Calculus 2502A - Advanced Calculus I
Fall 2014
§14.7: Local minima and maxima

Martin Frankland

November 17, 2014

In these notes, we discuss the problem of finding the local minima and maxima of a function.

1 Background and terminology

Definition 1.1. Let $f: D \to \mathbb{R}$ be a function, with domain $D \subseteq \mathbb{R}^n$. A point $\vec{a} \in D$ is called:

- a **local minimum** of f if $f(\vec{a}) \leq f(\vec{x})$ holds for all $\vec{x}$ in some neighborhood of $\vec{a}$.
- a **local maximum** of f if $f(\vec{a}) \geq f(\vec{x})$ holds for all $\vec{x}$ in some neighborhood of $\vec{a}$.
- a **global minimum** (or *absolute minimum*) of f if $f(\vec{a}) \leq f(\vec{x})$ holds for all $\vec{x} \in D$.
- a **global maximum** (or *absolute maximum*) of f if $f(\vec{a}) \geq f(\vec{x})$ holds for all $\vec{x} \in D$.

An **extremum** means a minimum or a maximum.

Definition 1.2. An interior point $\vec{a} \in D$ is called a:

- **critical point** of f if f is not differentiable at $\vec{a}$ or if the condition $\nabla f(\vec{a}) = \vec{0}$ holds, i.e., all first-order partial derivatives of f are zero at $\vec{a}$.
- **saddle point** of f if it is a critical point which is not a local extremum. In other words, for every neighborhood U of $\vec{a}$, there is some point $\vec{x} \in U$ satisfying $f(\vec{x}) < f(\vec{a})$ and some point $\vec{x} \in U$ satisfying $f(\vec{x}) > f(\vec{a})$.

Remark 1.3. Global minima or maxima are not necessarily unique. For example, the function $f: \mathbb{R} \to \mathbb{R}$ defined by $f(x) = \sin x$ has a global maximum with value 1 at the points:

$$\ldots, -\frac{3\pi}{2}, \frac{\pi}{2}, \frac{5\pi}{2}, \ldots = \left\{ \frac{\pi}{2} + 2k\pi \mid k \in \mathbb{Z} \right\}$$

and a global minimum with value -1 at the points:

$$\ldots, -\frac{\pi}{2}, \frac{3\pi}{2}, \frac{7\pi}{2}, \ldots = \left\{ \frac{3\pi}{2} + 2k\pi \mid k \in \mathbb{Z} \right\}.$$

For a sillier example, consider a *constant* function $g(\vec{x}) = c$ for all $\vec{x} \in D$. Then *every* point $\vec{x} \in D$ is simultaneously a global minimum and a global maximum of g.

Proposition 1.4. *Let $\vec{a}$ be an interior point of the domain D. If $\vec{a}$ is a local extremum of f, then any partial derivative of f which exists at $\vec{a}$ must be zero.*

Proof. Assume that the partial derivative $D_m f(\vec{a})$ exists, for some $1 \leq m \leq n$. Then the function of a single variable

$$g(x) := f(a_1, a_2, \ldots, \overbrace{x}^{m^{\text{th}} \text{ coordinate}}, \ldots, a_{n-1}, a_n)$$

is differentiable at $x = a_m$ and has a local extremum at $x = a_m$. Therefore its derivative vanishes at that point:

$$g'(a_m) = D_m f(\vec{a}) = 0.$$

□

Remark 1.5. The converse does not hold: a critical point of f need not be a local extremum. For example, the function $f(x) = x^3$ has a critical point at $x = 0$ which is a saddle point.

Upshot: In order to find the local extrema of f, it suffices to consider the *critical points* of f.

2 One-dimensional case

Recall the second derivative test from single variable calculus.

Proposition 2.1 (Second derivative test)**.** *Let $f\colon D \to \mathbb{R}$ be a function of a single variable, with domain $D \subseteq \mathbb{R}$. Let $a \in D$ be a critical point of f such that the first derivative f' exists in a neighborhood of a, and the second derivative $f''(a)$ exists.*

- *If $f''(a) > 0$ holds, then f has a local minimum at a.*
- *If $f''(a) < 0$ holds, then f has a local maximum at a.*
- *If $f''(a) = 0$ holds, then the test is inconclusive.*

Remark 2.2. The following examples illustrate why the test is inconclusive when the critical point a satisfies $f''(a) = 0$.

- The function $f(x) = x^4$ has a local minimum at 0.
- The function $f(x) = -x^4$ has a local maximum at 0.
- The function $f(x) = x^3$ has a saddle point at 0.

Example 2.3. Find all local extrema of the function $f\colon \mathbb{R} \to \mathbb{R}$ defined by

$$f(x) = 3x^4 - 4x^3 - 12x^2 + 10.$$

Solution. Note that f is twice differentiable on all of $\mathbb{R}$, and moreover the domain of f is open in $\mathbb{R}$ (so that there are no boundary points to check). Let us find the critical points of f:

$$\begin{aligned}
&f'(x) = 12x^3 - 12x^2 - 24x = 0 \\
\Leftrightarrow\, &x^3 - x^2 - 2x = 0 \\
\Leftrightarrow\, &x\left(x^2 - x - 2\right) = 0 \\
\Leftrightarrow\, &x(x+1)(x-2) = 0 \\
\Leftrightarrow\, &x = -1, 0, \text{ or } 2.
\end{aligned}$$

Let us use the second derivative test to classify these critical points:

$$\begin{aligned}
f''(x) &= 12\left(3x^2 - 2x - 2\right) \\
f''(-1) &= 12\,(3 + 2 - 2) = 12(3) > 0 \\
f''(0) &= 12\,(-2) < 0 \\
f''(2) &= 12\,(12 - 4 - 2) = 12(6) > 0
\end{aligned}$$

from which we conclude:

- $x = -1$ is a $\boxed{\text{local minimum}}$ of f, with value $f(-1) = 3 + 4 - 12 + 10 = \boxed{5}$.
- $x = 0$ is a $\boxed{\text{local maximum}}$ of f, with value $f(0) = \boxed{10}$.
- $x = 2$ is a $\boxed{\text{local minimum}}$ of f, with value $f(2) = 3(16) - 4(8) - 12(4) + 10 = \boxed{-22}$.

Remark 2.4. In this example, the function f has *no global maximum*, because it grows arbitrarily large: $\lim_{x\to+\infty} f(x) = +\infty$.

Moreover, the condition $\lim_{x\to-\infty} f(x) = +\infty$, together with the extreme value theorem, implies that f reaches a global minimum. Therefore, $x = 2$ is in fact the *global minimum* of f, with value $f(2) = -22$, as illustrated in Figure 1.

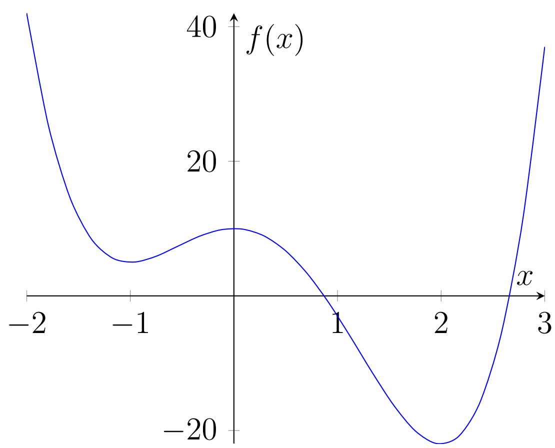

Figure 1: Graph of the function f in Example 2.3.

3 Two-dimensional case

Theorem 3.1 (Second derivatives test)**.** *Let $f\colon D \to \mathbb{R}$ be a function of two variables, with domain $D \subseteq \mathbb{R}^2$. Let $(a,b) \in D$ be a critical point of f such that the second partial derivatives f_{xx}, f_{xy}, f_{yx}, f_{yy} exist in a neighborhood of (a,b) and are continuous at (a,b), which implies in particular $f_{xy}(a,b) = f_{yx}(a,b)$. The* ***discriminant*** *of f is the 2×2 determinant*

$$\begin{aligned} D &:= \begin{vmatrix} f_{xx} & f_{xy} \\ f_{yx} & f_{yy} \end{vmatrix} \\ &= f_{xx} f_{yy} - f_{xy}^2. \end{aligned}$$

Now consider the discriminant $D = D(a,b)$ at the point (a,b).

- *If $D > 0$ and $f_{xx}(a,b) > 0$ hold, then f has a local minimum at (a,b).*
- *If $D > 0$ and $f_{xx}(a,b) < 0$ hold, then f has a local maximum at (a,b).*
- *If $D < 0$ holds, then f has a saddle point at (a,b).*
- *If $D = 0$ holds, then the test is inconclusive.*

Remark 3.2. The following examples illustrate why the test is inconclusive when the discriminant satisfies $D = 0$.

- The function $f(x,y) = x^2 + y^4$ has a (strict) local minimum at $(0,0)$.
- The function $f(x,y) = x^2$ has a (non-strict) local minimum at $(0,0)$.
- The function $f(x,y) = -x^2 - y^4$ has a (strict) local maximum at $(0,0)$.
- The function $f(x,y) = -x^2$ has a (non-strict) local maximum at $(0,0)$.
- The function $f(x,y) = x^2 + y^3$ has a saddle point at $(0,0)$.
- The function $f(x,y) = -x^2 + y^3$ has a saddle point at $(0,0)$.
- The function $f(x,y) = x^2 - y^4$ has a saddle point at $(0,0)$.

Example 3.3 (# 14.7.11)**.** Find all local extrema and saddle points of the function $f\colon \mathbb{R}^2 \to \mathbb{R}$ defined by

$$f(x,y) = x^3 - 12xy + 8y^3.$$

Solution. Note that f is infinitely many times differentiable on its domain $\mathbb{R}^2$. Let us find the critical points of f:

$$\begin{cases} f_x(x,y) = 3x^2 - 12y = 0 \\ f_y(x,y) = -12x + 24y^2 = 0 \end{cases}$$

$$\Leftrightarrow \begin{cases} 4y = x^2 \\ x = 2y^2 \end{cases}$$

$$\Rightarrow 4y = (2y^2)^2 = 4y^4$$

$$\Leftrightarrow y = y^4$$

$$\Leftrightarrow y(y^3 - 1) = 0$$

$$\Leftrightarrow y = 0 \text{ or } y^3 = 1$$

$$\Leftrightarrow y = 0 \text{ or } y = 1$$

$$y = 0 \Rightarrow x = 2(0)^2 = 0$$

$$y = 1 \Rightarrow x = 2(1)^2 = 2$$

which implies that $(0,0)$ and $(2,1)$ are the only points that could be critical points of f, and indeed they are.

Let us compute the discriminant of f:

$$\begin{aligned} f_{xx}(x,y) &= 6x \\ f_{xy}(x,y) &= -12 \\ f_{yy}(x,y) &= 48y \\ D(x,y) &= f_{xx}(x,y)f_{yy}(x,y) - f_{xy}(x,y)^2 \\ &= (6x)(48y) - (-12)^2 \\ &= 12^2\,(2xy - 1)\,. \end{aligned}$$

At the critical point $(0,0)$, the discriminant is:

$$D(0,0) = 12^2(0-1) < 0$$

so that $(0,0)$ is a $\boxed{\text{saddle point}}$.

At the critical point $(2,1)$, the discriminant is:

$$D(2,1) = 12^2(4-1) > 0$$

and we have $f_{xx}(2,1) = 6(2) = 12 > 0$, so that $(2,1)$ is a $\boxed{\text{local minimum}}$ with value $f(2,1) = 8 - 24 + 8 = \boxed{-8}$, as illustrated in Figure 2..

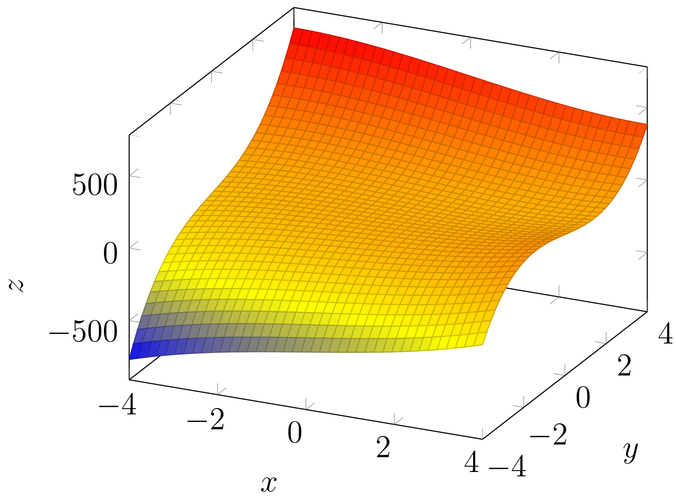

Figure 2: Graph of the function f in Example 3.3.

4 Why does it work?

4.1 Quadratic approximation

The second derivatives test relies on the quadratic approximation of f at the point (a, b), which is the Taylor polynomial of degree 2 at that point.

Recall that for functions of a single variable $f(x)$, the quadratic approximation of f at a is:

$$f(x) \approx f(a) + f'(a)(x-a) + f''(a)\frac{(x-a)^2}{2}.$$

Since the variable of interest is the *displacement* from the basepoint a, let us rewrite $h := x-a$ and thus:

$$f(a+h) \approx f(a) + f'(a)h + f''(a)\frac{h^2}{2}.$$

If a is a critical point of f, then we have $f'(a) = 0$ and the quadratic approximation becomes:

$$f(a+h) \approx f(a) + f''(a)\frac{h^2}{2}.$$

If $f''(a) \neq 0$ holds, then the quadratic term will determine the local behavior of f around a: $f''(a) > 0$ yields a local minimum while $f''(a) < 0$ yields a local maximum.

Idea:

- If $f''(a) > 0$ holds, then f qualitatively behaves like $f(a+h) = f(a) + h^2$ when h is small.
- If $f''(a) < 0$ holds, then f behaves like $f(a+h) = f(a) - h^2$.

For functions of two variables $f(x, y)$, the quadratic approximation at (a, b) is:

$$f(x,y) \approx f(a,b) + f_x(a,b)(x-a) + f_y(a,b)(y-b) + \\ + f_{xx}(a,b)\frac{(x-a)^2}{2} + f_{xy}(a,b)(x-a)(y-b) + f_{yy}(a,b)\frac{(y-b)^2}{2}$$

Again, consider the displacement from the basepoint (a, b) and write

$$\vec{h} := (h,k) := (x-a, y-b)$$

and thus:

$$f(a+h, b+k) \approx f(a,b) + f_x(a,b)h + f_y(a,b)k + f_{xx}(a,b)\frac{h^2}{2} + f_{xy}(a,b)hk + f_{yy}(a,b)\frac{k^2}{2}.$$

In vector notation, which is convenient in higher dimension, this can be rewritten as:

$$f(a+h, b+k) \approx f(a,b) + \nabla f(a,b) \cdot \vec{h} + \frac{1}{2}\vec{h}^T H_f(a,b)\vec{h}$$

where $H_f(a, b)$ is the **Hessian** of f at the point (a, b), defined as the matrix of second partial derivatives:

$$H_f = \begin{bmatrix} f_{xx} & f_{xy} \\ f_{yx} & f_{yy} \end{bmatrix}.$$

Here $\vec{h}^T$ denotes the row vector $\begin{bmatrix} h & k \end{bmatrix}$, which is the *transpose* of the column vector $\vec{h} = \begin{bmatrix} h \\ k \end{bmatrix}$.

If (a, b) is a critical point of f, then we have $f_x(a, b) = 0$ and $f_y(a, b) = 0$, and the quadratic approximation becomes:

$$\begin{aligned} f(a+h, b+k) &\approx f(a,b) + f_{xx}(a,b)\frac{h^2}{2} + f_{xy}(a,b)hk + f_{yy}(a,b)\frac{k^2}{2} \\ &= f(a,b) + \frac{1}{2}\vec{h}^T H_f(a,b)\vec{h}. \end{aligned}$$

4.2 Behavior close to a critical point

The Hessian $H_f(a, b)$ should be viewed as a symmetric bilinear form on the tangent space of the domain D at the point (a, b). Given two direction vectors $\vec{v}$ and $\vec{w}$, the Hessian outputs the number

$$\vec{v}^T H_f(a,b)\vec{w} = D_{\vec{w}} D_{\vec{v}} f(a,b)$$

where $D_{\vec{v}} f(a, b)$ denotes the derivative of f in the (non-normalized) direction $\vec{v}$ at the point (a, b). For example, taking standard basis vectors $\vec{v} = \vec{\imath}$ and $\vec{w} = \vec{\imath}$ yields the number

$$\vec{\imath}^T H_f(a,b)\vec{\imath} = D_{\vec{\imath}} D_{\vec{\imath}} f(a,b) = f_{xx}(a,b)$$

and taking $\vec{v} = \vec{\imath}$ and $\vec{w} = \vec{\jmath}$ yields the number

$$\vec{\imath}^T H_f(a,b)\vec{\jmath} = D_{\vec{\jmath}} D_{\vec{\imath}} f(a,b) = f_{xy}(a,b).$$

From linear algebra, we know that a symmetric bilinear form is classified, up to a change of basis, by its **signature**: how many of its eigenvalues are positive, zero, or negative. Let λ_1, λ_2 be the eigenvalues of $H_f(a,b)$. Depending on the signs of λ_1 and λ_2, the bilinear form $H_f(a,b)$ will be equivalent (up to a change of basis) to one of the following:

- $\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$ if λ_1 and λ_2 are both positive.
- $\begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}$ if one λ_i is positive and the other is negative.
- $\begin{bmatrix} -1 & 0 \\ 0 & -1 \end{bmatrix}$ if λ_1 and λ_2 are both negative.
- $\begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}$ if one λ_i is positive and the other is zero.
- $\begin{bmatrix} -1 & 0 \\ 0 & 0 \end{bmatrix}$ if one λ_i is negative and the other is zero.
- $\begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix}$ if λ_1 and λ_2 are both zero.

Upshot: To analyze the behavior of f close to a critical point (a,b), the first step is to find the signs of the eigenvalues of the Hessian $H_f(a,b)$.

The discriminant of f was defined as the determinant of the Hessian, which is the product of the eigenvalues:

$$D(a,b) = \det H_f(a,b) = \lambda_1 \lambda_2.$$

Definition 4.1. A critical point (a,b) of f is called **non-degenerate** if the Hessian $H_f(a,b)$ is non-degenerate, i.e., has only non-zero eigenvalues. Otherwise, the critical point is called **degenerate**, which means that the Hessian $H_f(a,b)$ has an eigenvalue 0.

Note that a critical point is degenerate if and only if the product of the eigenvalues is zero: $\lambda_1 \lambda_2 = 0$; equivalently, the discriminant is zero: $D(a,b) = 0$.

Example 4.2. For all the functions described in Example 3.2, the origin $(0,0)$ is a degenerate critical point of f. As we have seen, the function f can have all kinds of behaviors close to a degenerate critical point.

In contrast, close to a non-degenerate critical point, the behavior of f is dictated by the signature of the Hessian. If the discriminant satisfies $D(a,b) \neq 0$, then the critical point (a,b) is non-degenerate, and its signature is one of the following:

- $\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$
- $\begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}$
- $\begin{bmatrix} -1 & 0 \\ 0 & -1 \end{bmatrix}$.

The case of one positive eigenvalue and one negative eigenvalue happens if and only if the discriminant is negative: $D = \lambda_1 \lambda_2 < 0$.

In the case $D > 0$, there are still two possible signatures:

- $\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$
- $\begin{bmatrix} -1 & 0 \\ 0 & -1 \end{bmatrix}$.

To distinguish between the two cases, Sylvester's criterion (from linear algebra) tells us that both eigenvalues are positive if and only if the top left entry of the matrix is positive: $f_{xx}(a,b) > 0$.

Remark 4.3. For this last step, one could also use the equivalent condition $f_{yy}(a,b) > 0$, or that the trace of the matrix is positive: $f_{xx}(a,b) + f_{yy}(a,b) = \lambda_1 + \lambda_2 > 0$.

Summary: Here is a reinterpretation of the second derivatives test.

- $D = 0 \Leftrightarrow$ one of the eigenvalues λ_i is zero.
- $D < 0 \Leftrightarrow$ one of the eigenvalues is positive and the other is negative.
- $D > 0$ and $f_{xx} > 0 \Leftrightarrow$ both eigenvalues λ_i are positive.
- $D > 0$ and $f_{xx} < 0 \Leftrightarrow$ both eigenvalues λ_i are negative.

Idea:

- If both eigenvalues are positive, then f qualitatively behaves like $f(a+h, b+k) = f(a,b) + h^2 + k^2$ when h and k are small.
- If one eigenvalue is positive and the other is negative, then f behaves like $f(a+h, b+k) = f(a,b) + h^2 - k^2$.
- If both eigenvalues are negative, then f behaves like $f(a+h, b+k) = f(a,b) - h^2 - k^2$.