XIII. Badanie Ciał Krzywych

Można śmiało powiedzieć, że ładne, gładkie, krzywoliniowe ciało jest łatwiejsze dla oczu niż ciało o ostrych krawędziach. Można też powiedzieć, że poruszanie się po ładnej, gładkiej krzywej jest zazwyczaj lepsze niż poruszanie się w szarpanym stylu. Krzywe i ruch mają różne właściwości, które można zbadać, takie jak okresowość (powtarzające się wzory) i jednolitość. W tym rozdziale przyjrzymy się kilku krzywiznom, które można łatwo i szybko wygenerować, aby uzyskać bardzo praktyczne wartości. To, co czyni te krzywe ważniejszymi od ogólnych krzywych, to to, że można nimi łatwo manipulować. Takie krzywe mogą być używane do płynnego i wygodnego poruszania kamerą po ekranie; ponadto można użyć tych krzywych do wygenerowania gładkich powierzchni przypominających wzniesienia lub innych interesujących kształtów i geometrii.

Konkretnie tematy poruszane tu są następujące:

■ Krzywe Béziera

■ Węzły wektora

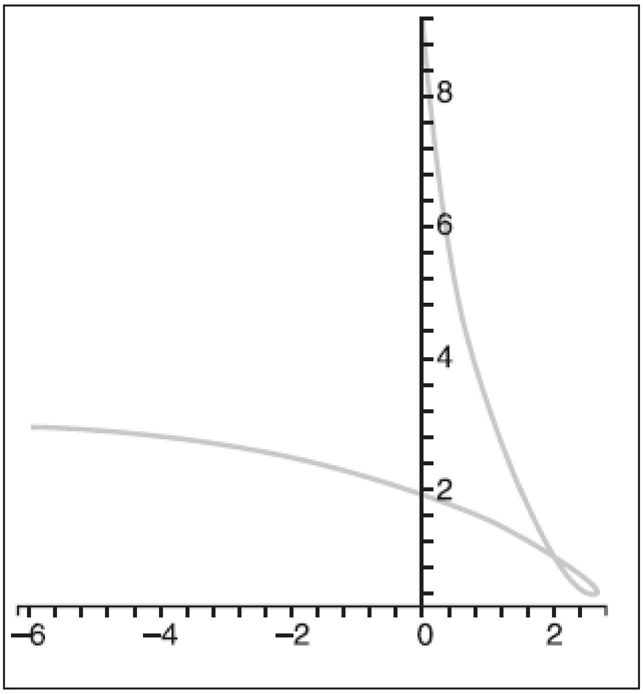
■ B-sklejane

■ NURB

■ Powierzchnie

Krzywe składane

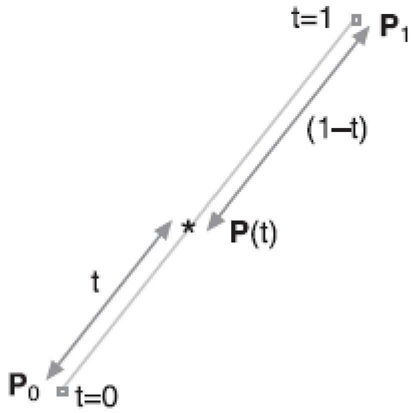
Krzywe, takie jak przedstawione na rysunku 13.1,



występują w różnych postaciach, kształtach i stylach. Szczególnie interesujące w tej sekcji są krzywe składane 2D. W skrócie, krzywa składana jest prostą wielomianową funkcją, która jest globalnie elastyczna i gładka. Co daje krzywa składana tej ekstremalnej elastyczności jest fakt, że są one opisane za pomocą punktów kontrolnych, które są wierzchołkami w przestrzeni, które służą do prowadzenia krzywej. (Można uważać punkty kontrolne za magnesy, przyciągając do nich krzywą.) Wielobok generowany przez zestaw punktów kontrolnych nazywany jest wielokątem kontrolnym. Ani punkty kontrolne, ani wielobok kontrolny nie są nieodłącznie częścią krzywej. Raczej oba służą jako przewodnik po zachowaniu się krzywej. Jak wspomniano wcześniej, krzywe składane są wielomianowymi funkcjami, które, jak dowiedziałeś się w części 12, tworzą ciekawą i użyteczną formę, gdy chcesz określić położenie wielomianu w przyrostach ustalonych odstępach czasu. Ponieważ krzywe składane mają wymaganą elastyczność, jednak nie mogą być funkcjami; muszą być one funkcjami częściowymi. Przecież widziałeś w Części 1, że funkcja jest ograniczona do monotonicznej funkcji jej parametrów, a zatem kształt taki jak okrąg nie może być wyrażony jako funkcja. (Oczywiście, możesz użyć brudnych sztuczek pokazanych w pierwszym rozdziale i sprawić, że moje poprzednie zdanie będzie fałszywe, ale ogólnie rzecz biorąc, nie jest prawdą, że możesz przekształcić dowolne równanie w funkcję.) W związku z tym, ponieważ funkcje są zbyt restrykcyjne, możesz użyj równań parametrycznych. Oferują elastyczność we wszystkich kierunkach, ale zachowują równanie, ponieważ są w rzeczywistości zestawem funkcji odpowiednich dla osi. Dlatego też wypiszesz krzywe sklejane używając równań parametrycznych, które nadal pozwalają na równania wielomianowe w każdej z osi. Istnieje kilka typów krzywych, w tym:

- Krzywe Béziera
- Kardynalne krzywe składane
- B-sklejane
- NURB

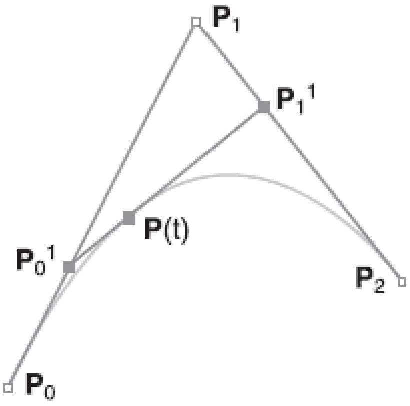
Każda krzywa omawiana tu może być postrzegany jako forma interpolacji. Powinieneś już znać kilka form interpolacji, takich jak liniowa, sferyczna i odwrotna. Wszystkie te interpolacje są ustalonymi przyrostami parametru w funkcji, jak pokazano na rysunku 13.2.



Zwykle są łatwiejsze w obsłudze, gdy są wyrażone w procentach, a więc przy parametrach w zakresie od 0 do 1.

Krzywa Béziera

Krzywa Béziera pochodzi od francuskiego matematyka Pierre'a Béziera, który w latach 70. pracował dla znanej europejskiej firmy motoryzacyjnej Renault. Bézier użył tych krzywych, aby łatwo opisać gładką powierzchnię, jaką dziś widzisz w samochodach. Oczywiście, jesteś bardziej zainteresowany wykorzystaniem tych krzywych do generowania rzeczy, takich jak ukształtowanie terenu lub ruch kamery, lub do reprezentowania innych typów obiektów w przestrzeni 3D. Na przykład Tomb Raider prawdopodobnie używa krzywych Béziera do ruchu kamery; pozwala to na płynne przejście pozycji bez wymuszania stałych pozycji kamery. Uzyskanie równania krzywej Béziera jest proste. Krzywa Béziera może być faktycznie postrzegana jako interpolacja liniowa dwóch równych interpolacji liniowych, jak pokazano na rysunku 13.3.



To brzmi jak bzdura, ale tak nie jest. Załóżmy na przykład, że wybieramy trzy wierzchołki a, b i c. Łącznie te wierzchołki mogą być określane jako punkty kontrolne krzywej. Ponadto możesz zdefiniować trójkąt generowany przez te trzy wierzchołki jako wielokąt kontrolny. Ponieważ krzywa Béziera jest interpolacją liniową dwóch interpolacji liniowych, należy najpierw określić, jakie są pierwsze dwie interpolacje liniowe. Ponieważ masz trzy wierzchołki, możesz zdefiniować swoją pierwszą interpolację jako interpolację liniową od a do b, a drugą interpolację liniową jako między b i c. Oprócz tego możesz pozwolić, aby ich równania miały wspólny parametr t. Tak więc, jeśli wybierzesz jakąkolwiek wartość dla t, otrzymasz dwie współrzędne - lub, jeśli wolisz, dwie interpolacje. Od tego momentu możesz po prostu interpolować od pierwszej współrzędnej do drugiej, używając dokładnie tego samego parametru t; to, co kończysz, to krzywa Béziera. Matematyka jest następująca:

$$f(t) = t\mathbf{b} + (1-t)\mathbf{a}$$
$$g(t) = t\mathbf{c} + (1-t)\mathbf{b}$$

$$\begin{aligned} h(t) &= t \cdot g(t) + (1-t) \cdot f(t) \\ &= t\left(t\mathbf{c} + (1-t)\mathbf{b}\right) + (1-t)\left(t\mathbf{b} + (1-t)\mathbf{a}\right) \\ &= \mathbf{c}t^2 + \mathbf{b}t - \mathbf{b}t^2 + t\mathbf{b} + \mathbf{a} - \mathbf{a}t - t^2\mathbf{b} - \mathbf{a}t + \mathbf{a}t^2 \\ &= t^2\left(\mathbf{a} - 2\mathbf{b} + \mathbf{c}\right) + t\left(-2\mathbf{a} + 2\mathbf{b}\right) + \mathbf{a} \end{aligned}$$

Jak na razie dobrze; możesz teraz wygenerować krzywą, która może być sterowana przez trzy punkty kontrolne. Zanim przejdziesz dalej, przyjrzyjmy się najpierw kilku właściwościom, które wykazuje ta krzywa. Na początek, jeśli ustawisz t = 0, otrzymasz pierwszy wierzchołek, który jest a. Podobnie, jeśli ustawisz t = 1, otrzymasz drugi punkt końcowy, którym jest c. Jeśli jednak spróbujesz wyizolować b, zrobisz to szybko zauważają, że nie można tego zrobić. Bez względu na to, jaką wartość wybierzesz dla t, b generalnie nie będzie na krzywej, ale nadal będzie mieć pewną "kontrolę ciągnięcia" nad ogólnym aspektem krzywej. Mówię "ogólnie" tylko dlatego, że mogłeś mieć krzywą Béziera, która przerodziła się w linię; w tej konkretnej sytuacji można rzeczywiście uzyskać krzywą, w której b jest częścią linii. Jednak z tym zauważonym wyjątkiem, dla kwadratowej krzywej Béziera (to jest drugiego stopnia), nie można wygenerować krzywej, dla której b jest rozwiązaniem krzywej. Więc wiesz, że masz do czynienia z trzema punktami kontrolnymi, ale co z czterema z nich? To też jest łatwe. Aby uzyskać równanie sześciennej krzywej Béziera (to znaczy trzeciego stopnia), po prostu weź dwie sześcienne krzywizny Béziera i liniowo interpoluj z jednego do drugiego. Proces przebiega następująco:

$$f(t) = t^2(\mathbf{a} - 2\mathbf{b} + \mathbf{c}) + t(-2\mathbf{a} + 2\mathbf{b}) + \mathbf{a}$$
$$g(t) = t^2(\mathbf{b} - 2\mathbf{c} + \mathbf{d}) + t(-2\mathbf{b} + 2\mathbf{c}) + \mathbf{b}$$

$$\begin{aligned} h(t) &= t \cdot g(t) + (1-t) \cdot f(t) \\ &= t\left(t^2(\mathbf{b} - 2\mathbf{c} + \mathbf{d}) + t(-2\mathbf{b} + 2\mathbf{c}) + \mathbf{b}\right) + (1-t)\left(t^2(\mathbf{a} - 2\mathbf{b} + \mathbf{c}) + t(-2\mathbf{a} + 2\mathbf{b}) + \mathbf{a}\right) \\ &= t^3(-\mathbf{a} + 2\mathbf{b} - 2\mathbf{c} + \mathbf{d}) + t^2(3\mathbf{a} - 6\mathbf{b} + 3\mathbf{c}) + t(-3\mathbf{a} + 3\mathbf{b}) + c \end{aligned}$$

W tym momencie powinieneś być w stanie zobaczyć, jak rozszerza się to na wyższe krzywe Béziera, jak pokazano na rysunku 13.4.

Wystarczy wziąć dwie krzywe o mniejszym stopniu i liniowo interpolować między nimi. Ten rodzaj definicji może być jednak dość kłopotliwy, aby wyprowadzić go za każdym razem, gdy chcesz użyć równania wyższego stopnia. Ze względu na ich formę, równanie krzywej Béziera można uzyskać w znacznie bardziej bezpośredniej metodzie z samą obserwacją, że współczynnik punktów kontrolnych można wyrazić w następujący sposób:

$$\begin{aligned} 1 &= 1 + t - t \\ &= \left(t + (1-t)\right)^n \\ &= \sum_{i=0}^{n} \binom{n}{i} t^i (1-t)^{n-i} \end{aligned}$$

Ponieważ krzywa Béziera definiowana jest jako interpolacja interpolacji (rekursywnie), można zapisać jej równanie jako iloczyn interpolacji liniowej. Jeśli rozwiążesz to równanie za pomocą współczynnika dwumianowego, możesz wyrazić współczynnik i-tego wierzchołka z n punktów kontrolnych. W konsekwencji współczynnik punktu kontrolnego można wyrazić następująco:

$$B_{i,n}(t) = \binom{n}{i} t^i (1-t)^{n-i}$$

Tak więc równanie krzywej Béziera można zapisać w następujący sposób: gdzie pi jest i-tym punktem kontrolnym:

$$p(t) = \sum_{i=0}^{n} B_{i,n}(t)\mathbf{p}_i$$

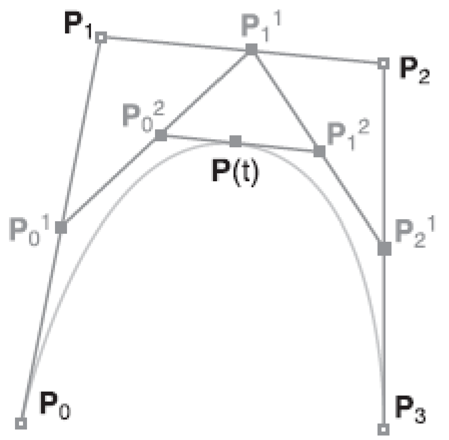
Daje to szybki i łatwy sposób obliczenia różnych czynników krzywej. Jeśli chcesz być nieco bardziej optymalny w czasie wykonywania, możesz zapisać równanie w typowej wielomianowej notacji lub możesz napisać to za pomocą wzoru Hornera, o którym mowa w poprzedniej części. W tej postaci krzywa jest wyrażana jako liniowa kombinacja punktów kontrolnych. Możesz na to spojrzeć jako przestrzeń wektorową, gdzie Bi, n to zestaw funkcji, które tworzą przestrzeń wektorową. Można również powiedzieć, że te funkcje są podstawowymi funkcjami krzywej Béziera, które można wykreślić. Jeśli to zrobisz, zauważysz, że suma krzywej w dowolnym punkcie wynosi 1, co jest wymogiem określonym przez definicję podstawy. Ponadto zauważysz, że funkcje mają zakres od 0 do 1. Ta późniejsza obserwacja oznacza, że ruch jednego z punktów kontrolnych wpłynie na całą długość krzywej, gdzie zmiana jest proporcjonalna do amplitudy funkcji bazowej. To powinno być jasne, jeśli po prostu przedstawisz matematykę w tych kategoriach, coś, co zrobię dalej. Do tej pory patrzyłeś na krzywe Béziera jako liniowe interpolacje liniowych interpolacji. To dobra optyka dla tej krzywej, ale nie jest jedyna. Innym eleganckim sposobem napisania równania krzywej Béziera jest zapisanie go w postaci macierzowej. W przypadku sześciennej krzywej Béziera wynikowe równanie w postaci macierzowej wygląda następująco:

$$p(t) = \begin{bmatrix} t^3 & t^2 & t & 1 \end{bmatrix} \begin{bmatrix} -1 & 3 & -3 & 1 \\ 3 & -6 & 3 & 0 \\ -3 & 3 & 0 & 0 \\ 1 & 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} P_1 \\ P_2 \\ P_3 \\ P_4 \end{bmatrix}$$

Ta prosta, ale mocna metoda napisania równania krzywej Béziera oświeca cię jeszcze jednym sposobem, aby spojrzeć na krzywe Béziera. Jeśli po prostu uważasz, że iloczyn dwóch pierwszych macierzy jest jedną macierzą, macie pomnożenie dwóch macierzy. Dokładniej, masz iloczyn macierzy jednego wiersza z jedną macierzą kolumn, przy czym ta ostatnia jest zbiorem punktów kontrolnych. W konsekwencji krzywą Béziera można zatem wyrazić jako sumę ważoną punktów kontrolnych. Punkty kontrolne mają różną wagę, koncepcję, którą ponownie można wizualizować jako siłę magnetyczną. Im mocniejszy magnes, tym silniejsze jest ciągnięcie, a im bliżej krzywa wygina się w kierunku punktu kontrolnego. Wrócę do tego, ale pamiętaj o tym, czytając bardziej zaawansowany materiał w tym rozdziale. Matryce nie są jedynym sposobem wyrażenia równania krzywej. W rzeczywistości, jak zobaczycie w tym rozdziale, wszystkie krzywe mogą być zapisane przy użyciu następującego ogólnego równania (z jednym zaznaczonym wyjątkiem), gdzie B jest zbiorem podstawowych funkcji, a P jest zbiorem punktów kontrolnych.

Ocena krzywej Béziera

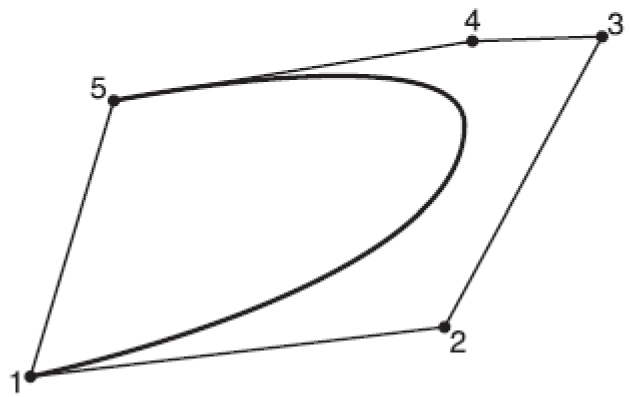
Istnieje wiele sposobów oceny krzywej Béziera; Metoda, której używasz, zależy od tego, w jaki sposób wykorzystasz krzywą. Na przykład, jeśli planujesz użyć krzywej Béziera do płynnego poruszania kamerą po całym świecie, możesz po prostu wybrać pozycję bieżącego bohatera jako pierwszy punkt kontrolny. Ostatnim punktem kontrolnym będzie pozycja, w której twoja postać będzie prawdopodobnie w x milisekundach. Na koniec wybierz inne punkty, aby wielokąt kontrolny nie przecinał się z czymkolwiek na świecie, a kamera przesunie się od pierwszego do ostatniego punktu, przesuwając w kierunku innych wierzchołków. W tej sytuacji prawdopodobnie chcesz wybrać parametr t tylko w zależności od liczby klatek na sekundę. Jeśli więc, na przykład, chcesz przejść od punktu a do z w ciągu dwóch sekund, po prostu skalowałeś czas tak, aby wynosił od 0 do 1 w dwie sekundy. Po prostu podziel czas przez 2, aby to osiągnąć. Jeśli zastanawiasz się nad renderowaniem krzywej Béziera - na przykład, aby stworzyć wzgórza do gry na snowboardzie - to istnieje kilka opcji. Jeśli chcesz mieć ustaloną liczbę interwałów na krzywej (lub chcesz być leniwy), możesz po prostu przekazać dalej - rozróżnić równanie, aby wypróbować krzywą Béziera. Z drugiej strony, jeśli wolisz dynamicznie ustawiać poziom jakości, możesz skorzystać z samej definicji krzywej Béziera, aby wymyślić dobrze znany algorytm. Pozwól mi najpierw przedstawić ideę kwadratowej krzywej Béziera. Koncepcja może następnie zostać rozszerzona podobnie do tego, co zostało zrobione w samej definicji krzywej Béziera. Algorytm faktycznie pochodzi z definicji krzywej. Krzywa jest interpolacją dwóch innych interpolacji, więc algorytm dzieli problem na dwie części. Najpierw ocenia środkowy punkt krzywej, oceniając połowiczną interpolację na pierwszej i drugiej krzywej, a następnie oceniając połowiczną interpolację interpolacji między połowicznymi interpolacjami. Jest to nieco skomplikowane, ale w sumie to, co w zasadzie robi, to ocena krzywej w połowie. Po znalezieniu tego wierzchołka algorytm sprawdza, czy powinien on kontynuować. To heurystyka jest zwykle oparta na dwóch sprawdzeniach, z których pierwszym jest to, że wygenerowała ona minimalną liczbę wierzchołków (zwykle co najmniej większą niż całkowita liczba punktów kontrolnych), a druga to zapewnienie, że dwa nachylenia między trzema sąsiednimi wierzchołkami są z grubsza równy. (Oczywiście, nachylenie nie powinno być dokładnie równe, chyba że masz do czynienia z linią, ale musisz ustawić poziom tolerancji, co oznacza, że nie powinna być żadnej różnicy nachylenia większej niż x. To w zasadzie określa, jak gładko krzywa będzie wyglądać.) Do tej pory masz tylko trzy wierzchołki na krzywej - pierwszy wierzchołek, punkt środkowy i ostatni wierzchołek - co nie jest zbyt użyteczne. Aby zwiększyć liczbę wierzchołków, należy zastosować ten proces rekursywnie, jak pokazano na rysunku 13.5,



gdzie dwa zestawy czterech wierzchołków mogą być ponownie wykorzystane do wygenerowania dwóch nowych krzywych Béziera. Dzielenie krzywej jako takiej daje cztery segmenty liniowe. Dwa pierwsze segmenty mogą zostać użyte do określenia innego punktu środkowego, podczas gdy dwa inne segmenty linii mogą zostać użyte do określenia kolejnego wierzchołka punktu środkowego. W wyniku tego procesu powinieneś efektywnie mieć pięć wierzchołków próbkowanych na linii. Powtarzaj ten proces, aż osiągniesz maksimum lub do momentu, gdy różnica nachylenia między dwoma sąsiednimi wierzchołkami będzie zadowalająca i skończysz. Jak wcześniej wspomniano, jest to interesująca technika do renderowania głównie dlatego, że można dynamicznie ustawiać tolerancję nachylenia, co z kolei wpływa na gładkość krzywej. Jeśli zastosujesz technikę różnicowania do przodu, nadal możesz uzyskać dobrze wyglądającą krzywą, ale krzywa ta może nie wyglądać tak dobrze, gdy jest mocno przyspieszona. Dzieje się tak głównie dlatego, że tempo zmiany krzywej jest bardzo wysokie w tym regionie, podczas gdy region, który jest bardziej lub mniej liniowy, nie potrzebuje tylu próbek.

Przecięcie i otoczka wypukła

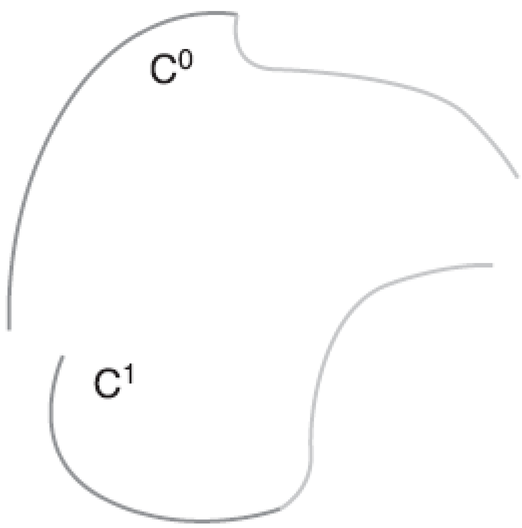
Jeśli użyjesz krzywych Béziera jako formy terenu, otrzymasz zdolność dynamicznego generowania gładkiego terenu i łatwego obliczania przecięcia dowolnego obiektu z terenem. Drugim typowym rozwiązaniem do generowania terenu jest mapa wysokości. W skrócie, jest to mapa parzyste-kwadratowa, która podaje wysokość mapy podaną <x, y>. Problem z mapami wysokości polega na tym, że nie dają one żadnych informacji między wartościami próbek <x, y> i są one ograniczone do dobrze określonego rozmiaru. Z drugiej strony, krzywe Béziera mogą generować duży fragment terenu z niewielką ilością danych i mogą dokładnie dać przecięcie czegoś, co na nim spadnie. Chociaż możesz już mieć dobry pomysł, jak to zrobić, musisz jeszcze zobaczyć, jak generować powierzchnie za pomocą Béziera, więc trzymajmy się teraz krzywych 2D. Na początek można obliczyć, czy obiekt znajduje się powyżej lub poniżej krzywej, stosując sztuczkę podaną w Części 4, która sugerowała generowanie homogenicznego równania i wstawianie wartości dla x i y w celu sprawdzenia, czy wynikowe równanie jest większe lub mniejsze od zera. To zadziała dobrze, jeśli potrzebujesz tylko testu przecięcia. Innym bardziej złożonym sposobem jest obliczenie przecięcia przez izolowanie dla parametru t oraz zbioru równań opisujących ruch obiektu, jak widać w części 10. Kolejny szybki test, można go zastosować w stosunku do wypukłego kadłuba, jak widać w części 12. Krzywe Béziera mają interesującą właściwość, że są ograniczone przez wypukły kadłub ich wielokąta kontrolnego, jak pokazano na rysunku 13.6.



Oznacza to, że w dowolnym momencie krzywa nigdy nie wyjdzie poza kadłub wypukły opisany przez uporządkowany zestaw punktów kontrolnych. To może być bardzo interesujące, jeśli myślisz o użyciu krzywej Béziera, na przykład do poruszania kamerą. Sugeruje to, że jeśli wybiorą wierzchołki świata tak, że wypukły kadłub nie przecina się z niczym innym na świecie, krzywa Béziera nigdy nie wejdzie "w" żaden obiekt. Jest tak, chyba że wolisz, żeby kamera gracza przechodziła przez ściany i zobaczyła, co leży po drugiej stronie. Kto wie? Oznacza to również, że jeśli wybierzesz wielokąt kontrolny taki, że jest wypukły, możesz wygenerować prosty test ograniczający pole. Z drugiej strony, jeśli grasz wieloma krzywymi Béziera w jednym konkretnym obszarze, możesz także wygenerować większe, wypukłe ciało wypukłe, na którym możesz testować kolizję. Tylko niebo ogranicza.

Kardynalne Krzywe Sklejane

To, co widzieliście do tej pory, nie jest krzywą sklejaną. Jeśli przyjrzymy się bliżej definicji krzywej sklejanej podanej wcześniej, zauważysz, że krzywa składana jest ciągły, fragmentaryczny. Krzywa Béziera jest ciągła w zakresie [0..1], ale nie jest ciągła w części (tzn. Przez połączenie dwóch Béziers). Jeśli dodasz dwie krzywe Béziera jedna po drugiej, nic nie gwarantuje, że będą one ciągłe. Minimalnie prawdopodobnie chcesz, aby twoja krzywa była klasą C^0. Jeśli nie, uzyskasz wrażenie teleportacji tu i tam na świecie. Możesz to łatwo osiągnąć za pomocą krzywej Béziera, po prostu zmuszając kolejną krzywą Béziera do rozpoczęcia od punktu końcowego pierwszej krzywej Béziera. Nie ma nic do tego. Z drugiej strony, jeśli chcesz, aby twoja krzywa była klasą C^1, to ty masz dużo pracy do zrobienia. Zasadniczo, musisz upewnić się, że pochodna w t = 1 dla pierwszej krzywej jest równa pochodnej w t = 0 dla drugiej krzywej. To nie brzmi jak poważny problem. Jeśli wyodrębnisz równania wychodzące z tego, będziesz wymyślił rozwiązanie, że drugi punkt drugiego Béziera musi być ustawiony na bardzo konkretną wartość. Mianowicie, dla sześciennej krzywej Béziera, ta wartość jest dwa razy ostatnim wierzchołkiem pierwszego Béziera, minus trzeci wierzchołek pierwszego Béziera. Pozostałe dwa punkty można dowolnie wybierać. Jest to uciążliwe, ponieważ musisz wyeliminować dwa stopnie swobody (tracisz pierwszy i drugi punkt każdej kolejnej krzywej Béziera), ale przynajmniej daje ci krzywą G^1. Różnicę przedstawiono na rysunku 13.7.



Na razie rozważę tylko krzywą sześcienną, ale wiem, że ten sam pomysł i proces stosuje się do każdej wyższej krzywej mocy. Innym podejściem jest wykonanie innej interpolacji przez wymuszenie zestawu ograniczeń, które chciałbyś zastosować w systemie. Jednym z ograniczeń, z których skorzystałaś na krzywej Béziera, była krzywa rozpoczęta i zakończona w pierwszym i ostatnim punkcie krzywej. odpowiednio. Innymi słowy, krzywa w jej najprostszej formie rozciągała się od wierzchołka P_k do P_{k+1}. Inną właściwością, którą można ustawić w celu wymuszenia ciągłości G^1, jest wymuszenie, aby pochodna pierwszego i ostatniego punktu miała własną stałą wartość. Dla punktu k wybrany wektor styczny to wektor generowany przez wektor $P_{k+1}P_{k-1}$. Innymi słowy, pochodna w P_k jest liniową kombinacją wektora generowanego przez P_{k+1} i P_{k-1}. Podobnie, dla t = 1, punkt końcowy, P_{k+1}, możesz zdefiniować pochodną jako tę generowaną przez wektor P_kP_{k+2}. W ujęciu matematycznym, ograniczenia ustawione naprzód dla krzywej parametrycznej C (t) są następujące:

$$C(0) = P_k$$
$$C(1) = P_{k+1}$$
$$C'(0) = s(P_{k+1} - P_{k-1})$$
$$C'(1) = s(P_{k+2} - P_k)$$

Od tego momentu zasadniczo zajmujesz się algebrą w celu znalezienia współczynnika matryca dla kardynalnej krzywej. Bardziej szczegółowo:

$$C(t) = \begin{bmatrix} t^3 & t^2 & t & 1 \end{bmatrix} \begin{bmatrix} a_x & a_y & a_z & a_w \\ b_x & b_y & b_z & b_w \\ c_x & c_y & c_z & c_w \\ d_x & d_y & d_z & d_w \end{bmatrix}$$

$$\begin{bmatrix} 0 & 0 & 0 & 1 \\ 1 & 1 & 1 & 1 \\ 0 & 0 & 1 & 0 \\ 3 & 2 & 1 & 0 \end{bmatrix} \begin{bmatrix} a_x & a_y & a_z & a_w \\ b_x & b_y & b_z & b_w \\ c_x & c_y & c_z & c_w \\ d_x & d_y & d_z & d_w \end{bmatrix} = \begin{bmatrix} P_k \\ P_{k+1} \\ s(P_{k+1} - P_{k-1}) \\ s(P_{k+2} - P_k) \end{bmatrix}$$

$$\begin{bmatrix} 0 & 0 & 0 & 1 \\ 1 & 1 & 1 & 1 \\ 0 & 0 & 1 & 0 \\ 3 & 2 & 1 & 0 \end{bmatrix}^{-1} \begin{bmatrix} 0 & 0 & 0 & 1 \\ 1 & 1 & 1 & 1 \\ 0 & 0 & 1 & 0 \\ 3 & 2 & 1 & 0 \end{bmatrix} \begin{bmatrix} a_x & a_y & a_z & a_w \\ b_x & b_y & b_z & b_w \\ c_x & c_y & c_z & c_w \\ d_x & d_y & d_z & d_w \end{bmatrix} = \begin{bmatrix} 0 & 0 & 0 & 1 \\ 1 & 1 & 1 & 1 \\ 0 & 0 & 1 & 0 \\ 3 & 2 & 1 & 0 \end{bmatrix}^{-1} \begin{bmatrix} P_k \\ P_{k+1} \\ s(P_{k+1} - P_{k-1}) \\ s(P_{k+2} - P_k) \end{bmatrix}$$

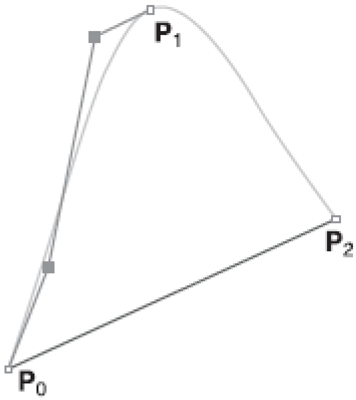
$$\begin{bmatrix} a_x & a_y & a_z & a_w \\ b_x & b_y & b_z & b_w \\ c_x & c_y & c_z & c_w \\ d_x & d_y & d_z & d_w \end{bmatrix} = \begin{bmatrix} 2 & -2 & 1 & 1 \\ -3 & 3 & -2 & -1 \\ 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & s & -s & 0 \\ -s & 0 & 0 & s \end{bmatrix} \begin{bmatrix} P_k \\ P_{k+1} \\ P_{k-1} \\ P_{k+2} \end{bmatrix}$$

$$= \begin{bmatrix} 2-s & s-2 & -s & s \\ s-3 & 3-2s & 2s & -s \\ 0 & s & -s & 0 \\ 1 & 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} P_k \\ P_{k+1} \\ P_{k-1} \\ P_{k+2} \end{bmatrix}$$

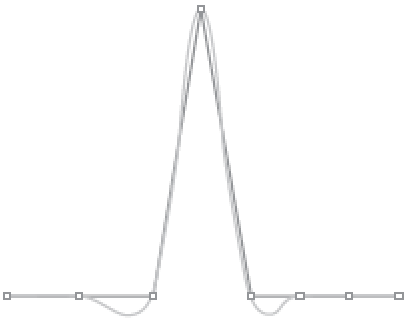
W konsekwencji równanie dla kardynalnej krzywej jest następujące:

$$C(t) = \begin{bmatrix} t^3 & t^2 & t & 1 \end{bmatrix} \begin{bmatrix} -s & 2-s & s-2 & s \\ 2s & s-3 & 3-2s & -s \\ -s & 0 & s & 0 \\ 0 & 1 & 0 & 0 \end{bmatrix} \begin{bmatrix} P_{k-1} \\ P_k \\ P_{k+1} \\ P_{k+2} \end{bmatrix}$$

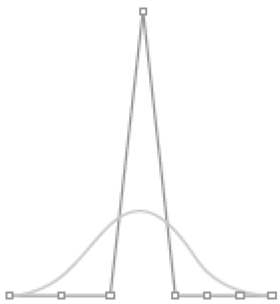
Być może zastanawiasz się, co oznacza. s jest niczym więcej niż nie znormalizowaną długością wektora stycznego. Im mniejszy s, tym bliżej krzywa będzie do punktów. Z drugiej strony, im większy s, tym bardziej krzywa odejdzie od punktów i tym bardziej zabawna będzie twoja ścieżka. Możesz wybrać dowolną wartość dla s, i znowu służy jako forma przyciągania magnetycznego w kierunku wielokąta kontrolnego. Heck, jeśli chcesz, możesz nawet wybrać inne wartości dla twoich dwóch ograniczeń. To jest piękno zrozumienia tutaj matematyki, a nie używania równań typu plug-and-play. Pojawia się interesująca koncepcja. Przypomnijmy, że z definicji renderowanie tej funkcji dla zakresu t = [0, 1] renderuje tylko od P_k do P_{k+1}. W rezultacie otrzymujemy szereg funkcji sześciennych, z których każda opisuje fragment całego sklejenia który z kolei jest opisany przez skończony zestaw punktów kontrolnych, jak pokazano na rysunku 13.8.



Co sprawia, że ta krzywa jest jeszcze lepsza niż krzywa Béziera? że jest to klasa G_1, i że tylko wierzchołki lokalne są dotknięte zmianą wierzchołków. Jeśli pamiętasz równanie Béziera, jego podstawowe funkcje wpływają na cały zakres krzywej. Jedynym sposobem na stworzenie krzywej, która została tylko lokalnie dotknięta, było zbudowanie krzywej Béziera. Patrząc na równanie tej konkretnej krzywej, widać, że biorąc pod uwagę dowolny punkt k, tylko cztery wierzchołki będą miały wpływ na segment krzywej. To jest naprawdę zgrabne, ponieważ oznacza to, że jeśli zmienisz wartość P_k, cała krzywa nie zacznie się zmieniać. Zamiast tego będzie to miało wpływ tylko na najwyżej trzy inne wierzchołki. Jest to naprawdę dobre, ponieważ oznacza, że możesz przesuwać wierzchołki bez zmiany ogólnego wyglądu krzywej. To równanie zostało wyprowadzone dla krzywej sześciennej. Można łatwo zastosować ten sam proces, generując nowy zestaw kolejnych lub mniej ograniczeń do naśladowania, ustanawiając liniowy system do rozwiązania, i, jak to opisano powyżej, rozwiązując macierz współczynników. Jeśli to zrobisz, otrzymasz większą lub mniejszą kwadratową matrycę, której rozmiar jest rzędem wielomianu minus jeden. Jeśli spojrzysz na to, co zostało zrobione wcześniej, ustawiono cztery ograniczenia. Oznacza to, że potrzebny był wielomian trzeciego stopnia, a w konsekwencji macierz 4 x 4. Różnica między normalną krzywą Béziera a krzywą kardynalną została przedstawiona na rysunku 13.9.



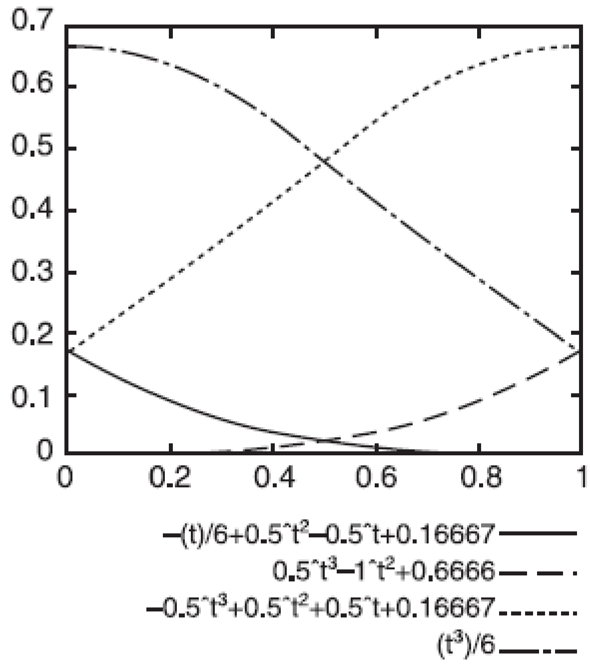
Krzywa Kardynalna



Krzywa Beziera

B-Sklejana

Do tej pory widziałeś krzywą Béziera, która jest doskonała, jeśli nie potrzebujesz krzywej klasy C_1. Z drugiej strony masz Krzywą Kręgową, która może być użyta dla krzywych klasy G_1, ale które wymagają trzy razy większej liczby obliczeń niż krzywej Béziera (pamiętaj, że średni wierzchołek jest używany czterokrotnie w macierzy). Nic tak naprawdę nie zmusza sklejonej do klasy G_2. W rzeczywistości, funkcja klasy C_2 może być interesująca, jeśli chcesz, aby przyspieszenie twojego ruchu pozostało takie samo. Właśnie o to chodzi w B-sklejonych. Zanim przyjrzysz się głębiej analizie B-sklejonych, zbadajmy rzeczy pod innym kątem: funkcje podstawowe. W przypadku krzywej Béziera funkcje podstawowe obejmują cały zakres krzywej. Zauważono, że efekt pojedynczego wierzchołka wpływa na całą krzywą. Niektóre funkcje podstawowe dla krzywej Béziera zostały przedstawione na rysunku 13.10. Z drugiej strony krzywa kardynalna ma funkcje podstawowe, które obejmują tylko niewielki zakres. Funkcje podstawowe dla B-sklejonych przedstawione na rysunku 13.11.



W poprzedniej części dowiedziałeś się, że sześcienna krzywa kardynalna jest taka, że każdy wierzchołek kontrolny oddziałuje na najwyżej cztery wierzchołki sterujące. Innymi słowy, dany wierzchołek kontrolny ma wpływ jedynie na odcinek krzywej otaczających wierzchołków kontrolnych. Nie może wpłynąć na całą krzywą, jak zrobiły to tylko krzywe Béziera. Jeśli masz zamiar wykreślić funkcje podstawowe, zauważysz, że jest to faktycznie reprezentowane na wykresach. Każda podstawowa funkcja obejmuje najwyżej cztery wierzchołki kontrolne. Można również zauważyć, że suma amplitudy funkcji w dowolnym punkcie czasu wynosi 1. Funkcja Béziera ma również tę właściwość, którą odnotowano algebraicznie. Aby uzyskać równanie dla B-sklejonych możesz myśleć o przestrzeni wektorowej, którą utworzysz w kategoriach funkcji podstawowych. Jeśli skupisz się na sześciennej wersji krzywej, możesz ustawić ścisły zestaw reguł, które muszą zostać spełnione. Krzywa podstawowa dla sklejonego sześciennego jest tworzona przez cztery wielomianów bazowych, jak pokazano na rysunku 13.11. (Dla każdej podstawy przyjmuje cztery punkty.) Dla tych czterech wielomianów powinno być oczywiste, że początek i koniec krzywej powinien wynosić 0. Ponadto ostatnia wartość pierwszego wielomianu powinna być równa pierwszej wartości drugi wielomian i tak dalej dla wszystkich wielomianów. Ponownie, to stwierdzenie po prostu wymusza ciągłość C^0. Możesz zastosować ten sam pomysł, aby wymusić ciągłość C^1 i C^2; to da ci 15 równań z 15 niewiadomymi. Ale ponieważ macie macierz bazową 4 x 4 do wypełnienia, brakuje wam jeszcze jednego warunku. Możesz dowolnie wybrać dowolny warunek, ale ten, który ostatecznie okaże się najbardziej przydatny, to fakt, że wszystkie wielomiany sumują się do 1 w t = 0, podobnie jak krzywe Béziera i Kardynała. W rzeczywistości nie zostanie to pokazane, ale jest to wymóg, który wymusza regułę wypukłego kadłuba. W sumie, zestaw równań, które masz, jest następujący:

Wrunki Pozycyjne	Warunki Nachylenia	Warunki Krzywizny
$0 = a(0)$	$0 = a'(0)$	$0 = a''(0)$
$a(1) = b(0)$	$a'(1) = b'(0)$	$a''(1) = b''(0)$
$b(1) = c(0)$	$b'(1) = c'(0)$	$b''(1) = c''(0)$
$c(1) = d(0)$	$c'(1) = d'(0)$	$c''(1) = d''(0)$
$d(1) = 0$	$d'(1) = 0$	$d''(1) = 0$

i

$$1 = a(0) + b(0) + c(0) + d(0)$$

Szesnaście równań z 16 niewiadomymi. W skrócie, jest to jedna podła i brzydka algebraiczna ekspresja. Oszczędzę ci żonglerki, przechodząc od razu do rozwiązania:

$$b(t) = \begin{bmatrix} t^3 & t^2 & t & 1 \end{bmatrix} \frac{1}{6} \begin{bmatrix} -1 & 3 & -3 & 1 \\ 3 & -6 & 3 & 0 \\ -3 & 0 & 3 & 0 \\ 1 & 4 & 1 & 0 \end{bmatrix} \begin{bmatrix} P_{i-2} \\ P_{i-1} \\ P_i \\ P_{i+1} \end{bmatrix}$$

Jedną z rzeczy, którą powinieneś zauważyć, po prostu podłączając wartości do równań, jest to, że nie zawsze jest możliwe, aby każdy wierzchołek Pi był częścią krzywej. To znaczy, że wymuszenie krzywizny na równi pomiędzy kawałkami sklejania ma swoją cenę. W konsekwencji B-sklejona nie można nazwać interpolacją tylko dlatego, że nie przechodzi od a do z, przechodząc przez zbiór punktów. W rzeczywistości może nawet nie przechodzić przez a i z! Może to być duża wada, jeśli planujesz mieć krzywą interpolującą, a nie tylko estetyczną. Wierzchołki nadal działają jak magnes w kierunku punktów kontrolnych, ale krzywa niekoniecznie kieruje się bezpośrednio do im. Zanim spojrzysz tutaj na swoje opcje, zobaczmy, jak możesz wykonać renderowanie takich krzywych.

Przekształcanie B-Sklejonych w odcinkowe Béziers'a

Renderowanie B-sklejonych może wydawać się monumentalnym zadaniem, ponieważ nie ma łatwego, oczywistego algorytmu. Oczywiście, możesz użyć algorytmu podziału podrzędnego, ale nie będziesz w stanie wykorzystać efektywnych przyspieszeń dzielenia, z których korzysta krzywa Béziera. Ale co by było, gdybyś mógł wyrazić B-sklejone jako krzywą Béziera? Zarówno B-sklejana, jak i krzywa Béziera są liniowymi przekształceniami, więc dlaczego nie? Następuje następująca logika:

$$b(t)=\begin{bmatrix} t^3 & t^2 & t & 1 \end{bmatrix}\frac{1}{6}\begin{bmatrix} -1 & 3 & -3 & 1 \\ 3 & -6 & 3 & 0 \\ -3 & 0 & 3 & 0 \\ 1 & 4 & 1 & 0 \end{bmatrix}\begin{bmatrix} P_{i-2} \\ P_{i-1} \\ P_i \\ P_{i+1} \end{bmatrix}$$

$$=\begin{bmatrix} t^3 & t^2 & t & 1 \end{bmatrix}\left(\begin{bmatrix} -1 & 3 & -3 & 1 \\ 3 & -6 & 3 & 0 \\ -3 & 3 & 0 & 0 \\ 1 & 0 & 0 & 0 \end{bmatrix}\begin{bmatrix} -1 & 3 & -3 & 1 \\ 3 & -6 & 3 & 0 \\ -3 & 3 & 0 & 0 \\ 1 & 0 & 0 & 0 \end{bmatrix}^{-1}\right)\frac{1}{6}\begin{bmatrix} -1 & 3 & -3 & 1 \\ 3 & -6 & 3 & 0 \\ -3 & 0 & 3 & 0 \\ 1 & 4 & 1 & 0 \end{bmatrix}\begin{bmatrix} P_{i-2} \\ P_{i-1} \\ P_i \\ P_{i+1} \end{bmatrix}$$

$$=\begin{bmatrix} t^3 & t^2 & t & 1 \end{bmatrix}\begin{bmatrix} -1 & 3 & -3 & 1 \\ 3 & -6 & 3 & 0 \\ -3 & 3 & 0 & 0 \\ 1 & 0 & 0 & 0 \end{bmatrix}\left(\frac{1}{6}\begin{bmatrix} 0 & 0 & 0 & 1 \\ 0 & 0 & \frac{1}{3} & 1 \\ 0 & \frac{1}{3} & \frac{2}{3} & 1 \\ 1 & 1 & 1 & 1 \end{bmatrix}\begin{bmatrix} -1 & 3 & -3 & 1 \\ 3 & -6 & 3 & 0 \\ -3 & 0 & 3 & 0 \\ 1 & 4 & 1 & 0 \end{bmatrix}\begin{bmatrix} P_{i-2} \\ P_{i-1} \\ P_i \\ P_{i+1} \end{bmatrix}\right)$$

$$=\begin{bmatrix} t^3 & t^2 & t & 1 \end{bmatrix}\begin{bmatrix} -1 & 3 & -3 & 1 \\ 3 & -6 & 3 & 0 \\ -3 & 3 & 0 & 0 \\ 1 & 0 & 0 & 0 \end{bmatrix}\left(\frac{1}{6}\begin{bmatrix} 1 & 4 & 1 & 0 \\ 0 & 4 & 2 & 0 \\ 0 & 2 & 4 & 0 \\ 0 & 1 & 4 & 1 \end{bmatrix}\begin{bmatrix} P_{i-2} \\ P_{i-1} \\ P_i \\ P_{i+1} \end{bmatrix}\right)$$

Czy to nie było takie proste? Wszystko, co musisz zrobić, aby przekonwertować punkty B-sklejone na punkty Béziera, to pomnożyć zbiór punktów przez macierz transformacji liniowej. W konsekwencji oznacza to, że każde B-sklejone jest w rzeczywistości połączeniem zestawu krzywych Béziera. Tak więc forma B-sklejona jest w rzeczywistości niczym innym jak ograniczeniem stosowanym do zbioru punktów Béziera. To samo można powiedzieć o kardynalnej sklejonej. Możesz konwertować punkty B-sklejone na punkty Béziera za pomocą następującej formuły:

$$\begin{bmatrix} BezierPt_0 \\ BezierPt_1 \\ BezierPt_2 \\ BezierPt_3 \end{bmatrix}=\frac{1}{6}\begin{bmatrix} 1 & 4 & 1 & 0 \\ 0 & 4 & 2 & 0 \\ 0 & 2 & 4 & 0 \\ 0 & 1 & 4 & 1 \end{bmatrix}\begin{bmatrix} P_{i-2} \\ P_{i-1} \\ P_i \\ P_{i+1} \end{bmatrix}$$

B-sklejone, które dotychczas widzieliście, pasują do grupy zwanej jednolitymi B-sklejonymi. Następna sekcja ma rozszerzyć tę "jednolitą" część.

Świat poza: rozszerzanie pojęcia miejsca za pomocą wektorów węzłów

Widzieliście, że można utworzyć funkcję, która ma lokalną kontrolę nad wierzchołkami. W szczególności funkcja sześcienna ma kontrolę nad czterema wierzchołkami sąsiada. Możesz rozszerzyć tę koncepcję, tak abyś miał kontrolę nad pięcioma, sześcioma lub dowolną liczbą wierzchołków. Aby to zrobić, po prostu wypisz zestaw wiązań, jak pokazano powyżej w sekcji B-sklejone i wyprowadź odpowiednią macierz. To naprawdę kłopotliwe zadanie. Zobaczysz inną metodę wyprowadzania podstawowych funkcji, ale zanim to zrobisz, chciałabym zrobić dźgnięcie inną koncepcją: wektory węzłów. Wektor węzłowy jest wektorem, który daje zakres wpływu i można go wykorzystać do stworzenia krzywej B-sklejonej, która niekoniecznie jest jednolita. Dla danego punktu i, możesz określić jego zakres za pomocą wektora węzłów, patrząc na jego dolną granicę vi i jej górną granicę v_{i+k+1}. Z wypustkami, które widziałeś do tej pory, tak naprawdę nie spojrzałeś na wpływ zakres w procentach [0..1], ale jako zasięg pod względem otaczających wierzchołków. Jest to odpowiednik równego dystansu do każdego wierzchołka, dlatego te typy sklejonych nazywa się "jednolitymi" B-sklejonymi. Oznacza to, że są one nazywane "jednolite", ponieważ mają swój wektor węzłowy , który wzrasta liniowo. Oczywiście, aby wektor węzeł mógł mieć jakikolwiek sens, musi się zwiększać, inaczej uzyskasz wyraźne niespójności. Ponadto, ponieważ twoja krzywa jest zdefiniowana w zakresie [0, 1], wektor węzłowy musi być również ograniczony przez [0, 1]. (Jeśli chodzi o bardziej ogólną definicję B - sklejonego, nie jest to wymagane, ponieważ równanie skaluje wszystko, jednak ze względu na prostotę, trzymajmy się bardziej praktycznego podejścia, to jest [0, 1].) Na przykład , jeśli chciałbyś przedstawić sześcienny B-sklejony, który widziałeś w przykładzie konwersji B-sklejonej do Béziera jako wektor węzłowy, dla sześciu wierzchołków napisałbyś wektor w następujący sposób:

$$\left[\frac{0}{9},\frac{1}{9},\frac{2}{9},\frac{3}{9},\frac{4}{9},\frac{5}{9},\frac{6}{9},\frac{7}{9},\frac{8}{9},\frac{9}{9}\right]$$

Zakres pierwszego wierzchołka jest zatem [0, 4/9], drugi wierzchołek ma zakres [1/9, 5/9] i tak dalej, aż do szóstego wierzchołka o zakresie [5/9, 1]. Jeśli wykonasz obliczenia matematyczne, zauważysz, że wektor węzłów ma zawsze tyle wartości, ile liczba wierzchołków plus stopień. Więc dla poprzedniego przykładu, miałeś 4 + 6 - 1 = 9 różnych wartości. Dzielenie dziesięciu wartości równo oznacza, że musisz zwiększyć o 1/9 każdy krok, zaczynając od 0. Ale tutaj jest podchwytliwe pytanie: Jak przedstawiasz krzywe Béziera używając tej notacji? Na szczęście nie jest to takie trudne. Przypomnijmy, że krzywe Béziera wpłynęły na całą krzywą z zakresem [0, 1]. Oznacza to, że biorąc pod uwagę stopień k, wybrany zakres powinien zawsze mieć zakres [0, 1]. Jedynym sposobem, aby to zrobić, jest ustawienie połowy wartości na 0, a drugiej na wartość 1. W ten sposób sześcienny Bézier ma wektor węzłowy

$$[0,0,0,0,1,1,1,1]$$

Te formy wektorów węzłowych nazywa się wektorami otwartymi, ponieważ powtarzają one wartość zakresu na początku i na końcu. Otwarte wektory są nadal uważane za jednolite wtedy i tylko wtedy, gdy usunięcie duplikatów na początku i na końcu daje jednorodny wektor węzłowy. Istnieją również wektory węzłowe, które nie są jednolite, ale nadal są częścią rodziny B-sklejonych. Koncepcja jest dokładnie taka sama, jak wersja jednolita, z tym wyjątkiem, że zasięg oddziaływania nie jest równy od jednego interwału do drugiego. Pomijam złożoną algebrę konwersji do relacji nawrotu i od niej i przedstawię równanie rekurencyjnie obliczając podstawowe funkcje B-sklejone, biorąc pod uwagę jego stopień k i liczbę punktów kontrolnych n. I-ta funkcja podstawowa wielomianu k stopnia B-sklejona $B_{i,k}$ z elementami wektora węzła x jest następująca:

$$B_{i,1} = \begin{Bmatrix} 1 & x_i \le t < x_{i+1} \\ 0 & else \end{Bmatrix}$$

$$B_{i,k}(t) = \frac{(t - x_i)\, B_{i,k-1}(t)}{x_{i+k-1} - x_i} + \frac{(x_{i+k} - t)\, B_{i+1,k-1}(t)}{x_{i+k} - x_{i+1}}$$

W tym momencie prawdopodobnie należałoby przyjrzeć się efektowi wektora węzłów na całej krzywej. Na początek, za pomocą krzywej Béziera, można ustalić uogólnienie, które można wymusić na sklejonej stopnia k, aby przejść przez początkowy i ostatni punkt kontrolny, jeśli wszystkie k pierwsze i ostatnie punkty kontrolne k są takie, że ich zasięg wpływa na pierwsze / ostatni punkt kontrolny. Mówiąc prościej, sklejoa przechodzi przez swój pierwszy i ostatni punkt kontrolny, jeśli jego wektor węzłowy zaczyna się od zer i kończy się na k. Oczywiście, jak wspomniano wcześniej, możesz mieć zakres, który nie jest [0, 1], a definicja odpowiednio się zmieni. Ale ponieważ zależy ci tylko na zakresie [0, 1], nie powinieneś zawracać sobie głowy innymi przypadkami, takimi jak czyste liczby całkowite, które czasami są używane w literaturze, aby "uprościć" rzeczy.

Zamknięcie luki

B-sklejone są świetne, ponieważ zapewniają ciągłość C^2, ale nie gwarantują, że którykolwiek z punktów kontrolnych będzie częścią końcowej krzywej. (Oczywiście, ponieważ możesz wyrazić krzywą Béziera jako B-sklejoną, istnieje co najmniej jedna krzywa, która pozwala na to i prawdopodobnie możesz wymyślić również kilka innych przykładów.) W rezultacie może być trudne zamknij kształt. Na przykład, jeśli chciałbyś wyrazić równanie okręgu lub coś podobnego za pomocą B-sklejonych, prawdopodobnie nie uznałbyś tego na początku za oczywisty problem. Rozwiązanie jest jednak zaskakująco proste. Wszystko, co musisz zrobić, aby zamknąć kształt, to dodać pierwsze k - 2 wierzchołki. To takie proste.

NURB

NURB, który jest ostatnim typem sklejonej, który zamierzam pokazać, oferuje największą kontrolę, ale jest także najcięższą sklejoną do obliczenia. Ideą krzywej NURB (niejednolitej wymiernej krzywej B-sklejonej) jest zdefiniowanie współczynników ciężaru dla każdego wierzchołka. Innymi słowy, dla punktu kontrolnego P_i, sklejony NURB definiuje go jako nowy punkt kontrolny Q_i, pokazany:

$$P_i = \begin{bmatrix} x_i \\ y_i \\ z_i \end{bmatrix} \approx w_i \begin{bmatrix} x_i \\ y_i \\ z_i \\ 1 \end{bmatrix} = Q_i$$

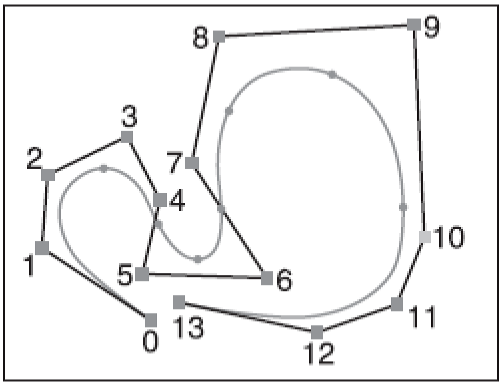
Niestety, nie możesz po prostu pomnożyć wierzchołka w ten sposób, nie martwiąc się o nic innego. Twoja krzywa jest zdefiniowana dla zakresu [0..1], a zastosowanie losowych multiplikacji po prostu nie zadziała. To, co musisz zrobić, aby zachować tę samą integralność i właściwości określone w poprzednich sekcjach, polega na renormalizacji krzywej. Oczywiście, gdybyś wybrał Q_i tak, że były już znormalizowane, nie byłoby to problemem. Ale jak wektor, ogólny przypadek zakłada nie znormalizowane wagi; w związku z tym musisz podzielić przez podział perspektywy, aby uzyskać końcowe równanie. Połóż wszystko razem, a otrzymasz następujące informacje:

$$f(t) = \frac{\sum_{i=1}^{n} P_i \cdot B_{i,k}(t)\, w_i}{\sum_{i=1}^{n} B_{i,k}(t)\, w_i}$$

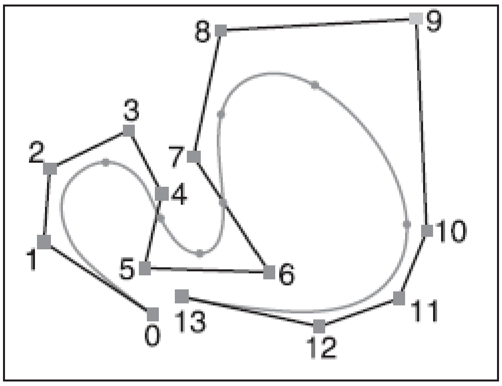
Ale czekajcie - czy nie mówiłem wcześniej, że każdą krzywą / splajn można zapisać w postaci sumowanego dodania funkcji bazowej pomnożonej przez punkt kontrolny? Cóż, niezupełnie. Jeśli pamiętasz, dodałem nawias "z jednym zauważonym wyjątkiem." Jeśli, na przykład, miałbyś wybrać otwarty jednolity B-splajn w trzecim stopniu, możesz zapisać go w reprezentacji macierzy, jak pokazano tutaj:

$$b_i(t) = \frac{\begin{bmatrix} t^3 & t^2 & t & 1 \end{bmatrix} \frac{1}{6} \begin{bmatrix} -1 & 3 & -3 & 1 \\ 3 & -6 & 3 & 0 \\ -3 & 0 & 3 & 0 \\ 1 & 4 & 1 & 0 \end{bmatrix} \begin{bmatrix} w_{i-2} & 0 & 0 & 0 \\ 0 & w_{i+1} & 0 & 0 \\ 0 & 0 & w_i & 0 \\ 0 & 0 & 0 & w_{i+1} \end{bmatrix} \begin{bmatrix} P_{i-2} \\ P_{i-1} \\ P_i \\ P_{i+1} \end{bmatrix}}{\begin{bmatrix} t^3 & t^2 & t & 1 \end{bmatrix} \frac{1}{6} \begin{bmatrix} -1 & 3 & -3 & 1 \\ 3 & -6 & 3 & 0 \\ -3 & 0 & 3 & 0 \\ 1 & 4 & 1 & 0 \end{bmatrix} \begin{bmatrix} w_{i-2} \\ w_{i+1} \\ w_i \\ w_{i+1} \end{bmatrix}}$$

W tej formie możesz prawdopodobnie zrozumieć nieco więcej, dlaczego NURB są o wiele bardziej kosztowne obliczeniowo niż B-sklejone. Mnożenie dwóch górnych matryc można by zmienić w produkt punktowy, ale nie byłaby to purystyczna reprezentacja macierzowa. Oczywiście elastyczność prawie zawsze ma swoją cenę, a taka jest cena dla NURB. Pytanie, które możesz słusznie zadać, brzmi: co możesz zrobić z NURB, co nie było możliwe przy poprzednich krzywych? Zaletą NURB jest to, że mogą one dokładnie przedstawiać sekcje stożkowe, a przez to okręgi, elipsy, hiperbole i tym podobne. Znacznie łatwiej jest myśleć o krzywej pod względem współczynników ważenia niż myśleć o zwiększaniu porządku krzywej lub zmianie dotknięty region za pomocą wektora węzłów. Piękno tej funkcji polega na tym, że może łatwo można przekształcić w równanie B-sklejone, jeśli weźmiesz pod uwagę znormalizowane ważone punkty jako dane wejściowe dla B-sklejone. W ten sposób można również konwertować z B-sklejonej na Béziera, a następnie renderować krzywą. W jaki sposób różne wartości ciężaru wpływają na ogólny wygląd krzywej? Na początek twoja krzywa jest nadal ograniczona przez zakres ustawiony w wektorze węzła. Jeśli spojrzysz na to algebraicznie, zauważysz, że ustawienie ciężaru na 0 to to samo, co po prostu zapomnienie o wierzchołku w ogóle. W związku z tym ustawienie jednego z wag na 0 będzie miało taki sam efekt, jak nigdy nie będzie określać punktu kontrolnego. Z drugiej strony, jeśli ustawisz tę samą wagę dla każdego innego wierzchołka, otrzymasz B-sklejoną. Ponieważ każda waga ma tę samą wartość, znormalizowany ciężar wynosi 1 i, jak widzieliśmy w sekcji B-sklejone, suma funkcji bazowych dla dowolnej wartości t wynosi 1 (określona przez definicję). Od tego momentu możesz zobaczyć, jak będą zachowywać się wartości od 1 do 0. Wartość bliższa 0 będzie mało ważna dla punktu kontrolnego, podczas gdy wartość zmierzająca w kierunku nieskończoności zasadniczo ignoruje każdy inny wierzchołek kontrolny. Jeśli wolisz patrzeć na to w inny sposób, wartość nieskończoności zasadniczo pozwala znormalizowanej wadze innych sterujących wierzchołków zmierzać w kierunku 0. Rysunki 13.12, 13.13 i 13.14 ilustrują wpływ wag na krzywą.



13.12



13.13

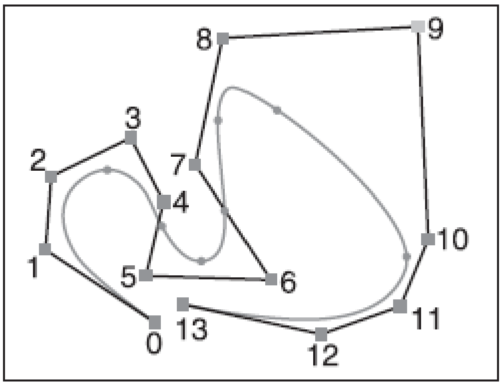


Figure 13.14

Używanie NURB do reprezentowania przekrojów stożkowych

To, co sprawia, że NURB jest szczególnie interesujące, to fakt, że dzięki nim można połączyć zalety kardynalnych i B-sklejonych. W rzeczywistości, jeśli się nad tym zastanowisz, prawdopodobnie zauważysz, że NURB są w rzeczywistości bardziej restrykcyjną wersją B-sklejonych, które z kolei są bardziej restrykcyjną wersją zbioru krzywych Béziera. Problem z B-sklejonymi i sklejonymi kardynalnymi polega jednak na tym, że nie mogą one łatwo reprezentować przekrojów stożkowych (elips, hiperboli itd.). Jednak z NURB można zdefiniować krzywą przechodzącą przez jej pierwszy i ostatni punkt za pomocą otwarty węzeł, ale nadal możesz użyć współczynników masy, aby uzyskać lepszą kontrolę nad krzywą. Oznacza to, że w przypadku NURB można uzyskać krzywą przechodzącą przez dany zestaw punktów i nadal mieć dobry poziom kontroli nad krzywizną krzywej, co jest szczególnie skuteczne w przypadku przekrojów stożkowych. Jeśli weźmiesz krzywą z tylko trzema wierzchołkami, wybranymi jako trzy punkty na trójkącie równobocznym (każda strona ma równą długość), możesz wymusić krzywą, aby przejść przez pierwszy i ostatni wierzchołek, i nadal możesz zmienić krzywiznę krzywa, po prostu zmieniając wagę środkowego wierzchołka, ustawiając wagę 1 na końcach. Przy takiej krzywej ustawienie średniej wartości 0 spowoduje narysowanie prostej od a do b. Z drugiej strony, jeśli ustawisz wagę 1, wygenerujesz krzywe paraboliczne. Ustawienie wartości masy powyżej 1 da krzywą hiperboliczną, natomiast ustawienie wartości poniżej 1 da krzywą eliptyczną. Możesz to sprawdzić, podłączając niektóre wartości wewnątrz krzywej i obliczając pochodną. Krzywa hiperboliczna dąży do nachylenia [+ /?] 1, nachylenie eliptyczne jest mniejsze niż 1 w wartości bezwzględnej, a parabola to pozostałe wartości. Jeśli chcesz zrobić kółko, powinieneś oczekiwać wagi mniejszej niż 1, ponieważ krąg jest łukiem eliptycznym. Jeśli wymusisz, że punkt środkowy krzywej znajduje się na okręgu o promieniu r, możesz określić wagę, która byłaby potrzebna, aby to zrobić. Oznaczałoby to, że okrąg powinien się składać z trzech oddzielnych krzywych o wadze [1 w 1 1 w 1 1 w 1] i wektora węzła [0 0 0 1 1 2 2 3 3 3]. Nie ma sensu przechodzić przez szczegóły tego dowodu; po prostu wiedz, że równanie, które otrzymujesz, jest następujące:

$$w = \cos\left(\frac{\pi}{6}\right) = \frac{1}{2}$$

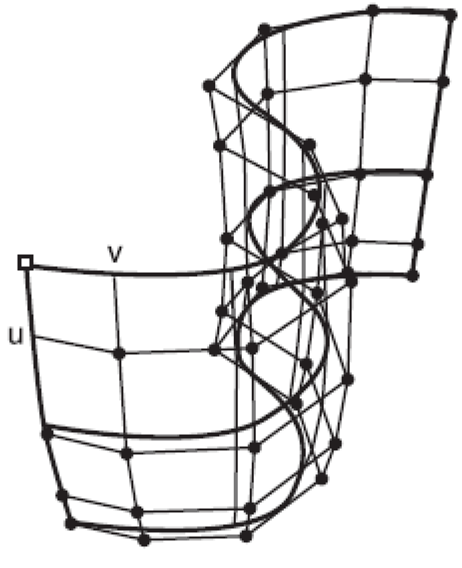
Oczywiście działa to, ale nie jest to bardzo wygodna forma dla twojego kręgu, głównie dlatego, że nie możesz go łatwo rozciągnąć, aby uzyskać elipsę. Z drugiej strony, jeśli masz kwadrat, możesz po prostu przetłumaczyć punkty kontrolne, aby uzyskać elipsę. To staje się teraz znacznie bardziej atrakcyjne. Oczywiście można również zastosować dowolną transformację, którą lubisz w tych punktach kontrolnych, aby obracać, tłumaczyć, skalować itd. Kiedy to robisz, po prostu upewnij się, że niczego nie założyłeś podczas obliczania współczynników wagowych; w przeciwnym razie wszystko się rozpadnie. Na przykład, nie można przeskalować w kręgu opisanym za pomocą siedmiu punktów kontrolnych, głównie dlatego, że złamałoby to ustawioną regułę w odniesieniu do trójkątów będących równobocznymi. A zatem, jak można obliczyć współczynniki ciężaru dla NURB o liczbie punktów k = 3? Dlaczego rzeczy zawsze muszą wyglądać bardziej skomplikowanie, niż są w rzeczywistości? Możesz aplikować dokładnie taką samą logikę, aby określić wagę, której potrzebujesz. Ogólnie rzecz biorąc, można to zrobić ogólnie dla n wierzchołków na okręgu. Rozwiązaniem tego konkretnego problemu jest w = -1. Ogólny przypadek będzie pozostawiony jako ćwiczenie. Możesz napisać kod, aby wyrenderować taką krzywą, jeśli chcesz sprawdzić, czy rzeczywiście zrobiłeś to poprawnie. Jak zawsze najlepszym sposobem rozwiązania tego problemu jest zastanowienie się nad tym, co wiesz o problemie. Jakie dane wejściowe możesz wykorzystać do rozwiązania problemu? Po wprowadzeniu danych wejściowych możesz przyjrzeć się związkom problemu, aby znaleźć rozwiązanie. Ostatnim krokiem jest przetestowanie go i wypróbowanie różnych danych wejściowych, aby zobaczyć, jak się zachowuje. Czasami możesz wymyślić coś, co jest bardziej zbliżone niż dokładne rozwiązanie, ale niekoniecznie jest problematyczne - szczególnie w przypadku gier, gdzie dokładność można z łatwością odłożyć na stronę wizualną, aby uzyskać większą liczbę klatek na sekundę.

Mądrze wykorzystując krzywe

Jedną rzeczą jest poznanie matematyki; to kolejny, aby wiedzieć, gdzie należy go używać. Te krzywe mogą znaleźć drogę do różnych rzeczy. Używanie ich zależy wyłącznie od Ciebie. Popularnym wyborem jest użycie tych krzywych do poruszania kamerą. Nikt nie lubi gwałtownego ruchu kamery; to po prostu wygląda nieprofesjonalnie. Na szczęście tego typu ruchy tak naprawdę nie wymagają tak dużej precyzji, więc aby je wyrenderować, nie potrzebujesz fantazyjnej krzywej. Tak długo, jak możesz wybrać punkty kontrolne tak, że nie uderzysz w ścianę, powinieneś być w porządku. Jak wspomniałem wcześniej, można użyć właściwości wypukłego kadłuba sklejonej, aby to zagwarantować. Oczywiście można zrobić o wiele więcej dzięki tym typom krzywych niż proste tłumaczenia kamerą. W końcu, w przeważającej części, krzywe te interpolują krzywe. Różnią się one od aproksymacja krzywych, ponieważ przechodzą one od jednej wartości do drugiej w dobrze określony sposób. Oczywiście nie oznacza to, że nie możesz używać obu jednocześnie. Na przykład możesz użyć krzywych Béziera do renderowania okręgu, jeśli chcesz. Po prostu zastosuj przybliżenie do równania parametrycznego koła, a będziesz na dobrej drodze. Znowu odnosi się to nie tylko do ruchu. Ponieważ te krzywe interpolują między różnymi wartościami, możesz użyć interpolacji dla jeszcze większej ilości rzeczy. Na przykład możesz użyć takiej krzywej, aby wygładzić obrót kamery - zdecydowanie "miło mieć". Jeśli chcesz, aby kamera była skierowana w kierunku, w którym zmierzasz, powinieneś być w stanie to zrobić obliczając wektor styczny sklejonej. Jest to jeden z najłatwiejszych problemów pochodnych, więc obliczanie go powinno być dla ciebie bardzo proste. Teren jest również bardzo atrakcyjną opcją dla takich łuków, ponieważ gwarantuje ciągłość i wymaga dużo mniej pamięci niż konwencjonalne mapy wysokości.

Powierzchnie

Bez wątpienia krzywe / sklejone są całkiem użyteczne - i można je jeszcze bardziej zwiększyć, przesuwając je do wyższych wymiarów, dodając jeden dodatkowy parametr. Dodanie dodatkowego parametru umożliwia zdefiniowanie powierzchni, która może być później używana w twoim świecie. Wspomina się o wspomnianej wcześniej grze snowboardowej. Piękno sklejonych w polega na tym, że możesz użyć czegoś takiego jak ekstremalnie długa B-sklejona, aby w zasadzie opisać cały świat. W przeciwieństwie do używania mapy wysokości, możesz używać powierzchni do renderowania tylko niewielkich części świata gry, które są potencjalnie widoczne w zależności od kamery, definiując w ten sposób bardzo długi świat z niewielkimi danymi, z dodatkowymi zaletami ciągłości i dokładności kolizji. . Powierzchnie mogą być również używane jako obiekty. Quake III przychodzi tu na myśl; wykorzystuje powierzchnie do generowania różnych obiektów świata. Tego rodzaju obiekty są w rzeczywistości dość łatwe do odkształcenia, ponieważ poruszanie punktem kontrolnym wpłynie na mały określony obszar zgodnie z twoimi wektorami węzłowymi. Jeśli chcesz mieć dodatkowe wyobrażenia, możesz nawet użyć sklejonej, aby przesunąć punkty kontrolne swojej powierzchni, aby wygenerować ciekawy efekt morfingu. Nie ma sensu patrzeć niezależnie na różne możliwe powierzchnie, głównie dlatego, że wszystkie są przedłużeniami ich pochodnej krzywej / sklejonej, a zatem mają te same właściwości i zasady. Rozważmy na przykład ciąg wygięty w różnych kierunkach, które zostały zdefiniowane przez punkty kontrolne. Teraz przełącz się na parametry drugiego wymiaru, rozszerzając ciąg na arkusz papieru. Podobnie jak w przypadku napisu, możesz zgiąć arkusz papieru zgodnie z zestawem punktów kontrolnych, jak pokazano na rysunku 13.15.



W rzeczywistości kartka papieru może być zdefiniowana za pomocą dwóch ciągów. Jeśli pozwolisz, aby pierwszy ciąg wędrował swobodnie wzdłuż osi X, zablokujesz jeden z parametrów. Stamtąd możesz użyć drugiego ciągu, aby zdefiniować inną krzywą prostopadłą do pierwszej, która może być postrzegana jako oś y lub z, w zależności od twojego punktu widzenia. Teraz masz dwa ciągi, których parametry są zdefiniowane ortogonalnie. Uwaga: nie powiedziałem, że krzywe były ortogonalne. Zamiast tego powiedziałem, że parametry łańcuchów są ortogonalne. Jeśli wolisz, możesz powiedzieć, że t, pierwszy parametr, nie jest w żaden sposób zależny od s, który możesz zdefiniować jako drugi parametr i na odwrót. Możesz także zobaczyć to mniej więcej jako podstawę, w której parametr t przesuwa się w jednym kierunku, podczas gdy s umożliwia swobodne poruszanie się w innym kierunku w 3D. Ideą powierzchni jest połączenie funkcji podstawowej $B_i(s)$ i $B_j(t)$. W związku z tym końcową funkcją podstawową dla parametru i, j jest

$$B_{i,j}(s,t) = B_i(s)B_j(t)$$

Jeśli przyjrzysz się dokładnie tej formule, zauważysz, że ma ona te same właściwości, co jednoparametrowa funkcja podstawowa. Jak można się spodziewać, ponieważ obie jednoparametryczne funkcje bazowe mają zakres [0, 1], mnożenie dwóch takich zakresów również da zakres [0, 1], co oznacza, że krzywa ponownie spełni wypukła właściwość kadłuba. Jeśli chodzi o równania, zachowują się one w dokładnie taki sam sposób, jak były używane, z wyjątkiem tego, że musisz obliczyć oba parametry. Tak więc było to równanie NURB

$$f(t) = \frac{\sum_{i=1}^{n} P_i \cdot B_{i,k}(t) w_i}{\sum_{i=1}^{n} B_{i,k}(t) w_i}$$

teraz staje się

$$f(s,t) = \frac{\sum_{j=1}^{m}\sum_{i=1}^{n} P_{i,j} \cdot B_{j,l}(s) \cdot B_{i,k}(t) w_{i,j}}{\sum_{j=1}^{m}\sum_{i=1}^{n} B_{j,l}(s) \cdot B_{i,k}(t) w_{i,j}}$$

Po pierwsze, zauważ, że funkcje podstawowe mogą faktycznie mieć różne moce wielomianowe. Jeden ma stopień k, podczas gdy drugi ma stopień l. Oznacza to, że jeśli chcesz mieć więcej kontroli lokalnej w jednej osi, możesz to łatwo zrobić. Z drugiej strony, jeśli wolisz pozostać symetryczny, zawsze możesz pozwolić k = l. Następnie zwróć uwagę, że masa i punkty kontrolne zostały pomnożone. Każdy punkt jest związany z wagą, więc to nic nowego, ale same punkty kontrolne przeskoczyły dodatkowy wymiar. Aby zrozumieć, dlaczego tak jest, po prostu pomyśl o różnicy między płaszczyzną a linią. Płaszczyznę można właściwie opisać jako zbiór linii wzdłuż linii. Podobnie tutaj powierzchnię można opisać za pomocą zestawu punktów kontrolnych lub, jeśli wolisz, zestawu punktów kontrolnych krzywej. Jeśli nadal masz problemy z wizualizacją powierzchni, pomyśl o czterech krzywych Béziera w 2D oddzielonych od siebie ustaloną odległością w 3D. Innymi słowy, masz cztery krzywe Béziera w 3D, które są równoległe do jednej z osi w 3D, ale poruszają się swobodnie w pozostałych dwóch osiach, które są oddzielone od siebie o stałą wartość. Teraz weź pierwsze punkty kontrolne każdej z krzywych Béziera i wygeneruj z nich kolejną krzywą Béziera. Zrób to dla wszystkich zestawów czterech punktów kontrolnych, a zaczniesz widzieć powierzchnię; właściwie to się tutaj dzieje. Więc jakie jest równanie dla powierzchni opartej na B-sklejonej, krzywej kardynalnej lub Béziera? Cóż, używa tego samego procesu, jak wyjaśniono wcześniej. Po prostu zmień podstawową funkcję i ustalone wartości tak, aby zajmowały się tymi dwoma parametrami i jesteś ustawiony. Jeśli masz już kształt w 2D, możesz łatwo rozszerzyć go w 3D, dodając więcej szczegółów na temat jego nowego wymiaru. Na przykład już wiesz, jak utworzyć koło w 2D. Jeśli chcesz utworzyć rurę w 3D, wystarczy określić punkty kontrolne na osi Z. Jeśli chcesz, aby rura miała równomierny promień, wystarczy skopiować punkty kontrolne, ale zmienić liniowo wartość z. Alternatywnie, możesz zacząć przekształcać ten kształt, zwiększając rozmiar koła przy różnych wartościach z, aby uzyskać coś takiego, powiedzmy, garnek. Garnek wykonany jest z małego koła u dołu i innego nieco większego okręgu u góry. Tymczasem ma on dużą okrągłą talię, co oznacza, że punkty kontrolne powinny być odpowiednio ułożone.

Rendering Powierzchnie

Jeśli chcesz obliczyć normalny wektor powierzchni do celów oświetleniowych i usuwania powierzchni ukrytych, powinieneś być w stanie to zrobić, obliczając wektor styczny funkcji in s i in oraz obliczając iloczyn iloczynu obu. Jest to coś, do czego powinieneś się teraz dobrze zapoznać. Następnie zwróć uwagę na wyznaczanie punktów wzdłuż krzywej. Dla uproszczenia rozważmy powierzchnię Béziera. Oczywiście można użyć poprzedniej metody analitycznej, która polegała na różnicowaniu w przód formuły i renderowaniu poprawek stałych w kwadratach parametrów. Druga technika polega na zastosowaniu przedstawionego poprzednio algorytmu podziału. W przypadku powierzchni algorytm musi działać nieco inaczej, aby uwzględnić dodatkowy wymiar wolności. Zasadniczo powierzchnia jest interpolacją interpolacji w dwóch kierunkach (zdefiniowanych przez s i t). Załóżmy, że pracujesz na powierzchni sześciennej (cztery wierzchołki kontrolne z czterema rzędami). Aby wyrenderować sześcienną powierzchnię, najpierw trzeba zastosować algorytm podziału na każdy wiersz. To da ci dwie nowe krzywe na rząd, a więc dwie nowe powierzchnie. Obróć dwie nowe powierzchnie o 90 stopni i zastosuj to samo do obu powierzchni. To da ci cztery powierzchnie. Zastosuj ten proces rekurencyjnie; możesz przestać, gdy nachylenie obu osi jest zadowalające.