Applications of Partial Derivatives

Optimization

Lecture 7: More about optimization: Restricted domain problems and constrained problems

Learning goals:

1. How to find extremes for a function whose domain is restricted to a subset of $\mathbb{R}^n$?
2. How to find extremes for a function whose variables must satisfy one or more constraint equations (Lagrange multilpiers)?
3. Extra topic: How does optimization justify the method of least squares (used in regression analysis)?

Where to find the material?

Corral 2.7 (Lagrange multipliers)
Guichard et friends 14.8 (Lagrange multipliers)
Active Calculus 10.7 (Restricted domains), 10.8 (Lagrange multipliers)
Adams-Essex 14.2, 14.3 and 14.5

Extreme values for a function defined on restricted domain

Recall: Extreme values for the function $f\colon D \subset \mathbb{R}^n \to \mathbb{R}$ can be

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

The Extreme Value Theorem

If f is a continuous function whose domain is a closed and bounded set in $\mathbb{R}^n$, then f has a global maximum and a global minimum.

Example

Find the extreme values of the function $f(x, y) = x^2ye^{-(x+y)}$ on a region T given by $x \geq 0$, $y \geq 0$, and $x + y \leq 4$.

Constrained extreme-value problems

Often in optimisation tasks, you want to impose constraints on the variables to be optimised.

Example: One wants to minimize the surface area of the can (cylinder shaped) (i.e. the used material) $A(h, r) = 2\pi rh + 2\pi r^2$ such that the volume $V(r, h) = \pi r^2 h$ is constant (for example 2dl).

Lagrange multipliers

A general optimization problem with constraint condition:

"Minimize $f(x, y)$ under the condition $g(x, y) = 0$."

- Note that if the problem has a solution, then at the solution point (a, b) the vectors ∇f and ∇g have must either be parallel or opposite (if $\nabla g(a, b) \neq 0$).
- Why? Because otherwise the function f would have a non-zero directional derivative in the direction that is tangent to the curve $g(x, y) = 0$ at the point (a, b), and therefore the minimum cannot be at the point (a, b).
- What if the task were to maximize $f(x, y)$ under the condition $g(x, y) = 0$?

Lagrange multipliers

- If an optimum point exists, it is a critical point of the Lagrange function
$$L(x, y, \lambda) = f(x, y) + \lambda g(x, y).$$
- The method also generalizes to three or more variable cases. For example, for three variables, the Lagrange function is
$$L(x, y, z, \lambda, \mu) = f(x, y, z) + \lambda g(x, y, z) + \mu h(x, y, z),$$
where f is the function to be minimized and the constraint conditions are $g(x, y, z) = 0$ and $h(x, y, z) = 0$.

Example 1

Minimize the function $f(x, y) = x^2 + y^2$ with the condition $g(x, y) = x^2 y - 16 = 0$.

- $\nabla g(x, y) = (2xy, x^2) \neq (0, 0)$ when $g(x, y) = 0$
- Form the Lagrange function

$$L(x, y, \lambda) = x^2 + y^2 + \lambda(x^2 y - 16).$$

- The equations for the critical points are

$$\begin{aligned} 0 &= \frac{\partial L}{\partial x} = 2x(1 + \lambda y), \\ 0 &= \frac{\partial L}{\partial y} = 2y + \lambda x^2, \\ 0 &= \frac{\partial L}{\partial \lambda} = x^2 y - 16, \end{aligned}$$

the last of which is always the constraint condition itself.

Example 1 (continues)

- The first equation gives $x = 0$ or $\lambda y = -1$, but $x = 0$ contradicts the third equation.
- Thus, from the second equation

$$0 = 2y^2 + \lambda y x^2 = 2y^2 - x^2.$$

- This gives $x = \pm\sqrt{2}y$, and $2y^3 = 16$, so $y = 2$.
- So there are two extreme values occur at $(x, y) = (\pm 2\sqrt{2}, 2)$. We have to find out by other means whether they are minima or maxima.

Example 2

Let us try to find the minimum of the function $f(x, y) = y$ with the condition $g(x, y) = y^3 - x^2 = 0$ using the method of Lagrange multipliers.

- It is clearly seen that the minimum $f(x, y) = 0$ is reached at the point $(0, 0)$ (use Geogebra to graph the $g(x, y) = 0$).
- Notice $g(x, y) = (-2x, 3y^2) = (0, 0)$ when $(x, y) = (0, 0)$.
- Let us form anyway the Lagrange function

$$L(x, y, \lambda) = y + \lambda(y^3 - x^2).$$

- For critical points of L we get the equations

$$-2\lambda x = 0, \quad 1 + 3\lambda y^2 = 0, \text{ and } y^3 - x^2 = 0.$$

- These equations contradict each other, so there is no solution.
- From this we see, the Lagrange multiplier method sees extreme values only in points where $\nabla g(x, y) \neq \mathbf{0}$.

Example with three variables

Find the extreme values of the function $f(x, y, z) = xy + 2z$ under the conditions $x + y + z = 0$ and $x^2 + y^2 + z^2 = 24$.

- Since f is continuous and the intersection of the given intersection sets is circular line (i.e. a bounded and closed set), then the extreme values exist.
- Form the Lagrange function

$$L(x, y, z, \lambda, \mu) = xy + 2z + \lambda(x + y + z) + \mu(x^2 + y^2 + z^2 - 24).$$

- The equations for the critical points of the Lagrange function are

$$
\begin{aligned}
&y + \lambda + 2\mu x = 0, \\
&x + \lambda + 2\mu y = 0, \\
&2 + \lambda + 2\mu z = 0, \\
&x + y + z = 0, \text{ ja} \\
&x^2 + y^2 + z^2 - 24 = 0.
\end{aligned}
$$

- Subtracting the first equality from the second leads to $(x - y)(1 - 2\mu) = 0$, so either $\mu = 1/2$ or $x = y$. Let's examine both cases.

- **Case I** ($\mu = 1/2$): based on the second and third equation.

$$x + \lambda + y = 0 \text{ and } 2 + \lambda + z = 0, \text{ so } x + y = 2 + z.$$

- The fourth equation ($x + y + z = 0$) gives $z = -1$ and so $x + y = 1$.
- From the last equation ($x^2 + y^2 + z^2 - 24 = 0$) we get $x^2 + y^2 = 24 - z^2 = 23$.
- Since $x^2 + y^2 + 2xy = (x + y)^2 = 1$, we get $2xy = 1 - (x^2 + y^2) = 1 - 23 = -22$ and thus $xy = -11$.
- Now $(x - y)^2 = x^2 + y^2 - 2xy = 23 + 22 = 45$, so $x - y = \pm 3\sqrt{5}$.
- Together with the equation $x + y = 1$ this gives two critical points

$$P_1 = \big((1+3\sqrt{5})/2, (1-3\sqrt{5})/2, -1\big) \text{ and } P_2 = \big((1-3\sqrt{5})/2, (1+3\sqrt{5})/2, -1\big).$$

- At each these points $f(x, y, z) = -11 - 2 = -13$.

- **Case II** ($x = y$): from the fourth equation we see that. $z = -2x$, and from the last equation $6x^2 = 24$, i.e. $x = \pm 2$.
- Thus, the critical points are

$$P_3 = (2, 2, -4) \text{ and } P_4 = (-2, -2, 4).$$

- We obtain

$$f(2, 2, -4) = 4 - 8 = -4 \text{ and } f(-2, -2, 4) = 4 + 8 = 12.$$

- Thus, the maximum of the function f is 12 and the minimum is -13.

Regression analysis

Application of Optimization

- Regression analysis aims to select the value of the parameter β such that the curve

$$y = f(x; \beta)$$

  is as close as possible to each observation point

$$(x_j, y_j) \in \mathbb{R}^2,\ j = 1, 2, \ldots, n.$$

- Such an optimally chosen curve is called the *regression model* $y = f(x; \beta)$, where the form of the function f is the chosen according tho the situation.
- As long as f is chosen, one solution to the curve fitting problem is to use the *least squares method*.

Least Square Method

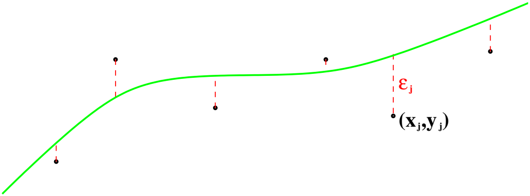
- The least squares method aims to minimize the error terms of the regression model ε_j

$$\varepsilon_j = y_j - f(x_j;\beta)\,, \quad j = 1, 2, \ldots, n$$

  the sum of squares, i.e. the function

$$F(\beta) = \sum_{j=1}^{n} \varepsilon_j^2 = \sum_{j=1}^{n} \big(y_j - f(x_j;\beta)\big)^2.$$

  by changing the parameter vector $\beta = (\beta_0, \beta_1, \ldots, \beta_m)$.
- **Question:** Why not minimize the expression $\sum_{j=1}^{n} |y_j - f(x_j;\beta)|$ instead of the square sum?



Green line = the graph of the function $f(x;\beta)$ with some fixed parameter $\beta = (\beta_1, \beta_2, \ldots, \beta_m)$.

Black dots = datapoints (x_j, y_j)

red dashed lines = error terms ε_j

Linear regression

In linear regression $f(x;\beta) = \beta_0 + \beta_1 x$ where $\beta = (\beta_0, \beta_1)$ and the sum of squares is

$$F(\beta_0, \beta_1) = \sum_i (y_i - \beta_0 - \beta_1 x_i)^2.$$

Find the point (β_0, β_1) such that $\nabla F(\beta_0, \beta_1) = 0$.

- Calculate the partial derivative of

$$\frac{\partial}{\partial \beta_0} F(\beta_0, \beta_1) = 2\big(\beta_1 \sum_i x_i + n\beta_0 - \sum_i y_i\big).$$

- Solve for the zero point

$$\beta_0 = \frac{1}{n} \sum_i y_i \frac{\beta_1}{n} \sum_i x_i = \bar{\mathbf{y}} - \beta_1 \bar{\mathbf{x}}$$

where $\bar{\mathbf{x}}$ is the arithmetic mean of the components of the data vector $\mathbf{x} = (x_1, x_2, \ldots, x_n)$.

- Next, let's calculate the partial derivative

$$\frac{\partial}{\partial \beta_1} F(\beta_0, \beta_1) = 2\big(\beta_0 \sum_i x_i + \beta_1 \sum_i x_i^2 - \sum_i x_i y_i\big).$$

- Substituting the expression for β_0, we get

$$n\bar{\mathbf{x}}\bar{\mathbf{y}} - n\beta_1 \bar{\mathbf{x}}^2 + \beta_1 \sum_i x_i^2 - \sum_i x_i y_i = 0.$$

- Solving β_1 from this equation we get

$$\beta_1 = \frac{n\bar{\mathbf{x}}\bar{\mathbf{y}} - \sum_i x_i y_i}{n\bar{\mathbf{x}}^2 - \sum_i x_i^2}$$

Example

Find the linear regression model using the method of the least squares for the data

x_i	0.0	1.0	2.0	3.0	4.0
y_i	2.10	1.92	1.84	1.71	1.64

Estimate y when $x = 5$.

- We obtain $\bar{\mathbf{x}} = 2.0$, $\bar{\mathbf{y}} = 1.842$, and
$$\beta_1 = \frac{-1.13}{10.0} = -0.113.$$
- Thus $\beta_0 = 1.842 + 0.113 \cdot 2.0 = 2.068$.
- Thus $y = -0.113x + 2.068$, and the desired estimate at the point $x = 5$ is $y = -0.113 \cdot 5 + 2.068 = 1.503$.