Gradient and optimization problems

O.M. Kiselev
o.kiselev@innopolis.ru

Innopolis university

March 3, 2023

An invariant form of differential

Geometrical sense of the partial derivatives

Surfaces on the $N+1$-dimension spaces

Extreme points on the surface

Invariant form of the differential

Consider the changing of coordinates for x, y:

$$x = x(u, v),\ \ y = y(u, v),$$

Below we suppose that the functions $x(u, v)$ and $y(u, v)$ are differentiable.

$$\begin{aligned} df &= \frac{\partial f}{\partial x} dx + \frac{\partial f}{\partial y} dy = \\ &= \frac{\partial f}{\partial x}\left(\frac{\partial x}{\partial u} du + \frac{\partial x}{\partial v} dv\right) + \frac{\partial f}{\partial y}\left(\frac{\partial y}{\partial u} du + \frac{\partial y}{\partial v} dv\right) = \\ &= \left(\frac{\partial f}{\partial x}\frac{\partial x}{\partial u} + \frac{\partial f}{\partial y}\frac{\partial y}{\partial u}\right) du + \left(\frac{\partial f}{\partial x}\frac{\partial x}{\partial v} + \frac{\partial f}{\partial y}\frac{\partial y}{\partial v}\right) dv \\ &= \frac{\partial f}{\partial u} du + \frac{\partial f}{\partial v} dv. \end{aligned}$$

Differential for the N-dimensional function

The changing of the variables in general form looks like:

$$X = X(U), \quad X(U) = (x_1(u_1, \ldots, u_n), \ldots, x_N(u_1, \ldots, u_N)).$$

In this case the differential has the same form:

$$df = \sum_{k=1}^{N} \frac{\partial f}{\partial x_k} dx_k = \sum_{k=1}^{N} \frac{\partial f}{\partial u_k} du_k.$$

As well as the differential is the primary (linear) part of the function changing then the vector

$$\vec{S} = \left(\frac{\partial f}{\partial x_1}, \ldots, \frac{\partial f}{ptx_N} \right)$$

defines the direction of the grows of the function for the given point X.

Gradient of the function

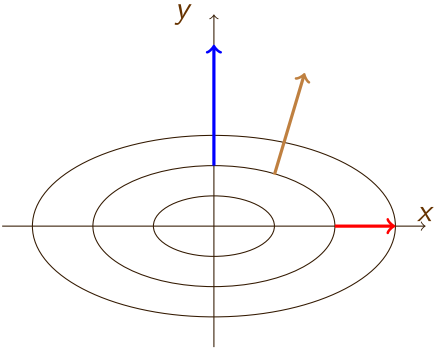
The vector $\vec{v} = \left(\frac{\partial f}{\partial x}, \frac{\partial f}{\partial y}\right)$ is called gradient of the function $f(x, y)$ at the point (x, y). The gradient can be written by following equivalent definitions:

$$\vec{\text{grad}}(f) \equiv \left(\frac{\partial f}{\partial x}, \frac{\partial f}{\partial y}\right);$$
$$\vec{\nabla} f = \left(\frac{\partial f}{\partial x}, \frac{\partial f}{\partial y}\right).$$

Define the differential of the independent variables as $\vec{dX} = (dx_1, dx_2, \dots, dx_N)$. The differential of the function can be written as scalar product:

$$(\vec{\nabla} f, \vec{dX}) = \sum_{k=1}^{N} \frac{\partial f}{\partial x_k} dx_k.$$

Example. Gradient of the function



$$f(x,y) = \frac{x^2}{4} + y^2,$$

$$\vec{\nabla} f = \left(\frac{x}{2}, 2y\right),$$

$$\vec{\nabla} f|_{(2,0)} = (1,0),$$

$$\vec{\nabla} f|_{(0,1)} = (0,2),$$

$$\vec{\nabla} f|_{(1,\sqrt{3}/2)} = (0.5, \sqrt{3}).$$

Derivative at given direction

Directional derivatives one can obtain by the following track:

1. Define direction $\vec{s} = (s_x, s_y)$ with the unit length: $|\vec{s}| = 1$.
2. Specify the certain point $A = (x_0, y_0)$.
3. Define the dependence of the coordinates $x = x_0 + s_x t,\ y = y_0 + s_y t$.
4. Substitute the dependence into the function $f(x, y) = f(x_0 + s_x t, y_0 + s_y t)$.
5. Find full derivative on t:

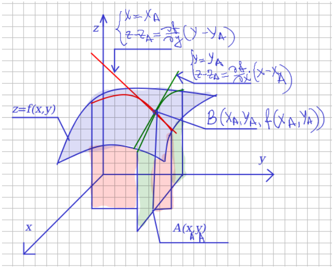
$$\frac{df}{dt} = \frac{\partial f}{\partial x}\frac{dx}{dt} + \frac{\partial f}{\partial y}\frac{dy}{dt} = \left(\frac{\partial f}{\partial x}s_x + \frac{\partial f}{\partial y}s_y\right)\Bigg|_{(x,y)=(x_0,y_0)}.$$

Geometry of surfaces

Let $f(x, y)$ be a definition for some surface in 3-dimensional space:$z = f(x, y)$.

- $f_x = \frac{\partial f}{\partial x}$ define the inclination along the direction of the x axis.
- $f_y = \frac{\partial f}{\partial y}$ define the inclination along the direction of the y axis.
- The vector $\vec{\Phi} = (f_x, f_y)$ defines a projection on the plane (x, y) of the surface inclination at the current point (x, y).

Geometrical sense of the partial derivatives



Let us consider surface $z = f(x, y)$. Consider a dissection of the surface by the plain $y = y_0$, $y_0 = \text{const}$. The intersection of the surface and plain defines the curve one-dimensional curve $z = f(x, y_0)$ and the angle of the tangent line for the curve at the point x_0 is $\frac{\partial f}{\partial x}$ The same for the curve $z = f(x_0, y)$ one gets the angle of the tangent line for the curve $z = f(x_0, y)$ is $\frac{\partial f}{\partial y}$.

A normal vector

Rewrite the equation for the surface in the form:

$$z - f(x, y) = 0.$$

The differential on the surface should be following:

$$dz - \frac{\partial f}{\partial x} dx - \frac{\partial f}{\partial y} dy = 0.$$

This equality should be fulfill for any curve on this surface $(x(t), y(t), z(t))$, then these equality is the scalar product for the vector $\vec{N} = \left(\frac{\partial f}{\partial x}, \frac{\partial f}{\partial y}, -1\right)$ and the vector of differential for any curve on the surface. As well as the differential defines the tangent lines for the surface, then $\vec{N}$ is a normal vector for the surface at the point $(x_0, y_0, f(x_0, y_0))$.

Equation for the tangent plain

From the course of analytic geometry the plain is defined by the normal vector $\vec{N}$ and a point. It yields:

$$\frac{\partial f}{\partial x}(x_0, y_0)(x - x_0) + \frac{\partial f}{\partial y}(x_0, y_0)(y - y_0) - (z - z_0) = 0.$$

This formula defines the tangent plain for given surface $z = f(x, y)$ at the point $(x_0, y_0, f(x_0, y_0))$.

Tangent plain for implicit case

The function $F(x, y, z) = 0$ defines two dimension manifold in general case.

$$dF \equiv \frac{\partial F}{\partial x} dx + \frac{\partial F}{\partial y} dy + \frac{\partial F}{\partial z} dz.$$

For any curve on this surface $(x(t), y(t, z(t))$:

$$\frac{\partial F}{\partial x} dx + \frac{\partial F}{\partial y} dy + \frac{\partial F}{\partial z} dz = 0.$$

The $\vec{V} = (x', y', z')$ define the differential of the curve, then $\vec{N} = \left(\frac{\partial F}{\partial x}, \frac{\partial F}{\partial y}, \frac{\partial F}{\partial z}\right)$ is a normal vector for the surface.
The tangent plain at the point (x_0, y_0, z_0) has the form:

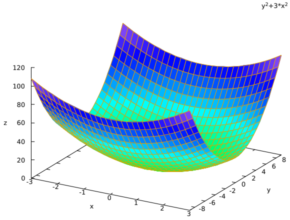
$$\frac{\partial F}{\partial x}(x_0, y_0, z_0)(x - x_0) + \frac{\partial F}{\partial y}(x_0, y_0, z_0)(y - y_0) + \frac{\partial F}{\partial z}(x_0, y_0, z_0)(z - z_0) = 0.$$

Definitions of local minima and maxima

A value A is a local maxima of $f(X)$ in the point $X^{(A)}$ if $\forall \epsilon > 0 \exists \delta \ \forall X: \ ||X - X^{(A)}|| < \delta, \ A - f(X) \geq 0.$

Vice versa. A value A is a local minima of $f(X)$ in the point $X^{(A)}$ if $\forall \epsilon > 0 \exists \delta \ \forall X: \ ||X - X^{(A)}|| < \delta, \ A - f(X) \leq 0.$

Definition of extreme point

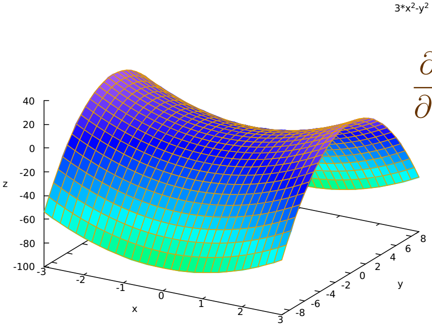


The point
of differentiable function $f(X)$
where all derivatives of first order
are zero is called extreme point.
For following functions point
$A = (0,0)$ is an extreme point:

$$f(x_1, x_2) = 3x_1^2 + x_2^2,$$
$$\frac{\partial f}{\partial x_1} = 6x_1, \ \frac{\partial f}{\partial x_1} = 2x_2;$$

$$f(x_1, x_2) = -3x_1^2 - x_2^2, \quad \frac{\partial f}{\partial x_1} = -6x_1, \ \frac{\partial f}{\partial x_1} = -2x_2;$$

Saddle point



$$f(x_1, x_2) = 3x_1^2 - x_2^2,$$

$$\frac{\partial f}{\partial x_1} = 6x_1, \ \frac{\partial f}{\partial x_1} = -2x_2.$$

Necessary conditions for the extreme points

Theorem

If $f(X)$ is differentiable, then $\frac{\partial f}{\partial x_k} = 0$, $\forall k \in \{1, \ldots, N\}$ at the interior maxima or minima point.

Proof. Suppose the $A = (0, 0)$ is maxima, and one of the partial derivative is not zero, then in the maxima point:

$$f(X) = f(0,0) + \frac{\partial f}{\partial x}(0,0)x + o(|x| + |y|),$$

then for $\frac{\partial f}{\partial x}(0,0)x > 0$ the $f(x, y) > f(0, 0)$, which contradict to the initial claim. For the another partial derivative one can consider by the same way. As result one gets the claim of the theorem.

Theorem about mixed second derivative

Theorem

Let $f(X)$ be differentiable function and all derivatives of the first and second order are continuous, then

$$\frac{\partial^2 f}{\partial x_k \partial x_j} = \frac{\partial^2 f}{\partial x_j \partial x_k}.$$

Proof. For simplicity consider the function of two variables $X = (x_1, x_2)$ and write the function using the Taylor formula for first and second variable by sequence:

$$f(x_1, x_2) = f(0, x_2) + \frac{\partial f}{\partial x_1}(0, x_2)x_1 + \frac{1}{2}\frac{\partial^2 f}{\partial x_1^2}(0, x_2)x_1^2 + o(x_1^2) =$$

Theorem about mixed derivative. Proof.

Define to be shorter $f_0 = f(0,0)$, $\frac{\partial f}{\partial x_k}\big|_{(0,0)} = \frac{\partial f_0}{\partial x_k}$, $\frac{\partial^2 f}{\partial x_k^2}\big|_{(0,0)} = \frac{\partial^2 f_0}{\partial x_k^2}$, $\frac{\partial^2 f_0}{\partial x_k \partial x_j} = \frac{\partial^2 f}{\partial x_k \partial x_j}\big|_{(0,0)}$.

$$f(x_1, x_2) = f_0 + \frac{\partial f_0}{\partial x_2} x_2 + \frac{1}{2}\frac{\partial^2 f_0}{\partial x_1^2} x_2^2 + O(x_2^2) +$$

$$\left(\frac{\partial f_0}{\partial x_1} + \frac{\partial^2 f_0}{\partial x_2 \partial x_1} x_2 + o(x_2)\right) x_1 + \frac{1}{2}\left(\frac{\partial^2 f_0}{\partial x_1^2} + o(x_2)\right) x_1^2 + o(x_1^2) =$$

$$f_0 + \frac{\partial f_0}{\partial x_1} x_1 + \frac{\partial f_0}{\partial x_2} x_2 +$$

$$\frac{1}{2}\left(\frac{\partial^2 f_0}{\partial x_1^2} x_1^2 + 2\frac{\partial^2 f_0}{\partial x_2 \partial x_1} x_1 x_2 + \frac{\partial^2 f_0}{\partial x_1^2} x_2^2\right) + o(x_1^2 + x_1 x_2 + x_2^2);$$

Theorem about mixed derivative. Proof.

The same formula one obtains for the case of using the Taylor formula by opposite case and the therm of order $x_2 x_1$:

$$f(x_1, x_2) \sim f_0 + \frac{\partial f_0}{\partial x_1} x_1 + \frac{\partial f_0}{\partial x_2} x_2 +$$
$$\frac{1}{2}\left(\frac{\partial^2 f_0}{\partial x_1^2} x_1^2 + 2 \frac{\partial^2 f_0}{\partial x_1 \partial x_2} x_2 x_1 + \frac{\partial^2 f_0}{\partial x_1^2} x_2^2 \right);$$

Due to the linear independence of the terms of order $x_1 x_2$ are equivalent. Then

$$\frac{\partial^2 f_0}{\partial x_1 \partial x_2} = \frac{\partial^2 f_0}{\partial x_2 \partial x_1}$$

Sufficient conditions for minima and maxima

Theorem

Let $f(X)$ be twice differentiable function of two variables and the point $A = (0, 0)$ is extreme point

- *if* $\left(\frac{\partial^2 f}{\partial x_1 \partial x_2}\right)^2 - \frac{\partial^2 f}{\partial x_1^2}\frac{\partial^2 f}{\partial x_2^2} < 0$ *and* $\frac{\partial^2 f}{\partial x_1^2} < 0$, *then* A *is a maxima;*
- *if* $\left(\frac{\partial^2 f}{\partial x_1 \partial x_2}\right)^2 - \frac{\partial^2 f}{\partial x_1^2}\frac{\partial^2 f}{\partial x_2^2} < 0$ *and* $\frac{\partial^2 f}{\partial x_1^2} > 0$ *then* A *is a minima;*
- *if* $\left(\frac{\partial^2 f}{\partial x_1 \partial x_2}\right)^2 - \frac{\partial^2 f}{\partial x_1^2}\frac{\partial^2 f}{\partial x_2^2} > 0$, *then* A *is a saddle point;*
- *if* $\left(\frac{\partial^2 f}{\partial x_1 \partial x_2}\right)^2 - \frac{\partial^2 f}{\partial x_1^2}\frac{\partial^2 f}{\partial x_2^2} = 0$, *then we need an additional studies.*

A sketch of proof for the second order criteria for maxima

in the extreme point $\vec{\nabla} f = 0$, then the change of the function in small neighborhood of the point is:

$$f(X) - f(0) = \frac{\partial^2 f_0}{\partial x_1^2}(dx)^2 + \left(\frac{\partial^2 f_0}{\partial x_1 \partial x_2}\right)^2 dxdy + \frac{\partial^2 f_0}{\partial x_2^2}(dy)^2 + \\ + o((dx)^2 + dxdy + (dy)^2) = \\ = \left(\frac{\partial^2 f_0}{\partial x_1^2} + \left(\frac{\partial^2 f_0}{\partial x_1 \partial x_2}\right)^2 \frac{dy}{dy} + \frac{\partial^2 f_0}{\partial x_2^2}\left(\frac{(dy)^2}{(dx)^2}\right)^2\right)(dx)^2 + \\ + o((dx)^2 + dxdy + (dy)^2).$$

In maxima point the difference should be negative. It implies:

$$\frac{\partial^2 f_0}{\partial x_1^2} + \left(\frac{\partial^2 f_0}{\partial x_1 \partial x_2}\right)^2 \frac{dy}{dx} + \frac{\partial^2 f_0}{\partial x_2^2}\left(\frac{(dy)^2}{(dx)^2}\right) < 0.$$

A sketch of proof for the second order criteria for maxima

Define $\varkappa = dy/dx$, then

$$\frac{\partial^2 f_0}{\partial x_1^2} + \left(\frac{\partial^2 f_0}{\partial x_1 \partial x_2}\right)^2 \varkappa + \frac{\partial^2 f_0}{\partial x_2^2} \varkappa^2 < 0.$$

The second-order expression is negative for any $\varkappa \in \mathbb{R}$ when

$$\left(\frac{\partial^2 f_0}{\partial x_1 \partial x_2}\right)^2 - \frac{\partial^2 f_0}{\partial x_1^2} \frac{\partial^2 f_0}{\partial x_2^2} < 0.$$

Optimal problems with constrains

Let add a constrain to the optimization problem. The simplest question of this type is follows.

Find the shorter distance from the origin to a plain $ax + by + cz + d = 0$**.**

The first step is define the function for the optimization. The distance from the origin is follows:

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

From geometrical point of view one should construct a sphere which touch the given plain.

Straight forward solution is follows. Define one of the variable, say $z = z(x, y)$ using the equation for the plain and substitute the definition into the function $f = f(x, y, z(x, y))$. Then find the extreme point for the function of two variables.

Lagrange multipliers

Let us consider the level of f the function should touch to the plain. Then the gradients of f and the constraint curve are collinear:

$$\vec{\nabla} f = -\lambda \vec{\nabla} \phi.$$

Additional condition is the constrain
$\phi(x, y, z) \equiv ax + by + cz + d = 0$.

$$\phi(x, y, z) = 0.$$

Define the Lagrange function (Lagrangian):

$$L(x, y, z) = f(x, y, z) + \lambda \phi(x, y, z).$$

Lagrange multipliers

The necessary conditions for the extreme points:

$$\vec{\nabla} L(x, y, z, \lambda) = 0.$$

Or the same

$$\frac{\partial f}{\partial x} + \lambda \frac{\partial \phi}{\partial x} = 0,$$
$$\frac{\partial f}{\partial y} + \lambda \frac{\partial \phi}{\partial y} = 0,$$
$$\frac{\partial f}{\partial z} + \lambda \frac{\partial \phi}{\partial z} = 0,$$
$$\phi = 0$$

Lagrange multipliers. Example

$$f(x,y,z) = 4x + 2y - z + 1 \to \min,\ (x,y,z) \in x^2 + y^2 - 4 = 0.$$
$$L = 4x + 2y - z + 1 + \lambda(x^2 + y^2 - 4);$$
$$\frac{\partial L}{\partial x} = 4 + 2\lambda x,\ \frac{\partial L}{\partial y} = 2 + 2\lambda y,\ \frac{\partial L}{\partial z} = -1,\ \frac{\partial L}{\partial \lambda} = x^2 + y^2 - 4.$$
$$A_1 = (-4/\sqrt{5}, -2/\sqrt{5}, -1),\ A_2 = (4/\sqrt{5}, 2/\sqrt{5}, -1).$$

The answer: $A_1 = (-4/\sqrt{5}, -2/\sqrt{5}, -1)$.

Lagrange multipliers. General case

$f = f(X)$ and constraints $\phi_k(X)$, $k = 1, \ldots, m$, then:

$$L = f(x) + \sum_{k=1}^{m} \lambda_k \phi_k(X).$$

The necessary condition for the extreme point:

$$\vec{\nabla} L(X, \Lambda) = 0, \;\; \Lambda = (\lambda_1, \ldots, \lambda_m).$$