

Non-Oscillatory Splines on Irregular Data*

*David E. Gilsinn[†], Marjorie A. McClain[‡], and
Christoph Witzgall[§]*

1 Introduction

In computer aided design (CAD), design-defined objects are routinely represented by surfaces that are C^1 or even C^2 . Terrain representation, however, poses a different challenge: The location of "crease lines" or "break lines", along which the actual terrain surfaces are not differentiable, are usually not known ahead of time, whereas in a CAD environment break lines tend to be specified as part of the design. Unspecified break lines, on the other hand, along with actual verticalities, tend to give rise to spurious oscillation and, what may be called, "Gibbs phenomena" in analogy to the phenomenon well known from the theory of Fourier series. Susceptibility to spurious oscillations and Gibbs phenomena are one of the reasons why the terrain modeling community has been slow to accept smooth surfaces. There has thus been a long quest for "non-oscillatory splines" which would obviate this pesky conundrum.

In 1994, Lavery [4, 5] proposed successful paradigms for univariate as well as bivariate non-oscillatory splines, which could be used for representing 2D or 3D data sets, respectively. Lavery introduced the term " L_1 splines" for his brand of nonoscillatory splines. The term L_1 splines is, however, frequently misinterpreted as minimizing an L_1 measure-of-fit when approximating a point set by, say, classical

*Contribution of the National Institute of Standards and Technology, a Federal agency, not subject to copyright.

[†]Mathematical and Computational Sciences Division, National Institute of Standards and Technology, 100 Bureau Drive, Stop 8910 Gaithersburg, MD 20899-8910.

[‡]Mathematical and Computational Sciences Division, National Institute of Standards and Technology, 100 Bureau Drive, Stop 8910 Gaithersburg, MD 20899-8910.

[§]Mathematical and Computational Sciences Division, National Institute of Standards and Technology, 100 Bureau Drive, Stop 8910 Gaithersburg, MD 20899-8910.

splines. For this reason, we prefer the term "Lavery splines".

Classical splines are characterized by their minimizing energy functionals. Lavery splines, on the other hand, minimize different functionals. In the bivariate case, especially, the computational effort of minimizing these functionals, however, exceeds the effort required by the classical approach by an order of magnitude. We have, therefore, considered an approximation to the calculation of Lavery splines, along with a prior modification of the functional proposed by Lavery for the univariate case. This modified univariate functional is an extension of Lavery's functional. We will present preliminary results for the univariate nonoscillatory splines. In Section 2 we discuss aspects of univariate Lavery splines in order to shed light on related issues for bivariate splines. For the issues related to the bivariate case see Witzgall et al. [9]. The algorithm proposed there for bivariate non-oscillatory splines requires solving large sparse systems of linear equations.

2 Univariate Spline Interpolation

In the univariate case, cubic "spline functions" are most commonly used and are considered here. They form a linear space

$$F$$

of piecewise cubic C^1 functions $f(x)$ defined locally over intervals between "knots"

$$x_0 < x_1 < \dots < x_n ,$$

that is, they consist of cubic polynomials

$$f_i(x) , x \in [x_{i-1}, x_i], i = 1, \dots, n .$$

Adjacent cubic polynomials are required to assume the same values y_i at common interior knots,

$$y_i = f_i(x_i) = f_{i+1}(x_i) .$$

This ensures continuity of the complete spline function $f(x)$ over the entire interval

$$I = [x_0, x_n] .$$

In addition, the polynomials are to assume the same slopes

$$m_i = f'_i(x_i) = f'_{i+1}(x_i) .$$

The spline functions $f(x)$ are thus continuously differentiable, that is, they belong to class C^1 . In what follows, the linear spaces

$$F' , F''$$

of first and second derivatives of spline functions are also considered, in spite of the fact that, at common knots, the second derivatives of adjacent cubic polynomials

may not agree, so that the spline functions $f(x) \in F$ are generally not twice differentiable at such knots. However, they are twice differentiable everywhere but on this set of measure zero. For the purposes of integration below, it does not matter that the function $f''(x)$ may not be defined for those arguments.

Each of the constituent cubic polynomials $f_i(x)$ is uniquely determined by the values y_{i-1} , y_i at the knots x_{i-1} , x_i and the slopes m_{i-1} , m_i at those locations (Figure 1), in fact, the polynomial is linear in these parameters. The entire function $f(x)$ is thus uniquely determined by its values and slopes at the knots, and it, too,

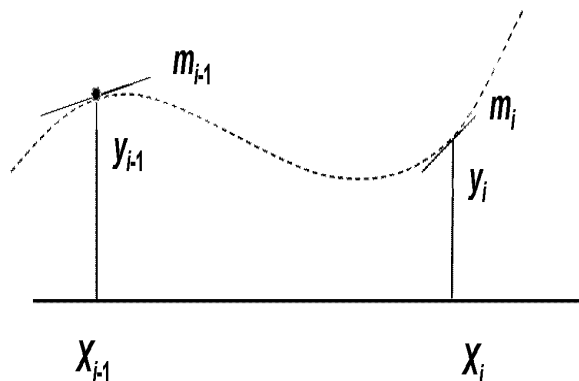


Figure 1. Hermite cubic polynomial determined by coordinates and slopes at end points.

depends linearly on these parameters, so that the space F is isomorphic to the $2(n+1)$ -dimensional vector space of values y_i and slopes m_i , $i = 0, \dots, n$.

We now turn to the task of interpolation. Here the values y_i at the knots x_i are fixed and specified. Given a particular specification of values y_i , a corresponding "interpolating spline function" depends only on the parameters m_i :

$$f(x) = f(m_0, m_1, \dots, m_n; x).$$

Collectively, these functions define affine manifolds or flats,

$$S, S', S'',$$

within the linear spaces F , F' , F'' , respectively. That is, if

$$m_i = \frac{m_i^{(1)} + m_i^{(2)}}{2}, \quad i = 0, 1, \dots, n,$$

then it follows that correspondingly

$$f(x) = f(m_0, m_1, \dots, m_n; x) = \frac{f(m_0^{(1)}, m_1^{(1)}, \dots, m_n^{(1)}; x) + f(m_0^{(2)}, m_1^{(2)}, \dots, m_n^{(2)}; x)}{2}.$$

The question then becomes, how to select slopes m_i so as to achieve a "satisfactory" interpolation. That selection is generally made by minimizing a functional on the affine space S defined by an integral. In this work, we reserve the term "spline" – as opposed to "spline function", for the results of such a minimization.

2.1 Paradigms for Univariate Splines

"Classical splines" are uniquely defined as those interpolating cubic spline functions which are (i) C^2 , that is, twice differentiable, such that

$$f_i''(x_i) = f_{i+1}''(x_i), \quad i = 1, \dots, n-1 \quad (1)$$

holds at interior knots, and for which (ii) the second derivative vanishes

$$f''(x_0) = f''(x_n) = 0 \quad (2)$$

at the two exterior knots. Holladay proved early on (see Ahlberg [1]) that classical splines are also defined as the unique minimizers of the "thin beam energy"

$$E(f) = \int_{x_0}^{x_n} f''(x)^2 dx \quad (3)$$

over the affine space S of all C^1 interpolating spline functions.

Condition (2) is familiar to structural engineers as the vanishing of the second derivative at the "free end" of a beam. It is, however, remarkable that this energy minimization enforces a higher level of compatibility across knots so that the minimizing C^1 spline functions are, in fact, in class C^2 . On the other hand, this stiffness contributes to the tendency of classical splines to produce spurious oscillations, Gibbs phenomena and other undesirable inflections.

Several attempts have been made to avoid these problems. Taking a clue from mechanics, Schweikert [6] and Cline [2, 3] have introduced "tension splines", where the arclength of the spline function is made part of the defining minimization. Reinsch (see Stoer and Bulirsch [7]) moved to "exponential splines" by adding a multiple of the square of first derivatives to the integrand in (3). In Figure 2 we compare the exponential spline of Reinsch against the classic spline. Those efforts were only partially successful. Drawbacks include the need to specify an additional

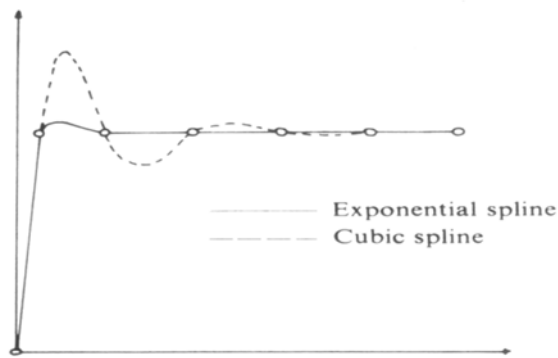


Figure 2. Exponential vs. classic spline.

parameter in order to balance conflicting minimization requirements, and the fact that these techniques are not readily generalized to the bivariate case.

Lavery splines, on the other hand, appear to avoid such shortcomings. In Figures 3 and 4 we compare Lavery splines against classic splines. We are particularly

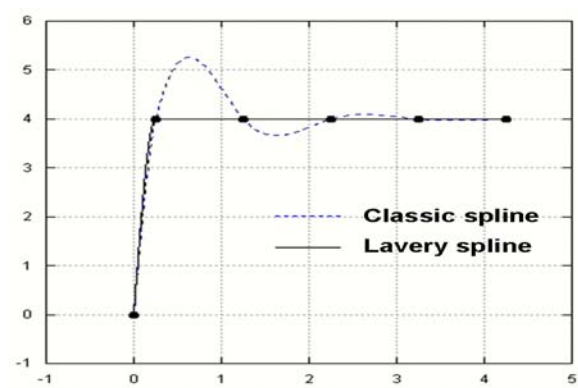


Figure 3. *Lavery spline vs. classic spline: Example 1.*

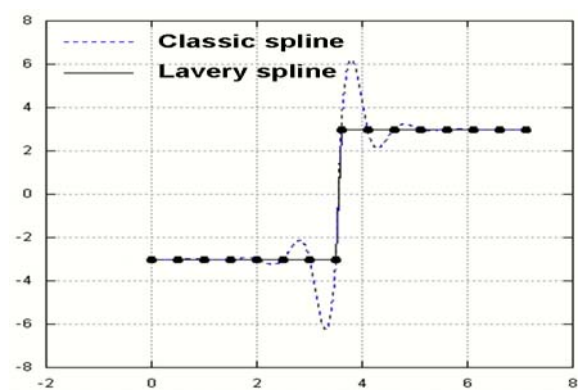


Figure 4. *Lavery spline vs. classic spline: Example 2.*

impressed with the performance of Lavery splines in the examples in Figures 3 and 4. Figures 5 and 6 in particular demonstrate the advantages of Lavery splines. What is the secret of Lavery splines? How are they defined as opposed to classical splines?

Lavery defines his nonoscillatory splines, in essence, as minimizing the integral of the absolute value rather than the square of the second derivative of a spline function:

$$e(f) = \int_{x_0}^{x_n} |f'''(x)| dx \quad (4)$$

over the given affine space S of interpolating spline functions. As pointed out in the beginning of Section 2.1, minimizing over S may be accomplished by minimizing

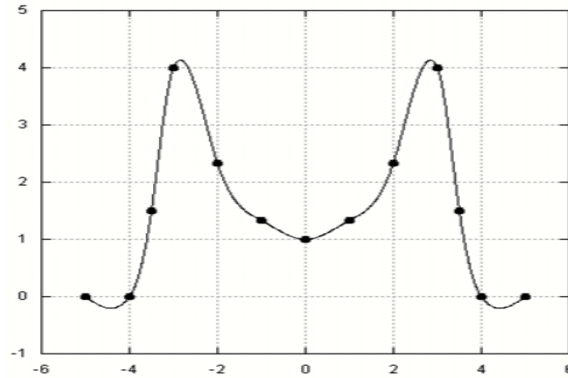


Figure 5. *Classic splines produce unnecessary variations in curvature.*

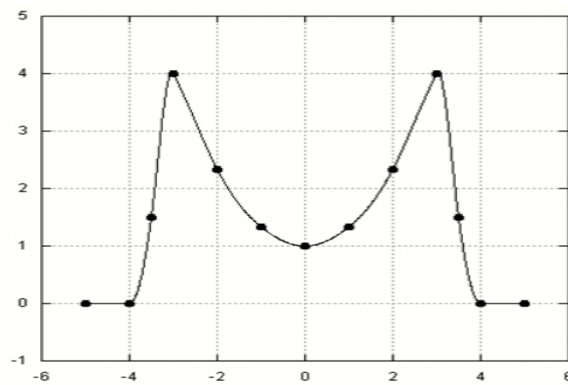


Figure 6. *Lavery spline has smooth curvature transitions.*

over the slopes m_i , $i = 0, \dots, n$ at the knots. This requires expressing the integral $e(f)$ in terms of these slopes:

$$e(f) = e(m_0, m_1, \dots, m_n).$$

2.2 Expressing Holladay and Lavery Integrals

For the purposes of this paper, we will refer to the integrals (3) and (4) as the "Holladay integral" and the "Lavery integral", respectively. The goal of this section is to derive expressions for the values of these integrals in terms of the slopes m_i , $i = 1, \dots, n$, at the knots of the spline function $f(x)$ under consideration. Both are the sums of the corresponding integrals of the second derivatives of the individual cubic polynomials $f_i(x)$, which constitute the spline function $f(x)$.

As pointed out in Section 2.1, each such cubic polynomial is uniquely defined by its end points (x_{i-1}, y_{i-1}) , (x_i, y_i) and its end slopes m_{i-1} , m_i . The

various formulas describing this polynomial are commonly referred to as "Hermite" formulas. For the versions used here, we introduce the quantities

$$\Delta_i = x_i - x_{i-1}$$

and

$$M_i = \frac{y_i - y_{i-1}}{x_i - x_{i-1}} = \frac{y_i - y_{i-1}}{\Delta_i},$$

where M_i represents the slope of the straight line between end points. Instead of referring to the variable x directly, the following formulas are in terms of the weights

$$\begin{aligned} \lambda_i = \lambda_i(x) &= \frac{x_i - x}{x_i - x_{i-1}} = \frac{x_i - x}{\Delta_i}, \\ \mu_i = \mu_i(x) &= \frac{x - x_{i-1}}{x_i - x_{i-1}} = \frac{x - x_{i-1}}{\Delta_i}, \end{aligned} \quad (5)$$

where $\lambda_i + \mu_i = 1$, and $\lambda_i, \mu_i \geq 0$, for x in the interval $[x_{i-1}, x_i]$. Such weights are often referred to as "barycentric coordinates". With these conventions, we find, for instance,

$$f_i(x) = \lambda_i y_{i-1} + \mu_i y_i + \lambda_i^2 \mu_i (m_{i-1} - M_i) \Delta_i - \lambda_i \mu_i^2 (m_i - M_i) \Delta_i. \quad (6)$$

Furthermore, by definition (5),

$$dx = -\Delta_i d\lambda_i = +\Delta_i d\mu_i, \quad (7)$$

the chain rule yields, using $(\lambda_i + \mu_i)^2 = 1$,

$$f'_i(x) = 6\lambda_i \mu_i M_i + (\lambda_i^2 - 2\lambda_i \mu_i) m_{i-1} + (\mu_i^2 - 2\lambda_i \mu_i) m_i.$$

An alternate expression for the first derivative is readily derived:

$$f'_i(x) = \lambda_i m_{i-1} + \mu_i m_i + \lambda_i \mu_i D_i. \quad (8)$$

Here the quantity

$$D_i = 6M_i - 3m_{i-1} - 3m_i = 6 \left(M_i - \frac{m_{i-1} + m_i}{2} \right), \quad (9)$$

vanishes if and only if the polynomial $f(x)$: has degree less than three. Because

$$\lambda_i \mu_i = \left(\frac{x_i - x}{\Delta_i} \right) \left(\frac{x - x_{i-1}}{\Delta_i} \right)$$

$-D_i$ is seen as the lead coefficient of $f'_i(x)$ as expressed in x . Consequently $-\frac{1}{3}D_i$ is the lead coefficient of $f_i(x)$. $D_i < 0$ indicates that the function is concave up to its inflection point and convex thereafter. Conversely, $D_i > 0$ indicates

and

$$c = \left(\frac{M_1}{\Delta_1}\right)^2 + \left(\frac{M_2}{\Delta_2}\right)^2 + \dots + \left(\frac{M_n}{\Delta_n}\right)^2.$$

Minimizing this expression for the Holladay energy integral is to solve the linear system of equations

$$2\mathbf{H}\vec{m} = \mathbf{M} \quad (15)$$

or

$$\begin{aligned} \frac{2}{\Delta_1}m_0 + \frac{1}{\Delta_1}m_1 &= \frac{3}{\Delta_1}M_1 \\ \frac{1}{\Delta_1}m_0 + \left(\frac{2}{\Delta_1} + \frac{2}{\Delta_2}\right)m_1 + \frac{1}{\Delta_2}m_2 &= \frac{3}{\Delta_1}M_1 + \frac{3}{\Delta_2}M_2 \\ \frac{1}{\Delta_2}m_1 + \left(\frac{2}{\Delta_2} + \frac{2}{\Delta_3}\right)m_2 + \frac{1}{\Delta_3}m_3 &= \frac{3}{\Delta_2}M_2 + \frac{3}{\Delta_3}M_3 \\ &\dots = \dots \\ \frac{1}{\Delta_n}m_{n-1} + \frac{2}{\Delta_n}m_n &= \frac{3}{\Delta_n}M_n \end{aligned} \quad (16)$$

The first equation, in fact, may be restated as

$$\frac{2}{\Delta_1}U_0 = f_1''(x_0) = 0,$$

that is, as the requirement (2) that second derivatives vanishes at the first knot. The last equation reflects the corresponding requirement at the last knot. The second equation is equivalent to

$$\frac{2}{\Delta_1}V_1 + \frac{2}{\Delta_2}U_2$$

which is the condition (1) for second order differentiability at the interior knot x_1 . The remaining equations similarly enforce compatibility of the one-sided second derivatives at the remaining interior knots. This confirms the result of Holladay.

The reader should note that the system (16) for classical splines differs from the linear system mostly offered in the current literature. There the second order differentiability of the classical splines is already assumed and the spline functions are formulated in terms of those second order derivatives, $n_i = f''(x_i)$, $i = 0, \dots, n$. However, for weighted classical splines to be encountered in Section 2.4, second order differentiability no longer holds, and a weighted version of the linear system (16) needs to be considered.

Expressing the Lavery integral

As to the integral of the absolute value of the second derivative, it is readily available in the case that the derivative does not change sign inside the subinterval:

$$e(f_i) = \int_{x_{i-1}}^{x_i} |f_i''(x)| dx = \left| \int_{x_{i-1}}^{x_i} f_i''(x) dx \right| = |f_i'(x_i) - f_i'(x_{i-1})| = |m_i - m_{i-1}|.$$

This includes the case in which the polynomial is of lesser degree than cubic, and thus has constant second derivatives. This case is signaled by the vanishing of the quantity D_i introduced earlier in (9).

The function $f_i''(x)$, however, is a linear function in x and, unless constant, changes sign at some location \hat{x} , which also marks the location of the inflection point of $f_i(x)$. Suppose this location falls into the interior of the subinterval:

$$x_{i-1} < \hat{x} < x_i.$$

Then the integral has to be calculated as the sum of two integrals of linear functions:

$$e(f_i) = \left| \int_{x_{i-1}}^{\hat{x}} f_i''(x) dx \right| + \left| \int_{\hat{x}}^{x_i} f_i''(x) dx \right|.$$

The integrals between the absolute value bars are of opposite signs, so that the total integral can be written as the absolute value of a difference of integrals

$$e(f_i) = \left| \int_{x_{i-1}}^{\hat{x}} f_i''(x) dx - \int_{\hat{x}}^{x_i} f_i''(x) dx \right|.$$

These integrals can be separately evaluated as differences of slopes. Let

$$\hat{m}_i = f_i'(\hat{x})$$

denote the inflection slope, then – in the case of an interior inflection –

$$e(f_i) = |2\hat{m}_i - m_{i-1} - m_i|. \quad (17)$$

The quantities U_i , V_i depend linearly on the two slopes m_{i-1} , m_i . Conversely, those slopes can be expressed in terms of U_i , V_i .

$$m_{i-1} = M_i - \frac{2}{3}U_i + \frac{1}{3}V_i, \quad m_i = M_i + \frac{1}{3}U_{i-1} - \frac{2}{3}V_i. \quad (18)$$

The next step is to express the inflection slope in terms of U_i , V_i . To this end, we express the inflection argument by its barycentric weights:

$$\hat{x} = \hat{\lambda}_i x_{i-1} + \hat{\mu}_i x_i, \quad (19)$$

From the Hermite expression (10) for $f_i''(x)$ it follows that

$$\hat{\lambda}_i = \frac{V_i}{D_i}, \quad \hat{\mu}_i = \frac{U_i}{D_i}, \quad (20)$$

where the denominator $D_i = U_i + V_i = 6M_i - 3m_{i-1} - 3m_i$ has already been encountered in (9) as a quantity that vanishes if and only if the polynomial in question is parabolic or linear. Thus $D_i = 0$ leads back to the previous case of no sign changes by the second derivative.

The weights $\hat{\lambda}_i$, $\hat{\mu}_i$ may now be inserted into the expression (8) for the derivative $f'_i(x)$. This gives

$$\hat{m}_i = \frac{V_i}{D_i} m_{i-1} + \frac{U_i}{D_i} m_i + \frac{V_i U_i}{D_i}.$$

Substituting for m_{i-1} , m_i according to (18) yields

$$\hat{m}_i = M_i + \frac{U_i^2 + V_i^2 - V_i U_i}{3D_i}$$

as well as

$$e(f_i) = \frac{U_i^2 + V_i^2}{3|D_i|} = \frac{\hat{\mu}_i^2 + \hat{\lambda}_i^2}{3} |D_i|. \quad (21)$$

in view of (20). In terms of the slopes m_{i-1} , m_i :

$$e(f_i) = \left| \frac{18M_i^2 - 18M_i(m_{i-1} + m_i) + 5m_{i-1}^2 + 8m_{i-1}m_i + 5m_i^2}{6M_i - 3m_{i-1} - 3m_i} \right|. \quad (22)$$

The expressions (21) and (22) for $e(f_i)$ are also valid if the inflections occur at the ends x_{i-1} , x_i of the interval of definition, reducing to $e(f_i) = |m_i - m_{i-1}|$, in accordance with earlier results..

We are now ready to examine the full Lavery integral. At first blush, all that remains to be done is to sum over the partial integrals $e(f_i)$ in their various forms. We will show, however, that many terms of the expressions (17) cancel each other out as these partial integrals are added together. To this end, we distinguish five separate kinds of polynomials $f_i(x)$ depending on their behavior in the interior of the interval between its knots, $x_{i-1} < x < x_i$:

- "Linear"; here $f_i''(x) = 0$ throughout and $e(f_i) = 0$
- "Convex"; here $f_i''(x) > 0$ in the interior of the interval $[x_{i-1}, x_i]$ and $e(f_i) = m_i - m_{i-1}$
- "Concave"; here $f_i''(x) < 0$ in the interior of the interval $[x_{i-1}, x_i]$ and $e(f_i) = m_{i-1} - m_i$
- "Convex-concave"; here $f_i''(x) > 0$ for $x < \hat{x}$, $f_i''(x) < 0$ for $x > \hat{x}$ and $e(f_i) = +\hat{m} - m_{i-1} - m_i$, also $D_i > 0$
- "Concave-convex"; here $f_i''(x) < 0$ for $x < \hat{x}$, $f_i''(x) > 0$ for $x > \hat{x}$ and $e(f_i) = -\hat{m} + m_{i-1} + m_i$, also $D_i < 0$

The last two categories are the ones with an interior inflection point \hat{x} and inflection slope \hat{m} .

The interior inflection points considered so far may not be the only inflection points of the spline function $f(x)$. Inflections occur also at knots x_i , $0 < i <$

n if a concave or convex-concave polynomial is followed by a convex or convex-concave polynomial and, analogously, if a convex or concave-convex polynomial is followed by a concave or concave-convex polynomial. We will refer to such knots as "inflection knots". To make matters more complicated, however, inflection may occur along an entire stretch of consecutive linear polynomials of equal slope, the inflection slope in this case, provided there are adjacent nonlinear polynomials at both ends of such a stretch exhibiting the same convexity/concavity pattern that would cause an inflection at a knot. In this case, we choose an arbitrary knot, say, the leftmost one in the linear stretch, as the inflection knot representing the inflection.

Clearly, slopes at interior knots that are not inflections cancel out as the expressions (17) for the Lavery integrals $e(f_i)$ for the partial spline functions $f_i(x)$ are added up. See Figure 7 for an example. This leads to

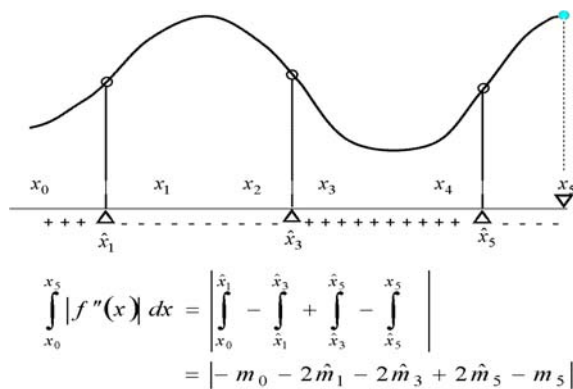


Figure 7. Lavery integral as expressed by inflection slopes and end slopes.

Observation A : (23)

The Lavery integral of a cubic spline function is the absolute value of an alternating sum of the inflection slopes and of the end slopes m_0, m_n . Let

$$\hat{m}_1, \hat{m}_2, \hat{m}_3, \dots, \hat{m}_L,$$

be the sequence of all inflection points identified above, sorted from left to right by their location. Then

$$e(f) = \left| m_0 + 2 \sum_{l=1}^L (-1)^l \hat{m}_l - (-1)^L m_n \right|.$$

(Note that the indices l of \hat{m}_l do not refer to the interval in which they are located).

2.3 Properties of Lavery Splines

In this section, we gather some information about Lavery Splines, namely, interpolating spline functions in the affine space S which minimize their respective Lavery integrals. A first general observation concerns the convexity of the Lavery integral.

The Holladay and Lavery integrals (3) and (4) of a piecewise cubic spline function,

$$f(x) = f(m_0, m_1, \dots, m_n; x)$$

are functions of the slope specifications m_i :

$$E(f) = E(m_1, m_2, \dots, m_n), \quad e(f) = e(m_1, m_2, \dots, m_n).$$

The quadratic function $E(f)$ representing the Holladay integral can be shown to be positive definite and, therefore, strictly convex. Its minimum is unique on the affine manifold S interpolating spline functions. The restriction to S is, of course, necessary as the value of $E(f)$ would not change if the spline function $f(x)$ were modified by adding a linear function in x . Adding a non-zero linear function to the function $f(x)$ would not preserve its interpolation property.

Convexity and Uniqueness of Inflection Points

Next, we establish the convexity of the Lavery integral. It may be viewed as the extension of the L_1 vector norm to a norm on the linear space F'' of second derivatives of spline functions:

$$e(f) = \|f''\|_1.$$

The following generic seminorm properties are easily verified for the piecewise linear functions in F'' ,

$$\begin{aligned} \|f''\|_1 = 0 &\iff f'' = 0 \\ |\alpha| \|f''\|_1 &= \|\alpha f''\|_1 \end{aligned}$$

along with the triangle inequality,

$$\|f^{(1)''} + f^{(2)''}\|_1 \leq \|f^{(1)''}\|_1 + \|f^{(2)''}\|_1. \quad (24)$$

Suppose the two spline functions $f^{(1)}$, $f^{(2)}$ are actually two interpolating spline functions and, therefore, in the affine space S . Then their mean is again in S and, from the triangle inequality (24),

$$\left\| \frac{f^{(1)''} + f^{(2)''}}{2} \right\|_1 \leq \frac{\|f^{(1)''}\|_1 + \|f^{(2)''}\|_1}{2}.$$

In terms of Lavery integrals,

$$e\left(\frac{m_0^{(1)} + m_0^{(2)}}{2}, \dots, \frac{m_n^{(1)} + m_n^{(2)}}{2}\right) \leq \frac{e(m_1^{(1)}, \dots, m_n^{(1)}) + e(m_1^{(2)}, \dots, m_n^{(2)})}{2},$$

which establishes convexity. Contrary to the Holladay functional which is strictly convex, the Lavery integral is not. As a result, uniqueness does not follow and, in fact, does not hold, as an example in Section ?? will show. Such minima of a convex function, however, must form a convex set. This leads immediately to

Observation B : (25)

Any positive linear combination of Lavery splines for the same interpolation problem, – in particular, their mean –, is again a Lavery spline for this problem.

Note that, in general,

$$\int \frac{|f^{(1)''}(x)| + |f^{(2)''}(x)|}{2} dx \geq \int \frac{|f^{(1)''}(x) + f^{(2)''}(x)|}{2} dx .$$

If both $f^{(1)}$ and $f^{(2)}$ are Lavery splines for the same interpolation problem, then so is their mean, and all three functions

$$f^{(1)''} , f^{(2)''} , \frac{f^{(1)''} + f^{(2)''}}{2}$$

return the same optimal value for their Lavery integrals. Thus equality holds in the above relation. This implies that both $f^{(1)''}(x)$ and $f^{(2)''}(x)$ have the same sign pattern:

$$f^{(1)''}(x) \geq 0 \text{ if and only if } f^{(2)''}(x) \geq 0 .$$

This can be rephrased as

Observation C : (26)

Two Lavery splines for the same optimization problem share essentially the same inflections: if one of them has an inflection point at $x = \hat{x}$, then so has the other unless it is linear at this point.

Free Ends of Lavery Splines

Here we will examine the free ends of Lavery splines, in particular, the cubic polynomial $f_1(x)$ and the corresponding first summand $e(f_1)$ of the Lavery integral $e(f)$. As the end slope m_0 may vary freely, it must optimize $e(f_1)$ while keeping the slope m_1 fixed. This fact determines the behavior of Lavery splines at free ends. As seen in the previous section, the first summand $e(f_1)$ is a convex function in the variables m_0, m_1 . If the slope m_1 is held fixed, $e(f_1)(m_0)$ is convex as a function in m_0 alone. If the fixed slope equals the straight-line slope, then $m_0 = m_1 = M_1$ obviously represents the optimal value for m_0 , since $f_1(x)$ in that case is a straight line with $e(f_1) = 0$. We suppose, therefore, that

$m_1 \neq M_1$, and we examine the case, that $f_1(x)$ has an inflection \hat{x} in the interval between the first two knots, $x_0 \leq \hat{x} \leq x_1$.

Consider the function

$$\tilde{e}(f_1)(m_0) = \frac{U_1^2 + V_1^2}{D_1},$$

where \hat{m} as well as U_1, V_1 also depend on the variable m_0 . Clearly, the absolute value of $\tilde{e}(f_1)$ is given by $e(f_1)$. Note that

$$\frac{\partial \tilde{e}(f_1)}{\partial m_0} = \frac{V_1^2 - U_1^2 - 6V_1U_1}{D_1^2} = \hat{\lambda}_1^2 - \hat{\mu}_1^2 - 6\hat{\lambda}_1\hat{\mu}_1 = 6\hat{\lambda}_1^2 - 4\hat{\lambda}_1 - 1 \quad (27)$$

in view of (20), and

$$\frac{\partial U_1}{\partial m_0} = -2, \quad \frac{\partial V_1}{\partial m_0} = -1, \quad \frac{\partial D_1}{\partial m_0} = -3.$$

Solving the quadratic equation for $\hat{\lambda}$, and taking into account that $0 \leq \hat{\lambda} \leq 1$, yields

$$\hat{\lambda}_1 = \frac{2 + \sqrt{10}}{6} = 0.860378, \quad \hat{\mu}_1 = \frac{4 - \sqrt{10}}{6} = 0.139622. \quad (28)$$

The value of m_0 for which this value for $\hat{\lambda}_1$ is realized can be inferred from the general definition (20) of an inflection defining the barycentric coordinate $\hat{\lambda}_1$ in terms of the slopes m_0, m_1 , as follows:

$$\hat{\lambda}_1(6M_1 - 3m_0 - 3m_1) = 3M_1 - m_0 - 2m_1, \quad (29)$$

which – for the particular value (28) of $\hat{\lambda}_1$ – yields the corresponding end-slope

$$m_0 = \frac{10 - \sqrt{10}}{5}M_1 - \frac{5 - \sqrt{10}}{5}m_1. \quad (30)$$

This value for m_0 represents a locally unique stationary value of $\tilde{e}(f_1)(m_0)$.

Now $\tilde{e}(f_1)(m_0) = 0$ would imply $e(f_1)(m_0) = 0$, and consequently linearity, that is, $m_0 = m_1 = M_1$, which has been ruled out. By continuity, $\tilde{e}(f_1)(m_0)$ is either always positive or always negative, – in other words, either

$$e(f_1)(m_0) = +\tilde{e}(f_1)(m_0)$$

or

$$e(f_1)(m_0) = -\tilde{e}(f_1)(m_0).$$

This implies that the value (30) for m_0 is also a locally unique stationary value of $e(f_1)(m_0)$. In view of the convexity of this function, it is also its minimizer. At the last end we differentiate

$$\hat{e}(f_n) = \frac{U_n^2 + V_n^2}{3D_n} \quad (31)$$

and find

$$\frac{\partial \hat{e}(f_n)}{\partial m_n} = \frac{U_n^2 - V_n^2 - 6U_n V_n}{D_n^2} = 6\hat{\mu}_n^2 - 4\hat{\mu}_n - 1. \quad (32)$$

Symmetrically, we thus have

$$\hat{\mu}_n = \hat{\lambda}_1, \quad \hat{\lambda}_n = \hat{\mu}_1. \quad (33)$$

The equation

$$\hat{\mu}_n (6M_n - 3m_{n-1} - 3m_n) = 3M_n - 2m_{n-1} - m_n \quad (34)$$

thus yields a symmetric relationship to (30):

$$m_n = \frac{10 - \sqrt{10}}{5} M_n - \frac{5 - \sqrt{10}}{5} m_{n-1}. \quad (35)$$

This establishes

Observation D : (36)

The free ends of Lavery splines are either linear functions or they contain an inflection in the interior of their interval of definition. The locations \hat{x}_1 , \hat{x}_n of these inflections are universally given, respectively, by

$$\hat{x}_0 = \frac{2 + \sqrt{10}}{6} x_0 + \frac{4 - \sqrt{10}}{6} x_1, \quad \hat{x}_n = \frac{4 - \sqrt{10}}{6} x_{n-1} + \frac{2 + \sqrt{10}}{6} x_n.$$

Observation D enables us to determine universal values for partial Lavery integrals at the end-intervals of Lavery splines. By (20), (22), (30), and again by (33), which implies

$$\hat{\mu}_1^2 + \hat{\lambda}_1^2 = \hat{\lambda}_n^2 + \hat{\mu}_n^2 = \frac{10 - \sqrt{10}}{9}, \quad (37)$$

we find

$$e(f_1) = \frac{10 - \sqrt{10}}{27} |D_1|; \quad e(f_n) = \frac{10 - \sqrt{10}}{27} |D_n|. \quad (38)$$

Substituting for m_0 and m_n , respectively in $D_1 = 6M_1 - 3m_0 - 3m_1$ and $D_n = 6M_n - 3m_{n-1} - 3m_n$, yields

$$D_1 = \frac{3\sqrt{10}}{5} (M_1 - m_1), \quad D_n = \frac{3\sqrt{10}}{5} (M_n - m_{n-1}), \quad (39)$$

giving rise to

Observation E : (40)

A necessary, but far from sufficient, condition for a spline function f to be a Lavery spline is that

$$e(f_1) = \frac{2}{9} (\sqrt{10} - 1) |M_1 - m_1|, \quad e(f_n) = \frac{2}{9} (\sqrt{10} - 1) |M_n - m_{n-1}| \quad (41)$$

Examples of Non-unique Lavery Splines

In this section, we present an example in which the Lavery splines are not unique. Consider the three points (Figures 8, 9, 10):

$$\begin{aligned}
 P_0 &= (x_0, y_0) = (-1, -1) \\
 P_1 &= (x_1, y_1) = (0, 0) \\
 P_2 &= (x_2, y_2) = (+1, -1)
 \end{aligned}
 \tag{42}$$

The associated interpolating cubic spline functions $f(x)$ are then defined by their

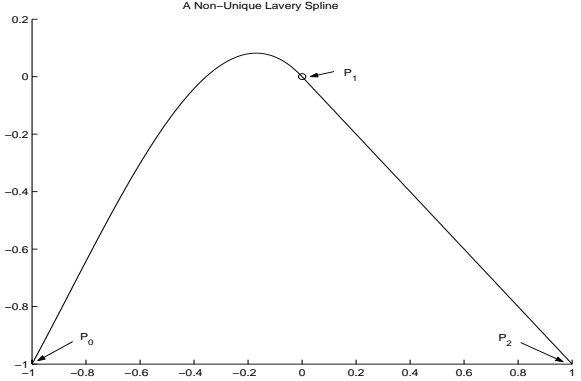


Figure 8. One of three Lavery splines for the same interpolation problem.

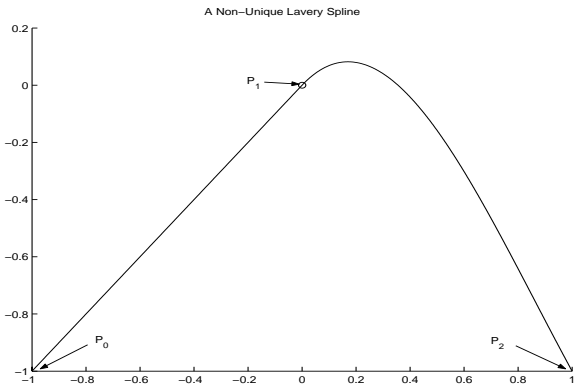


Figure 9. Second of three Lavery splines for the same interpolation problem.

slopes

$$m_0, m_1, m_2$$

at these points. There are two subintervals with cubic polynomials

$$f_1(x), f_2(x),$$

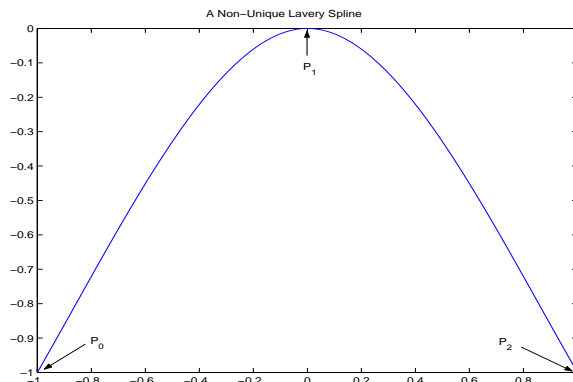


Figure 10. Third of three Lavery splines for the same interpolation problem.

each of them a free end. This determines the coordinates \hat{x}_1 , \hat{x}_2 at inflections:

$$\hat{x}_1 = -\frac{2 + \sqrt{10}}{6}, \quad \hat{x}_2 = +\frac{2 + \sqrt{10}}{6}.$$

Note that both partial Lavery integrals are end-integrals. For an interpolating spline function $f(x)$ to be a Lavery spline it will be necessary by **Observation E** that

$$e(f_1) = \frac{2}{9}(\sqrt{10} - 1)|1 - m_1|, \quad e(f_2) = \frac{\sqrt{10} - 1}{9}|-1 - m_1| \quad (43)$$

so that

$$e(f) = \frac{2}{9}(\sqrt{10} - 1)(|1 - m_1| + |-1 - m_1|). \quad (44)$$

Clearly

$$-1 \leq m_1 \leq 1 \quad (45)$$

implies

$$|1 - m_1| + |-1 - m_1| = |(1 - m_1) + (-1 - m_1)| = 2|m_1| \leq 2, \quad (46)$$

whereas either $m_1 < -1$ or $m_1 > 1$ would imply

$$|1 - m_1| + |-1 - m_1| = 2 \geq 2|m_1|. \quad (47)$$

Thus condition (45) characterizes all Lavery splines for the example.

Consider now any slope m_1 from -1 through $+1$. For $m_1 = -1$, we have by (30) and in view of $M_1 = 1$:

$$m_0 = \frac{+15 - 2\sqrt{10}}{5}, \quad m_2 = -1.$$

Symmetrically, we find for $m_1 = +1$ that

$$m_0 = +1, \quad m_2 = \frac{-15 + 2\sqrt{10}}{5}.$$

These slopes, respectively, determine the two extreme Lavery splines, because each partial function is at a free end, and is optimized according to Observation D (36) in the previous section. The two resulting Lavery splines are extreme in that each has a straight line segment as a partial function.

Using Hermite's formula (6) and substituting for x , we find for the choice $m_1 = -1$,

$$\begin{aligned} f_1(x) &= -\frac{2\sqrt{10}}{5}x^3 - \frac{10 + 2\sqrt{10}}{5}x^2 - x \\ f_2(x) &= -x. \end{aligned}$$

For $m_1 = +1$, the resulting Lavery spline is the symmetric image of the previous one:

$$\begin{aligned} f_1(x) &= +x \\ f_2(x) &= \frac{2\sqrt{10}}{5}x^3 - \frac{10 + 2\sqrt{10}}{5}x^2 + x. \end{aligned}$$

Both splines are shown in Figures 8, 9. The self-symmetric spline from the choice $m_1 = 0$ is shown in Figure 10. In the latter case,

$$m_0 = \frac{10 - \sqrt{10}}{5} = -m_2,$$

and

$$\begin{aligned} f_1(x) &= -\frac{\sqrt{10}}{5}x^3 - \frac{5 + \sqrt{10}}{5}x^2 \\ f_2(x) &= +\frac{\sqrt{10}}{5}x^3 - \frac{5 + \sqrt{10}}{5}x^2. \end{aligned}$$

This Lavery spline is the mean of the two splines with linear free ends. This is an instance of observation B about positive linear combinations of Lavery splines in Section 2.3.

2.4 Computing Univariate Lavery Splines

We now turn our attention to the computation of Lavery splines. The commonly used approach (see Lavery [4, 5]) is to minimize the Lavery integral in discretized form, say,

$$\int_{x_0}^{x_n} |f''(x)| dx = \sum_i \frac{\Delta_i}{k_i} \sum_{k=1}^{k_i} \left| f''\left(x_{i-1} + \frac{k-1}{k_i} \Delta_i\right) \right|, \quad (48)$$

where the integrand is sampled in each subinterval $[x_{i-1}, x_i]$ at k_i equidistant points. Now

$$x_{i-1} + \frac{k-1}{k_i} \Delta_i = \frac{(k_i - k + 1)x_{i-1} + (k-1)x_i}{k_i} = \lambda_{i,k} x_{i-1} + \mu_{i,k} x_i \quad \lambda_{i,k} + \mu_{i,k} = 1.$$

Thus

$$\begin{aligned} \int_{x_0}^{x_n} f''(x) dx &= \sum_i \frac{\Delta_i}{k_i} \sum_{k=1}^{k_i} |\lambda_{i,k} U_i + \mu_{i,k} V_i| \\ &= \sum_i \frac{\Delta_i}{k_i} \sum_{k=1}^{k_i} |\lambda_{i,k} (3M_i - 2m_{i-1} - m_i) + \mu_{i,k} (3M_i - m_{i-1} - 2m_i)| . \end{aligned}$$

The Lavery integral is discretized as a sum of absolute values of the variables m_i of the minimization. Minimizing such an expression is a well-known Linear Programming problem. Many satisfactory methods for solving it are available, such as the Simplex Method or Interior Point Methods. The accuracy of the discretization increases with the number of sample points, but computational effort increases accordingly.

For that reason, and also to motivate an analogous approach in the bivariate case, we are proposing to minimize a different approximation to the Lavery integral, one that takes advantage of the ease of computation offered by energy minimization.

By the mean value theorem of integral calculus, there exist arguments u_i such that

$$\int_{x_{i-1}}^{x_i} f_i''(x)^2 dx = \Delta_i (f_i''(u_i))^2, \quad i = 1, \dots, n .$$

we then propose to approximate the Lavery integral by the following Riemann sum:

$$\int_{x_0}^{x_n} |f''(x)| dx \approx \sum_i \Delta_i |f_i''(u_i)| . \quad (49)$$

In contrast to the approximation (48) by discretization, this approximation does not offer the option of further refinement, unless the interpolation problem itself is changed by adding additional knots and ordinates. In a sense, it approximates the Lavery paradigm itself rather than the Lavery integral. We still use, however, the term "approximate Lavery integral" for our proposed approximation (49).

An Iterative Algorithm for Approximate Lavery Splines

For the purpose of computation, we rewrite the non-zero terms in (49)

$$\Delta_i \sqrt{f_i''(u_i)^2} = \sqrt{\Delta_i} \frac{\Delta_i f_i''(u_i)^2}{\sqrt{\Delta_i f_i''(u_i)^2}} = \frac{\sqrt{\Delta_i}}{\sqrt{\int_{x_{i-1}}^{x_i} f_i''(x)^2 dx}} \int_{x_{i-1}}^{x_i} f_i''(x)^2 dx \quad (50)$$

This expression suggests an iterative approach. Starting with the classic spline $f^{(0)}(x)$, a sequence of interpolating spline functions is generated in hopes to converge towards the approximate Lavery integral (49),

$$f^{(0)}(x), f^{(1)}(x), \dots, f^{(l)}(x), \dots$$

with associated partial Holladay integrals

$$E^{f_i^{(l)}} = \int_{x_{i-1}}^{x_i} \left(f_i^{(l)''}(x) \right)^2 dx, \quad i = 1, \dots, n.$$

At each step $l = 0, 1, 2, \dots$, weights

$$w_i^{(l)} = \sqrt{\frac{\Delta_i}{E(f_i^{(l)})}}, \quad i = 1, \dots, n, \quad (51)$$

are introduced. Given the function $f^{(l)}$, the subsequent function $f^{(l+1)}$ is determined as the solution to the following minimization problem:

$$\min \sum_i w_i^{(l)} \int_{x_{i-1}}^{x_i} \left(f_i^{(l+1)''}(x) \right)^2 dx.$$

This approach raises the question what to do if $E(f_i^{(l)}) = 0$?

Simply ignoring such terms may prematurely lock in straight line segments between knots. What first comes to mind is to specify a limit $\epsilon > 0$ and boost lower values of $E(f_i^{(l)})$ to this level. A more diligent procedure might be to start with all weights at value 1, – the weight setting that yields the initial classical spline according to Holliday's observation –, and then progressively increase the use of the weights given by (51). Such strategies remain to be explored.

In general, adding up partial Holliday integrals, each with weight, say, w_i leads again to an expression of the energy of a physical structure: a collection of thin beams of different thicknesses given by w_i , respectively, and welded together at knots. Minimizing this energy expression requires an adjustment to the linear system (15).

The weighted energy of the partial spline functions $f_i(x)$ is just the product of the straight Holliday integral and the respective weight:

$$E_W(f_i) = w_i E(f_i).$$

The total weighted energy is thus given by

$$E_W(f) = \sum_i E_w(f_i) = \sum_i w_i E(f_i).$$

This means that in the expression (14) for $E(f_i)$, the factor $\frac{4}{\Delta_i}$ is affected as it is multiplied by w_i . In other words, the substitution,

$$\frac{1}{\Delta_i} \rightarrow \frac{w_i}{\Delta_i},$$

transforms the linear system (15) for minimizing $E(f)$ into the linear system for minimizing $E_W(f)$:

$$\frac{2w_1}{\Delta_1} m_0 + \frac{w_1}{\Delta_1} m_1 = \frac{3w_1}{\Delta_1} M_1$$

$$\begin{aligned}
\frac{w_1}{\Delta_1} m_0 + \left(\frac{2w_1}{\Delta_1} + \frac{2w_2}{\Delta_2} \right) m_1 + \frac{w_2}{\Delta_2} m_2 &= \frac{3w_1}{\Delta_1} M_1 + \frac{3w_2}{\Delta_2} M_2 \\
\frac{w_2}{\Delta_2} m_1 + \left(\frac{2w_2}{\Delta_2} + \frac{2w_3}{\Delta_3} \right) m_2 + \frac{w_3}{\Delta_3} m_3 &= \frac{3w_2}{\Delta_2} M_2 + \frac{3w_3}{\Delta_3} M_3 \\
&\dots = \dots \\
\frac{w_n}{\Delta_n} m_{n-1} + \frac{2w_n}{\Delta_n} m_n &= \frac{3w_n}{\Delta_n} M_n
\end{aligned} \tag{52}$$

Note that the weighted classic splines are, in general, not twice differentiable at the knots. However, due to the fact that the first and the last equations have common factors w_o , w_n , respectively, they are equivalent to the corresponding equations in the Holladay system (15): free ends thus have zero second derivatives in the weighted case, too. Again, this is to be expected from Physics.

Note also that some commonly used methods for determining classic splines such as B-splines or linear systems formulated in terms of second derivatives at knots do not carry over to the weighted case. However, the above linear system is still "banded", and many excellent methods are known for its solution. To solve this system we used the venerable Gauss-Seidel method, not just for ease of programming, but also because it seems to work for bivariate weighted splines. Its advantage lies in the fact that the matrix of the linear system need not be changed and can be read, so to speak, in sequence. This is important for the very large systems likely to arise in the bivariate case. In the univariate case, the convergence behavior is well understood (See Varga [8]). Using an iterative method for solving the class of linear systems above will result in a two-tiered iteration procedure: an "outer" iteration, developing new sets of weights, and an "inner" iteration, solving the resulting linear system. Such procedures can be "balanced", that is, the inner iteration may be terminated at a lower accuracy level during the early stages of the outer iteration and may be carried to a higher level of accuracy as the outer method approaches convergence. This is an added advantage of an iterative method for solving the linear systems at hand.

Note finally, that the approximate Lavery method proposed here will definitely not converge to the optimal Lavery integral. This is because the approximate solutions are based on weighted classic splines, and therefore have vanishing second derivatives at the free ends. The second derivatives of Lavery splines, on the other hand, assume their inflections in interiors of the end interval (See **Observation D** (36)).

Observation F : (53)

Unless a free end function of an approximate Lavery spline is linear, it disagrees with the corresponding end function of the true Lavery spline in that the latter has an inflection in the interior of at least one end interval, whereas the former assumes its inflection at the end knot.

The justification for introducing the approximate Lavery splines concept lies

in its computational ease and the fact that it appears to retain the anti-oscillatory dynamic of the original Lavery concept. In fact, the examples in Figures 3 and 4 were calculated using the approximate algorithm outlined in this section.

3 Conclusions

In this paper we have investigated some properties of non-oscillatory splines introduced by John Lavery [4, 5]. These splines, called in this paper Lavery splines, minimize what we have termed the Lavery integral (4). We have seen that the minimizing spline for (4) does indeed model sharp edges and jumps in data without introducing the "Gibbs phenomenon" at the corners. We have shown that the Lavery integral and the associated Lavery splines satisfy a number of properties. First, we have shown that the Lavery integral of a cubic spline function is the absolute value of an alternating sum of inflection slopes and of the end slopes. Next, we showed that any positive linear combination of Lavery splines for the same interpolation problem is again a Lavery spline for the same problem. Furthermore, two Lavery splines for the same optimization problem share essentially the same inflection points and that the free ends of Lavery splines are either linear functions or they contain an inflection point in the interval of definition.

Two algorithms for estimating Lavery splines have also been considered. The first algorithm introduced by John Lavery [4, 5] reduces to solving a least absolute value minimization problem for which he used an interior point method for linear programming to obtain the minimum spline coefficients. The absolute value minimization was based on a "discretization" which leads to a very large number of variables as sufficiently smaller discretizations were considered. The extension of this method to bivariate Lavery splines also led to computationally intensive compute times even for moderate data sizes. The authors have introduced a modified approach based on an iterated weighted least squares algorithm. Although the minimizing spline this algorithm produces is not a Lavery spline it is an approximation that also produces sharp edges without the "Gibbs phenomenon."

Bibliography

- [1] J.H. Ahlberg, E. N. Nilson, and J. L. Walsh, The Theory of Splines and Their Applications, Academic Press, New York, 1967.
- [2] A. K. Cline, Scalar-and Planar- Valued Curve Fitting Using Splines Under Tension, Communications of the ACM, 17 (1974), 218-220.
- [3] A. K. Cline, Algorithm 476 - Six subprograms for curve fitting using splines under tension, Communications of the ACM 17 (1974), 220-223.
- [4] J. E. Lavery, Univariate cubic L_p splines and shape preserving, multiscale interpolation by univariate cubic L_1 splines, Comput. Aided Design 17 (2000), 319-336.
- [5] J. E. Lavery, Shape-preserving, multiscale fitting of univariate data by cubic L_1 smoothing splines, Comput. Aided Geom. Design, 17 (2000), 715-727.
- [6] D. G. Schweikert, An interpolation curve using a spline in tension, Journal of Mathematics and Physics, 45 (1966), 312-317.
- [7] J. Stoer and B. Bulirsch, Introduction to Numerical Analysis, Springer-Verlag, New York, 1993.
- [8] R. S. Varga, Matrix Iterative Analysis, Prentice-Hall, Inc., Englewood Cliffs, New Jersey, 1962.
- [9] C. Witzgall, D. E. Gilsinn, M. A. McClain, An Examination of New Paradigms for Spline Approximation, journal of Research of the National Institute of Standards and Technology, 2006 (To Appear).