Appendix E : Note on regular curves in Euclidean spaces

In Section III.5 of the course notes we posed the following question:

> *Suppose that U is a connected open subset of $\mathbf{R}^n$ and x, $y \in U$. Is there a continuous curve $\gamma : [a,b] \to U$ such that $\gamma(a) = x$, $\gamma(b) = y$, the curve γ has (continuous) derivatives of all orders, the tangent vector $\gamma'(t)$ is never zero?*

We have shown that x and y can be connected by a broken line curve, and therefore one can partition $[a,b]$ into subintervals such that the additional conditions hold on each subinterval. A little physical experimentation — either with pencil and paper or with wires or strings — strongly indicates that one can modify the broken line curve into a smooth curve with the desired properties. The purpose here is to give a mathematical justification for this physical intuition. Although the underlying concepts are not particularly difficult, verifying that everything works as expected turns out to be more demanding than one might initially expect.

The first step is a refinement of the result on joining points in a connected open subset by broken lines.

Definition. Given a line segment $\gamma : [a,b] \to \mathbf{R}^n$, its direction vector given by the difference $\gamma(b) - \gamma(a)$.

Definition. Suppose that $n \geq 2$. A broken line in $\mathbf{R}^n$ is said to be a *strictly broken line* if it is a concatenation of line segments such that the direction vectors of adjacent line segments are always linearly independent. We need the condition on n in order to ensure the existence of a linearly independent set of two vectors; on the other hand, if U is open and connected in $\mathbf{R}$ then we already know that every pair of points in U form the endpoints of a closed segment that is entirely contained in U, so there is nothing that needs to be proven in this case.

PROPOSITION. *If U is an open connected subset of $\mathbf{R}^n$ and x and y are distinct points of U, then there is a strictly broken line $\gamma : [a,b] \to U$ such that $\gamma(a) = x$ and $\gamma(b) = y$.*

Proof. First of all, we might as well eliminate all constant curves from the list of segments that appear in the construction of the broken line curve, for if we do so we shall still have a broken line joining x to y. Let us desribe broken lines with no constant pieces as reduced broken lines.

Imitating the proof of the result on joining two points by broken line curves, we consider a binary relation $\equiv$ such that $u \equiv v$ if and only if there is a strictly broken line joining u to v. This relation is symmetric, and it also reflexive (but we cannot use the constant curve here; instead, given $z \in U$ choose $\varepsilon > 0$ so that $N_\varepsilon(z) \subset U$ and consider a small square inside U that has z as one of its vertices). The main thing that needs to be checked is transitivity.

Suppose now that $x \equiv y$ and $y \equiv z$, let α and β be reduced broken lines joining x to y and y to z respectively, and let $\mathbf{q}$ and $\mathbf{r}$ be the direction vectors for the last segment in α and the first segment in β respectively. If $\mathbf{q}$ and $\mathbf{r}$ are linearly independent then $x \equiv z$ follows immediately, so suppose this is not the case. Choose $\varepsilon > 0$ such that $N_\varepsilon(y) \subset U$, and choose points y_- and y_+ that are contained in $N_\varepsilon(y)$ and lie on last segment in α and the first segment in β respectively but are not endpoints of these segments. Let $\mathbf{p}$ be a unit vector that is perpendicular to the one-dimensional subspace spanned by $\mathbf{q}$ and $\mathbf{r}$.

Let $\alpha^{\#}$ and $\beta^{\#}$ be the strictly broken lines joining x to y_- and y_+ to z by removing the appropriate small pieces from α and β. Suppose first that $\mathbf{q}$ and $\mathbf{r}$ are positive multiples of each other. Then the concatenation of $\alpha^{\#}$, the segment joining y_- to $y + (\varepsilon/2)\mathbf{p}$, the segment joining the latter to y_+, and $\beta^{\#}$ will be a strictly broken line joining x to z; geometrically, we have just arranged for the curve to take a detour around the point y. As usual, it may be helpful to draw a picture to visualize the construction.

If $\mathbf{q}$ and $\mathbf{r}$ are negative multiples of each other, we again want to arrange a detour so that the curve misses y, but the construction requires slightly more work. Specifically, take the concatenation of $\alpha^{\#}$, the segment joining y_- to $y + (\varepsilon/2)\mathbf{p}$, the segment joining $y + (\varepsilon/2)\mathbf{p}$ to $y - (\varepsilon/2)\mathbf{p}$, the segment joining the latter to y_+, and $\beta^{\#}$; the resulting curve will be a strictly broken line joining x to z.■

Bump functions

We need some tools for taking differentiable functions defined on pieces of a space and constructing something out of the pieces that is globally differentiable. Such constructions are needed in a wide range of geometric and analytical contexts, and the following fact from single variable calculus contains the crucial insight..

PROPOSITION. *There is an infinitely differentible function* $f : \mathbf{R} \to \mathbf{R}$ *such that* $f(x) > 0$ *for* $x > 0$ *and* $f(x) = 0$ *for* $x < 0$.

Since this function is infinitely differentiable it follows that the higher order derivatives satisfy $f^{(n)}(0) = 0$ for all n even though the function is not contstant in any open neighborhood of zero. In particular, this means that there cannot be an infinite series expansion for f as a convergent power series.

Sketch of Proof. Consider the function

$$f(t) = \exp\left(-\frac{1}{t^2}\right)$$

which is defined and infinitely differentiable for $t > 0$. If we can show that $f^{(n)}(0) = 0$ for all n, then we can extend f to an infinitely differentiable function on the whole real line by taking it to be zero for $t \leq 0$.

Since the iterated derivatives of f have the form $g \cdot f$ where g is a rational function of t (use the Leibniz rule repeatedly), the result will follow if we can show $\lim_{t \to 0+} g(t) \cdot f(t) = 0$. This is a straightforward (but perhaps somewhat messy) consequence of L'Hospital's Rule.■

The previous result allows us to construct a large number of infinitely differentiable functions that are constant on entire intervals.

PROPOSITION. *There is an infinitely differentiable function* $B : \mathbf{R} \to \mathbf{R}$ *such that* $B(t) > 0$ *for* $t \in (0, 1)$ *and* $B(t) = 0$ *elsewhere.*

Proof. Take $B(t) = f(t) \cdot f(1 - t)$.■

PROPOSITION. *There is an infinitely differentiable function* $C : \mathbf{R} \to \mathbf{R}$ *such that* $C(t) = 0$ *for* $t \leq 0$, C *is strictly increasing on* $[0, 1]$, *and* $C(t) = 1$ *for* $t \geq 1$.

Proof. Let $L = \int_0^1 B(t)\,dt$ and set $C(t) = \int_0^t B(u)\,du/L$.■

COROLLARY. *A similar result holds if* $[0, 1]$ *is replaced by an arbitrary closed interval* $[a, b]$.■

The following special case yields a method for smoothing out an arbitrary (strictly) broken line.

PROPOSITION. *Let* $\mathbf{a}$ *and* $\mathbf{b}$ *be linearly independent vectors in* $\mathbf{R}^n$ *for some* $n \geq 2$, *and let* $r > 0$ *such that* $N_r(0)$ *contains* $\mathbf{a}$ *and* $\mathbf{b}$. *Suppose that* $\gamma : [-1, 1] \to N_r(0)$ *is the broken line curve defined by* $\gamma(t) = -t\mathbf{a}$ *if* $t \leq 0$ *and* $\gamma(t) = t\mathbf{b}$ *if* $t \geq 0$. *Then there is an* $\varepsilon > 0$ *and an infinitely differentiable curve* $\beta : [-1, 1] \to N_r(0)$ *such that* β' *is never zero and* $\beta = \gamma$ *on* $[-1, -1 + \varepsilon) \cup (1 - \varepsilon, 1]$.

Proof. Let $\varphi : \mathbf{R} \to [0, 1]$ be an infinitely differentiable function such that $\varphi(t) = 1$ for $|t| \geq 1 - \varepsilon$, $\varphi(t) = 0$ for $|t| < \varepsilon$, and φ is strictly increasing on $[-1 + \varepsilon, -\varepsilon]$ and strictly decreasing on $[\varepsilon, 1 - \varepsilon]$. Also, let $\alpha ; [-1, 1] \to N_r(0)$ define the segment joining $\mathbf{a}$ to $\mathbf{b}$ (the image of α lies in $N_r(0)$ because the latter is convex). An explicit formula for this curve is

$$\alpha(t) \quad = \quad \left(\frac{1 - t}{2}\right)\mathbf{a} \; + \; \left(\frac{1 + t}{2}\right)\mathbf{b} \, .$$

Consider the curve

$$\beta(t) \quad = \quad \varphi(t) \cdot \gamma(t) \; + \; (1 - \varphi(t)) \cdot \alpha(t)$$

which agrees with γ on $[-1, -1 + \varepsilon) \cup (1 - \varepsilon, 1]$ and with α on $(-\varepsilon, \varepsilon)$; by the convexity of $N_r(0)$ this curve is contained in the latter.. Since both γ and α are infinitely differentiable away from zero, the same is true for β, and since β is equal to the infinitely differentiable curve α near zero it follows that β is infinitely differentiable there too. This establishes everything except the assertion that β' is never zero.

The first step towards showing the statement about β' is to write out the latter explicitly:

$$\beta'(t) \quad = \quad \varphi(t) \cdot \gamma'(t) \; + \; (1 - \varphi(t)) \cdot \alpha'(t) \; + \; \varphi'(t) \, (\gamma(t) - \alpha(t))$$

It follows that $\beta'(t) = 0$ if and only if the following holds:

$$\varphi'(t) \, (\alpha(t) - \gamma(t)) = \alpha'(t) + \varphi(t) \, (\gamma'(t) - \alpha'(t))$$

By construction we know that $\alpha'(t) = (\mathbf{b} - \mathbf{a})/2$. There are now two cases depending upon whether $-1 + \varepsilon < t < -\varepsilon < 0$ or $1 - \varepsilon > t > \varepsilon > 0$. If $|t| < \varepsilon$ or $t \geq 1 - \varepsilon$ then b is equal to α or γ, and hence $\beta'(t)$ is nonzero at all such points.

Suppose first that $-1 + \varepsilon < t < -\varepsilon < 0$. Then $\gamma(t) = -t\mathbf{a}$ and $\gamma'(t) = -\mathbf{a}$. Furthermore, in this case we have $\varphi'(t) > 0$. The condition $\beta'(t) = 0$ then specializes to the following equation:

$$\frac{\varphi'(t)}{2}\Big((1 + t) \cdot (\mathbf{a} + \mathbf{b})\Big) \quad = \quad -\frac{1}{2}\Big(\big(\varphi(t) + 1\big)\mathbf{a} \; + \; \big(\varphi(t) - 1\big)\mathbf{b} \Big)$$

Since $\mathbf{a}$ and $\mathbf{b}$ are linearly independent, the coefficients of these vectors on both sides of the equation are equal. Consider what this means for the coefficients of $\mathbf{a}$. The coefficient on the left hand side is positive because $\varphi'(t) > 0$ and $t > -1$, but the coefficient on the right hand side is negative because $\varphi(t) > 0$. This means that there cannot be any values of t such that $-1 + \varepsilon < t < \varepsilon < 0$ and $\beta'(t) = 0$.

Suppose now that $1 - \varepsilon > t > \varepsilon > 0$. We then have $\gamma(t) = t\mathbf{b}$ and $\gamma'(t) = \mathbf{b}$; furthermore, we now have $\varphi'(t) < 0$. The condition $\beta'(t) = 0$ now specializes to the following equation:

$$\frac{\varphi'(t)}{2}\Big((1-t)\cdot(\mathbf{a}+\mathbf{b})\Big) \quad = \quad \frac{1}{2}\Big(\,\big(\varphi(t)-1\big)\mathbf{a} \;+\; \big(\varphi(t)+1\big)\mathbf{b}\,\Big)$$

Once again the coefficients of **a** and **b** on both sides of the equation must be equal. Consider what this means for the coefficients of **b**. The coefficient on the left hand side is negative because $\varphi'(t) > 0$ and $t < 1$, but the coefficient on the right hand side is positive because $\varphi(t) > 0$. This means that there cannot be any values of t such that $1 - \varepsilon > t > \varepsilon > 0$ and $\beta'(t) = 0$. This completes the proof that $\beta'(t)$ is nonzero for all values of t in the interval $[-1, 1]$.■

Some elementary changes of variables yield the following more general version of the preceding construction:

PROPOSITION. *Let $\varepsilon > 0$, let $\mathbf{u}$, $\mathbf{v}$ and $\mathbf{w}$ be vectors in $\mathbf{R}^n$ (where $N \geq 2$) such that $\mathbf{u} - \mathbf{v}$ and $\mathbf{w} - \mathbf{v}$ are linearly independent and $\mathbf{u}, \mathbf{w} \in N_\varepsilon(\mathbf{v})$, let $c < d < e$ in $\mathbf{R}$, and suppose that $\gamma : [c, e] \to N_\varepsilon(\mathbf{v})$ is the continuous curve that maps $[c, d]$ linearly to the closed segment joining $\mathbf{u}$ to $\mathbf{v}$ and also maps $[d, e]$ linearly to the closed segment joining $\mathbf{v}$ to $\mathbf{w}$. Then there is a $\delta > 0$ and an infinitely differentiable curve $\beta : [c, e] \to N_\varepsilon(\mathbf{v})$ such that β' is never zero and $\beta = \gamma$ on $[c, c + \delta) \cup (e - \delta, e]$.*

Sketch of proof. Without loss of generality we may assume that the point d is the midpoint of the interval by taking a smaller interval $[d - \eta,\, d + \eta]$ (note that if we can prove the result for the smaller interval it automatically holds for the larger one!). We can then reparametrize the curve's domain to the interval $[-1, 1]$ by translation and multiplication by a positive constant, and the vectors **a** and **b** become suitable multiples of $\mathbf{u} - \mathbf{v}$ and $\mathbf{w} - \mathbf{v}$ respectively. We can then apply the construction in the proposition, and after doing so we reparametrize the smooth curve so that its domain is equal to the interval $[d - \eta,\, d + \eta]$.■

The main result

PROPSITION. *Let U be an open connected subset of $\mathbf{R}^n$ where $n \geq 2$, and let $\mathbf{u}, \mathbf{v} \in U$. Then there is an infinitely differentiable curve $\gamma; [0, 1] \to U$ such that $\gamma(0) = \mathbf{u}$, $\gamma(1) = \mathbf{v}$, and γ' is never zero on $[0, 1]$.*

Proof. We know that ther eis a strictly broken line curve α joining **u** to **v**. There are finitely many points on this curve where it changes direction, and these correspond to finitely many points in the unit interval. Let $0 < t_1 \;\cdots\; < t_k < 1$ be the set of all such points taken in order.

For each i there is an $\varepsilon_i > 0$ such that

$$W_i = N_\varepsilon\big(\alpha(t_i)\big) \quad \subset \quad U$$

and there are associated numbers $\eta_i > 0$ such that

(1) the intervals $(t_i - 2\eta_i,\, t_i + 2\eta_i)$ have disjoint closures,

(2) the curve α maps each interval $(t_i - 2\eta_i,\, t_i + 2\eta_i)$ into the neighborhood W_i.

We may now use the previous construction to modify α on each of the closed intervals $[t_i - \eta_i,\, t_i + \eta_i]$ such that resulting curve γ is infinitely differentiable on the corresponding open intervals and has a nonzero tangent vector at all points of these intervals. On the other hand, by construction the new curve agrees with α on an open subset containing

$$[0,1] \quad - \quad \bigcup_{i=1}^{k} (t_i - \eta_i, t_i + \eta_i)$$

and on this set we know that α is infinitely differentiable and has nonzero tangent vectors everywhere. Theefore the curve γ is infinitely differentiable at all points of $[0,1]$ and has a nonzero tangent vector for all points t in that interval.■