Constrained Space Curve Interpolation with Constraint Planes

V.P. Kong and B.H. Ong*

School of Mathematical Sciences, Universiti Sains Malaysia, 11800 Penang, Malaysia

Abstract

We consider the interpolation of a given set of ordered space data points by a smooth curve in the presence of a set of finite or infinite constraint planes, where the polyline joining consecutive data points does not intersect with the constraint planes. The geometrical properties of the Bézier rational cubics are characterized and exploited in the derivation of conditions for the interpolant to avoid crossing the constraint planes. A method is presented for the construction of the G^2 constrained piecewise rational cubic interpolant which is local.

Keywords: Constrained interpolation; Rational cubic splines; G^2 space curve; Constraint planes

1. Introduction

Curves occur in modelling of paths of robots and particles, abstract and physical processes in economics, finance, social and physical sciences, and in numerous other situations. When a non-negative physical quantity is visualized by an interpolating curve, the latter should not admit negative values. Investigation on non-negativity preserving interpolation has been pursued to address this problem and much has been done in this respect, for example (Schmidt & Hess, 1988; Wever, 1988; Opfer & Oberle, 1988; Dougherty et al., 1989; Lahtinen, 1993). A number of attempts have also been made to consider the more general problem, range restricted interpolation, where the interpolant could be parametric or non-parametric, and as the name suggests, the interpolant is restricted to lie within a specified region. Arbitrary linear lines and quadratic curves instead of just the horizontal linear lines have been used as the boundaries of the admissible region, see (Goodman et al., 1991; Ong & Unsworth, 1992; Butt & Brodlie, 1993; Zhang & Cheng, 2001; Meek et al., 2003). The interpolating curves considered in the references cited above are planar curves.

In (Goodman et al., 1991) interpolation to planar data points that lie on one side of one or more linear lines has been considered for generating a G^2 parametric rational cubic spline which also lies on the same side of these lines. Meek et al. (2003), extend this result to generate for a given set of ordered planar points lying on one side of a polyline, a planar G^2 interpolating curve to these data which lies on the same side of the polyline as the data. This allows the polyline as a more general constraint where all the data points need not lie on one side of the infinite line through each of its edges. These two schemes generate curves which are shape preserving in the sense that they have the minimal number of inflections consistent with the data.

We consider the problem of constrained interpolation by parametric space curves with planes as constraints. This extends the works of (Goodman et al., 1991) and (Meek et al. 2003) from

* Corresponding author.
E-mail address : bhong@cs.usm.my *Fax*: 6 04 6570910

the two dimensional to the three dimensional setting. We characterize the corresponding geometrical structures of the Bézier rational cubics to derive a method which generates constrained G^2 interpolating space curves that avoid intersection with given constraint planes. Given a finite sequence of space data points and constraint planes where the polyline connecting consecutive data points does not intersect with these constraint planes, the interpolating curve also does not cross these constraint planes. This type of constrained interpolation could be useful in problems like designing a smooth space curve within a specified region or generating a smooth robot's path in space that avoids obstacles (corners or polyhedral objects) in the navigation of mobile robots (McKerrow, 1991).

In Section 2, we describe the geometrical properties of the parametric rational cubic Bézier space curves of the form $\boldsymbol{R}(t, \alpha, \beta)$, $t \in [0,1]$ in (2.1) with positive weights α, β, where the control polygon is not planar. Three geometrical structures, the β-surface, α-surface and the t-triangle are obtained from $\boldsymbol{R}(t, \alpha, \beta)$ by respectively holding β, α and t fixed while allowing the other two to vary. Each family of these structures fill the interior of the tetrahedron with the control points of $\boldsymbol{R}(t, \alpha, \beta)$ as its vertices. The "nested" and "layered" properties of these structures are exploited in the construction of the constrained interpolating curve in avoiding the intersection with the given constraint planes. Section 3 derives the necessary and sufficient conditions for a rational cubic Bézier to touch a plane. The construction of the constrained interpolating curve together with an algorithm is given Section 4. To conclude and to illustrate the method two graphical examples are presented in Section 5.

2.. The parametric Bézier rational cubic space curves

Consider the parametric rational cubic Bézier curve in space

$$\boldsymbol{R}(t, \alpha, \beta) = \frac{\alpha(1-t)^3 \boldsymbol{A} + 3(1-t)^2 t \boldsymbol{B} + 3(1-t)t^2 \boldsymbol{C} + \beta t^3 \boldsymbol{D}}{\alpha(1-t)^3 + 3(1-t)^2 t + 3(1-t)t^2 + \beta t^3}, \quad 0 \le t \le 1, \tag{2.1}$$

with the weights α, $\beta > 0$, and the control points $\boldsymbol{A}, \boldsymbol{B}, \boldsymbol{C}, \boldsymbol{D} \in \Re^3$. With α, β as fixed and abbreviating $\boldsymbol{R}(t, \alpha, \beta)$ as $\boldsymbol{R}(t)$, we have

$$\begin{aligned} &\boldsymbol{R}(0) = \boldsymbol{A}, && \boldsymbol{R}(1) = \boldsymbol{D}, \\ &\boldsymbol{R}'(0) = 3(\boldsymbol{B}-\boldsymbol{A})/\alpha, && \boldsymbol{R}'(1) = 3(\boldsymbol{D}-\boldsymbol{C})/\beta, \\ &\kappa(0) = \frac{2\alpha \left\| (\boldsymbol{B}-\boldsymbol{A}) \times (\boldsymbol{C}-\boldsymbol{B}) \right\|}{3\left\| \boldsymbol{B}-\boldsymbol{A} \right\|^3}, && \kappa(1) = \frac{2\beta \left\| (\boldsymbol{C}-\boldsymbol{B}) \times (\boldsymbol{D}-\boldsymbol{C}) \right\|}{3\left\| \boldsymbol{D}-\boldsymbol{C} \right\|^3}, \end{aligned} \tag{2.2}$$

where $\kappa(0)$ and $\kappa(1)$ are the curvatures of $\boldsymbol{R}(t)$ at 0 and 1, and $\| \cdot \|$ denotes the magnitude of a vector. The curve $\boldsymbol{R}(t)$, $t \in [0,1]$ lies in the convex hull of its control points, i.e. the tetrahedron $\boldsymbol{ABCD}$.

As the case of the parametric Bézier rational cubic in $\Re^2$ has been explored in (Meek et al., 2003), let us assume that $\boldsymbol{A}, \boldsymbol{B}, \boldsymbol{C}, \boldsymbol{D}$ are not coplanar for otherwise the rational cubic would be a planar curve. We shall first describe some simple and significant geometric properties of $\boldsymbol{R}(t, \alpha, \beta)$ which can be obtained by considering its first partial derivatives.

2.1 Cone-shaped β-surface and α-surface

For $t \in (0,1)$ and $\beta > 0$, $\lim_{\alpha \to 0} \boldsymbol{R}(t,\alpha,\beta) = \dfrac{3(1-t)^2 \boldsymbol{B} + 3(1-t)t\boldsymbol{C} + \beta t^2 \boldsymbol{D}}{3(1-t)^2 + 3(1-t)t + \beta t^2}$ and $\lim_{\alpha \to \infty} \boldsymbol{R}(t,\alpha,\beta) = \boldsymbol{A}$. The former is rational quadratic in t and rational linear in β, and is denoted as $\boldsymbol{R}(t, 0, \beta)$.

$$\frac{\partial \boldsymbol{R}}{\partial \alpha}(t, \alpha, \beta) = \frac{(1-t)^3}{W(t,\alpha,\beta)^2}\,[3(1-t)^2 t + 3(1-t)t^2 + \beta t^3]\,[\boldsymbol{A} - \boldsymbol{R}(t, 0, \beta)],$$

where $W(t, \alpha, \beta) = \alpha(1-t)^3 + 3(1-t)^2 t + 3(1-t)t^2 + \beta t^3$. Thus for any fixed $t_0 \in (0,1)$ and $\beta_0 > 0$, as α increases from 0 to ∞, $\boldsymbol{R}(t_0, \alpha, \beta_0)$ moves along a linear line from the point $\boldsymbol{R}(t_0, 0, \beta_0)$ towards $\boldsymbol{A}$. This geometric property gives rise to a cone-shaped constant β surface which we shall refer to as the β_0-surface (Figure 1(a)). This surface consists of open linear line segments indexed by $t \in (0,1)$ where the corresponding open line segment connects $\boldsymbol{R}(t, 0, \beta_0)$ to $\boldsymbol{A}$. Moreover the β_0-surface also represents the family of rational cubic curves indexed by α, $\{\boldsymbol{R}(t, \alpha, \beta_0),\ t \in (0,1) : 0 < \alpha < \infty\}$, as shown in Figure 1(b). As another consequence of the above geometric property, this family of curves is nested, i.e. each curve of the family does not intersect any other curve in the family.

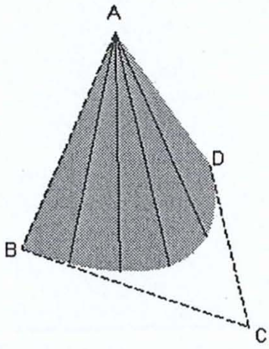


Figure 1(a). β_0-surface, $\boldsymbol{R}(t, \alpha, \beta_0)$, $\alpha \in (0, \infty)$, $t \in (0,1)$, for fixed $\beta_0 > 0$, with some locus lines where $t = 0.2, 0.4, 0.6, 0.8$.

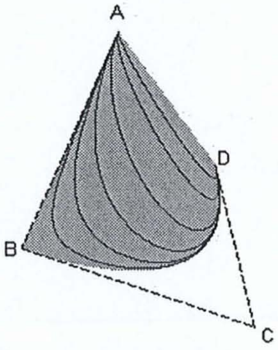


Figure 1(b). Nested curves $\boldsymbol{R}(t, \alpha, \beta_0)$, $t \in (0,1)$, for a fixed $\beta_0 > 0$ and six fixed values of α, which lie on β_0-surface.

Similarly, for $t \in (0,1)$ and $\alpha > 0$, $\lim_{\beta \to 0} \boldsymbol{R}(t,\alpha,\beta) = \dfrac{\alpha(1-t)^2 \boldsymbol{A} + 3(1-t)t\boldsymbol{B} + 3t^2 \boldsymbol{C}}{\alpha(1-t)^2 + 3(1-t)t + 3t^2}$ which is denoted as $\boldsymbol{R}(t,\alpha,0)$ and $\lim_{\beta \to \infty} \boldsymbol{R}(t,\alpha,\beta) = \boldsymbol{D}$. The partial derivatives of $\boldsymbol{R}(t, \alpha, \beta)$ with respect to β is

$$\frac{\partial \boldsymbol{R}}{\partial \beta}(t, \alpha, \beta) = \frac{t^3\,[\alpha(1-t)^3 + 3(1-t)^2 t + 3(1-t)t^2]}{W(t,\alpha,\beta)^2}\,[\boldsymbol{D} - \boldsymbol{R}(t,\alpha,0)].$$

For any fixed $t_0 \in (0,1)$ and $\alpha_0 > 0$, as β increases from 0 to ∞, $\boldsymbol{R}(t_0, \alpha_0, \beta)$ moves from the point $\boldsymbol{R}(t_0, \alpha_0, 0)$ linearly towards $\boldsymbol{D}$. This yields a cone-shaped constant α surface named the α_0-surface (Figure 2(a)) consisting of open straight line segments joining $\boldsymbol{R}(t, \alpha_0, 0)$ to $\boldsymbol{D}$, for every $t \in (0,1)$. The α_0-surface also represents the family of nested curves $\{\boldsymbol{R}(t, \alpha_0, \beta),\ t \in (0,1) : 0 < \beta < \infty\}$, as shown in Figure 2(b).

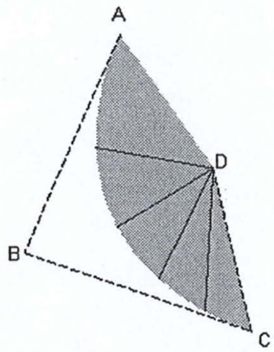


Figure 2(a). α_0-surface, $\boldsymbol{R}(t, \alpha_0, \beta)$, $\beta \in (0, \infty)$, $t \in (0,1)$, for fixed $\alpha_0 > 0$, with some locus lines where $t = 0.2, 0.4, 0.6, 0.8$.

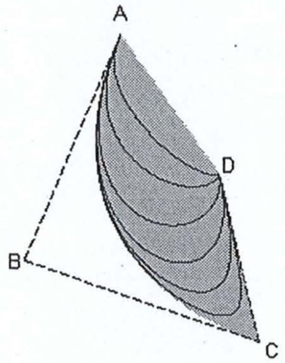


Figure 2(b). Nested curves $\boldsymbol{R}(t, \alpha_0, \beta)$, $t \in (0,1)$, for a fixed $\alpha_0 > 0$ and six fixed values of β, which lie on α_0-surface.

Consider now the limiting surfaces which are obtained when $\alpha \to 0$ or $\beta \to 0$ in $\boldsymbol{R}(t, \alpha, \beta)$, i.e.

$$\boldsymbol{R}(t, 0, \beta) = \frac{3(1-t)^2 \boldsymbol{B} + 3(1-t)t\boldsymbol{C} + \beta t^2 \boldsymbol{D}}{3(1-t)^2 + 3(1-t)t + \beta t^2}, \quad 0 < t < 1,\ 0 < \beta < \infty,$$

$$\boldsymbol{R}(t, \alpha, 0) = \frac{\alpha(1-t)^2 \boldsymbol{A} + 3(1-t)t\boldsymbol{B} + 3t^2 \boldsymbol{C}}{\alpha(1-t)^2 + 3(1-t)t + 3t^2}, \quad 0 < t < 1,\ 0 < \alpha < \infty,$$

which actually coincide with the interior of triangles $\boldsymbol{BCD}$ and $\boldsymbol{ABC}$ respectively. We have $\lim_{\beta\to\infty} \boldsymbol{R}(t, 0, \beta) = \boldsymbol{D}$ and $\lim_{\beta\to 0} \boldsymbol{R}(t, 0, \beta) = (1-t)\boldsymbol{B} + t\boldsymbol{C}$ which is denoted as $\boldsymbol{S}(t)$.

The partial derivative of $\boldsymbol{R}(t, 0, \beta)$ with respect to β is

$$\frac{\partial \boldsymbol{R}}{\partial \beta}(t, 0, \beta) = \frac{3(1-t)t^2}{V(t,\beta)^2}[\boldsymbol{D} - [(1-t)\boldsymbol{B} + t\boldsymbol{C}]]$$

where $V(t,\beta) = 3(1-t)^2 + 3(1-t)t + \beta t^2$. As β increases from 0 to ∞, for any fixed $t \in (0,1)$, $\boldsymbol{R}(t, 0, \beta)$ moves along a linear line from point $\boldsymbol{S}(t)$ towards $\boldsymbol{D}$. Moreover, for each fixed $\beta > 0$, $\boldsymbol{R}(t, 0, \beta)$, $t \in (0,1)$, is a rational quadratic Bézier curve joining $\boldsymbol{B}$ to $\boldsymbol{D}$ and hence it is planar and convex. Thus $\{\boldsymbol{R}(t, 0, \beta),\ t \in (0,1) : 0 < \beta < \infty\}$ is a nested family of planar curves, indexed by β, which fills the interior of triangle $\boldsymbol{BCD}$ (Figure 3). By this nested property and that a β-surface is made up of open line segments joining $\boldsymbol{A}$ to points of the rational quadratic curve $\boldsymbol{R}(t, 0, \beta)$, $t \in (0,1)$, we obtain that the cone-shaped

β-surfaces, $\beta \in (0, \infty)$, form a nested family which fills the interior of tetrahedron ***ABCD*** (Figure 4). Here a family of surfaces is nested if each surface of the family does not intersect any other surface in the family.

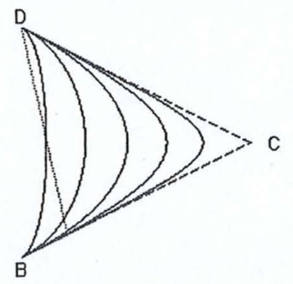


Figure 3. Nested property of $\{ \boldsymbol{R}(t, 0, \beta), \ t \in (0,1) : \ \beta \in (0,\infty) \}$ and the linear line segment $\boldsymbol{R}(0.2, 0, \beta), \beta \in (0, \infty)$.

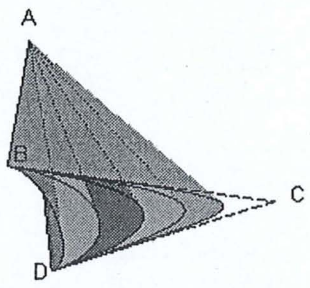


Figure 4. Nested property of the β-surfaces.

Similarly $\{ \boldsymbol{R}(t, \alpha, 0), \ t \in (0,1) \ : 0 < \alpha < \infty \}$, indexed by α, is a family of nested planar curves that fills the interior of the triangle ***ABC***, and the family of α-surfaces is nested and fills the interior of tetrahedron ***ABCD***.

Observe that the rational cubic curve $\boldsymbol{R}(t, \alpha, \beta), \ t \in (0,1)$ lies on the α-surface and β-surface, hence it is the intersection of α-surface and β-surface (Figure 5).

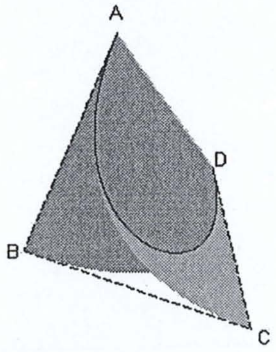


Figure 5. Curve $\boldsymbol{R}(t, \alpha, \beta), \ t \in (0,1)$, is the intersection of the α-surface and β-surface.

2.2 *t*-triangle

Now for any fixed $t_0 \in (0,1)$,

$$\boldsymbol{R}(t_0, \alpha, \beta) = \frac{\alpha(1-t_0)^3}{W_0}\boldsymbol{A} + \frac{3(1-t_0)t_0}{W_0}\left[(1-t_0)\boldsymbol{B} + t_0\boldsymbol{C}\right] + \frac{\beta t_0^{\,3}}{W_0}\boldsymbol{D}$$

where $W_0 = \alpha(1-t_0)^3 + 3(1-t_0)^2 t_0 + 3(1-t_0)t_0^{\,2} + \beta t_0^{\,3}$. As each $\boldsymbol{R}(t_0, \alpha, \beta)$, $\alpha, \beta \in (0, \infty)$, is a convex combination of the three points ***A***, $\boldsymbol{S}_0 = (1-t_0)\boldsymbol{B} + t_0\boldsymbol{C}$ and ***D***, thus $\{\boldsymbol{R}(t_0, \alpha, \beta) : \alpha, \beta \in (0,\infty)\}$ is the interior of triangle $\boldsymbol{AS_0D}$.

Recall that

$$R(t_0, \alpha, 0) = \frac{\alpha(1-t_0)^2 A + 3t_0\left[(1-t_0)B + t_0 C\right]}{\alpha(1-t_0)^2 + 3(1-t_0)t_0 + 3t_0^2}, \quad \alpha \in (0, \infty),$$

is the open linear line segment S_0A. Similarly, $R(t_0, 0, \beta)$, $\beta \in (0, \infty)$, is the open linear line segment S_0D. The interior of triangle AS_0D shall be referred as the t_0-triangle (Figure 6). The collection of all the layered t-triangles, $t \in (0,1)$, fills the interior of the tetrahedron ***ABCD***.

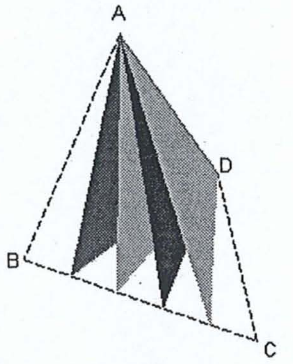


Figure 6. Layered t-triangles, $t = 0.2, 0.4, 0.6, 0.8$.

2.3 Uniqueness of representation

Based upon the above discussion, we have characterized the interior of the tetrahedron whose vertices are the control points A, B, C, D of $R(t, \alpha, \beta)$ which are not coplanar, as either one of the following structures:

(i) the set of points $R(t, \alpha, \beta)$, $t \in (0,1)$, $\alpha, \beta \in (0, \infty)$.

(ii) the set of nested α-surfaces, $\alpha \in (0, \infty)$

(iii) the set of nested β-surfaces, $\beta \in (0, \infty)$

(iv) the set of layered t-triangles, $t \in (0,1)$.

Indeed every interior point of the tetrahedron $ABCD$ has a unique representation of the form $R(t, \alpha, \beta)$ as noted in Proposition 1.

Proposition 1 Given a point G in the interior of the tetrahedron $ABCD$ where A, B, C and D are not coplanar, then exists a unique rational cubic curve $R(t, \alpha, \beta)$ of (2.1) passing through it, i.e. there exists a unique $t_0 \in (0,1)$, $\alpha_0, \beta_0 > 0$ such that $R(t_0, \alpha_0, \beta_0) = G$.

Proof : We shall first find the t-triangle on which the point G lies. Let the point of projection of A onto the triangle BCD through point G be denoted as Q_1 (see Figure 7). AQ_1 is the line segment $R(t_0, \alpha, \beta_0), \alpha \in (0, \infty)$ for some $\beta_0 > 0$ and $t_0 \in (0,1)$. Then the point D is projected through Q_1 to the edge BC, with the point obtained be denoted as S_0. The value t_0 can be found easily from the linear equation $S_0 = (1-t_0)B + t_0 C$. β_0 can then be determined from the linear equation $Q_1 = R(t_0, 0, \beta_0)$. Lastly, since the line AQ_1 passes through G, then $\alpha = \alpha_0 > 0$ can be found from the linear equation $G = R(t_0, \alpha, \beta_0)$. Thus there exists t_0, α_0, β_0 such that $R(t_0, \alpha_0, \beta_0) = G$.

We can also find a solution $R(t_1, \alpha_1, \beta_1) = G$ by first projecting D through G to a point Q_2 which lies on triangle ABC, and then projecting A through Q_2 to a point S on BC. From this sequence of operations, the values of t_1, then α_1 and finally β_1 can be computed easily as above.

We now show that $(t_0, \alpha_0, \beta_0) = (t_1, \alpha_1, \beta_1)$. As we know the interior of the tetrahedron $ABCD$ consists of layered t-triangles, $t \in (0,1)$, which are the interior of triangles ASD, where S is an interior point of BC. Since these t-triangles do not intersect one another, so G lies on one and only one of the t-triangles, hence $t_0 = t_1$. We shall show that the two tuples (α_0, β_0) and (α_1, β_1) are identical. The first tuple is based on the β-surface and the projection point Q_1, while the second tuple is based on the α-surface and the projection point Q_2. Observe that G lies on the t_0-triangle which is the interior of the triangle AS_0D with point $S_0 = (1-t_0)B + t_0 C$, Q_1 and Q_2 lie on DS_0 and AS_0 respectively. The line between the points Q_1 and A is represented by $R(t_0, \alpha, \beta_0)$, $\alpha > 0$, while the line between Q_2 and D is $R(t_0, \alpha_1, \beta)$, $\beta > 0$. Since both lines pass through G, *i.e.* the point G is the intersection point of lines $R(t_0, \alpha, \beta_0)$ and $R(t_0, \alpha_1, \beta)$, hence $\alpha_0 = \alpha_1$ and $\beta_1 = \beta_0$.

To show uniqueness, suppose that $R(t^*, \alpha^*, \beta^*) = G$. Then t^*, α^* and β^* can be computed by either one of the two approaches described above and thus the solution set is unique. □

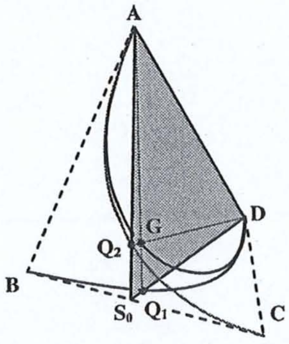


Figure 7. Geometric view of the unique representation $R(t, \alpha, \beta)$ for a space point G.

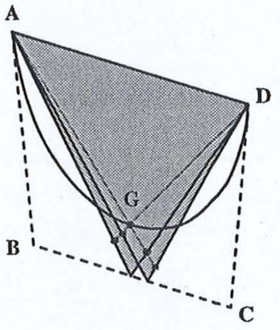


Figure 8. Geometric view of 2 of the infinite representations $R(t, \alpha, \beta)$ for the point G in the planar quadrilateral $ABCD$.

Observe that $G = R(t_0, \alpha_0, \beta_0)$ lies on the t_0-triangle. If A, B, C, D are coplanar, we obtain an infinite set of t-triangles that contained G, hence there are an infinite number of representation of G (Figure 8) in terms of t, α and β. However, if we fix either one of the weights α or β, or the ratio of these two weights, or the t value, then a unique representation can be achieved when the polyline $\boldsymbol{ABCD}$ is C-shaped, but if the polyline $\boldsymbol{ABCD}$ is S-shaped, then the representation need not be unique (see Meek et. al, 2003).

2.4 Partially linear (α, β)-surface

Now suppose that we scale both weights α and β while keeping their ratio fixed. Consider

$$\boldsymbol{R}(t, \lambda\alpha, \lambda\beta) = \frac{\lambda\alpha(1-t)^3 \boldsymbol{A} + 3(1-t)^2 t \boldsymbol{B} + 3(1-t)t^2 \boldsymbol{C} + \lambda\beta t^3 \boldsymbol{D}}{\lambda\alpha(1-t)^3 + 3(1-t)^2 t + 3(1-t)t^2 + \lambda\beta t^3}, \quad t \in (0,1),$$

where $\alpha, \beta > 0$ are fixed and $\lambda \in (0, \infty)$.

$$\frac{\partial \boldsymbol{R}}{\partial \lambda}(t, \lambda\alpha, \lambda\beta) = \frac{3(1-t)t}{\eta(t,\lambda,\alpha,\beta)}[\alpha(1-t)^3 + \beta t^3]\left\{\frac{\alpha(1-t)^3 \boldsymbol{A} + \beta t^3 \boldsymbol{D}}{\alpha(1-t)^3 + \beta t^3} - [(1-t)\boldsymbol{B} + t\boldsymbol{C}]\right\}$$

where $\eta(t,\lambda,\alpha,\beta) = \lambda\alpha(1-t)^3 + 3(1-t)^2 t + 3(1-t)t^2 + \lambda\beta t^3$.

For $t \in (0,1)$, $\lim_{\lambda\to\infty} \boldsymbol{R}(t,\lambda\alpha,\lambda\beta) = \dfrac{\alpha(1-t)^3 \boldsymbol{A} + \beta t^3 \boldsymbol{D}}{\alpha(1-t)^3 + \beta t^3}$ and $\lim_{\lambda\to 0} \boldsymbol{R}(t,\lambda\alpha,\lambda\beta) = (1-t)\boldsymbol{B} + t\boldsymbol{C}$.

Let us denote these two limits respectively as $\boldsymbol{R}(t,\infty,\infty)$ and $\boldsymbol{R}(t,0,0)$, then for any fixed $t \in (0,1)$, as λ varies from 0 to ∞, $\boldsymbol{R}(t,\lambda\alpha,\lambda\beta)$ moves along a linear line from $\boldsymbol{R}(t,0,0)$ towards $\boldsymbol{R}(t,\infty,\infty)$. These open linear segments form a partially linear surface which is referred to as an (α,β)-surface as shown in Figure 9(a). Moreover, this (α,β)-surface can be regarded as the family of curves $\{\boldsymbol{R}(t,\lambda\alpha,\lambda\beta),\ t\in(0,1) : \lambda \in (0,\infty)\}$ as shown in Figure 9(b). When λ decreases, $\boldsymbol{R}(t,\lambda\alpha,\lambda\beta)$, $t \in (0,1)$, approaches the line segment $\boldsymbol{BC}$ while when λ increases, it will approach the line segment $\boldsymbol{AD}$. Thus this family of curves $\{\boldsymbol{R}(t,\lambda\alpha,\lambda\beta),\ t\in(0,1) : \lambda\in(0,\infty)\}$ is nested.

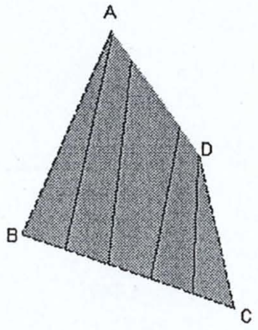


Figure 9(a). Partially linear $\alpha\beta$-surface $\boldsymbol{R}(t,\lambda\alpha,\lambda\beta)$, $\lambda\in(0,\infty)$, $t\in(0,1)$ with some locus lines where $t = 0.2, 0.4, 0.6, 0.8$.

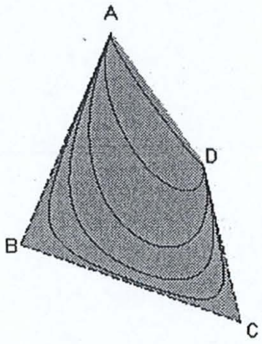


Figure 9(b): Nested curves $\boldsymbol{R}(t,\lambda\alpha,\lambda\beta)$, $t\in(0,1)$, on partially linear (α,β)-surface.

3. The point of contact between the rational cubic and a plane

Let us recall the following result quoted from (Meek et al., 2003) on the zeros of multiplicities 2 or 3 of a cubic Bernstein polynomial on $(0,1)$.

Lemma 1. *Let* $g(t)=a(1-t)^3+3bt(1-t)^2+3ct^2(1-t)+dt^3$, $t\in[0,1]$, $a, b, c, d\in\Re$.

(i) g *has a zero of multiplicity* 2 *in* (0, 1) *if and only if*

$$(ad-bc)^2=4(ac-b^2)(bd-c^2)$$

and

$$(ac-b^2)(ad-bc)<0 \ \text{[equivalently } (bd-c^2)(ad-bc)<0\text{]}.$$

This zero occurs at $\xi\in(0,1)$ *where* $\dfrac{\xi}{\xi-1}=\dfrac{2(ac-b^2)}{ad-bc}=\dfrac{ad-bc}{2(bd-c^2)}<0$.

(ii) g *has a zero of multiplicity of* 3 *in* (0, 1) *if and only if* $\dfrac{a}{b}=\dfrac{b}{c}=\dfrac{c}{d}<0$.

The zero occurs at $\xi\in(0,1)$ *where* $\dfrac{\xi}{\xi-1}=\dfrac{a}{b}=\dfrac{b}{c}=\dfrac{c}{d}<0$.

First we note that a smooth rational function with a non vanishing denominator has a multiple zero if and only if its numerator has a zero of the same multiplicity at the same point.

Consider the parametric rational cubic $\boldsymbol{R}(t,\alpha,\beta)$, $t\in[0,1]$, of (2.1) where $\alpha,\beta>0$ are fixed. $\boldsymbol{R}(t,\alpha,\beta)$ touches the plane $z=0$ at $t=t_0\in(0,1)$ if for $\boldsymbol{k}=(0,0,1)$,

(i) $\boldsymbol{k}\bullet\boldsymbol{R}(t_0,\alpha,\beta)=0$ (where $\bullet$ denotes the dot product)

(ii) there exists $\varepsilon>0$ such that

$$\boldsymbol{k}\bullet\boldsymbol{R}(t,\alpha,\beta)>0,\ \forall t\in(t_0-\varepsilon,t_0+\varepsilon)\setminus\{t_0\}$$

$$\text{or } \boldsymbol{k}\bullet\boldsymbol{R}(t,\alpha,\beta)<0,\ \forall t\in(t_0-\varepsilon,t_0+\varepsilon)\setminus\{t_0\}.$$

Statement (i) together with statement (ii) above are equivalent to that the rational cubic function

$h(t)=\boldsymbol{k}\bullet\boldsymbol{R}(t,\alpha,\beta)$, $t\in(0,1)$, has a zero of multiplicity 2 or 3 at $t=t_0$.

More generally for $\boldsymbol{R}(t,\alpha,\beta)$, $t\in(0,1)$, to touch at $t=t_0\in(0,1)$ any plane $ax+by+cz+d=0$ whose normal vector is $\boldsymbol{n}$, the above two conditions apply with $\boldsymbol{k}$ being replaced by $\boldsymbol{n}$.

Furthermore, the Bézier representation is affine invariant and via rotation and translation which are affine transformations, any given constraint plane may be transformed to the plane $z=0$. With these observations, we derive directly from Lemma 1 the necessary and sufficient conditions for the rational cubic $\boldsymbol{R}(t,\alpha,\beta)$, $t\in[0,1]$ to touch a given plane as given in Proposition 2.

Proposition 2. *Consider the parametric rational cubic Bézier curve* $\boldsymbol{R}(t,\alpha,\beta)$, $0\le t\le 1$, *of* (2.1) *where* $\alpha,\beta>0$ *and a given constraint plane. Then* $\boldsymbol{R}(t,\alpha,\beta)$ *touches the constraint plane at* $\xi\in(0,1)$ *if and only if one of the following two holds.*

(i) $(\alpha \beta a d - b c)^2 = 4(\alpha a c - b^2)(\beta b d - c^2)$ *and*

$(\alpha a c - b^2)(\alpha \beta a d - b c) < 0$ [*equivalently* $(\beta b d - c^2)(\alpha \beta a d - b c) < 0$]

(ii) $\dfrac{\alpha a}{b} = \dfrac{b}{c} = \dfrac{c}{\beta d} < 0$

where a, b, c, d are respectively the signed distances of the control points **A**, **B**, **C**, **D** of $\boldsymbol{R}(t, \alpha, \beta)$ *from the constraint plane, with points on one side of constraint plane having positive distances while points on the other side having negative distances.*

With condition (i), $\xi \in (0,1)$ *is given by* $\dfrac{\xi}{\xi - 1} = \dfrac{2(\alpha ac - b^2)}{\alpha \beta ad - bc} = \dfrac{\alpha \beta ad - bc}{2(\beta bd - c^2)} < 0$ *while with condition* (ii), $\xi \in (0,1)$ *is given by* $\dfrac{\xi}{\xi - 1} = \dfrac{\alpha a}{b} = \dfrac{b}{c} = \dfrac{c}{\beta d} < 0$.

4. Construction of the constrained interpolating curve

Suppose that we are given an ordered sequence of distinct data points $\{ \boldsymbol{I}_i : 0 \le i \le n \}$ in space and a finite collection of constraint planes. The polyline which joins consecutive data points does not intersect with any of the constraint planes. We would like to construct a G^2 piecewise rational cubic interpolating curve which does not cross any given constraint plane but may touch these planes or pass through points along the boundaries of these planes.

The constraint planes considered may be infinite or finite, and may have finite number of boundaries which are finite or infinite linear line segments. Thus our constraint plane may be described as one of the following types:

(PI) an infinite plane $ax + by + cz + d = 0$ without boundary, where $a, b, c, d \in \Re$

(PII) an infinite plane with a finite number of linear line segments as its boundaries

(PIII) a finite plane bounded by a finite number of finite linear line segments.

We assume without loss of generality that each constraint plane is convex for otherwise it can be subdivided into two or more convex constraint planes.

There are two main steps in constructing the constrained interpolating curve :

(i) generation of the initial default G^2 interpolating curve

(ii) modification of some curve segments if necessary so that the resulting interpolating curve does not intersect any of the constraint planes.

Here we shall concentrate on the second step and shall not deliberate on the first step though the final interpolant is of course dependent on the initial default curve. The default curve may be any G^2 piecewise rational cubics with positive weights or it may also be G^1 instead of G^2 if we only require G^1 continuity. We use the shape preserving G^2 interpolation scheme in (Goodman & Ong, 1997) to generate the initial piecewise rational cubic interpolating curve. The scheme preserves convexity and inflections which are defined according to some criteria, and it also preserves collinearity, coplanarity and sign of the torsion. When there are three or more consecutive collinear data points, a linear line segment is generated to interpolate these data points and the interpolating curve to all the data points may not be curvature continuous but is G^1 at the junction point where a linear curve segment meets a non-linear curve segment. We refer the interested reader to that paper for details of the construction. Each curve segment generated is expressed in a form similar to (2.1). It differs from (2.1) only in a multiplicative factor of 3 for α and β. Thus we can easily reexpress the curve segment in the form (2.1). Henceforth we may assume that the curve segments of the default curve are of the form (2.1). Some of the curve segments of the default interpolant could have crossed the given constraint planes. We apply the properties and results presented in Sections 2 and 3 to identify and modify the curve segments which have crossed any of the given constraint planes by scaling the corresponding weights α or β, or both α and β. A curve segment which does not cross any given constraint plane will not be modified except when the G^2 continuity at any of its end points is disrupted by the modification on its adjacent segments.

4.1 Modification of curve segments

Each segment will be treated locally in the order of the given data points. Denote the i-th interpolating curve segment between $\boldsymbol{I}_i$ and $\boldsymbol{I}_{i+1}$ as $\boldsymbol{R}_i$. By performing some or all of the three following operations on each curve segment $\boldsymbol{R}_i$, the set Φ_i of constraint planes which intersect with the interior of the control tetrahedron $\boldsymbol{ABCD}$ of the curve segment $\boldsymbol{R}_i$ is identified. Here for the sake of simplicity, the subscript i for the control points of $\boldsymbol{R}_i$ are omitted.

For each given constraint plane,

(i) determine the position of each control point $\boldsymbol{A}, \boldsymbol{B}, \boldsymbol{C}, \boldsymbol{D}$ of $\boldsymbol{R}_i$ relative to the constraint plane

(ii) test whether each of the edges $\boldsymbol{AB}, \boldsymbol{AC}, \boldsymbol{BC}, \boldsymbol{BD}$ and $\boldsymbol{CD}$ intersect with the interior of the constraint plane

(iii) test whether the boundaries of the constraint plane, if any, intersect with the interior of the triangular faces of the tetrahedron $\boldsymbol{ABCD}$.

Steps (ii) and (iii) above often involve computing the intersection point between a line segment and a plane. Finding the intersection point of a line segment with a plane when two points of the line segment lie on opposite sides of the plane is a simple direct computation (see Appendix A).

Three types of modification are used : scaling α, scaling β or scaling both with a factor $\lambda > 1$. The type of modification applied on curve segment $\boldsymbol{R}_i$ is determined based upon the positions of its control points relative to all the constraint planes in Φ_i. It should be made clear that in determining the positions of the control points relative to a given constraint plane, the constraint plane is assumed to be the full infinite plane of type (PI) above irrespective of whether it is finite or infinite. Thus to avoid confusion, let us denote by Φ_i^* the set of constraint planes obtained by extending each constraint plane in Φ_i, be it finite or infinite, to the full infinite plane of type (PI). In any other context in this paper, by a constraint plane it shall mean the actual constraint plane, be it finite or infinite. We classify the relative positions of the control points $\boldsymbol{A}, \boldsymbol{B}, \boldsymbol{C}, \boldsymbol{D}$ of the curve segment $\boldsymbol{R}_i$ with respect to a constraint plane $\mathcal{P}$ in Φ_i^* into two cases :

(CI) $\boldsymbol{A}, \boldsymbol{B}, \boldsymbol{C}, \boldsymbol{D}$ are coplanar and lie on the plane $\mathcal{P}$

(CII) Any other case which is not case (CI).

The curve modification to case (CI) can be reduced to the 2 dimensional problem considered in (Meek et al., 2003) where a Bézier rational cubic curve segment in $\Re^2$ is constrained so that it does not intersect a given linear line segment but may touch the the line segment or pass through the end point of the line segment. Thus we shall omit here and not repeat the discussion about the curve modification for the case (CI) and refer the reader to (Meek et al., 2003) for the details of the treatment for that case. Henceforth we shall consider case (CII) . We note that the case where $\boldsymbol{A}, \boldsymbol{C}$ and $\boldsymbol{D}$ lie on a constraint plane $\mathcal{P}$ in Φ_i^* would not happen as this would contradict the fact that $\mathcal{P}$ intersects the interior of the tetrahedron $\boldsymbol{ABCD}$. Similarly that $\boldsymbol{A}$, $\boldsymbol{B}$ and $\boldsymbol{D}$ lie on $\mathcal{P}$ would not happen. The case (CII) on the relative positions of the control points with respect to a constraint plane $\mathcal{P}$ in Φ_i^* is subdivided into three sub groups as listed below.

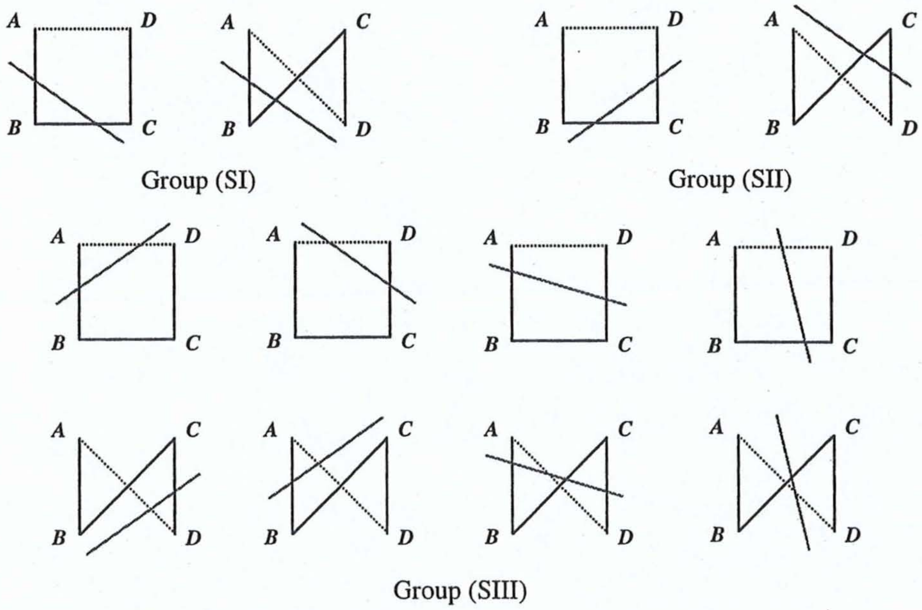


Figure 10. Positions of the control points relative to the constraint plane.

(SI) The point $\boldsymbol{B}$ lies on the opposite side of the constraint plane $\mathcal{P}$ as points $\boldsymbol{A}, \boldsymbol{C}, \boldsymbol{D}$. Here the points $\boldsymbol{A}, \boldsymbol{C}$ or $\boldsymbol{D}$ may lie on the plane $\mathcal{P}$ but not all three together.

(SII) The point **C** lies on opposite side of the constraint plane $\mathcal{P}$ as points $\boldsymbol{A}, \boldsymbol{B}, \boldsymbol{D}$. Here the points $\boldsymbol{A}, \boldsymbol{B}$ or $\boldsymbol{D}$ may lie on $\mathcal{P}$ but not all three together.

(SIII) Any other case which is not case (SI) nor case (SII).

Figure 10 illustrates the classification above. In each sub figure, the unlabelled line segment denotes the constraint plane.

The positions of the control points of a curve segment $\boldsymbol{R}_i$ relative to the constraint planes in Φ_i^* vary from plane to plane. If with respect to all the planes in Φ_i^*, the positions of the control points fall solely in group (SI), then the curve segment is modified, if necessary, by scaling its weight α. If the positions of the control points relative to all the planes in Φ_i^* fall only in group (SII), then modification, if necessary, proceeds by scaling the weight β. When the positions of the control points fall in group (SIII) or have a mixed bahaviour with respect to different constraint planes in Φ_i^*, i.e. not purely in one group, then modification, if necessary, proceeds by scaling both weights, α and β, of the curve segment.

For simplicity, first we consider a single piece of rational cubic $\boldsymbol{R}(t, \alpha, \beta)$, $t \in (0, 1)$ and a single constraint plane. With fixed $\alpha, \beta > 0$, as λ decreases from ∞ to 0, the curves $\boldsymbol{R}(t, \lambda\alpha, \beta)$, $\boldsymbol{R}(t, \alpha, \lambda\beta)$ and $\boldsymbol{R}(t, \lambda\alpha, \lambda\beta)$, $t \in (0, 1)$ move away from **AD** to approach respectively the limiting rational quadratic curves $\boldsymbol{R}(t, 0, \beta)$, $\boldsymbol{R}(t, \alpha, 0)$, $t \in (0, 1)$ and its control polygon **$ABCD$**. In view of this behaviour and that the given constraint plane does not intersect **AD**, if we consider say the curve $\boldsymbol{R}(t, \lambda\alpha, \beta)$, $t \in (0, 1)$ with λ decreasing from ∞ to 0, this rational cubic curve will first hit the constraint plane, if at all, on the boundary of the constraint plane or at a point where it touches this constraint plane. The curves $\boldsymbol{R}(t, \alpha, \lambda\beta)$ and $\boldsymbol{R}(t, \lambda\alpha, \lambda\beta)$, $t \in (0, 1)$ have the same behaviour. Thus the following operations are performed on the curve segment $\boldsymbol{R}(t, \alpha, \beta)$ with respect to the constraint plane to determine whether the curve segment $\boldsymbol{R}(t, \alpha, \beta)$ crosses the constraint plane and if so the value of the scaling factor λ on the weight α, or β or both α and β. in accordance to the case (SI), (SII) or (SIII).

Process (a): Find the λ value of the rational cubic, if any, which touches the constraint plane by using Proposition 2.

Process (b): For each boundary line of the given constraint plane, find the λ value of the rational cubic, if any, which passes through some point of the boundary line.

If no $\lambda \in (0, \infty)$ is obtained from processes (a) and (b), or the largest λ obtained is less than 1, then the curve segment $\boldsymbol{R}(t, \alpha, \beta)$ does not cross the constraint plane. If the largest λ value from processes (a) and (b) exceeds 1, the curve segment $\boldsymbol{R}(t, \alpha, \beta)$, $t \in [0, 1]$, has crossed that constraint plane and it will be modified by scaling the weight α or β or both α and β accordingly by that λ value so that the modified curve segment does not intersect the constraint plane. Let us take a closer look at the determination of λ above for a single piece of rational cubic $R(t, \alpha, \beta)$, $t \in [0,1]$ with respect to a single constraint plane.

(i) **Determination of λ for $R(t, \lambda\alpha, \beta)$**

In this case, modification of the curve $R(t, \alpha, \beta)$, $t \in [0,1]$ is obtained by scaling α. A boundary line of the constraint plane, if exists, is an infinite or a finite line segment. It may intersect with the β-surface at most at two points except the case when it contains part of a linear line segment on the β-surface, in which case the intersection is that part of the line segment itself consisting of an infinite number of points but not the point ***A***. However this line segment when extended linearly will pass through ***A***. In the latter case, among all the curves $R(t, \lambda\alpha, \beta)$, $t \in [0,1]$, which pass through the intersection points, the curve which passes through the intersection point nearest to ***A*** has the largest λ value (see Figure 11 (d)) because as seen in Section 2.1 when λ increases, $R(t, \lambda\alpha, \beta)$ moves along a linear path towards ***A*** for each $t \in (0,1)$.

Figure 11 shows the β-surface which intersects with a constraint plane. The bold line through ***G***, ***H*** or ***F*** is the intersection curve of the constraint plane and the β-surface. The default interpolating curve which is denoted as a dotted curve intersects the constraint plane at ***G*** and ***H*** as in (a) and (c) of Figure 11. ***T*** is a contact point on the constraint plane to a rational cubic curve from the family $\{R(t, \lambda\alpha, \beta), t \in [0,1] : \lambda \in (0, \infty)\}$. ***F*** is a point on a boundary line of the constraint plane and on the β-surface. The curve drawn in bold is the rational cubic curve from the family which passes through ***F*** as in (b), (c) and (d) or touches the constraint plane at ***T*** as in (a). The λ value of the curve through the point ***T*** is determined by process (a) above while that through the point ***F*** is determined by process (b).

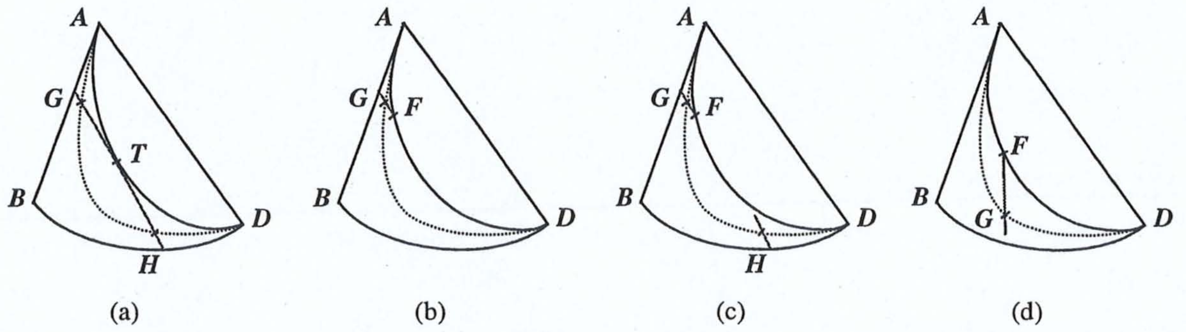


Figure 11. The default curve and the modified curve on the β-surface.

The λ value for process (a), if exists, is computed easily by using Proposition 2. To determine the λ for process (b), the t value of the point $R(t, \lambda\alpha, \beta)$ which lies on the β-surface and the boundary of the constraint plane is first obtained by solving a quadratic equation and then the corresponding λ is obtained from a linear equation. The solution, if exists, may not be unique and there may be two sets of admissible solutions for (t, λ) where $t \in (0,1)$ and $\lambda \in (0, \infty)$. There is one special case as illustrated in (d) of Figure 11 where a boundary line of the constraint plane contains a line segment on the β-surface. In this case, among the different curves $R(t, \lambda\alpha, \beta)$, $t \in [0,1]$ which pass through a point on this line segment, the curve through the point on this line segment which is nearest to A has the highest λ value. Some details of the computation involved in finding t and λ is given in Appendix B.

Processes (a) and (b) are performed to determine the λ values *i.e.* λ of the curves that pass through the point F or T if exists. No positive λ value is obtained if the constraint plane does not intersect the β-surface, otherwise if the largest λ value obtained, say $\lambda = \overline{\lambda}$, is greater than 1, it is used as the scaling factor for the weight α of the curve and the resulting curve will not cross the constraint plane. Moreover, in view of the geometrical properties described in Section 2.1, for any higher λ value, i.e. $\lambda > \overline{\lambda}$, the corresponding curve $R(t, \lambda\alpha, \beta)$, $t \in [0,1]$, will also not cross that constraint plane. If the λ value obtained is not greater than 1, then the curve $R(t, \alpha, \beta)$, $t \in [0,1]$ does not cross the constraint plane.

However for any increase in β value to the curve $R(t, \overline{\lambda}\alpha, \beta)$, $t \in [0,1]$, the resulting curve moves closer to D, but it might intersect the constraint plane though unlikely.

In the 2D case of (Meek et al., 2003), the interpolating curve segment is in $\Re^2$ and the constraints are finite or infinite linear line segments. The λ value for the curve tangent to a constraint line, if exists, is obtained by the 2D version of Proposition 2. However for the analogy to process (b), the situation is more definite for the 2D case since a boundary of a constraint line in 2D if exists is a point instead of a linear line segment as a boundary of a constraint plane. It is thus more direct to identify a 2D rational cubic curve which passes through a specified point than to identify a 3D rational cubic that passes through some point on a specified linear line segment. Nevertheless the t value for the 2D case is also obtained by solving a quadratic equation as in the 3D case.

(ii) **Determination of λ for $R(t, \alpha, \lambda\beta)$**

This is similar to case (i).

(iii) **Determination of λ for $R(t, \lambda\alpha, \lambda\beta)$**

The boundary of the constraint plane may intersect with the (α, β)-surface at most at two points except the case when it contains part of a linear line segment on the (α, β)-surface, in which case the intersection is that part of the line segment itself consisting of an infinite number of points. In the latter case, among all the curves $R(t, \lambda\alpha, \lambda\beta)$, $t \in [0,1]$, which pass through the intersection points, the curve which passes through the intersection point nearest to AD has the largest λ value because as seen in Section 4.2, when λ increases $R(t, \lambda\alpha, \lambda\beta)$ moves towards $\dfrac{\alpha(1-t)^3 A + \beta t^3 D}{\alpha(1-t)^3 + \beta t^3}$ along a linear path for each $t \in (0,1)$.

The modification was based on the family of curves $\{R(t, \lambda\alpha, \lambda\beta) : \lambda \in (0, \infty)\}$ which form the partially linear (α, β)-surface. The λ values are determined by the above processes (a) and (b). The procedure to determine λ is similar to that of the case (i) above. The λ value for process (a), if exists, is computed easily by using Proposition 2. For process (b), first the t value of the point $\boldsymbol{R}(t, \lambda\alpha, \lambda\beta)$ which lies on the boundary line segment of the constraint plane is determined by solving a polynomial equation of degree 4, then the corresponding λ value is obtained from a linear equation. If no positive λ value is obtained or the largest λ value obtained is not greater than 1, then no modification is necessary on the curve segment since $\boldsymbol{R}(t, \alpha, \beta)$, $t \in [0,1]$ does not intersect the constraint plane, otherwise the largest λ value, say $\lambda = \overline{\lambda} > 1$, is used as the scaling factor for the weights α and β of the curve and the resulting curve will not cross the constraint plane. Any curve of the family with higher λ value than λ^* will no more cross the constraint plane. Figure 12 shows the (α, β)-surface whose intersection with the constraint plane is the bold curve segment through $\boldsymbol{G}$, $\boldsymbol{H}$ or $\boldsymbol{F}$. The default interpolating curve is drawn as a dotted curve while the modified curve is drawn as a bold curve. $\boldsymbol{T}$ is a contact point on the constraint plane to a rational cubic on the (α, β)-surface. $\boldsymbol{F}$ is a point on a boundary line of the constraint plane and the (α, β)-surface. The λ of the curve through $\boldsymbol{T}$ is determined by process (a) while that through $\boldsymbol{F}$ by process (b).

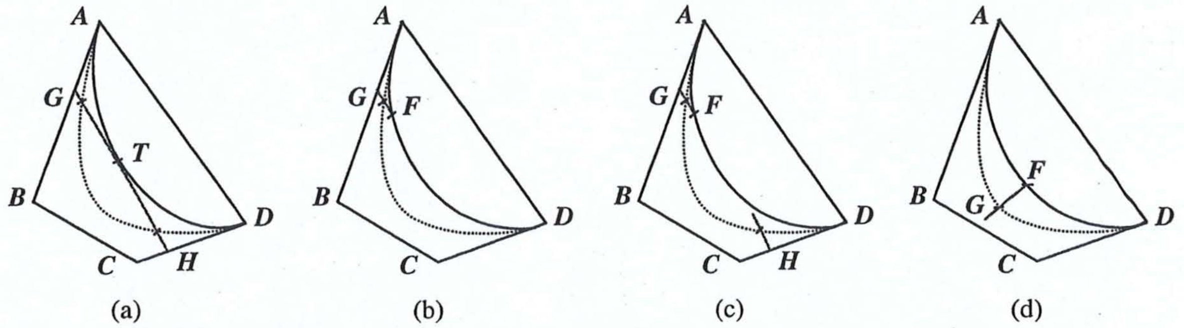


Figure 12. The default curve and the modified curve on the (α, β)-surface.

(iv) **Multiple constraint planes**

For multiple constraint planes, the set Φ_i of all the constraint planes which intersect with the interior of the control tetrahedron of the curve segment $\boldsymbol{R}_i$ is identified and the type of modification if any is determined. Each constraint plane in Φ_i is processed one by one by the above λ determination processes and the largest λ obtained from among all these constraint planes, if any and if it is greater than 1, is used as the scaling factor for the corresponding weight(s) so that the resulting modified curve does not cross any constraint plane.

4.3. Restoring G^2 continuity

When a curve segment $\boldsymbol{R}_i$, is modified by increasing one or both of its weights, the curvature at the corresponding end point of the curve segment changes according to (2.2) and this disrupts the G^2 continuity between adjacent curve segments. If the weight α of the curve segment $\boldsymbol{R}_i$ has been scaled by a factor λ_i, then to restore G^2 continuity the weight β of the preceding segment $\boldsymbol{R}_{i-1}$ has to be scaled by the same factor λ_i, while if the weight β

of the curve segment $\boldsymbol{R}_i$ have been scaled by a factor λ_i, then the weight α of the succeeding segment $\boldsymbol{R}_{i+1}$ has to be scaled by the same factor λ_i.

4.4 Algorithm for Constrained Interpolation in 3D

Based upon the above discussion, we have adopted the following approach to generate a G^2 interpolating space curve that avoids intersection with the given constraint planes. Though we have not included case (CI) in the algorithm below, it could actually be treated along with case (CII). The scaling factor λ for the weight(s) of a curve segment of case (CI) can be obtained by reducing this case to the 2D case of (Meek et al., 2003) and applying the approach there.

Algorithm Given a finite set of space data points $\{ \boldsymbol{I}_i : 0 \leq i \leq n \}$, and a finite set $\mathcal{H}$ of constraint planes where the polyline joining the data points does not intersect the constraint planes.

1. For $i = 0, 1, \cdots, n-1$,

   construct the rational cubic curve segment $\boldsymbol{R}_i$ of the form (2.1) joining consecutive data $\boldsymbol{I}_i$ and $\boldsymbol{I}_{i+1}$ by using the piecewise rational cubic Bézier G^2 shape preserving interpolation scheme of (Goodman & Ong, 1997) or other G^2 piecewise rational cubic Bézier interpolation scheme where the weights of the rational cubic are positive.

2. For $i = 0, 1, \cdots, n-1$,

   2.1 identify the set Φ_i of all the constraint planes from $\mathcal{H}$ which intersect the interior of the tetrahedron whose vertices are the control points of $\boldsymbol{R}_i$

   2.2 if $\Phi_i \neq \emptyset$, determine the type of modification on $\boldsymbol{R}_i$ according to the classification on the positions of its control point relative to the constraint planes in Φ_i^* into three groups as described in Section 4.1, i.e. group (SI), group (SII) or group (SIII).

3. For $i = 0, 1, \cdots, n-1$,

   initialize $\lambda_A^i = 1$ and $\lambda_D^i = 1$.

4. For $i = 0, 1, \cdots, n-1$,

   4.1 for each constraint plane in Φ_i,

      4.1.1 perform process (a) on $\boldsymbol{R}_i$

      4.1.2 perform process (b) on $\boldsymbol{R}_i$ to each of its boundary line segments

   4.2 if there exists $\lambda \in (0, \infty)$ from step 4.1.1 or step 4.1.2

      let $\overline{\lambda}$ be the maximum among all the λ's obtained

      if $\overline{\lambda} > 1$,

         case of (SI): let $\lambda_i^A = \overline{\lambda}$

case of (SII): let $\lambda_i^D = \overline{\lambda}$

case of (SIII): let $\lambda_i^A = \overline{\lambda}$ and $\lambda_i^D = \overline{\lambda}$

5. For $i = 0, 1, \cdots, n-1$

   let $\lambda_i^A = \max\{\lambda_i^A, \lambda_{i-1}^D\}$ and $\lambda_i^D = \max\{\lambda_{i+1}^A, \lambda_i^D\}$ where $\lambda_{-1}^D = \lambda_n^A = 1$

   if $\lambda_i^A > 1$, scale the weight α of $\boldsymbol{R}_i$ by λ_i^A

   if $\lambda_i^D > 1$, scale the weight β of $\boldsymbol{R}_i$ by λ_i^D

6. If in step 5, there is some $\lambda_i^A > 1$ or some $\lambda_i^D > 1$, then repeat step 3 to step 5 until $\lambda_i^A = 1$ and $\lambda_i^D = 1$, $\forall\ i = 0, 1, \cdots, n-1$.

5. Graphical Examples

We illustrate our scheme with two numerical examples. In all the figures, the data points are represented by the symbol $\circ$ and the inner Bézier control points are denoted by the symbol $+$.

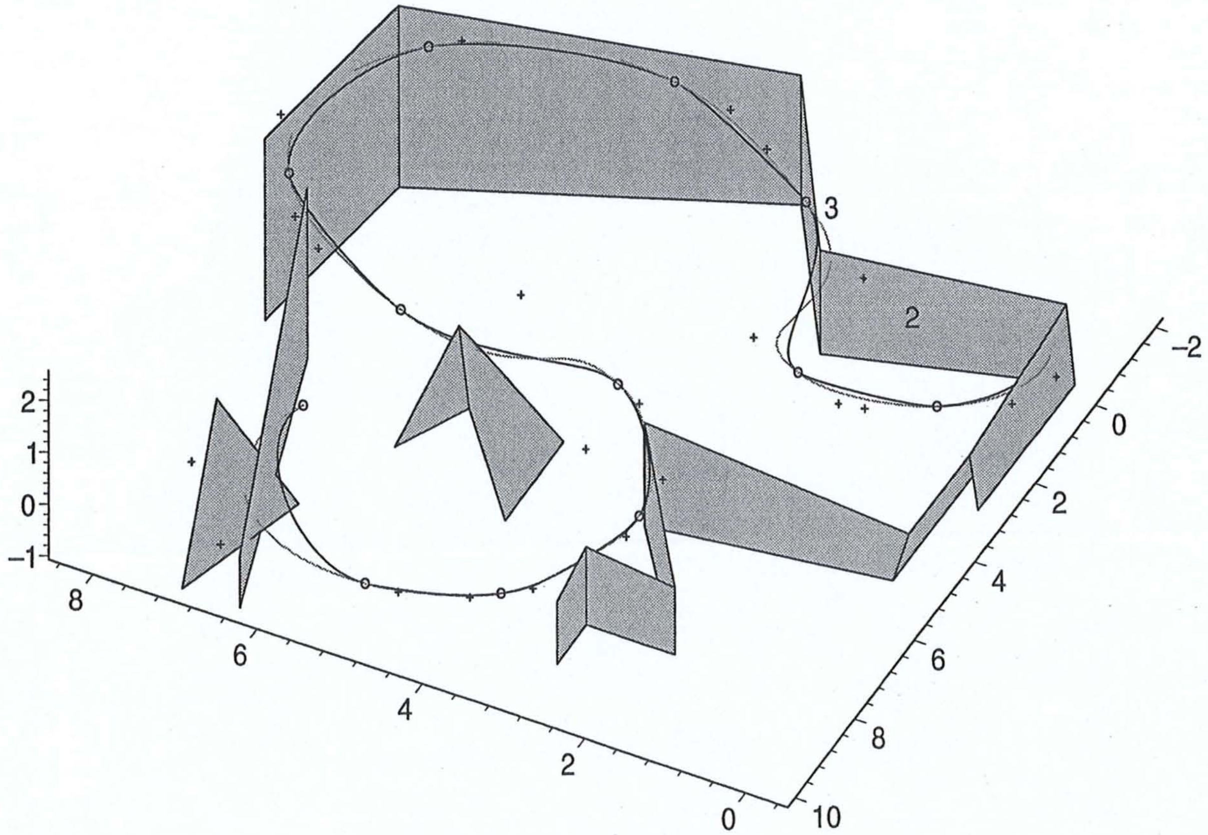


Figure 13. Example 1.

The first example consists of 13 data points with the point (0, 0, 0) as the first point. The open default curve which is generated by the scheme in (Goodman & Ong, 1997) is drawn in grey in Figure 13. There are 14 constraint planes. Curve segments 1, 5, 6 and 7 cross one constraint plane each while curve segments 3, 8, 9 and 12 cross two constraint planes each. All these curve segments are modified by scaling the weights α and β. Two iterations are required. In the first iteration, curve segment 3 crosses the constraint planes 2 and 3 as indicated in Figure 13. Two λ values for scaling α and β are identified, i.e. $\lambda = 3.076397$ is obtained by identifying the rational cubic curve which passes through a point on a boundary of constraint plane 2 while $\lambda = 5.312762$ is obtained from the process of identifying the tangent point on constraint plane 3 to the corresponding curve segment. Since the tangent point lies outside of the finite constraint plane 3, $\lambda = 3.076397$ is selected as the scaling factor for the modification. However with this scaling factor the modified curve segment still crosses constraint plane 3. When its weights are further scaled by $\lambda = 1.726943$, then the resulting curve is tangent to this constraint plane at a tangent point which lies on the plane Thus a second iteration is required to fix the curve segment 3. No other curve segments need further modification at this iteration.

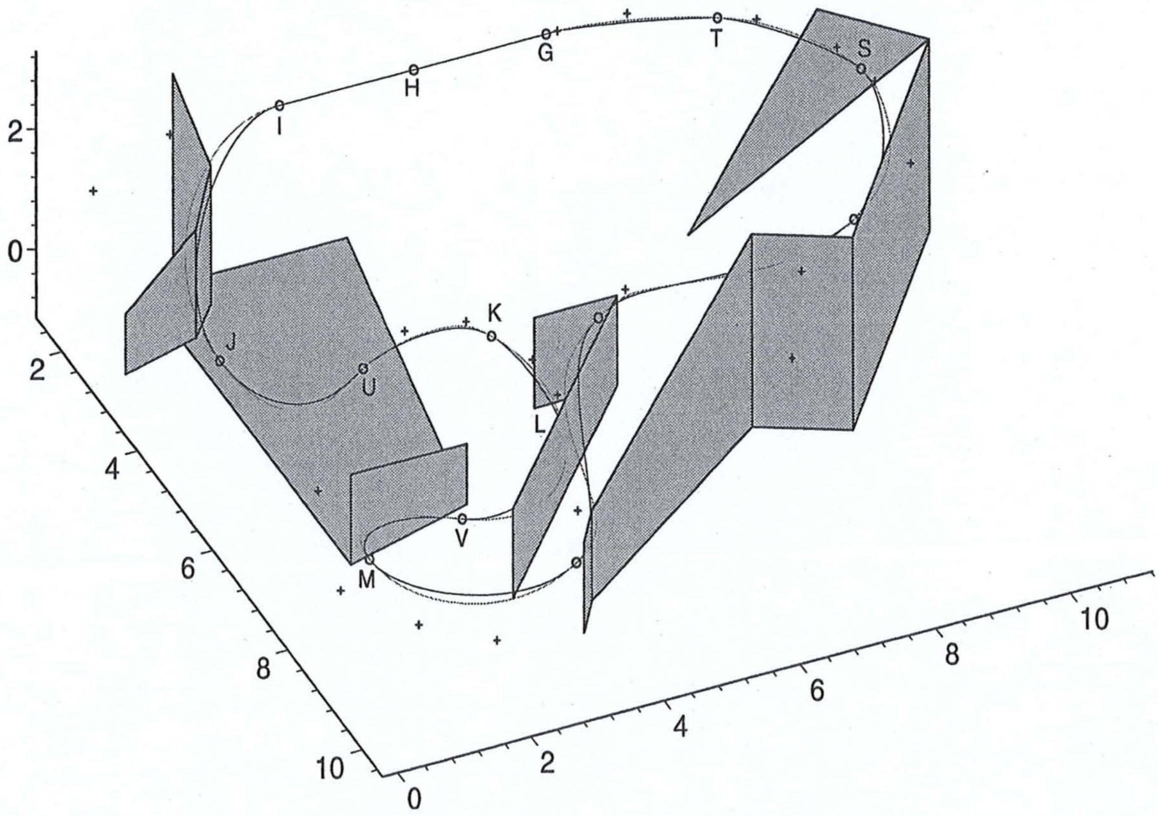


Figure 14. Example 2.

The second example is a closed curve (in Figure 14) interpolating 12 data points with ***S*** and ***T*** as the first and second data points. Data points denoted by ***G***, ***H*** and ***I*** are collinear. To preserve collinearity, the scheme in (Goodman & Ong, 1997) generates a linear line segment through these three points and the interpolating curve is G^1 at these points since the curvature of the curve is not continuous at these points. Data points ***J***, ***K***, ***L*** and ***M*** are coplanar and the scheme in (Goodman & Ong, 1997) which preserves coplanarity generates a planar curve segment between ***K*** and ***L*** while two curve segments each are generated between ***J*** and ***K***, ***L*** and ***M***, with the joint of the two segments at ***U*** and ***V*** respectively, so the total number of segments of the default interpolating curve is 14. Curve segments 1, 6, 8, 9, 10 and 14 cross one constraint plane each while curve segments 12 and 13 cross two constraint planes each. Curve segment 5 which is between ***I*** and ***J*** crosses two constraint planes and goes from the "front" to the "back" of one finite constraint plane without crossing it. These curve segments are modified by scaling both of their weights α and β except curve segment 9 is modified by scaling only the weight α. The modified curve obtained after one iteration does not cross any of the constraint planes.

Acknowledgements

The financial support of the Fundamental Research Grant of Malaysia is gratefully acknowledged.

Appendix A

Find point of intersection of line segment PQ with plane $n_1 x + n_2 y + n_3 z + d = 0$ where P and Q lie on opposite side of the plane.

The normal to the plane is $N = \langle n_1, n_2, n_3 \rangle$.

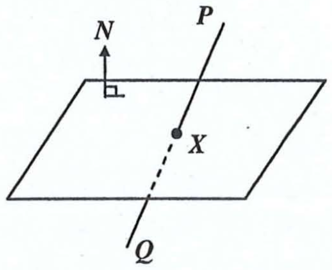
Let X be the point of intersection on PQ, F be the foot of the perpendicular line from P to the plane. Then

$$X = (1-s)P + sQ\ , \quad \text{for some } s \in (0,1)\ , \quad \text{and} \quad N \bullet F + d = 0\ .$$

Substituting X into the equation of the plane, we have

$N \bullet (P + s\,PQ) + d = 0$ and thus

$$s = \frac{-d - N \bullet P}{N \bullet PQ} = \frac{N \bullet PF}{N \bullet PQ} \in (0,1)\ .$$



Appendix B

Suppose that PQ is a boundary line segment of a constraint plane. We would like to find, if any, the value of $t \in (0,1)$ and $\lambda > 0$ such that the point $R(t, \lambda\alpha, \beta)$ is a point of PQ. Denote the point $R(t, \lambda\alpha, \beta) = R = (R_x, R_y, R_z)$

Since $PQ \times PR = \mathbf{0}$, we have

$$(R_x - P_x)(Q_y - P_y) = (R_y - P_y)(Q_x - P_x) \qquad \text{(B.1)}$$

$$(R_x - P_x)(Q_z - P_z) = (R_z - P_z)(Q_x - P_x) \qquad \text{(B.2)}$$

$$(R_y - P_y)(Q_z - P_z) = (R_z - P_z)(Q_y - P_y) \qquad \text{(B.3)}$$

As $$R(t, \lambda\alpha, \beta) = \frac{E}{W} = \frac{\lambda\alpha A(1-t)^3 + 3B\,t(1-t)^2 + 3C\,t^2(1-t) + \beta D\,t^3}{\lambda\alpha(1-t)^3 + 3t(1-t)^2 + 3t^2(1-t)^3 + \beta t^3}$$

we obtain

$$\overline{PQ}_y(E_x - P_xW) = \overline{PQ}_x(E_y - P_yW) \qquad (B.4)$$

$$\overline{PQ}_z(E_x - P_xW) = \overline{PQ}_x(E_z - P_zW) \qquad (B.5)$$

$$\overline{PQ}_z(E_y - P_yW) = \overline{PQ}_y(E_z - P_zW) \qquad (B.6)$$

where the components of a general point or vector K is denoted by $K = (K_x, K_y, K_z)$ and those of the vector ST by $(\overline{ST}_x, \overline{ST}_y, \overline{ST}_z)$.

From (B.4) we get

$$\lambda\alpha(1-t)^3[\overline{PQ}_y\overline{PA}_x - \overline{PQ}_x\overline{PA}_y] = 3t(1-t)^2[\overline{PQ}_x\overline{PB}_y - \overline{PQ}_y\overline{PB}_x] +$$
$$3t^2(1-t)[\overline{PQ}_x\overline{PC}_y - \overline{PQ}_y\overline{PC}_x] + \beta t^3[\overline{PQ}_x\overline{PD}_y - \overline{PQ}_y\overline{PD}_x] \qquad (B.7)$$

From (B.5) we get

$$\lambda\alpha(1-t)^3[\overline{PQ}_z\overline{PA}_x - \overline{PQ}_x\overline{PA}_z] = 3t(1-t)^2[\overline{PQ}_x\overline{PB}_z - \overline{PQ}_z\overline{PB}_x] +$$
$$3t^2(1-t)[\overline{PQ}_x\overline{PC}_z - \overline{PQ}_z\overline{PC}_x] + \beta t^3[\overline{PQ}_x\overline{PD}_z - \overline{PQ}_z\overline{PD}_x] \qquad (B.8)$$

From (B.6) we get

$$\lambda\alpha(1-t)^3[\overline{PQ}_y\overline{PA}_z - \overline{PQ}_z\overline{PA}_y] = 3t(1-t)^2[\overline{PQ}_z\overline{PB}_y - \overline{PQ}_y\overline{PB}_z] +$$
$$3t^2(1-t)[\overline{PQ}_z\overline{PC}_y - \overline{PQ}_y\overline{PC}_z] + \beta t^3[\overline{PQ}_z\overline{PD}_y - \overline{PQ}_y\overline{PD}_z] \qquad (B.9)$$

Denote $PQ \times PK = (\overline{PQK}_x, \overline{PQK}_y, \overline{PQK}_z)$ for $K = A, B, C$ or D.

From (B.7) if $\overline{PQ}_y\overline{PA}_x - \overline{PQ}_x\overline{PA}_y = -\overline{PQA}_z = 0$, then

$$0 = 3t(1-t)^2[\overline{PQB}_z] + 3t^2(1-t)[\overline{PQC}_z] + \beta t^3[\overline{PQD}_z]$$

i.e. $0 = 3b + 3(c-2b)t + (3b - 3c + d\beta)t^2$ (B.10)

where $b = \overline{PQB}_z$, $c = \overline{PQC}_z$, $d = \overline{PQD}_z$.

Similarly

From (B.8) if $\overline{PQ}_z\overline{PA}_x - \overline{PQ}_x\overline{PA}_z = \overline{PQA}_y = 0$, then

$$0 = 3b + 3(c-2b)t + (3b - 3c + d\beta)t^2 \qquad (B.11)$$

where $b = \overline{PQB}_y$, $c = \overline{PQC}_y$, $d = \overline{PQD}_y$.

From (B.9) if $\overline{PQ}_y\overline{PA}_z - \overline{PQ}_z\overline{PA}_y = \overline{PQA}_x = 0$, then

$$0 = 3b + 3(c-2b)t + (3b - 3c + d\beta)t^2 \qquad (B.12)$$

where $b = \overline{PQB}_x$, $c = \overline{PQC}_x$, $d = \overline{PQD}_x$.

If $\overline{PQA}_x \neq 0$, $\overline{PQA}_y \neq 0$ and $\overline{PQA}_z \neq 0$, then from (B.7) and (B.8) we obtain

$$0 = 3b + 3(c-2b)t + (3b - 3c + d\beta)t^2 \qquad (B.13)$$

where

$$b = [\overline{PQB}_z][\overline{PQA}_y] - [\overline{PQB}_y][\overline{PQA}_z]$$
$$c = [\overline{PQC}_z][\overline{PQA}_y] - [\overline{PQC}_y][\overline{PQA}_z]$$
$$d = [\overline{PQD}_z][\overline{PQA}_y] - [\overline{PQD}_y][\overline{PQA}_z].$$

The value of t can be obtained from one of the quadratic equations (B.10)-(B.13) whichever is applicable.

From (B.7) we get

$$\lambda = \frac{3t(1-t)^2[\overline{PQB}_z] + 3t^2(1-t)[\overline{PQC}_z] + \beta t^3[\overline{PQD}_z]}{\alpha(1-t)^3[-\overline{PQA}_z]}$$

From (B.8) we get

$$\lambda = \frac{3t(1-t)^2[\overline{PQB}_y] + 3t^2(1-t)[\overline{PQC}_y] + \beta t^3[\overline{PQD}_y]}{\alpha(1-t)^3[-\overline{PQA}_y]}$$

From (B.9) we get

$$\lambda = \frac{3t(1-t)^2[\overline{PQB}_x] + 3t^2(1-t)[\overline{PQC}_x] + \beta t^3[\overline{PQD}_x]}{\alpha(1-t)^3[-\overline{PQA}_x]}$$

The value of λ can be computed by using any one of the above three equations as long as the denominator is non zero.

If $\overline{PQA}_x = \overline{PQA}_y = \overline{PQA}_z = 0$, then P, Q and A are collinear and so the boundary line PQ contains part of a line segment on the β-surface. Denote the point on PQ which is nearest to A and lies on the β-surface as F. Then from $F = R(t, \lambda\alpha, \beta)$

$$F_x = \frac{\lambda\alpha A_x(1-t)^3 + 3B_x t(1-t)^2 + 3C_x t^2(1-t) + \beta D_x t^3}{\lambda\alpha(1-t)^3 + 3t(1-t)^2 + 3t^2(1-t)^3 + \beta t^3}$$

$$\lambda\alpha(1-t)^3(F_x - A_x) = 3t(1-t)^2(B_x - F_x) + 3t^2(1-t)(C_x - F_x) + \beta t^3(D_x - F_x)$$

so $$\lambda = \frac{3t(1-t)^2(B_x - F_x) + 3t^2(1-t)(C_x - F_x) + \beta t^3(D_x - F_x)}{\alpha(1-t)^3(F_x - A_x)} \quad \text{if } F_x - A_x \neq 0.$$

We may use the y or z component instead of the x component above to find λ as long as the denominator is non-zero .

References

Butt S. & Brodlie K.W., 1993. Preserving positivity using piecewise cubic interpolation. Computer Graphics 17, 55-64.

Randall L. Dougherty, Alan Edelman & James M. Hyman, 1989. Nonnegativity-, monotonicity-, or convexity-preserving cubic and quintic Hermite interpolation. Mathematics of Computation, Vol. 52, No. 186, 471-494.

Goodman T. N. T. & Ong B. H., 1997. Shape preserving interpolation by G^2 curves in three dimensions. In: A. Le Méhaute & L.L. Schumaker (Eds.), Curves and Surfaces with Applications in CAGD. Vanderbilt University Press, Nashville & London, 151-158.

Goodman T. N. T., Ong B. H. & Unsworth K., 1991. Constrained interpolation using rational cubic splines. In: Farin G. (Ed.), Nurbs for Curve and Surface Design. SIAM, Philadelphia, 59-74.

Lahtinen Aatos, 1993. Positive Hermite interpolation by quadratic splines. SIAM J. Math. Anal., Vol 24, No. 1, 223-233.

Meek D. S., Ong B. H. & Walton D. J., 2003. Constrained interpolation with rational cubics. Computer Aided Geometric Design 20, 253-275.

Ong B. H. & K. Unsworth K., 1992. On non-parametric constrained interpolation. In: Tom Lyche & Larry L. Schumaker, (Eds.), Mathematical Methods in Computer Aided Geometric Design II. Academic Press, Inc., 419-430.

Opfer G. & Oberle H. J., 1988. The derivation of cubic splines with obstacles by methods of optimization and optimal control. Numer. Math. 52, 17-31.

Press W. H., Teukolsky, S. A., Vetterling W. T. & Flannery B. P., 2002. Numerical Recipes in C++, 2nd ed.. Cambridge University Press, Cambridge.

Schmidt J.W. & Hess W., 1988. Positivity of cubic polynomials on intervals and positive spline interpolation. Bit 28, 340-362.

Wever U., 1988. Non-negative exponential splines. Computer Aided Design, Vol. 20, No.1, 11-16.

Zhang Caiming & Cheng Fuhua, 2001. Constrained shape-preserving curve interpolation with minimum curvature. Preprint.