Introduction to contact geometry

Jessica Zhang

December 6, 2020

Many first-order ordinary differential equations can be written explicitly, i.e., as $G(x,z) = z'(x)$. A solution corresponds to a curve which is everywhere tangent to the associated *slope field*, which assigns a line of slope $G(x,z)$ to every point $(x,z) \in \mathbb{R}^2$. We call each such a vector a *line element*.

Observe that the line element of slope p_0 at (x,z) corresponds to the kernel of the differential form

$$\alpha = dz - p_0 dx.$$

To see this, simply note that $\ker\alpha$ is simply the space spanned by

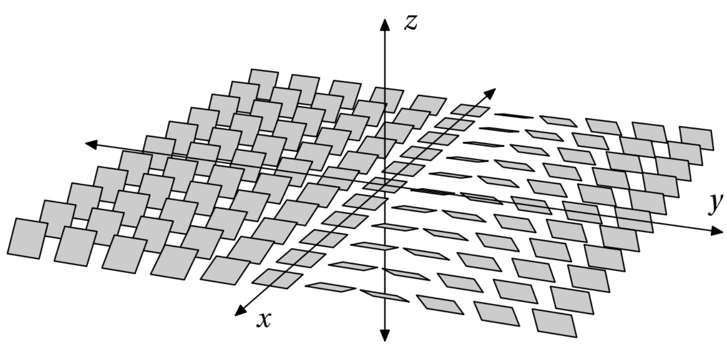
$$\frac{\partial}{\partial x} + p_0 \frac{\partial}{\partial z},$$

which in turn corresponds to the line of slope p_0.

Now consider the space of *all* line elements. One way to think of this space is by identifying the line element of slope p at (x,z) to the point $(x,z,p) \in \mathbb{R}^3$. Note that since the slope is now variable (i.e., p is a variable, not just x and z), the differential form α can be thought of as a 3-form over $\mathbb{R}^3$. In particular, set

$$\omega = dz - p dx$$

and let $\xi = \ker\omega$, as shown below.



Recall that we began with solutions to differential equations. Let $z = z(x)$ be a function in x. More generally than the explicit form $G(x,z) = z'$, we can write a first-order differential equation as $F(x,z,z') = 0$ for some function F. Note that there is an associated curve to $z(x)$ in our copy of $\mathbb{R}^3$, namely

$$x \mapsto (x, z(x), z'(x)).$$

The tangent vector to this curve at any given point $(x_0, z(x_0), z'(x_0))$ is contained in ξ at that point. In particular, a solution to a differential equation corresponds to an *integral curve* of ξ.

Example 1. Consider the differential equation $F : z'(x) - z(x) = 0$. Its solutions are of the form $z(x) = ae^x$, where a is any constant. Its associated curve is thus

$$\gamma : x \mapsto (x, ae^x, ae^x).$$

The tangent vector is the vector corresponding to taking the derivative of each component, so

$$\gamma'(x) = (1, ae^x, ae^x).$$

Note that the tangent vector lives in the tangent space, which has basis $\left\{\frac{\partial}{\partial x}, \frac{\partial}{\partial z}, \frac{\partial}{\partial p}\right\}$, so this means that

$$\gamma'(x) = \frac{\partial}{\partial x} + ae^x \frac{\partial}{\partial z} + ae^x \frac{\partial}{\partial p}.$$

At a given point $\gamma(x_0) = (x_0, z_0, p_0)$ on this curve, we know that

$$\begin{aligned}(dz - pdx)_{\gamma(x_0)}(\gamma'(x_0)) &= (dz - pdx)_{\gamma(x_0)} \left(\frac{\partial}{\partial x}\bigg|_{\gamma(x_0)} + ae^x \frac{\partial}{\partial z}\bigg|_{\gamma(x_0)} + ae^x \frac{\partial}{\partial p}\bigg|_{\gamma(x_0)} \right) \\ &= (dz - p_0 dx) \left(\frac{\partial}{\partial x}\bigg|_{\gamma(x_0)} + ae^x \frac{\partial}{\partial z}\bigg|_{\gamma(x_0)} + ae^x \frac{\partial}{\partial p}\bigg|_{\gamma(x_0)} \right).\end{aligned}$$

But recall that dz is 0 on $\frac{\partial}{\partial x}$ and $\frac{\partial}{\partial p}$, and is 1 on $\frac{\partial}{\partial z}$, and similarly for dx. Thus this is exactly equal to

$$p_0 dx \left(\frac{\partial}{\partial x} \right) + dz \left(ae^x \frac{\partial}{\partial z} \right) = -p_0 + ae^x = 0,$$

since $p_0 = ae^x$. Hence this curve is indeed an integral curve.

Example 2. The second example differential equation we will use is the closely related equation $G : z' - zx = 0$. Its solutions are of the form $z(x) = ae^{x^2/2}$, where a is again a constant. We can check that the tangent vector of the associated curve $\delta : x \mapsto (x, ae^{x^2/2}, axe^{x^2/2})$ is

$$\delta'(x) = (1, axe^{x^2/2}, ae^{x^2/2}(x^2 + 1)).$$

Now for each point $\delta(x_0)$, we can check that

$$(dz - pdx)(\delta'(x_0)) = dz \left(ax_0 e^{x_0^2/2} \frac{\partial}{\partial z} \right) - p_0 dx \left(\frac{\partial}{\partial x} \right) = ax_0 e^{x_0^2/2} - ax_0 e^{x_0^2/2} = 0.$$

(The general idea is that $p_0 = z'(x_0)$ and the $\frac{\partial}{\partial z}$ term in the tangent vector is equal to $z'(x_0)$ as well, so applying $dz - pdx$ to the tangent vector will give us $z'(x_0) - z'(x_0) = 0$ for every point x_0.)

Consider some diffeomorphism

$$f : (x, z, p) \mapsto (X, Z, P)$$

of $\mathbb{R}^3$. Define a function F_1, known as the *transformed differential equation*, which corresponds to F after this diffeomorphism. In particular, define F_1 so that

$$F_1(X, Z, P) = F(x, z, p).$$

This is equivalent to setting F_1 to be the function with $F_1 \circ f = F$.

A solution to F as a differential equation would be $z = z(x)$ so that $F(x, z, z') = 0$. In particular, we set $p = z'(x)$. Restricting our attention to this case, we get a curve

$$f : (x, z(x), z'(x)) \mapsto (X(x), Z(x), P(x)).$$

One natural question to ask is whether or not these transformed coordinates also satisfy the transformed differential equation.

Recall that integral curves of ξ correspond to solutions, so if f preserves integral curves of $dz - pdx$, it must also preserve solutions. It isn't immediately obvious, however, what it means to "preserve integral curves." The following theorem make this idea formal and shows that it is, indeed, the correct intuition.

Theorem 3. *Let f be a* contact transformation*, i.e., a map*

$$f : (x, z, p) \mapsto (X, Z, P)$$

such that f takes integral curves of $\ker(dz - pdx)$ to integral curves of $\ker(dZ - PdX)$. Suppose $x \mapsto z(x)$ is a solution to $F(x, z, z') = 0$. Consider the transformed curve

$$x \mapsto (X(x), Z(x), P(x)) := f(x, z(x), z'(x)).$$

If $(X'(x), Z'(x)) \neq (0, 0)$ for all x, then $Z(x)$ can be thought of as a function of $X(x)$. The curve $X \mapsto Z(X)$ is then a solution to the transformed equation

$$F_1\left(X, Z, \frac{dZ}{dX}\right) = 0.$$

Proof. Since $\gamma : x \mapsto (x, z(x), z'(x))$ is an integral curve of $\ker(dz - pdx)$, we know that $x \mapsto (X(x), Z(x), P(x))$ is an integral curve of $\ker(dZ - PdX)$. Hence we know, in particular, that

$$Z'(x) - PX'(x) = 0.$$

If $X'(x) = 0$, then $Z'(x) = 0$, contradicting our "regularity" condition that $(X'(x), Z'(x)) \neq 0$. Thus $X'(x)$ is nonzero, and so the inverse function theorem implies that we can write x as a function $x(X)$.

The chain rule implies that

$$\frac{dZ}{dX}(X) = \frac{dZ}{dx}(x(X)) \cdot \frac{dx}{dX}(X).$$

Now note that $X(x(X)) = X$, and so $X'(x(X))x'(X) = 1$. Hence it follows that

$$\frac{dx}{dX}(X) = \frac{1}{X'(x(X))},$$

and so we find that

$$\frac{dZ}{dX}(X) = \frac{Z'(x(X))}{X'(x(X))} = P(x(X)).$$

In other words, we know that X, Z, P have the same relationship that x, z, p had on the curve γ, namely that $p = z'(x)$. Hence it follows that

$$F_1\left(X, Z, \frac{dZ}{dX}\right) = F_1(X, Z, P) = F(x, z, z') = 0,$$

as desired. □

But exactly which functions f are contact transformations, anyway? They're precisely those functions which take vectors in $\ker(dz - pdx)$ to vectors in $\ker(dZ - PdX)$, and which take vectors of $T\mathbb{R}^3$ that are *not* in $\ker(dz - pdx)$ to vectors *not* in $\ker(dZ - PdX)$.

Hence at each point, we must have $dZ - PdX$ equal to some nonzero multiple of $dz - pdx$. (Note that X, Z, P are not a different set of coordinates for $\mathbb{R}^3$ and are instead functions of the coordinates x, z, p of $\mathbb{R}^3$. In particular, the differential form $dZ - PdX$ is a 1-form with x, z, p as variables.) In other words, contact transformations $f : (x, z, p) \mapsto (X, Z, P)$ are exactly those functions such that there is a (smooth) nowhere zero function $g : \mathbb{R}^3 \to \mathbb{R}$ with

$$dZ - PdX = gdz - gpdx.$$

In fact, to more explicitly relate this function g to f, we can note that $X = x \circ f$, where the function x represents the x-coordinate of a point, and similarly for Z and P. Hence

$$dZ - PdX = d(z \circ f) - (p \circ f)d(x \circ f),$$

which is precisely the pullback $f^*(dz - pdx)$. This means that contact transformations are functions so that there is a nowhere zero function g with

$$f^*(dz - pdx) = gdz - gpdx = g(dz - pdx).$$

Example 4. Consider the function

$$f : (x, z, p) \mapsto (p, px - z, x).$$

In this case, we know that

$$dZ - PdX = d(px - z) - xdp.$$

But because the differential respects addition/subtraction and follows the product rule, we know that

$$d(px - z) = pdx + xdp - dz,$$

from which we conclude that

$$dZ - PdX = pdx - dz = -1(dz - pdx).$$

Hence this is a contact transformation.

Indeed, consider how f affects transformations to the differential equations in Examples 1 and 2. Recall the definition of

$$F : z' - z = 0.$$

The transformed equation would be

$$F_1 : (xp - z)' - (xp - z) = 0.$$

To see this, simply note that

$$F_1(f(x, z, p)) = F_1(p, px - z, x) = z' - z,$$

which is exactly F. The solutions to F correspond to curves $x \mapsto (x, ae^x, ae^x)$. Using f, these transform to curves $x \mapsto (ae^x, (x - 1)ae^x, x)$.

Now observe that, thinking of the z-coordinate as a function of the x-coordinate, we do in fact find that the p-coordinate corresponds to the derivative:

$$\frac{\partial (x - 1)ae^x}{\partial ae^x} = x.$$

Moreover, this curve corresponds to a solution to F_1. This is because

$$(ae^x x - (x - 1)ae^x)' - (ae^x x - (x - 1)ae^x) = (ae^x)' - ae^x = 0.$$

We can do something similar with the differential equation G from Example 2. In particular, we end up with the transformed equation

$$G_1 : (xp - z)' - p(xp - z) = 0.$$

A solution $x \mapsto (x, ae^{x^2/2}, axe^{x^2/2})$ of G is mapped by f to the curve $x \mapsto (axe^{x^2/2}, ax^2e^{x^2/2} - ae^{x^2/2}, x)$. Observe that, again, the p-coordinate is the derivative of the z-coordinate, with respect to the x-coordinate $axe^{x^2/2}$. We can moreover verify that this curve solves G_1. (In fact, the curve will always solve the transformed equation; the nontrivial part, and the part which Theorem 3 shows, is that the $P = Z'(X)$, to use our old notation.)

Example 5. For a more interesting example of a contact transformation, we begin with a geometric construction. First, let $\gamma : x \mapsto (x, z(x))$ be a curve in $\mathbb{R}^2$. Let δ be one of the two curves in $\mathbb{R}^2$ which lies parallel to γ at a distance $k > 0$ away. Hence δ is given by $X \mapsto (X, Z(X))$ with the requirements

$$\begin{aligned}(X - x)^2 + (Z - z)^2 &= k^2 \\ (\delta(X) - \gamma(x)) \cdot \gamma'(x) &= 0 \\ \delta'(X) &= \gamma'(x).\end{aligned}$$

The first two requirements basically say that δ is a distance k away from γ. (The $\cdot$ is the Euclidean dot product.) The third requirement ensures that the slopes are the same, so the two curves are indeed parallel.

Using our definitions of $p = z'(x)$ and $P = Z'(X)$, and hence of $(1, p) = \gamma'(x)$ and $(1, P) = \delta'(X)$, we can rewrite the conditions as follows:

$$\begin{aligned} (X - x)^2 + (Z - z)^2 &= k^2 \\ X - x + (Z - z)p &= 0 \\ P &= p \end{aligned}$$

We can solve this system (for example, by first solving for $X - x$, and then for $Z - z$) to get

$$X = x \pm \frac{kp}{\sqrt{1 + p^2}}, \qquad Z = z \pm \frac{p}{\sqrt{1 + p^2}},$$

where the $\pm$ comes from the fact that there were two options for δ.

Now observe[1] that

$$dX = dx \pm d\left(\frac{kp}{\sqrt{1 + p^2}}\right) = dx \pm \frac{k}{(\sqrt{1 + p^2})^{3/2}} dp.$$

Similarly, we find that

$$dZ = dz \pm \frac{kp}{(\sqrt{1 + p^2})^{3/2}} dp,$$

from which we conclude that

$$dZ - PdX = dZ - pdX = dz - pdx.$$

Hence this function f, which corresponds to this shifting of a curve γ in $\mathbb{R}^2$, is a contact transformation.

Example 6. As a non-example of a contact transformation, consider the dilation

$$f : (x, z, p) \mapsto (2x, 2z, 2p).$$

This would mean that

$$dZ - PdX = d(2z) - (2p)d(2x) = 2dz - 4pdx,$$

which is not always a constant multiple of $dz - pdx$.

To see how this non-contact transformation violates Theorem 3, consider the function G defined in Example 2, namely

$$G : z' - zx = 0.$$

The transformed differential equation G_1 is thus given by

$$G_1 : \frac{z'}{2} - \frac{zx}{4} = 0.$$

Recall that solutions of G correspond to curves $x \mapsto (x, ae^{x^2/2}, axe^{x^2/2})$. Applying f to these curves, we get the transformed curves $x \mapsto (2x, 2ae^{x^2/2}, 2axe^{x^2/2})$. If we could just replace z' with the p-coordinate, these curves would solve G_1. However, writing $X = 2x$ and hence $2ae^{x^2/2} = 2ae^{X^2/8}$, we find that

$$\frac{d}{dX} 2ae^{X^2/8} = \frac{1}{2} aXe^{X^2/8} \neq aXe^{X^2/8},$$

which is the p-coordinate in terms of X. To put it another way, the z-coordinate of the transformed curve is $z(x) = 2ae^{x^2/2}$, which is not a solution to G_1. This issue arises from the fact that, strictly speaking, the transformed curve does not correspond to a solution to a differential equation, since the p-coordinate is not $z'(x)$.

[1] I may or may not have used Wolfram Alpha.

This gives us some idea of why we might care about this differential form $dz - pdx$ (or, more accurately, about its kernel): A function which maps the integral curves of $\ker(dz - pdx)$ to integral curves of $\ker(dZ - Pdx)$ maps solutions to solutions under the corresponding change of coordinates.

We call $\ker(dz - pdx)$ a *standard contact structure*. It lives in $\mathbb{R}^3$, but was created by looking at tangent vectors in $\mathbb{R}^2$. When $M = \mathbb{R}^n$ and has coordinates $(x_1, \ldots, x_{n-1}, z)$, the standard contact structure lives in $\mathbb{R}^{2n-1} = (x_1, p_1, \ldots, x_{n-1}, p_{n-1}, z)$ and looks like $\ker(dz - \sum p_i dx_i)$.

In general, let M be a smooth n-manifold M. We can take a hyperplane (dimension $n-1$ subspace of the tangent space) V_p at each point p. Each such hyperplane is called a *contact element* (when $M = \mathbb{R}^2$, we called these line elements), and the space of all (p, V_p) is called the *space of contact elements*. This is a $(2n-1)$-dimensional space and, it turns out, can be associated to the kernel of a differential form as well. We call this kernel the *natural contact structure* on the space of contact elements.[2]

The modern definition of a contact structure has been abstracted from this geometric setting. We now define a contact structure on $(2n-1)$-dimensional manifold M as the kernel of some 1-form α which satisfies the condition that $\alpha \wedge (d\alpha)^n \neq 0$ everywhere. Note that a contact structure can only be defined on odd-dimensional spaces. This isn't inconsistent with our earlier construction of a contact structure on $\mathbb{R}^2$ because the contact structure itself actually lived in $\mathbb{R}^3$.

As a note, this definition is not actually equivalent to the definition of the geometric definition of a natural contact structure, since there are many contact structures (in the modern definition) which cannot be obtained by the contact element construction. In fact, part of why we call $\ker(dz - pdx)$ the *standard* contact structure is because the differential form definition allows for other contact structures on $\mathbb{R}^3$.

Example 7. As an example of a nonstandard contact structure on $\mathbb{R}^3$, consider giving $\mathbb{R}^3$ cylindrical coordinates (r, θ, z). Let β be the 1-form

$$\beta = \cos r \; dz + r \sin r \; d\theta.$$

We can check that

$$\beta \wedge d\beta = \left(1 + \frac{\sin r \cos r}{r}\right) dV.$$

For any $r > 0$, however, we can check that $1 + \frac{\sin r \cos r}{r}$ is positive, and so $\ker \beta$ is indeed a contact structure.

We can check that the planes of $\ker \beta$ end up rotating along rays perpendicular to the z-axis, similarly to $\ker(dz - pdx)$. However, whereas $\ker \beta$ makes infinitely many full rotations, the standard contact structure never fully rotates (which can be seen by the fact that p_0 never reaches ∞). Thus we can at least intuitively see that the two contact structures must be globally different.

That being said, the two contact structures still *are* very similar. As mentioned, both involve planes rotating in rays perpendicular to the z-axis; the only difference is in how quickly they rotate. In fact, this property of being locally very similar is not unique to the contact structures at all.

Theorem 8 (Darboux)**.** *Let M be any $(2n-1)$-manifold and let ξ be some contact structure defined on M. Fix some point $p \in M$. Moreover, let ξ_{st} be the standard contact structure on $\mathbb{R}^{2n-1}$. Then there exist open neighborhoods $U \subseteq M$ of p and $V \subseteq \mathbb{R}^{2n-1}$ of the origin in $\mathbb{R}^{2n-1}$ such that there is a diffeomorphism $f : U \to V$ such that $df : \xi|_U \to \xi_{st}|_V$.*

[2] Strictly speaking, the construction we went through does not allow for *all* hyperplanes (which are just tangent lines in the $\mathbb{R}^2$ case). In particular, we only allowed for finite slope. If we allowed for any slope, we would end up looking at a contact structure on $\mathbb{R}^2 \times \mathbb{PR}$, where $\mathbb{PR}$ is the real projective line.