

SIGNATURE CURVES STATISTICS OF DNA SUPERCOILS

CHERI SHAKIBAN and PETER LLOYD

*Department of Mathematics, University of St. Thomas
St. Paul, MN 55105, USA*

Abstract. In this paper we describe the Euclidean signature curves for two dimensional closed curves in the plane and their generalization to closed space curves. The focus will be on discrete numerical methods for approximating such curves. Further we will apply these numerical methods to plot the signature curves related to three-dimensional simulated DNA supercoils. Our primary focus will be on statistical analysis of the data generated for the signature curves of the supercoils. We will try to establish some relationships between the statistics and the shape of the signature curve. The hope is that these findings will provide a more efficient way for computers to search for and identify signature curves corresponding to similar DNA supercoils.

1. Introduction

When one begins the examination of signature curves, it is helpful to understand why they are so important. Signature curves are most useful in the study of computer vision applications because they allow any object to be represented by a unique curve which is invariant under Euclidean transformations such as rotation. The ability for a computer to *view* objects as invariant curves holds much promise for future developments in the field of artificial intelligence. Signature curves can also be used in medical imaging devices such as CAT or MRI scans. There are also many military and civil defense systems that can employ the use of signature curves for object recognition purposes. Signature curves can be calculated for curves that are described by functions algebraically but this process can be difficult as the calculations get very tedious. However they can simply be created numerically by methods described by Calabi *et al*, [2]. These numerical methods can also be applied to find the signature curves of an image taken with a camera or other imaging device. One can find the boundary of such an image using a segmentation algorithm such as the method of active contours or snakes described in [4], and thereby create the corresponding signature curve.

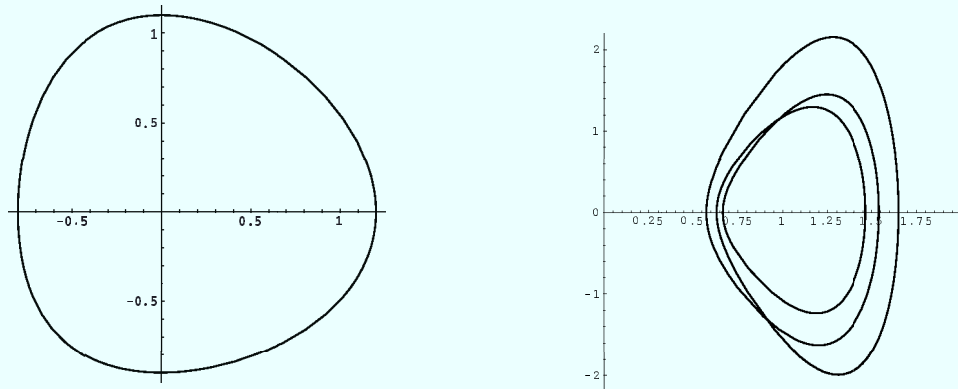


Figure 1. The Original curve C and its Signature curve S

Algebraically, the Euclidean signature curve associated to planar curve C in \mathbb{R}^2 that is parametrized by arclength $s = s(t) = \int_0^t \sqrt{x_t^2 + y_t^2}$, i.e. $C = \{x(s), y(s)\}$, is the curve $S = \{\kappa, \kappa_s\}$, where $\kappa = \kappa(s) = x_s y_{ss} - y_s x_{ss}$ is the curvature as a function of arclength and κ_s is its derivative with respect to arclength.

The Euclidean signature curve is then displayed by simply plotting (κ, κ_s) . It is important to note however, that this formula for curvature only applies to Euclidean signature curves. In order to plot *affine signature curves*, the affine curvature and arclength must be used, see [2].

Figure 1 shows a graph of a curve C in its parametric form $x(t) = \cos t + \frac{1}{5} \cos^2 t$, $y(t) = \sin t + \frac{1}{10} \sin^2 t$ followed by the graph of its algebraic Euclidean signature curve S . Note that if the original curve is rotated or translated in any direction, the signature curve S remains unchanged.

Although the formulated approach to finding curvature is effective for explicitly parametrized curves, it is not used in computer vision applications as it is not so easy to fit a function to a segmented image. Thus, it becomes important that curvature be approximated by some numerical method that respects the underlying symmetry group. The process for approximating discrete curvature begins by placing three successive mesh points on the original curve (image). Three mesh points is optimal because the Euclidean curvature, as defined above, is a second order differential function. Let A , B , and C denote the three points, and let $a = d(A, B)$, $b = d(B, C)$, $c = d(A, C)$ be the distances between them. The circle passing through the points A , B and C in Figure 2 is an approximation of the osculating circle at point B . As a result, $\tilde{\kappa} = \tilde{\kappa}(A, B, C)$, which is the reciprocal of the radius of this circle, acts as an approximation to the curvature at point B . We can apply Heron's formula to find the radius of the circle passing through these three points,

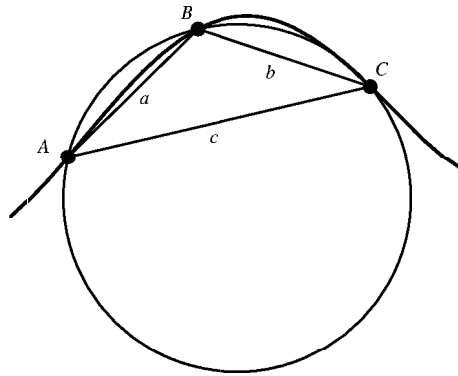


Figure 2. Original curve with points A , B , and C and the circle going through these points

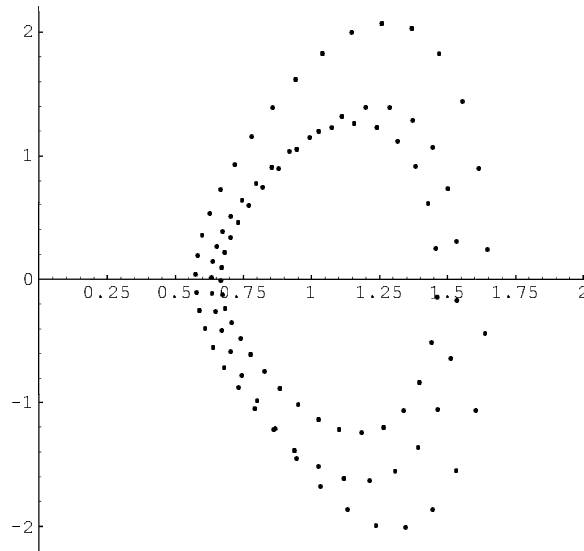


Figure 3. The discrete version of the signature curve obtained numerically for the curve C of Figure 1

and thus find a good approximation of the curvature.

$$\kappa(A, B, C) \approx \tilde{\kappa}(A, B, C) = \frac{4\Delta}{abc} = \pm \frac{4\sqrt{s(s-a)(s-b)(s-c)}}{abc}$$

where $s = \frac{1}{2}(a + b + c)$.

The first derivative of curvature can also be approximated using a centered difference formula

$$\tilde{\kappa}_s(P_{i-2}, P_{i-1}, P_i, P_{i+1}, P_{i+2}) = \frac{\tilde{\kappa}(P_i, P_{i+1}, P_{i+2}) - \tilde{\kappa}(P_{i-2}, P_{i-1}, P_i)}{d(P_{i+1}, P_{i-1})}.$$

This formula is only accurate when the points are fairly regularly spaced around the curve, a more general formula was found by Mireille Boutin [1]. Thus, the entire numerically discrete approximation to the Euclidean signature curve can be plotted by

$$(\tilde{\kappa}(P_{i-1}, P_i, P_{i+1}), \tilde{\kappa}_s(P_{i-2}, P_{i-1}, P_i, P_{i+1}, P_{i+2})).$$

Figure 4 gives the smooth and discrete versions of the signature curve obtained numerically for the curve given in polar form $r(\theta) = 3 + \frac{1}{10} \cos 3\theta + \frac{1}{40} \cos 7\theta$.

As was stated previously, signature curves have many applications today in the field of medical imaging. One reason for this is that the signature curve for an object remains the same despite any rigid movement of the object, e.g. tilting at different angles. This fact is extremely useful for DNA imaging, since DNA is constantly in motion. In particular however, signature curves seem to be well suited for characterizing DNA supercoils. DNA supercoiling occurs when the ends of DNA strands are not allowed to rotate, which in effect causes the DNA strands to coil over each other. Since DNA supercoils are space curves, their signature curves are parameterized by the four basic differential invariants: curvature κ , torsion, τ , and their derivatives with respect to arclength κ_s, τ_s . However, it has been proved [5], that a complete signature requires only three of them, namely, κ, κ_s, τ . The numerical invariant expressions for the differential invariants can be found in [1]. For the time being, we examine signature curves by analyzing curvature versus the derivative of curvature and torsion versus the derivative of torsion.

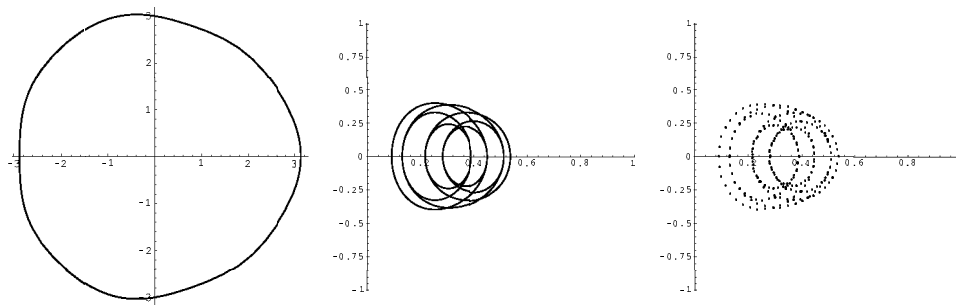


Figure 4. The original curve and its smooth and discrete signature curves

One difficulty when considering signature curves is how a computer will interpret these curves and decide if they are representing similar or the same objects. Computers do not actually *see* signature curves the same way humans do, but they see them rather as a set of points that the curve is composed of. One possible way of interpreting these points is by applying various statistical measurements to them. For the most part these measurements will be unique to the data they represent, and thus provide an excellent way for computers to compare and identify the signature curve for a particular object. For this study we use the simulated points which were generated from a model to form three-dimensional DNA supercoils by the Numerical Analysis group at École Polytechnique Fédérale de Lausanne (EPFL) under the supervision of John Maddocks [3]. By using the purely numeric approach developed in [1] coupled with a *Mathematica* program, we graph the signature curves using curvature κ versus the derivative of curvature κ_s and torsion τ versus the derivative of torsion τ_s . Along with these curves, histograms for κ , κ_s , τ and τ_s are also plotted and the following statistics for each of these variables are provided: mean, harmonic mean, median, variance, standard deviation, sample range, mean deviation, median deviation, quartile deviation, skewness, quartile skewness, and the Kurtosis excess [6].

One problem we are facing is how to compare both the derivative and non-derivative versions of the histograms. However, since both histograms and their accompanying statistical findings are representing the signature curve, it seems reasonable to take the difference or the sum of certain statistics to be a good indicator of an overall statistical measurement for a particular curve. We apply this method first by taking $|\kappa SD - \kappa_s SD|$ where κSD and $\kappa_s SD$ represent the standard deviations for the curvature and its derivative, respectively. We can then see that signature curves whose difference in standard deviations were low are much more spread out, than those whose difference in standard deviations are relatively high. One way to show this beyond simple observation is to examine the Kurtosis excesses from both histograms where the standard deviations are taken as Kurtosis excess to measure the “peakedness” of the histogram. Thus, if the histogram is more peaked, the Kurtosis excess is high, and vice versa. Because of this, the sum of the Kurtosis excesses can be taken to get overall measurement of how strong the peakedness is between the two histograms. The graph in Figure 5 represents a linear regression plot of the difference in standard deviations versus the sum of the Kurtosis excesses. As one can see there is a fairly strong correlation between the two variables. However, this strong linear correlation may have to do in part with several outliers that were within the data.

Therefore, the difference in standard deviation for torsion and its derivative should also be considered. As one can see from Figure 6, there is a stronger linear regression (larger R -Sq(adj)) present inspire the fact that the torsion and derivative

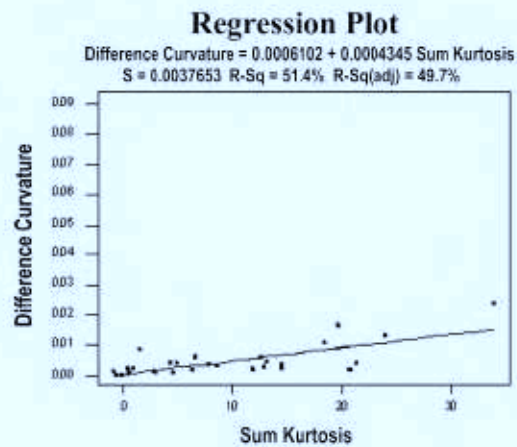


Figure 5. Difference Curvature Standard Deviation versus Sum of Curvature Kurtosis Excesses

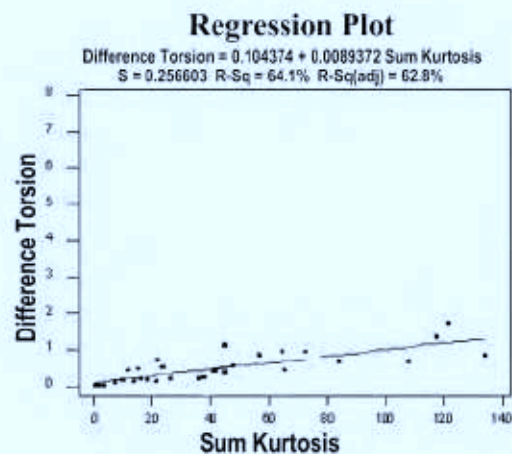


Figure 6. Difference Torsion Standard Deviation versus Sum of Torsion Kurtosis Excesses

of torsion histograms have more severe peaks to them, than the histograms for curvature and its derivative. This strong linear regression is a good statistical measure for representing the original curve and its signature curve.

As one can see the original graph in Figure 7 is very elliptical, and thus the signature curves of curvature versus the derivative of curvature and torsion versus the derivative of torsion are very spread out. Similarly, if one examines the histograms for both curvature and torsion, they will see that the histograms are fairly uniform, and do not have any large spikes in them. On the other hand, the original

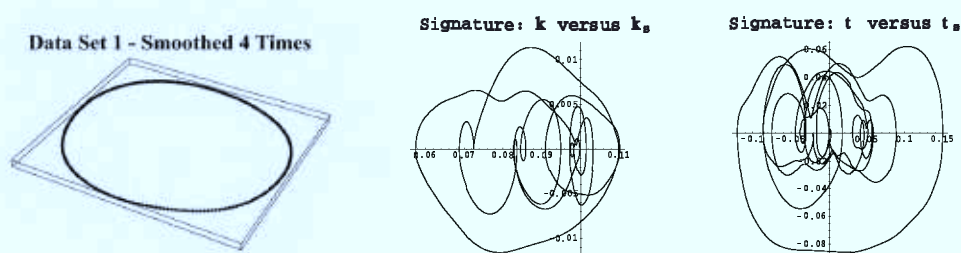


Figure 7. The original curve and its signature curves

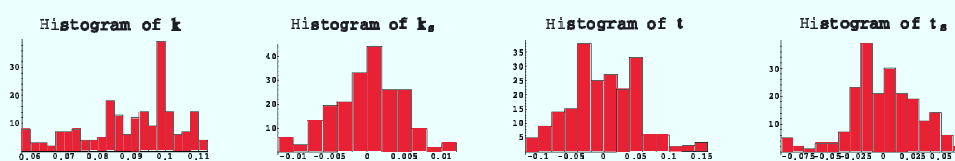


Figure 8. The Histograms for the curve in Figure 7

curve in Figure 9 is not elliptical at all, and thus its signature curves are much less spread out than those in Figure 7. In addition, when the signature curve is tightly wound around the origin as in Figure 9, large arcing lines will represent any outlying points on the histogram. This is especially true when the sample range for the histograms representing the signature curve is large compared to the mean value for the histograms' points.

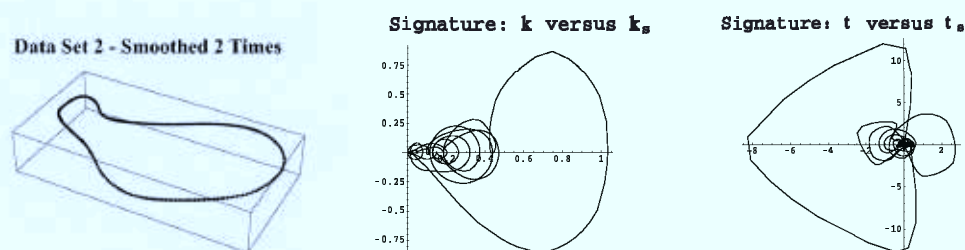


Figure 9. The original curve and its signature curves

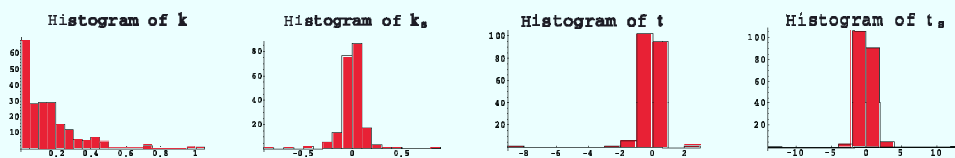


Figure 10. The Histograms for the curve in Figure 9

Although much work has gone into these findings, there is still much that can be done in terms of a statistical analysis concerning three-dimensional Euclidean signature curves. In fact, due to the existence of heteroscedastic responses in the observed variances, one could use a variance-stabilizing transformation such as \sqrt{y} for the regression model [6]. Further, it would be interesting to actually apply some of these findings to real-life DNA supercoils and see how accurate by computer can identify these supercoils. This experiment could be run to judge just how effective statistics are as a way to identify signature curves for particular objects.

References

- [1] Boutin M., *Numerically Invariant Signature Curves*, Int. J. Computer Vision **40** (2000) 235–248.
- [2] Calabi E., Olver P., Shakiban C., Tannenbaum A. and Haker S., *Differential and Numerically Invariant Signature Curves Applied to Object Recognition*, Int. J. Computer Vision **26** (1998) 107–135.
- [3] Hoffman K., Manning R. and Maddocks J., *Link, Twist, Energy, and the Stability of DNA Minicircles*, Biopolymers, to appear.
- [4] Kichenassamy S., Kumar A., Olver P., Tannenbaum A. and Yezzi A., *Conformal Curvature Flows: From Phase Transitions to Active Vision*, Arch. Rat. Mech. Anal. **134** (1996) 275–301.
- [5] Shakiban C., *Invariant Signature Curves in 3D*, Preprint, 2003.
- [6] Mendenhall W., *A Second Course in Statistics. Regression Analysis*, Sixth Edition, Pearson Education, Inc., Upper Saddle River, NJ, 2003.