

Unconstrained Optimization

Philip Pennance¹

March 2022.

1. Introduction.

Recall that in the calculus of a single variable, the quadratic term in a Taylor Series may be used to check critical points. In the case of a twice continuously differentiable function f of one variable, Taylor's theorem states that

$$f(p+h) = f(p) + f'(p)h + \frac{f''(p)}{2}h^2 + R_2(h)$$

where the remainder R_2 satisfies.

$$\lim_{h \rightarrow 0} \frac{R_2(h)}{h^2} = 0 \quad (1)$$

At a critical point, $f'(p) = 0$ and

$$f(p+h) = f(p) + \frac{f''(p)}{2}h^2 + R_2(h)$$

It follows from (1), that for sufficiently small h , $f''(p)$ and $f(p+h) - f(p)$ have the same sign. Hence

(a) $f''(p) > 0$ indicates a local minimum.

(b) $f''(p) < 0$ indicates a local maximum.

If $f''(p) = 0$ at a critical point and $f^{(3)}(p) \neq 0$ then Taylor's theorem yields

$$f(p+h) = f(p) + \frac{f^{(3)}(p)}{3!}h^3 + O(h^4)$$

In this case it is easily seen that $x = a$ is neither a maximum or minimum.

The use of a Taylor expansion to evaluate critical points extends to the case of more than one variable. In this case the quadratic term $\frac{f''(p)}{2}h^2$ is replaced by a quadratic form, called the *Hessian*, the determinant of which is used to examine the nature of the critical point. Details now follow.

2. Definition.

Let $U \subseteq \mathbb{R}^n$ be an open set and

$$f : U \rightarrow \mathbb{R}.$$

A point $p \in U$ is:

(a) a *critical point* if $\nabla f(p) = 0$.

(b) a *local maximum* of f if

$$f(X) \leq f(p)$$

for all X in some open ball with center p and contained in U .

(c) a *local minimum* of f if

$$f(X) \geq f(p)$$

for all X in some open ball with center p and contained in U .

3. Claim.

Let U be open and $f : U \rightarrow \mathbb{R}$ be differentiable. If p is a local maximum or minimum, then p is a critical point.

Proof.

Let $h \in \mathbb{R}^n$ be a non zero vector. Since U is open

$$p + th \in U$$

for all $t \in \mathbb{R}$ small enough. Define

$$g(t) = f(p + th)$$

If p is a local maximum, then for all t sufficiently small

$$f(p + th) \leq f(p)$$

I.e.,

$$g(t) \leq g(0)$$

Hence 0 is a local maximum of g and so $g'(0) = 0$. But

$$\begin{aligned} g'(t) &= \frac{d}{dt}f(p + th) \\ &= \langle \nabla f(p + th), h \rangle \\ &= 0. \end{aligned}$$

¹<https://pennance.us>

Putting $t = 0$ yields

$$\langle \nabla f(p), h \rangle = 0.$$

Since this holds for all vectors h it must be (exercise) that

$$\nabla f(p) = O.$$

Hence p is a critical point. The proof in the case that p is a minimum is similar.

4. Example.

Let

$$f(x, y) = e^{-(x^2+y^2)}.$$

Then

$$\nabla f(x, y) = (-2x, -2y)e^{-(x^2+y^2)}$$

The only critical point is $(0, 0)$ which is clearly a maximum.

5. Definition.

A *quadratic form* is a homogeneous quadratic polynomial in n variables.

6. Examples.

The following are quadratic forms in 1, 2, 3 variables respectively.

$$(a) \quad g(x) = 4x^2$$

$$(b) \quad g(x, y) = 4x^2 + 2xy - 3y^2$$

$$(c) \quad g(x, y, z) = 4x^2 + 2xy - 3y^2 + 7z^2$$

7. Matrix Representation.

Let g be a quadratic form in two variables. Then

$$\begin{aligned} g(h_1, h_2) &= ah_1^2 + 2bh_1h_2 + ch_2^2 \\ &= \begin{pmatrix} h_1 & h_2 \end{pmatrix} \begin{pmatrix} a & b \\ b & a \end{pmatrix} \begin{pmatrix} h_1 \\ h_2 \end{pmatrix} \\ &= \mathbf{h}^t \mathbf{A} \mathbf{h}, \end{aligned}$$

$$\text{where } \mathbf{h} = \begin{pmatrix} h_1 \\ h_2 \end{pmatrix} \text{ and } \mathbf{A} = \begin{pmatrix} a & b \\ b & a \end{pmatrix}.$$

8. Definition.

A quadratic form $g : \mathbb{R}^2 \rightarrow \mathbb{R}$ is

(a) *positive definite* if

$$\begin{aligned} g(h) &\geq 0, \quad \forall h \in \mathbb{R}^2 \text{ and} \\ g(h) &= 0 \iff h = O. \end{aligned}$$

(b) *negative definite* if

$$\begin{aligned} g(h) &\leq 0, \quad \forall h \in \mathbb{R}^2 \text{ and} \\ g(h) &= 0 \iff h = O. \end{aligned}$$

9. Lemma. Let g be a positive definite quadratic form on $\mathbb{R}^2$. Then there is a constant $m > 0$ such that

$$g(h) \geq m\|h\|^2. \quad (2)$$

Proof.

Recall that on a closed and bounded set, a continuous function has both a maximum and a minimum. It follows that g , restricted to the unit circle has a minimum m . Hence, for all $h \in \mathbb{R}^2$,

$$\begin{aligned} g(h) &= g\left(\|h\| \cdot \frac{h}{\|h\|}\right) \\ &= \|h\|^2 \cdot g\left(\frac{h}{\|h\|}\right) \\ &\geq m\|h\|^2. \end{aligned}$$

10. Claim.

Let $g : \mathbb{R}^2 \rightarrow \mathbb{R}$ be the quadratic form given by

$$g(x, y) = ax^2 + 2bxy + cy^2$$

where $a \neq 0$. Let $A = \begin{pmatrix} a & b \\ b & c \end{pmatrix}$. The following are equivalent.

(a) g is positive definite.

(b) $\det A > 0$ and $a > 0$.

Proof.

Suppose that $\det A > 0$ and $a > 0$. Then

$$\begin{aligned} g(x, y) &= a \left[x^2 + \frac{2by}{a} x \frac{c}{a} y^2 \right] \\ &= a \left[\left(x + \frac{by}{a} \right)^2 - \left(\frac{by}{a} \right)^2 + \frac{c}{a} y^2 \right] \\ &= a \left(x + \frac{by}{a} \right)^2 + \frac{y^2}{a} (ac - b^2) \\ &= a \left(x + \frac{by}{a} \right)^2 + \frac{y^2}{a} \det A \\ &\geq 0 \end{aligned}$$

with equality if and only if $x = y = 0$.

Conversely suppose that g is positive definite. Then $a = g(1, 0) > 0$. Notice that the term $\left(x + \frac{by}{a} \right)^2$ is zero along the line $ax + by = 0$. Choice of the point $(-b, a)$ on this line yields

$$g(-b, a) = a \det A,$$

Since $g(-b, a) > 0$, it must be that a and $\det A$ have the same sign, which must be positive since $a > 0$.

11. Remark.

In a similar manner, it may be shown that g is negative definite if and only if $\det A > 0$ and $a < 0$.

12. Claim. (Taylor's Theorem) Let $U \subseteq \mathbb{R}^2$ be open and $f \in \mathcal{C}^{(3)}(U)$, $p \in U$. Then, for $\vec{h} \in \mathbb{R}^2$ sufficiently small

$$f(p + \vec{h}) = f(p) + (\vec{h} \cdot \nabla) f(p) + \frac{1}{2!} (\vec{h} \cdot \nabla)^2 f(p) + R_2. \quad (3)$$

where

$$\lim_{\|\vec{h}\| \rightarrow 0} \frac{R_2}{\|\vec{h}\|^2} = 0.$$

Proof. See appendix.

13. Definition.

The quadratic form

$$\begin{aligned} \mathcal{H}_p(\vec{h}) &= (\vec{h} \cdot \nabla)^2 f(p) \\ &= D_1^2 f(p) h_1^2 + D_1 D_2 f(p) h_1 h_2 \\ &\quad + D_2 D_1 f(p) h_2 h_1 + D_2^2 f(p) h_2^2 \end{aligned}$$

is called the *Hessian*.

The matrix of $\mathcal{H}_p(h)$,

$$\begin{pmatrix} D_1^2 f(p) & D_1 D_2 f(p) \\ D_2 D_1 f(p) & D_2^2 f(p) \end{pmatrix}$$

is called the *Hessian matrix* and its determinant, the *Hessian determinant*.

14. Claim.

Let U be open in $\mathbb{R}^2$ and $f : U \rightarrow \mathbb{R}$ a real valued differentiable function such that all third partial derivatives are continuous. Let $p \in U$ be a critical point of f and let $\mathcal{H}_p$ be the Hessian of f at p .

- (a) If $\mathcal{H}_p$ positive definite, then p is a relative minimum.
- (b) If $\mathcal{H}_p$ negative definite, then p is a relative maximum.

Proof.

Suppose that $\mathcal{H}_p$ is positive definite. Then by (2), there exists a positive constant m such that

$$\mathcal{H}_p(h) \geq m \|h\|^2.$$

Since p is a critical point, $\nabla f(p) = \vec{0}$, and Taylor's theorem yields

$$f(p + h) = f(p) + \frac{1}{2!} \mathcal{H}_p(h) + R_2$$

Hence

$$\begin{aligned} \frac{f(p + h) - f(p)}{\|h\|^2} &= \frac{\mathcal{H}_p(h)}{2\|h\|^2} + \frac{R_2}{\|h\|^2} \\ &\geq \frac{m}{2} + \frac{R_2}{\|h\|^2}. \end{aligned}$$

Since $\lim_{x \rightarrow 0} \frac{R_2}{\|h\|^2} = 0$, it follows, just as in the one dimensional case, that $f(p + h) - f(p) \geq 0$. Hence p is a local minimum. The proof in the case that $\mathcal{H}_p$ is negative definite is similar.

15. Example.

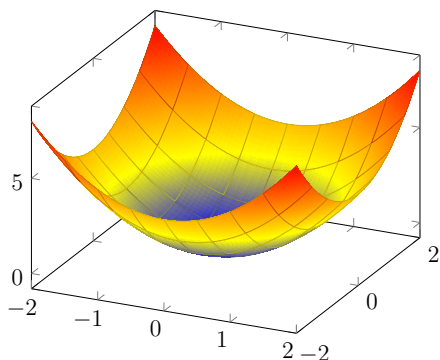
Let $f(x, y) = x^2 + y^2$. Then $\nabla f(x, y) = O$ if and only if $(x, y) = (0, 0)$. Hence $(0, 0)$ is the only critical point. The Hessian

$$\begin{aligned}\mathcal{H}_p(h) &= \begin{pmatrix} D_1^2 f(p) & D_1 D_2 f(p) \\ D_2 D_1 f(p) & D_2^2 f(p) \end{pmatrix} \\ &= \begin{pmatrix} 2 & 0 \\ 0 & 2 \end{pmatrix}\end{aligned}$$

Since

$$\det \mathcal{H}_p(h) > 0, \text{ and } D_1^2 f(O) > 0$$

the Hessian is positive definite and the critical point is a minimum.



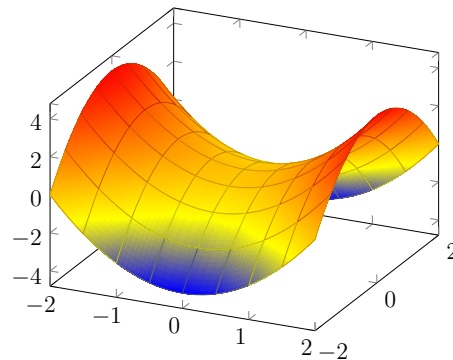
16. Example.

Let $f(x, y) = x^2 - y^2$. Then $\nabla f(x, y) = O$

if and only if $(x, y) = (0, 0)$. Hence $(0, 0)$ is the only critical point. The Hessian

$$\begin{aligned}\mathcal{H}_p(h) &= \begin{pmatrix} D_1^2 f(p) & D_1 D_2 f(p) \\ D_2 D_1 f(p) & D_2^2 f(p) \end{pmatrix} \\ &= \begin{pmatrix} 2 & 0 \\ 0 & -2 \end{pmatrix}\end{aligned}$$

Since $\det \mathcal{H}_p(h) < 0$, the origin is neither a maximum nor a minimum.



17. Definition.

Let U be open in $\mathbb{R}^2$ and $f : U \rightarrow \mathbb{R}$ a real valued differentiable function such that all third partial derivatives are continuous. Let $p \in U$ be a critical point of f . If $\det \mathcal{H}_p < 0$, then p is called a *saddle point*.

Appendix – Taylor’s Theorem in $\mathbb{R}^2$

Philip Pennance² Version: – March 2022.

1. Claim.

Let $U \subseteq \mathbb{R}^2$ be open and $f \in \mathcal{C}^{(3)}(U)$,
 $p \in U$. Then, for $h \in \mathbb{R}^2$ sufficiently small

$$f(p+h) = f(p) + (h \cdot \nabla)f(p) + \frac{1}{2!}(h \cdot \nabla)^2 f(p) + R_2. \quad (4)$$

where

$$\lim_{\|h\| \rightarrow 0} \frac{R_2}{\|h\|^2} = 0.$$

Proof.

Let

$$\gamma(t) = p + th, \quad 0 \leq t \leq 1.$$

and

$$g(t) = f(\gamma(t)).$$

Notice,

$$g(0) = f(p).$$

$$g(1) = f(p+h).$$

$$\begin{aligned} g'(t) &= \nabla f(\gamma(t)) \cdot \gamma'(t) \\ &= h_1 D_1 f(\gamma(t)) + h_2 D_2 f(\gamma(t)). \end{aligned}$$

$$\begin{aligned} g'(0) &= h_1 D_1 f(p) + h_2 D_2 f(p) \\ &= (h_1 D_1 + h_2 D_2) f(p) \\ &= (h \cdot \nabla) f(p). \end{aligned}$$

$$\begin{aligned} g''(t) &= h_1 [D_1 f(\gamma(t))]' + h_2 [D_2 f(\gamma(t))]' \\ &= h_1 \nabla D_1 f(\gamma(t)) \cdot \gamma'(t) \\ &\quad + h_2 \nabla D_2 f(\gamma(t)) \cdot \gamma'(t) \\ &= \sum_{i=1}^2 \sum_{j=1}^2 h_i h_j D_i D_j f(\gamma(t)). \end{aligned}$$

$$\begin{aligned} g''(0) &= (h_1 D_1 + h_2 D_2)^2 f(p) \\ &= (h \cdot \nabla)^2 f(p). \end{aligned}$$

Similarly,

$$g^{(3)}(0) = (h \cdot \nabla)^3 f(p).$$

By Taylor’s theorem with integral form of the remainder

$$g(1) = g(0) + g'(0)t + \frac{g''(0)}{2!}t^2 + R_2,$$

where,

$$R_2 = \int_0^1 \frac{g^{(3)}(t)}{2!} dt.$$

Equation (4) follows on substituting for $g(0)$, $g'(0)$ and $g''(0)$. Finally,

$$\begin{aligned} |R_2| &\leq \frac{1}{2} \int_0^1 (h_1 D_1 + h_2 D_2)^3 f(\gamma(t)) dt \\ &\leq \frac{\|h\|^3}{2} \int_0^1 |(D_1 + D_2)^3 f(\gamma(t))| dt \\ &\leq \frac{\|h\|^3}{2} M. \end{aligned}$$

where

$$M = \max\{|(D_1 + D_2)^3 f(\gamma(t))| : 0 \leq t \leq 1\}.$$

It follows that

$$\lim_{\|h\| \rightarrow 0} \frac{R_2}{\|h\|^2} = 0.$$

2. Remark. If $\det \mathcal{H}_p = 0$, then the nature of p is indeterminate.

²<https://pennance.us> Reference: A Serge Lang, Calculus of Several Variables

Exercises

Philip Penance³

1. Let $f(x, y) = 2x^2 + y^2 - xy - 7y$.
 - (a) Show that $(1, 4)$ is the only critical point of f .
 - (b) Test the critical point by means of the Hessian.
 - (c) Verify your calculation in (b) by examining $f(1 + h, 4 + k) - f(1, 4)$.
2. Compute the second degree Taylor approximation

$$f(p + h) = f(p) + \langle h, \nabla \rangle f(p) + \frac{1}{2!} \langle h, \nabla \rangle^2 f(p)$$

at $p = (2, 1)$ for the function $f(x, y) = 1 + x + 2y^2 + x^3$.

3. Compute the second degree Taylor expansion at $p = (0, 0)$ for the function $f(x, y) = e^x \cos y$.
4. Find the relative maxima, minima and saddle points of the function f defined by

$$f(x, y) = x^3 + 3xy^2 - 3x^2 - 3y^2 + 4.$$

5. Let $f(x, y) = (x - 1)^2 + y^2$.
 - (a) Find the critical points of f and test them by means of the Hessian.
 - (b) Find the Taylor expansion of f of order 2 about the origin.
6. Find the absolute maxima and minima of the function $f(x, y) = x^2 - xy + y^2$ in the rectangle $S = [-1, 1] \times [-1, 1]$.

³<https://penance.us>