Applications of Partial Derivatives
Optimization

Lecture 6: Optimization - Extreme values

Learning goals:

1. What are extreme values?
2. What are necessary conditions for extreme values?
3. How to classify critical points?

Where to find the material?

Corral 2.5

Guichard et friends 14.7

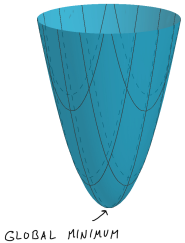
Active Calculus 10.7

Adams-Essex 14.1

Extreme values

$f: D \to \mathbb{R}$, where $D \subset \mathbb{R}^n$, has

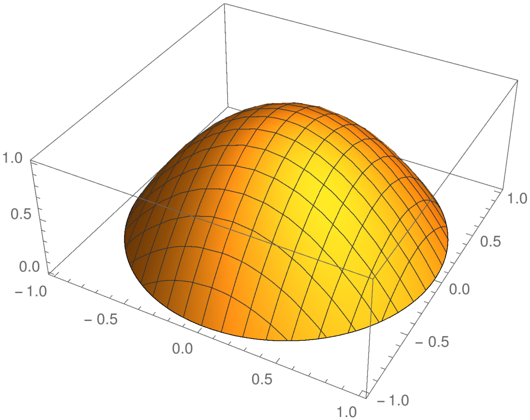
- a **local maximum** at the point $\mathbf{x}_0$ in its domain D if $f(\mathbf{x}) \leq f(\mathbf{x}_0)$ for all points x in the domain D that are *sufficiently close* to the point $\mathbf{x}_0$
- a **global maximum** (or **absolute maximum**) at the point $\mathbf{x}_0$ in its domain D if $f(\mathbf{x}) \leq f(\mathbf{x}_0)$ for all points x in the domain D
- a **local minimum** at the point $\mathbf{x}_0$ in its domain D if $f(\mathbf{x}) \geq f(\mathbf{x}_0)$ for all points x in the domain D that are *sufficiently close* to the point $\mathbf{x}_0$
- a **global minimum** (or **absolute minimum**) at the point $\mathbf{x}_0$ in its domain D if $f(\mathbf{x}) \geq f(\mathbf{x}_0)$ for all points x in the domain D



Necessary conditions for extrema values

- Recall: for single variable $f: I \to \mathbb{R}$ extremas can occur
  1. at the critical points of the function f: i.e. at the points where $f'(x) = 0$,
  2. at points where the derivative is not defined, and
  3. on the edge of the set I
- Next, we generalize the conditions of the function $f: D \subset \mathbb{R}^n \to \mathbb{R}$ to the case of $f: D \subset \mathbb{R}^n \to \mathbb{R}$.
- For multivariable function extrema values can occur:
  1. at the critical points of the function f i.e. at the points where $\nabla f(\mathbf{x}) = \mathbf{0}$,
  2. at points where ∇f is not defined, and
  3. on the boundary of the domain D
- A critical point that is not maximum or minimum is a **saddle point**

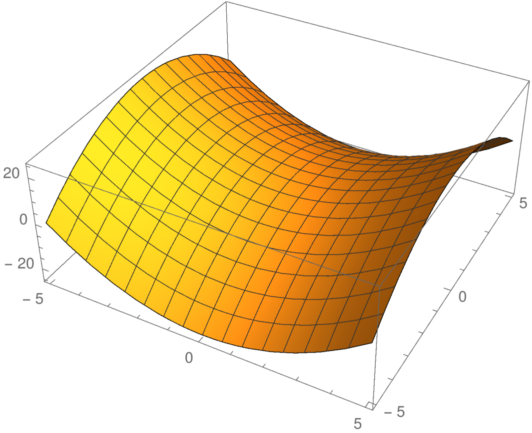
Example 1



The function $f(x, y) = 1 - x^2 - y^2$ has a local maximum $f(0, 0) = 1$ at point $(0, 0)$. This point is a critical point of the function f, because

$$\nabla f(0, 0) = (-2x, -2y)\Big|_{(0,0)} = \mathbf{0}.$$

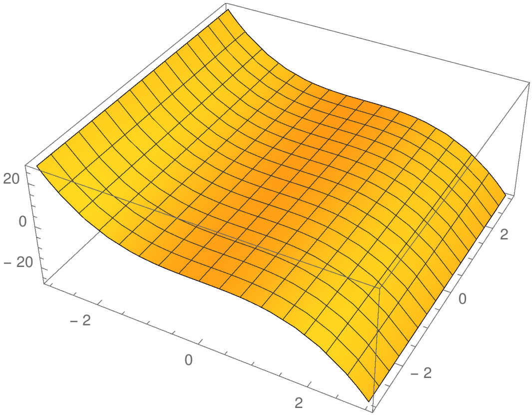
Example 2



Function $f(x, y) = y^2 - x^2$ has a saddle point at $(0, 0)$. The point is a critical point, because

$$\nabla f(0, 0) = (-2x, 2y)\Big|_{(0,0)} = \mathbf{0}.$$

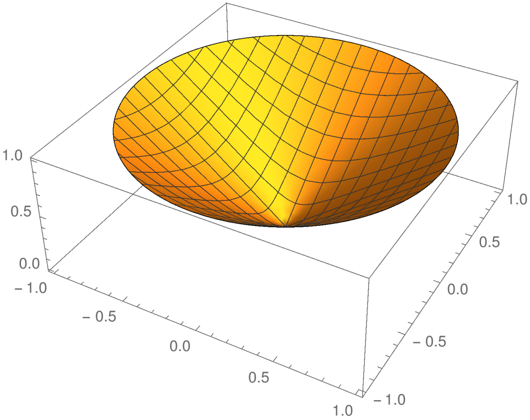
Example 3



All the points on the line $x = 0$ are saddle points for a function $f(x, y) = -x^3$.

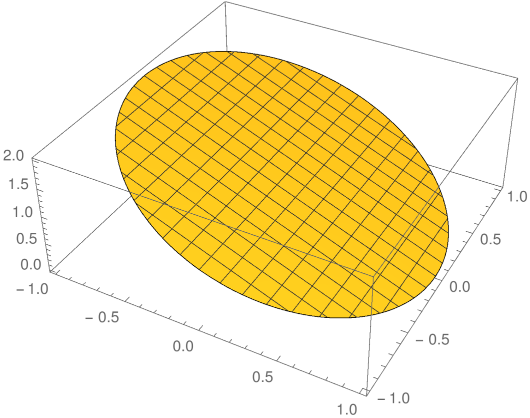
$$\nabla f(0, y) = (-3x^2, 0)\Big|_{(0,y)} = \mathbf{0} \text{ for all } y \in \mathbb{R}.$$

Example 4



Function $f(x, y) = \sqrt{x^2 + y^2}$ has a global minimum $f(0, 0) = 0$ at the point $(0, 0)$. The function f is continuous, but its gradient ∇f is not defined at $(0, 0)$.

Example 5



The function $f(x, y) = 1 - x$ does not have extremes, if it is defined in the whole plain $D = \mathbb{R}^2$.
If we take $D = \{(x, y) : x^2 + y^2 \leq 1\}$, then the function has maximum $f(-1, 0) = 2$ and minimum $f(1, 0) = 0$ at the boundary of D.

Classification of critical values: introduction

- Consider the quantity

$$\Delta f = f(\mathbf{x} + \mathbf{h}) - f(\mathbf{x})$$

  at the critical point $\mathbf{x} \in D$.
- If Δf takes only positive values (when $\|\mathbf{h}\|$ is small), the point $\mathbf{x}$ is a (local) minimum, and if it takes only negative values, the point $\mathbf{x}$ is a (local) maximum. If Δf changes sign, then the point $\mathbf{x}$ is neither a minimum nor a maximum (= saddle point).

The second derivative test

- We want to see how the change of function changes $\rightarrow$ second derivative test
- The single variable second derivative test:
  1. If $f''(x) < 0$, the function f has a local maximum at the point x.
  2. If $f''(x) > 0$, the function f has a local minimum at the point x.
  3. If $f''(x) = 0$, the test does not give an answer and the question must be solved in another way.
- Next, we will generalize this idea to the multivariable case.

Second derivative for multivariable function

- Let $f\colon D \subset \mathbb{R}^n \to \mathbb{R}$ be a function $f\colon D \subset \mathbb{R}^n \to \mathbb{R}$ with continuous second-order partial derivatives.
- The natural derivative of the function f is a gradient which is itself a vector-valued function $\nabla f\colon \mathbb{R}^n \to \mathbb{R}^n$.
- Thus the second derivative of a function f is a matrix, which we call the **Hessian matrix**

$$H_f(\mathbf{x}) = \begin{bmatrix} \frac{\partial^2}{\partial x_1^2} f(\mathbf{x}) & \frac{\partial^2}{\partial x_2 \partial x_1} f(\mathbf{x}) & \dots & \frac{\partial^2}{\partial x_n \partial x_1} f(\mathbf{x}) \\ \frac{\partial^2}{\partial x_1 \partial x_2} f(\mathbf{x}) & \frac{\partial^2}{\partial x_2^2} f(\mathbf{x}) & \dots & \frac{\partial^2}{\partial x_n \partial x_2} f(\mathbf{x}) \\ \vdots & \vdots & & \vdots \\ \frac{\partial^2}{\partial x_1 \partial x_n} f(\mathbf{x}) & \frac{\partial^2}{\partial x_2 \partial x_n} f(\mathbf{x}) & \cdots & \frac{\partial^2}{\partial x_n^2} f(\mathbf{x}) \end{bmatrix}.$$

- Since all second order partial derivatives of f are continuous, the order of the derivation can be changed, and the matrix is symmetric.

Hessian matrix

Why are we interested in the Hessian matrix?

- The Gradient allows us to write an linear (first-order) approximation for the function $f\colon D \subset \mathbb{R}^n \to \mathbb{R}$,
- The Hessian matrix gives a quadratic refinement:

$$f(\mathbf{x}+\mathbf{h}) \approx f(\mathbf{x}) + \mathbf{h}\cdot\nabla f(\mathbf{x}) + \frac{1}{2}\mathbf{h}H_f(\mathbf{x})\mathbf{h}^T,$$

  where the (horizontal) vector $\mathbf{h} = (h_1, h_2, \ldots, h_n)$ is small.
- This is actually just a new way of writing second-order Taylor approximation
- A term of the form $\mathbf{z}A\mathbf{z}^T$ is the quadratic form for a $n \times n$-matrix A where $\mathbf{z}$ is the n-row vector.
- Thus for the point $\nabla f(\mathbf{x}) = 0$, we have approximation

$$\Delta f = f(\mathbf{x}+\mathbf{h}) - f(\mathbf{x}) \approx \frac{1}{2}\mathbf{h}H_f(\mathbf{x})\mathbf{h}^T.$$

  and by thinking that $\mathbf{h} \approx 0$ we can use this to determinate the type of the critical value.

Definiteness of the matrix

What does the positivity or negativity of an symmetric matrix mean?
If A is a symmetric matrix, then

- A is called positive definite if its all eigenvalues are positive.
- A is said to be negative definite if its all eigenvalues are negative.
- A is said to be indefinite, if it has at least one positive and one negative eigenvalue.
- A is a positive semidefinite if its eigenvalues are nonnegative.
- A matrix A is negative semidefinite if all its eigevalues are nonpositive.

Positive/negative definite matrices have many of the same properties as positive/negative real numbers.

Definiteness of the matrix and the quadratic form

The definiteness or indefiniteness of a symmetric matrix A is inherited to the corresponding quadratic form.

- If A is a positive definite, then $\mathbf{x}A\mathbf{x}^T > 0$ for all nonzero (horizontal) vectors $\mathbf{x} \in \mathbb{R}^n$.
- If A is a negative definite, then $\mathbf{x}A\mathbf{x}^T < 0$ for all nonzero (horizontal) vectors $\mathbf{x} \in \mathbb{R}^n$.
- If A is an indefinite, then $\mathbf{x}A\mathbf{x}^T$ gets both negative and positive values.

These can be proved by diagonalizing the symmetric matrix A to the form $A = U^T \Lambda U$, where the diagonal matrix Λ contains the eigenvalues of A

Second derivative test for multivariable functions

Theorem

Let $f\colon D \subset \mathbb{R}^n \to \mathbb{R}$ be a function with continuous second partial derivatives around the critical point $\mathbf{x} \in D$. Then:

a) If $H_f(\mathbf{x})$ is a positive definite, then f has a local minimum at $\mathbf{x}$.

b) If $H_f(\mathbf{x})$ is a negative definite, then f has a local maximum at $\mathbf{x}$.

c) If $H_f(\mathbf{x})$ is an indefinite, then f has a saddle point at $\mathbf{x}$.

d) Otherwise, the test gives no information about the function f.

These follow from the approximation $f(\mathbf{x}+\mathbf{h}) - f(\mathbf{x}) \approx \frac{1}{2}\mathbf{h}H_f(\mathbf{x})\mathbf{h}^T$ when $\mathbf{h} \approx 0$.

Example when $n = 2$

Find and classify the critical points of the function

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

There is a simpler test for two variable functions based on the determinant of the Hessian matrix, see Guichard et friends 14.7

Example when $n = 3$

- Find and classify the critical points of the function

$$f(x, y, z) = x^2 y + y^2 z + z^2 - 2x.$$

- The equations for the critical points ($\nabla f = 0$) are

$$\begin{aligned} 0 &= \partial_1 f(x, y, z) = 2xy - 2, \\ 0 &= \partial_2 f(x, y, z) = x^2 + 2yz, \\ 0 &= \partial_3 f(x, y, z) = y^2 + 2z. \end{aligned}$$

- Solving these we see that the only critical point of the function f is $P = (1, 1, -1/2)$.

Example when $n = 3$ continues

- The Hessian matrix is $H_f(1, 1, -1/2) = \begin{bmatrix} 2 & 2 & 0 \\ 2 & -1 & 2 \\ 0 & 2 & 2 \end{bmatrix}$.
- Let's calculate the eigenvalues of the matrix using MATLAB, for example

```
>> a = [2 2 0 ; 2 -1 2 ; 0 2 2]
a =
     2     2     0
     2    -1     2
     0     2     2
>> eig(a)
ans =
  -2.7016
   2.0000
   3.7016
```

- So the function f has a saddle point at the point $P = (1, 1, -1/2)$.