THIELE CENTRE
FOR APPLIED MATHEMATICS IN NATURAL SCIENCE

Shape Discrimination by Total Curvature with a view to Cancer Diagnostics

R. J. Gardner, A. Hobolth, E. B. V. Jensen
and F. B. Sørensen

Research Report No. 02 I January 2004

Shape Discrimination by Total Curvature with a view to Cancer Diagnostics

This Thiele Research Report is also Research Report number 440 in the Stochastics Series at Department of Mathematical Sciences, University of Aarhus, Denmark.

SHAPE DISCRIMINATION BY TOTAL CURVATURE, WITH A VIEW TO CANCER DIAGNOSTICS

R. J. GARDNER, A. HOBOLTH, E. B. V. JENSEN, AND F. B. SØRENSEN

Abstract. This paper studies the use of total curvature for shape discrimination of objects via profiles of their planar sections (not assumed star shaped). Methods of estimating total curvature from observation of a finite number of points on the boundary of the object are investigated, including a simple discrete approximation method and various interpolation methods. Total curvature is capable of revealing shape differences on a local scale, as demonstrated by the analysis of two data sets of malignant and normal or benign tumor cell nuclear profiles.

1. Introduction

Diagnostic examination of tissue samples from neoplasms is performed every day by pathologists. A number of morphological features are subjectively evaluated in the diagnostic discrimination between benign and malignant cell proliferations. On the macroscopic scale, malignant neoplasms (cancers) appear less ordered than benign and normal tissue from which the malignant proliferations may have derived. On the microscopic scale, the nuclear chromatin texture, the nuclear size and variability of nuclear size are recognized cytological parameters that are evaluated in the diagnosis of malignancy. Furthermore, the architectural pattern of the tumor histology is used in the process of malignancy grading. Stereological methods of estimating some of these parameters have been developed in [13, 14, 15, 22]. The shape of cell nuclei also appears to be an important parameter for some types of cancers. Often, the boundary of cell nuclei from cancer tissue are more irregular on a local scale; see [4, 9, 12].

In the shape literature, various types of shape descriptors of planar objects have been studied; see [1] and [21, Part II] for a review. The most familiar is the geometric shape ratio $4\pi A/B^2$, where A is the area of the object and B its boundary length. Most of these shape descriptors are not directed specifically to reveal shape differences on a local scale. For instance, cell nuclei with a locally irregular boundary may have a small value of $4\pi A/B^2$ but such a small value can also be obtained for a very elongated, but locally smooth object.

In [12] (see also [10, 11]), parametric shape modelling is studied in relation to cancer diagnostics. Using the suggested p-order model, global and local shape differences can be detected. However, the model described in [12] only applies to star-shaped objects. For certain types of cancer this is a serious limitation because the boundaries of cancer cell nuclei can be very irregular and certainly not in general star shaped.

1991 *Mathematics Subject Classification.* Primary: 53A04, 92C50; secondary: 52A40, 62P10.

Key words and phrases. Local stereology, total curvature, cancer diagnostics.

First author supported in part by U.S. National Science Foundation Grant DMS-0203527. Third author supported in part by MaPhySto, funded by a grant from the Danish National Research Foundation.

In the present paper we study the use of total curvature as a means of shape discrimination, with particular reference to cancer diagnostics. Total curvature is, as a shape discriminator, surely local and can be defined for non-star-shaped objects. Furthermore, for convex objects, the total curvature is always equal to 2π. Total curvature can therefore also be regarded as a measure of convexity or non-convexity.

In Section 2, we define total curvature and provide various formulas for calculating it. In Section 3, estimation of total curvature from discrete data is discussed, and in Section 4 indicators of convexity are considered. Examples of the use of total curvature in cancer diagnostics are given in Sections 5 and 6. Clinical consequences of our work are outlined in Section 7.

2. Total curvature and Fenchel's theorem

Let P be a closed polygon in $\mathbb{R}^n$, that is, a polygonal arc with vertices $p_0, \dots, p_{m-1}, p_m$, where $p_m = p_0$ and $p_i \neq p_{i+1}$ for $i = 0, \dots, m-1$. Let ψ_i be the exterior angle at p_i, that is, the angle between the prolongation of the edge $[p_{i-1}, p_i]$ and the edge $[p_i, p_{i+1}]$, or, alternatively, between the vectors $p_i - p_{i-1}$ and $p_{i+1} - p_i$. (See Figure 1; throughout the paper indices will be taken modulo m.) Then the *total curvature* $M(P)$ of P is defined to be

$$M(P) = \sum_{i=0}^{m-1} \psi_i.$$

Let C be any closed curve in $\mathbb{R}^n$. The *total curvature* $M(C)$ of C is defined by

$$M(C) = \sup\{M(P) : P \text{ is a polygon inscribed in } C\}. \tag{1}$$

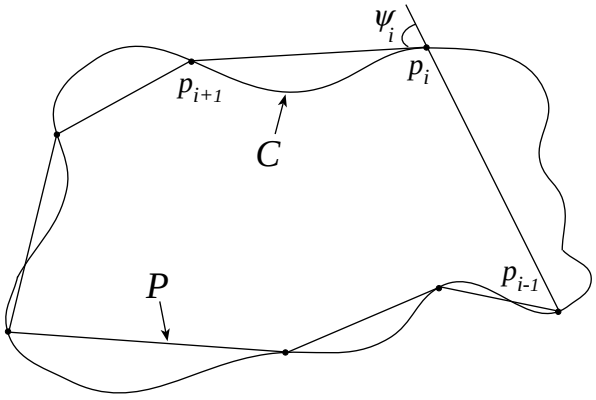

Figure 1. A polygon P inscribed in a curve C.

There is a convenient alternative definition when the curve C is C^2; if $\kappa(s)$ is the signed curvature of C at the point where the arc length, measured from some base point in some preferred direction around C, is s, then

$$M(C) = \int_C |\kappa(s)|\, ds. \tag{2}$$

More generally, if C is piecewise C^2 on arcs A_i, $i = 0, \ldots, m-1$, whose union is C, then

$$M(C) = \sum_{i=0}^{m-1} \int_{A_i} |\kappa(s)|\, ds + \sum_{i=0}^{m-1} \psi_i, \tag{3}$$

where ψ_i is the exterior angle at the point p_i where A_{i-1} and A_i meet. The definition of ψ_i is the intuitive one, namely, the angle between the appropriately directed tangent vectors to A_{i-1} and A_i at p_i.

The above definitions are well known, but the concept of total curvature is surprisingly elusive in the literature (at least in book form). Our definition (1) coincides with that utilized by Milnor [18], who attributes it to the mathematician R. H. Fox. The alternative definition (2) is more widely used, but of course it applies only to C^2 curves. The fact that (1), (2), and (3) are consistent is proved by Milnor [18, Section 2].

Proposition 2.1. *Let C be a closed curve in $\mathbb{R}^n$. Then*

$$M(C) \geq 2\pi,$$

with equality if and only if C is the boundary of a two-dimensional compact convex set.

In this paper we shall only employ Proposition 2.1 when $n = 2$ and refer to the result as *Fenchel's theorem*. In fact, Fenchel proved the theorem for C^2 curves and $n \leq 3$, and the extension to C^2 curves and $n \geq 2$ was carried out by K. Borsuk. The general result stated above is established by Milnor [18, Theorem 3.4], where references can also be found. Far-reaching extensions to manifolds are stated in [2, Section 29].

A possible point of confusion arises because some authors, for example Santaló [20, p. 113], define total curvature as in (2) (or, more generally as in (3)) but with the absolute value signs removed. In this case, when C is a simple closed curve in $\mathbb{R}^2$, the expression is always equal to $\pm 2\pi$, the sign depending on the orientation of the curve; this is a special case of the Gauss-Bonnet theorem (see, for example, [7]).

Some authors refer to the *theorem of turning tangents* when discussing such results. This is because

$$M(C) = \int_C |\kappa(s)|\, ds = \int_C \|T'(s)\|\, ds, \tag{4}$$

where $T(s)$ denotes the unit tangent vector to C at s. It follows that $M(C)$ measures how much the tangent turns as a point moves around C, or, in other words, how crooked the boundary of C is.

For the sequel it will be convenient to have some more explicit formulas in hand. If P is the closed polygon defined above, we clearly have

$$M(P) = \sum_{i=0}^{m-1} \arccos\left(\frac{(p_i - p_{i-1}) \cdot (p_{i+1} - p_i)}{\|p_i - p_{i-1}\|\, \|p_{i+1} - p_i\|} \right). \tag{5}$$

Let A be an arc in $\mathbb{R}^2$, parameterized by a parameter t, so that $A = A(t) = (f(t), g(t))$ for $0 \leq t \leq 1$. Then arc length is given by

$$s = \int_0^s \left(f'(t)^2 + g'(t)^2 \right)^{1/2} dt,$$

so we have

$$\frac{ds}{dt} = \left(f'(t)^2 + g'(t)^2\right)^{1/2}.$$

The formula for the curvature $\kappa(t)$ is (see, for example, [8, p. 11])

$$\kappa(t) = \frac{f'(t)g''(t) - g'(t)f''(t)}{(f'(t)^2 + g'(t)^2)^{3/2}},$$

and it follows that

$$\int_A |\kappa(s)|\, ds = \int_0^1 \left| \frac{f'(t)g''(t) - g'(t)f''(t)}{f'(t)^2 + g'(t)^2} \right| dt.$$

The total curvature of a piecewise C^2 curve may be calculated from this and (3) by the formula

$$M(C) = \sum_{i=0}^{m-1} \int_0^1 \left| \frac{f_i'(t)g_i''(t) - g_i'(t)f_i''(t)}{f_i'(t)^2 + g_i'(t)^2} \right| dt + \sum_{i=0}^{m-1} \psi_i, \tag{6}$$

where $A_i = A_i(t) = (f_i(t), g_i(t))$ for $0 \le t \le 1$ and

$$\psi_i = \arccos\left(\frac{f_i'(0)f_{i-1}'(1) + g_i'(0)g_{i-1}'(1)}{\left[(f_i'(0)^2 + g_i'(0)^2)\left(f_{i-1}'(1)^2 + g_{i-1}'(1)^2\right)\right]^{1/2}} \right),$$

for $i = 0, \dots, m-1$.

It may happen that the curve C is a simple closed curve containing the origin in its interior that bounds a region star shaped with respect to the origin. In this case we can describe C as a function $r(\theta)$, $0 \le \theta \le 2\pi$, in polar coordinates. When C is C^2, we have

$$M(C) = \int_0^{2\pi} |\gamma'(\theta)|\, d\theta = \int_0^{2\pi} \left| \frac{r^2 + 2(r')^2 - rr''}{r^2 + (r')^2} \right| d\theta, \tag{7}$$

where the previous integrand is the absolute value of the curvature of C in polar coordinates and $0 \le \gamma(\theta) < 2\pi$ is the angle between the tangent line to C at the point (r, θ), directed in the counterclockwise sense, and the positive x-axis.

3. Total curvature from discrete data

Suppose that the closed curve C in $\mathbb{R}^2$ is only accessible via a finite set of points $p_0, \dots, p_{m-1}$ in C, where it is assumed that these points occur in order of increasing arc length around C measured from some fixed base point. How can $M(C)$ be estimated?

Perhaps the simplest method is to define

$$M_D(C) = M(P), \tag{8}$$

where P is the polygon with vertices $p_0, \dots, p_{m-1}$ and $M(P)$ is given by (5). Then $M_D(C)$ is an estimate of $M(C)$.

In practice, however, the fact that total curvature is not continuous in the Hausdorff metric may sometimes cause extra difficulties. If D and E are closed curves in $\mathbb{R}^2$, we can estimate how close they are by the *Hausdorff distance*

$$\delta_H(D', E') = \max\{\max_{x \in D'} d(x, E'), \max_{x \in E'} d(x, D')\}$$

between the compact sets D' and E' they enclose, where $d(x, M)$ denotes the distance between the point x and the set M. The "cookie cutter" on the right in Figure 2 is close in this sense to

the circle on the left, but has much larger total curvature. Suitable polygonal approximations to these curves would clearly exhibit the same phenomenon. We will address this problem in Section 5.

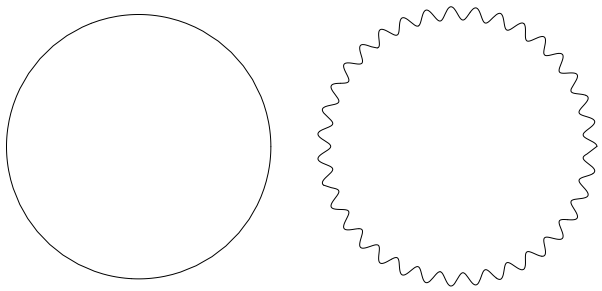

FIGURE 2. Closed curves with very different total curvatures.

Another natural method of estimating $M(C)$ is to generate a suitable simple closed curve C_m that interpolates the points $p_0, \ldots, p_{m-1}, p_0$, and then calculate the total curvature of C_m. Suppose that C_m is piecewise C^2, the union of C^2 arcs A_i with endpoints p_i and p_{i+1}, $i = 0, \ldots, m-1$. Then we can estimate $M(C)$ by $M_A(C)$, where

$$M_A(C) = M(C_m), \tag{9}$$

and $M(C_m)$ is given by (3).

Cubic spline and Fourier interpolation are discussed in the Appendix. Figure 3 illustrates these two interpolation procedures. On the left is a closed polygon with 30 vertices (actually one of the cell nuclear profiles in the data set shown in Figure 4). The center image depicts the cubic spline interpolation, and the right-hand image shows the Fourier interpolation. We let $M_{AS}(C)$ and $M_{AF}(C)$ be the estimates of $M(C)$ based on cubic spline and Fourier interpolation, respectively.

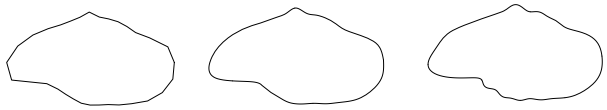

FIGURE 3. Approximating polygon, cubic spline, and Fourier interpolation.

Yet another approach to estimate $M(C)$ is to compute the natural discrete approximation to the integrals in the various expressions for $M(C)$. We illustrate in the situation where C is given as a function $r(\theta)$ in polar coordinates as above, and the given points on C are of the form $p_i = (r(\theta_i), \theta_i)$, $i = 0, \ldots, m-1$, where

$$0 \le \theta_0 < \cdots < \theta_{m-1} < 2\pi$$

holds and moreover $\theta_{i+1} - \theta_i = \Delta$ for $i = 0, \ldots, m-1$. (The assumption that the angles are equally spaced can be avoided and is merely a matter of convenience.) We have

$$r'(\theta_i) \approx \frac{r(\theta_{i+1}) - r(\theta_i)}{\Delta}$$

and

$$r''(\theta_i) \approx \frac{r(\theta_{i+2}) - 2r(\theta_{i+1}) + r(\theta_i)}{\Delta^2},$$

for $i = 0, \ldots, m-1$. Substituting these quantities into the discrete form of (7), we obtain

$$M_I(C) = \Delta \sum_{i=0}^{m-1} \left| 1 + \frac{r(\theta_{i+1})^2 - r(\theta_i) r(\theta_{i+2})}{r(\theta_i)^2 \Delta^2 + \left(r(\theta_{i+1}) - r(\theta_i)\right)^2} \right|. \tag{10}$$

Then $M_I(C)$ can be used to estimate $M(C)$.

4. Indicators of convexity

The convexity of a set E in $\mathbb{R}^n$ can be tested by examining its two-dimensional sections through a fixed point. It is easy to see, for example, that E is convex if and only if $E \cap S$ is convex for each two-dimensional plane containing the origin. In this way, the problem is essentially reduced to its two-dimensional version. (As far as we know, the related problem of estimating the total curvature of a set E in $\mathbb{R}^n$ from measurements of sections $E \cap S$ is open.)

For a closed curve C in $\mathbb{R}^2$, Proposition 2.1 shows that

$$K(C) = M(C) - 2\pi \geq 0,$$

with equality if and only if C is the boundary of a convex body. Each of $M(C)$ and $K(C)$ is therefore a *measure of convexity* in the sense of [16, Section 5], that is, a real-valued function f defined on a collection $\mathcal{E}$ of sets such that for some fixed value α, $f(E) = \alpha$ for a set $E \in \mathcal{E}$ if and only if E is convex. (In our case we take $\mathcal{E}$ to be the collection of all sets E whose boundary is a closed curve.) Preference is given here to $K(C)$ merely because it is easier to assess when given in numerical form.

Suppose that each closed curve C is only accessible via a finite set of points p_i, $i = 0, \ldots, m-1$ on its boundary. Define $K_D(C) = M_D(C) - 2\pi$, $K_{AS}(C) = M_{AS}(C) - 2\pi$, $K_{AF}(C) = M_{AF}(C) - 2\pi$, and $K_I(C) = M_I(C) - 2\pi$.

Clearly

$$K_D(C) = M(P) - 2\pi \geq 0,$$

with equality if and only if P, the closed polygon with vertices p_i, $i = 0, \ldots, m-1$, is the boundary of a convex polygon. Similarly,

$$K_{AS}(C) = M(C_m) - 2\pi \geq 0,$$

with equality if and only if the interpolating curve C_m is the boundary of a convex body. As noted in the Appendix, problems may arise in the case of Fourier interpolation if the interpolating curve C_m contains the origin; but if this is not the case, we again have

$$K_{AF}(C) = M(C_m) - 2\pi \geq 0,$$

with equality if and only if C_m is the boundary of a convex body.

Thus each of the above three quantities, while not satisfying the definition of a measure of convexity, can be regarded as an *indicator of convexity*, a term that we deliberately do not define rigorously.

The quantity $K_I(C)$ is more problematic. If the points $p_i = (r(\theta_i), \theta_i)$, $i = 0, \ldots, m-1$, lie on a circle centered at the origin, then $K_I(C) = 0$. However, it is not generally true that $K_I(C) \geq 0$; for example, when $m = 4$ and $r(\theta_i) = 1$, 2, 6, and 28 for $\theta_i = \pi i/2$, and $i = 0$, 1,

2, and 3, respectively, we have $K_I(C) = -0.0194$. Moreover, when $m = 4$ and $r(\theta_i) = 1$, 2, 6, and 27 for $\theta_i = \pi i/2$, and $i = 0$, 1, 2, and 3, respectively, we have $K_I(C) = 0.1108$, so by continuity there are non-circular curves C such that $K_I(C) = 0$. It is also easy to find curves C that bound convex bodies for which $K_I(C) > 0$. Thus while experimental evidence indicates that it is rare for $K_I(C)$ to be negative, and while we expect $K_I(C)$ to be generally larger when C bounds a non-convex region, it is not such a straightforward indicator of convexity as the three considered above.

5. Experimental data

The indicators of convexity in the previous section were tested on two sets of experimental data. In the first set, measurements $r(\theta_i)$, where $\theta_i = 2\pi i/30$, $i = 0, \ldots, 29$, were taken of profiles of two samples of star-shaped cell nuclei. Here a *profile* means a two-dimensional section of the cell nucleus, and the measurement $r(\theta_i)$ gives the distance from a fixed point inside the nucleus to its boundary in the direction θ_i. Each sample consisted of 27 cell nuclear profiles from a malignant and a benign tumor of the human skin - a malignant melanoma and its benign counterpart, a benign melanocytic nevus. The profiles were observed on 4 μm thick histological sections of plastic embedded tissue biopsies. The histopathological discrimination between these two types of tumors represents an often-encountered problem in histopathological routine diagnostic practice. The profiles were observed using light microscopy. The measurements are normalized so that the average value for each nucleus is 1.

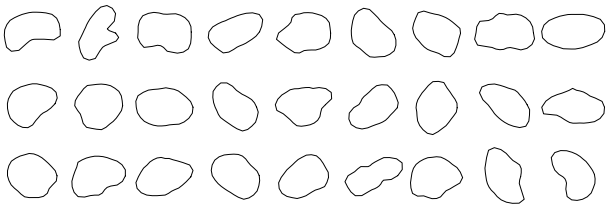

Figure 4. Data Set 1: Tumor cell nuclear profiles from a malignant melanoma of the human skin.

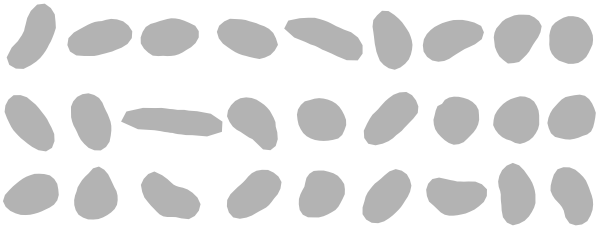

Figure 5. Data Set 1: Tumor cell nuclear profiles from a benign melanocytic nevus of the human skin.

The 30 data points for each nuclear profile form the vertices of a closed polygon that can be used to approximate the profile. Figures 4 and 5 show these polygonal approximations for the two samples. These profiles have also been analyzed in [9, 13]. We assign to each nuclear profile in each sample a number j, $j = 1, \ldots, 27$, in the natural way. For example, the sixth nuclear profile in the second row in each sample shown in Figures 4 and 5 is assigned the number 15.

A second set of data is shown below in Figures 6 and 7. Both samples consist of 50 cell nuclei, observed on 4 μm thick histological sections of plastic embedded tissue biopsies. The first sample consists of nuclear profiles in a malignant T-cell lymphoma, and the second sample represents nuclear profiles of normal T-lymphocytes obtained from the interfollicular T-zone of a hyperplastic lymph node. The profiles were again observed by light microscopy. A search algorithm was used to identify approximately 250 (the precise number varies from profile to profile) pixels on the boundary, each identified by a point with integer coordinates. The approximately 250-sided polygons with these points as vertices are shown in Figures 6 and 7. Numbers are assigned to each nuclear profile in each sample first by row and then by column in the way described above. As can be seen from malignant tumor cell nuclear profile numbers 7, 12, and 47, the profiles are generally not star shaped.

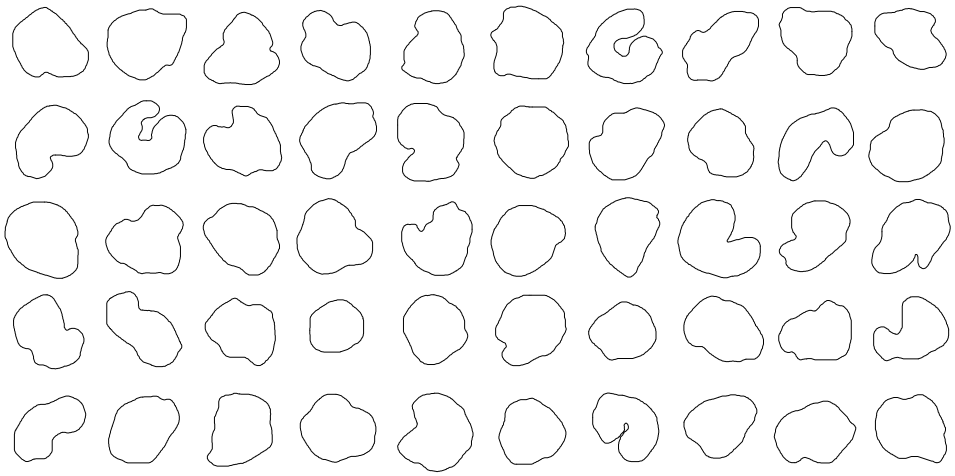

FIGURE 6. Data Set 2: Cell nuclear profiles from a malignant T-cell lymphoma.

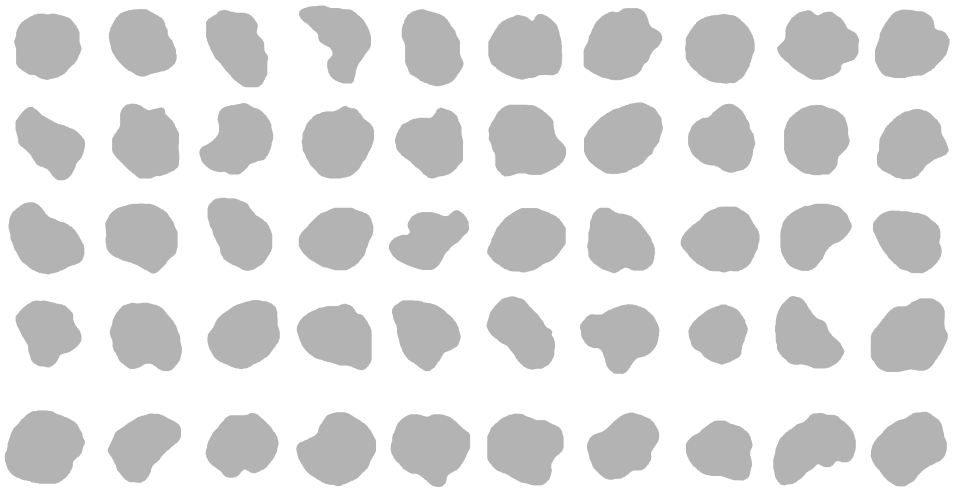

FIGURE 7. Data set 2: Cell nuclear profiles from normal T-lymphocytes.

6. Results

Beginning with the first data set, Figure 8 shows the distribution of each indicator of convexity for the malignant and benign cell nuclear profiles. We tested for each indicator of convexity the hypothesis that the mean value for the benign tumor cell nuclear profiles is equal to the mean value for the malignant tumor cell nuclear profiles, against the alternative hypothesis that the former mean is less than the latter. This was done with an ordinary t-test, applied to the log-transformed data which could be regarded as normally distributed. The resulting p-values for the indicators $K_D(C)$, $K_{AS}(C)$, $K_{AF}(C)$, and $K_I(C)$ were all below 10^{-3}. Based on this test alone, one could claim that each of the four indicators successfully distinguishes between the malignant and benign tumor cell nuclear profiles.

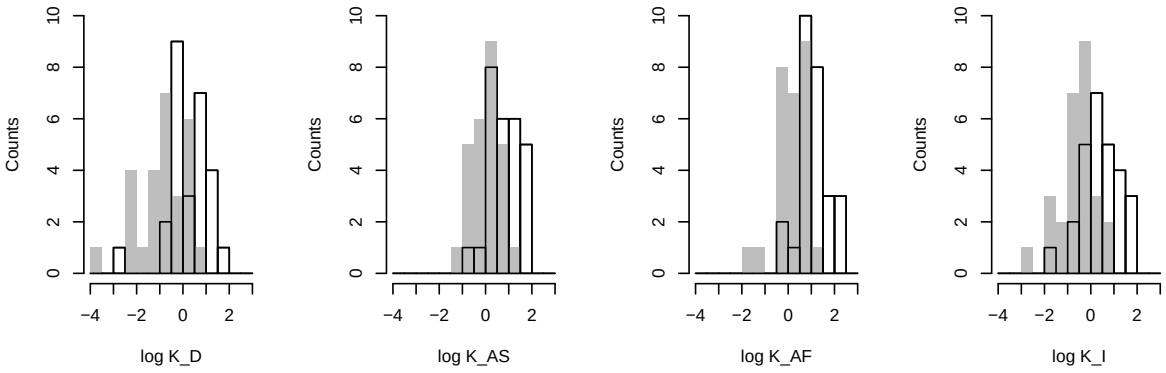

Figure 8. Distributions for malignant (white) and benign (grey) tumor cell nuclear profiles from Data Set 1.

If each indicator is truly estimating convexity, each pair of indicators should be strongly correlated for both the malignant and benign tumor cell nuclear profiles. In Figure 9, the value of each indicator is plotted against the corresponding value of any other indicator. In each plot, the correlation coefficients for the malignant and benign tumor cell nuclear profiles, respectively, are also indicated. The results show the expected strong (and statistically significant) correlation between the indicators.

In the case of the benign tumor cell nuclear profiles, the correlation coefficients involving the indicator $K_{AS}(C)$ are somewhat smaller. This is probably a consequence of the fact that the spline fit we employ yields a curve that is generally not C^2 at one point (see the Appendix for details), resulting in a relatively higher value especially for profiles that are nearly convex. In the case of the malignant tumor cell nuclear profiles, the correlation coefficients involving the indicator $K_{AF}(C)$ are somewhat smaller. This is probably due to the fact that the Fourier interpolation does not preserve convexity. Often the interpolating curve will tend to oscillate between positive and negative curvature more than the original; malignant tumor cell nuclear profile 18 is an extreme example, resulting in an $K_{AF}(C)$ value much larger than those of the other indicators, see also Figure 3. The cubic spline interpolation generally seems better at preserving convexity. The benign tumor cell nuclear profile 12 is an outlier, though it seems to be close to convex in Figure 5. Figure 10 shows that the boundary meanders along the two

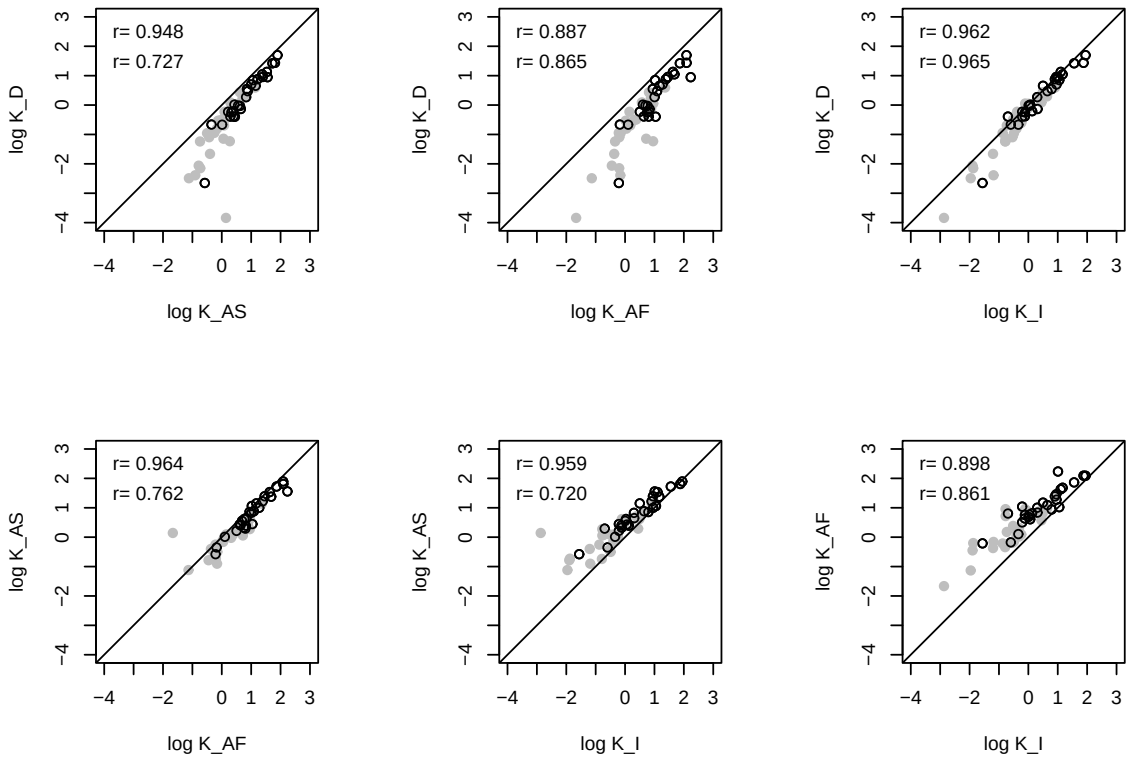

FIGURE 9. Bivariate distributions and correlation coefficients for pairs of indicators and malignant (white) and benign (grey) tumor cell nuclear profiles from Data Set 1.

nearly straight sides of this long, thin nuclear profile, which causes the indicators to increase as they accumulate contributions from the parts of the boundary with positive or negative curvature.

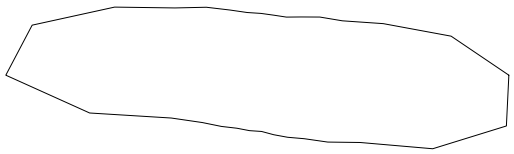

FIGURE 10. Approximating polygon for benign tumor cell nuclear profile 12.

This latter phenomenon influenced our different treatment of the second data set. We focused solely on the indicator $K_D(C)$, since our results for the first data set show that it should serve as well as any of the others. Tables 1 and 2 show quantities labelled $K_D(C)[n]$, where $n = 30$, 40, 50, 60, and $n = \text{all}$. Here "$n = \text{all}$" means we are computing $M(P)$ from (5) using all 250 or so vertices of the polygons shown in Figures 6 and 7. For other values of n, we select n vertices of each polygon with indices that are "equally spaced" and compute $M(P)$ from (5) using only these vertices. Specifically, if $p_1, \ldots, p_m$ is the set of all vertices of

Nuclear section	$K_D(C)[30]$	$K_D(C)[40]$	$K_D(C)[50]$	$K_D(C)[60]$	$K_D(C)[\text{all}]$
1	2.0590	2.9551	3.2456	3.3380	47.9075
2	1.8144	3.4742	4.1426	4.6028	38.0010
3	4.2602	5.7165	6.9114	7.6445	46.0629
4	4.6298	4.6922	5.8948	6.0484	45.8929
5	2.7515	4.2656	5.4225	7.2089	48.6369
6	3.5551	5.0903	6.4781	6.7929	42.1838
7	8.6546	9.7279	10.2741	12.1720	62.6602
8	3.7211	5.2781	6.3524	6.7588	51.5887
9	2.7754	3.3878	4.9737	6.3190	35.3330
10	4.8233	5.7369	6.1232	6.8854	41.2185
mean (all 50)	3.4668	4.3700	5.1903	6.0039	45.1406

TABLE 1. Data Set 2: An indicator of convexity for nuclear profiles from a malignant T-cell lymphoma.

a given polygon, we select those with indices

$$1 + \left\langle \frac{i(m-1)}{n} \right\rangle,$$

for $i = 0, 1, \ldots, n-1$, where $\langle x \rangle$ denotes the integer nearest to x (rounding up in the case of a half integer). Due to the size of the data sets, we give only the values for the first 10 nuclear profiles in each sample, but the mean values listed are taken over all 50 nuclear profiles. Initially, we computed for each profile only $K_D(C)[\text{all}] = M(P[\text{all}]) - 2\pi$, where $P[\text{all}]$ is the corresponding polygon with all 250 or so vertices shown in Figures 6 and 7, and $M(P[\text{all}])$ is calculated by (5) with $P = P[\text{all}]$. The relatively large numbers in the last columns in Tables 1 and 2 indicate that there are very many small oscillations in the boundaries of these polygons. For example, $K_D(C)[\text{all}] = 45.2115$ for normal T-lymphocyte nuclear profile number 8, which appears to be nearly convex in Figure 7. Thus the "noise" produced by these oscillations tends to obscure the true differences in total curvatures of the boundaries of the underlying sections. For this reason, we pruned the numbers of vertices of the defining polygons down to 30, 40, 50, or 60 and recalculated the indicator, with the results shown in the first four columns of Tables 1 and 2. Note that the small values for normal T-lymphocyte nuclear profile number 8 in these columns reflect much more accurately the true total curvature.

To determine an appropriate number of vertices to select, we adopted the following procedure based on artificial *test profiles*. These are curves defined by equations from which the exact total curvature can be calculated. These can be processed by the pixel-searching algorithm mentioned in Section 5 to produce approximating polygons with about 250 vertices. For various values, n vertices can be selected and the corresponding values of $K_D(C)[n]$ computed and compared to each other and the true value. Examples of the processed test sections used are illustrated in Figure 11. On the left is an ellipse C_1 and on the right a curve C_2 with polar coordinate equation $r = 1 + (\cos 4t)/5$, $0 \le t \le 2\pi$. Of course, the selection of test sections must be done appropriately, to reflect the general shapes of the nuclear profiles under consideration.

Nuclear section	$K_D(C)[30]$	$K_D(C)[40]$	$K_D(C)[50]$	$K_D(C)[60]$	$K_D(C)$[all]
1	1.8843	2.4189	3.5436	4.5881	36.7899
2	0.7177	1.6724	1.7865	3.9492	41.1426
3	2.6482	3.7745	4.1270	5.3468	41.3884
4	4.9628	6.2477	7.4093	8.1417	44.6059
5	1.5126	2.1565	3.2621	3.3866	39.2501
6	2.6179	4.1476	4.5150	5.1322	44.0384
7	4.0745	4.0371	4.8487	6.6306	45.7032
8	0.0786	0.4571	1.4246	2.3964	45.2115
9	5.3890	6.8319	7.6325	8.1063	56.4050
10	1.8296	3.2040	3.6161	4.1344	47.1799
mean (all 50)	2.1450	2.9764	3.7094	4.6439	43.0405

TABLE 2. Data Set 2: An indicator of convexity for nuclear profiles of T-lymphocytes from a hyperplastic lymph node.

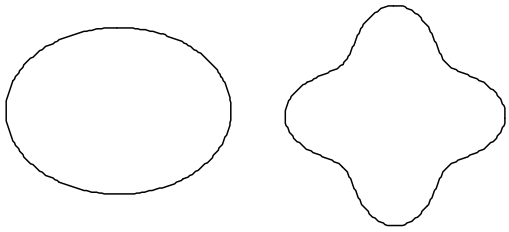

FIGURE 11. Processed test profiles.

We have $M(C_1) = 2\pi$ and by direct computation using (7), $M(C_2) = 12.6405\cdots$. The quantities $M_D(C_1)[n] - 2\pi$ and $M_D(C_2)[n] - 12.6405$ were computed for $n = 5, 10, \ldots, 120$ and are displayed in Figure 12 on the left and right, respectively. As expected, when n is too large, $M_D(C_i)[n]$ is significantly larger than the true value $M(C_i)$, for $i = 1, 2$. Observe also that for the non-convex test section with boundary C_2, $M_D(C_2)[n]$ is significantly *smaller* than $M(C_2)$ when $n \leq 20$, an effect caused by the approximating polygon $P[n]$ becoming more convex as n becomes smaller. Thus in general a balance must be achieved, and our test profiles indicate that values of n between 30 and 60 are appropriate.

We tested for the second data set, and for each $n = 30$, 40, 50 and 60, the hypothesis that the mean value of $K_D(C)[n]$ for the normal T-lymphocyte nuclear profiles is equal to the mean value of $K_D(C)[n]$ for the malignant T-cell nuclear profiles, against the alternative hypothesis that the former mean is less than the latter. This was done with an ordinary t-test, applied to log-transformed data. The resulting p-values were all below 10^{-3}.

Based on this test, we can again claim that the indicator $K_D(C)$ successfully distinguishes between the malignant T-cell and normal T-lymphocyte nuclear profiles. To be more specific, the differences between the mean values for the two samples are highly significant when anywhere between 30 and 60 vertices are used for the approximating polygons. This and the

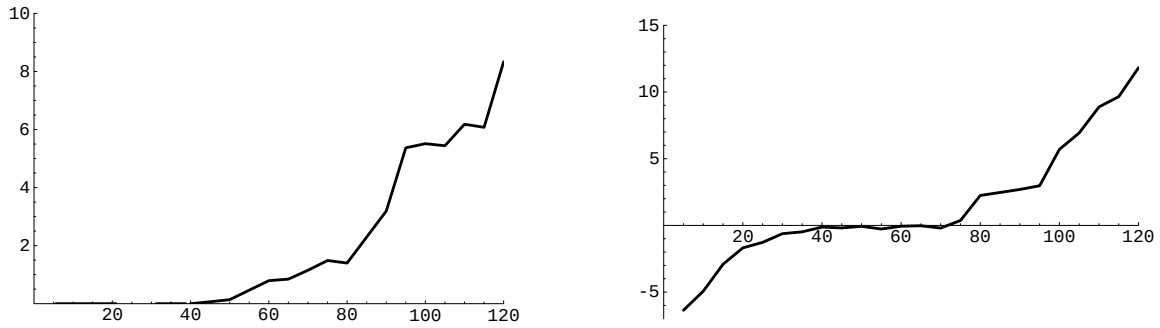

FIGURE 12. $M_D(C_1)[n] - 2\pi$ (left) and $M_D(C_2)[n] - 12.6405$ (right) against n.

very high correlation between the values of $K_D(C)[n]$ for $n = 30$, 40, 50, and 60 (we omit the details) suggest that the procedure is robust in that the choice of n is not too critical.

7. Perspectives

The examples used in this paper focus on the cytological aspects of the transformation from normal or benign to malignant nuclear characteristics as evidenced by two-dimensional profiles. In the routine practice of histopathology the cytological evaluation of nuclear features is mandatory for tentatively diagnosing a neoplastic cell population as malignant, and both nuclear size and nuclear shape are variables used in the diagnosis. If such qualitative judgements can be quantified in a diagnostic setting, objective and reproducible diagnosis of malignancy is a possibility in a (semi)-automated image analysis setting. Moreover, the shape discrimination technique based on total curvature is expected to be applicable in a much broader context, for example in malignancy grading at the histological architectural level and not only at the cytological level. Thus, one could use the technique for quantifying the degree of glandular irregularity in adenocarcinomas, and in this way provide objective malignancy grading data of prognostic impact in common malignant neoplasms like breast cancer.

8. Appendix

Given a finite set of points $p_0, \ldots, p_{m-1}$ on a closed curve C in $\mathbb{R}^2$, there are many methods of generating a simple closed curve C_m that interpolates the points $p_0, \ldots, p_{m-1}, p_0$. A standard one is cubic spline interpolation. Between each successive pair of points p_i and p_{i+1} a cubic curve is defined, of the form

$$A_i(t) = (a_{i0} + a_{i1}t + a_{i2}t^2 + a_{i3}t^3, b_{i0} + b_{i1}t + b_{i2}t^2 + b_{i3}t^3),$$

for $0 \leq t \leq 1$. There are many different formulas for the coefficients of these cubics. It is possible to generate a curve C_m that is C^2. It is also possible to guarantee that various features of the polygon P are preserved by C_m. An ideal interpolation would be one that results in a C^2 curve C_m that is simple if P is simple and does not have any unnecessary kinks. The latter can be achieved if C_m also preserves local convexity and concavity of P. As hints to the quite substantial literature on this topic we refer the reader to [6, 17, 19] and the references in these papers.

We did not attempt an optimal interpolation here, instead preferring to utilize a ready-made implementation, the `SplineFit` function in Mathematica. This produces a curve C_m that is

C^2 except possibly at the initial point p_0. Accordingly, we calculated $M_{AS}(C)$ by

$$M_{AS}(C) = M(C_m), \tag{11}$$

where C_m is the curve produce by Mathematica's `SplineFit` function and $M(C_m)$ is given by (6). Note that in this case $\psi_i = 0$ for $i = 1, \dots, m-1$, but ψ_0 is not in general zero and must be calculated. The coefficients of all the cubic polynomials involved in the interpolation can be obtained in Mathematica, making the computations possible.

We also employed Fourier interpolation, when C is a function $r(\theta)$, $0 \le \theta \le 2\pi$ in polar coordinates. Then $p_i = (r(\theta_i), \theta_i)$, $i = 0, \dots, m-1$, where

$$0 \le \theta_0 < \cdots < \theta_{m-1} < 2\pi.$$

We followed the approach via splines on the circle described in [23, pp. 21–24]. The method produces approximations to given data that minimize a certain energy function that depends on a parameter λ; if we set $\lambda = 0$, the curves obtained are in fact interpolations, agreeing exactly with the given data at the corresponding values of θ.

For simplicity, we suppose that m is even and $\theta_i = (2\pi i)/m$, $i = 0, \dots, m-1$. The Fourier coefficients for the interpolation are given by

$$a_0 = \frac{1}{m} \sum_{i=0}^{m-1} r(\theta_i),$$

$$a_{m/2} = \frac{1}{m} \sum_{i=0}^{m-1} r(\theta_i) \cos \pi i,$$

and for $k = 1, \dots, m/2 - 1$,

$$a_k = \frac{\sqrt{2}}{m} \sum_{i=0}^{m-1} r(\theta_i) \cos k\theta_i$$

and

$$b_k = \frac{\sqrt{2}}{m} \sum_{i=0}^{m-1} r(\theta_i) \sin k\theta_i.$$

Then the Fourier interpolation curve C_m has radial function

$$r_{C_m}(\theta) = a_0 + \sum_{k=1}^{m/2-1} \left(\sqrt{2} a_k \cos k\theta + \sqrt{2} b_k \sin k\theta \right) + a_{m/2} \cos(m\theta/2), \tag{12}$$

for $0 \le \theta \le 2\pi$.

While $r_{C_m}(\theta)$ is clearly infinitely differentiable as a function of θ, it is unfortunately not true that the curve C_m is always smooth. A specific example is obtained by taking $m = 2$, $r(\theta_0) = 1$ and $r(\theta_1) = 0$, which results in

$$r_{C_1}(\theta) = \frac{1}{2}(1 + \cos \theta).$$

This curve is a cardioid, which has a cusp at the origin. Other problems can arise when the curve given by (12) passes through the origin. It is easy to generate examples where the curve obtained is not even a simple curve.

The formula for curvature in polar coordinates (inside the absolute value signs in the integrand in (7)) shows that the curve C_m will be a simple smooth curve provided it does not pass through the origin. This is easy to check and was always the case for our data. Accordingly, we calculated $M_{AF}(C)$ by

$$M_{AF}(C) = M(C_m), \tag{13}$$

where $M(C_m)$ is given by (7).

References

[1] W. Alt, *Statistics and dynamics of cellular shape changes*, in: On Growth and Form. Spatio-Temporal Formation in Biology, M. A. J. Chaplain, G. D. Singh, and J. C. McLachlan, eds., Wiley, Chichester, pp. 287–307.

[2] Y. D. Burago and V. A. Zalgaller, *Geometric Inequalities*, Springer, Berlin, 1980.

[3] R. J. Gardner, *Geometric Tomography*, Cambridge University Press, New York, 1995.

[4] M. A. J. Chaplain, G. D. Singh, and J. C. McLachlan (eds.), *On Growth and Form. Spatio-Temporal Pattern Formation in Biology*, Wiley, Chichester, 1999.

[5] W. Fenchel, *Über Krümmung und Windung geschlossener Raumkurven*, Math. Ann. **101** (1929), 238–252.

[6] T. N. T. Goodman and K. Unsworth, *Shape-preserving interpolation by parametrically defined curves*, SIAM J. Numer. Anal. **25** (1988), 1453–1465.

[7] D. H. Gottlieb, *All the way with Gauss-Bonnet and the sociology of mathematics*, Amer. Math. Monthly **103** (1996), 457–469.

[8] A. Gray, *Modern Differential Geometry of Curves and Surfaces*, CRC Press, Boca Raton, 1993.

[9] A. Hobolth and E. B. V. Jensen, *Modelling stochastic changes in curve shape, with an application to cancer diagnostics*, Adv. in Appl. Probab. **32** (2000), 344–362.

[10] A. Hobolth and E. B. V. Jensen, *Stereological analysis of shape*, Image Anal. Stereol. **21** (2002), 23–29.

[11] A. Hobolth, J. Pedersen, and E. B. V. Jensen, *A deformable template model, with special reference to elliptical templates*, J. Math. Imaging Vision **17** (2002), 131–137.

[12] A. Hobolth, J. Pedersen, and E. B. V. Jensen, *A continuous parametric shape model*, Ann. Inst. Statist. Math. **55** (2003), 227–242.

[13] E. B. V. Jensen and F. Sørensen, *A note on stereological estimation of the volume-weighted second moment of particle volume*, J. Microsc. **164** (1991), 21–27.

[14] M. Ladekarl, *Quantitative histopathology in ductal carcinoma of the breast: prognostic value of mean nuclear size and mitotic counts*, Cancer **75** (1995), 2114–2122.

[15] M. Ladekarl, T. Bæk-Hansen, R. H. Nielsen, C. Mouritzen, U. Henriques, and F. B. Sørensen, *Objective malignancy grading of squamous cell carcinoma of the lung. Stereologic estimates of the mean nuclear size are of prognostic value, independent of clinical stage of disease*, Cancer **76** (1995), 797–802.

[16] P. Mani-Levitska, *Characterizations of convex sets*, in: Handbook of Convex Geometry, P. M. Gruber and J. M. Wills, eds., North-Holland, Amsterdam, 1993, pp. 19–41.

[17] C. Manni, *On shape preserving C^2-Hermite interpolation*, BIT **41** (2001), 127–148.

[18] J. W. Milnor, *On the total curvature of knots*, Ann. of Math. **52** (1950), 248–257.

[19] M. Sakai, *Osculatory interpolation*, Comput. Aided Geom. Design **18** (2001), 739–750.

[20] L. A. Santaló, *Integral Geometry and Geometric Probability*, Addison-Wesley, Reading, MA, 1976.

[21] D. Stoyan and H. Stoyan, *Fractals, Random Shapes and Point Fields*, Wiley, Chichester, 1998.

[22] F. B. Sørensen, *Objective histopathological grading of cutaneous malignant melanomas by stereological estimation of nuclear volume: prediction of survival and disease-free period*, Cancer **63** (1989), 1784–1798.

[23] G. Wahba, *Spline models from observational data*, Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA, 1990.

Department of Mathematics, Western Washington University, Bellingham, WA 98225-9063
E-mail address: Richard.Gardner@wwu.edu

Bioinformatics Research Center, Department of Computer Science, University of Aarhus, Ny Munkegade, DK–8000 Aarhus C, Denmark
E-mail address: asger@daimi.au.dk

Laboratory for Computational Stochastics, Department of Mathematical Sciences, University of Aarhus, Ny Munkegade, DK–8000 Aarhus C, Denmark
E-mail address: eva@imf.au.dk

University Institute of Pathology, Aarhus University Hospital, Aarhus Amtssygehus, Tage-Hansens Gade 2, DK-8000 Aarhus C, Denmark
E-mail address: flemming.soerensen@aas.auh.dk