34

Blossoms Are Polar Forms

by Lyle Ramshaw

January 2, 1989

Systems Research Center
130 Lytton Avenue
Palo Alto, California 94301

Systems Research Center

DEC's business and technology objectives require a strong research program. The Systems Research Center (SRC) and three other research laboratories are committed to filling that need.

SRC began recruiting its first research scientists in 1984 — their charter, to advance the state of knowledge in all aspects of computer systems research. Our current work includes exploring high-performance personal computing, distributed computing, programming environments, system modelling techniques, specification technology, and tightly-coupled multiprocessors.

Our approach to both hardware and software research is to create and use real systems so that we can investigate their properties fully. Complex systems cannot be evaluated solely in the abstract. Based on this belief, our strategy is to demonstrate the technical and practical feasibility of our ideas by building prototypes and using them as daily tools. The experience we gain is useful in the short term in enabling us to refine our designs, and invaluable in the long term in helping us to advance the state of knowledge about those systems. Most of the major advances in information systems have come through this strategy, including time-sharing, the ArpaNet, and distributed personal computing.

SRC also performs work of a more mathematical flavor which complements our systems research. Some of this work is in established fields of theoretical computer science, such as the analysis of algorithms, computational geometry, and logics of programming. The rest of this work explores new ground motivated by problems that arise in our systems research.

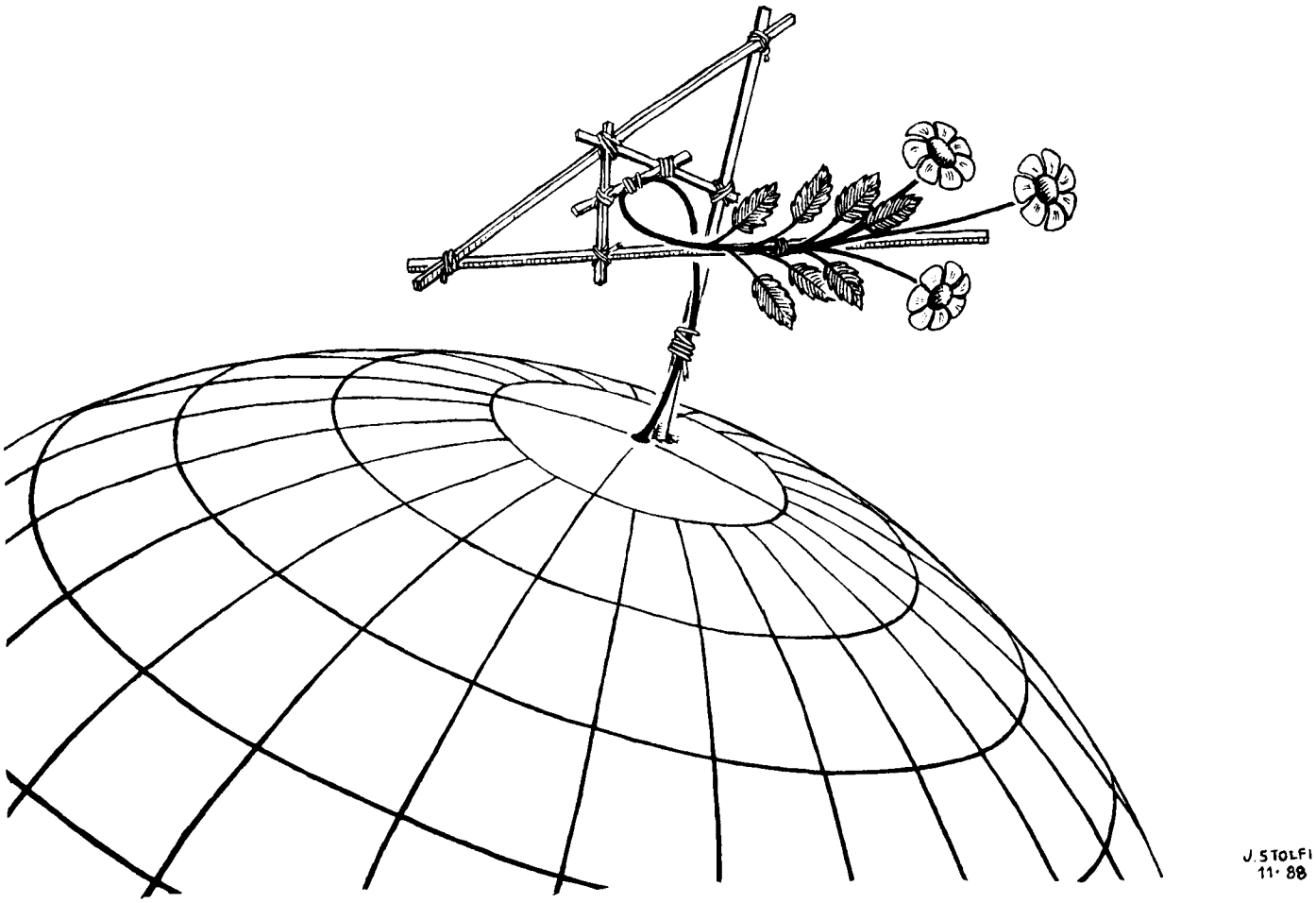
DEC has a strong commitment to communicating the results and experience gained through pursuing these activities. The Company values the improved understanding that comes with exposing and testing our ideas within the research community. SRC will therefore report results in conferences, in professional journals, and in our research report series. We will seek users for our prototype systems among those with whom we have common research interests, and we will encourage collaboration with university researchers.

Robert W. Taylor, Director

Blossoms Are Polar Forms

Lyle Ramshaw

January 2, 1989



Perspective

Splines are the standard technique for mathematically modeling smooth shapes—such as curves in the plane or surfaces in space—in computer-aided geometric design. The basic idea behind splines is to cut the shape up into pieces and to model each piece with polynomials, but to constrain the polynomials so that the joints between the pieces are guaranteed to be smooth. There is a substantial body of mathematical theory about splines, associated with such names as Bézier, de Casteljau, and de Boor.

Polynomials, of course, are used in many places in mathematics. There is a standard construction that converts a polynomial of degree n in one variable into an equivalent polyomial in n variables that has degree 1 in each variable separately. The latter polynomial is called the *polar form* of the former.

Recently, a new and perhaps simpler approach to the theory of splines has emerged, based on converting each polynomial piece of the spline into its polar form. This paper is a review of the resulting polarized approach to splines. The intended reader already knows something about Bézier curves and is willing to try to understand a sentence that includes the word "tensor."

Author's abstract

Consider the functions $H(t) := t^2$ and $h(u,v) := uv$. The identity $H(t) = h(t,t)$ shows that H is the restriction of h to the diagonal $u = v$ in the uv-plane. Yet, in many ways, a bilinear function like h is simpler than a homogeneous quadratic function like H. More generally, if $F(t)$ is some n-ic polynomial function, it is often helpful to study the *polar form* of F, which is the unique symmetric, multiaffine function $f(u_1, \ldots, u_n)$ satisfying the identity $F(t) = f(t, \ldots, t)$. The mathematical theory underlying splines is one area where polar forms can be particularly helpful, because two pieces F and G of an n-ic spline meet at r with C^k parametric continuity if and only if their polar forms f and g satisfy

$$f(u_1, \ldots, u_k, \underbrace{r, \ldots, r}_{n-k}) = g(u_1, \ldots, u_k, \underbrace{r, \ldots, r}_{n-k})$$

for all u_1 through u_k.

This polarized approach to the theory of splines emerged in rather different guises in three independent research efforts: Paul de Faget de Casteljau called it "shapes through poles"; Carl de Boor called it "B-splines without divided differences"; and Lyle Ramshaw called it "blossoming." This report reviews the work of de Casteljau, de Boor, and Ramshaw in an attempt to clarify the basic principles that underlie the polarized approach. It also proposes a consistent system of nomenclature as a possible standard.

Contents

1 Introduction

The term "spline" is used in computer-aided geometric design to refer to a family of techniques for modeling smooth curves and surfaces with parametric, piecewise polynomial functions; see the survey article by Böhm, Farin, and Kahmann [4] or the textbook by Bartels, Beatty, and Barsky [1]. Recently, a new *polarized approach* to the mathematical theory underlying splines has emerged from the work of Paul de Faget de Casteljau [11] and others. The primary benefits of the polarized approach are simplicity and clarity. The polarized approach gives perspicuous labels to the points in diagrams and thus makes it easier to understand and to explain the standard theorems and algorithms involving splines. This report concentrates on spline curves, which are the easiest case.

The polarized approach is based on the Polarizing Principle, which is a standard technique used in various places in mathematics. What the Polarizing Principle does, as discussed in Section 2, is to convert polynomial behavior into multiaffine behavior (or homogeneous polynomial behavior into multilinear behavior). When applied to one of the polynomial pieces F of a spline, the Polarizing Principle converts that piece into its *polar form*, which is a symmetric, multiaffine function f.

The power of polarization begins to show in Section 3, where we study the de Casteljau Algorithm. When the de Casteljau Algorithm is used to compute a point $F(u)$ on a polynomial curve, all of the points that arise—input points, intermediate points, and output points—are values of the polar form f of F. Labeling each point as a value of f helps to clarify the iterated phases of affine interpolations that constitute the de Casteljau Algorithm. Section 4 shows that the same labeling also works for the de Boor Algorithm, which is essentially the result of generalizing the de Casteljau Algorithm from the *plain case* to the *progressive case*. Having established that polarization is at least a source of good labels, Section 5 discusses a variety of methods for computing polar forms.

If a new approach to an established theory is to be convincing, it must provide its own proofs of the important theorems, not just its own labels for the important diagrams. Polarization can provide clearer and more enlightening proofs when it is supplemented by two other mathematical principles: homogenization and the tensor product construction.

Homogenization is a fancy name for the standard idea that the *affine world*, where practitioners of CAGD prefer to work, can be viewed as a restricted subset of the *linear world*, where mathematicians are more at home. For example, an affine transformation of the plane is often represented as a 3-by-3 matrix whose last column (or row) is the unit vector $(0, 0, 1)$. This representation depends on the fact that homogenizing an affine map of the plane produces a linear map of 3-space, as discussed in Section 6.

The tensor product construction is a more powerful and more subtle idea. It takes polar forms, which are symmetric, multiaffine functions, and converts them into *tensor forms*, which are affine functions. By using both polarization

and the tensor product construction, we can convert a polynomial map into something much simpler: an affine map. The price that we pay for simplifying the map is making its domain more complex. If the affine space P is the domain of an n-ic function F, the domain of the polar form of F is the Cartesian n^{th} power of P (which is written P^n), while the domain of the tensor form of F is the n^{th} *symmetric tensor power* of P (which we shall write $\bigodot^n P$, by analogy with the standard notation $\bigotimes^n P$ for the asymmetric case). Sections 7 and 8 present a concrete construction of the symmetric tensor power space $\bigodot^n P$, based on formal polynomials. Section 9 discusses tensors more abstractly; it relates the symmetric tensors that we use in the polarized approach to the alternating tensors that many people know about from integration theory. We finally put tensors to work for us in Section 10, where we use them to give an easy proof of an important dangling theorem from Section 4.

The polar forms and tensor forms of single polynomials are enlightening, but they are not new. The novelty in the polarized approach comes from the interaction between polarization and the continuity conditions that are enforced at the joints of a spline. Section 11 shows that the requirement of parametric k^{th}-order continuity at a joint of a spline curve corresponds to a wonderfully simple constraint on the polar forms of the two joining segments. Two non-adjacent segments of a spline curve are still related to each other, as long as the number of intervening knots does not exceed the degree; Section 12 extends the analysis of Section 11 to this case, arriving at a similarly simple constraint on the polar forms of two non-adjacent spline segments. Section 12 then concludes by proving one version of the basic theorem about spline curves: that the B-splines are a basis.

Section 13 touches on spline surfaces. In some sense, spline curves—and bipolynomial spline surfaces, which are just spline curves of spline curves—are too easy. It will be interesting to see, in the coming years, if the power and simplicity of the polarized approach leads to any progress on the more subtle problems of polynomial spline surfaces.

The polarized approach first appeared, in three different guises, in the independent work of three people: Paul de Faget de Casteljau, Carl de Boor, and Lyle Ramshaw. One goal of this report is to clarify the interrelationships between these three threads. (I am unavoidably biased by the fact that I understand my own work the best. I apologize for that bias and for any ways in which I have inadvertently misrepresented the work of de Casteljau or de Boor.) The last three sections briefly discuss the related work of de Casteljau, de Boor, and Ramshaw. Section 14 discusses the "dual functionals" of de Boor from the framework of this report and identifies them as bags of polar arguments. Section 15 mentions Ramshaw's concept of "overloading" and his handling of the rational case. And Section 16 talks about de Casteljau's technique for easily designing quasi-interpolant spline methods that are as reproductive as possible, given their other parameters.

In a work of this sort, which attempts to point out the connections between

different people's work and to propose standards, it is particularly important to be clear and precise. Therefore, we will pause from time to time to discuss terminology and notation.

Note on "spline": Some authors, de Casteljau among them, reserve the word "spline" to refer to those piecewise polynomials of degree at most n that have C^{n-1} continuity at each joint between two adjacent pieces. That is, they demand that splines have only simple knots. We shall use the word "spline" in its broader sense, in which knots of multiplicity m—that is, joints with C^{n-m} continuity—are allowed, even when $m > 1$.

Note on "n-ic": The polynomial $H(t) := t^2$ is a polynomial of degree 2 and, hence, is not a polynomial of degree 3. Nevertheless, is it correct to say that "H is a cubic"? The adjective "n-ic" sometimes means "of degree precisely n" and sometimes means "of degree at most n." In this report, we shall assume the latter, more inclusive meaning. For example, if one of the segments of a cubic spline actually has degree less than 3—even if all of the segments do—we will still feel free to refer to that spline and to each of its segments as cubic.

Note on "order": Some authors use "order" rather than "degree" when defining splines, because "order" has the sense of "at most" built into it. A polynomial has *order* $n + 1$ if its degree is at most n. Unfortunately, there are off-by-one problems: If a cubic spline is assembled from pieces of order 4, why isn't it called a quartic spline? In the polarized approach, the degree bound n actually counts something—the polar form has n arguments—so the off-by-one problems would be particularly severe. We shall stick with degree.

Math note: All the spaces in this report—affine and linear—are assumed to be finite-dimensional.

2 The Polarizing Principle

The polarized approach to splines is based on the classical mathematical principle of polar forms, which says, in essence, that we can trade one parameter of degree n for n symmetric parameters, each of degree 1. Before discussing an example of a polar form, we pause to clarify our nomenclature for functions of degree 1.

Note on "linear" versus "affine": The word "linear" in mathematics sometimes implies homogeneity and sometimes doesn't. A polynomial $F(x)$ is linear if it has the form $F(x) = ax + b$, where $b \neq 0$ is usually allowed. If we reinterpret $F: \mathbb{R} \to \mathbb{R}$ as a transformation of a 1-dimensional linear space, however, we must have $b = 0$ in order for F to be called linear; that is, F must also be

homogeneous. To avoid confusion, the word "linear" in this report always implies homogeneity. When homogeneity is not implied, we use the word "affine." Under this convention, the polynomial $F(x) := x+1$ is not linear, but it is affine.

More abstractly, an *affine combination* is a linear combination whose scalar coefficients sum to 1. A function is *affine* if it commutes with affine combinations, while an *affine space* (or *flat*) is a set of points that is closed under affine combinations. An *affine frame* for an affine space P is a set of points in P that are affinely independent and whose affine span is all of P; that is, an affine frame is the analog of a linear basis. Every affine frame for a p-dimensional space contains precisely $p+1$ points.

Now, an example of a polar form: Consider the cubic polynomial $G(t) := t^3+3t^2-6t-8$. A polar form for G is a symmetric, triaffine polynomial $g(u,v,w)$ that satisfies the identity $g(t,t,t) = G(t)$. In order for g to be triaffine, that is, an affine function of each of its three arguments when the other two are held fixed, it must have the form

$$g(u,v,w) := c_1uvw + c_2uv + c_3uw + c_4vw + c_5u + c_6v + c_7w + c_8$$

for some real constants c_1 through c_8. To make the identity $g(t,t,t) = G(t)$ hold, we must have $c_1 = 1$, $c_2 + c_3 + c_4 = 3$, $c_5 + c_6 + c_7 = -6$, and $c_8 = -8$. To make $g(u,v,w)$ a symmetric function of its three arguments, we must have $c_2 = c_3 = c_4$ and $c_5 = c_6 = c_7$. We are left with the unique choice

$$g(u,v,w) := uvw + uv + uw + vw - 2u - 2v - 2w - 8. \tag{2.1}$$

Conversely, suppose that we are given the trivariate polynomial $g(u,v,w) := uvw + uv + uw + vw - 2u - 2v - 2w - 8$. The identity $G(t) = g(t,t,t)$ then trivially determines the unique cubic polynomial $G(t) := t^3 + 3t^2 - 6t - 8$.

From these two observations, we deduce that the two polynomials $G(t)$ and $g(u,v,w)$ are actually two different aspects of the same entity. That entity can be viewed either as a cubic function of the single parameter t or as a symmetric, triaffine function of the three parameters u, v, and w. The quantity $G(t) = g(t,t,t)$ varies cubically as a function of t because, when t varies, all three of u, v, and w are varying in parallel.

The correspondence between G and g doesn't depend upon the particular coefficients involved, nor does it depend in any essential way upon the fact that G is cubic. More generally, we have the following, which is the nonhomogeneous, univariate case of the Polarizing Principle.

Theorem 2.2 *Univariate polynomials $F(t)$ of degree at most n are equivalent to symmetric, n-affine polynomials $f(u_1,\ldots,u_n)$ in the sense that, given a polynomial of either type, there exists a unique polynomial of the other type that satisfies the correspondence identity $F(t) = f(t,\ldots,t)$.*

Definition 2.3 If $F(t)$ is a polynomial of degree at most n, the *polar form* of F (or the *n-polar form*, if the intended degree bound n isn't obvious from the context) is the unique symmetric, n-affine polynomial $f(u_1, \ldots, u_n)$ that corresponds to F via the identity $F(t) = f(t, \ldots, t)$. A value $f(u_1, \ldots, u_n)$ of the polar form is a *polar value* of F, and each u_i that helps to determine such a value is a *polar argument* to F. In contrast, F itself is the *diagonal form* of F; a value $F(t)$ is a *diagonal value* of F; and the t that determines a diagonal value is a *diagonal argument* to F. Note that diagonal values are a special case of polar values, the case in which all n of the polar arguments are equal.

Exercise 2.4 Compute the 4-polar form h that results when the polynomial $G(t) = t^3 + 3t^2 - 6t - 8$ above is viewed as a degenerate quartic. (The answer $h(u, v, w, x)$ has fifteen terms, two of which are $uvw/4$ and $uv/2$.) Note that h is related to the 3-polar form g of G given in Equation 2.1 by the formula

$$h(u, v, w, x) = \frac{g(u, v, w) + g(u, v, x) + g(u, w, x) + g(v, w, x)}{4}.$$

Convince yourself that this identity expresses the 4-polar form of any cubic in terms of its 3-polar form.

The Polarizing Principle applies to multivariate polynomials also, but the concept of a degree bound is more complicated in the multivariate case. We can either bound the total degree in all the variables or bound the degree in each variable separately. Those different ways of bounding the degree give rise to different polar forms, as discussed in Section 13.

A homogeneous polynomial has a polar form that is multilinear, not merely multiaffine. This fact isn't very interesting, however, until one turns to the multivariate case, since a univariate polynomial $F(t)$ must be a scalar multiple of t^n in order to be homogeneous of degree n. The n-polar form of t^n is, of course, the n-linear polynomial $u_1 \cdots u_n$.

While polar forms are new to spline theory, their use has long been standard in other areas of mathematics. For example, consider quadratic forms and bilinear forms in linear algebra. It is well known that, for each quadratic form $F: V \to \mathbb{R}$ on a linear space V, there is a unique symmetric, bilinear form $f: V \times V \to \mathbb{R}$ that satisfies the identity $F(\mathbf{z}) = f(\mathbf{z}, \mathbf{z})$. This fact is precisely the Polarizing Principle applied to multivariate polynomials that are homogeneous of total degree 2.

The Polarizing Principle can also be applied to more than one polynomial at a time. In the case of interest to us, the polynomials involved are the coordinates of a parametric, polynomial curve. Let L denote a 1-dimensional affine space, that is, a line. A *polynomial curve* is a function $F: L \to Q$ from the parameter space L to some affine object space Q with the property that, for u in L, each coordinate of the point $F(u)$ in Q is given by a polynomial in u. To polarize a polynomial curve F, we polarize each coordinate polynomial separately, using

the same degree bound n for all coordinates. The *n-polar form* of F that results is the unique symmetric, n-affine function $f: L^n \to Q$ satisfying the identity $F(u) = f(u, \ldots, u)$.

Math note: The discussion above presumes that a "polynomial function" is, by definition, a function that is given by polynomials when expressed in coordinates. Abstractly, it is better to turn things around by defining an n-ic polynomial function to be the diagonal of an n-affine function. Note that this definition is coordinate-free and hence extends easily to infinite-dimensional spaces. One then uses arguments like those in Theorem 2.2 to prove that, in the finite-dimensional case, the polynomial functions are precisely those functions whose coordinates are given by polynomials.

Even though many areas of mathematics use polar forms, the terminology based on the word "polar" seems to have fallen out of favor in the last half-century. While algebra books up through van der Waerden [29] generally used the name "polar form," more recent books often leave the correspondence of Theorem 2.2 nameless. (For example, Lang doesn't mention polar forms, although he does refer to the "polarization identity" in a related context [15].) Because Ramshaw was not aware of "polar form," he proposed an alternative system of nomenclature in which f was called the "blossom" of F [19, 20]. One advantage of Ramshaw's proposal was that the term "blossoming" suggests the revealing of hidden structure, which describes pretty well what happens when we convert from F to f. The term "polarization," on the other hand, suggests concentration into opposing extremes, which is not at all the right idea. While "blossoming" has a certain charm, "polar form" is a better name for f than "blossom."

The words "pole" and "polar" are used in various places in mathematics that are not at all related to the Polarizing Principle. For example, consider the poles of a sphere, polar coordinate systems, and the poles of a complex analytic function. Those uses don't seem likely to be confused with the uses proposed in Definition 2.3.

The words "pole" and "polar" are also used in projective geometry, when discussing conic sections [27], higher order plane curves [25], and quadric surfaces [2]. There is more chance of confusion in this case, because these uses *are* related to the Polarizing Principle. To understand the relationship between the "polar" in projective geometry and the "polar" in Definition 2.3, we must distinguish between the two different ways that polynomial functions are used to model shapes: parametric models and implicit models.

Suppose that we want to model a shape S in an affine space Q, where $\dim(S) = s$ and $\dim(Q) = q$. A *parametric model* for S defines S as the range of a function $F: P \to Q$, where P is an s-dimensional parameter space. For example, the quadratic function $G: \mathbb{R} \to \mathbb{R}^2$ given by $G(u) := \langle u, u^2 \rangle$ is a parametric model for the standard parabola in the plane. An *implicit* or *Cartesian model*

for a shape S, on the other hand, defines S by the formula $S = F^{-1}(0)$, where $F: Q \to R$ is a function from the object space Q to a result space R of dimension $q - s$. The quadratic function $H: \mathbb{R}^2 \to \mathbb{R}$ given by $H(\mathbf{z}) = H(\langle x, y \rangle) := y - x^2$ is an implicit model of the same standard parabola.

In this report, we are applying the Polarizing Principle to the pieces of splines, which are parametric models of shapes. In the parabola example, the polar form of the parametric model G is the symmetric, biaffine function $g: \mathbb{R} \times \mathbb{R} \to \mathbb{R}^2$ given by $g(u_1, u_2) := \langle (u_1 + u_2)/2, u_1 u_2 \rangle$.

Exercise 2.5 Verify that the point $g(u_1, u_2)$ is, in fact, the intersection of the tangent lines to the parabola at the points $G(u_1)$ and $G(u_2)$.

Projective geometry, on the other hand, models its shapes implicitly. In the parabola example, if we use a bound of 2 on the total degree, the polar form of the parabola's implicit model H is the symmetric, biaffine mapping $h: \mathbb{R}^2 \times \mathbb{R}^2 \to \mathbb{R}$ given by

$$h(\mathbf{z}_1, \mathbf{z}_2) = h\big(\langle x_1, y_1 \rangle, \langle x_2, y_2 \rangle\big) := \frac{y_1 + y_2}{2} - x_1 x_2.$$

The most interesting thing about h is where it assumes the distinguished value 0. If we fix $\mathbf{z}_1$, the equation $h(\mathbf{z}_1, \mathbf{z}_2) = 0$ constrains $\mathbf{z}_2$ to lie on some line, call it $\ell_2 := \{\mathbf{z}_2 \mid h(\mathbf{z}_1, \mathbf{z}_2) = 0\}$. If $\mathbf{z}_1$ is outside the parabola, then ℓ_2 is precisely the line joining the two points where lines through $\mathbf{z}_1$ are tangent to the parabola. If $\mathbf{z}_1$ lies on the parabola, then ℓ_2 is the tangent line through $\mathbf{z}_1$. And if $\mathbf{z}_1$ is inside the parabola, then ℓ_2 is entirely outside it. The correspondence between $\mathbf{z}_1$ and ℓ_2 is a duality between points and lines called the *polarity* of the conic. Of the dual pair $\mathbf{z}_1$ and ℓ_2, the point $\mathbf{z}_1$ is called the *pole* and the line ℓ_2 is called the *polar* [27].

Going up by 1 in dimension, an implicit model of a quadric surface in 3-space defines a duality between pole points and polar planes in a similar way [2].

Going up in degree instead of in dimension, suppose that $H: \mathbb{R}^2 \to \mathbb{R}$ is an implicit model of an algebraic plane curve C of degree n, for $n > 2$. The polar form of H is a symmetric, n-affine map $h: (\mathbb{R}^2)^n \to \mathbb{R}$. If we fix k of the polar arguments of f at $\mathbf{w}_1$ through $\mathbf{w}_k$, the remaining function

$$(\mathbf{z}_{k+1}, \ldots, \mathbf{z}_n) \mapsto h(\mathbf{w}_1, \ldots, \mathbf{w}_k, \mathbf{z}_{k+1}, \ldots, \mathbf{z}_n)$$

is symmetric and $(n - k)$-affine, and it is hence the polar form of an implicit model of an algebraic plane curve D of degree $n - k$. When $\mathbf{w}_1 = \cdots = \mathbf{w}_k = \mathbf{w}$, the curve D is called the k^{th} *polar curve* of the pole point $\mathbf{w}$ with respect to C; an $(n-2)^{\text{nd}}$ polar is a conic and an $(n-1)^{\text{st}}$ polar is a line [25].

Open problem 2.6 If C is a polynomial (or rational) plane curve that is modeled both implicitly and parametrically, study the relationships between the polar forms of the two models of C. For example, if G and H are parametric and implicit models of a common parabola, as above, the identity $h(g(u, u), g(u, v)) = 0$

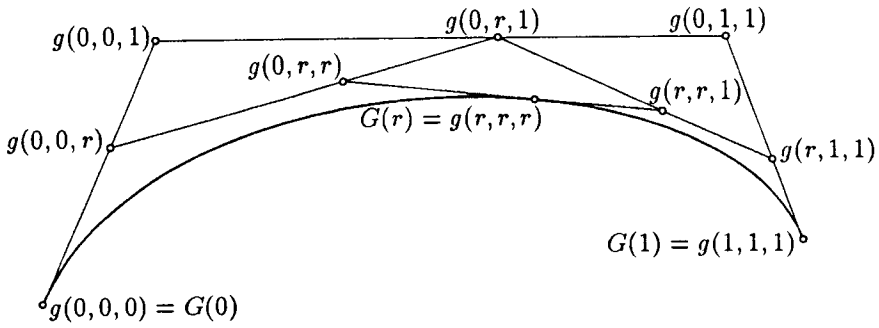


Fig. 1: The de Casteljau Algorithm, plain case, computes a diagonal value

holds between the polar forms g and h. Note that the curve C must be rational, that is, of genus zero, in order to have a rational parametric model. It must lie in a plane in order to be implicitly modeled by a single polynomial, rather than by some non-principal ideal.

Math note: We are tacitly assuming in this report that the coefficient field is the real numbers $\mathbb{R}$. While any field of characteristic zero would work just as well, fields of finite characteristic cause trouble when polarizing a polynomial whose degree equals or exceeds the characteristic. Some authors define polar forms slightly differently than we have done in order to address this problem. For example, van der Waerden [29] defines the polar form f of a quadratic form F on a linear space by using the correspondence identity $f(\mathbf{z},\mathbf{z}) = 2F(\mathbf{z})$, rather than $f(\mathbf{z},\mathbf{z}) = F(\mathbf{z})$, because he wants to deal with fields of characteristic 2.

3 The plain case of the de Casteljau Algorithm

What does polarization have to do with splines? We begin to trace the connections in this section by studying the de Casteljau Algorithm from a polarized point of view.

Note: To reduce the overloading on the comma, we shall borrow from the programming language Pascal by writing $(x\,..\,y)$ for an open interval and $[x\,..\,y]$ for a closed interval.

Let G be a cubic parametric curve and let g be the polar form of G. Suppose that we know the four polar values $g(0,0,0)$, $g(0,0,1)$, $g(0,1,1)$, and $g(1,1,1)$ and that we want to compute, from them, the diagonal value $G(r) = g(r,r,r)$ for some r in $[0\,..\,1]$. Figure 1 shows one way to proceed. (The four input points $g(0,0,0)$, $g(0,0,1)$, $g(0,1,1)$, and $g(1,1,1)$ need not be coplanar. If they are not, then G is a twisted cubic curve, lying in the affine 3-space that they span.)

Since the polar form g is affine in its third argument, the line joining the points $g(0,0,0)$ and $g(0,0,1)$ must be the line $g(0,0,u)$ for all u, and the point $g(0,0,r)$ must be located r of the way from $g(0,0,0)$ to $g(0,0,1)$ along this line. The points $g(0,r,1)$ and $g(r,1,1)$ can be constructed similarly. These three affine interpolations constitute the first phase of the algorithm.

During the second phase, we need to appeal to the symmetry of g as well as to its triaffineness. Since g is symmetric, the first phase output $g(0,0,r)$ could equally well be labeled $g(0,r,0)$. Then, since g is affine in its third argument, the point $g(0,r,r)$ must be r of the way from $g(0,r,0)$ to $g(0,r,1)$. The point $g(r,r,1)$ can be constructed similarly.

The third phase finishes the job, constructing $g(r,r,r)$ by interpolating between $g(0,r,r) = g(r,r,0)$ and $g(r,r,1)$.

Of course, this algorithm is not new. The four starting points $g(0,0,0)$, $g(0,0,1)$, $g(0,1,1)$, and $g(1,1,1)$ are called the *Bézier points* of the cubic curve segment $G([0\,..\,1])$, and the algorithm above for computing $G(r)$ from the Bézier points is the *de Casteljau Algorithm.* The polarized labels, however, are new. Note that all of the geometric structure of the diagram is captured by those labels, together with the facts that the polar form g is symmetric and triaffine.

Much of the importance of Bézier points comes from the fact that they can be used as handles that control the shape of the curve.

Theorem 3.1 *If* $\mathbf{a}$, $\mathbf{b}$, $\mathbf{c}$, *and* $\mathbf{d}$ *are any four points in an affine space* Q, *there exists a unique cubic curve* $F: L \to Q$ *whose polar form* $f: L^3 \to Q$ *satisfies* $f(0,0,0) = \mathbf{a}$, $f(0,0,1) = \mathbf{b}$, $f(0,1,1) = \mathbf{c}$, *and* $f(1,1,1) = \mathbf{d}$.

Proof: The de Casteljau Algorithm makes it obvious that there can't be more than one such cubic F, since it allows us to compute any diagonal value $F(r) = f(r,r,r)$ for r in $[0\,..\,1]$ by performing affine interpolations that depend upon only the known values $\mathbf{a}$, $\mathbf{b}$, $\mathbf{c}$, $\mathbf{d}$, and r. These diagonal values $F(r)$ are more than enough to completely determine F.

Proving the existence of some such cubic F is harder. There are two strategies: one computational and the other conceptual.

In the computational strategy, we construct the required cubic curve F by performing the same affine interpolations that the de Casteljau Algorithm would have performed, had we assumed that some F existed. The first phase of affine interpolations produces the points $(1-r)\mathbf{a}+r\mathbf{b}$, $(1-r)\mathbf{b}+r\mathbf{c}$, and $(1-r)\mathbf{c}+r\mathbf{d}$. The second phase interpolates between these points in pairs, while the third phase produces the final output point

$$(1-r)^3\mathbf{a} + 3r(1-r)^2\mathbf{b} + 3r^2(1-r)\mathbf{c} + r^3\mathbf{d}.$$

This formula defines a cubic curve; call it $F(r)$. With a little thought, we can

write down the polar form f of F:

$$\begin{aligned} f(u,v,w) := {} & (1-u)(1-v)(1-w)\mathbf{a} \\ & + \big(u(1-v)(1-w) + (1-u)v(1-w) + (1-u)(1-v)w\big)\mathbf{b} \\ & + \big(uv(1-w) + u(1-v)w + (1-u)vw\big)\mathbf{c} \\ & + uvw\,\mathbf{d}. \end{aligned}$$

Note that $f(u,v,w)$ is symmetric, is triaffine, and satisfies $f(r,r,r) = F(r)$; therefore, f is indeed the polar form of F. Since it is easy to check that $f(0,0,0) = \mathbf{a}$, $f(0,0,1) = \mathbf{b}$, $f(0,1,1) = \mathbf{c}$, and $f(1,1,1) = \mathbf{d}$, we conclude that a satisfactory cubic curve F does exist. □

The conceptual strategy uses tensors and linear algebra to prove that F must exist without explicitly constructing it. One attraction of the conceptual strategy is that it generalizes from the plain case to the progressive case more gracefully than does the computational strategy. We will return to the conceptual strategy in Section 10.

4 The de Casteljau Algorithm, progressive case

The de Casteljau Algorithm can be generalized in several ways. In this section, we explore four successive generalizations.

Note on "bag": It will be useful to have a name for the entire collection, say $(u_1, \ldots, u_n)$, of n polar arguments that determines a polar value. This collection isn't a sequence, because order doesn't matter. But it also isn't a set, because repeated elements do matter. We shall call the collection $(u_1, \ldots, u_n)$ a *bag* of polar arguments; some authors use the term "multiset."

First, the starting parameter interval in the de Casteljau Algorithm doesn't have to be $[0\,..\,1]$, nor is it necessary for the new parameter value r to lie inside the starting interval. For example, we could also build Figure 1 by treating the four Bézier points $g(0,0,0)$, $g(0,0,r)$, $g(0,r,r)$, and $g(r,r,r)$ of the segment $G([0\,..\,r])$ as the given data and treating 1 as the new parameter value. It would then take us three phases of affine extrapolations to construct the rest of the diagram, working from left to right. In what follows, we shall denote the starting parameter interval by $[s\,..\,t]$ and the new parameter value by u.

Second, the de Casteljau Algorithm easily extends from cubics to n-ic polynomial curves for any n. In the n-ic case, we perform n phases of affine interpolations, each of which adds one more copy of the desired diagonal argument u to the bag of polar arguments. After n phases, we have added n copies of u to the bag, that is, we have constructed the point $f(u, \ldots, u) = F(u)$.

Third, the de Casteljau Algorithm can be used to compute polar values of F just as easily as diagonal values. To compute the polar value $f(u_1, \ldots, u_n)$,

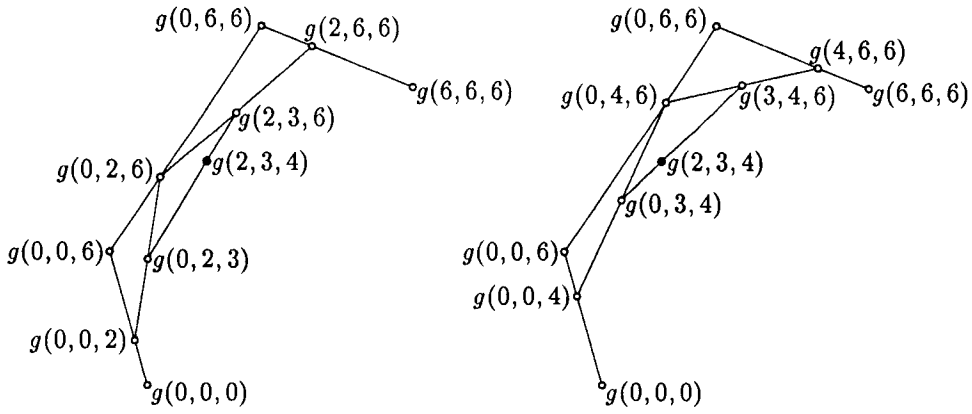


Fig. 2: Computing $g(2,3,4)$ in two different ways

we control each phase of affine interpolations with a different polar argument, say the i^{th} phase with u_i. The left half of Figure 2 shows this version of the de Casteljau Algorithm computing the polar value $g(2,3,4)$ from the four Bézier points of the cubic segment $G([0\,..\,6])$. The right half of Figure 2 shows what happens when the polar arguments are used in the reverse order. Note that the final result is the same, even though the intermediate points and lines are different.

Our fourth—and most important—generalization of the de Casteljau Algorithm deals with the structure of the bags of polar arguments in the $n+1$ input polar values. So far, we have been starting the de Casteljau Algorithm with Bézier points; when $n = 3$, we started with the polar values $g(s,s,s)$, $g(s,s,t)$, $g(s,t,t)$, and $g(t,t,t)$, for some $s \neq t$. We shall refer to that situation henceforth as the *plain case*. More generally, we can start with $g(r_1,r_2,r_3)$, $g(r_2,r_3,r_4)$, $g(r_3,r_4,r_5)$, and $g(r_4,r_5,r_6)$, where the values r_1 through r_6 satisfy the six disequalities given in the following table:

	r_4	r_5	r_6
r_1	$\neq$		
r_2	$\neq$	$\neq$	
r_3	$\neq$	$\neq$	$\neq$

We shall refer to this fancier situation as the *progressive case*, because the four bags of polar arguments (r_1,r_2,r_3), (r_2,r_3,r_4), (r_3,r_4,r_5), and (r_4,r_5,r_6) are generated as a window of width three progresses along the sequence $[r_1,\ldots,r_6]$. Note that the progressive case reduces to the plain case when we set $r_1 = r_2 = r_3 = s$ and $r_4 = r_5 = r_6 = t$; in particular, all six disequalities reduce to the single disequality $s \neq t$.

Figure 3 shows the de Casteljau Algorithm in the cubic, progressive case computing the polar value $g(u_1,u_2,u_3)$. Note that the six disequalities in the

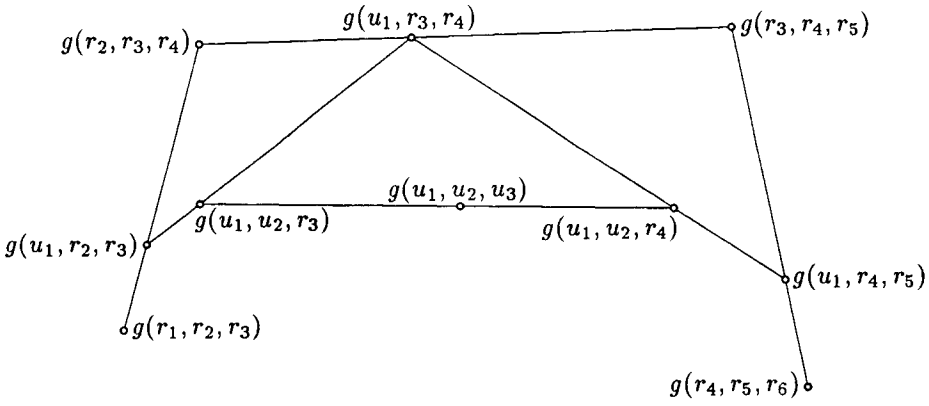


Fig. 3: The de Casteljau Algorithm, progressive case, computes a polar value

table above are just what is needed to guarantee that none of the six affine interpolations involved in the de Casteljau Algorithm entails dividing by zero. For example, one of the interpolations in the first phase locates the point $g(u_1, r_2, r_3)$ by interpolating between $g(r_1, r_2, r_3)$ and $g(r_2, r_3, r_4)$. If we had $r_1 = r_4$, it would follow that $g(r_1, r_2, r_3) = g(r_2, r_3, r_4)$ and we would be unable to calculate the location of $g(u_1, r_2, r_3)$ from that single datum (except in the special case $u_1 = r_1$).

Definition 4.1 For any n, we shall call a sequence of $2n$ numbers $[r_1, \ldots, r_{2n}]$ *progressable* if the $\binom{n+1}{2}$ disequalities $r_j \neq r_{n+i}$ for $1 \leq i \leq j \leq n$ are satisfied.

Let F be an n-ic curve and let $[r_1, \ldots, r_{2n}]$ be a progressable sequence. The progressive case of the de Casteljau Algorithm takes, as input, the $n + 1$ polar values of F whose argument bags consist of n consecutive elements from the progressable sequence—in symbols, the polar values $f(r_{k+1}, \ldots, r_{k+n})$ for k in $[0 \,..\, n]$. It performs n phases of affine interpolations, where the i^{th} interpolation of the j^{th} phase for j in $[1 \,..\, n]$ and i in $[0 \,..\, n - j]$ interpolates between $f(u_1, \ldots, u_{j-1}, r_{i+j}, \ldots, r_{i+n})$ and $f(u_1, \ldots, u_{j-1}, r_{i+j+1}, \ldots, r_{i+n+1})$ to produce $f(u_1, \ldots, u_j, r_{i+j+1}, \ldots, r_{i+n})$. The 0^{th} interpolation of the n^{th} phase produces the output value $f(u_1, \ldots, u_n)$.

Theorem 4.2 *Let $[r_1, \ldots, r_{2n}]$ be a progressable sequence of points in the affine line L, and let $\mathbf{a}_0$ through $\mathbf{a}_n$ be arbitrary points in an affine space Q. There exists a unique n-ic curve $F: L \to Q$ whose polar form $f: L^n \to Q$ satisfies $f(r_{k+1}, \ldots, r_{k+n}) = \mathbf{a}_k$ for k in $[0 \,..\, n]$.*

Proof (half given here, half postponed): The uniqueness of F follows easily from the de Casteljau Algorithm, just as it did in Theorem 3.1. Once again, it is the existence of F that is the hard part. And once again, there are two proof

strategies: computational and conceptual. Unfortunately, the computational strategy gets fairly complicated in the progressive case. We leave it to the following exercises, and we defer our proof of existence in the progressive case until Section 10, where we will use the conceptual strategy.

Exercise 4.3 Give a computational proof of the existence of F in the n-ic, plain case, hence generalizing the proof of Theorem 3.1 from degree 3 to degree n. In particular, show that the interpolations that the de Casteljau Algorithm would perform to compute $f(u_1, \ldots, u_n)$ from $f(s, \ldots, s)$, $f(s, \ldots, s, t)$, $\ldots$, $f(t, \ldots, t)$ if F did exist result in the output value

$$\sum_{\substack{I \cap J = \emptyset \\ I \cup J = \{1, \ldots, n\}}} \prod_{i \in I} \left(\frac{t - u_i}{t - s} \right) \prod_{j \in J} \left(\frac{u_j - s}{t - s} \right) \mathbf{a}_{|J|}.$$

Use this expression to define a symmetric, n-affine function $f: L^n \to Q$, and let F be its diagonal. Then, for k in $[0 \,..\, n]$, verify that

$$f(\underbrace{s, \ldots, s}_{n-k}, \underbrace{t, \ldots, t}_{k}) = \mathbf{a}_k.$$

Exercise 4.4 Give a computational proof of the existence of F in the full-blown n-ic, progressive case. Tracing through the de Casteljau Algorithm, you will find that its output is $\mathbf{b}_{0,n}(u_1, \ldots, u_n)$, where $\mathbf{b}_{i,j}(u_1, \ldots, u_j)$ for j in $[0 \,..\, n]$ and i in $[0 \,..\, n - j]$ is given by

$$\mathbf{b}_{i,j} := \begin{cases} \mathbf{a}_i & \text{if } j = 0 \\ \dfrac{r_{i+n+1} - u_j}{r_{i+n+1} - r_{i+j}} \mathbf{b}_{i,j-1} + \dfrac{u_j - r_{i+j}}{r_{i+n+1} - r_{i+j}} \mathbf{b}_{i+1,j-1} & \text{otherwise.} \end{cases}$$

(This recurrence is the heart of the de Boor Algorithm.) The output function $\mathbf{b}_{0,n}(u_1, \ldots, u_n)$ is clearly n-affine, but it is not at all clear that it is symmetric. Explicitly verify that $\mathbf{b}_{0,n}$ is symmetric by studying what happens when two adjacent arguments u_j and u_{j+1} are interchanged. Then make the definition $f := \mathbf{b}_{0,n}$ and verify that $f(r_{k+1}, \ldots, r_{k+n}) = \mathbf{a}_k$ for k in $[0 \,..\, n]$.

5 Methods of polarization

Our study of the de Casteljau Algorithm shows that polar forms can be an enlightening way to view spline algorithms. In this section, we will try to gain more perspective on the mathematical process of polarization by studying eight different methods for computing polar information about a polynomial from diagonal information.

5.1 Using differential geometry

One of the prettiest and most geometric ways to compute polar values applies to non-degenerate polynomial curves and uses differential geometry. For example, consider a twisted cubic curve $F(t)$, lying in some 3-space, and suppose that we want to compute the polar value $f(u, v, w)$. If u, v, and w are distinct, it turns out that $f(u, v, w)$ is precisely the intersection of the three osculating planes to the curve $F(t)$ at $t = u$, $t = v$, and $t = w$. If two of the three polar arguments are equal, say $u = v$, the resulting point $f(u, u, w)$ must lie in the osculating plane to F at u twice, which means that it must lie on the tangent line at u. Thus, $f(u, u, w)$ is the intersection of the tangent line at u with the osculating plane at w. When all three polar arguments are equal, the resulting point $f(u, u, u)$ lies in the osculating plane to F at u three times, which means that it must be $F(u)$, as we already knew. These ideas can be extended to curves of degree greater than 3 by considering osculating flats of dimension greater than 2.

Unfortunately, this method fails whenever the n-ic curve F is degenerate in the sense that F lies in some flat of dimension less than n. For example, if F is a cubic curve that lies in a plane, then all of the osculating planes of F coincide, so we can't use their intersections to compute polar values. A single cubic polynomial is even more degenerate, since it corresponds to a cubic curve that lies in a line.

5.2 Using the elementary symmetric functions

Algebraic methods of polarization have the advantage that they are not confused by degeneracies. In Section 2, we found that the polar form $g(u, v, w)$ of the cubic polynomial $G(t) = t^3 + 3t^2 - 6t - 8$ is given by

$$g(u, v, w) = uvw + 3\left(\frac{uv + uw + vw}{3}\right) - 6\left(\frac{u + v + w}{3}\right) - 8.$$

In general, suppose that $F(t)$ is an n-ic polynomial given to us explicitly as a linear combination of powers of t. Let $\sigma_k(u_1, \ldots, u_n)$ denote the k^{th} elementary symmetric function of n arguments, that is, the n-variate polynomial that is computed by choosing k of the u_i and forming their product, then summing those products over all $\binom{n}{k}$ possible choices. Note that σ_k is a symmetric, n-affine function, whose diagonal is given by $\sigma_k(t, \ldots, t) = \binom{n}{k} t^k$. In particular, the n-polar form of the monomial t^k is precisely $\sigma_k / \binom{n}{k}$. Thus, we can compute the polar form $f(u_1, \ldots, u_n)$ of F by replacing each power t^k in the linear combination for F with $\sigma_k(u_1, \ldots, u_n) / \binom{n}{k}$.

5.3 Using probability theory

The method of elementary symmetric functions can be reinterpreted probabilistically. Suppose that the polar arguments u_1 through u_n are fixed. The quantity

$\sigma_k(u_1, \ldots, u_n)/\binom{n}{k}$ then denotes the average value of the product of k of the polar arguments, the average being taken over all $\binom{n}{k}$ ways of picking which k of the polar arguments to include in the product. This average is a quite reasonable thing to substitute for the simple value t^k, given that the n polar arguments u_1 through u_n have replaced the single diagonal argument t.

5.4 Using the Taylor series of an auxiliary function

If we like, we can hide the complexity of sums with $\binom{n}{k}$ terms by finding those same sums in the Taylor series of an appropriately-defined auxiliary function. Let $h(x)$ be the auxiliary function $h(x) = \prod_{1 \le i \le n}(u_i - x)$. Note that h is an n-ic polynomial with the polar arguments u_1 through u_n as its roots. If we compute the $(n-k)^{\text{th}}$ derivative of $h(x)$ and then evaluate that derivative at $x = 0$, we find that

$$D^{(n-k)}h(0) = (-)^{n-k}(n-k)!\,\sigma_k(u_1, \ldots, u_n).$$

Thus, we can compute polar forms by substituting for t^k the quantity

$$\frac{\sigma_k(u_1, \ldots, u_n)}{\binom{n}{k}} = \frac{k!\,\big((-D)^{n-k}h\big)(0)}{n!}.$$

Carl de Boor used this method to compute polar values (in his terminology, to define the "dual functionals") [9, 10].

5.5 Saving symmetry for last

Instead of hiding the sums, another alternative is to simplify them, at least conceptually, by combining them all into one huge sum with $n!$ terms. In this method, we achieve first multiaffineness and then symmetry. For example, consider the cubic polynomial $G(t) := t^3 + 3t^2 - 6t - 8$. The function $h_1(u, v, w) := uvw + 3uv - 6u - 8$ is obviously triaffine, and it obviously satisfies the identity $G(t) = h_1(t, t, t)$; but h_1 isn't symmetric. The same is true of the function $h_2(u, v, w) := (u-2)(v+1)(w+4)$, since $G(t)$ happens to factor as $G(t) = (t-2)(t+1)(t+4)$. We can compute the polar form g by explicitly symmetrizing some function like h_1 or h_2; that is, we define $g(u, v, w)$ to be

$$\frac{h(u,v,w) + h(u,w,v) + h(v,u,w) + h(v,w,u) + h(w,u,v) + h(w,v,u)}{6}.$$

5.6 Polarization formulas with rational coefficients

The algebraic methods that we have discussed so far start with some explicit diagonal formula for F. Another important class of methods are those that start, instead, with a finite number of diagonal values of F. Of course, once we know

any four diagonal values of a cubic curve such as F, say $F(r_i)$ for i in $[0\,..\,3]$, we can use Lagrange interpolation to construct a diagonal formula for F:

$$F(t) = \sum_{0 \le j \le 3} F(r_j) \prod_{\substack{0 \le i \le 3 \\ i \neq j}} \frac{t - r_i}{r_j - r_i}.$$

We can then find $f(u,v,w)$ by substituting $\sigma_k / \binom{3}{k}$ for t^k in the Lagrange interpolant. Things work out particularly neatly if we choose the r_i cleverly.

Perhaps the neatest choice, due to de Casteljau [12], is to pick $r_1 = u$, $r_2 = v$, and $r_3 = w$, while leaving r_0 arbitrary. By a wonderful coincidence, the term involving $F(r_0)$ drops out, leaving the following 3-term polarization formula for cubic curves:

$$f(u,v,w) = \frac{(w-v)^2 F(u)}{3(w-u)(u-v)} + \frac{(w-u)^2 F(v)}{3(w-v)(v-u)} + \frac{(v-u)^2 F(w)}{3(v-w)(w-u)}.$$

Among other things, this result implies that the four points $G(0)$, $G(r)$, $G(1)$, and $g(0,r,1)$ in Figure 1 must be coplanar.

Analogous wonderful coincidences can be engineered for curves of any degree n, but odd degrees work out more neatly than even degrees. If we express the polar value $f(u_1, \ldots, u_n)$ as an affine combination of the diagonal values $F(r_0)$ through $F(r_n)$, it turns out that the term involving $F(r_0)$ drops out precisely when $\det M = 0$, where the matrix $M = (m_{ij})$ for i and j in $[1\,..\,n]$ is given by $m_{ij} = r_i - u_j$. When n is odd, choosing $r_i = u_i$ makes M a skew-symmetric matrix of odd order, which guarantees $\det M = 0$. The case $n = 3$ is de Casteljau's formula above, while the case $n = 1$ is the trivial formula $f(u) = F(u)$. When n is even, however, it takes more work to find choices for the r_i that make $\det M = 0$. When $n = 2$, the $F(r_0)$ term drops out if the pair of points (r_1, r_2) is harmonically conjugate to the pair (u_1, u_2). For example, choosing $r_1 = (2u + v)/3$ and $r_2 = 2u - v$ gives the 2-term formula

$$f(u,v) = \frac{9F\left(\frac{2u+v}{3}\right) - F(2u-v)}{8}.$$

Math note: Instead of $r_1 = (2u + v)/3$, it would be more natural to choose $r_1 = (u + v)/2$. Unfortunately, this forces the corresponding choice $r_2 = \infty$. The right way to interpret the quantity $F(\infty)$ in this context involves extending F into a homogeneous quadratic $F_\star$, as discussed in Section 6, leading to the homogenized polarization formula

$$f(u,v) = f_\star(\bar{u}, \bar{v}) = \frac{F_\star(\bar{u} + \bar{v}) - F_\star(\bar{u} - \bar{v})}{4}.$$

5.7 Polarization formulas with constant coefficients

One difficulty with the cubic formula in the previous section is that its coefficients are rather complicated. By tolerating four terms instead of only three, we can

simplify the coefficients all the way to constants. In particular, making the choices $r_0 = (u+v+w)/3$, $r_1 = u+v-w$, $r_2 = u+w-v$, and $r_3 = v+w-u$ results in the cubic polarization formula

$$f(u,v,w) = \frac{27F\left(\frac{u+v+w}{3}\right) - F(u+v-w) - F(u+w-v) - F(v+w-u)}{24}.$$

(One advantage of extending F into the homogeneous cubic $F_\star$, as discussed in Section 6, is that it allows us to cancel the $27 = 3^3$ outside against the $1/3$ inside: $27F_\star\big((\bar{u}+\bar{v}+\bar{w})/3\big) = F_\star(\bar{u}+\bar{v}+\bar{w})$.)

Polarization formulas with constant coefficients exist for all n, but they—at least, the ones that I know—have exponentially many terms. The simplest n-ic formula [21] has $2^n - 1$ terms:

$$f(u_1,\ldots,u_n) = \frac{1}{n!} \sum_{\substack{S\subseteq\{1,\ldots,n\} \\ k=|S|}} (-)^{n-k} k^n F\left(\frac{1}{k}\sum_{i\in S} u_i\right).$$

The number of terms can be reduced to 2^{n-1} when n is odd, as in the case $n = 3$ above, by using $+1$ and -1 as the coefficients on the polar arguments u_i, rather than 1 and 0:

$$f(u_1,\ldots,u_n) = \frac{1}{2^{n-1}n!} \sum_{\substack{S\cup T=\{1,\ldots,n\} \\ S\cap T=\emptyset \\ |S|-|T|=k>0}} (-)^{|T|} k^n F\left(\frac{1}{k}\left(\sum_{i\in S} u_i - \sum_{j\in T} u_j\right)\right).$$

The same reduction can also be achieved when n is even, but homogenization or some other technique must be used to deal with the problem of diagonal arguments, such as $u_1 - u_2$ when $n = 2$, that cannot be rescaled to become affine combinations of the u_i.

Math note: One advantage of the polarization formulas in this section is that they involve only affine combinations. They express a polar value as a fixed affine combination of diagonal values, whose diagonal arguments are fixed affine combinations of the original polar arguments. As a result, they can be used to polarize surfaces (and k-folds for $k > 2$) as well as curves. The cubic formula in Section 5.6, on the other hand, applies only to curves, since it has coefficients that are rational functions of the polar arguments.

5.8 Using integration by parts

Our final method of polarization is a weird one, an example of the complex disguises in which polarization can hide. It is the method that E. T. Y. Lee used in his short proof of the Oslo Algorithm based on de Boor's "dual functionals" [17]. Let $h(x) = \prod_{1\le i\le n}(u_i - x)$ be the auxiliary function from Section 5.4, and let

$\varphi(x)$ be any smooth (say C^∞) function with the properties that $\varphi(x) = 0$ whenever x is sufficiently small and $\varphi(x) = 1$ whenever x is sufficiently large. If we define $g(x)$ by the formula $g(x) := \varphi(x)h(x)$, it turns out that, regardless of the detailed behavior of φ, we have

$$f(u_1, \ldots, u_n) = \frac{1}{n!} \int_{-\infty}^{\infty} F(x) g^{(n+1)}(x)\, dx,$$

where $g^{(n+1)}$ denotes the $(n+1)^{\text{st}}$ derivative of g. The proof begins by noting that the integral on the right-hand side must be a symmetric, n-affine function of the polar arguments u_i, because the auxiliary function $h(x)$ is. Therefore, it suffices to verify the identity in the diagonal case, in which $h(x) = (u - x)^n$. Checking that the right-hand side evaluates to $F(u)$ in the diagonal case is an exercise in integration by parts, which we omit.

6 Homogenizing

In order for polarization to achieve its full potential when applied to splines, it has to be helped by two other classical mathematical principles: homogenization and the symmetric tensor product construction. In the following sections, we consider those accessory principles briefly. This section deals with homogenization.

There are two mathematical worlds in which mathematics of our current flavor can be done: the affine world and the linear world. The affine world consists of affine spaces, affine maps, and polynomial maps. The linear world consists of linear spaces, linear maps, and homogeneous polynomial maps. Practitioners of CAGD prefer to work in the affine world, while mathematicians prefer to work in the linear world. There are good reasons behind both preferences, as we will see shortly. Homogenization is the glue that connects the two worlds. To be more accurate, the affine world is really a subset of the linear world, looked at through a kind of blinders. Homogenization is the process that takes an object in the affine world and constructs the corresponding object in the linear world, in essence, removing the blinders.

Practitioners of CAGD use homogenization frequently, although they may not realize it. For example, an affine transformation of p-space is generally represented as a $(p+1)$-by-$(p+1)$ matrix whose last column (or row) is a unit vector. This representation is precisely an instance of homogenization. If $F\colon P \to P$ is an affine transformation of the p-dimensional affine space P, the homogenized form of F is a linear transformation $F_\star\colon P_\star \to P_\star$ of the linear space $P_\star$, where $\dim(P_\star) = p + 1$. Practitioners represent the affine map F by storing the matrix of the linear map $F_\star$.

Consider spaces to begin with. We homogenize an affine space by extending it to become a linear space of the next higher dimension.

Definition 6.1 If P is an affine space, the *linearization* of P is a linear space $P_\star$ together with a linear functional $\omega\colon P_\star \to \mathbb{R}$, called *weight*, that satisfies $P = \omega^{-1}(1)$. Let $P_c = \omega^{-1}(c)$ denote the set of all elements of the linearization $P_\star$ that are *c-heavy*, that is, that have weight c. Each P_c is a hyperplane parallel to $P = P_1$. Careful people doing geometry in P distinguish between the 1-heavy elements of P_1, which are called *points*, and the 0-heavy elements of P_0, which are called *vectors* (more precisely, *free vectors on* P). For example, adding a vector to a point gives a point.

Linearization is a simple process when considered in coordinates. In the most common case, the affine space P comes equipped with a Cartesian coordinate system. This means that a particular point $\mathbf{z}$ in P has been chosen along with a basis of vectors $\delta_1, \ldots, \delta_p$ for the linear space P_0. Every point $\mathbf{u}$ in P can then be written uniquely in the form

$$\mathbf{u} = \langle u^1, \ldots, u^p \rangle = \mathbf{z} + \sum_{1 \le i \le p} u^i \delta_i.$$

(The coordinates of u are indexed with superscripts rather than with subscripts because, as we will see later, a point u in P is actually a contravariant 1-tensor on P, and convention dictates that contravariant indices go upstairs.) To extend this coordinate system into a coordinate system on all of $P_\star$, it suffices to allow the coefficient of $\mathbf{z}$ to be different from 1. That is, an arbitrary element e of the linearization $P_\star$ can be written uniquely in the form

$$e = \langle e^1, \ldots, e^p; e^0 \rangle = e^0 \mathbf{z} + \sum_{1 \le i \le p} e^i \delta_i.$$

Note that the new coordinate e^0 measures weight; that is, $\omega(e) = e^0$. Following standard practice, we shall write the weight coordinate last, even though we index it with 0.

Exercise 6.2 Show that a barycentric coordinate system for an affine space P can be extended to its linearization $P_\star$ simply by removing the restriction that the coordinates must sum to 1. In the resulting coordinate system for $P_\star$, show that weight is given by the sum of the coordinates.

Note: So far, we have been writing particular points in the affine line L simply as real numbers, tacitly assuming a Cartesian coordinate system on L. But once we remove our blinders to reveal that L is actually a line lying in the 2-dimensional linear space $L_\star$, we must distinguish between the real number u and the point in L whose coordinate is u, which we shall do by writing the latter as $\bar{u}$. That is, we have $\bar{u} = \langle u; 1 \rangle$. For example, as shown in Figure 4, the expression $\bar{3}$ denotes a point in $L = L_1$, while $\bar{1} + \bar{2} = 2\,\overline{1.5}$ denotes a 2-heavy element of $L_\star$. Also, we shall use δ to denote the unit vector on L with positive sense; that is, $\delta = \langle 1; 0 \rangle = \bar{1} - \bar{0}$.

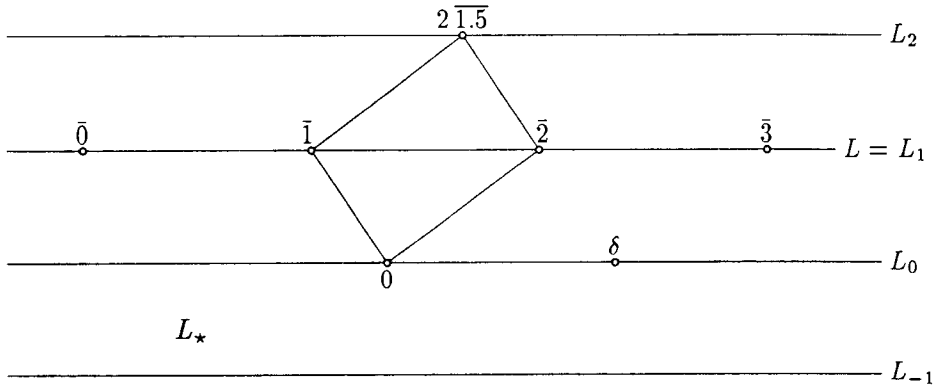


Fig. 4: The linearization $L_\star$ of the affine line L

Next, we consider an affine map $F\colon P \to Q$ from one affine space to another. Homogenizing the map F means extending it to a linear map $F_\star\colon P_\star \to Q_\star$ from one linearization to the other. To see that $F_\star$ exists and is unique, choose an affine frame $[\mathbf{r}_0, \ldots, \mathbf{r}_p]$ for P. Since $[\mathbf{r}_0, \ldots, \mathbf{r}_p]$ is also a linear basis for $P_\star$, the conditions $F_\star(\mathbf{r}_i) := F(\mathbf{r}_i)$ determine a unique linear map $F_\star$. This linear map $F_\star$ must agree with F on all of P because every point in P is an affine combination of the $\mathbf{r}_i$.

In addition to being linear, the homogenized map $F_\star$ is special in another way: it is *weight-preserving*. That is, since $\omega(F(\mathbf{u})) = \omega(\mathbf{u}) = 1$ for all $\mathbf{u}$ in P, we must have $\omega(F_\star(e)) = \omega(e)$ for all e in $P_\star$. Thus, the homogenized forms of affine maps are weight-preserving, linear maps.

In coordinates, homogenizing a map corresponds to making its defining polynomials be homogeneous. For example, consider the affine map $G(t) := 2t + 3$ from L to Q, where $\dim(Q) = 1$. More formally, we have $G(\bar{t}) = \overline{2t+3}$, that is to say, $G(\langle t; 1\rangle) = \langle 2t+3; 1\rangle$. To homogenize G, we replace the weight coordinate of the argument by a variable, say w, and we add a factor of w to each constant term in the defining polynomials: $G_\star(\langle t; w\rangle) = \langle 2t + 3w; w\rangle$. Note that $G_\star$ has the matrix $\left(\begin{smallmatrix} 2 & 0 \\ 3 & 1 \end{smallmatrix}\right)$ (or perhaps $\left(\begin{smallmatrix} 2 & 3 \\ 0 & 1 \end{smallmatrix}\right)$, depending upon your conventions); the last column has the special form $\binom{0}{1}$ precisely because $G_\star$ is weight-preserving.

In an analogous way, homogenizing an n-ic polynomial map $F\colon P \to Q$ means extending it to become a homogeneous n-ic polynomial map $F_\star\colon P_\star \to Q_\star$. There is always a unique way to perform this extension, and the resulting extension $F_\star$ has the special property that it is *weight-exponentiating*, that is, $\omega(F_\star(e)) = \omega(e)^n$ for all e in $P_\star$. In coordinates, we multiply each term in the defining polynomials by whatever power of the weight coordinate of the argument is necessary to bring the total degree up to n. For example, the homogenized form of the standard parabola $G(\langle t; 1\rangle) = \langle t, t^2; 1\rangle$ is given by $G_\star(\langle t; w\rangle) = \langle tw, t^2; w^2\rangle$.

The linear world has a simpler formal structure than the affine world, but the affine world involves one less dimension, because the affine blinders prevent us from seeing all of the elements whose weights differ from 1. Practitioners of CAGD prefer to work in the affine world when possible in order to save that dimension. Since they deal primarily with curves and surfaces, it is a significant conceptual advantage for them that the corresponding domain spaces in the affine world have dimensions 1 and 2, rather than 2 and 3. Mathematicians, on the other hand, deal with general k-folds. For them, the formal simplicity of the linear world outweighs the disadvantage that the domain of the linear-world version of a k-fold has dimension $k+1$. Once we understand homogenization, we can get the best of both worlds by moving back and forth as appropriate.

How does homogenization interact with polarization, for polynomial curves? If $F: L \to Q$ is an n-ic curve, we can polarize F to produce $f: L^n \to Q$ or we can homogenize F to produce $F_\star: L_\star \to Q_\star$. Using both polarization and homogenization, in either order, we can also transform F into a symmetric, multilinear map $f_\star: (L_\star)^n \to Q_\star$. For example, the standard parabola $G(t) = \langle t, t^2 \rangle$ has the homogenized polar form $g_\star$ given by

$$g_\star(\langle t_1; w_1 \rangle, \langle t_2; w_2 \rangle) = \left\langle \frac{t_1 w_2 + t_2 w_1}{2}, t_1 t_2; w_1 w_2 \right\rangle .$$

A homogenized polar form $f_\star$ is *weight-multiplicative*, that is, it satisfies the identity $\omega(f_\star(e_1, \ldots, e_n)) = \prod_i \omega(e_i)$.

Thus, a polynomial curve F has four different forms—either F, $F_\star$, f, or $f_\star$—depending upon what combination of homogenization and polarization have been applied. We shall call these forms the *diagonal form*, the *homogenized diagonal form*, the *polar form*, and the *homogenized polar form*. In the next two sections, we will arrive at our grand total of six forms by adding two more to the list: the *tensor form* $f^\otimes$ and the *homogenized tensor form* $f^\otimes_\star$.

7 Modeling symmetric tensors as Z-polynomials

If $F: L \to Q$ is an n-ic curve, the homogenized polar form of F is a multilinear map $f_\star: (L_\star)^n \to Q_\star$. We would like to understand $f_\star$ thoroughly. The tensor product construction is a general technique in mathematics that replaces multilinear maps with linear maps on a more complicated domain space [18]. It behooves us to apply the tensor product construction to $f_\star$.

For readers who already understand tensors, the only challenge involved—and it isn't hard—is the symmetry of $f_\star$. As a result of that symmetry, we will work with symmetric tensors and a symmetrized form of the tensor product operation, analogous to the alternating tensors and the alternized form of the tensor product that are used in integration theory.

Unfortunately, tensors are a wee bit subtle and many people do not understand them or are even scared by them. Spline theory should be a good chance

for such people to get over their fears, because the symmetric variant of the tensor product construction is easier to understand than the asymmetric or alternating variants. In each case, tensors behave algebraically like some flavor of polynomials. In the symmetric case, it is ordinary polynomials; in the asymmetric case, it is polynomials in non-commuting variables; in the alternating case, it is polynomials in skew-commuting variables.

One way to start thinking about tensors is to consider putting a geometric structure onto bags of polar arguments. Let $F: L \to Q$ be a cubic curve, and let $f: L^3 \to Q$ be its polar form. Regardless of which F we have chosen, we know that

$$f(\bar{1},\bar{4},\bar{5}) = \frac{f(\bar{0},\bar{4},\bar{5}) + f(\bar{2},\bar{4},\bar{5})}{2}.$$

(Recall that we now distinguish between the point $\bar{u}$ in L and the scalar u.) We also know more complicated facts, such as the relationship

$$f(\bar{r},\bar{r},\bar{r}) = (1-r)^3 f(\bar{0},\bar{0},\bar{0}) + 3r(1-r)^2 f(\bar{0},\bar{0},\bar{1}) + 3r^2(1-r) f(\bar{0},\bar{1},\bar{1}) + r^3 f(\bar{1},\bar{1},\bar{1}).$$

Note that f is a free variable in these formulas. It would be simpler if the bags of polar arguments themselves actually lay in some affine space. Then, a statement like

$$(\bar{1},\bar{4},\bar{5}) = \frac{(\bar{0},\bar{4},\bar{5}) + (\bar{2},\bar{4},\bar{5})}{2},$$

without the f's, could be interpreted as a geometric fact about that space. Thus, we would like to view the bags $(\bar{u},\bar{v},\bar{w})$ as points arranged in an affine space in such a way that precisely the same geometric facts hold of the points $(\bar{u},\bar{v},\bar{w})$ as must hold of the image points $f(\bar{u},\bar{v},\bar{w})$ for any cubic curve F.

First warning: One obvious candidate for an affine space containing the bags $(\bar{u},\bar{v},\bar{w})$ is the Cartesian power space L^3. But L^3 is *not* the space that we desire. It is true that the relationship $(\bar{1},\bar{4},\bar{5}) = \big((\bar{0},\bar{4},\bar{5}) + (\bar{2},\bar{4},\bar{5})\big)/2$, which must hold in the space we desire, does hold in L^3. But the relationship $(\bar{0},\bar{1},\bar{2}) = (\bar{0},\bar{2},\bar{1})$ must also hold in the space we desire, and it doesn't hold in L^3. Furthermore, the relationship $(\bar{1},\bar{1},\bar{1}) = \big((\bar{0},\bar{0},\bar{0}) + (\bar{2},\bar{2},\bar{2})\big)/2$ must not hold in the space we desire, and it does hold in L^3.

The space that we desire is actually the *third symmetric tensor power* of L, which we shall write $\bigodot^3 L$. In this section, however, instead of being abstract and talking about tensors, we will be concrete and model $\bigodot^3 L$ as a certain space of formal polynomials. Don't be fooled into thinking that this concrete model is the sole truth. There are other models, and, from an abstract point of view, no one model is any better than any other.

In our concrete model, $\bigodot^3 L := \{Z^3 + aZ^2 + bZ + c \mid a,b,c \in \mathbb{R}\}$ will denote the affine 3-space consisting of all monic, cubic polynomials in a formal variable Z. We will view a bag $(\bar{u},\bar{v},\bar{w})$ as corresponding to the polynomial $(Z-u)(Z-v)(Z-w)$ in $\bigodot^3 L$. That is, $(\bar{u},\bar{v},\bar{w})$ corresponds to the unique monic, cubic Z-polynomial that has the three real numbers u, v, and w as its

roots. Note that the Z-polynomial $(Z-1)(Z-4)(Z-5)$ does lie halfway in between $(Z-0)(Z-4)(Z-5)$ and $(Z-2)(Z-4)(Z-5)$ in $\bigodot^3 L$. Similarly, note that $(Z-0)(Z-1)(Z-2) = (Z-0)(Z-2)(Z-1)$. But $(Z-1)^3$ is not halfway in between $(Z-0)^3$ and $(Z-2)^3$. So far, so good.

Second warning: Every bag $(\bar{u}, \bar{v}, \bar{w})$ corresponds to a Z-polynomial in $\bigodot^3 L$, but not every Z-polynomial in $\bigodot^3 L$ corresponds to a bag. That is to say, not every monic, cubic polynomial has three real roots. This is one of the main subtleties behind tensors: In order to turn the collection of all bags $(\bar{u}, \bar{v}, \bar{w})$ into an affine space with the proper structure, we have to add in some other points as well, points that don't correspond to any bag. A Z-polynomial that has three real roots is called a *simple 3-tensor*, while the others are *compound 3-tensors*. Remember: Not every tensor is simple.

Math note: We could eliminate the problem of compound tensors when computing the tensor powers $\bigodot^n L$ of the affine line L by enlarging our coefficient field from the real numbers to the complex numbers. But if $\dim(P) > 1$, there is no way to avoid compound tensors when computing the tensor powers $\bigodot^n P$ of P, even if the coefficient field is algebraically closed. That is, most multivariate polynomials don't split into affine factors, even over the complex numbers.

The following theorem shows that the space of Z-polynomials $\bigodot^n L$ has precisely the geometric structure that we desired.

Theorem 7.1 *Let L be the affine line, let $\bigodot^n L$ denote the space of all monic, n-ic polynomials in the formal variable Z, and let Q be any affine space. Symmetric, n-affine maps $f\colon L^n \to Q$ and affine maps $f^{\odot}\colon \bigodot^n L \to Q$ are equivalent in the sense that, given a map of either type, a unique map of the other type exists that satisfies the correspondence identity $f(\bar{u}_1, \ldots, \bar{u}_n) = f^{\odot}\big((Z-u_1)\cdots(Z-u_n)\big)$.*

Proof: The easy direction of the proof assumes that we are given the affine map $f^{\odot}$. The correspondence identity then determines all of the values of f, and it does so in a way that guarantees that f will be symmetric and n-affine.

The hard direction starts with a symmetric, n-affine function f. By evaluating f on the $n+1$ argument bags that consist entirely of $\bar{0}$'s and $\bar{1}$'s, we can determine the values that we would like $f^{\odot}$ to have on the polynomials $(Z-0)^{n-i}(Z-1)^i = Z^{n-i}(Z-1)^i$ for i in $[0..n]$. Since the matrix of coefficients of these polynomials is triangular with nonzeros on the diagonal, we deduce that these polynomials form an affine frame for $\bigodot^n L$. Therefore, there exists a unique affine map $f^{\odot}$ that has the correct values on these $n+1$ polynomials. It remains to check that the correspondence identity always holds.

From whatever function $f^{\odot}$ we have obtained, we can use the correspondence identity as in the easy direction of the proof to define a symmetric, n-affine function g. This function g must agree with the original f at least on the argument bags that consist entirely of $\bar{0}$'s and $\bar{1}$'s. In other words, f and g must have the same Bézier points. But then the plain case of de Casteljau's

Algorithm shows that f and g must be identical, so the correspondence identity always holds. □

Exercise 7.2 If f and $f^\odot$ are maps that make the correspondence identity of Theorem 7.1 hold in every diagonal case, that is, $F(\bar{u}) = f^\odot\big((Z-u)^n\big)$ for all u, show that every polar case must also hold.

Definition 7.3 If $F\colon L \to Q$ is a polynomial curve of degree at most n, the affine map $f^\odot\colon \bigodot^n L \to Q$ that corresponds to F via the identity $F(\bar{u}) = f^\odot\big((Z-u)^n\big)$ is the *tensor form* of F. An argument to $f^\odot$ is a *tensor argument* to F, and a value of $f^\odot$ is a *tensor value* of F.

Polar values $f(\bar{u}_1, \ldots, \bar{u}_n)$ of F are the special case of tensor values $f^\odot(e)$ in which the tensor argument $e = (Z-u_1)\cdots(Z-u_n)$ is a simple n-tensor. Diagonal values $F(\bar{u})$ of F are the special case of tensor values in which the tensor argument $e = (Z-u)^n$ is a perfect n^{th} power.

Exercise 7.4 Let $G\colon L \to Q$ be the standard parabola in the plane Q given by $G(t) := \langle t, t^2\rangle$. Show that $g^\odot(Z^2 + aZ + b) = \langle -a/2, b\rangle$. Conclude that $g^\odot$ is an affine isomorphism between $\bigodot^2 L$ and Q. The points of Q inside the parabola are precisely the tensor values $g^\odot(e)$ of G whose tensor arguments e in $\bigodot^2 L$ are compound.

8 Adding X's and Y's to the Z-polynomials

The Z-polynomials are a satisfactory concrete model for the symmetric tensor powers $\bigodot^n L$ of the affine line L. But we have to use somewhat different polynomials when modeling either the tensor powers $\bigodot^n P$ of an affine space P with $\dim P > 1$ or the tensor powers of a linear space.

First, let's pause to refine our understanding of the Z-polynomial model. Theorem 7.1 associates the Z-polynomial $(Z-u_1)\cdots(Z-u_n)$ with the bag of polar arguments $(\bar{u}_1, \ldots, \bar{u}_n)$, all in one step. It is simpler instead to associate, for all u, the affine Z-polynomial $(Z-u)$ with the point $\bar{u}$ in L. We then define the Z-polynomial associated with a bag of points to be the product of the Z-polynomials associated with the points separately. With this convention, the Z-polynomial model for the symmetric tensors boils down to the simple rule $\bar{u} \mapsto (Z-u)$.

Putting things another way, we can use the association $\bar{u} \mapsto (Z-u)$ to define a commutative and associative product operation on the points of the line L. Each point corresponds to a monic, affine Z-polynomial, while a product of n points corresponds to a monic Z-polynomial of degree n. This multiplication is called the *symmetric tensor product* at the element level. It doesn't seem to have a standard symbol; we shall write $\bar{u}_1 \odot \cdots \odot \bar{u}_n$ for the symmetric tensor product of the points $\bar{u}_i$. In the model $\bar{u} \mapsto (Z-u)$, the tensor product $\bar{u}_1 \odot \cdots \odot \bar{u}_n$ is modeled by the formal polynomial $(Z-u_1)\cdots(Z-u_n)$.

The rule $\bar{u} \mapsto (Z - u)$ isn't the only way to model tensors as formal polynomials. The rule $\bar{u} \mapsto (Z + u)$ would work just as well, for example. In the model that results from this rule, it is the negatives of the roots of a Z-polynomial that are the points in the associated bag, rather than the roots themselves.

Homogenization provides a more interesting alternative model. Let X be a second formal variable, and consider the model $\bar{u} \mapsto (uX + Z)$. A bag of n points is associated, in this model, with a formal polynomial that is homogeneous of degree n in X and Z and in which the coefficient of Z^n is 1.

One advantage of the homogenized model $\bar{u} \mapsto (uX + Z)$ is that it allows us to combine points by taking linear combinations, as well as affine combinations. Consider, for example, the linear combination $\bar{3} - \bar{2} = \delta$, where $\delta := \langle 1; 0 \rangle = \bar{1} - \bar{0}$ denotes the unit vector on L that has positive sense, as shown in Figure 4. In our original model $\bar{u} \mapsto (Z - u)$, the vector $\delta = \bar{3} - \bar{2}$ is modeled by the formal polynomial $(Z - 3) - (Z - 2) = -1$, that is, by the scalar -1. It seems unwise to use scalars as models for anything but themselves. The homogenized model $\bar{u} \mapsto (uX + Z)$, on the other hand, models δ by $(3X + Z) - (2X + Z) = X$, which is fine.

In fact, the homogenized model $\bar{u} \mapsto (uX + Z)$ extends easily to give us a model for symmetric tensors in the linear world, that is, tensors on $L_\star$: With each element $\langle t; w \rangle$ of $L_\star$, we associate the polynomial $(tX + wZ)$. Having made the association $\langle t; w \rangle \mapsto (tX + wZ)$, we can then combine the elements of $L_\star$ however we like with addition, scalar multiplication, and the symmetric tensor product. The resulting abstract algebra is called the *symmetric tensor algebra* on $L_\star$ [16], which we shall write $\bigodot L_\star$. We are modeling $\bigodot L_\star$ as the algebra of all polynomials in X and Z. An *n-tensor* on L is an element of $\bigodot L_\star$ whose modeling polynomial is homogeneous of degree n. In particular, a 1-tensor on L is just an element of $L_\star$. The *weight* of an n-tensor is the coefficient of Z^n in its formal polynomial. The affine symmetric tensor power space $\bigodot^n L$ consists precisely of all 1-heavy n-tensors on L. The linear symmetric tensor power space $\bigodot^n L_\star$ consists of all n-tensors of any weight.

Exercise 8.1 Prove the linear-world variant of Theorem 7.1; that is, show that symmetric, n-linear maps $f_\star : (L_\star)^n \to Q_\star$ are equivalent to linear maps $f_\star^\odot : \bigodot^n L_\star \to Q_\star$.

A second advantage of the homogenized model $\bar{u} \mapsto (uX + Z)$ is that it extends easily to higher-dimensional affine spaces. Suppose that we want to model the symmetric tensor powers $\bigodot^n P$ of an affine plane P. Inventing a third formal variable Y, we can use the model $\langle u, v \rangle \mapsto (uX + vY + Z)$. More generally, we can associate the polynomial $(uX + vY + wZ)$ with the element $\langle u, v; w \rangle$ of the linearization $P_\star$. The linear space $\bigodot^n P_\star$ is then modeled as the set of all polynomials that are homogeneous of degree n in X, Y, and Z. The affine space $\bigodot^n P$ is the 1-heavy part of $\bigodot^n P_\star$, that is, those polynomials in $\bigodot^n P_\star$ whose coefficient of Z^n is 1.

Diagonal Form	**Homogenized Diagonal Form**
$F: L \to Q$	$F_\star: L_\star \to Q_\star$
degree at most n	homogeneous of degree n weight-exponentiating
Polar Form	**Homogenized Polar Form**
$f: L^n \to Q$	$f_\star: (L_\star)^n \to Q_\star$
symmetric, n-affine	symmetric, n-linear weight-multiplicative
Tensor Form	**Homogenized Tensor Form**
$f^\odot: \bigodot^n L \to Q$	$f_\star^\odot: \bigodot^n L_\star \to Q_\star$
affine	linear weight-preserving

Fig. 5: The six forms of the polynomial curve F

Exercise 8.2 If P is an affine plane, the tensor square $\bigodot^2 P$ of P under the model $\langle u, v\rangle \mapsto (uX + vY + Z)$ is the 5-dimensional affine space

$$\bigodot^2 P = \{aXY + bXZ + cYZ + dX^2 + eY^2 + Z^2 \mid a, b, c, d, e \in \mathbb{R}\}.$$

Note that at most 4 dimension's worth of the polynomials in this space can model simple 2-tensors on P, since the expression $(u_1X + v_1Y + Z)(u_2X + v_2Y + Z)$ has only four parameters. Characterize the simplicity of a 2-tensor on P as an algebraic relationship among the coefficients a through e. (Answer: $a^2 + b^2e + c^2d = abc + 4de$.)

Concrete models are a good way to develop one's intuition about tensors, especially in the symmetric case, where they can be based on formal polynomials. It is also important to understand tensors from an abstract perspective, however. We tackle that task in the next section. In particular, we will go to some lengths to clarify the analogy between the symmetric case of the tensor product construction, which we need, and the alternating case, which arises in integration theory. We note in passing, however, that we can now convert any polynomial curve into six different forms, as shown in Figure 5. Polarization takes us from the first row to the second row; tensoring, from the second row to the third row; homogenizing, from the first column to the second column.

9 Symmetric tensors in the abstract

Let P be an affine space. From an abstract point of view, the n^{th} *symmetric tensor power* of P is an affine space A together with a symmetric, multiaffine map $\varphi: P^n \to A$ that has the following property:

For any affine space Q and for any symmetric, multiaffine map $f: P^n \to Q$, there exists a unique affine map $f^{\odot}: A \to Q$ that satisfies $f = f^{\odot} \circ \varphi$.

This property implies that the maps f and $f^{\odot}$ are in one-to-one correspondence.

If any such pair $[A, \varphi]$ of a space A and map φ exists, it is easy to show that any other such pair $[B, \psi]$ will be isomorphic to it in a unique way. Therefore, it makes sense to reserve the special name $\bigodot^n P$ for the space A and the special notation $(\mathbf{u}_1, \ldots, \mathbf{u}_n) \mapsto \mathbf{u}_1 \odot \cdots \odot \mathbf{u}_n$ for the map φ. In the last two sections, we constructed a concrete model for A and φ based on formal polynomials and hence demonstrated that the symmetric tensor power does exist.

The tensor product construction comes in many different flavors. Perhaps the most basic is the one that converts a multilinear map $h: V_1 \times \cdots \times V_n \to W$ (not necessarily symmetric) into a linear map $h^{\otimes}: V_1 \otimes \cdots \otimes V_n \to W$ via the correspondence identity $h(\mathbf{v}_1, \ldots, \mathbf{v}_n) = h^{\otimes}(\mathbf{v}_1 \otimes \cdots \otimes \mathbf{v}_n)$. If the special case in which $V_1 = \cdots = V_n = V$, the domain of $h^{\otimes}$ is written $\bigotimes^n V$ and its elements are called *n-contravariant tensors* on V.

Math note: One advantage of the linear world over the affine world is that the linear world has a duality: The dual space V^* of the linear space V is the set of all linear functionals from V to $\mathbb{R}$. The elements of the asymmetric tensor product space

$$\bigotimes_n^m V := \underbrace{V \otimes \cdots \otimes V}_{m} \otimes \underbrace{V^* \otimes \cdots \otimes V^*}_{n}$$

are called *m-contravariant, n-covariant tensors* on V. That is, in tensor theory, *contravariant* means primal and *covariant* means dual. (See Dodson and Poston [13] for the story of how those names were chosen.) All of the tensors that we use in this report are purely primal, that is, purely contravariant. We shall therefore abbreviate the term "n-contravariant tensor" to "n-tensor."

If $[\delta_1, \ldots, \delta_p]$ is a basis for V, every n-tensor e in $\bigotimes^n V$ can be written uniquely in the form

$$e = \sum_{1 \le k_1, \ldots, k_n \le p} e^{k_1 \ldots k_n} \, \delta_{k_1} \otimes \cdots \otimes \delta_{k_n}.$$

One way to think about this is to say that the tensor e can be modeled as a formal polynomial that is homogeneous of degree n in the non-commuting formal variables δ_1 through δ_p. It is more common, however, to think about the p^n coordinates $(e^{k_1 \ldots k_n})$ as forming an n-dimensional array. It is this latter point of view that lies behind the physics definition of a tensor as an array of numbers that transforms according to a certain rule.

A particular n-tensor e is called *symmetric* if its coordinate array is symmetric, that is, if interchanging the values of k_i and k_j does not change the value of the coordinate $e^{k_1 \ldots k_n}$. The n-tensor e is called *alternating* if its coordinate array is skew-symmetric, that is, if interchanging k_i and k_j negates $e^{k_1 \ldots k_n}$ (and hence $e^{k_1 \ldots k_n} = 0$ when $k_i = k_j$ for $i \neq j$).

Math note: Things might be simpler if the tensor e were also called "skew-symmetric," instead of being called "alternating." The term "alternating" is preferred, however, because of a distinction that arises in the case of a coefficient field of characteristic 2. A bilinear map $h: V^2 \to W$ is called *alternating* if $h(\mathbf{v}, \mathbf{v}) = 0$ for all $\mathbf{v}$, while h is called *skew-symmetric* if $h(\mathbf{u}, \mathbf{v}) + h(\mathbf{v}, \mathbf{u}) = 0$ for all $\mathbf{u}$ and $\mathbf{v}$. If the characteristic is not 2, the two notions are equivalent. In characteristic 2, however, "skew-symmetric" and "symmetric" are equivalent; every alternating map is symmetric; but not every symmetric map is alternating.

Integration theory deals with multilinear functions that are alternating, such as the determinant. (Actually, the things that one integrates over a manifold are differential forms, which are alternating covariant tensor fields. Here, we discuss the simpler case of alternating contravariant tensors.) The alternating case of the tensor product construction converts an alternating, multilinear map $h: V^n \to W$ into a linear map $h^\wedge: \bigwedge^n V \to W$ via the correspondence $h(\mathbf{v}_1, \ldots, \mathbf{v}_n) = h^\wedge(\mathbf{v}_1 \wedge \cdots \wedge \mathbf{v}_n)$. The space $\bigwedge^n V$ is called the n^{th} *alternating tensor power* of V or the n^{th} *exterior power* of V.

If $[\delta_1, \ldots, \delta_p]$ is a basis for V as above, each element e of the alternating tensor power space $\bigwedge^n V$ can be written uniquely in the form

$$e = \sum_{1 \le k_1 < \cdots < k_n \le p} e^{k_1 \ldots k_n}\, \delta_{k_1} \wedge \cdots \wedge \delta_{k_n}.$$

We can think of this sum as a formal polynomial in the skew-commuting variables δ_1 through δ_p, or we can think of the coefficients $\left(e^{k_1 \ldots k_n}\right)$ for $1 \le k_1 < \cdots < k_n \le p$ as forming a triangular array. But the most common model for the space $\bigwedge^n V$ is yet a third alternative: We identify $\bigwedge^n V$ with the subset of $\bigotimes^n V$ consisting of all alternating n-tensors via the rule

$$\delta_{k_1} \wedge \cdots \wedge \delta_{k_n} := \mathrm{Alt}(\delta_{k_1} \otimes \cdots \otimes \delta_{k_n}). \tag{9.1}$$

In this rule, Alt denotes the alternizer, the function that projects all of $\bigotimes^n V$ down onto the subspace of alternating n-tensors; in particular,

$$\mathrm{Alt}(\delta_{k_1} \otimes \cdots \otimes \delta_{k_n}) = \frac{1}{n!} \sum_{\rho \in S_n} \mathrm{sgn}(\rho)\, \delta_{k_{\rho(1)}} \otimes \cdots \otimes \delta_{k_{\rho(n)}},$$

where S_n denotes the symmetric group on $\{1, \ldots, n\}$ and $\mathrm{sgn}(\rho)$ is $+1$ if the permutation ρ is even and -1 if ρ is odd. In the model resulting from Equation 9.1, the alternating tensor product at the element level is defined by alternizing the asymmetric tensor product. For example, $\bar{3} \wedge \bar{5} = \mathrm{Alt}(\bar{3} \otimes \bar{5}) = (\bar{3} \otimes \bar{5} - \bar{5} \otimes \bar{3})/2$. Converting to the standard Cartesian basis $[\bar{0}, \delta]$ for $L_\star$, we have

$$\bar{3} \wedge \bar{5} = \frac{(\bar{0} + 3\delta) \otimes (\bar{0} + 5\delta) - (\bar{0} + 5\delta) \otimes (\bar{0} + 3\delta)}{2} = \bar{0} \otimes \delta - \delta \otimes \bar{0}.$$

The polarized approach to splines deals with multilinear functions that are symmetric, such as the homogenized polar form $f_\star$ of a polynomial curve. The appropriate theory is completely analogous to the alternating case. We can convert a symmetric, multilinear map $h: V^n \to W$ into a linear map $h^\odot: \bigodot^n V \to W$ via the correspondence $h(\mathbf{v}_1, \ldots, \mathbf{v}_n) = h^\odot(\mathbf{v}_1 \odot \cdots \odot \mathbf{v}_n)$, where the space $\bigodot^n V$ is the n^{th} *symmetric tensor power* of V.

Each element e of $\bigodot^n V$ can be uniquely expressed in the form

$$e = \sum_{1 \le k_1 \le \cdots \le k_n \le p} e^{k_1 \ldots k_n} \, \delta_{k_1} \odot \cdots \odot \delta_{k_n}.$$

We can think of this sum as a formal polynomial in the commuting variables δ_1 through δ_p, as we did in Section 8, or we can think of the coefficients $\left(e^{k_1 \ldots k_n}\right)$ for $1 \le k_1 \le \cdots \le k_n \le p$ as forming a triangular array. (The triangular array is a bit larger in the symmetric case, since the indices k_i are allowed to be equal.) A third alternative is to identify $\bigodot^n V$ with the subset of $\bigotimes^n V$ consisting of all symmetric n-tensors, via the rule

$$\delta_{k_1} \odot \cdots \odot \delta_{k_n} := \mathrm{Sym}(\delta_{k_1} \otimes \cdots \otimes \delta_{k_n}), \tag{9.2}$$

where Sym denotes the symmetrizing projection given by

$$\mathrm{Sym}(\delta_{k_1} \otimes \cdots \otimes \delta_{k_n}) := \frac{1}{n!} \sum_{\rho \in S_n} \delta_{k_{\rho(1)}} \otimes \cdots \otimes \delta_{k_{\rho(n)}}.$$

In this third view, the symmetric tensor product at the element level is defined by symmetrizing the asymmetric tensor product. For example, $\bar{3} \odot \bar{5} = \mathrm{Sym}(\bar{3} \otimes \bar{5}) = (\bar{3} \otimes \bar{5} + \bar{5} \otimes \bar{3})/2$. In terms of the Cartesian basis $[\bar{0}, \delta]$, we have $\bar{3} \odot \bar{5} = \bar{0} \otimes \bar{0} + 4\,\bar{0} \otimes \delta + 4\,\delta \otimes \bar{0} + 15\,\delta \otimes \delta$.

Math note: The nonzero coordinates of an alternating n-tensor are partitioned into equivalence classes of size $n!$, whose elements differ at most in sign. Some authors try to make certain coordinate formulas cleaner in the alternating case by multiplying the right-hand side of Equation 9.1 by $n!$ [28]. The price that they pay for this rescaling is a messier definition of the alternating tensor product at the element level. In particular, if e_1 is an alternating k-tensor and e_2 is an alternating l-tensor, the rescaled model has $e_1 \wedge e_2 = \left((k+l)!/k!\,l!\right) \mathrm{Alt}(e_1 \otimes e_2)$. Other authors get the same rescaled model in a different way by omitting the division by $n!$ in the definition of Alt [18]. Indeed, that division must be omitted in the case of finite characteristic, where we might have $n! = 0$. But omitting it destroys the desirable property that $\mathrm{Alt}(e) = e$ for an alternating tensor e.

A symmetric n-tensor can have all of its coordinates nonzero, and the equivalence classes into which they are partitioned are of different sizes, depending upon how many indices are repeated and how often. Hence, the analogous rescaled model is less attractive in the symmetric case.

The asymmetric and symmetric cases of the tensor product construction make perfect sense in the affine world, as well as in the linear world, and we shall use the same symbols: $\otimes$ and $\odot$. Defining the alternating case in the affine world would be tricky, since the fundamental condition $h(\mathbf{v},\mathbf{v}) = 0$ is not an affine relation; we won't bother to try.

Exercise 9.3 If P is an affine space of dimension p and Q is an affine space of dimension q, verify the following dimension formulas by presenting either a linear basis or an affine frame, as appropriate:

$$\begin{aligned} \dim(P_\star \otimes Q_\star) &= (p+1)(q+1) & \dim(P \otimes Q) &= pq + p + q \\ \dim(\textstyle\bigotimes^n P_\star) &= (p+1)^n & \dim(\textstyle\bigotimes^n P) &= (p+1)^n - 1 \\ \dim(\textstyle\bigwedge^n P_\star) &= \tbinom{p+1}{n} & &\text{(not applicable)} \\ \dim(\textstyle\bigodot^n P_\star) &= \tbinom{p+n}{n} & \dim(\textstyle\bigodot^n P) &= \tbinom{p+n}{n} - 1. \end{aligned}$$

10 Poles and pole systems

It is time for us to reap some reward from our investment in the conceptual complexity of tensors. In this section, we use tensors to complete the proof of Theorem 4.2 via the conceptual approach. We first pause to promulgate a valuable notational convention.

The symmetric tensor product operation at the element level is a commutative product; for example, $\bar{2} \odot \bar{3} = \bar{3} \odot \bar{2}$. As a result, it doesn't hurt to omit the special multiplication symbol and to write this product simply $\bar{2}\,\bar{3}$. That is, when points, vectors, or other symmetric tensors are multiplied together henceforth, the symmetric tensor product is the intended flavor of multiplication. With this convention, we can use exponential notation to write the k^{th} Bézier point of the n-ic curve segment $F([0\,..\,1])$ in the form $f^\odot(\bar{0}^{\,n-k}\,\bar{1}^{\,k})$, which is a dramatic improvement over our previous notation

$$f(\underbrace{\bar{0},\ldots,\bar{0}}_{n-k},\underbrace{\bar{1},\ldots,\bar{1}}_{k}) = f^\odot(\underbrace{\bar{0}\odot\cdots\odot\bar{0}}_{n-k}\odot\underbrace{\bar{1}\odot\cdots\odot\bar{1}}_{k}).$$

Note that the success of this convention depends critically on the fact that we are distinguishing between the scalar u and the point $\bar{u}$ in the affine line L. For example, $\bar{2}\,\bar{3}$ is a 1-heavy 2-tensor on L; it is not at all the same thing as $\bar{6}$.

Recall that Theorem 4.2 dealt with the progressive case of the de Casteljau Algorithm. We suppose that $[\bar{r}_1,\ldots,\bar{r}_{2n}]$ is a progressable sequence of points in L, that Q is some affine space, and that $\mathbf{a}_0$ through $\mathbf{a}_n$ are points in Q. We would like to conclude that there exists a unique n-ic curve $F\colon L \to Q$ whose polar form satisfies $f(\bar{r}_{k+1},\ldots,\bar{r}_{k+n}) = \mathbf{a}_k$ for k in $[0\,..\,n]$. The uniqueness of F follows easily from the de Casteljau Algorithm, but we have not yet demonstrated the existence of F.

Proof (of existence in Theorem 4.2): Using polarization and the tensor product construction, we know that n-ic curves $F: L \to Q$ are equivalent to affine maps $f^{\odot}: \bigodot^n L \to Q$ via the correspondence $F(\bar{u}) = f^{\odot}(\bar{u}^n)$. Thus, in tensored form, our problem is as follows: Show that there exists a unique affine map $f^{\odot}$ that satisfies the constraints $f^{\odot}(\bar{r}_{k+1} \cdots \bar{r}_{k+n}) = \mathbf{a}_k$ for k in $[0 \,..\, n]$. (By our new convention, the expression $\bar{r}_{k+1} \cdots \bar{r}_{k+n}$ is shorthand for the n-tensor $\bar{r}_{k+1} \odot \cdots \odot \bar{r}_{k+n}$ on L.)

But affine maps are quite simple things. Let $[e_0, \ldots, e_n]$ denote some sequence of n-tensors on L. Our ability to specify an affine map $f^{\odot}$ by specifying the values $f^{\odot}(e_k)$ depends upon the geometric relationships that hold among the e_k in $\bigodot^n L$. First, an affine map $f^{\odot}$ satisfying $f^{\odot}(e_k) = \mathbf{a}_k$ exists for all sequences $[\mathbf{a}_0, \ldots, \mathbf{a}_n]$ precisely when the e_k are affinely independent. Second, the equations $f^{\odot}(e_k) = \mathbf{a}_k$ are enough to completely determine the affine map $f^{\odot}$, leaving no room for further choices, precisely when the e_k affinely span all of $\bigodot^n L$.

The relevant n-tensors e_k for the proof of Theorem 4.2 are given by $e_k := \bar{r}_{k+1} \cdots \bar{r}_{k+n}$ for k in $[0 \,..\, n]$. From the progressive case of the de Casteljau Algorithm, we already know that the affine map $f^{\odot}$ is uniquely determined. Therefore, we may conclude that the tensors $e_k = \bar{r}_{k+1} \cdots \bar{r}_{k+n}$ must span all of $\bigodot^n L$. But there are $n+1$ of them, and $\bigodot^n L$ is an affine space of dimension n. By linear algebra, the only way that the e_k can span is if they are also affinely independent. Therefore, we may conclude that some satisfactory affine map $f^{\odot}$ will always exist. □

Through the tensor product construction, we can bring linear algebra to bear on the problem, and linear algebra lets us deduce by simple counting that F exists if and only if F is unique. Roughly speaking, since the $n+1$ constraints on F are spread out enough that they determine all of the values of F uniquely, they must also be spread out enough that they don't make any contradictory demands about any particular values of F.

Exercise 10.1 Reprove the fact that the n-tensors $e_k := \bar{r}_{k+1} \cdots \bar{r}_{k+n}$ are affinely independent by expanding them in coordinates and computing a determinant. In particular, use the formula $\bar{r}_k = \bar{0} + r_k \delta$ to show that

$$e_k = \bar{r}_{k+1} \cdots \bar{r}_{k+n} = \sum_{0 \le l \le n} \sigma_l(r_{k+1}, \ldots, r_{k+n})\, \bar{0}^{n-l} \delta^l,$$

where σ_l denotes the l^{th} elementary symmetric function, as in Section 5.2. Conclude that the e_k will be independent if and only if $\det(M) \neq 0$, where $M = (m_{kl})$ is the matrix given by $m_{kl} = \sigma_l(r_{k+1}, \ldots, r_{k+n})$ for k and l in $[0 \,..\, n]$. Prove that

$$\det(M) = \prod_{1 \le i \le j \le n} (r_{n+i} - r_j)$$

by imitating the computation of the Vandermonde determinant, and use the progressability of the sequence $[\bar{r}_1, \ldots, \bar{r}_{2n}]$ to conclude, once again, that the e_k are affinely independent.

Another way to describe Theorem 4.2 is as a result about "polar interpolation," as opposed to "diagonal interpolation." If $\bar{u}_0$ through $\bar{u}_n$ are $n+1$ distinct points in L and $\mathbf{a}_0$ through $\mathbf{a}_n$ are arbitrary points in an affine space Q, the Lagrange Interpolation Formula computes the unique n-ic curve $F: L \to Q$ satisfying the diagonal conditions $F(\bar{u}_i) = \mathbf{a}_i$. That is, the Lagrange Formula solves the problem of *diagonal interpolation*. In the corresponding problem of *polar interpolation*, we start with $n+1$ bags of polar arguments to F, say B_0 through B_n, and we try to specify a curve F using the polar conditions $f(B_k) = \mathbf{a}_k$. Let b_k denote the simple n-tensor on L that corresponds to the bag B_k. Polar interpolation is more subtle than diagonal interpolation because it is not enough for the bags B_k merely to be distinct. Instead, the tensors b_k must be affinely independent, so that they form an affine frame for $\bigodot^n L$. The situation in which the b_k do form a frame is important enough that it deserves its own nomenclature.

Definition 10.2 If B_0 through B_n are bags of n points in L with the property that the corresponding simple n-tensors b_0 through b_n form an affine frame for $\bigodot^n L$, then an n-ic curve F can be uniquely specified by arbitrarily specifying the polar values $f(B_k) = f^\odot(b_k)$. In this situation, we shall call the sequence of polar values $[f(B_0), \ldots, f(B_n)]$ a *pole system* for F. A *pole* is one of the polar values $f(B_k)$ in a pole system. Thus, a pole is a polar value that we may use as a handle to control the shape of the function.

The plain and progressive cases of the de Casteljau Algorithm are two sources of poles and pole systems.

Definition 10.3 A *plain frame* for $\bigodot^n L$ is an affine frame consisting of the tensors $\bar{s}^{n-k}\bar{t}^k$ for k in $[0 \,..\, n]$ and $\bar{s} < \bar{t}$. The images under $f^\odot$ of the elements of a plain frame are precisely the Bézier points of the n-ic segment $F([\bar{s} \,..\, \bar{t}\,])$; they form a *plain pole system*. Each Bézier point is a *plain pole*.

A *progressive frame* for $\bigodot^n L$ is an affine frame consisting of $\bar{r}_{k+1} \cdots \bar{r}_{k+n}$ for k in $[0 \,..\, n]$, where $[\bar{r}_1, \ldots, \bar{r}_{2n}]$ is a progressable sequence. The image under $f^\odot$ of the elements of a progressive frame form a *progressive pole system*, and each point in a progressive pole system is a *progressive pole*.

Note on "pole," "plain," and "progressive": De Casteljau introduced the term "pole" [11], but he meant by it any polar value, whether or not that polar value was part of a pole system. He also invented the wonderful name "progressive pole." I am proposing the name "plain pole" to replace his "simple pole," in order to avoid confusion with simple tensors and with the simple poles of a complex analytic function.

11 The polarized view of continuity

While the polar form of one polynomial curve in isolation is worth studying, the real power of polarization in the theory of splines comes from the simple way

that parametric continuity constraints at a joint between two polynomial pieces interact with their polar forms. Our next goal is to study that interaction.

Open problem 11.1 A simple interaction arises only when the continuity constraints are parametric. Several weaker notions of continuity for spline curves have generated a lot of interest recently, under the general rubric of "geometric continuity." One of those notions corresponds to parametric continuity after some regular reparameterization, while the other is a weaker notion proposed by Boehm [3]. Saying anything interesting about either type of geometric continuity from the polarized approach is an important open problem.

Before considering parametric continuity, we must establish some connection between polarizing and differentiating. This is easy to do, now that we also know about homogenizing and tensoring. Indeed, we wouldn't have been able to proceed without homogenizing in some form, since, if the values of a polynomial curve F are points in the affine space Q, then the values $F'(\bar{u})$ of the derivative of F are vectors on Q, that is, 0-heavy elements of the linearization $Q_\star$.

Let $F: L \to Q$ be an n-ic curve, and let $f_\star^\odot: \bigodot^n L_\star \to Q_\star$ be the homogenized tensor form of F. Since $F(\bar{r}) = f_\star^\odot(\bar{r}^n)$, we have

$$F'(\bar{r}) = \frac{d}{dr} f_\star^\odot(\bar{r}^n).$$

Since the map $f_\star^\odot$ is linear, it commutes with the operator d/dr. We then apply the rule for differentiating an n^{th} power:

$$F'(\bar{r}) = f_\star^\odot\left(\frac{d}{dr}\bar{r}^n\right) = f_\star^\odot\left(n\bar{r}^{n-1}\frac{d}{dr}\bar{r}\right).$$

The value $d\bar{r}/dr$ is simply $\delta = \langle 1;0\rangle$, the unit vector on L. The final result is

$$F'(\bar{r}) = n f_\star^\odot(\delta\bar{r}^{n-1}).$$

Thus, to evaluate the derivative of F at $\bar{r}$, we replace one of the n copies of $\bar{r}$ among the polar arguments of F by a vector that gives the direction of differentiation; and we multiply by an annoying scale factor of n.

The reason for the annoying factor is that, when we vary the argument $\bar{r}$ of $F(\bar{r})$, all n of the arguments of the polar form $f(\bar{r},\ldots,\bar{r})$ vary in parallel. If only one of them varied, the rate of change in the result would be

$$\frac{d}{du_i} f(\bar{u}_1,\ldots,\bar{u}_n) = f_\star(\bar{u}_1,\ldots,\bar{u}_{i-1},\delta,\bar{u}_{i+1},\ldots,\bar{u}_n).$$

But when all n of them vary in parallel, the rate of change is magnified by a factor of n.

If we compute higher derivatives by repeating the process above, we find that $F^{(k)}(\bar{r}) = n^{\underline{k}} f_\star^\odot(\delta^k\bar{r}^{n-k})$, where the *falling-factorial power* $n^{\underline{k}}$ out front is an

abbreviation for the annoying scale factor of $n^{\underline{k}} := n(n-1)\cdots(n-k+1)$. In particular, note that the n^{th} derivative of the n-ic polynomial F has the constant value $F^{(n)}(\bar{r}) = n! f^{\odot}_{\star}(\delta^n)$.

If we like, we can dehomogenize these formulas by rewriting the vector δ as a difference of points. For example, substituting $\delta := (\bar{t} - \bar{s})/(t-s)$ and $\bar{r} := \bar{s}$ gives the standard formula for the k^{th} end-point derivative of an n-ic segment $F([\bar{s} \,..\, \bar{t}\,])$ as a k^{th} difference of its plain poles (that is, of its Bézier points):

$$F^{(k)}(\bar{s}) = \frac{n^{\underline{k}}}{(t-s)^k} \sum_{0 \le i \le k} \binom{k}{i} (-)^{k-i} f^{\odot}(\bar{s}^{\,n-i}\, \bar{t}^{\,i}).$$

Exercise 11.2 Extend these differentiation results to the case of a map $F\colon P \to Q$ with $\dim(P) > 1$ whose total degree is bounded by n. For example, show that the second directional derivative $D_\xi D_\eta F(\mathbf{u})$ of F at $\mathbf{u}$ in the directions ξ and η (where ξ and η are vectors on P) is given by $D_\xi D_\eta F(\mathbf{u}) = n^{\underline{2}} f^{\odot}_{\star}(\xi \eta\, \mathbf{u}^{n-2})$.

Theorem 11.3 *Two polynomial curves $F(\bar{u})$ and $G(\bar{u})$, each of degree at most n, agree to k^{th} order at a point $\bar{r}$ if and only if their n-polar forms f and g have the property that*

$$f(\bar{u}_1, \ldots, \bar{u}_k, \underbrace{\bar{r}, \ldots, \bar{r}}_{n-k}) = g(\bar{u}_1, \ldots, \bar{u}_k, \underbrace{\bar{r}, \ldots, \bar{r}}_{n-k})$$

for all $\bar{u}_1$ through $\bar{u}_k$, that is, if and only if F and G agree on all bags of polar arguments that contain at most k values different from $\bar{r}$.

For example, in the case $k = 0$, we are told that F and G agree to 0^{th} order if and only if $f(\bar{r}, \ldots, \bar{r}) = g(\bar{r}, \ldots, \bar{r})$, that is, if and only if $F(\bar{r}) = G(\bar{r})$. In the case $k = n$, we are told that F and G agree to n^{th} order, that is, are identical, if and only if all of their polar values agree.

Proof: The curves F and G agree to k^{th} order at $\bar{r}$ precisely when $F^{(i)}(\bar{r}) = G^{(i)}(\bar{r})$ for i in $[0 \,..\, k]$. Translating the i^{th} condition into tensor terms gives $f^{\odot}_{\star}(\delta^i \bar{r}^{\,n-i}) = g^{\odot}_{\star}(\delta^i \bar{r}^{\,n-i})$; the annoying numeric factor of $n^{\underline{i}}$ appears on both sides, and hence cancels out. Thus, F and G agree to k^{th} order if and only if the relationship $f^{\odot}_{\star}(e\bar{r}^{\,n-k}) = g^{\odot}_{\star}(e\bar{r}^{\,n-k})$ holds whenever e has the form $e = \delta^i \bar{r}^{\,k-i}$ for i in $[0 \,..\, k]$.

On the other hand, the identity

$$f(\bar{u}_1, \ldots, \bar{u}_k, \underbrace{\bar{r}, \ldots, \bar{r}}_{n-k}) = g(\bar{u}_1, \ldots, \bar{u}_k, \underbrace{\bar{r}, \ldots, \bar{r}}_{n-k})$$

corresponds to the relationship $f^{\odot}_{\star}(e\bar{r}^{\,n-k}) = g^{\odot}_{\star}(e\bar{r}^{\,n-k})$ for $e = \bar{u}_1 \cdots \bar{u}_k$. Thus, the polar-form condition in Theorem 11.3 is satisfied if and only if the relationship $f^{\odot}_{\star}(e\bar{r}^{\,n-k}) = g^{\odot}_{\star}(e\bar{r}^{\,n-k})$ holds whenever e is the product of k points in L.

The set of all e in $\bigodot^k L_\star$ for which the identity $f_\star^\odot(e\bar{r}^{n-k}) = g_\star^\odot(e\bar{r}^{n-k})$ holds must be a linear subspace S of $\bigodot^k L_\star$. The k-tensors $\delta^i \bar{r}^{k-i}$ for i in $[0..k]$ form a basis for $\bigodot^k L_\star$. Thus, F and G agree to k^{th} order if and only if $S = \bigodot^k L_\star$. But there are also bases for $\bigodot^k L_\star$, such as $[\bar{s}^k, \bar{s}^{k-1}\bar{t}, \ldots, \bar{t}^k]$ for $\bar{s} \neq \bar{t}$, each of whose elements is a product of k points. Thus, it is also the case that the polar-form condition is satisfied if and only if $S = \bigodot^k L_\star$. We conclude that agreement to k^{th} order and the polar-form condition are equivalent. □

Exercise 11.4 Prove that two n-ic curves F and G agree to k^{th} order at $\bar{s}$ if and only if each of the first $k+1$ plain poles of the segment $F([\bar{s}\,..\,\bar{t}])$ coincides with the corresponding plain pole of $G([\bar{s}\,..\,\bar{t}])$.

We can rephrase the result in Theorem 11.3 as follows: Two n-ic curves agree to k^{th} order at $\bar{r}$ if and only if their polar forms agree on all bags of polar arguments that are superbags of

$$B = (\underbrace{\bar{r}, \ldots, \bar{r}}_{n-k}).$$

This concept is begging to be polarized.

Definition 11.5 If $F, G\colon L \to Q$ are two n-ic curves and $B = (\bar{r}_1, \ldots, \bar{r}_{n-k})$ is a bag of $n-k$ points in L, we say that F and G *agree on all superbags of* B if the identity

$$f(\bar{u}_1, \ldots, \bar{u}_k, \bar{r}_1, \ldots, \bar{r}_{n-k}) = g(\bar{u}_1, \ldots, \bar{u}_k, \bar{r}_1, \ldots, \bar{r}_{n-k})$$

holds for all $\bar{u}_1$ through $\bar{u}_k$ in L.

Note that the condition "agreement on all superbags of B" becomes stronger as the bag B gets smaller. For example, agreeing on all superbags of $(\bar{r}_1, \ldots, \bar{r}_n)$ means merely that a single polar value is the same, while agreeing on all superbags of the empty bag means identity. In the next section, we will find that any two segments of a spline curve agree on all superbags of the bag of intervening knots.

Exercise 11.6 Show that two cubic curves F and G agree on all superbags of $(\bar{u}, \bar{v})$ for $\bar{u} \neq \bar{v}$ if and only if each of the middle two plain poles of the segment $F([\bar{u}\,..\,\bar{v}])$ coincides with the corresponding plain pole of $G([\bar{u}\,..\,\bar{v}])$.

12 The polarized view of spline curves

We now tackle spline curves with parametric continuity constraints from the polarized approach.

Let $[\bar{t}_i]$ be a nondecreasing sequence of knots in the affine line L, under the standard convention that a knot $\bar{t}_{i+1} = \bar{t}_{i+2} = \cdots = \bar{t}_{i+m}$ of multiplicity $m \geq 1$ is the parameter value of a joint in the spline where C^{n-m} continuity is required. The largest multiplicity that we allow in a knot sequence is $m = n+1$, corresponding to C^{-1} continuity, that is, to a completely unconstrained joint.

An n-ic spline curve F with the knot sequence $[\bar{t}_i]$ is a piecewise polynomial function $F: L \to Q$ with one piece for each pair of distinct, adjacent knots. That is, if $\bar{t}_i < \bar{t}_{i+1}$, the spline curve F follows a single n-ic polynomial F_i over the nonempty interval $(\bar{t}_i \,..\, \bar{t}_{i+1})$. The n-ic segment $F_i([\bar{t}_i \,..\, \bar{t}_{i+1}])$ joins on, at its right end, to the left end of the segment $F_{i+m}([\bar{t}_{i+m} \,..\, \bar{t}_{i+m+1}])$, where m is the largest integer such that $\bar{t}_{i+1} = \cdots = \bar{t}_{i+m}$. In order for the assembled segments $[F_i]$ to form a spline, this joint is required to have at least C^{n-m} continuity. By Theorem 11.3, this is equivalent to the requirement that F_i and F_{i+m} must agree on all superbags of $(\bar{t}_{i+1}, \ldots, \bar{t}_{i+m})$.

Two non-adjacent segments F_i and F_j of a spline F are still related to each other, as long as the total number $j - i$ of intervening knots is at most n. One of the advantages of the polar approach to splines is that it gives us a simple way to describe the relationship between F_i and F_j in general, regardless of the multiplicities of the $j - i$ intervening knots: The curves F_i and F_j must agree on all superbags of $(\bar{t}_{i+1}, \ldots, \bar{t}_j)$.

Theorem 12.1 *Let $[\bar{t}_i]$ be a knot sequence in L and, for each i with $\bar{t}_i < \bar{t}_{i+1}$, let $F_i: L \to Q$ be an n-ic curve. The sequence of curves $[F_i]$ forms a spline with the knot sequence $[\bar{t}_i]$ if and only if the following condition holds: For all i and j with $i \leq j \leq i+n$, $\bar{t}_i < \bar{t}_{i+1}$, and $\bar{t}_j < \bar{t}_{j+1}$, the curves F_i and F_j agree on all superbags of the bag $(\bar{t}_{i+1}, \ldots, \bar{t}_j)$ of intervening knots.*

Proof: We already observed that Theorem 11.3 handles one direction of the proof. If the superbag condition holds for all pairs F_i and F_j, it follows that any two adjacent segments F_i and F_{i+m} must agree on all superbags of the bag $(\bar{t}_{i+1}, \ldots, \bar{t}_{i+m})$. By Theorem 11.3, this means that F_i and F_{i+m} agree to $(n-m)^{\text{th}}$ order at the separating knot $\bar{t}_{i+1} = \bar{t}_{i+m}$, so F_i and F_{i+m} do fit together properly.

In the other direction, suppose that all of the F_i fit together to form a spline; we must show that the superbag condition holds for all pairs of segments F_i and F_j. Let m_1, m_2, and so forth be the multiplicities of the knots that intervene between F_i and F_j, so that $\sum_k m_k = j - i$. Since the segment F_i joins its right-neighbor F_{i+m_1} with C^{n-m_1} continuity at $\bar{t}_{i+1} = \bar{t}_{i+m_1}$, Theorem 11.3 tells us that F_i and F_{i+m_1} must agree on all superbags of $(\bar{t}_{i+1}, \ldots, \bar{t}_{i+m_1})$. In a similar way, F_{i+m_1} must agree with its right-neighbor $F_{i+m_1+m_2}$ on all superbags of $(\bar{t}_{i+m_1+1}, \ldots, \bar{t}_{i+m_1+m_2})$. And so on. By transitivity, we conclude that F_i and F_j will agree on all superbags of $(\bar{t}_{i+1}, \ldots, \bar{t}_j)$. □

Suppose that F is an n-ic spline curve. What does Theorem 12.1 say about the values of the various polar forms f_i on the argument bag $B_k :=$

$(\bar{t}_{k+1},\ldots,\bar{t}_{k+n})$, which consists of n consecutive knots? If i and j are any two indices that lie in $[k\,..\,k+n]$, the bag $(\bar{t}_{i+1},\ldots,\bar{t}_j)$ will be a sub-bag of B_k. Therefore, by Theorem 12.1, all of the f_i for i in $[k\,..\,k+n]$ must agree on a common output value for the input bag B_k. The resulting polar values $\mathbf{a}_k := f_i(\bar{t}_{k+1},\ldots,\bar{t}_{k+n})$ are important, because they can be used as handles that control the shape of the spline curve F.

Theorem 12.2 *Let n be a nonnegative integer and let $[\bar{t}_k]$ be a knot sequence in the affine line L whose knots have multiplicity at most $n+1$. For any affine space Q and for any sequence $[\mathbf{a}_k]$ of points in Q, there exists a unique n-ic spline curve $F\colon L \to Q$ with knot sequence $[\bar{t}_k]$ that satisfies*

$$\mathbf{a}_k = f_i(\bar{t}_{k+1},\ldots,\bar{t}_{k+n}) \qquad \text{for} \quad k \le i \le k+n. \tag{12.3}$$

Proof: Let $(\bar{t}_i\,..\,\bar{t}_{i+1})$ be a nonempty knot interval, and consider the sequence of $2n$ knots $S := [\bar{t}_{i-n+1},\ldots,\bar{t}_{i+n}]$, centered at the interval $(\bar{t}_i\,..\,\bar{t}_{i+1})$. Since

$$\bar{t}_{i-n+1} \le \cdots \le \bar{t}_i < \bar{t}_{i+1} \le \cdots \le \bar{t}_{i+n},$$

every element of the first half of S is distinct from every element of the second half, which implies that S is progressable. Hence, the points $f_i(\bar{t}_{k+1},\ldots,\bar{t}_{k+n})$ for k in $[i-n\,..\,i]$ constitute a progressive pole system for f_i. By Theorem 4.2, there is a unique n-ic curve F_i that satisfies $\mathbf{a}_k = f_i(\bar{t}_{k+1},\ldots,\bar{t}_{k+n})$ for k in $[i-n\,..\,i]$.

Since the constraints in Equation 12.3 uniquely determine every segment F_i, we deduce that there can't be more than one satisfying spline F. To prove that there is one, we must show that the F_i determined by Equation 12.3 do fit together properly to form a spline. Using Theorem 12.1, fix any i and j with $i < j < i+n$, $\bar{t}_i < \bar{t}_{i+1}$, and $\bar{t}_j < \bar{t}_{j+1}$. We must show that the segments F_i and F_j agree on all superbags of $(\bar{t}_{i+1},\ldots,\bar{t}_j)$. From Equation 12.3, we know that both $f_i(\bar{t}_{k+1},\ldots,\bar{t}_{k+n})$ and $f_j(\bar{t}_{k+1},\ldots,\bar{t}_{k+n})$ are equal to $\mathbf{a}_k$ for k in $[j-n\,..\,i]$; hence, they are also equal to each other. That is to say, f_i and f_j agree on the bags formed by the rows of the following parallelogram:

$$\begin{array}{ccccccccccc}
\bar{t}_{j-n+1} & \bar{t}_{j-n+2} & \cdots & \bar{t}_i & \bar{t}_{i+1} & \cdots & \bar{t}_j & & & & \\
 & \bar{t}_{j-n+2} & \cdots & \bar{t}_i & \bar{t}_{i+1} & \cdots & \bar{t}_j & \bar{t}_{j+1} & & & \\
 & & \ddots & \vdots & \vdots & & \vdots & \vdots & \ddots & & \\
 & & & \bar{t}_i & \bar{t}_{i+1} & \cdots & \bar{t}_j & \bar{t}_{j+1} & \cdots & \bar{t}_{i+n-1} & \\
 & & & & \bar{t}_{i+1} & \cdots & \bar{t}_j & \bar{t}_{j+1} & \cdots & \bar{t}_{i+n-1} & \bar{t}_{i+n}
\end{array}$$

To finish the proof, we cut the middle $j-i$ columns out of this parallelogram and collapse the two remaining triangles to form a smaller parallelogram.

Let $l := n-j+i$ and let g_i denote the symmetric, multiaffine function of l arguments formed by fixing $(j-i)$ of the arguments of f_i at $\bar{t}_{i+1}$ through $\bar{t}_j$. That is, define g_i by $g_i(\bar{u}_1,\ldots,\bar{u}_l) := f_i(\bar{u}_1,\ldots,\bar{u}_l,\bar{t}_{i+1},\ldots,\bar{t}_j)$. Let g_j

denote the analogous restriction of f_j. Proving that F_i and F_j agree on all superbags of $(\bar{t}_{i+1}, \ldots, \bar{t}_j)$ is equivalent to showing that g_i and g_j are identical. Since f_i and f_j agree on the rows of the parallelogram above, we deduce that g_i and g_j agree on each sequence of l consecutive knots from the sequence $T := [\bar{t}_{j-n+1}, \ldots, \bar{t}_i, \bar{t}_{j+1}, \ldots, \bar{t}_{i+n}]$ of $2l$ knots. Furthermore, the sequence T is progressable, since each element of its left half is strictly less than each element of its right half. Thus, g_i and g_j have a progressive pole system in common, so they must be identical. □

The sequence $[\mathbf{a}_k]$ in Theorem 12.2 plays the same role for an entire spline curve that a pole system plays for a single curve; that is, the points $\mathbf{a}_k$ are handles with which we can control the shape of the spline. Since the sequence $[\mathbf{a}_k]$ includes a system of progressive poles for each segment F_i, we shall refer to the points $\mathbf{a}_k$ as the *progressive poles* of the spline F. The $\mathbf{a}_k$ are also called the *de Boor points* of F or the *coefficients in the B-spline expansion* of F.

With Theorem 12.2, we end our development of the theory of spline curves via the polarized approach. The reader must now judge whether or not the polarized approach is a good thing. The polarized labels certainly bring a new clarity to the de Casteljau and de Boor Algorithms. It is unfortunate that the associated proofs aren't simpler. The computational proof strategy, in Exercise 4.4, seems both tedious and a bit mysterious; it shows that $\mathbf{b}_{0,n}$ is symmetric without explaining why. The conceptual strategy is both elegant and enlightening, but it depends upon the tensor product construction, which is rather subtle, even in the symmetric case.

The reader might enjoy comparing the polarized approach to the last great revolution in spline theory, which was knot insertion algorithms. Some years ago, people realized that many different problems about splines could be solved by knot insertion. To a devotee of the polarized approach, this is hardly surprising: A knot insertion algorithm is just one particular, algorithmic way of computing polar values of the segments of a spline from that spline's progressive poles.

13 Polynomial and bipolynomial spline surfaces

Spline surfaces are significantly more subtle than spline curves.

Suppose that we want to model a surface parametrically by giving each coordinate of a point on the surface as a polynomial in two real parameters, say u and v. There are two different types of degree bounds that we can impose on the coordinate polynomials: Either we can bound the total degree in u and v jointly or we can bound the degrees in u and in v separately.

If we bound the total degree by n, the resulting map is properly thought of as a function $F: P \to Q$, where P is an affine parameter plane and Q is the object space. In particular, the degree bound is not affected by the choice of the Cartesian coordinate system $\langle u, v \rangle$ for P. The obvious name for such surfaces is *polynomial surfaces* of degree at most n. They are sometimes called "triangular-

patch surfaces"; but one can cut a patch of any shape out of any surface, at least in principle.

If we bound the degree in u by m and the degree in v by n separately, the resulting degree bounds do depend upon our choice of coordinate system in the parameter plane. Indeed, the resulting surface is best thought of as a function $G: U \times V \to Q$, where U and V are two different affine parameter lines. For each fixed u in U, the function $v \mapsto G(u; v)$ is an n-ic curve; for each fixed v in V, the function $u \mapsto G(u; v)$ is an m-ic curve. I like to call such a surface G a *bipolynomial surface* of degree $(m; n)$. The term "bipolynomial" generalizes the names "biquadratic" and "bicubic", which are the standard terms for the particular cases $m = n = 2$ and $m = n = 3$.

Note that every n-ic polynomial surface is also a bipolynomial surface of degree $(n; n)$ and that every $(m; n)$-ic bipolynomial surface is also a polynomial surface of degree $(m+n)$. Thus, the distinction between the two types of surfaces arises only when particular degree bounds are involved.

Exercise 13.1 Show that the polar form f of an n-ic polynomial surface F takes, as arguments, n points in the parameter plane P. On the other hand, show that the polar form g of an $(m; n)$-ic bipolynomial surface G takes m points in U and n points V as arguments.

Bipolynomial surfaces are often called "tensor-product surfaces," but that name is the result of incomplete understanding. Both types of surfaces can be converted into affine maps by means of the tensor product construction. An n-ic polynomial surface $F: P \to Q$ has the tensor form $f^{\odot}: \bigodot^n P \to Q$, while an $(m; n)$-ic bipolynomial surface $G: U \times V \to Q$ has the tensor form $g^{\otimes}: \bigodot^m U \otimes \bigodot^n V \to Q$. The only difference between the two cases is that tensoring F involves only the symmetric case of the tensor product construction, while tensoring G requires one instance of the asymmetric case as well as the symmetric case.

A bipolynomial spline surface is just a spline curve of spline curves. Polarization and tensoring handle such spline surfaces just as neatly as they handle spline curves. The only difficulty is a notational one: Since two different affine lines are now involved, we must keep track of which polar arguments come from U and which from V, as we multiply them together to form tensor arguments. One choice is to keep track positionally. In this scheme, the tensor $e = \bar{0}\,\bar{1} \otimes \bar{2}\,\bar{3}\,\bar{4}$ is the product of five points: $\bar{0}$ and $\bar{1}$ in U and $\bar{2}$, $\bar{3}$, and $\bar{4}$ in V. (The tensor e is a tensor argument for a bipolynomial surface of degree $(2; 3)$.) On the other hand, we can continue to think of the multiplication operation on points as commutative if we adopt some other convention to keep track of where the points come from. I like the notations $\grave{u}$, read "u-in," for the point $\langle u \rangle$ in U and $\acute{v}$, read "v-out," for the point $\langle v \rangle$ in V. With this convention, the tensor e could be written as the simple product $e = \grave{0}\,\grave{1}\,\acute{2}\,\acute{3}\,\acute{4}$, with its five factors in any order.

Open problem 13.2 The other major category of regular spline surface are the box spline surfaces. Ramshaw [23] discusses this case at length but with little success. Find a natural and helpful polar form for a patch of a box spline surface. In particular, your polar form must have the appropriate constraints on the higher-order derivatives of the surface built-in, in some fashion.

Open problem 13.3 Use polar forms in an enlightening way to study spline surfaces based on irregular partitions of the parameter plane.

14 B-splines and the work of Carl de Boor

In the last three sections of this report, we consider the three pioneers of the polarized approach in turn, presenting their perspectives briefly. Carl de Boor comes first, so it is time to talk about B-splines.

The traditional approach to the theory of spline curves places a great deal of weight on B-splines, although we have hardly mentioned them so far. Every knot sequence $[\bar{t}_k]$ has an associated sequence $[B_{j,n+1}(\bar{u})]$ of n-ic B-splines. (The second index is $n+1$, rather than n, because convention dictates that B-splines be indexed by order, rather than by degree.) The j^{th} member of this family is the real-valued spline function $B_{j,n+1}: L \to \mathbb{R}$ whose progressive poles $[\mathbf{b}_{j,k}]$ are given by

$$\mathbf{b}_{j,k} := \begin{cases} 1 & \text{if k=j} \\ 0 & \text{otherwise.} \end{cases}$$

If F is the n-ic spline curve on the knot sequence $[\bar{t}_k]$ with progressive poles $[\mathbf{a}_k]$, the B-spline $B_{j,n+1}$ quantifies the influence exerted by the progressive pole $\mathbf{a}_j$ on the value of F. That is, we can expand any spline as an affine combination of its progressive poles, with the B-splines as the coefficients:

$$F(\bar{u}) = \sum_j B_{j,n+1}(\bar{u})\,\mathbf{a}_j.$$

To say more or less the same thing in another way, we can express any real-valued spline function F as a linear combination of B-splines, where the coefficients are the progressive poles $[\mathbf{a}_j]$ of F. Thus, the B-splines form a basis for the space of all real-valued splines on the knot sequence $[\bar{t}_k]$; that is the way that a traditionalist would rephrase our Theorem 12.2.

The B-splines can be defined in three different ways. They were originally defined by a formula involving divided differences [6]. The other two definitions are both recurrences, and they are, in some sense, dual to each other.

One of the recurrences is called *the B-spline recurrence* [7]:

$$B_{j,1}(\bar{u}) = \begin{cases} 1 & \text{if } \bar{u} \in (\bar{t}_j \,..\, \bar{t}_{j+1}) \\ 0 & \text{otherwise} \end{cases}$$

$$B_{j,n+1}(\bar{u}) = \frac{t_{j+n+1} - u}{t_{j+n+1} - t_{j+1}} B_{j+1,n}(\bar{u}) + \frac{u - t_j}{t_{j+n} - t_j} B_{j,n}(\bar{u}).$$

It expresses a B-spline of degree n as a linear combination of B-splines of degree $n-1$; but note that the linear combination is not an affine combination.

The other recurrence is the one the lies at the heart of the de Boor Algorithm [8]. To compute $B_{j,n+1}(\bar{u})$, first find i so that $\bar{u}$ lies in $(\bar{t}_i \,..\, \bar{t}_{i+1})$, and let $B_{j,n+1,i}$ denote the n-ic polynomial that the B-spline $B_{j,n+1}(\bar{v})$ follows for $\bar{v}$ in $(\bar{t}_i \,..\, \bar{t}_{i+1})$. Adopting the abbreviation $\alpha_{k,l}(\bar{u})$ for the polar value

$$\alpha_{k,l}(\bar{u}) := b_{j,n+1,i}(\bar{t}_{k+1}, \ldots, \bar{t}_{k+n-l}, \underbrace{\bar{u}, \ldots, \bar{u}}_{l})$$

of $B_{j,n+1,i}$ whenever k lies in the interval $[i+l-n..i]$, we have $B_{j,n+1}(\bar{u}) = \alpha_{i,n}(\bar{u})$ where

$$\alpha_{k,0}(\bar{u}) = \begin{cases} 1 & \text{if } k = j \\ 0 & \text{otherwise} \end{cases}$$

$$\alpha_{k,l+1}(\bar{u}) = \frac{t_{k+n-l+1} - u}{t_{k+n-l+1} - t_k}\alpha_{k-1,l}(\bar{u}) + \frac{u - t_k}{t_{k+n-l+1} - t_k}\alpha_{k,l}(\bar{u}).$$

Note that the linear combinations in this recurrence are actually affine combinations.

One of the benefits of the polarized approach is that it lets us study spline theory without needing to talk about divided differences. Indeed, Carl de Boor and Klaus Höllig invented the polarized approach in order to convert from defining B-splines by means of the divided-difference formula to defining them by means of the B-spline recurrence [9, 10]. The *dual functionals* $\lambda_{i,n+1}$ that they define correspond precisely to polar evaluation. That is, if F is an n-ic spline, we have $\lambda_{i,n+1}(F) = f_j(\bar{t}_{i+1}, \ldots, \bar{t}_{i+n})$, where j is chosen so that the interval $(\bar{t}_j \,..\, \bar{t}_{j+1})$ contains their parameter τ. In particular, note that their dual functionals are dual only in the sense that any evaluation map is dual. For example, the map μ_3 defined by $\mu_3(F) := F(3)$ is a "dual functional" that performs diagonal evaluation.

Of the two recurrences for the B-splines, it isn't clear which is the best one to take as the definition and how to verify the other one most efficiently. Ramshaw [20] and deBoor [9, 10] each took a stab at this problem. More recently, Hans-Peter Seidel has made further progress [26].

15 The work of Ramshaw

There are two ideas in Ramshaw's monograph [20] that are worth mentioning here: overloading and the rational case.

Each segment of a spline curve F is a polynomial curve F_k and has a polar form f_k. It gets annoying, after a while, to specify the subscript k all the time. Note that the exact value of k often doesn't matter. We know from Theorem 12.1 that all of the polar forms f_k for k in $[i..j]$ agree on all superbags

of $(\bar{t}_{i+1}, \ldots, \bar{t}_j)$. Thus, any k in $[i \,..\, j]$ would work equally well in an expression of the form $f_k(\bar{t}_{i+1}, \ldots, \bar{t}_j, \bar{u}_1, \ldots, \bar{u}_{n-j+i})$.

This annoyance suggests that it might be worthwhile to establish a convention whereby an appropriate subscript k is chosen automatically, based on the bag of polar arguments. Note that the diagonal form F of the spline involves just such a convention: We interpret $F(\bar{u})$ to mean $F_k(\bar{u})$ where k is chosen so that the interval $(\bar{t}_k \,..\, \bar{t}_{k+1})$ contains $\bar{u}$. A convention of this kind is analogous to *overloading* in a programming language, where a single function symbol is used to denote one of several functions, depending upon the types of the supplied arguments. Thus, we can say that $F(\bar{u})$ is the *overloaded diagonal form* of the spline $[F_k]$. Analogously, we can invent an *overloaded polar form* $f(\bar{u}_1, \ldots, \bar{u}_n)$ for the spline $[F_k]$.

Overloading is more subtle in the polar case than in the diagonal case, because there are n polar arguments to consider. If those arguments are completely arbitrary, in fact, there is no reasonable convention for automatically choosing an appropriate k. But there are conventions that work well when the polar arguments are sufficiently close together. Ramshaw defines and studies two flavors of overloading, which he calls *tame* and *wild* [22].

Ramshaw also considers what the polarized approach can say about rational curves and surfaces, instead of polynomial ones [24]. A rational map is a function whose coordinates are given by the quotients of polynomials. It is best to rewrite those quotients, if necessary, so that all of the denominators are the same, that is, to put all of the coordinates over a common denominator. This common denominator plays the role of a generalized weight coordinate. Polynomial maps are the special case of rational maps in which this weight coordinate is identically 1. Roughly speaking, the polarized approach treats rational maps as follows: A rational map has a polar form that is multiprojective and a tensor form that is projective. There is a subtle point, however, about degenerate rational functions, that is, rational functions of degree m that are being considered as n-ic for some $n > m$; the polar forms of such maps are not uniquely determined.

16 The seminal work of Paul de Faget de Casteljau

Paul de Faget de Casteljau deserves the lion's share of the credit for inventing the polarized approach. He also realized from the start the connection with the mathematical concept of polar forms.

The first half of his book [11] is a polarized approach to the theory of spline curves. In comparison with this report, de Casteljau focuses more on patterns and techniques, less on definitions and properties. He represents his splines throughout using an overloaded polar form.

In the second half of his book, de Casteljau builds on the correspondence between a spline and its progressive poles to present a simple technique for deriving a wide class of spline methods that are elsewhere called *quasi-interpolants.*

Most of the curve methods used in practice fall into this class. Deriving quasi-interpolant methods with de Casteljau's technique is like shopping: You pick the degree n, the level of parametric continuity k, and the width—that is, the number w of input points upon which each segment of the spline will depend. From de Casteljau's technique, you get a rule that tells you where to put each of the plain poles of each spline segment. The spline method that you get will be optimal in a sense that involves the concept of reproductive power. A spline method has *reproductive power* r if, whenever the method is given input data that all comes from a single polynomial curve of degree at most r, the approximating spline that the method outputs is precisely that polynomial curve. The spline methods given by de Casteljau's technique have the highest level r of reproductive power possible, given their other parameters (n, k, w). Except for several classes of special cases, this best possible r is given by

$$r = w - \left\lceil \frac{n+1}{n-k} \right\rceil .$$

In the special cases, the $n + 1$ in the numerator is reduced to n.

Note on "reproduction": The term "reproduction" comes from the literature on quasi-interpolants. De Casteljau used the French word "restitution," which his translator rendered into English sometimes as "reconstruction" and sometimes as "restoration."

Open problem 16.1 One of the limitations on de Casteljau's technique is that it does not allow its user to demand at the outset that the resulting spline method be interpolatory. Some collections of parameters (n, k, w) do produce spline methods whose output curves interpolate their input data, but others do not. If we demand interpolation, it appears that the highest reproductive power possible is given by

$$r = w - \left\lceil \frac{n-1}{n-k-1} \right\rceil .$$

(For example, it is impossible to achieve C^{n-1} continuity for an interpolatory n-ic spline method with bounded width.) Find a simple technique, analogous to de Casteljau's, that generates interpolatory spline methods that are as reproductive as possible.

References

[1] Richard H. Bartels, John C. Beatty, and Brian A. Barsky. *An Introduction to Splines for use in Computer Graphics and Geometric Modeling.* Morgan Kaufmann Publishers, Inc., 95 First Street, Los Altos, California (1987).

[2] Maxime Bôcher. *Introduction to higher algebra*, pages 114–128. Macmillan (1907).

[3] Wolfgang Boehm. Smooth Curves and Surfaces. In Gerald E. Farin, editor, *Geometric Modeling: Algorithms and New Trends*, pages 175–184. SIAM (1987).

[4] Wolfgang Böhm, Gerald Farin, and Jürgen Kahmann. A survey of curve and surface methods in CAGD. *Computer Aided Geometric Design* **1**, 1–60 (1984).

[5] Carl de Boor. *A Practical Guide to Splines.* Springer-Verlag (1978).

[6] Page 108.

[7] Page 131.

[8] Page 147.

[9] Carl de Boor. B(asic)-Spline Basics. First portion of course notes for course #5, Extension of B-Spline Curve Algorithms to Surfaces, at ACM SIGGRAPH 1986.

[10] Carl de Boor and Klaus Höllig. B-Splines Without Divided Differences. In Gerald E. Farin, editor, *Geometric Modeling: Algorithms and New Trends*, pages 21–27. SIAM (1987).

[11] Paul de Faget de Casteljau. *Formes à pôles*, volume 2 of the series Mathématiques et CAO. Hermes Publishing, 51 rue Rennequin, 75017 Paris, France (1985). Also published in English translation as *Shape Mathematics and CAD* by Kogan Page Ltd., 120 Pentonville Road, London N1 9JN (1986).

[12] Page 59 in the French, page 50 in the English.

[13] C. T. J. Dodson and T. Poston. *Tensor Geometry: The Geometric Viewpoint and its Uses*, paperback edition, page 83. Pitman, London (1979).

[14] Serge Lang. *Algebra*, second edition. Addison-Wesley (1984).

[15] Page 520.

[16] Pages 586–588.

[17] E. T. Y. Lee. Some remarks concerning B-splines. *Computer Aided Geometric Design* **2**, pages 307–311 (1985).

[18] The Mathematical Society of Japan. *Encyclopedic Dictionary of Mathematics*, second edition, article 256 (Linear Spaces), sections H through O. MIT Press (1987).

[19] Lyle Ramshaw. Béziers and B-splines as multiaffine maps. In R. A. Earnshaw, editor, *Theoretical Foundations of Computer Graphics and CAD*, pages 757–776. NATO ASI Series F, volume 40, Springer-Verlag (1988).

[20] Lyle Ramshaw. *Blossoming: A Connect-the-Dots Approach to Splines.* Digital Equipment Corporation Systems Research Center research report #19 (1987).

[21] Equation (2.3), page 11.

[22] Section 16, pages 86, 87.

[23] Section 23, pages 131–139.

[24] Section 24, pages 140–146.

[25] George Salmon. *Higher Plane Curves*, third edition, pages 49–52. Chelsea (1879).

[26] Hans-Peter Seidel. A new multiaffine approach to B-splines. Accepted for publication in *Computer Aided Geometric Design.*

[27] A. Seidenberg. *Lectures in Projective Geometry*, pages 115–121 and 184–193. Van Nostrand (1962).

[28] Michael Spivak. *Calculus on Manifolds*, page 79 and Problem 4–1 on page 84. Benjamin/Cummings (1965).

[29] B. L. van der Waerden. *Algebra*, volume 2, translated from the German fifth edition, page 20. Frederick Ungar, New York (1970).

SRC Reports

"A Kernel Language for Modules and Abstract Data Types."
R. Burstall and B. Lampson.
Research Report 1, September 1, 1984.

"Optimal Point Location in a Monotone Subdivision."
Herbert Edelsbrunner, Leo J. Guibas, and Jorge Stolfi.
Research Report 2, October 25, 1984.

"On Extending Modula-2 for Building Large, Integrated Systems."
Paul Rovner, Roy Levin, John Wick.
Research Report 3, January 11, 1985.

"Eliminating **go to**'s while Preserving Program Structure."
Lyle Ramshaw.
Research Report 4, July 15, 1985.

"Larch in Five Easy Pieces."
J. V. Guttag, J. J. Horning, and J. M. Wing.
Research Report 5, July 24, 1985.

"A Caching File System for a Programmer's Workstation."
Michael D. Schroeder, David K. Gifford, and Roger M. Needham.
Research Report 6, October 19, 1985.

"A Fast Mutual Exclusion Algorithm."
Leslie Lamport.
Research Report 7, November 14, 1985.

"On Interprocess Communication."
Leslie Lamport.
Research Report 8, December 25, 1985.

"Topologically Sweeping an Arrangement."
Herbert Edelsbrunner and Leonidas J. Guibas.
Research Report 9, April 1, 1986.

"A Polymorphic λ-calculus with Type:Type."
Luca Cardelli.
Research Report 10, May 1, 1986.

"Control Predicates Are Better Than Dummy Variables For Reasoning About Program Control."
Leslie Lamport.
Research Report 11, May 5, 1986.

"Fractional Cascading."
Bernard Chazelle and Leonidas J. Guibas.
Research Report 12, June 23, 1986.

"Retiming Synchronous Circuitry."
Charles E. Leiserson and James B. Saxe.
Research Report 13, August 20, 1986.

"An $O(n^2)$ Shortest Path Algorithm for a Non-Rotating Convex Body."
John Hershberger and Leonidas J. Guibas.
Research Report 14, November 27, 1986.

"A Simple Approach to Specifying Concurrent Systems."
Leslie Lamport.
Research Report 15, December 25, 1986. Revised January 26, 1988

"A Generalization of Dijkstra's Calculus."
Greg Nelson.
Research Report 16, April 2, 1987.

"*win* and *sin*: Predicate Transformers for Concurrency."
Leslie Lamport.
Research Report 17, May 1, 1987.

"Synchronizing Time Servers."
Leslie Lamport.
Research Report 18, June 1, 1987.

"Blossoming: A Connect-the-Dots Approach to Splines."
Lyle Ramshaw.
Research Report 19, June 21, 1987.

"Synchronization Primitives for a Multiprocessor: A Formal Specification."
A. D. Birrell, J. V. Guttag, J. J. Horning, R. Levin.
Research Report 20, August 20, 1987.

"Evolving the UNIX System Interface to Support Multithreaded Programs."
Paul R. McJones and Garret F. Swart.
Research Report 21, September 28, 1987.

"Building User Interfaces by Direct Manipulation."
Luca Cardelli.
Research Report 22, October 2, 1987.

"Firefly: A Multiprocessor Workstation."
C. P. Thacker, L. C. Stewart, and E. H. Satterthwaite, Jr.
Research Report 23, December 30, 1987.

"A Simple and Efficient Implementation for Small Databases."
Andrew D. Birrell, Michael B. Jones, and Edward P. Wobber.
Research Report 24, January 30, 1988.

"Real-time Concurrent Collection on Stock Multiprocessors."
John R. Ellis, Kai Li, and Andrew W. Appel.
Research Report 25, February 14, 1988.

"Parallel Compilation on a Tightly Coupled Multiprocessor."
Mark Thierry Vandevoorde.
Research Report 26, March 1, 1988.

"Concurrent Reading and Writing of Clocks."
Leslie Lamport.
Research Report 27, April 1, 1988.

"A Theorem on Atomicity in Distributed Algorithms."
Leslie Lamport.
Research Report 28, May 1, 1988.

"The Existence of Refinement Mappings."
Martín Abadi and Leslie Lamport.
Research Report 29, August 14, 1988.

"The Power of Temporal Proofs."
Martín Abadi.
Research Report 30, August 15, 1988.

"Modula-3 Report."
Luca Cardelli, James Donahue, Lucille Glassman, Mick Jordan, Bill Kalsow, Greg Nelson.
Research Report 31, August 25, 1988.

"Bounds on the Cover Time."
Andrei Broder and Anna Karlin.
Research Report 32, October 15, 1988.

"A Two-view Document Editor with User-definable Document Structure."
Kenneth Brooks.
Research Report 33, November 1, 1988.

by Lyle Ramshaw

Systems Research Center
130 Lytton Avenue
Palo Alto, California 94301