A Relationship Between Minimum Bending Energy and Degree Elevation for Bézier Curves

David Eberly
Geometric Tools, LLC
http://www.geometrictools.com/

Created: October 24, 2010

Contents

1 Introduction

A Bézier curve of degree d is of the form

$$\mathbf{X}(t) = \sum_{i=0}^{d} B_i(t)\mathbf{P}_i = \sum_{i=0}^{d} c_{d,i}(1-t)^{d-i}t^i\mathbf{P}_i \tag{1}$$

where $\mathbf{P}_i$ are control points and $B_i(t)$ are the Bernstein polynomials with $c_{d,i} = d!/[i!(d-i)!]$ and $t \in [0,1]$. Let us assume that minimally $\mathbf{P}_0$ and $\mathbf{P}_d$ are specified; that is, the endpoints of the curve are known quantities. The other control points may be selected as necessary, depending on the situation.

The motivation for this document is that conditions are imposed whereby some, but not all, of the control points are determined. The unknown control points are selected to minimize the bending energy,

$$E(P) = \int_0^1 \left|\mathbf{X}''(t)\right|^2 \, dt \tag{2}$$

where P represents the collection of $d+1$ control points. The indices of the n known control points are $\{k_0, \ldots, k_{n-1}\}$ and the indices of the m unknown control points are $\{u_0, \ldots, u_{m-1}\}$ with the union of the two sets being $\{0, \ldots, d\}$; necessarily $n + m = d + 1$.

This type of problem arose in a surface-fitting application, where samples of the form $(x, y, f(x, y))$ were specified, the (x, y) pairs were connected using a Delaunay triangulation, and a globally C^2 graph with local control had to be constructed using degree 9 Bézier triangle patches. Some of the control points of the patches were determined by the samples, but the remaining control points had to be chosen one way or another. The choice was made to minimize bending energy of the final graph. This document shows how the minimization of bending energy is related to degree elevation, but for curves. The extension to surface patches is discussed in a separate document.

2 Degree Elevation

The Bézier curve of degree d is specified in Equation (1). The same curve may be represented as a Bézier curve of degree $d+1$. The idea is that $1 = (1-t) + t$ for all t, so the degree d curve can be multiplied by 1 using this identity. For example, consider the linear curve

$$\mathbf{X}(t) = (1-t)\mathbf{P}_0 + t\mathbf{P}_1 \tag{3}$$

Multiplying by $(1-t) + t$ leads to

$$\begin{aligned} \mathbf{X}(t) &= [(1-t)+t]\,[(1-t)\mathbf{P}_0 + t\mathbf{P}_1] \\ &= (1-t)^2\mathbf{P}_0 + 2(1-t)t\left(\tfrac{\mathbf{P}_0+\mathbf{P}_1}{2}\right) + t^2\mathbf{P}_1 \\ &= (1-t)^2\mathbf{Q}_0 + 2(1-t)t\mathbf{Q}_1 + t^2\mathbf{Q}_2 \end{aligned} \tag{4}$$

where $\mathbf{Q}_0 = \mathbf{P}_0$, $\mathbf{Q}_2 = \mathbf{P}_1$, and $\mathbf{Q}_1 = (\mathbf{Q}_0 + \mathbf{Q}_2)/2$. Equation (4) is formally a quadratic Bézier curve, but it still represents the line segment defined by Equation (3).

Generally, the degree-elevated curve is

$$\mathbf{X}(t) = \sum_{i=0}^{d+1} c_{d+1,i}(1-t)^{d+1-i}t^i\mathbf{Q}_i \tag{5}$$

where

$$\mathbf{Q}_i = \frac{i}{d+1}\mathbf{P}_{i-1} + \left(1 - \frac{i}{d+1}\right)\mathbf{P}_i \tag{6}$$

for $0 \le i \le d+1$ and with the convention that $\mathbf{P}_{-1} = \mathbf{0}$ and $\mathbf{P}_{d+1} = \mathbf{0}$.

3 Minimizing Bending Energy

To illustrate, consider a quadratic Bézier curve

$$\mathbf{X}(t) = (1-t)^2\mathbf{P}_0 + 2(1-t)t\mathbf{P}_1 + t^2\mathbf{P}_2$$

where $\mathbf{P}_0$ and $\mathbf{P}_2$ are known but $\mathbf{P}_1$ must be chosen to minimize the bending energy. The first derivative of the curve is

$$\mathbf{X}'(t) = -2(1-t)\mathbf{P}_0 + 2(1-2t)\mathbf{P}_1 + 2t\mathbf{P}_2$$

and the second derivative of the curve is

$$\mathbf{X}''(t) = 2\left(\mathbf{P}_0 - 2\mathbf{P}_1 + \mathbf{P}_2\right)$$

The bending energy is

$$E = \int_0^1 \left|\mathbf{X}''(t)\right|^2\, dt = 4\left|\mathbf{P}_0 - 2\mathbf{P}_1 + \mathbf{P}_2\right|^2$$

The minimum is obtained when

$$\mathbf{P}_0 - 2\mathbf{P}_1 + \mathbf{P}_2 = \mathbf{0}$$

in which case

$$\mathbf{P}_1 = \frac{\mathbf{P}_0 + \mathbf{P}_2}{2}$$

Observe that this condition shows that the curve that minimizes the bending energy is a line segment, which is consistent with your intuition; see Equation (4).

Now consider a cubic Bézier curve

$$\mathbf{X}(t) = (1-t)^3\mathbf{P}_0 + 3(1-t)^2t\mathbf{P}_1 + 3(1-t)t^2\mathbf{P}_2 + t^3\mathbf{P}_3$$

where $\mathbf{P}_0$ and $\mathbf{P}_3$ are known but $\mathbf{P}_1$ and $\mathbf{P}_2$ must be chosen to minimize the bending energy. The first derivative of the curve is

$$\mathbf{X}'(t) = -3(1-t)^2\mathbf{P}_0 + 3(1-t)(1-3t)\mathbf{P}_1 + 3t(2-3t)\mathbf{P}_2 + 3t^2\mathbf{P}_3$$

and the second derivative of the curve is

$$\mathbf{X}''(t) = 6\left((1-t)\mathbf{P}_0 + (3t-2)\mathbf{P}_1 + (1-3t)\mathbf{P}_2 + t\mathbf{P}_3\right)$$

Define P to be the 4×1 column vector whose ith row is formally $\mathbf{P}_i$; then

$$
\begin{aligned}
\left|\mathbf{X}''(t)\right|^2 &= 36 \begin{bmatrix} \mathbf{P}_0^{\mathrm{T}} & \mathbf{P}_1^{\mathrm{T}} & \mathbf{P}_2^{\mathrm{T}} & \mathbf{P}_3^{\mathrm{T}} \end{bmatrix} \begin{bmatrix} 1-t \\ 3t-2 \\ 1-3t \\ t \end{bmatrix} \begin{bmatrix} 1-t & 3t-2 & 1-3t & t \end{bmatrix} \begin{bmatrix} \mathbf{P}_0 \\ \mathbf{P}_1 \\ \mathbf{P}_2 \\ \mathbf{P}_3 \end{bmatrix} \\
&= 36\, P^{\mathrm{T}} \begin{bmatrix} (1-t)^2 & (1-t)(3t-2) & (1-t)(1-3t) & (1-t)t \\ (1-t)(3t-2) & (3t-2)^2 & (3t-2)(1-3t) & (3t-2)t \\ (1-t)(1-3t) & (3t-2)(1-3t) & (1-3t)^2 & (1-3t)t \\ (1-t)t & (3t-2)t & (1-3t)t & t^2 \end{bmatrix} P
\end{aligned}
$$

The bending energy is obtained by integrating the matrix terms,

$$
E = \int_0^1 \left|\mathbf{X}''(t)\right|^2 \, dt = 6\, P^{\mathrm{T}} \begin{bmatrix} 2 & -3 & 0 & 1 \\ -3 & 6 & -3 & 0 \\ 0 & -3 & 6 & -3 \\ 1 & 0 & -3 & 2 \end{bmatrix} P = 6\, P^{\mathrm{T}} M P
$$

where the last equality defines the 4×4 symmetric matrix M.

The matrix M has eigenvalues 10, 6, and 0 (multiplicity 2). The eigenspace for eigenvalue 10 is 1-dimensional and is spanned by the unit-length eigenvector $\mathbf{V}_0 = (1, -3, 3, -1)/\sqrt{20}$. The eigenspace for eigenvalue 6 is 1-dimensional and is spanned by the unit-length eigenvector $\mathbf{V}_1 = (1, -1, -1, 1)/2$. The eigenspace for eigenvalue 0 is 2-dimensional and is spanned by the unit-length and orthogonal eigenvectors $\mathbf{V}_2 = (1, 0, -1, -2)/\sqrt{6}$ and $\mathbf{V}_3 = (4, 3, 2, 1)/\sqrt{30}$. The set $\{\mathbf{V}_0, \mathbf{V}_1, \mathbf{V}_2, \mathbf{V}_3\}$ is right-handed and orthonormal, so the eigenvectors may be written as the columns of a rotation matrix

$$
R = \begin{bmatrix} \mathbf{V}_0 & \mathbf{V}_1 & \mathbf{V}_2 & \mathbf{V}_3 \end{bmatrix}
$$

The symmetric matrix factors to $M = RDR^{\mathrm{T}}$, where $D = \mathrm{Diagonal}(10, 6, 0, 0)$.

Define

$$
\begin{bmatrix} \mathbf{Q}_0 \\ \mathbf{Q}_1 \\ \mathbf{Q}_2 \\ \mathbf{Q}_3 \end{bmatrix} = Q = R^{\mathrm{T}} P = R^{\mathrm{T}} \begin{bmatrix} \mathbf{P}_0 \\ \mathbf{P}_1 \\ \mathbf{P}_2 \\ \mathbf{P}_3 \end{bmatrix}
$$

so that the bending energy is

$$
E = 6\, P^{\mathrm{T}} M P = 6\, P^{\mathrm{T}} R D R^{\mathrm{T}} P = 6\, Q^{\mathrm{T}} D Q = 60|\mathbf{Q}_0|^2 + 36|\mathbf{Q}_1|^2
$$

The minimum of E occurs when $\mathbf{Q}_0 = \mathbf{0}$ and $\mathbf{Q}_1 = \mathbf{0}$. Some algebra will show that

$$\mathbf{P}_0 - 3\mathbf{P}_1 + 3\mathbf{P}_2 - \mathbf{P}_3 = \mathbf{0}$$
$$\mathbf{P}_0 - \mathbf{P}_1 - \mathbf{P}_2 + \mathbf{P}_3 = \mathbf{0}$$

Notice that the coefficients of the control points in the first equation are the components of $\mathbf{V}_0$ and the coefficient of the control points in the second equation are the components of $\mathbf{V}_1$. In a loose mathematical description, this says that the formal vector P is orthogonal to the eigenvectors of M corresponding to the eigenvalues 10 and 6, and in this sense P formally appears to be in the eigenspace for eigenvalue 0.

The control points $\mathbf{P}_1$ and $\mathbf{P}_2$ are the unknowns, so solving the equations leads to

$$\mathbf{P}_1 = \frac{2\mathbf{P}_0 + \mathbf{P}_3}{3}, \quad \mathbf{P}_2 = \frac{\mathbf{P}_0 + 2\mathbf{P}_3}{3}$$

A quick check will show you that for this choice of control points, the cubic Bézier curve is simply the degree-elevated linear Bézier curve.

4 The General Problem

In the example of a cubic Bézier curve, we specified the control points $\mathbf{P}_0$ and $\mathbf{P}_3$ and showed that in minimizing the bending energy, we had two linear equations in the two unknowns $\mathbf{P}_1$ and $\mathbf{P}_2$. As expected, the minimum bending energy occurs when the control points are collinear.

Consider a modified problem where we specify $\mathbf{P}_0$, $\mathbf{P}_1$, and $\mathbf{P}_3$. The only unknown is $\mathbf{P}_2$. This problem amounts to specifying the positions of the endpoints of the curve and the derivative at the first endpoint. When $\mathbf{P}_1$ is *not* chosen to be on the line segment connecting the endpoints, the bending energy E cannot be zero (because you have to bend the curve away from a line segment). Setting some $\mathbf{Q}_i = \mathbf{0}$ will not work here as it did when specifying only the endpoints.

Instead, let's permute the rows of P so that the unknown control points occur first. The matrix M must be adjusted accordingly. Define the permutation matrix J so that

$$\hat{P} = \begin{bmatrix} \mathbf{P}_2 \\ \mathbf{P}_0 \\ \mathbf{P}_1 \\ \mathbf{P}_3 \end{bmatrix} = \begin{bmatrix} 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} \mathbf{P}_0 \\ \mathbf{P}_1 \\ \mathbf{P}_2 \\ \mathbf{P}_3 \end{bmatrix} = JP$$

Define $\hat{M} = JMJ^{\mathrm{T}}$. The bending energy is then

$$\frac{E}{6} = \hat{P}^{\mathrm{T}} \hat{M} \hat{P} = \hat{P}^{\mathrm{T}} \left[\begin{array}{c|c} A & B \\ \hline B^{\mathrm{T}} & C \end{array} \right] \hat{P} \tag{7}$$

where A is the 1×1 upper-left entry of $\hat{M}$, C is the 3×3 lower-right block of $\hat{M}$, and B is the 1×3 upper-right block of $\hat{M}$. Define U be the first row of $\hat{P}$ (the unknown) and define K to be the last three rows of $\hat{P}$ (the knowns), so U is 1×1 and K is 3×1. Equation (7) becomes

$$\frac{E}{6} = U^{\mathrm{T}} A U + 2 K^{\mathrm{T}} B^{\mathrm{T}} U + K^{\mathrm{T}} C K \tag{8}$$

The binding energy is nonnegative, which implies E as a function of U is a paraboloid and has a global minimum that occurs when its gradient is zero,

$$AU + BK = 0 \tag{9}$$

Solve the linear system for U. When A is invertible, it is clear what the solution is. If A is not invertible, there must be infinitely many solutions to the equation–the graph of E must be a parabolic cylinder.

In the current example,

$$\hat{M} = \left[\begin{array}{r|rrr} 6 & 0 & -3 & -3 \\ \hline 0 & 2 & -3 & 1 \\ -3 & -3 & 6 & 0 \\ -3 & 1 & 0 & 2 \end{array}\right]$$

so Equation (9) is

$$\mathbf{0} = 6\mathbf{P}_2 + (0\mathbf{P}_0 - 3\mathbf{P}_1 - 3\mathbf{P}_3)$$

which has solution $\mathbf{P}_2 = (\mathbf{P}_1 + \mathbf{P}_3)/2$. Notice that $\mathbf{X}'(1) = 3(\mathbf{P}_1 - \mathbf{P}_3)/2$. The intuition is that the choice of $\mathbf{P}_2$ should attempt to "straighten out" as much as possible the curve near $t = 1$. In this case, $\mathbf{P}_2$ is on the line segment connecting $\mathbf{P}_1$ and $\mathbf{P}_3$.

This approach works even for the previous example with two knowns and two unknowns. The permutation matrix is

$$J = \begin{bmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

which leads to $\hat{P} = (\mathbf{P}_1, \mathbf{P}_2, \mathbf{P}_0, \mathbf{P}_3)$, $U = (\mathbf{P}_1, \mathbf{P}_2)$, $K = (\mathbf{P}_0, \mathbf{P}_1)$, and

$$\hat{M} = \left[\begin{array}{rr|rr} 6 & -3 & -3 & 0 \\ -3 & 6 & 0 & -3 \\ \hline -3 & 0 & 2 & 1 \\ 0 & -3 & 1 & 2 \end{array}\right]$$

Equation (9) is

$$\begin{bmatrix} \mathbf{0} \\ \mathbf{0} \end{bmatrix} = \begin{bmatrix} 6 & -3 \\ -3 & 6 \end{bmatrix} \begin{bmatrix} \mathbf{P}_1 \\ \mathbf{P}_2 \end{bmatrix} + \begin{bmatrix} -3 & 0 \\ 0 & -3 \end{bmatrix} \begin{bmatrix} \mathbf{P}_0 \\ \mathbf{P}_3 \end{bmatrix}$$

Inverting the 2×2 matrix A leads to the solution $\mathbf{P}_1 = (2\mathbf{P}_0 + \mathbf{P}_3)/3$ and $\mathbf{P}_2 = (\mathbf{P}_0 + 2\mathbf{P}_3)/3$.

5 Computing Matrix M

The first derivative of the Bézier curve of Equation (1) is

$$\mathbf{X}'(t) = \sum_{i=0}^{d} c_{d,i} \left[(1-t)^{d-i} i t^{i-1} - (d-i)(1-t)^{d-1-i} t^{i}\right] \mathbf{P}_i \tag{10}$$

and the second derivative is

$$\begin{aligned} \mathbf{X}''(t) &= \textstyle\sum_{i=0}^{d} c_{d,i} \left[(1-t)^{d-i} i(i-1) t^{i-2} - 2(d-i)(1-t)^{d-1-i} i t^{i-1} + (d-i)(d-1-i)(1-t)^{d-2-i} t^{i}\right] \mathbf{P}_i \\ &= \textstyle\sum_{i=0}^{d} f_i(t) \mathbf{P}_i \end{aligned} \tag{11}$$

where the last equality defines the functions $f_i(t)$. The entry m_{ij} of the matrix M is

$$m_{ij} = \int_0^1 f_i(t) f_j(t)\, dt$$

The integrand consists of a summation involving 9 polynomial terms of the form $t^p(1-t)^n$ for nonnegative integer powers p and n. We need to integrate these. The integrals are values of the *Beta function*,

$$B(p+1, n+1) = \int_0^1 t^p (1-t)^n \, dt = \frac{p!\, n!}{(p+n+1)!}$$

whose formula is well known in statistics. The coefficients $c_{d,i} = d!/[(d-i)!i!]$, so the integration appears to involve symbolic manipulation of a lot of factorials. However, you can set up a two-variable recursion,

$$B(p+1, n+1) = \frac{p}{p+n+1} \frac{(p-1)!\, n!}{((p-1)+n+1)!} = \frac{p}{p+n+1} B(p, n+1)$$

and

$$B(p+1, n+1) = \frac{n}{p+n+1} \frac{p!\, (n-1)!}{(p+(n-1)+1)!} = \frac{n}{p+n+1} B(p+1, n)$$