

3 Geometric interpretation of the first order ODE. Existence and uniqueness theorem

It was stated that our main goal for the first half of the course is to learn *analytical* methods of solving ODE. Especially, of solving the first order ODE in the form

$$y' = f(x, y). \quad (1)$$

Sometimes we will need the initial condition

$$y(x_0) = y_0, \quad x_0, y_0 \in \mathbf{R}. \quad (2)$$

In this lecture, however, we will discuss the geometric interpretation of the problem (1)–(2). This geometric interpretation allows us to use the graphical method to obtain approximate solutions to (1)–(2). First, recall that ϕ is a solution to (1)–(2) on the interval $I = (a, b)$ if $\phi(x_0) = y_0$, $\phi \in C^{(1)}(I; \mathbf{R})$ (this means that ϕ has a continuous first derivative everywhere on I) and satisfies (1) (i.e., $\phi'(x) \equiv f(x, \phi(x))$ for any $x \in I$).

Definition 1. *The graph of the solution $y = \phi(x)$ to (1)–(2) is called an integral curve of equation (1).*

If we change the initial conditions, we obtain different integral curves. Can we figure out an approximate form of the integral curves of (1) without actually solving (1)? The answer is “yes,” and it is based on the notion of the direction (or slope) field for the equation (1).

Recall from Calculus that the geometric meaning of the derivative at the point x_0 is the value of function \tan of the angle at which the tangent line at x_0 intersects the x axis (this is what we usually call “slope”). Therefore, from (1) it follows that for any point (\tilde{x}, \tilde{y}) , for which the right hand side of (1) is defined, we can calculate the slope of the graph of $y(\tilde{x})$ at this point, which is given simply by $f(\tilde{x}, \tilde{y})$. Now change the coordinates (\tilde{x}, \tilde{y}) and we will find the slope of $y(\tilde{x})$ at another point. These slopes can be represented graphically as small line segments, all together, these line segments constitute *the direction (slope) field* of a given ODE. Now consider a curve which is tangent at each point to the given direction field (it means this curve is tangent to the line segments at each point). The key fact is that this curve is an integral curve of the ODE. Indeed, at each point the slope $y'(x)$ of this curve is given by construction as $f(x, y)$, which exactly means that this function solves ODE. Therefore, if we are able to plot the direction field, we can sketch the integral curves of our equation.

There are two main ways to plot the direction field. The first one, which is suitable for a computer, is to divide the area in (x, y) plane into rectangular lattice, and calculate the slopes at all the points in the vertices of this lattice.

Example 2. Consider as an example the equation

$$y' = -\frac{x}{y}.$$

The direction field, built by computer, is presented in the figure below.

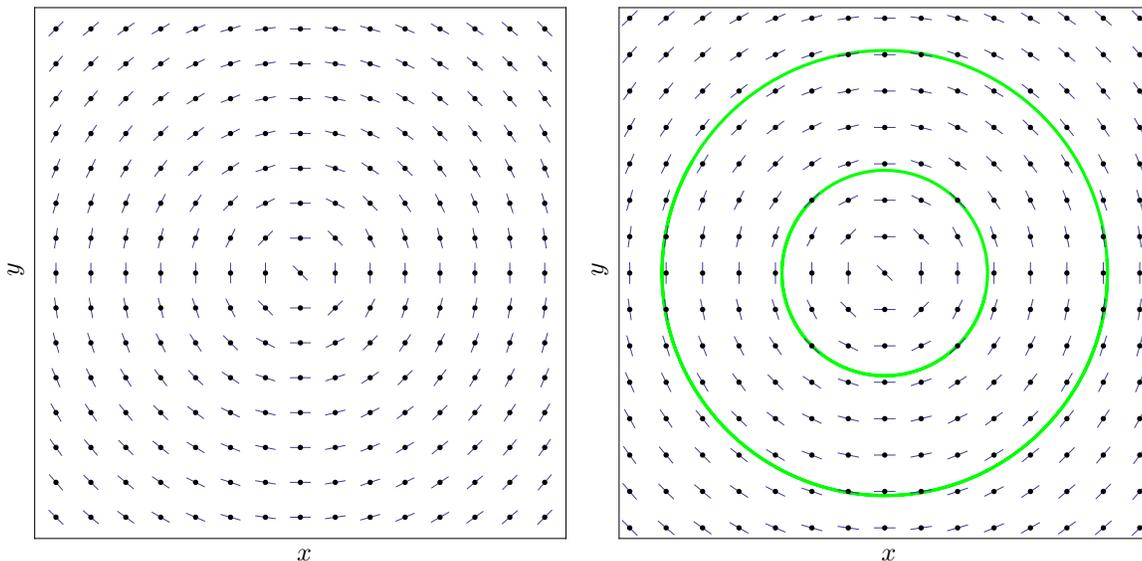


Figure 1: The direction field of the equation $y' = -x/y$. On the right two integral curves are also shown. Note that they are tangent to the direction field at each point.

An easier way for a human being to get an idea about the direction field is to consider *the isoclines* for the equation (1). The isoclines (“iso-cline” means lines of the “equal slope”) are defined by the equations

$$f(x, y) = C, \quad C = \text{constant}. \quad (3)$$

If we fix C , we find an implicitly defined curve $f(x, y) = C$, on every point of which the direction field is the same and has the slope C . We can plot this curve and put the line segments with the slope C on it. For example, if we take $C = 1$ and get a curve $f(x, y) = 1$ this means that at each point on this curve the direction field has the slope 1 (angle $\pi/4$). Now change C and find another curve $f(x, y) = C$, at every point of which the slopes are again the same, but different from the previous. Changing the constant C we can find enough isoclines so that the general pattern of the direction field becomes obvious.

Consider the same example

$$y' = -\frac{x}{y}.$$

The family of isoclines is given by

$$-\frac{x}{y} = C,$$

or, more convenient,

$$y = -\frac{x}{C},$$

which defines the family of straight lines passing through the origin. Take $C = 1$, we obtain $y = -x$. This means that on the line $y = -x$, at every point, the direction field has the slope 1 (see the left panel on the figure below). Now consider some other values of C , for instance, $C \rightarrow 0, C = -1, C \rightarrow \infty$. The corresponding isoclines will be $x = 0, y = x, y = 0$ (see the figure).

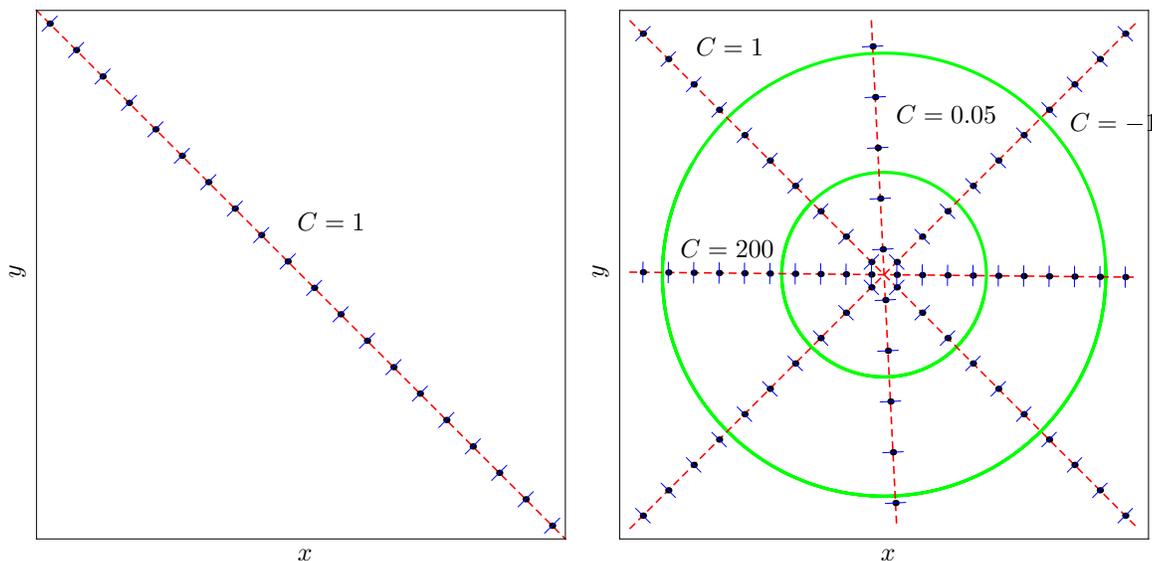


Figure 2: The direction field of the equation $y' = -x/y$ sketched by the isocline method. Left panel: The isocline corresponding to the slope $C = 1$ is shown in the dashed red. Four isoclines and two integral curves in the right panel are shown.

Example 3. Here is one more example:

$$y' = x^2 - y.$$

The family of isoclines is

$$y = x^2 - C.$$

In the figure below we have the direction field plotted by the computer method (on the left) and by the isocline method (on the right).

Now we can state the geometric interpretation of the problem (1)–(2): Geometrically, *to solve the IVP (1)–(2) means to find an integral curve belonging to the direction field of the equation and passing through the point with the coordinates (x_0, y_0) .*

Now try to think about the following questions: Can two isoclines intersect? Can they be tangent to each other? The first question is simple: if we assume that the right-hand side of the equation (1) is continuous both in x and y , then the isoclines cannot intersect. Indeed, assume the opposite, let two isoclines intersect with non-zero angle. This means that at the point of intersection we must have two different slopes, which is impossible for the continuous $f(x, y)$, hence the contradiction. The second question is much more subtle and requires the theory which is beyond our course. Therefore, we just state the most important facts without going into the details of justifying them.

First, let us assume that we do have two isoclines tangent to each other at some point (x_0, y_0) . What are some not very nice consequences of this fact? Assume that our differential equation models some real world process, and we are given the initial condition (x_0, y_0) . Since there are at least two integral curves passing through this point, by solving the equation we will have more than one solution. Which means, from the practical point of view, that we do not have any solutions, since we cannot use our mathematical model as a predictive tool. Therefore, we need to identify conditions, which

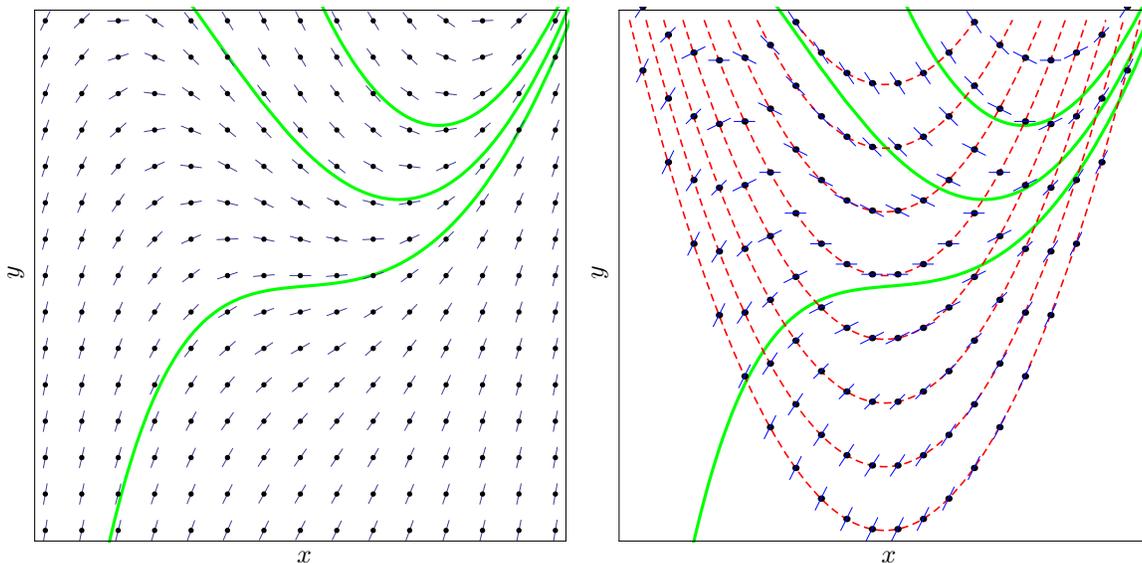


Figure 3: The direction field of the equation $y' = x^2 - y$ sketched by the computer method (on the left) and by the isocline method (on the right). The green lines are numerically obtained integral curves.

would allow us to avoid such situations. These conditions are provided by the existence and uniqueness theorem for the IVP (1)–(2), which is of paramount importance for the theory of differential equations.

Theorem 4 (The existence and uniqueness theorem). *Consider problem (1)–(2) and assume that function $f(x, y)$ is continuous in x and continuously differentiable in y in a neighborhood of (x_0, y_0) , including this point. Then there exists an $\varepsilon > 0$ such that for $(x_0 - \varepsilon, x_0 + \varepsilon)$ there exists one and only one integral curve of the equation (1) passing through the point (x_0, y_0) .*

This theorem shows that, for example, the integral curves in our last example do not intersect and are not tangent. Now consider an example, which is not covered by the existence and uniqueness theorem.

Example 5. Consider the equation

$$y' = \sqrt{y}.$$

The direction field together with a few integral curves are shown in the figure below.

Here the integral curves are tangent at certain points, and the reason for this is that \sqrt{y} does not have a continuous first derivative with respect to y for $y = 0$ (indeed, $(\sqrt{y})'_y = \frac{1}{2\sqrt{y}}$, which is undefined for $y = 0$).

To prove rigorously that the integral curves do indeed tangent to one another in certain cases (the figure cannot serve as a proof!), we solve the equation. This equation is separable, and we can solve it as

$$\begin{aligned} \int \frac{dy}{\sqrt{y}} &= \int dx \implies \\ 2\sqrt{y} &= x + C \implies \\ y &= \left(\frac{x}{2} + C\right)^2, \quad \frac{x}{2} + C \geq 0. \end{aligned}$$

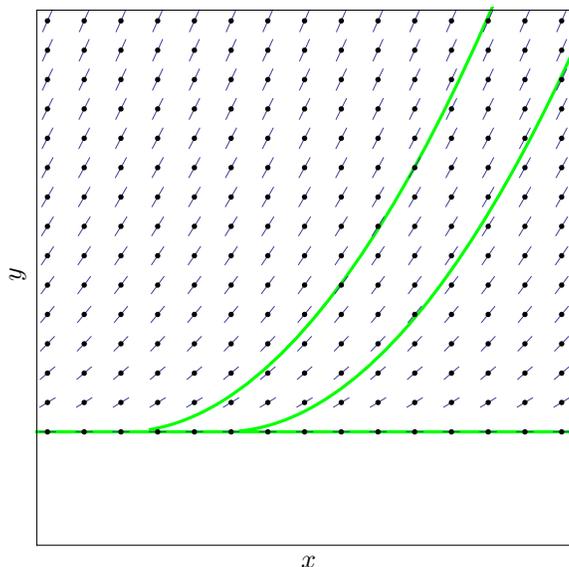


Figure 4: The direction field of the equation $y' = \sqrt{y}$ and three integral curves.

Note that $y = 0$ is also a solution.

Now consider the function

$$y(x) = \begin{cases} 0, & x < 0 \\ \frac{x^2}{4}, & x \geq 0. \end{cases}$$

This function is continuous and has a continuous first derivative (check this). It passes through the origin, and satisfies the equation, hence it is a solution. On the other hand the function $y = 0$ is also a solution passing through the origin. Hence, we have two solutions whose integral curves are tangent at the point $x = 0$ (see the figure). *Q:* Can you construct one more function passing through the origin and satisfying the ODE, and different from the two we discussed? Can you construct infinitely many such functions?

The last thing to discuss here is the meaning of the constant ε in the statement of the existence and uniqueness theorem. Its presence actually implies that the theorem is essentially *local*, and the existence and uniqueness is guaranteed only for some small interval of values of x . This is indeed important as the following example shows.

Example 6. Consider the equation

$$y' = 1 + y^2.$$

Its solution (you should verify it) is given by

$$y(x) = \tan(x + C).$$

For the initial condition $y(0) = 0$ we find $C = 0$ and $y(x) = \tan x$. Recall that if $x \rightarrow \pi/2$ then $\tan x \rightarrow \pm\infty$, therefore the solutions approach infinity on the finite interval of x . Again, in the equation itself the right hand side is not only continuous, but infinitely many times differentiable for any points (x, y) , and still the actual solutions cannot be continued infinitely. This is why we need ε in the statement of the existence and uniqueness theorem.

3.1 *For a mathematically inclined student

I am not planning to prove the existence and uniqueness theorem in our course (if you are interested in the proof you can consult many of the standard ODE textbooks, for example, I especially like for the clarity of exposition, Hirsch, M. W., & Smale, S. (1974). *Differential equations, dynamical systems, and linear algebra*), since it would take too much time. Here I would like to give two very particular examples of the existence and uniqueness proof.

First, consider again the equation

$$y' = ay,$$

where a some constant. We know now that the general solution to this equation is $y(x) = Ce^{ax}$, where C is an arbitrary constant.

Proposition 7. *There are no other solutions to $y' = ay$ except for Ce^{ax} .*

Proof. Indeed, assume that $u(x)$ is also a solution and consider function

$$u(x)e^{-ax}.$$

Its derivative is

$$\frac{d}{dx}(u(x)e^{-ax}) = u'(x)e^{-ax} - au(x)e^{-ax} = au(x)e^{-ax} - au(x)e^{-ax} = 0,$$

which implies that

$$u(x)e^{-ax} = C,$$

or

$$u(x) = Ce^{ax}.$$

■

This proposition *proves* that the only solution to the IVP

$$y' = ay, \quad y(x_0) = y_0$$

is

$$y(x) = y_0e^{a(x-x_0)}.$$

Actually, we can consider even more general case.

Proposition 8. *Consider a separable ODE*

$$y' = f_1(x)f_2(y)$$

in the rectangular Q defined by $a < x < b$, $c < y < d$. Assume that f_1 and f_2 are continuous in Q , and additionally $f_2(y) \neq 0$ for any $c < y < d$. Then through every point $(x_0, y_0) \in Q$ passes one and only one solution.

Proof. Assume first that there is solution $\phi(x)$ to our equation such that $\phi(x_0) = y_0$. It means that

$$\frac{d\phi}{dx}(x) = f_1(x)f_2(\phi(x)),$$

or, since $f_2(y) \neq 0$,

$$\frac{d\phi(x)}{f_2(\phi(x))} = f_1(x) dx.$$

Integrate the last equality from x_0 to x and find

$$\int_{\phi(x_0)=y_0}^{\phi(x)} \frac{d\phi(\xi)}{f_2(\phi(\xi))} = \int_{x_0}^x f_1(\xi) d\xi.$$

Let $F_2(y)$ be any antiderivative of $1/f_2(y)$ and $F_1(x)$ be any antiderivative of $f_1(x)$. Then the last equality takes the form

$$F_2(\phi(x)) - F_2(y_0) = F_1(x) - F_1(x_0).$$

Since $F_2(y)$ is a monotonous function (since its derivative does not change sign), the last equality can be solved for $\phi(x)$:

$$\phi(x) = F_2^{-1}(F_2(y_0) + F_1(x) - F_1(x_0)).$$

Therefore, by assuming the existence of solution $\phi(x)$ we were able to present it in a concrete form, which implies that such solution is *unique*: all functions in the solution are determined through the equation itself and the initial conditions.

On the other hand it is straightforward to show that $\phi(x)$ in this form actually is a solution. For this we need to check the initial conditions and show that that the expression for $\phi(x)$ satisfies the equation. ■

The last proposition may seem too abstract and not very informative. However, it actually shows how non uniqueness can appear in such equations. For this it is necessary and sufficient to have some \hat{y} such that $f_2(\hat{y}) = 0$ and the improper integral

$$\int_{y_0}^y \frac{d\eta}{f_2(\eta)}$$

converges when $y \rightarrow \hat{y}$. In this case there are infinitely many integral curves passing through some points of Q , they all tangent to the integral curve $y = \hat{y}$. If, however, this integral *diverges*, then the uniqueness is not broken.