Variational Subdivision for
Natural Cubic Splines

Joe Warren and Henrik Weimer

**Abstract.** This paper explores the intrinsic link between natural cubic splines and subdivision. Natural cubic splines are defined via the variational problem of minimizing a simple approximation of bending energy. A subdivision scheme is derived which converges to the minimizer of this particular variational problem.

§1. Introduction

Geometric design is the study of the representation of shapes with mathematical models. Today, curved shapes are most commonly described using a parametric representation. These are based on the weighting of some number of control points with appropriate parametric basis functions. The most successful family of such basis functions is without any doubt the B-spline basis [7,5].

More recently subdivision has evolved as a novel approach for representing shape [2,3]. In this framework smooth shapes are represented as the limit of a repeated weighted averaging process of control points.

This paper exposes the intrinsic link between the two concepts: Starting with the variational definition of natural cubic spline curves, a representation by means of a subdivision process is derived. The principles underlying this derivation are applicable to other variational problems in perfect analogy. Thus, the method presented in this paper can be used to derive subdivision schemes which produce the minimizers of variational problems.

1.1. Natural Cubic Splines

Historically, a spline was a thin, flexible piece of wood used in drafting. The draftsman attached the spline to a sequence of anchor points on a drafting table. The spline was then allowed to slide through the anchor points and assume a smooth, minimum energy shape.
In the 1940’s and 50’s, mathematicians realized that the behavior of a spline could be modeled mathematically. Let the shape of the spline be modeled by the graph of a function \( F(t) \) over a domain interval \([a, b]\). The bending energy at parameter value \( t \in [a, b] \) is roughly proportional to the value of the second derivative of \( F \) with respect to \( t \). The total energy associated with the function \( F \) on the interval \([a, b]\) is thus approximately the integral of the square of the second derivative of \( F \)

\[
\mathcal{E}[F] = \int_a^b F_{{ttt}}(t)^2 \, dt. \tag{1}
\]

The effects of the anchor points on the spline are modeled by constraining \( F \) to satisfy the additional interpolation conditions \( F(t_i) = v_i \). If the vector \( T_0 \) denotes the parameter values for the interpolation conditions \( \{t_0, t_1, \ldots, t_n\} \) and \( c_0 \) denotes the vector of the interpolation values \( \{v_0, v_1, \ldots, v_n\} \), then these conditions can be stated more concisely in vector notation as

\[
F(T_0) = c_0. \tag{2}
\]

Functions that minimize (1) and satisfy (2) are called natural cubic splines.

1.2. Associated Nested Spline Spaces

As the interpolation values \( c_0 \) vary, the minimizing solutions \( F(t) \) form a linear spline space \( V(T_0) \) for a fixed knot sequence \( T_0 \subset [a, b] \). Natural cubic splines satisfy the simple differential equation

\[
F_{{tttt}}(t) = 0 \quad \text{for all} \quad t \in [a, b] \tag{3}
\]

which can be derived directly from (1) using the Euler-Lagrange equations. Therefore, the possible solutions are exactly the cubic polynomials in \( t \).

The Euler-Lagrange condition (3) does not take into account the interpolation conditions (2) on \( F \). If \( F(t) \) is a natural cubic spline which satisfies the interpolation conditions (2) then \( F \) is actually a piecewise cubic polynomial function. The breaks between the polynomial pieces occur at the parameter values for the interpolation conditions, \( t_i \). The polynomial pieces satisfy the differential equation (3) between two adjacent knots \( t_i \) and \( t_{i+1} \) but the differential equation may not hold at the actual knots.

**Theorem 1.** Let \( V(T_0) \) and \( V(T_1) \) be the spline spaces of minimizers of (1) with respect to the interpolation condition (2) over \([a, b]\) corresponding to knot sequences \( T_0 \) and \( T_1 \) respectively. Then, \( T_0 \subset T_1 \Rightarrow V(T_0) \subset V(T_1) \).

**Proof:** Given any \( F(t) \in V(T_0) \) consider \( \hat{F}(t) \in V(T_1) \) with \( \hat{F}(T_1) = F(T_1) \). Assuming \( \mathcal{E}[F] > \mathcal{E}[\hat{F}] \) yields a contradiction to \( \mathcal{E}[F] \) being minimal because \( \hat{F}(T_0) = c_0 \) and \( \hat{F} \) has smaller energy, i.e. \( F \notin V(T_0) \). Conversely, assuming \( \mathcal{E}[F] < \mathcal{E}[\hat{F}] \) contradicts \( \mathcal{E}[\hat{F}] \) being minimal because \( F \) itself interpolates \( F(T_1) \), i.e. \( F \notin V(T_1) \). Therefore, \( \mathcal{E}[F] = \mathcal{E}[\hat{F}] \). \( \square \)
Theorem 2. Let $N_0(t)$ and $N_1(t)$ be vectors of basis functions for the spline spaces $V(T_0)$ and $V(T_1)$ respectively. Then $V(T_0) \subset V(T_1)$ implies $N_0(t) = N_1(t)S_0$ for some matrix $S_0$, called subdivision matrix.

Proof: $N_0$ is a basis for $V(T_0)$. As $V(T_0) \subset V(T_1)$ we also have $N_0 \in V(T_1)$. As $N_1$ is basis for $V(T_1)$ we can represent $N_0$ in terms of this basis. A column of $S_0$ contains the coefficients for basis functions in $N_1$ to represent one particular function in $N_0$. □

Any function $F(t) \in V(T_0)$ can be represented in terms of the basis $N_0(t)$, $F(t) = N_0(t)p_0$ for some vector of coefficients $p_0$. Due to the nesting of the spline spaces $V(T_0)$ and $V(T_1)$ $F(t)$ can also be expressed in terms the basis $N_1(t)$ for $V(T_1)$. $F(t) = N_1(t)p_1$. Because $N_0(t) = N_1(t)S_0$ is is possible to express the coefficients $p_1$ of $F(t)$ in terms of the coefficients $p_0$ as

$$p_1 = S_0p_0.$$  (4)

Theorem 3. If $T_0, T_1, T_2, \cdots$ defines an infinite sequence of finer and finer knot sequences which grows dense in $\mathbb{R}$ and $p_{k+1} = S_k p_k$ then the piecewise linear functions with values $p_k$ over the knots $T_k$ uniformly converge to $F(t)$ as $k \to \infty$.

Proof: See [6].

The B-Spline basis [1,4,5] is the only basis for natural cubic splines with local support. It has been studied to a great extent and was successful in many real world applications.

§2. Finite Element Solution

This section presents a derivation of natural cubic splines in terms of a finite element process starting from the specification of the variational problem (1). In particular we will derive B-splines over the domain $[0,4]$ with initial knots at $T_0 = \{0,1,2,3,4\}$. The derivations presented here are general and can be used in the irregular case in perfect analogy. However, we chose to address the uniform, bounded case in this expository discussion for the sake of clearness and simplicity.

Solving a variational problem with finite elements involves three major steps:

- Choose a set of finite element basis functions $B_k(t) = \{b_1^k, b_2^k, \cdots\}$ and define a continuous version of the level $k$ solution to $F_k(t) = B_k(t)p_k$ in vector notation where $p_k$ denotes the set of coefficients of level $k$.
- Measure the energy of $F_k(t)$ via $E[F_k] = p_k^T E_k p_k$.
- Minimize $p_k^T E_k p_k$ by solving $E_k p_k = 0$.

2.1. Finite Element Basis Functions

Due to the structure of the variational problem (1) the finite element basis functions have to lie in the Sobolev space $H_2([0,4])$, i.e. they have to be piecewise quadratic functions with non-zero, square integrable derivatives up to order two.
For reasons of computational stability and simplicity the basis functions should be centered over the knots, e.g. for level 0 the i-th basis function should have maximum magnitude over the knot \( t_i \).

A possible choice for these basis functions are quadratic piecewise Bézier curves. The Bézier control coefficients for the finite element basis \( B_0(t) \) for level 0 are shown in figure 1 together with the resulting basis functions for level 0 and 1. Basis functions for finer grids are derived similarly using a refined knot vector.

![Bézier control coefficients](image)

**Fig. 1.** Bézier control coefficients for the finite element basis functions of level 0 and finite element basis of level 0 and 1 over the domain [0, 4]. The basis functions are piecewise quadratic and centered over the knots.

These finite element basis functions can be used to define a continuous solution \( F_k(t) \) for level \( k \) by weighting the coefficients \( p_k \) of the level \( k \) solution with the basis functions of level \( k \), i.e.

\[
F_k(t) = B_k(t)p_k.
\]

### 2.2. Energy Matrix and Inner Product

Using the continuous representation (5) of the level \( k \) solution of the finite element process, the energy of \( p_k \) can be assessed as \( \mathcal{E}[F_k] \). This can be expressed as a quadratic form

\[
\mathcal{E}[F_k] = p_k^T \mathbf{E}_k p_k
\]

where \( \mathbf{E}_k \) is a symmetric, positive definite matrix called the energy matrix. Defining an inner product

\[
< f, g > = \int_a^b f(t)g(t)dt
\]

we get \( \mathbf{E}_k = (e_{ij}^k)_{n \times n} \) where \( e_{ij}^k = < b_i^k(t), b_j^k(t) > \). For the level 0 grid \( T_0 \) with knots at the integers over the domain \([0, 4]\),

\[
\mathbf{E}_0 = \begin{pmatrix}
1 & -2 & 1 & 0 & 0 \\
-2 & 5 & -4 & 1 & 0 \\
1 & -4 & 6 & -4 & 1 \\
0 & 1 & -4 & 5 & -2 \\
0 & 0 & 1 & -2 & 1 \\
\end{pmatrix}.
\]
Energy matrices $E_k$ for $k > 0$ have a similar structure. In particular, interior rows are scaled shifts of the center row of $E_0$.

### 2.3. Minimization with Interpolated Values

For $\mathcal{E}[F_k]$ to be minimal the control points $p_k$ have to minimize $p_k^T E_k p_k$. This in turn implies that the derivative of this expression, $2E_k p_k$, is zero.

However, $p_k$ is not completely unknown. Some of the entries in $p_k$ correspond to interpolation conditions $c_0$. Thus, $p_k$ can be partitioned into a known part $p_k^n = c_0$ and an unknown part $p_k^u$ as

$$p_k = \begin{pmatrix} p_k^n \\ p_k^u \end{pmatrix}.$$  

Using this decomposition of $p_k$, equation (6) rewrites to

$$\mathcal{E}[F_k] = p_k^T E_k p_k = \begin{pmatrix} p_k^n & p_k^u \end{pmatrix} \begin{pmatrix} E_k^{nn} & E_k^{nu} \\ E_k^{un} & E_k^{uu} \end{pmatrix} \begin{pmatrix} p_k^n \\ p_k^u \end{pmatrix} = p_k^n^T E_k^{nn} p_k^n + 2p_k^u^T E_k^{nu} p_k^n + p_k^u^T E_k^{uu} p_k^u,$$

$E_k^{nn}$ contains the entries of $E_k$ which have row and column indices corresponding to known grid values, $E_k^{nn}$ contains entries of $E_k$ which have row indices of unknown and column indices of known coefficients and so on.

Now, $p_k^n^T E_k^{nn} p_k^n$ is constant, $2p_k^u^T E_k^{nu} p_k^n$ is linear in the unknowns $p_k^u$ and $p_k^u^T E_k^{uu} p_k^u$ is a quadratic form in the unknowns. Therefore, the derivative of (6) with respect to the unknowns $p_k^u$ can be expressed as $2F_k^{nu} p_k^n + 2F_k^{uu} p_k^n$ and the minimizer of (6) is the solution $p_k^u$ to

$$E_k^{nu} p_k^n + E_k^{uu} p_k^n = 0.$$  

Using block matrix notation (9) can be rewritten as

$$\begin{pmatrix} E_k^{nu} & E_k^{uu} \\ \end{pmatrix} \begin{pmatrix} p_k^n \\ p_k^u \end{pmatrix} = 0.$$  

Together with the interpolation constraints we get

$$\begin{pmatrix} E_k^{nu} & E_k^{uu} \\ 0 & I \end{pmatrix} \begin{pmatrix} p_k^n \\ p_k^u \end{pmatrix} = \begin{pmatrix} 0 \\ c_0 \end{pmatrix}$$  

where $I$ is the identity matrix.

The solution $(p_k^n, p_k^u)$ converges to the solution of the original variational problem as $k \to \infty$, see [6]. Furthermore, the original interpolation conditions $c_0$ are satisfied because $p_k$ contains these values explicitly as $p_k^n$. 
§3. Minimization with Interpolated Differences

The second row of the system (10) enforces $p^u_k$ to interpolate the $c_0$. Therefore, all solutions to (10) interpolate the $c_c$. Unfortunately this leads to undesirable properties of the solution $p_k$ to this system. In particular, changing just one of the entries in $c_0$ results in a global change in the solution. This section presents a new scheme which leads to solutions with much nicer properties.

To this effect, the interpolation constraints are replaced by constraints on the energy of the solution. Equation (10) can be rephrased to

$$
\begin{pmatrix}
E^{uu}_k & E^{un}_k \\
E^{nu}_k & E^{nn}_k
\end{pmatrix}
\begin{pmatrix}
p^u_k \\
p^n_k
\end{pmatrix}
= \begin{pmatrix}
0 \\
E_{cP_0}
\end{pmatrix},
$$

(11)

where $p_c$ is some initial set of control points centered over the initial knots $T_0$. Note that (11) forces the $p_k$ to be chosen such that the energy values over the knots $T_0$ are the same as the original energy values at these knots, i.e. $E_{cP_0}$. All remaining entries in $E_k p_k$ must be zero.

Condition (11) can be expressed more concisely using the notion of an upsampling matrix $U_{k-1}$ which replicates coefficients associated with knots in $T_{k-1}$ to the next finer grid $T_k$ and forces zero coefficients at the knots $T_k - T_{k-1}$. Now (11) can be restated as

$$
E_k p_k = U_{k-1} U_{k-2} \cdots U_0 E_{cP_0}.
$$

(12)

Again, equation (12) states that the $p_k$ are chosen such that $E_k p_k$ reproduces the energy $E_{cP_0}$ over the original knots $T_0$ and forces zero energy at all knots in $T_k - T_c$.

Two such conditions (12) for levels $k$ and $k + 1$ can be assembled as

$$
E_k p_k = U_{k-1} \cdots U_0 E_{cP_0}
$$

$$
E_{k+1} p_{k+1} = U_k \cdots U_c E_{cP_0}.
$$

Combining them yields $E_{k+1} p_{k+1} = U_k E_k p_k$. Finally, applying the definition of the subdivision matrix, $p_{k+1} = S_k p_k$, yields

$$
E_{k+1} S_k = U_k E_k.
$$

(13)

Note that equation (13) allows us to express the solution to the variational problem (1) in a subdivision scheme defined by the subdivision matrices $S_k$. Because the columns of $E_{k+1}$ are not linearly independent this equation does not uniquely determine $S_k$. However, enforcing sparsity of $S_k$ yields a unique solution, the subdivision matrices for uniform natural cubic splines.

The action of the subdivision matrices $S_k$ satisfying (13) can be understood as follows: $S_k$ produces control coefficients $p_{k+1}$ for the knots $T_{k+1}$ such that the differences $E_k p_k$ of the level $k$ grid are maintained at the old knots in $T_k$ and zero differences are forced at all new knots in $T_{k+1} - T_k$. Thus,
telescoping this equation and taking the limit yields a surface which replicates the differences $E_0 p_c$ of the original level 0 grid at the knots in $T_0$ and has zero difference everywhere else.

Expanding equation (13) for $k = 0$ and solving for a sparse $S_0$ yields

$$S_c = \frac{1}{8} \begin{pmatrix}
8 & 0 & 0 & 0 \\
4 & 4 & 0 & 0 \\
1 & 6 & 1 & 0 \\
0 & 4 & 4 & 0 \\
0 & 1 & 6 & 1 \\
0 & 0 & 4 & 4 \\
0 & 0 & 0 & 8
\end{pmatrix}.$$ 

Subdivision matrices for finer grids, $k > 0$ have a similar structure: interior columns of $S_k$ are shifts of the center column of $S_c$. Note that $S_k$ is sparse, i.e. the basis functions induced by the scheme are local. Furthermore, rows in $S_k$ are all positive and sum to one. Hence, the scheme is affinely invariant and the limit curve lies in the convex hull of the control points. Finally, the first and last rows of $S_k$ are unit vectors, i.e. the limit curve interpolates the first and last control point.

The subdivision scheme defined by the $S_k$ is not interpolating. Initial control points $p_0$ must be chosen such that the limit curve satisfies (2). This can be accomplished using the interpolation matrix $I_c$ which contains samples of the basis functions induced by the subdivision scheme at the knots $T_0$. If $c_c$ is the set of interpolation conditions over $T_0$ then we get $p_0$ by $c_0 = I_c p_0$. For the subdivision scheme presented here we get

$$I_0 = \frac{1}{6} \begin{pmatrix}
6 & 0 & 0 & 0 & 0 \\
1 & 4 & 1 & 0 & 0 \\
0 & 1 & 4 & 1 & 0 \\
0 & 0 & 1 & 4 & 1 \\
0 & 0 & 0 & 0 & 6
\end{pmatrix}.$$ 

An application of the scheme is shown in figure 2.

Fig. 2. Subdivision for natural cubic splines.
§4. Conclusion and Future Work

This paper presented a binary subdivision scheme for uniform bounded B-splines. The scheme is stationary, i.e. the same subdivision masks are applied at different levels of the subdivision process. The method presented here extends to higher order variational problems even in higher dimensions. The steps to construct a subdivision scheme which converges to the minimizer of a variational problem can be outlined as follows: First, find appropriate finite element basis functions. Next, compute the energy matrix based on the inner product (7). Finally, find subdivision matrices $S_k$ by solving $E_{k+1}S_k = U_k E_k$. This system is usually rank deficient and a subdivision scheme with particularly nice properties, e.g. sparseness, can be chosen among the possible solutions.

In the future we plan to derive a subdivision scheme for thin plate splines on irregular grids and link the resulting scheme to the representation of these surfaces in terms of radial basis functions. Also, we plan to investigate the link between variational subdivision methods and multiresolution analysis.

Acknowledgments. This work has been supported in part under National Science Foundation grant CCR-9500572, Texas Advanced Technology Program grant 003604-010 and by Western Geophysical.

References


Joe Warren and Henrik Weimer
Department of Computer Science
Rice University
P.O. Box 1892
Houston, TX 77005-1892
{jwarren,henrik}@rice.edu