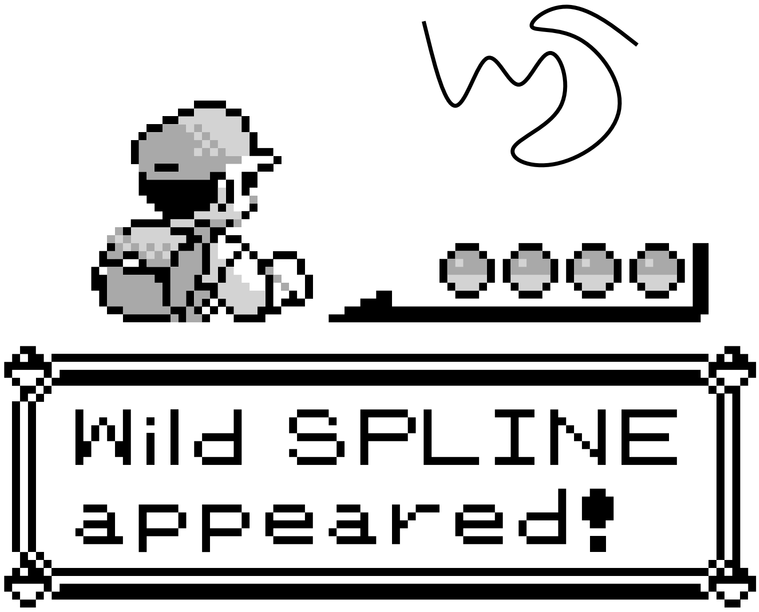
Catmull-Rom 101*

Dányi Bence

2013. március 23.



*Megkapó cím, tudom. Valójában bevezetés a CR görbe rejtelmeibe :-)

1. Görbe

A 2D görbe egy függvény, ami $f : \mathbb{R} \to \mathbb{R}^2$ alakú. Magyarul megeszik egy valós számot (ez lesz a t paramétere), a hasában megemészti, és *két* valós számot tob ki (ezek lesznek a t paraméterhez tartozó $(x; y)$ koordinátái).

1.1. Spline

A spline egy olyan görbe, ami maximum harmadfokú, és kb $C1$-$C2$ folytonos. Mi ez a Cn folytonosság ? (*tl;dr*: ennyiszer tudod lederiválni)

- $C0$ folytonos: a görbe megtörik, de nem szakad meg (meg tudod rajzolni anélkül, hogy felemelnéd a tollad).
- $C1$ folytonos: a görbe nem törik meg, mindenhol sima. (nincsenek sarkai)
- $C2$ folytonos: a görbe *nagyon* sima, azaz kétszer deriválható (mindenhol tudsz érintőkört rajzolni hozzá).

A splineokat többnyire kontrollpontokkal (knots) szeretjük megadni. A spline *szegmensekből* áll, egy szegmens a i. és $i + 1$. csomópont között áll. Ha a spline interpolációs, akkor át is döfi ezeket a pontokat, ha approximációs, akkor csak arra felé megy, de valószínűleg nem megy át rajta.

A splineok megadásánál két dolgot kell a görbéhez megmondani: *milyen* pontokon haladjon át (ha approximál, akkor nem garantáltan megy át rajta), és az egyes pontokba *mikor* érjen oda (amennyiben a görbe bemenő paraméterét úgy képzeljük el, mint az időt, magát a görbét pedig mint egy pontszerű test pályája által leírt útvonalat).

Most intepolációs $C1$ folytonos (majdnem teljesen $C2$) görbékkel foglalkozunk, igaziból tudunk $C2$ folytonos görbét csinálni ezekből az adatokból, de viszonylag bonyolult lineáris egyenletrendszert kell megoldani, szóval nem szeretjük annyira.

1.2. Catmull-Rom „spline”

A CR görbe többnyire $C2$ folytonos, azaz mindenhol, kivéve a csomópontokban (ott csak $C1$). Cserébe egyszerű számolni, mert a megadásánál azt is megmondjuk, (azon kívül, hogy *milyen* pontokon *mikor* haladjon át) hogy *milyen* sebességgel (érintővel) érkezzen az egyes pontokba. Így sokkal egyszerűbb egyenleteket kell megoldani.

2. Kezdjünk neki!

Meg akarjuk konstruálni a CR görbét. Az egyszerűség kedvéért kezdjük úgy, hogy egy dimenzióban dolgozunk, egy szegmenssel, és az időtaromány $[0;1]$. Az egész egy olyan koordinátarendszerben zajlik, ahol a vízszintes tengely az idő (t), a függőleges pedig $f(t)$.

Mit tudunk, mit keresünk? Keressük azt az $f(t)$ függvényt, ami a $t = 0$ időpontban egy általunk választott $f(t)$ értéket vesz fel, tehát $f(0) = x_0$, $t = 1$ időpontban pedig egy másik, általunk választottat ($f(1) = x_1$). Ez lesz a két csomópontunk. Megszorításokat teszünk f deriváltjára is: a $t = 0$ időpontban $f'(0) = v_0$ „sebességgel", azaz meredekséggel induljon (az x_0 értékből), és $t = 1$ időpontban az x_1 pontba $f'(1) = v_1$ sebességgel érkezzen. Ezekből a megkötésekből ki fogjuk tudni számolni a megoldást. Tehát

a) $f(0) = x_0$

b) $f(1) = x_1$

c) $f'(0) = v_0$

d) $f'(1) = v_1$

Hogyan fog kinézni a függvényünk? Harmadfokú, tehát az alakja

$$f(t) = a_3t^3 + a_2t^2 + a_1t + a_0$$

Lesz, nyilván az együtthatókat keressük. Helyettesítsünk be $t = 0$-t, ekkor megkapjuk, hogy $f(0) = a_0$, tehát $a_0 = x_0$ (az a pont miatt). Maradt három ismeretlen.

Deriváljuk le f-et:

$$f'(t) = 3a_3t^2 + 2a_2t + a_1$$

Helyettesítsünk be megint $t = 0$-t, kijön, hogy $f'(0) = a_1$, azaz $a_1 = v_0$ (a c pont miatt). Már csak két ismeretlen van! Írjuk föl f-et, és a deriváltját:

$$\begin{aligned}
f(t) &= a_3t^3 + a_2t^2 + v_0t + x_0 \\
f'(t) &= 3a_3t^2 + 2a_2t + v_0
\end{aligned}$$

Helyettesítsünk be $t = 1$-t, kijön, hogy $f(1) = a_3 + a_2 + v_0 + x_0$, azaz $a_3 + a_2 + v_0 + x_0 = x_1$ (b miatt). A derivált $f(1) = 3a_3 + 2a_2 + v_0$ lesz, amiből következik, hogy $3a_3 + 2a_2 + v_0 = v_1$ (d miatt).

$$\begin{aligned}
a_3 + a_2 + v_0 + x_0 &= x_1 \qquad (1) \\
3a_3 + 2a_2 + v_0 &= v_1 \qquad (2)
\end{aligned}$$

Vonjuk ki (2) egyenletből 2(1)-t:

$$a_3 - v_0 - 2x_0 = v_1 - 2x_1$$

Fejezzük ki a_3-t:

$$a_3 = v_1 - 2x_1 + v_0 + 2x_0$$

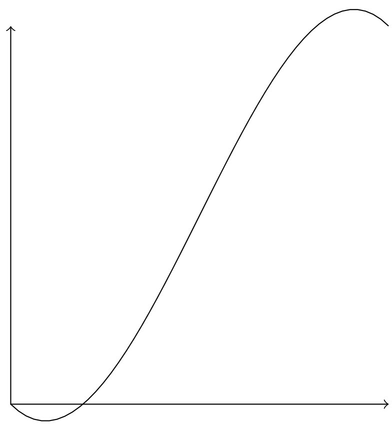
Helyettesítsünk vissza (1) egyenletbe, és határozzuk meg a_2-t:

$$a_2 = 3x_1 - 2v_0 - 3x_0 - v_1$$

Készen vagyunk, az együtthatók:

$$\begin{aligned}
a_0 &= x_0 \\
a_1 &= v_0 \\
a_2 &= -3x_0 - 2v_0 + 3x_1 - v_1 \\
a_3 &= 2x_0 + v_0 - 2x_1 + v_1
\end{aligned}$$

Próbáljuk ki! Legyen $f(0) = 0, f(1) = 1, f'(0) = -1, f'(1) = -1$, tehát $f(t) = -4t^3 + 6t^2 - t$:



Csillámgyönyörű! Mi történik, ha több szegmensünk van? Illetve ha a sebességeket nem akarjuk/tudjuk megmondani? Nézzük tovább...

3. Bedurvulunk

A CR görbét azért szeretjük, mert egyszerű: a csomópontok sebességére jó képletet ad:

$$v_i = \frac{\frac{r_{i+1}-r_i}{t_{i+1}-t_i} + \frac{r_i-r_{i-1}}{t_i-t_{i-1}}}{2}$$

Szemléletesen tehát az i csomópontban a sebesség a két szomszédos pont által meghatározott egyenes meredeksége (ha az időintervallumok egyenlőek, de ezt nem feltételül akarjuk).

Meg szeretnénk oldani azt is, hogy az időtartomány tetszőleges legyen, magyarán ne $[0;1]$ között fusson, hanem bárhol ($[t_0;t_1]$ intervallumon). Ezt hívják okos emberek *non-uniform* spline-nak.

Fussunk neki még egyszer. Egy olyan függvényt keresünk, aminek $f(t_0)$-ban adott az értéke és a meredeksége (az eddigi $f(0)$ helyett), $f(t_1)$-ban szintén (az eddig $f(1)$ helyett). A paraméter $[t_0;t_1]$ között fog futni. Hogyan fog kinézni a függvényünk?

$$f(t) = a_3(t-t_0)^3 + a_2(t-t_0)^2 + a_1(t-t_0) + a_0$$

Ha behelyettesítünk $f(t_0)$-t, akkor $a_0 = x_0, a_1 = v_0$ hasonlóan az előzőekhez kijön. Maradt két ismeretlen, írjuk föl szokás szerint az egyenleteket (a függvény és deriváltja értékét mondják meg a t_1 helyettesítéskor):

$$\begin{aligned} a_3(t_1-t_0)^3 + a_2(t_1-t_0)^2 + v_0(t_1-t_0) + x_0 &= x_1 &\quad (3) \\ 3a_3(t_1-t_0)^2 + 2a_2(t_1-t_0) + v_0 &= v_1 &\quad (4) \end{aligned}$$

(3)-t leosztjuk $(t_1 - t_0)$-val.

$$a_3(t_1-t_0)^2 + a_2(t_1-t_0) + v_0 + \frac{x_0}{t_1-t_0} = \frac{x_1}{t_1-t_0} \quad (5)$$

Most (4)-ből kivonunk 2(5)-t.

$$a_3(t_1-t_0)^2 - v_0 - \frac{2x_0}{t_1-t_0} = v_1 - \frac{2x_1}{t_1-t_0}$$

Kifejezzük a_3-t.

$$a_3 = \frac{v_1+v_0}{(t_1-t_0)^2} + \frac{2(x_0-x_1)}{(t_1-t_0)^3}$$

Visszahelyettesítjük (5)-be, és kifejezzük a_2-t (nem részletezem, négy alapművelettel különösebb trükkök nélkül kijön)

$$a_2 = \frac{3(x_1-x_0)}{(t_1-t_0)^2} - \frac{v_1+2v_0}{t_1-t_0}$$

Összeállt a kép, a *non-uniform* harmadfokú görbék együtthatói:

$$\begin{aligned} a_0 &= x_0 \\ a_1 &= v_0 \\ a_2 &= \frac{3(x_1-x_0)}{(t_1-t_0)^2} - \frac{v_1+2v_0}{t_1-t_0} \\ a_3 &= \frac{2(x_0-x_1)}{(t_1-t_0)^3} + \frac{v_1+v_0}{(t_1-t_0)^2} \end{aligned}$$

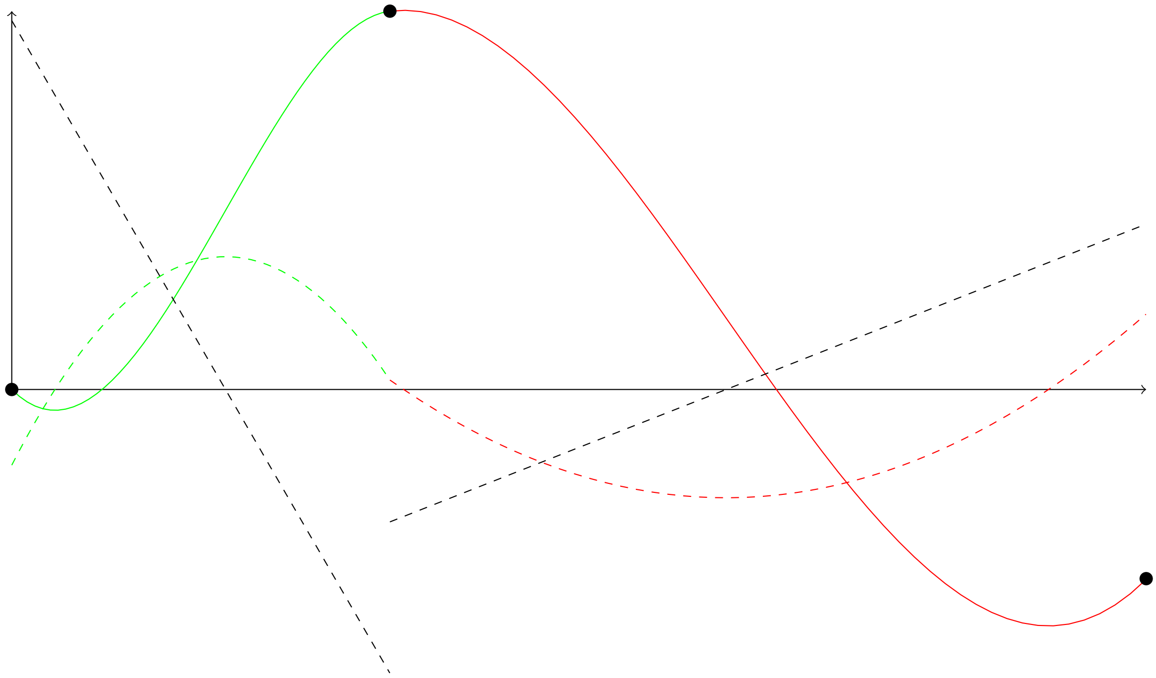
Minden eszköz rendelkezésünkre áll, hogy több szegmensből álló, tetszőleges paraméterezésű 1D görbét (oké, egy sima egyváltozós függvényt) faragjunk!

Legyen a görbénk $r(t)$, a görbe menjen át $r(0) = 0, r(1) = 1, r(3) = -0.5$ pontokon (megadtuk a csomópontokat, és az időpontokat is!). A sebesség képlete az első és utolsó pontra nem működik (hisz nincsen az első „előtti" illetve utolsó „utáni" pont), legyen a kezdősebesség -1, a végsebesség pedig 1.

A közbenső pont sebessége a képlet alapján

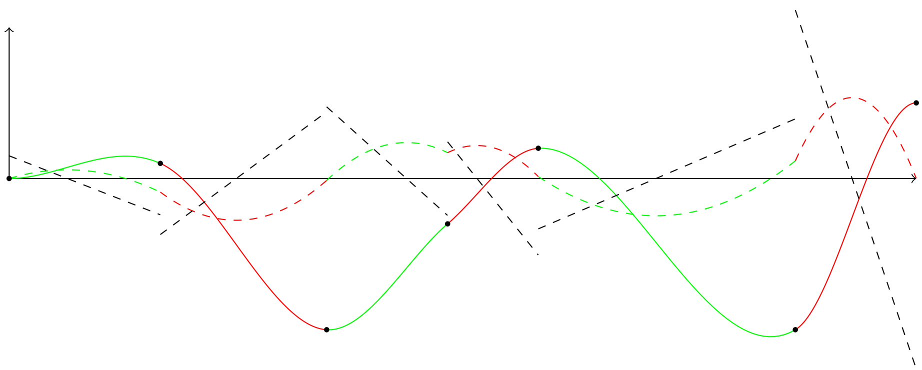
$$v_1 = \left(v_i = \frac{\frac{r_{i+1}-r_i}{t_{i+1}-t_i} + \frac{r_i-r_{i-1}}{t_i-t_{i-1}}}{2} \right) = \frac{\frac{r_2-r_1}{t_2-t_1} + \frac{r_1-r_0}{t_1-t_0}}{2} = \frac{\frac{-0.5-1}{3-1} + \frac{1-0}{1-0}}{2} = 0.125$$

Innen az első szegmens $f_0(t) = -2.875t^3 + 4.875t^2 - t$, a második szegmens pedig $f_1(t) = 0.656(t-1)^3 - 1.750(t-1)^2 + 0.125(t-1) + 1$. Amikor a az $f(t)$ t-je $[0; 1]$ között jár, akkor az f_0-n futunk végig $[0; 1]$-en, mikor $[1; 3]$ között, akkor pedig f_1-en futunk végig $[1; 3]$-n. A deriváltak is fel vannak rajzolva (színes szaggatottan az első, fekete szaggatottan a második), de csak szemléltetésképpen, függőlegesen össze vannak nyomva. Látszik, hogy az első derivált folytonos, és a kontrollpontokon kívül mindenhol deriválható, a második derivált ennek megfelelően a kontrollpontokban szakad, de egyébként folytonos. Ha nem Catmull-Rom görbét rajzoltunk volna, hanem az egyenletek megoldásakor a sebességeket is ismeretlennek fogtuk volna föl (és ezzel egyidőben a második deriváltakra is tettünk volna kitételeket), akkor a megoldásban a második derivált is folytonos lenne, így a görbe *mindenhol* C2 folytonos volna.



$$\begin{aligned} f_0(t) &= (-2.875)(t-0)^3 + (4.875)(t-0)^2 + (-1.000)(t-0) + (0.000) \\ f_1(t) &= (0.656)(t-1)^3 + (-1.750)(t-1)^2 + (0.125)(t-1) + (1.000) \end{aligned}$$

3.1. Hosszabb példa

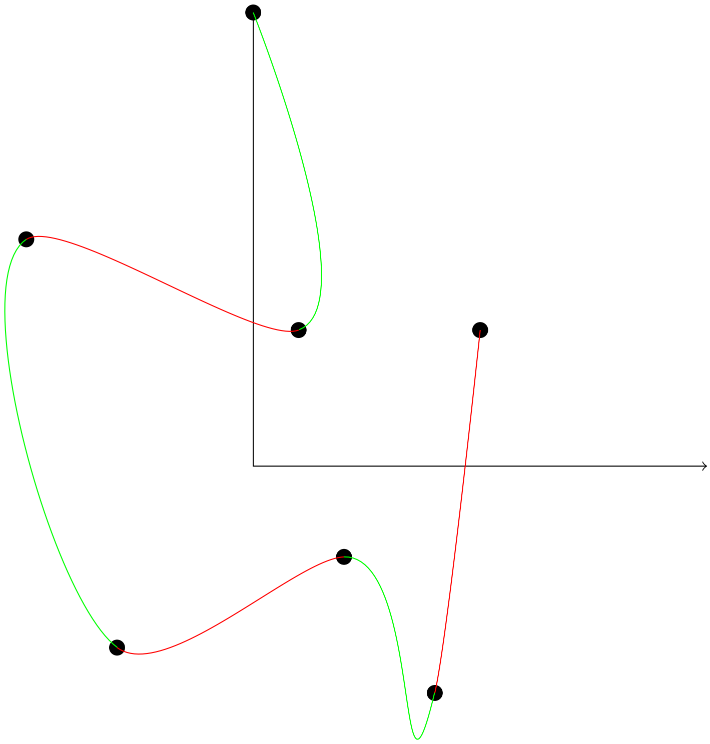


$$
\begin{aligned}
f_0(t) &= (-0.650)(t-0)^3 + (0.750)(t-0)^2 + (0.000)(t-0) + (0.000) \\
f_1(t) &= (1.229)(t-1)^3 + (-1.852)(t-1)^2 + (-0.450)(t-1) + (0.100) \\
f_2(t) &= (-1.497)(t-2.1)^3 + (2.370)(t-2.1)^2 + (-0.062)(t-2.1) + (-1.000) \\
f_3(t) &= (-2.080)(t-2.9)^3 + (1.213)(t-2.9)^2 + (0.854)(t-2.9) + (-0.300) \\
f_4(t) &= (0.713)(t-3.5)^3 + (-1.665)(t-3.5)^2 + (0.064)(t-3.5) + (0.200) \\
f_5(t) &= (-4.946)(t-5.2)^3 + (5.570)(t-5.2)^2 + (0.585)(t-5.2) + (-1.000)
\end{aligned}
$$

4. Közel a cél

Most pedig általánosítsunk 2 dimenzióra. Eddig a kontrollpontok úgy működtek, hogy „*Ha t itt jár, akkor ilyen x valós értéket vegyen föl*". 2D-ben *teljesen* ugyanígy működik, egyik képlethez sem kell hozzányúlni: az x vektor lesz (eddig is az volt, 1 dimenziós!), és minden ugyanúgy működik!

Ennek analógiájára működik a 3D görbe, sőt, tetszőleges dimenzióra tudunk így általánosítani.



$$
\begin{aligned}
f_{0_x}(t) &= (-0.384)(t-0)^3 + (0.531)(t-0)^2 + (0.000)(t-0) + (0.000)\\
f_{0_y}(t) &= (0.707)(t-0)^3 + (-1.334)(t-0)^2 + (0.000)(t-0) + (1.000)\\
f_{1_x}(t) &= (2.004)(t-1.2)^3 + (-2.075)(t-1.2)^2 + (-0.387)(t-1.2) + (0.100)\\
f_{1_y}(t) &= (-1.944)(t-1.2)^3 + (1.981)(t-1.2)^2 + (-0.149)(t-1.2) + (0.300)\\
f_{2_x}(t) &= (-0.220)(t-1.9)^3 + (0.691)(t-1.9)^2 + (-0.345)(t-1.9) + (-0.500)\\
f_{2_y}(t) &= (0.697)(t-1.9)^3 + (-1.268)(t-1.9)^2 + (-0.232)(t-1.9) + (0.500)\\
f_{3_x}(t) &= (-0.501)(t-3.1)^3 + (0.667)(t-3.1)^2 + (0.361)(t-3.1) + (-0.300)\\
f_{3_y}(t) &= (-0.861)(t-3.1)^3 + (1.315)(t-3.1)^2 + (-0.264)(t-3.1) + (-0.400)\\
f_{4_x}(t) &= (0.109)(t-4)^3 + (-0.304)(t-4)^2 + (0.344)(t-4) + (0.200)\\
f_{4_y}(t) &= (0.494)(t-4)^3 + (-0.881)(t-4)^2 + (0.011)(t-4) + (-0.200)\\
f_{5_x}(t) &= (-0.933)(t-5.5)^3 + (0.533)(t-5.5)^2 + (0.167)(t-5.5) + (0.400)\\
f_{5_y}(t) &= (-10.000)(t-5.5)^3 + (6.800)(t-5.5)^2 + (0.700)(t-5.5) + (-0.500)
\end{aligned}
$$