U.S. Department of Commerce
National Oceanic and Atmospheric Administration
National Weather Service
National Centers for Environmental Prediction
5830 University Research Court
College Park, MD 20740

Office Note 480
doi:10.7289/V59K485T

Recovery of aircraft vertical motion profiles from incomplete data – an application of the method of splines

R. James Purser*, Yanqiu Zhu
IM Systems Group, Rockville, Maryland

and Bradley A. Ballish
Environmental Modeling Center

November 21, 2014

* email: jim.purser@noaa.gov

Abstract

It has been determined that the biases in temperature reports from Aircraft Meteorological Data Relay (AMDAR) measurements exhibit a dependence, among other predictors, upon the vertical component, dh/dt, of the aircraft motion. While some aircraft report frequently and to the nearest second, and so provide data that are easily and directly applicable to the bias-correction problem, some other aircraft report infrequently and only to within the nearest minute, which makes it much harder to deduce their instantaneous vertical velocity component at each of their reporting times during their irregular ascending or descending flight trajectories. The problem of interpolating the likely vertical motion from such incomplete data seems suited to a solution by application of the method of numerical splines, which we describe in this short note.

1. Introduction

Experience with the assimilation of aircraft-derived temperature measurements, such as those derived from the Aircraft Meteorological Data Relay (AMDAR) system (Schwarz and Benjamin, 1995; Rienecker et al. 2011), has revealed that, among the predictors of measurement bias, are some of the aircraft flight parameters themselves (Ballish and Kumar, 2008; Dee and Uppala 2009). In particular, the aircraft altitude, h, and the vertical velocity component, dh/dt, at each time, t, are important predictors in this respect. While many aircraft report accurate altitudes at times accurate to the nearest second and at a frequent rate during the ascending and descending phases of flight, owing to historical bandwidth limitations, other aircraft report rather less frequently and at times report effectively only to the nearest minute. From these latter styles of reports it is problematic to extract the critical vertical motion information with sufficient accuracy and reliability to establish the proper bias corrections for the other reported data. Nevertheless, we must make the effort to deduce what we can from the limited data available, and one attractive approach to dealing with this problem is to exploit and adapt the classical curve-fitting method of numerical "splines" (Schoenberg, 1946; Curry and Schoenberg, 1947; de Boor, 1978; Wahba, 1990).

In this brief note we outline in the next section the method of deriving useful spline functions that accord with a variational principle and extend their application, in section 3, to situations where the constraints occur in the form of inequalities rather than the more common equality constraints at the nodes of the spline. We further show how the intrinsic "energy" of the spline, the quantity minimized in its variational definition, can be used to decide the choice of solution in cases, relevant to our aircraft trajectory problem, where several distinct plausible options present themselves for the route threading the inequality constraints. Finally, we discuss the possible extensions of the spline method that could be adopted for other kinds of constraints or data types.

2. The principles of splines

(*a*) *Equality constraints*

We approach the problem of constructing a spline from the standpoint of minimizing a suitable quadratic form on the derivatives. The traditional drafting spline is a thin elastic ruler, pinned by weights or pivots to pass through a preselected set of "knots" or "nodes", resulting in an elastostatic configuration in which the integrated elastic energy is minimized. The classical numerical idealization of this spline (Schoenberg 1946) is the curve constrained at the given nodes such that the energy idealization obtained by integrating the square of the second derivative of lateral displacement is minimized. The assumption being made here to justify the elastostatic analogy is that the lateral deflection, y, is always sufficiently small to make the first derivative negligible in the sense, $|dy/dx| \ll 1$. The mathematical theory of splines is covered in, for example, Ahlberg et al. (1967), de Boor (1978), while "splines in tension" were introduced by Schweikert (1966). Additional variational aspects and generalizations are discussed in Wahba (1990).

In the idealization in which the spline is considered to possess tension an additional component of energy needed to bend the spline into its constrained configuration is identified with the work done against the tensile force to gather the additional needed length, approximated in the numerical idealization by half the integral of the square of the first derivative, i.e., $\int \frac{1}{2}(dy/dx)^2\,dx$. In these numerical idealizations, we find that the classical (untensioned) spline is piecewise cubic polynomial, while the tensioned generalizations involves exponential functions, as we shall discover below from formal solutions of the energy integrals.

In general, the tensioned spline must minimize an energy proportional to

$$E = \int \frac{1}{2T^2}\left(\frac{dy}{dx}\right)^2 + \frac{1}{2}\left(\frac{d^2y}{dx^2}\right)^2 dx. \tag{2.1}$$

where the square of characteristic x-scale, T, of the spline emerges as the ratio between the spline's intrinsic "bending stiffness" and the imposed tension. But it must also respect the given constraints, which we can define at the m nodes, $\hat{x}_i$, as the list of constraints:

$$y(\hat{x}_i) = \hat{y}_i, \qquad \text{for} \qquad i = 1, \ldots m. \tag{2.2}$$

We can formally incorporate these constraints into the variational integral by means of delta-function contributions modulated by a set of Lagrange multipliers, J_i:

$$E = \int \left[\frac{1}{2T^2}\left(\frac{dy}{dx}\right)^2 + \frac{1}{2}\left(\frac{d^2y}{dx^2}\right)^2 + \sum_i^m (\hat{y}_i - y(x))J_i\delta(x-\hat{x}_i)\right] dx \tag{2.3}$$

The interpretation of the quantities, J_i, in the elastostatic analogy is that they are the transverse forces applied to the spline at the nodes, $\hat{x}_i$. The standard Euler-Lagrange procedure then tells us that the solution must obey the fourth-order elliptic equation:

$$-\frac{1}{T^2}\frac{d^2y}{dx^2} + \frac{d^4y}{dx^4} - \sum_i^m J_i\delta(x-\hat{x}_i) = 0, \tag{2.4}$$

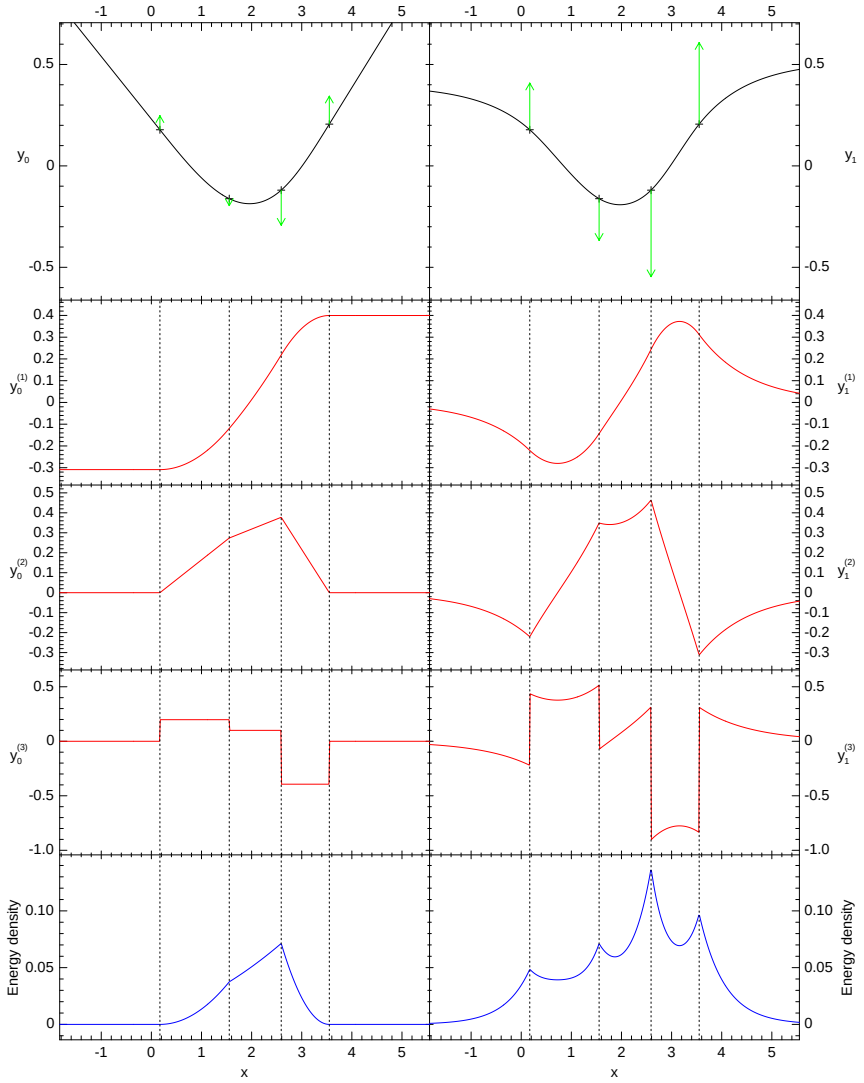

Figure 1. A comparision of the behavior of the untensioned cubic polynomial spline (left panels) and the standardized tensioned spline (right panels) for a simple case of four constraints (shown by the "+" symbols in the top panels. In successive rows we have the spline function, its 1st, 2nd and 3rd derivative, and the energy density profile. The arrows in the top panels indicate the "forces" at the constraints, that is, the jump in the 3rd-derivative at each node..

with the J_i set to make (2.2) true. In the generic case of positive finite T, which is the case of greater interest to us, a convenient simplification is achieved by expressing the independent variable, x, in units of T itself. If we assume this has been done, then the solution between any consecutive pair of the constraints takes the equivalent forms,

$$y(x) = a + bx + c\exp(x) + d\exp(-x), \tag{2.5}$$

or

$$y(x) = p + qx + rC(x) + sS(x), \tag{2.6}$$

where we define:

$$C(x) = \cosh(x) - 1 \tag{2.7a}$$
$$S(x) = \sinh(x) - x. \tag{2.7b}$$

In the limiting case where the spline tension vanishes, we do not rescale x but we achieve a more significant simplification through the omission of the first term in (2.3), and hence also the first term of (2.4). The hyperbolic, or exponential functions disappear and in their place we are left with their limiting forms, the cubic polynomials:

$$y(x) = p + qx + r\frac{x^2}{2} + s\frac{x^3}{6}. \tag{2.8}$$

(Another limiting case, where the bending stiffness, but not the tension, vanishes, leads to the piecewise linear solution we expect from what has effectively become tensioned string; this case is of no further interest to us.)

The qualitative distinction between the behavior of ordinary and tensioned splines is visible in the comparative plots of these splines and their derivatives shown in Fig. 1 and discussed in the caption. The tensioned spline is valued for its inherent suppression of large excursions in the broader gaps between constraints, as well as outside the spatial range of the constraints where, unlike the ordinary spline, any residual slope still incurs a penalty of energy (whose density is plotted in the bottom panels) and therefore causes the spline to "level off" exponentially.

As we observe by integrating (2.4), there appears a discontinuity in the third derivative of magnitude:

$$\left.\frac{d^3y}{dx^3}\right|_{\hat{x}_{i+}} - \left.\frac{d^3y}{dx^3}\right|_{\hat{x}_{i-}} = J_i, \tag{2.9}$$

at each node, where the coefficients sets, $\{a, b, c, d\}$ or $\{p, q, r, s\}$, change crossing each node, and where $\hat{x}_{i-}$ and $\hat{x}_{i+}$ denote the left and right side evaluations of the discontinuity in the function, crossing the node $\hat{x}_i$. The derivatives below third-order, as well as the value itself, remain continuous. For the energy to be minimized, it must certainly remain finite, and this is only possible when, for $x < \hat{x}_1$, the coefficients $b = 0$ and $d = 0$ (equivalently, $q = r = s$) while, for $x > \hat{x}_m$, the coefficients $b = 0$ and $c = 0$ (equivalently, $q = -r = s$). It is then easy to confirm that, if we adopt $\hat{x}_1$ for a local origin, the portion of the energy integral coming from $x < \hat{x}_1 = 0$ is just $q_1^2/2$, where q_1 is $dy/dx|_{x_1}$, and, likewise, adopting $\hat{x}_m$ as a local origin for the calculation of the energy integral over the other infinite segment, $x > x_m$, that portion of the integrated energy comes to $q_m^2/2$. The contributions to the energy integral coming from each interior segment is more complicated to evaluate, and we refer to the appendix for algebraic details. But we can immediately see that, in the generic interior segment $[\hat{x}_i, \hat{x}_{i+1}]$, the spline function's quartet of coefficients, say $\{p, q, r, s\}$, about a suitable origin, can be expressed as linear functions of the spline value and its first derivative at the two ends of this segment. Therefore the energy contribution, E_i, coming from this segment must be a quadratic expression in only these same four quantities. Moreover, the symmetries (energy unchanged by a constant change in y, and unchanged by mirror reflection of the spline function about the midpoint of the interval) substantially simplify the algebra since we know that terms in the spline values can

only come from each *difference*, $\hat{y}_{i+1} - \hat{y}_i$, multiplied by itself, or that difference multiplied by the local endpoint *sum*, $\hat{q}_i + \hat{q}_{i+1}$ of the spline derivatives, where we define $\hat{q}_i = dy/dx|_{\hat{x}_i}$. Once we know all the $\hat{q}_i$, then since we already know the $\hat{y}_i$ (which are given) it is straightforward to interpolate to interior values of $\hat{x}_i \le x \le \hat{x}_{i+1}$.

The spline equations are "solved" as a tridiagonal energy matrix linear system for the m-vector $\boldsymbol{q}$ of the components, $\hat{q}_i$, by seeking the minimum of the energy defined by:

$$E = \frac{1}{2}\boldsymbol{q}^T \boldsymbol{Q}\, \boldsymbol{q} + (\boldsymbol{\sigma q})^T \boldsymbol{R\delta}\, \boldsymbol{p} + \frac{1}{2}(\boldsymbol{\delta p})^T \boldsymbol{P}\, \boldsymbol{\delta p}, \tag{2.10}$$

whose solution involves the easily solved symmetric tridiagonal linear system:

$$\boldsymbol{Qq} = \boldsymbol{v}, \tag{2.11}$$

where

$$\boldsymbol{v} = -\boldsymbol{\sigma}^T \boldsymbol{R}\, \boldsymbol{\delta p}, \tag{2.12}$$

where $\boldsymbol{p}$ is the vector of nodal spline value components, $\hat{y}_i$, the bidiagonal operator, $\boldsymbol{\delta}$, is the difference matrix whose ith row comprises column elements at positions i and $i+1$ of -1 and $+1$ respectively, and where the bidiagonal operator, $\boldsymbol{\sigma}$, is the pairwise sum matrix whose ith row comprises column elements at positions i and $i+1$ of $+1$ and $+1$. The $\boldsymbol{R}$ and $\boldsymbol{P}$ are diagonal matrices of dimension $m-1$. The derivation of the components of $\boldsymbol{Q}$, $\boldsymbol{P}$ and $\boldsymbol{R}$ appears in the appendix. The energy E is computed along with vector, $\boldsymbol{q}$ of nodal derivatives, but, in the more general applications of the spline method governed by inequality constraints, dealt with in the next subsection, we also need the vector $\boldsymbol{J}$ of "jumps" in the 3rd derivative, as defined by (2.9). The components J_i may be obtained from the nearest three consecutive q_i and two Δp_i by means of a formula also given in the appendix.

In the special limiting case obtained when the tension is allowed to vanish, there is no rescaling of x to be done, and the computations further simplify with piecewise cubic polynomial solutions. This case is also dealt with in the appendix. There is another special case – where the spline stiffness, relative to the tension, is allowed to become negligible, and in this case the spline acts like tensioned string with a simpler solution that is piecewise linear between nodes and constant beyond both ends. Since, in this limit, the derivative is not uniquely defined at nodes (where we need estimates of it), this special case is hardly of direct interest to us, although, because it involves considerably simpler computations, it could be of value in suggesting the form of likely solutions in the more complicated spline solutions constrained by a series of inequalities, which we discuss next.

(*b*) *Inequality constraints*

We have already noted that the J_i are interpreted as the transverse "forces" needed to keep the spline passing through all the active constraints. Clearly therefore, if we replace equality constraints, $p_i = \hat{y}_i$, by inequality constraints of the form,

$$(p_i - \hat{y}_i) b_i \ge 0, \tag{2.13}$$

where the b_i are a series of "signs", ± 1, we expect the force J_i to vanish whenever $p_i \ne \hat{y}_i$ in the solution, and the sign of J_i, if nonzero, to match b_i whenever the constraint is "activated" in the

sense that $p_i = \hat{y}_i$ in the solution. Such inequalities are particularly simple examples of linear inequality constraints and, if any spline solution can be found that is consistent with all these constraints, then there is certainly one amongst these solutions that also minimizes the energy. A spline conforming to the inequalities (but not necessarily the optimal one) is said to be "feasible". This is standard technical usage of the word in general mathematical optimization under inequality constraints (see, for example, Rockafellar 1970). A viable solution algorithm is one that starts with a guaranteed feasible spline and, step-wise, switches the inequality constraints between active "on" and inactive "off" mode in such a way that feasibility is maintained and spline energy progressively decreases. One algorithm that follows this prescription will now be described.

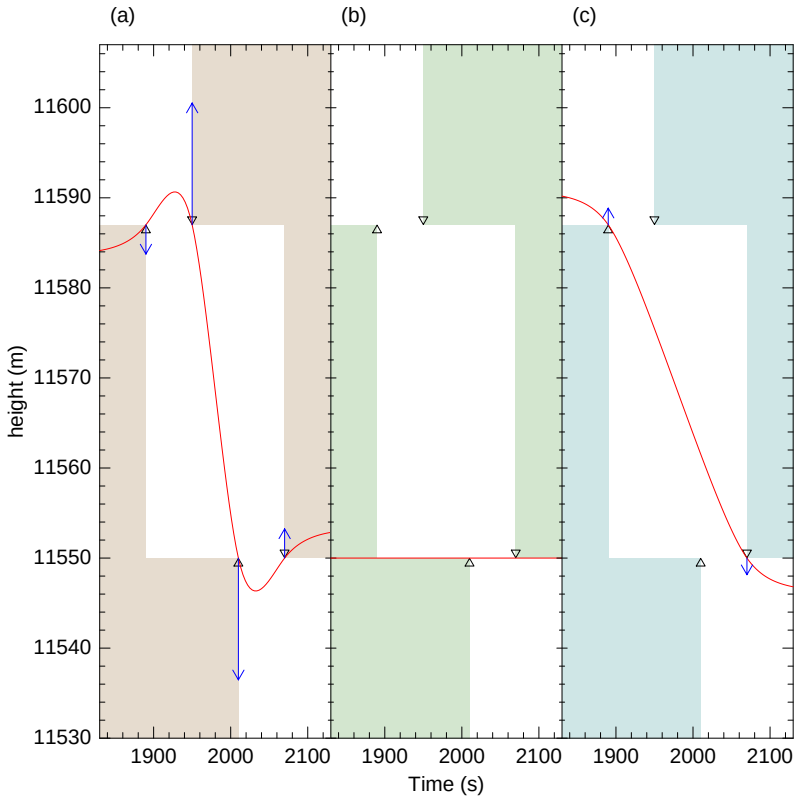

Figure 2. Stages in the solution of a tensioned spline problem constrained by inequalities equivalent to a succession of slalom gates, here shown for a trivial case of just two gates. Initially (panel (a)), the spline is fitted to as many compatible constraints as possible and the signs of each 3rd-derivative "jump" (the blue arrows) are compared to the signs implied by the inequalities (indicated by the oriented triangles). In this example, all the signs are wrong, so all but one of the constraints are switched off (b). The new spline fails to pass through the first gate, the positional inequality being wrong for the first gatepost. The constraint associated with the first gatepost is switched back on, and the new and final solution, panel (c), is found to obey all the inequalities (the directions of the blue arrows conform with those of the oriented triangles where the constraints are "on", while the positional relations elsewhere are also of the correct sense).

First we step along the abscissa of increasing x and, as soon as we encounter an inequality constraint at a *new* value of x, we force the activation of *one* constraint there. Thus, there may be some potential nodes that are duplicated, $\hat{x}_i = \hat{x}_j$, for $i \neq j$, in which case, we activate only

the constraint i as an equality condition, $p_i = \hat{y}_i$ and leave constraint j as a satified inequality. But for every x that *is* a node, at least one constraint there is active at the start of the procedure. We solve the spline, that is, we minimize the energy under the prescribed equality constraints and obtain our starting feasible, though generally not necessarily optimal, spline solution. The crucial step in this "A" iteration is to examine all the quantities, $b_i J_i$, for those constraints that are "on" and determine whether any are negative. If not, we are done – our solution is both feasible and energetically optimal with the jump conditions (in the third derivative) all of the correct sign (or possibly zero in degenerate special cases).

But let us suppose that there are violations of the jump conditions found during this "A" iteration. First, we record the actual values, say, $p_i^{(A)} = y(\hat{x}_i)$, of this feasible spline at all of the nodes, $\hat{x}_i$. Then we deactivativate, i.e., switch "off", those constraints just found to have violated the jump conditions and enter what we refer to as the "B" iteration. In the "B" iteration, we solve for what is possibly *not* a feasible spline, for which the nodal values are now $p_i^{(B)} = y(\hat{x}_i)$, and we examine those constraints that are switched "off" to see whether any of these $p_i^{(B)}$ violate the prescribed positional inequalities. If not, we exit the B-iteration. But if a positional inequality is violated, there is at least some largest positive "under-relation" factor, $u < 1$, such that the affine mixture of the two most recent splines is once again feasible; this new feasible spline replaces the one with the nodal values designated $p_i^{(A)}$:

$$p_i^{(A')} = p_i^{(A)} + u[p_i^{(B)} - p_i^{(A)}], \tag{2.14}$$

and, since the spline functions are linear, their derivatives, q_i, at all the nodes are similarly interpolated:

$$q_i^{(A')} = q_i^{(A)} + u[q_i^{(B)} - q_i^{(A)}]. \tag{2.15}$$

The largest such u able to recover the property of "feasibility" causes equality, $p_j^{(A')} = \hat{y}_j$ for one of the hitherto "off" constraints, j, which is therefore activated (or reactivated) to "on". With the new feasible state designated as the new A-spline, the B-iteration is repeated. Thus, the B-iterations switch constraints "on" (but only one at a time) while the A-iterations can only switch constraints "off". Each successive return to the A-loop in this procedure diminishes the spline energy, so a convergence is guaranteed.

The interplay between the A-loop and B-loop steps is illustrated in the sequence shown in Fig. 2. Here there are two "slalom gates" through which the spline is required to descend, so this is a relatively trivial case.

In a less trivial case we can look at the effect of changing the time scale parameter, T for the tensioned splines: The only difference between the left and right panel of Fig. 3 is that, on the left we use $T = 300s$ and on the right, $T = 30s$. Clearly, the effects for this relatively straightforward profile are minimal in this case, except in the two regions of extrapolation within the exterior segments, bottom and top.

In descending profiles, we might more often see a pause in the change of altitude, and in such cases, the differences between choices of the time constant tend to be more pronounced.

Figure 4 illustrates significant sensitivity of the time constant, T. At the two altitudes (near 7000m and 3000m) of brief hesitation in the otherwise monotonic descent, the spline solution with the longer time constant (left panel) "bounces" in order to keep the acceleration

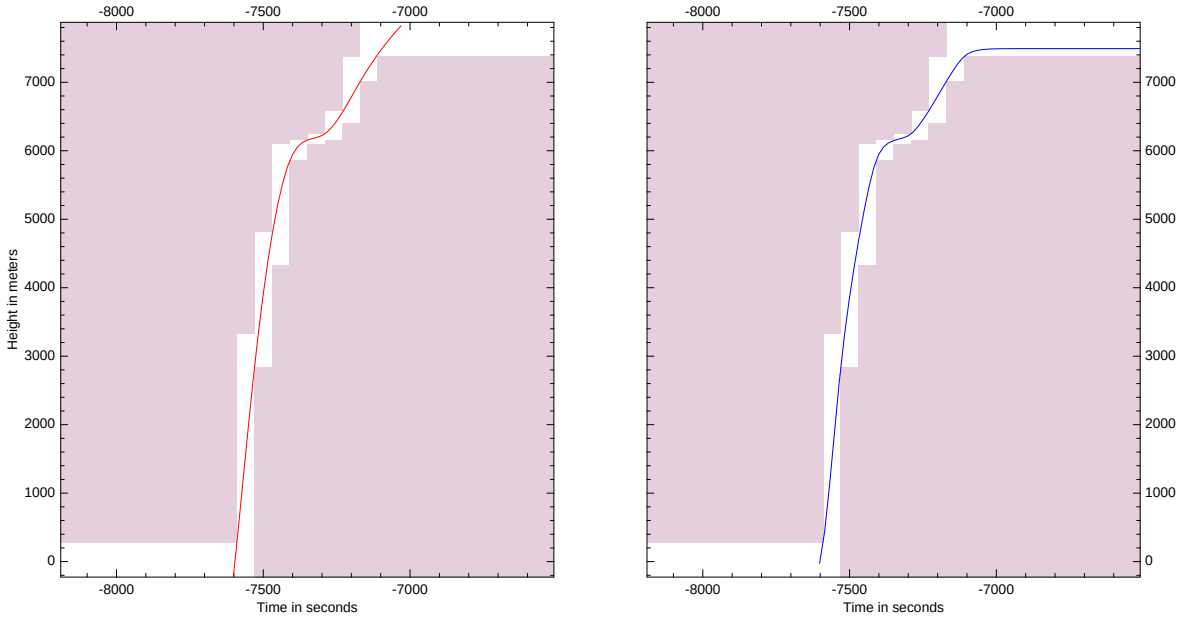

Figure 3. Tensioned spline problem for an ascending flight. The left panel shows the results when the time constant of the spline is $T = 300$ seconds; on the right is the corresponding solution with $T = 30$ seconds.

relatively small. When the tension is increased, and hence the time constant T diminished (right panel), the integrated acceleration becomes larger, the direction changes becoming more abrupt, as the solution is found that maintains a smaller integrated squared-vertical velocity, but the hesitation in the descent near 7000m is no longer a feature of the interpolated profile; it was probably merely an artifact in the case (left panel) with larger $T = 300s$.

(*c*) *Route combinatorics and final route selection*

A prerequisite for the solution of a given slalom spline is a specification of the "mode of passage", the sequence of binary alternatives of either descent or ascent, through each of the slalom gates. The complete set of geometrically possible combinations in a flight with many gates becomes too numerous for it to be practical to test all of them. Fortunately, we seldom err in using the heuristic decision to force "ascent"/"descent" upon the middle gate of any consecutive three whose relative altitudes reveal them to be in unequivocal ascending/descending order. But there are questionable cases, such as the gates at the beginning and end of a flight, where the mode of passage is ambiguous.

In those cases where the slalom gates follow directly from one to the next without intermission, geometrical considerations limit the number of combinations possible, regardless of the above-mentioned heuristic restrictions on ascent or descent. These geometrical restrictions can be conveniently coded, for each gate, as a two-digit trinary number, which we refer to as the "route code", with values between 0 and 8, whose trinary digits are referred to as the "option codes" (with values from 0, 1, or 2). The code works as follows. If the preceding gate is threaded in the descending sense then the geometrical or heuristic constraint is encoded in

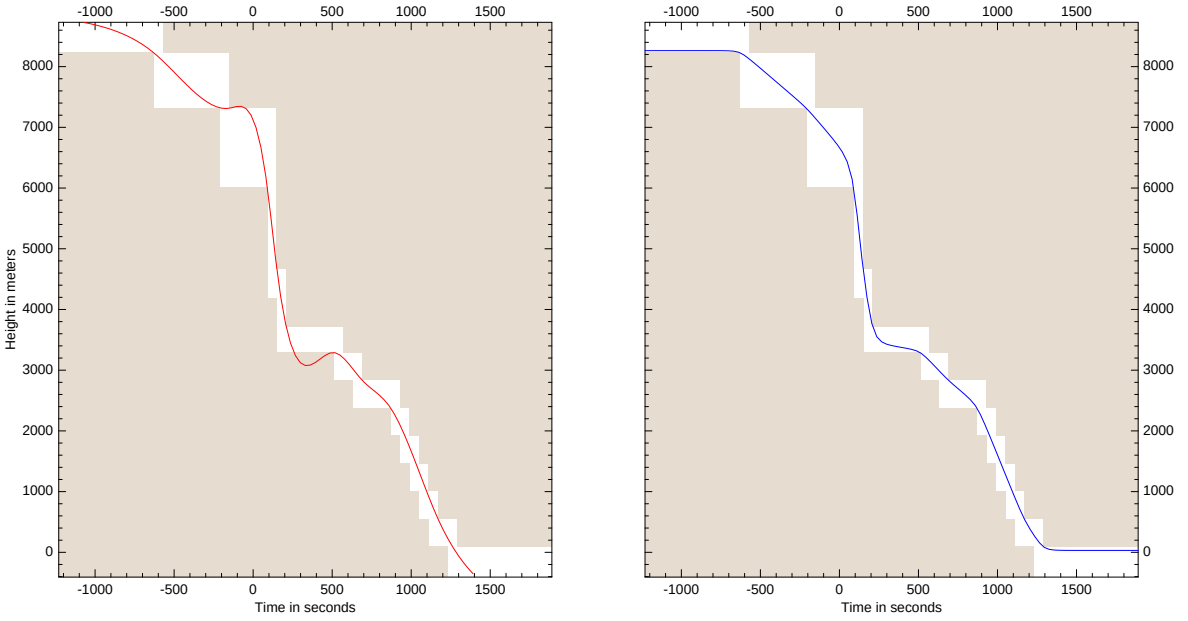

Figure 4. Tensioned spline problem for a descending flight. The left panel shows the results when the time constant of the spline is $T = 300$ seconds; on the right is the corresponding solution with $T = 30$ seconds.

the option code, O_D; alternatively, if the preceding gate is threaded in the ascending sense, the constraints become those encoded by option code, O_A. If an option code is 0, then, when it is in effect, passage through the present gate can be in either sense; when it is 1, passage must be descending through the present gate; when it is 2, passage must be ascending though the present gate. The route code for each gate is just,

$$C = O_D + 3 \times O_A. \tag{2.16}$$

If there is no restriction, regardless of the state of passage through the preceding gate, or if there is no preceding gate (as at the beginning of a flight), then both option codes are zero and so is the route code, C. If the heuristic considerations led us to dictate passage in the descending sense, regardless of the sense in which the preceding gate was threaded, then the resulting route code must be $C = 1 + 3 \times 1 = 4$; conversely, dictating ascent unconditionally means specifying that $C = 2 + 3 \times 2 = 8$. The less trivial examples come from the cases where consecutive gates, each possibly containing several distinct height measurements to give the gate depth (in height) as well as breadth (in time), are illustrated in the figures, 5–8. In Fig. 5 we see consecutive and adjoining gateways with a gap in height between them, the second being higher than the first. We are interested in the route code for the second gateway. If the sense of passage through the first gateway is upwards, as it is in the first two panels, the trajectory can either ascend (first panel) or descend (second panel) through the second gateway, so the "ascending" option code is $O_A = 0$. But if the threading of the first gateway is "descending", as in the third panel, the mode of passage through the second gateway can only be upwards, so the "descending" option code is $O_D = 2$. The result is a route code for this second gateway of $C = O_D + 3 \times O_A = 2$.

The situation in Fig. 6 is just the height-reversed image of that in Fig. 5 and the analogous arguments determine that, with $O_D = 0$ and $O_A = 1$, the route code is $C = 0 + 3 \times 1 = 3$.

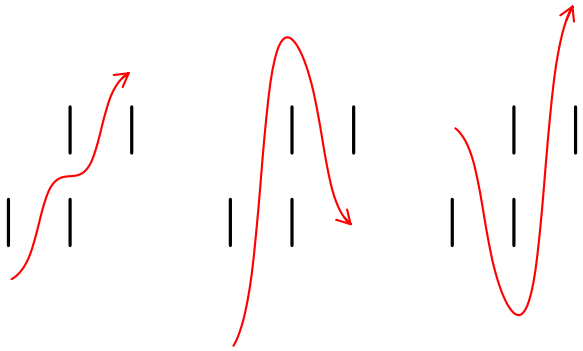

Figure 5. The three geometrically possible routes through a pair of adjoining consecutive gateways where the second resides wholly above the first. The route code is $C = 2$, as explained in the text.

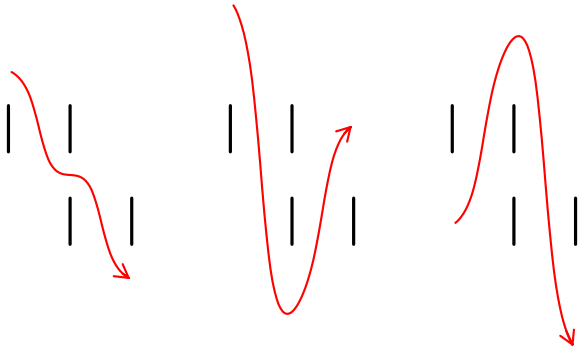

Figure 6. The three geometrically possible routes through a pair of adjoining consecutive gateways where the second resides wholly below the first. The route code is $C = 3$ in this case.

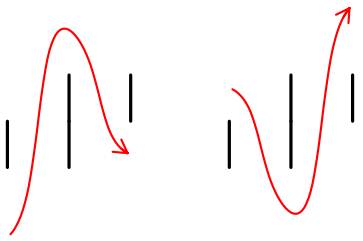

Figure 7. The two geometrically possible routes through a pair of adjoining consecutive gateways. The second gateway lies mainly above the first but since they overlap or touch in height, there is no gap that would allow passage in the same sense through both gateways. The route code is therefore $C = 5$, as explained in the text.

There can be cases where consecutive gateways adjoin in time and overlap in their heights. Fig. 7 shows such an example, the second gateway being mostly at a higher altitude but without the definite gap in height that would allow a trajectory to slip through gateways ascending

through both. In this case descent through the first implies ascent through the second, and vice versa, so the option codes are $O_D = 2$ and $O_A = 1$ and the route code is $C = 5$. Fig. 8 shows that, although we might reverse of the height relationships, we do not alter the form of the geometrical restriction, and the route code remains $C = 5$. In fact, one or other of the gateways' height range might completely overlap that of its neighbor and the geometrical forced alternation of the mode of passage would still be in effect.

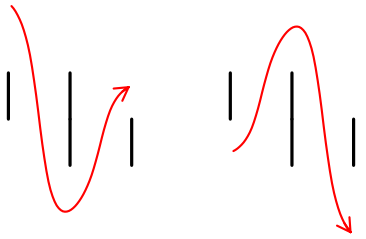

Figure 8. Although the height relationship between the two consecutive and adjoining gateways is the opposite of that shown in Fig 7, the same geometrical restrictions apply and the route code is, again, $C = 5$.

When the sequence of route codes are all determined, it is a very straightforward matter to step through each allowable route systematically, find the slalom spline solution for that route, together with its spline-energy, and finally to select the single route and associated spline solution that globally minimizes the spline energy. Moreover, in practice, the number of combinations we are obliged to test in this way rarely exceed a modest handful (most typically four), since heuristic decisions force either uniform descent or ascent in the interior portions of almost all the flights we encounter in these presorted aircraft datasets.

The natural question arises: how well does the slalom spline provided with 60 second-wide gates artificially constructed from data with, say, 6 second-wide tolerances, match the trajectory and vertical velocity component of the spline applied directly to the more informative data? In most of the smooth and monotonic ascents and descents the fit seems to be very good. Fig. 9 provides a real data case of an ascent (left panel) where, near the top of the climb, the hestitation caught by the finer data (blue curve) is not quite captured after coarsening (red curve). In such a case, it is very hard to see how it would be possible, even in principle, to recover this missing feature and its accompanying vertical velocity profile. On the other hand, for the descent case shown in the right panel the match between the two estimated trajectories is remarkably good. But while this result is encouraging as a validation of our spline approach, the value of acquiring the more precise temporal data if possible, is also convincingly confirmed by this exercise.

3. Discussion and conclusions

We have formulated and described a method of recovering aircraft vertical trajectory profiles from data assumed accuracte in height but incomplete in time, using a method of tensioned splines and the selection criterion of minimizing the spline's "energy" to identify the best topology of trajectory consistent with these equality and inequality constraints. The algorithm appears to be perfectly robust and its numerical efficiency has so far proven to be adequate.

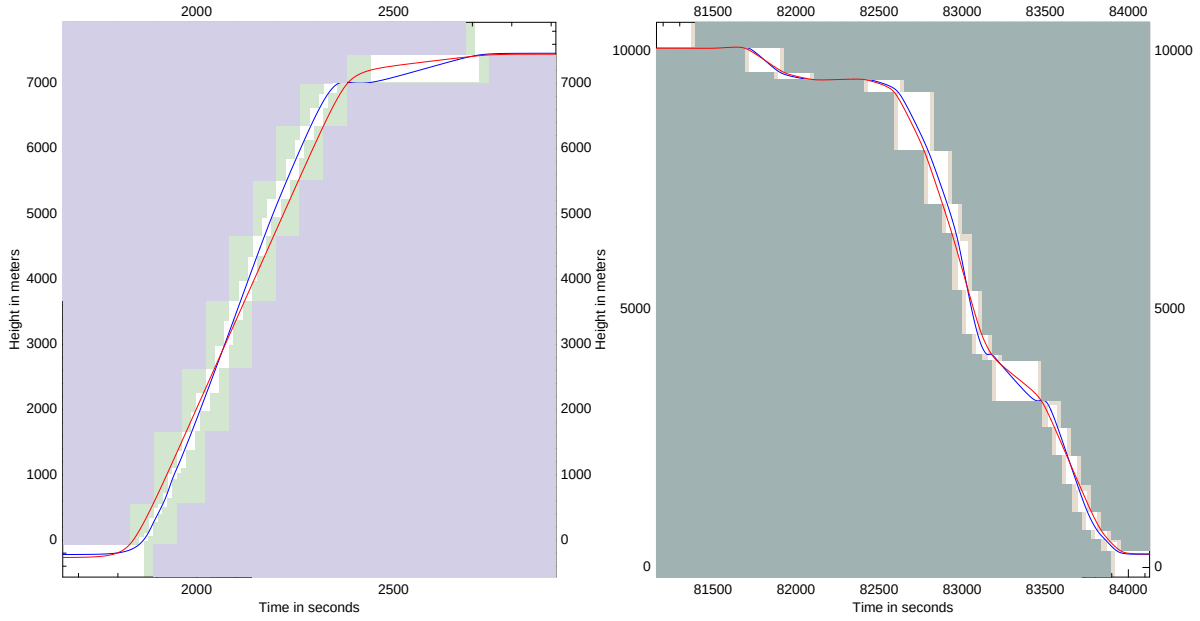

Figure 9. Two cases (ascending and descending) where the data given initially at the nearest second are artificially degraded to sparser data known only to the nearest minute. The "forbidden" regions of the respective slalom splines are shaded in different tones and superimposed. The curves fitting the original fine data, with gates of width 6s, are drawn in blue; the curves threading the coarser slalom gates following the artifial degradation of the temporal information to gates of width 60s are drawn in red.

However, should it ever become necessary to process much longer trajectory segments than those that are typical of present operations, it might become necessary to refine the present algorithm in order to circumvent the likely growth in the number of combinations needing to be tested. This could easily be done by dividing a long trajectory into short overlapping sections, with a relatively minor modification to the slalom spline solver to accommodate the inevitable situation where intermediate reconnected trajectory constraint configurations will not generally correspond to "feasible" splines (i.e., they will generally not correctly thread all the gateways when we simply concatenate consecutive sections of constraints). If a "recapture" algorithm is appended to the suite to deal with this kind of problem, then it might also be worth investigating whether a compuational speed-up can be obtained by finding an interim solution to each spline problem using the more conventional untensioned (cubic) splines to provide a close "first guess" to the final tensioned spline solver. The cubic splines are algebraically simpler and their computional solution times are found to be approximately two and a half times faster than those of the equivalent tensioned splines (algebraic function evaluations being cheaper than evaluations of hyperbolic or exponential functions).

We have assumed in these estimation and interpolation procedures that, although the time data are uncertain (plus or minus thirty seconds, typically) the height data are precise. Strictly, it is closer to the spirit of conventional data assimilation to assume that the height data themselves contain some random error. So another future generalization which might be desirable to investigate is the case where the constraints at "gateposts" remain one-sided, but are also progressive in their effect (i.e., conditional "weak" constraints). The implicit "switches" ac-

tivating or deactivating the constraints would remain, but the character of the optimization, once these switches are set, would then change to one without any Lagrange formal multiplier variables (which would instead have become absorbed in a more purely quadratic variational minimization).

Acknowledgments

This paper benefitted from the helpful comments of Drs. Andrew Collard, Jacob Carley and Paul Van Delst.

Appendix A

Constructing the coefficients of the linear system defining the spline

We shall imagine the abscissa x of the spline function to run from left to right and consider first the spline solution in the semi-infinite segment left of $\hat{x}_1$. If we redefine the x coordinate so that this first node is the origin, $\hat{q}_1$ is the derivative, $y^{(1)} \equiv dy/dx$, there. Then, defining higher derivatives similarly, the form of the spline function and its first two derivatives in this segment when x is scaled to make $T = 1$ must be:

$$\begin{aligned} y(x) &= \hat{p}_1 + \hat{q}_1(\exp(x) - 1) && \text{(A.1a)} \\ y^{(1)}(x) &= \hat{q}_1 \exp(x) && \text{(A.1b)} \\ y^{(2)}(x) &= \hat{q}_1 \exp(x) && \text{(A.1c)} \\ y^{(3)}(x) &= \hat{q}_1 \exp(x), && \text{(A.1d)} \end{aligned}$$

and hence the energy integral in the "segment 0" is just:

$$\begin{aligned} E_0 &= \int_{-\infty}^{0} \frac{1}{2}(y^{(1)})^2 + \frac{1}{2}(y^{(2)})^2 \, dx \\ &= \hat{q}_1^2 \int_{-\infty}^{0} \exp(2x) \, dx \\ &= \frac{1}{2}\hat{q}_1^2. && \text{(A.2)} \end{aligned}$$

Similarly, in the semi-infinite segment to the right of the final node, $\hat{x}_m$, we can also redefine the coordinate x to make this end node the origin, whereupon the general form of the spline function passing through $\hat{p}_m = \hat{y}_m$ is:

$$\begin{aligned} y(x) &= \hat{p}_m + \hat{q}_m(1 - \exp(-x)), && \text{(A.3a)} \\ y^{(1)}(x) &= \hat{q}_m \exp(-x), && \text{(A.3b)} \\ y^{(2)}(x) &= -\hat{q}_m \exp(-x), && \text{(A.3c)} \\ y^{(3)}(x) &= \hat{q}_m \exp(-x), && \text{(A.3d)} \end{aligned}$$

and the energy contribution of this segment is

$$E_m = \frac{1}{2}\hat{q}_m^2. \qquad \text{(A.4)}$$

The equation of the standardized tensioned spline between nodes, and the corresponding energy contribution, are more complicated to compute, but we start by resetting the coordinate origin to the segment midpoint so that we can exploit symmetries. We suppose the half-width of the segment to be X_i. It is algebraically convenient to introduce special functions:

$$\begin{aligned} C(x) &\equiv \cosh(x) - 1, && \text{(A.5a)} \\ S(x) &\equiv \sinh(x) - x, && \text{(A.5b)} \\ K(x) &\equiv xC(x) - S(x) = x\cosh(x) - \sinh(x), && \text{(A.5c)} \end{aligned}$$

and write the equation of the spline and its derivatives in segment i between nodes $\hat{x}_i = -X_i$ and $\hat{x}_{i+1} = +X_i$ as:

$$\begin{aligned} y(x) &= p + qx + rC(x) + sS(x), && \text{(A.6a)} \\ y^{(1)}(x) &= q + r\sinh(x) + sC(x), && \text{(A.6b)} \\ y^{(2)}(x) &= r\cosh(x) + s\sinh(x), && \text{(A.6c)} \\ y^{(3)}(x) &= r\sinh(x) + s\cosh(x). && \text{(A.6d)} \end{aligned}$$

We define $p_- = \hat{p}_i = y(-X_i)$, $p_+ = \hat{p}_{i+1} = y(X_i)$, $q_- = \hat{q}_i = y^{(1)}(-X_i)$, $q_+ = \hat{q}_{i+1} = y^{(1)}(X_i)$. We define consecutive sums and differences:

$$\begin{aligned} (\sigma p)_i &\stackrel{\text{def}}{=} p_- + p_+, && \text{(A.7a)} \\ (\delta p)_i &\stackrel{\text{def}}{=} p_+ - p_-, && \text{(A.7b)} \\ (\sigma q)_i &\stackrel{\text{def}}{=} q_- + q_+, && \text{(A.7c)} \\ (\delta q)_i &\stackrel{\text{def}}{=} q_+ - q_-, && \text{(A.7d)} \end{aligned}$$

the spline function in this interval can be expressed equivalently as the sum of functions with either odd or even symmetry:

$$y(x) = \frac{(\sigma p)_i}{2} + \frac{(\delta p)_i F(x) + (\sigma q)_i G(x)}{2K(X_i)} + \frac{(\delta q)_i H(x)}{2\sinh(X_i)}, \qquad \text{(A.8)}$$

where we define odd functions F and G and even function H, and their derivatives, by:

$$\begin{aligned} F(x) &= xC(X_i) - S(x), && \text{(A.9a)} \\ F^{(1)}(x) &= C(X_i) - C(x), && \text{(A.9b)} \\ F^{(2)}(x) &= -\sinh(x), && \text{(A.9c)} \\ F^{(3)}(x) &= -\cosh(x). && \text{(A.9d)} \end{aligned}$$

$$\begin{aligned} G(x) &= XS(x) - xS(X_i), && \text{(A.10a)} \\ G^{(1)}(x) &= XC(x) - S(X_i), && \text{(A.10b)} \\ G^{(2)}(x) &= X\sinh(x), && \text{(A.10c)} \\ G^{(3)}(x) &= X\cosh(x). && \text{(A.10d)} \end{aligned}$$

$$
\begin{aligned}
H(x) &= C(x) - C(X_i), && \text{(A.11a)}\\
H^{(1)}(x) &= \sinh(x), && \text{(A.11b)}\\
H^{(2)}(x) &= \cosh(x), && \text{(A.11c)}\\
H^{(3)}(x) &= \sinh(x). && \text{(A.11d)}
\end{aligned}
$$

The terms that contribute to the segment energy,

$$
E_i = \frac{1}{2}\int_{-X_i}^{X_i} y^{(1)}y^{(1)} + y^{(2)}y^{(2)}\,dx, \tag{A.12}
$$

for the interior segment reduce to:

$$
E_i = \frac{I_{FF}(\delta p)_i^2}{4K^2(X_i)} + \frac{I_{FG}(\delta p)_i(\sigma q)_i}{2K^2(X_i)} + \frac{I_{GG}(\sigma q)_i^2}{4K^2(X_i)} + \frac{I_{HH}(\delta q)_i^2}{4\sinh^2(X_i)}, \tag{A.13}
$$

where we define the four integrals,

$$
\begin{aligned}
I_{FF} \equiv \frac{1}{2}\int_{-X_i}^{X_i} F^{(1)}(x)F^{(1)}(x) + F^{(2)}(x)F^{(2)}(x)\,dx &= \cosh(X_i)K(X_i), && \text{(A.14a)}\\
I_{FG} \equiv \frac{1}{2}\int_{-X_i}^{X_i} F^{(1)}(x)G^{(1)}(x) + F^{(2)}(x)G^{(2)}(x)\,dx &= -\sinh(X_i)K(X_i), && \text{(A.14b)}\\
I_{GG} \equiv \frac{1}{2}\int_{-X_i}^{X_i} G^{(1)}(x)G^{(1)}(x) + G^{(2)}(x)G^{(2)}(x)\,dx &= X\sinh(X_i)K(X_i), && \text{(A.14c)}\\
I_{HH} \equiv \frac{1}{2}\int_{-X_i}^{X_i} H^{(1)}(x)H^{(1)}(x) + H^{(2)}(x)H^{(2)}(x)\,dx &= \cosh(X_i)\sinh(X_i). && \text{(A.14d)}
\end{aligned}
$$

Hence,

$$
E_i = \frac{(\delta p)_i^2\cosh(X_i)}{4K(X_i)} - \frac{(\delta p)_i(\sigma q)_i\sinh(X_i)}{2K(X_i)} + \frac{(\sigma q)_i^2 X\sinh(X_i)}{4K(X_i)} + \frac{(\delta q)_i^2\cosh(X_i)}{4\sinh(X_i)}, \tag{A.15}
$$

and we can identify how these terms fit into the components of diagonal matrices $\boldsymbol{P}$ and $\boldsymbol{R}$ of section 2:

$$
\begin{aligned}
\boldsymbol{P}_i &= \frac{\cosh(X_i)}{2K(X_i)}, && \text{(A.16a)}\\
\boldsymbol{R}_i &= -\frac{\sinh(X_i)}{2K(X_i)}. && \text{(A.16b)}
\end{aligned}
$$

The elements of the tridiagonal, $\boldsymbol{Q}$, combining these interior contributions with the exterior contributions (A.2) and (A.4) can now be seen to be:

$$
\begin{aligned}
Q_{i,i} &= \hat{Q}_{i-1} + \hat{Q}_i, \qquad 1 \le i \le m && \text{(A.17a)}\\
Q_{i,i+1} = Q_{i+1,i} &= \tilde{Q}_i, \qquad 1 \le i < m, && \text{(A.17b)}
\end{aligned}
$$

where,

$$
\hat{Q}_i = \begin{cases} 1 & : \quad i = 0, \quad i = m \\ \frac{X_i\sinh(X_i)}{2K(X_i)} + \frac{\cosh(X_i)}{2\sinh(X_i)} & : \quad 0 < i < m \end{cases}, \tag{A.18}
$$

and

$$\tilde{Q}_i = \frac{\cosh(X_i)}{2\sinh(X_i)} - \frac{X_i \sinh(X_i)}{2K(X_i)}. \tag{A.19}$$

The untensioned (cubic polynomial) splines can be obtained as the formal limit of the tensioned splines as the tension goes to zero; in this case we can make the formal substitutions in the expressions above to obtain to correct formulae in this limit (and with no scaling of x):

$$\begin{aligned}
\cosh(x) &\rightarrow 1, && \text{(A.20a)}\\
\sinh(x) &\rightarrow x, && \text{(A.20b)}\\
C(x) &\rightarrow x^2/2, && \text{(A.20c)}\\
S(x) &\rightarrow x^3/6, && \text{(A.20d)}\\
K(x) &\rightarrow x^3/3. && \text{(A.20e)}
\end{aligned}$$

From these limiting forms we deduce that, for the purely cubic splines,

$$\begin{aligned}
F(x) &= \frac{3X_i^2 x - x^3}{6}, && \text{(A.21a)}\\
F^{(1)}(x) &= \frac{X_i^2 - x^2}{2}, && \text{(A.21b)}\\
F^{(2)}(x) &= -x, && \text{(A.21c)}\\
F^{(3)}(x) &= -1. && \text{(A.21d)}
\end{aligned}$$

$$\begin{aligned}
G(x) &= \frac{X_i x^3 - X_i^3 x}{6}, && \text{(A.22a)}\\
G^{(1)}(x) &= \frac{3X_i x^2 - X_i^3}{6}, && \text{(A.22b)}\\
G^{(2)}(x) &= X_i x, && \text{(A.22c)}\\
G^{(3)}(x) &= X_i. && \text{(A.22d)}
\end{aligned}$$

$$\begin{aligned}
H(x) &= \frac{x^2 - X_i^2}{2}, && \text{(A.23a)}\\
H^{(1)}(x) &= x, && \text{(A.23b)}\\
H^{(2)}(x) &= 1, && \text{(A.23c)}\\
H^{(3)}(x) &= 0. && \text{(A.23d)}
\end{aligned}$$

The functional form taken by y in the interval becomes:

$$y(x) = \frac{(\sigma p)_i}{2} + \frac{(\delta p)_i(3X_i^2 x - x^3) + (\sigma q)_i(X_i x^3 - X_i^3 x)}{4X_i^3} + \frac{(\delta q)_i(x^2 - X_i^2)}{4X_i}. \tag{A.24}$$

The integrals that contribute to the energy in this limiting case become:

$$I_{FF} = \frac{X_i^3}{3}, \tag{A.25a}$$

$$I_{FG} = -\frac{X_i^4}{3}, \tag{A.25b}$$

$$I_{GG} = \frac{X_i^5}{3}, \tag{A.25c}$$

$$I_{HH} = X_i. \tag{A.25d}$$

The energy integrated over interior segment i is the simplified form:

$$E_i = \frac{1}{2}\int_{-X_i}^{X_i} y^{(2)} y^{(2)}\, dx, \tag{A.26}$$

which evaluates to:

$$E_i = \frac{3(\delta p)_i^2}{4X_i^3} - \frac{2(\delta p)_i(\sigma q)_i}{3X_i^2} + \frac{3(\sigma q)_i^2}{4X_i} + \frac{(\delta q)_i^2}{4X_i}. \tag{A.27}$$

Now,

$$\boldsymbol{P}_i = \frac{3}{2X_i^3}, \tag{A.28a}$$

$$\boldsymbol{R}_i = -\frac{3}{2X_i^2}. \tag{A.28b}$$

In this case,

$$\hat{Q}_i = \begin{cases} 0 & : \quad i = 0 \quad \text{or} \quad i = m \\ \frac{2}{X_i} & : \quad 0 < i < m \end{cases}, \tag{A.29}$$

and

$$\tilde{Q}_i = \frac{1}{X_i}. \tag{A.30}$$

References

Ahlberg, J. H., E. N. Nilson, and J. L. Walsh — 1967 — *The Theory of Splines and their Applications.* Academic Press, New York.

Ballish, B., and K. Kumar — 2008 — Systematic Differences in Aircraft and Radiosonde Temperatures. *Bull. Amer. Meteor. Soc.,* **89,** 1689–1708.

Curry, H. B., and I. J. Schoenberg — 1947 — On spline distributions and their limits: the Pólya distribution functions, Abstract 380t, *Bull. Amer. Math. Soc.,* **53,** 1114.

de Boor, C. — 1978 — *A Practical Guide to Splines.* Springer, New York. 392 pp.

Dee, D.P., and S. Uppala — 2009 — Variational bias correction of satellite radiance data in the ERA-Interim reanalysis. *Quart. J. R. Meteor. Soc.,* **135,** 1830–1841.

Rienecker, M. M., M. J. Suarez, R. Gelaro, R. Todling, J. Bacmeister, E. Liu, M. G. Bosilovich, S. D. Schubert, L. Takacs, G.-K. Kim, S. Bloom, J. Chen, D. Collins, A. Conaty, A. da Silva, et al. — 2011 — MERRA - NASA's Modern-Era Retrospective Analysis for Research and Applications. *J. Climate,* **24,** 3624–3648, doi:10.1175/JCLI-D-11-00015.1. Link. [Part of AMS MERRA Special Collection]

Rockafellar, R. T.	1970	*Convex Analysis.* Princeton University Press. 451 pp.
Schoenberg, I. J.	1946	Contributions to the problem of approximation of equidistant data by analytic functions. *Quart. Appl. Math.,* **4,** 45–99, 112–141.
Schwartz, B. and S. G. Benjamin	1995	A comparison of temperature and wind measurements from ACARS-Equipped aircraft and rawinsondes. *Wea. Forecasting,* **10,** 528–544.
Schweikert, D. G.	1966	An interpolation curve using a spline in tension, *J. Math. Phys.,* **45,** 312–317.
Wahba, G.	1990	*Spline Models for Observational Data*, SIAM. CBMS-NSF Regional Conference Series in Applied Mathematics, Vol. 59, 165 pp.