, b)(y - b)^2$$

Verify that (1)  $Q(a, b) = f(a, b)$ , (2)  $Q$  has the same first- and second- order partial derivatives as  $g$  at  $(a, b)$ .

Therefore, at a small  $\Delta x$  around  $(a, b)$ :

$$\begin{aligned} Q(a + \Delta x, b + \Delta x) &= g(a, b) + g_x(a, b)\Delta x + g_y(a, b)\Delta x \\ &\quad + \frac{1}{2}g_{xx}(a, b)\Delta x^2 + g_{xy}(a, b)\Delta x\Delta y + \frac{1}{2}g_{yy}(a, b)\Delta y^2 \end{aligned} \quad (4)$$

This can be written in a more compact form, which also suits any function from  $R^n$  to  $R^n$  with the input vectors  $\mathbf{x}$  and  $\Delta \mathbf{x} \in R^n$ :

$$g(\mathbf{x} + \Delta \mathbf{x}) \approx Q(\mathbf{x} + \Delta \mathbf{x}) = g(\mathbf{x}) + \nabla g(\mathbf{x})\Delta \mathbf{x} + \frac{1}{2}\Delta \mathbf{x}^T H(g(\mathbf{x}))\Delta \mathbf{x} \quad (5)$$

where  $H(g(\mathbf{x}))$  is called the Hessian matrix of  $g$  at  $\mathbf{x}$ .

$$\mathbf{H} = \begin{bmatrix} \frac{\partial^2 f}{\partial x_1^2} & \frac{\partial^2 f}{\partial x_1 \partial x_2} & \cdots & \frac{\partial^2 f}{\partial x_1 \partial x_n} \\ \frac{\partial^2 f}{\partial x_2 \partial x_1} & \frac{\partial^2 f}{\partial x_2^2} & \cdots & \frac{\partial^2 f}{\partial x_2 \partial x_n} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial^2 f}{\partial x_n \partial x_1} & \frac{\partial^2 f}{\partial x_n \partial x_2} & \cdots & \frac{\partial^2 f}{\partial x_n^2} \end{bmatrix}.$$

---

<sup>1</sup>I mainly refer from the Calculus textbook Stewart 6E

For a function in  $R^2$ , it is simply  $H = \begin{bmatrix} \frac{\partial^2 f}{\partial x^2} & \frac{\partial^2 f}{\partial x \partial y} \\ \frac{\partial^2 f}{\partial x \partial y} & \frac{\partial^2 f}{\partial y^2} \end{bmatrix}$ . This is the matrix in your textbook.

## 2. Test for local min/max at the point $\mathbf{x}$ .

The test here is based on the quadratic approximation of a function reviewed above. In calculus, you probably have learned it as the Second Derivative Test. Now with the concept of positive definiteness in linear algebra, you know why the test works.<sup>2</sup>

First, you should find the critical points where the first derivatives are all zero. This will lead to  $\nabla g = \mathbf{0}$ . Therefore, suppose  $\mathbf{x}$  is a critical point, the equation 5 is now simply:

$$g(\mathbf{x} + \Delta\mathbf{x}) \approx Q(\mathbf{x} + \Delta\mathbf{x}) = g(\mathbf{x}) + \frac{1}{2}\Delta\mathbf{x}^T H(g(\mathbf{x}))\Delta\mathbf{x} \quad (6)$$

Thus, for small displacements  $\Delta\mathbf{x}$ , the Hessian tells us how the function behaves around the critical point  $\mathbf{x}$ .

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

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

Now suppose that the Hessian  $H(g(\mathbf{x}))$  is indefinite, but  $\Delta\mathbf{x}^T H(g(\mathbf{x}))\Delta\mathbf{x} \neq 0$  for  $\Delta\mathbf{x} \neq \mathbf{0}$ . Equivalently, this is the same thing as saying that  $H(g(\mathbf{x}))$  has mixed positive and negative (but all nonzero) eigenvalues. Then (for small  $|\Delta\mathbf{x}|$ ) the value of  $g(\mathbf{x} + \Delta\mathbf{x})$  decreases or increases as you move away from the critical point, depending on which direction you take, so  $\mathbf{x}$  is a saddle point. (Take a look at equation 3 of the textbook 6.5 if you wonder what it is to do with eigenvalues. Just think about change of basis.)

Lastly, suppose that there exists some  $\Delta\mathbf{x} \neq \mathbf{0}$  such that  $\Delta\mathbf{x}^T H(g(\mathbf{x}))\Delta\mathbf{x} = 0$ . This is true if and only if  $H(g(\mathbf{x}))$  has a eigenvalue that is equal to 0. In this case the test fails: along this direction we aren't really sure whether the function  $g$  is increasing or decreasing as we move away from  $\mathbf{x}$ ; our second order approximation isn't good enough and we need higher order approximation terms to decide. Or, you can use other algebraic methods to find out. The example included is one such case.

What I've described for you here is the intuition for the general situation on  $R^n$ , but in  $R^2$ , the test becomes a bit simpler. In  $R^2$  we can only have two (possibly identical) eigenvalues  $\lambda_1$  and  $\lambda_2$  for  $H(g(\mathbf{x}))$ , since it is a 2 by 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.

In the  $R^2$  case:

- (a)  $\det(H(g(\mathbf{x}))) = 0$  means that there is a zero eigenvalue and so the test fails.
- (b)  $\det(H(g(\mathbf{x}))) < 0$  means that both eigenvalues are of different signs, so we have a saddle point at  $\mathbf{x}$ .

---

<sup>2</sup>Credit goes to: <https://math.stackexchange.com/questions/1985889/why-how-does-the-determinant-of-the-hessian-matrix-combined-with-the-2nd-deriva>

(c)  $\det(H(g(\mathbf{x}))) > 0$  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 using the trace, it actually suffices to just use the top left entry  $\frac{\partial^2 g}{\partial x^2}$  of  $H(g(\mathbf{x}))$  to decide.

In other words,  $\frac{\partial^2 g}{\partial x^2} > 0$  means both eigenvalues are positive (local min at a  $\mathbf{x}$ ), whereas  $\frac{\partial^2 g}{\partial x^2} < 0$  means both eigenvalues are negative (local max at  $\mathbf{x}$ ).

This is probably the form you were taught in a calculus course.

### 3. Example

**Question:**  $f(x, y) = \frac{1}{4}x^4 + x^2y + y^2$ . Find the critical point(s) and do the second derivative test.

**Solution:** We first find the critical points by setting first derivatives to 0: we need  $x^2 + 2y = 0$ . This gives the critical points.

Hessian matrix at points  $(x, -\frac{1}{2}x^2)$ ,  $H = \begin{bmatrix} 2x^2 & 2x \\ 2x & 2 \end{bmatrix}$ . We can use the determinant tests to see if  $H$  is positive definite or not. We find that  $\det(H) = 0$ , thus our second derivative test failed to tell us if  $(x, -\frac{1}{2}x^2)$  are local min/max.

However, we notice that the function  $f$  can be written as  $f = (\frac{1}{2}x^2 + y)^2$ .  $f = (\frac{1}{2}x^2 + y)^2 \geq 0$  and is equal to 0 when  $x^2 + 2y = 0$ . This shows that the points on the curve  $x^2 + 2y = 0$  are local minimums, and they are actually also global minimums.