Determining Flexibility of Molecules Using Resultants of Polynomial Systems

Robert H. Lewis[1] and Evangelos A. Coutsias[2]

[1] Fordham University, New York, NY 10458, USA

[2] University of New Mexico, Albuquerque, NM 87131, USA

We solve systems of multivariate polynomial equations in order to understand flexibility of three dimensional objects, including molecules.

Protein flexibility is a major research topic in computational chemistry. In general, a polypeptide backbone can be modeled as a polygonal line whose edges and angles are fixed while some of the dihedral angles in it can vary freely. It is well known that a segment of backbone with fixed ends will be (generically) flexible if it includes more than six free torsions. Resultant methods have been applied successfuly to this problem, see [3], [4].

In this work we focus on non-generically flexible structures (like a geodesic dome) that are rigid but become continuously movable under certain relations. A long history: Cauchy (1812), Bricard (1896), Connelly (1978).

In our previous work [8], we began a new approach to understanding flexibility, using not numeric but symbolic computation. We describe the geometry of the object with a set of multivariate polynomial equations, which we solve with resultants. Resultants were pioneered by Bezout, Sylvester, Dixon, and others. The resultant appears as a factor of the determinant of a matrix containing multivariate polynomials. Given the resultant, we described [8] an algorithm that examines it and determines relations for the structure to be flexible. We discovered in this way the conditions of flexibility for an arrangement of quadrilaterals in Bricard [1], which models molecules. Here we significantly extend the algorithm and the molecular structures.

• The flavor here is **algebraic** with **symbolic computation** of polynomials and matrices containing polynomials defined over **exact ground rings**.

• It is **not** numerical computing, graph theory, circuits, trees, Boolean anything.

• One way to solve systems of **multivariate polynomial equations** is **resultants**.

• Big problem: compute **determinant** of matrix with polynomial entries.

• Potential Application: design of nano-machine.

1 Introduction and Summary of Previous Work

Generic protein flexibility has been a major research topic in computational chemistry for a number of years, and it has a key role for many important functions of proteins as molecular machines [10]. In general, a polypeptide backbone can be modeled as a polygonal line whose edges and angles are fixed while some of the dihedral angles formed by successive triplets of edges can vary freely. It is well known that a segment of backbone both of

whose ends are fixed will be (generically) flexible if it includes more than six free torsions. Resultant methods have been applied successfuly to this problem, see [3], [4] and the references therein. In this work we focus on non-generically flexible structures that are rigid but become continuously movable if certain symmetries and relations exist. This subject has a long history. In 1812, Cauchy considered flexibility of three dimensional polyhedra, where each joint can pivot or hinge. He proved that if the polyhedron is convex it must be rigid [2]. But following Bricard's study of flexible non-convex intercrossing octahedra [1], in 1978 Connelly found genuine non-convex flexible polyhedra [5] that can live in 3 space (without self-intersections).

In our previous work [8], we began a new approach to understanding flexibility, using symbolic computation instead of numerical calculation. We describe the geometry of the object or molecule with a set of multivariate polynomial equations. Solving a system of multivariate polynomial equations is a classic, difficult problem. The approach via resultants was pioneered by Bezout, Sylvester, Dixon [7], and others [6]. The resultant res appears as a factor of the determinant det of a matrix containing multivariate polynomials. But often det is too large to compute, even though res is relatively small. I developed a method

that overcomes the problem [9], called Dixon-EDF. Given the resultant, we described [8] an algorithm that examines res and determines relations for the structure to be flexible.

We discovered in this way the conditions of flexibility for a significant arrangement of quadrilaterals in [1].

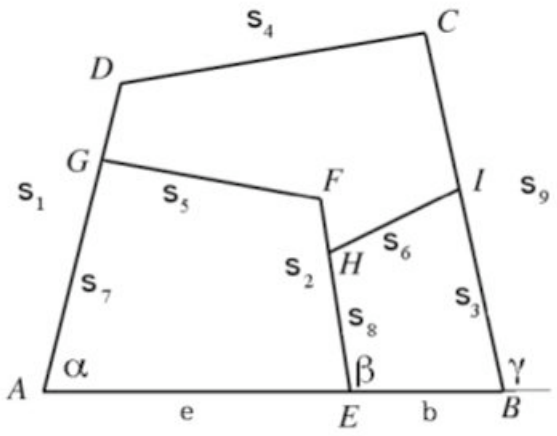

Fig. 1. Bricard's quadrilaterals.

The system in our original formulation had six equations in six variables and eleven parameters. The resultant res, a function of one variable ca and the eleven parameters, has 190981 terms. If the figure is flexible, there are infinitely many possible values for ca. That implies that every coefficient relative to ca in res must vanish. We developed an algorithm $Solve$ to search for relations among the parameters that will kill these coefficients

and so produce flexibility. As lengths of sides in a geometric figure, the parameters cannot be zero, nor can there be relations using only negative coefficients, such as $s_1 = -s_3 - s_4\, s_9$. These facts simplify the algorithm. *Solve* succeeds in three minutes on a desktop computer.

Briefly, *Solve* works as follows:

Algorithm *Solve(f, x):* Given a polynomial f in a variable x and a number n of parameters s_i, find relations on the parameters that make the entire polynomial vanish. Our problem is solved by invoking $Solve(res, ca)$, $n = 11$. Outline:

- Kill each coefficient *coef* of x in turn, starting at the highest degree. Do so by looking for contents, linear parameters to solve for, or a difference of squares. When a substitution is found, plug it in, reducing the degree of f. Continue.
- Also try to kill the coefficient *coef* by invoking the entire algorithm on it, relative to each variable in *coef*. So, this step of *Solve* works by calling $Solve(coef, s_i)$ within a loop.
- Use suitable data structures to keep track of all the substitutions.

Here is a simple example. If res were

$$(s_9\, s_8 - 2s_7\, s_6)ca^2 + ({s_4}^2\, s_9 - 2{s_3}^2\, s_7)ca + s_8 - s_6$$

one solution would be the relations $s_9 = 2s_7, s_8 = s_6, s_4 = s_3$.

1.1 First new result

We have now analyzed Bricard's original formulation of the quadrilaterals problem [1] in terms of three equations, with fifteen parameters and three variables.

$$a_1\,t_1^2\,t_2^2 + b_1\,t_1^2 + 2c_1\,t_1\,t_2 + d_1\,t_2^2 + e_1 = 0,$$

$$a_2\,t_2^2\,t_3^2 + b_2\,t_2^2 + 2c_2\,t_2\,t_3 + d_2\,t_3^2 + e_2 = 0,$$

$$a_3\,t_1^2\,t_3^2 + b_3\,t_1^2 + 2c_3\,t_1\,t_3 + d_3\,t_3^2 + e_3 = 0$$

The t_i are the half-angle tangents of the three base angles α, β, γ. As before, these equations result from elementary analytic geometry. The parameters $a_i, b_i, c_i, \ldots$ are quadratic functions of the eleven sides. For example,

$$a_1 = e^2 + s_2^2 + s_7^2 - s_5^2 - 2e\,s_2 + 2e\,s_7 - 2s_2\,s_7$$

which is a product of two linear terms. This is the form of the equations as derived by Bricard.

The resultant of this system has 5685 terms. Shall we apply our flexibility searching algorithm as before? It is more subtle, as now we must try relations like $a_1 = 0$ or $a_1 = -d_3 - e_2$. When the parameters were actually the sides, substitutions like this made no sense and were excluded, thereby streamlining the search. We have modified algorithm *Solve* to include these cases, with

great success. Although the physically meaningful flexible conformations of the cyclohexane are well known ("chair" versus "boat"), this appears to be the first fully algebraic approach for their derivation, as well as for deriving Bricard's flexible octahedra. Moreover, the identical set of equations arises in other contexts, and a variant (which includes the "missing" terms, such as $t_1^2 t_2$, t_3, etc.) gives the conformational equations of a protein or nucleic acid backbone [3] [4].

1.2 Second new result

Next we consider the cylo-octane molecule, pictured in figure 2.

Chemically relevant solutions fix the (bond) angles between the paler lines, introducing four constraint equations in the variables τ_i. To save space, we show one equation here; the other three are similar.

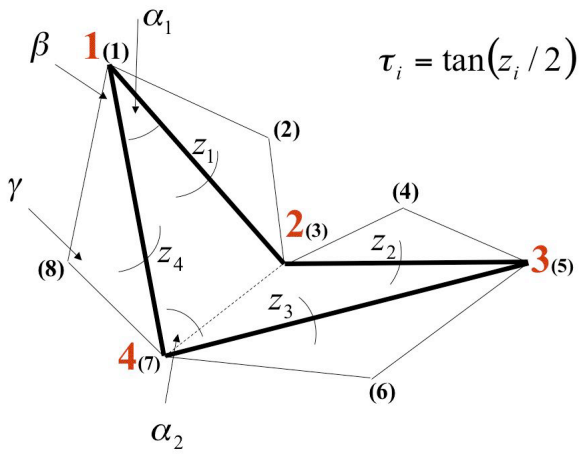

Fig. 2. Geometry of Octane Molecule.

$$-t\beta^4\,{\tau_4}^2\,{\tau_1}^2 - 4\,t\alpha_1\,t\beta^3\,{\tau_4}^2\,{\tau_1}^2 + 6\,t\beta^2\,{\tau_4}^2\,{\tau_1}^2 + 4\,t\alpha_1\,t\beta\,{\tau_4}^2\,{\tau_1}^2 - {\tau_4}^2\,{\tau_1}^2 - t\beta^4\,{\tau_1}^2 + 4\,t\alpha_1{}^2\,t\beta^2\,{\tau_1}^2 + 2\,t\beta^2\,{\tau_1}^2 - {\tau_1}^2 - 8\,t\alpha_1{}^2\,t\beta^2\,\tau_4\,\tau_1 - 8\,t\beta^2\,\tau_4\,\tau_1 - t\beta^4\,{\tau_4}^2 + 4\,t\alpha_1{}^2\,t\beta^2\,{\tau_4}^2 + 2\,t\beta^2\,{\tau_4}^2 - {\tau_4}^2 - t\beta^4 + 4\,t\alpha_1\,t\beta^3 + 6\,t\beta^2 - 4\,t\alpha_1\,t\beta - 1 = 0$$

Here $\tau_i = \tan(z_i/2)$, $t\beta = \tan(\beta/2)$, and $t\alpha_i = \tan(\alpha_i/2)$.

We use the Dixon resultant to eliminate τ_2, τ_3, and τ_4. An important **special case** is when the basic quadrilateral (heavy black lines) is planar. The equations then simplify quite a bit, and we can describe all the solutions of this case. The Dixon matrix is 24×24. 57% of the entries are 0. On average there are 41 terms per entry. It takes Dixon-EDF 3 minutes 38 seconds to compute the resultant for τ_1, which has 21715 terms. It is degree 32 in τ_1 but has only even degree terms.

In the **general case** (three dimensional space) we have also made significant progress. The Dixon matrix is 64×64. 64% of the entries are 0. On average there are 107 terms per entry. The determinant of the Dixon matrix here, were it ever computed, would have many billions of terms. But our Dixon-EDF techniques [9] discover its hundreds of factors in about 67 hours of CPU time. The largest has 4872161 terms. Using some of these factors, we have verified some known chemical arrangements. We seem to have found new interesting flexible cases. Work is ongoing.

References

1. Bricard, Raoul, Mémoire sur la théorie de l'octaèdre articulé, J. Math. Pures Appl. 3 (1897), p. 113 - 150
   (English translation: http://www.math.unm.edu/∼vageli/papers/bricard.pdf).
2. Cauchy, A. L. Sur les polygones et les polyhedres. Second Memoire. Journal de l'École Polytechn. **9** (1813), pp. 8.
3. Coutsias, E. A., C. Seok, M. P. Jacobson and K. A. Dill A Kinematic View of Loop Closure, J. Comput. Chem., 25 (2004), no. (4), p. 510 - 528.
4. Coutsias, E. A., C. Seok, M. J. Wester and K. A. Dill, Resultants and loop closure, Int. J. Quantum Chem. 106 (2005), no. (1), p. 176 - 189.
5. Cromwell, P. R. *Polyhedra*. New York: Cambridge Univ. Press, 1997. p. 222 - 224.
6. Cox, D., J. Little, D. O'Shea. Using Algebraic Geometry. Graduate Texts in Mathematics, 185. Springer-Verlag. New York, 1998.
7. Dixon, A. L. The eliminant of three quantics in two independent variables. Proc. London Math. Soc., **6** (1908) p. 468 - 478.
8. Lewis, R. H. and E. A. Coutsias, Algorithmic Search for Flexibility Using Resultants of Polynomial Systems; in Automated Deduction in Geometry, 6th International Workshop, ADG 2006. Springer-Verlag. LNCS **4869** p. 68 - 79 (2007).

9. Lewis, R. H., Heuristics to accelerate the Dixon resultant. Mathematics and Computers in Simulation **77**, Issue 4, p. 400-407, April 2008.
10. Thorpe, M., M. Lei, A. J. Rader, D. J. Jacobs, L. Kuhn. Protein flexibility and dynamics using constraint theory. J. Molecular Graphics and Modelling **19** (2001) p. 60 - 69.