

5-1-2004

Geometry of Protein Structures. I. Why Hyperbolic Surfaces are a Good Approximation for Beta-Sheets

Boguslaw Stec

Vladik Kreinovich

University of Texas at El Paso, vladik@utep.edu

Follow this and additional works at: http://digitalcommons.utep.edu/cs_techrep

 Part of the [Computer Engineering Commons](#)

Comments:

UTEP-CS-04-15.

Published in *Geombinatorics*, 2005, Vol. 15, Vol. 1, pp. 18-27.

Recommended Citation

Stec, Boguslaw and Kreinovich, Vladik, "Geometry of Protein Structures. I. Why Hyperbolic Surfaces are a Good Approximation for Beta-Sheets" (2004). *Departmental Technical Reports (CS)*. Paper 299.

http://digitalcommons.utep.edu/cs_techrep/299

This Article is brought to you for free and open access by the Department of Computer Science at DigitalCommons@UTEP. It has been accepted for inclusion in Departmental Technical Reports (CS) by an authorized administrator of DigitalCommons@UTEP. For more information, please contact lweber@utep.edu.

GEOMETRY OF PROTEIN STRUCTURES. I. WHY HYPERBOLIC SURFACES ARE A GOOD APPROXIMATION FOR BETA-SHEETS

Boguslaw Stec^{1,2} and Vladik Kreinovich^{1,3}

¹Bioinformatics Program

²Chemistry Department

³Department of Computer Science

University of Texas at El Paso

El Paso, TX 79968, USA

bstec@utep.edu, vladik@cs.utep.edu

Abstract. *Protein structure is invariably connected to protein function. To analyze the structural changes of proteins, we should have a good description of basic geometry of proteins' secondary structure. A beta-sheet is one of important elements of protein secondary structure that is formed by several fragments of the protein that form a surface-like feature. The actual shapes of the beta-sheets can be very complicated, so we would like to approximate them by simpler geometrical shapes from an approximating family. Which family should we choose? Traditionally, hyperbolic (second order) surfaces have been used as a reasonable approximation to the shape of beta-sheets. In this paper, we show that, under reasonable assumptions, these second order surfaces are indeed the best approximating family for beta-sheets.*

Introduction. Proteins are biological polymers that perform most of the life's function. A single chain polymer (protein) is folded in such a way that forms local substructures called secondary structure elements. In order to study the structure and function of proteins it is extremely important to have a good geometrical description of the proteins structure. There are two important secondary structure elements: alpha helices and beta-sheets. A part of the protein structure where different fragments of the polypeptide align next to each other in extended conformation forming a surface-like feature defines a secondary structure called a *beta pleated sheet*, or, for short, a *beta-sheet*; see, e.g., (Branden et al. 1999).

Beta-sheets are coming in many forms and shapes. In some

cases, we have a cylinder-like structure called a *beta-barrel* that is “closed” in one dimension and “open” in the other, but in most cases, we have a surface that is open in both directions.

The actual shapes of the beta-sheets can be very complicated, so we would like to approximate them by simpler shapes from an approximating family. Which family should we choose?

Traditionally, hyperbolic (second order) surfaces have been used as a reasonable approximation to the shape of beta-sheets; see, e.g., (Novotny et al. 1984). However, it is not clear whether they are indeed a good approximating family.

Of course, the more parameters we allow, the better the approximation. So, the question can be reformulated as follows: for a given number of parameters (i.e., for a given dimension of approximating family), which is the best family?

In this paper, we formalize and solve this problem. Specifically, we show that, under reasonable assumptions, these second order surfaces are indeed the best low-parameter approximating family for beta-sheets.

Formalizing the problem. All proposed families of sets have analytical (or piece-wise analytical) boundaries, so it is natural to restrict ourselves to such families. By definition, when we say that a piece of a boundary is analytical, we mean that it can be described by an equation $F(x) = 0$ for some analytical function $F(x) = F(x_1, x_2, x_3) = a_0 + a_1 \cdot x_1 + a_2 \cdot x_2 + a_3 \cdot x_3 + a_{11} \cdot x_1^2 + a_{12} \cdot x_1 \cdot x_2 + \dots$. So, in order to describe a family, we must describe the corresponding class of analytical functions $F(x_1, x_2, x_3)$.

Since we are interested in finite-dimensional families of sets, it is natural to consider finite-dimensional families of functions, i.e., families of the type $\{C_1 \cdot F_1(x) + \dots + C_d \cdot F_d(x)\}$, where $F_i(x)$ are given analytical functions, and C_1, \dots, C_d are arbitrary (real) constants.

For example, a general second-order surface can be described by the formula

$$a_0 + \sum_{i=1}^3 a_i \cdot x_i + \sum_{i=1}^3 \sum_{j=1}^3 a_{ij} \cdot x_i \cdot x_j = 0, \quad (1)$$

with $F_1(x) = 1$, $F_2(x) = x_1$, $F_3(x) = x_2$, $F_4(x) = x_3$, $F_5(x) = x_1^2$, \dots , and $d = 10$ parameters $a_0, a_1, a_2, a_3, a_{11}, a_{12}, a_{13}, a_{22}, a_{23}$, and a_{33} .

The question is: which of such families is the best?

When we say “the best”, we mean that on the set of all such families, there must be a relation \geq describing which family is better or equal in quality. This relation must be transitive (if A is better than B , and B is better than C , then A is better than C). This relation is not necessarily asymmetric, because we can have two approximating families of the same quality. However, we would like to require that this relation be *final* in the sense that it should define a unique *best* family A_{opt} (i.e., the unique family for which $\forall B (A_{\text{opt}} \geq B)$). Indeed:

- If none of the families is the best, then this criterion is of no use, so there should be *at least one* optimal family.
- If *several* different families are equally best, then we can use this ambiguity to optimize something else: e.g., if we have two families with the same approximating quality, then we choose the one which is easier to compute. As a result, the original criterion was not final: we get a new criterion ($A \geq_{\text{new}} B$ if either A gives a better approximation, or if $A \sim_{\text{old}} B$ and A is easier to compute), for which the class of optimal families is narrower. We can repeat this procedure until we get a final criterion for which there is only one optimal family.

It is reasonable to require that the relation $A \geq B$ should not change if we change the coordinate system and/or the measuring unit, i.e., if we shift, rotate, and/or re-scale ($x \rightarrow \lambda \cdot x$) all the points x ; in other word, the relation $A \geq B$ should be shift-, rotation- and scale-invariant.

Now, we are ready for the formal definitions.

Definition 1. Let $d > 0$ be an integer. By a d -dimensional family, we mean a family A of all functions of the type $\{C_1 \cdot F_1(x) + \dots + C_d \cdot F_d(x)\}$, where $F_i : R^3 \rightarrow R$ are given analytical functions, and C_1, \dots, C_d are arbitrary (real) constants.

Definition 2. By an *optimality criterion*, we mean a transitive relation \geq on the set of all d -dimensional families. We say that a criterion is *final* if there exists one and only one optimal family, i.e., a family A_{opt} for which $\forall B (A_{\text{opt}} \geq B)$. We say that a criterion \geq is *shift-* (corr., rotation- and scale-invariant) if for every two families A and B , $A \geq B$ implies $TA \geq TB$, where TA is a shift (rotation, scaling) of the family A .

We have already mentioned that to describe general second order surfaces in the above way, we need $d = 10$ parameters. We show that among all possible families with $d = 10$, second order surfaces are indeed the best approximating family. Moreover, we show that this family remains optimal even if we allow higher-dimensional families, up to $d = 12$. (With larger number of parameters, we may get new possible approximating families as well.)

Proposition 1. ($d \leq 12$) Let \geq be a final optimality criterion which is shift-, rotation-, and scale-invariant, and let A_{opt} be the corresponding optimal family. Then, every function $F(x)$ from this family A_{opt} is a quadratic polynomial.

Comment. Thus, the corresponding surface $F(x) = 0$ is a second-order surface. This result is in good accordance with the experimental data described, e.g., in (Branden et al. 1999).

A natural next question is: do we need *all* quadratic surfaces to describe protein shapes, or a subclass is sufficient? The following results provides a mathematical background for answering this question:

Proposition 2. Let \geq be a final optimality criterion which is shift-, rotation-, and scale-invariant, and let A_{opt} be the corresponding optimal family of quadratic polynomials. Then, we have 4 possibilities:

- (a) A_{opt} consists of all linear functions, so the corresponding surfaces $F(x) = 0$ are planes;
- (b) surfaces $F(x) = 0$ are planes and spheres;

- (c) A_{opt} consists of all the quadratic functions (1) for which $a_{11} + a_{22} + a_{33} = 0$;
- (d) A_{opt} consists of all possible quadratic surfaces.

Comment. Since some observed surfaces are cylindrical, and cylindrical functions do not belong to the classes (a)–(c), we thus conclude that, in general, to approximate the shapes of beta-sheets, we need to use all possible quadratic surfaces.

Proof of Propositions 1 and 2. This proof is similar to the ones from (Kreinovich et al. 2000) and (Nguyen et al. 1997).

1. Let us first show that the optimal family A_{opt} is itself shift-, rotation-, and scale-invariant.

Indeed, let T be an arbitrary shift, rotation, or scaling. Since A_{opt} is optimal, for every other family B , we have $A_{\text{opt}} \geq T^{-1}B$ (where T^{-1} means the inverse transformation). Since the optimality criterion \geq is invariant, we conclude that $TA_{\text{opt}} \geq T(T^{-1}B) = B$. Since this is true for every family B , the family TA_{opt} is also optimal. But since our criterion is final, there is only one optimal family and therefore, $TA_{\text{opt}} = A_{\text{opt}}$. In other words, the optimal family is indeed invariant.

2. Let us now show that all functions from A_{opt} are polynomials.

Indeed, every function $F \in A_{\text{opt}}$ is analytical, i.e., can be represented as a Taylor series (sum of monomials). Let us combine together monomials $c \cdot x_1^{d_1} \cdot x_2^{d_2} \cdot x_3^{d_3}$ of the same total degree $k = d_1 + d_2 + d_3$; then we get $F(x) = F_0(x) + F_1(x) + \dots + F_k(x) + \dots$, where $F_k(x)$ is the sum of all monomials of degree k . Let us show, by induction over k , that for every k , the function $F_k(x)$ also belongs to A_{opt} .

Let us first prove that $F_0(x) \in A_{\text{opt}}$. Since the family A_{opt} is scale-invariant, we conclude that for every $\lambda > 0$, the function $F(\lambda z)$ also belongs to A_{opt} . For each term $F_k(x)$, we have $F_k(\lambda \cdot x) = \lambda^k \cdot F_k(x)$, so $F(\lambda \cdot x) = F_0(x) + \lambda \cdot F_1(x) + \dots \in A_{\text{opt}}$. When $\lambda \rightarrow 0$, we get $F(\lambda \cdot x) \rightarrow F_0(x)$. The family A_{opt} is finite-dimensional hence closed; so, the limit $F_0(x)$ also belongs to A_{opt} . The induction base is proven.

Let us now suppose that we have already proven that for all $k < s$, we have $F_k(x) \in A_{\text{opt}}$. Let us prove that $F_s(x) \in A_{\text{opt}}$. For that, let us take $G(x) = F(x) - F_1(x) - \dots - F_{s-1}(x)$. We already know that $F_1, \dots, F_{s-1} \in A_{\text{opt}}$; so, since A_{opt} is a linear space, we conclude that $G(x) = F_s(x) + F_{s+1}(x) + \dots \in A_{\text{opt}}$.

The family A_{opt} is scale-invariant, so, for every $\lambda > 0$, the function $G(\lambda \cdot x) = \lambda^s \cdot F_s(x) + \lambda^{s+1} \cdot F_{s+1}(x) + \dots$ also belongs to A_{opt} . Since A_{opt} is a linear space, the function

$$H_\lambda(x) = \lambda^{-s} \cdot G(\lambda \cdot x) = F_s(x) + \lambda \cdot F_{s+1}(x) + \lambda^2 \cdot F_{s+2}(x) + \dots$$

also belongs to A_{opt} .

When $\lambda \rightarrow 0$, we get $H_\lambda(x) \rightarrow F_s(x)$. The family A_{opt} is finite-dimensional hence closed; so, the limit $F_s(x)$ also belongs to A_{opt} . The induction is proven.

Now, monomials of different degree are linearly independent; therefore, if we have infinitely many non-zero terms $F_k(x)$, we would have infinitely many linearly independent functions in a finite-dimensional family A_{opt} – a contradiction. Thus, only finitely many monomials $F_k(x)$ are different from 0, and so, $F(x)$ is a sum of finitely many monomials, i.e., a polynomial.

3. Let us prove that if a function $F(x)$ belongs to A_{opt} , then its partial derivatives $F_{,i}(x) = \partial F / \partial x_i$ also belong to A_{opt} .

Indeed, since the family A_{opt} is shift-invariant, for every $h > 0$, we get $F(\dots, x_{i-1}, x_i + h, x_{i+1}, \dots) \in A_{\text{opt}}$. Since this family is a linear space, we conclude that a linear combination

$$\frac{F(\dots, x_{i-1}, x_i + h, x_{i+1}, \dots) - F(\dots, x_{i-1}, x_i, x_{i+1}, \dots)}{h}$$

of two functions from A_{opt} also belongs to A_{opt} . Since the family A_{opt} is finite-dimensional, it is closed and therefore, the limit $F_{,i}(x, y)$ of such linear combinations also belongs to A_{opt} .

4. Due to Parts 2 and 3 of this proof, if any polynomial from A_{opt} has a non-zero part F_k of degree $k > 0$, then it also has a non-zero part $(F_k)_{,i}$ of degree $k-1$. Similarly, it has non-zero parts of degrees $k-2, \dots, 1, 0$.

So, in all cases, A_{opt} contains a non-zero constant and a non-zero linear function $F_1(x) = \sum a_i \cdot x_i$. We can now use the fact that the family A_{opt} is rotation-invariant. For every i , let T be a rotation which transforms the vector $a = (a_1, a_2, a_3)$ into the i -th axis, then we conclude that $F_1(Tx) = c \cdot x_i \in A_{\text{opt}}$, and hence $x_i \in A_{\text{opt}}$. So, the family A_{opt} contains at least 4 linearly independent functions: a non-zero constant, x_1 , x_2 , and x_3 .

5. We will now prove, by reduction to a contradiction, that functions from A_{opt} cannot contain terms of third or higher order. Due to Part 4 of this proof, if $F \in A_{\text{opt}}$ has a part of degree > 3 , then A_{opt} also contains a polynomial F_3 all of whose monomials are of degree 3. Thus, it is sufficient to show that A_{opt} cannot contain such a polynomial.

5.1. Indeed, let us assume that A_{opt} contains such a polynomial $F_3(x_1, x_2, x_3)$. Due to rotation-invariance, for every rotation T , the family A_{opt} also contains the polynomial $F_3(Tx)$. In particular, if, as T , we take a 180° rotation around the x_3 -axis, i.e., the transformation $x_1 \rightarrow -x_1, x_2 \rightarrow -x_2, x_3 \rightarrow x_3$, then we conclude that A_{opt} contains the polynomial $F_3^*(x) = F_3(-x_1, -x_2, x_3)$.

Since A_{opt} is a linear space, it also contains polynomials $F^+(x) = (F_3(x) + F_3^*(x))/2$ and $F^-(x) = (F_3(x) - F_3^*(x))/2$. The combination $F^+(x)$ contains all the terms for which the overall degree in x_1 and x_2 is even – or, equivalently, for which the degree in x_3 is odd. Similarly, $F^-(x)$ contains all the monomials for which the degree in x_3 is even. Since $F_3 = F^+ + F^- \neq 0$, at least one of the functions F^+ and F^- is non-zero.

Thus, by selecting the non-zero of the two functions (or F^+ if both are non-zero), we can conclude that A_{opt} contains a polynomial in which either all monomials are even in x_3 or all monomials are odd in x_3 – i.e., all the monomials have the same parity w.r.t. x_3 .

5.2. By considering rotations around x_1 and x_2 , we can similarly split the resulting function, and end up with a new function $F \in A_{\text{opt}}$ in which all the monomials have the same parity w.r.t. x_1 , the same parity w.r.t. x_2 , and the same parity w.r.t. x_3 .

Since all monomials in F have an overall degree 3 (i.e., odd), we

must have either all 3 degrees w.r.t. x_i odd, or one odd and two even.

5.3. In the first case, the only possible monomial is $x_1 \cdot x_2 \cdot x_3$, because this is the only way to represent 3 as the sum of 3 odd natural numbers; thus, $x_1 \cdot x_2 \cdot x_3 \in A_{\text{opt}}$. Rotating 45° around x_3 , we transform $x_1 \cdot x_2$ into $x_1^2 - x_2^2$, so we conclude that the family A_{opt} contains $x_1^2 \cdot x_3 - x_2^2 \cdot x_3$, a polynomial of the second type. Thus, it is sufficient to consider the second case.

5.4. Let us consider the second case. If the degree w.r.t. x_1 is odd, then, as one can easily check, the general form of such a polynomial is $c_1 \cdot x_1^3 + c_2 \cdot x_1 \cdot x_2^2 + c_3 \cdot x_1 \cdot x_3^2$, where at least one of the coefficients c_i is different from 0.

Due to rotation invariance, A_{opt} must contain two similar polynomials in which x_2 (correspondingly, x_3) have an odd degree, such as

$$c_1 \cdot x_2^3 + c_2 \cdot x_2 \cdot x_1^2 + c_3 \cdot x_2 \cdot x_3^2.$$

Thus, A_{opt} contains at least 3 linearly independent monomials of third degree.

5.5. According to Part 3 of our proof, A_{opt} contains partial derivatives of its members. If $c_2 \neq 0$, then we conclude that $F_{,2} = 2c_2 \cdot x_1 \cdot x_2 \in A_{\text{opt}}$ hence $x_1 \cdot x_2 \in A_{\text{opt}}$. Similarly, if $c_3 \neq 0$, then $x_1 \cdot x_3 \in A_{\text{opt}}$. Finally, if $c_2 = c_3 = 0$, this means that $c_1 \neq 0$, so, from $F_{,1} = 3c_1 \cdot x_1^2 \in A_{\text{opt}}$, we can conclude that $x_1^2 \in A_{\text{opt}}$.

In all three cases, A_{opt} contains either $x_i \cdot x_j$ for some $i \neq j$, or x_i^2 .

5.6. If A_{opt} contains, e.g., $x_1 \cdot x_2$, then, due to rotation invariance, it also contains $x_1 \cdot x_3$ and $x_2 \cdot x_3$. Rotating by 45° , we conclude that $x_1^2 - x_2^2 \in A_{\text{opt}}$ and $x_1^2 - x_3^2 \in A_{\text{opt}}$. Overall, we thus have 5 linearly independent quadratic polynomials in A_{opt} . Together with 4 constant and linear polynomials (from Part 4) and 3 cubic polynomials (from Part 5.4), we get a total of $4 + 3 + 5 = 13$ linearly independent functions in A_{opt} – which contradicts to our assumption that $d \leq 12$.

5.7. If A_{opt} contains, e.g., x_1^2 , then, due to rotation invariance, it also contains x_2^2 and x_3^2 . Since A_{opt} is a linear space, it contains,

e.g., $x_1^2 - x_2^2$. Rotating by 45° , we conclude that $x_1 \cdot x_2 \in A_{\text{opt}}$ and similarly, that $x_1 \cdot x_3 \in A_{\text{opt}}$ and $x_2 \cdot x_3 \in A_{\text{opt}}$. Overall, we thus have 6 linearly independent quadratic polynomials $x_i \cdot x_j$ in A_{opt} . Together with 4 constant and linear polynomials (from Part 4) and 3 cubic polynomials (from Part 5.4), we get a total of $4 + 3 + 6 = 14$ linearly independent functions in A_{opt} – which contradicts to our assumption that $d \leq 12$.

Proposition 1 is proven.

6. Let us now prove Proposition 2. We have already proven that the family A_{opt} contains all linear function, so if A_{opt} contains nothing else, we get the case (a).

Let us now consider the case when A_{opt} contains at least one non-linear quadratic function. Since A_{opt} is a linear space, it is sufficient to consider homogenous quadratic expressions

$$\sum a_{ij} \cdot x_i \cdot x_j \in A_{\text{opt}}.$$

Each such expression can be transformed, by an appropriate rotation, into a diagonal form $\sum \lambda_i \cdot y_i^2$. Since the family A_{opt} is rotation-invariant, it contains the corresponding function $\sum \lambda_i \cdot x_i^2$.

If for every function $F \in A_{\text{opt}}$, all three eigenvalues λ_i are equal to each other, then the functions $F(x)$ have the form $a_0 + \sum a_i \cdot x_i + \lambda \cdot (x_1^2 + x_2^2 + x_3^2)$. In this case, the equation $F(x) = 0$ describes a sphere, so we are in the case (b).

The only remaining case is when there exists a function $F \in A_{\text{opt}}$ for which at least two eigenvalues are different, e.g., $\lambda_1 \neq \lambda_2$. In this case, due to rotation-invariance, the family A_{opt} also contains not only the original function $F(x) = \lambda_1 \cdot x_1^2 + \lambda_2 \cdot x_2^2 + \lambda_3 \cdot x_3^2$, but also the rotated version of this function $F^*(x) = \lambda_1 \cdot x_2^2 + \lambda_2 \cdot x_1^2 + \lambda_3 \cdot x_3^2$. Since A_{opt} is a linear space, it also contains their linear combination $(F(x) - F^*(x))/(\lambda_1 - \lambda_2) = x_1^2 - x_2^2$.

Due to rotation-invariance, we conclude that A_{opt} also contains $x_1^2 - x_3^2$. Applying invariance w.r.t. 45° rotations, we conclude that $x_1 \cdot x_2$, $x_2 \cdot x_3$, and $x_1 \cdot x_3$ also belong to A_{opt} . So, A_{opt} contains 4 linearly independent functions 1, x_1 , x_2 , and x_3 , and 5 linearly independent quadratic functions. Thus, the dimension of this family is at least $4 + 5 = 9$.

Ovrecall, the dimension of the set Q of all quadratic functions is 10. So, we had two possibilities: either $\dim(A_{\text{opt}}) = 10$, in which case $A_{\text{opt}} = Q$ (case (d)), or $\dim(A_{\text{opt}}) = 9$, in which case A_{opt} is a linear combination of the above 9 functions – this is exactly case (c). Proposition 2 is proven.

Open problem. We described optimal 12-D families. What is 12 parameters are not enough? What are the best 13-, 14-, etc.- dimensional families? From the proof, we can conclude that these optimal families consist of *algebraic* sets, i.e., sets with boundary $F(x) = 0$ for a polynomial F , but a more specific description is desirable.

Acknowledgments. This work was supported in part by NASA under cooperative agreement NCC5-209, by the Future Aerospace Science and Technology Program (FAST) Center for Structural Integrity of Aerospace Systems, effort sponsored by the Air Force Office of Scientific Research, Air Force Materiel Command, USAF, under grant F49620-00-1-0365, by NSF grants EAR-0112968, EAR-0225670, and EIA-0321328, and by the Army Research Laboratories grant DATM-05-02-C-0046.

The authors are very thankful to Michael Zakharevich for helpful discussions.

References

C.-I. Branden and J. Tooze, *Introduction to Protein Structure*, Garland Publ., N.Y., 1999.

V. Kreinovich and J. Wolff von Gudenberg, “An optimality criterion for arithmetic of complex sets”, *Geombinatorics*, 2000, Vol. 10, No. 1, pp. 31–37.

H. T. Nguyen and V. Kreinovich, *Applications of continuous mathematics to computer science*, Kluwer, Dordrecht, 1997.

J. Novotny, R. E. Bruccoleri, and J. Newell, “Twisted hyperboloid (Strophoid) as a model of beta-barrels in proteins”, *J. Mol. Biol.*, 1984, Vol. 177, pp. 567–573.