

A Theorem about the Algebraic Structure underlying Orthogonal Graph Invariants

Oswaldo Araujo* and Daniel A. Morales

Departamentos de Matemática and Química, Facultad de Ciencias, Universidad de los Andes,
Mérida 5101, Venezuela

Received February 22, 1996[®]

In this work we prove a theorem which makes it possible to obtain a basis of graph theoretical invariants. The differences with Randić's method of orthogonalization are pointed out.

1. INTRODUCTION

It is well-known that graph theoretical invariants are important descriptors of molecular structure. However, different descriptors frequently represent, to a greater or lesser degree, the same aspect of molecular structure. Thus, in correlations constructed in the traditional way we will not be able to separate and identify different structural factors that contribute to the molecular property in question.

Randić introduced the concept of orthogonal molecular descriptors in order to solve this problem.^{1–4} Using the word orthogonal to mean the nonexistence of linear correlation, Randić established a criterion of orthogonality based on the lack of intercorrelation among the descriptors.

The above criterion suggests an analogy between the inner product of two vectors and the correlation coefficient. When the correlation coefficient is 1, then the two invariants are collinear and will represent the same aspect of molecular structure. When the correlation coefficient is zero, the invariants will represent different aspects of molecular structure. In this way a method that is analogous to the Gram–Schmidt orthogonalization process is obtained.

Further developments of the concept and importance of the orthogonalization of graph theoretical invariants has been given by Trinajstić and co-workers.^{5–7} These authors report on how structure–property and QSAR models can be improved using an orthogonalized basis of descriptors and on the importance of ordering the invariants in the orthogonalization procedure.

In this work we show that the underlying structure of some graph theoretical invariants, used to describe molecular structure, is the vector space $\mathbb{Q}(\sqrt{2}, \sqrt{3})$ over the field \mathbb{Q} of the rational numbers. On $\mathbb{Q}(\sqrt{2}, \sqrt{3})$ we define a symmetric bilinear form and then proceed to use the Gram–Schmidt orthogonalization process. In this way we formalize Randić's idea of orthogonalizing molecular descriptors which is based on residuals of stepwise multiple regression analysis; the results presented here make it possible to obtain a basis of graph theoretical invariants.

2. THE THEOREM

If \mathbb{Q} is the field of rational numbers, then we have

$$\mathbb{Q} \subset \mathbb{Q}(\sqrt{2}) \subset \mathbb{Q}(\sqrt{2}, \sqrt{3})$$

It is easy to see that $\{1, \sqrt{2}\}$ is a basis for the vector space $\mathbb{Q}(\sqrt{2})$ over \mathbb{Q} and that $\{1, \sqrt{3}\}$ is a basis for the vector space

$\mathbb{Q}(\sqrt{2}, \sqrt{3})$ over $\mathbb{Q}(\sqrt{2})$. Consequently, the dimension of the vector space $\mathbb{Q}(\sqrt{2}, \sqrt{3})$ over \mathbb{Q} is 4, and the set $\{1, \sqrt{2}, \sqrt{3}, \sqrt{6}\}$ is a basis for $\mathbb{Q}(\sqrt{2}, \sqrt{3})$ over \mathbb{Q} .⁸

The higher connectivity indices ${}^v\chi$, based on the connectivity index of Randić, are given by the following formula⁹

$${}^v\chi = \sum_{\text{paths}} \frac{1}{\sqrt{d_1 d_2 \dots d_{v+1}}} \quad (1)$$

where d_i is the degree of vertex i of the molecular graph.

Theorem. In a molecular graph, the connectivity indices ${}^v\chi$ ($v = 1, 2, \dots$) belong to the vector space $\mathbb{Q}(\sqrt{2}, \sqrt{3})$ over \mathbb{Q} .

Demonstration. First, we observe that any term in ${}^v\chi$ is of the form $q, q\sqrt{2}, q\sqrt{3}$, or $q\sqrt{6}$ with $q \in \mathbb{Q}$. In fact, since the degree of a carbon atom is 1, 2, 3, or 4, when we calculate any term of ${}^v\chi$ the only numbers that can appear are of the form $1/\sqrt{2^{e_1} 3^{e_2}}$, where e_1 and e_2 are non-negative integers. Therefore, if $e_1 \neq 0$ and $e_2 = 0$, $q = 1/\sqrt{2^{e_1}}$, then $q \in \mathbb{Q}$ or $q = p\sqrt{2}$ with $p \in \mathbb{Q}$. If $e_1 = 0$ and $e_2 \neq 0$ we have $q = 1/\sqrt{3^{e_2}}$; then $q \in \mathbb{Q}$ or $q = p\sqrt{3}$ with $p \in \mathbb{Q}$. Finally, if e_1 and e_2 both are different from zero we have $q = 1/\sqrt{2^{e_1} 3^{e_2}}$, then $q \in \mathbb{Q}$ or $q = p\sqrt{6}$ with $p \in \mathbb{Q}$.

On the other hand, the set $\{1, \sqrt{2}, \sqrt{3}, \sqrt{6}\}$ forms a basis for the \mathbb{Q} -vector space $\mathbb{Q}(\sqrt{2}, \sqrt{3})$. This concludes the proof.

Corollary. The connectivity indices ${}^v\chi$ ($v = 1, 2, \dots$) belong to the \mathbb{Q} -vector space $\mathbb{Q}(\sqrt{2})$ if and only if the chemical graph does not have tertiary carbon atoms.

3. GENERALIZATION OF THE GRAM–SCHMIDT ORTHOGONALIZATION PROCESS

We begin this section by recalling some definitions and facts about bilinear forms. We use here notation and terminology from Hoffman and Kunze.¹⁰

Let V be a vector space over the field K . Let $f: V \times V \rightarrow K$ be a bilinear form on the vector space V . We say that f is **symmetric** if $f(u, v) = f(v, u)$ for all vectors u, v in V . If f is any symmetric bilinear form we say that u and v are **orthogonal** with respect to f if $f(u, v) = 0$.

The following theorem is a generalization of the Gram–Schmidt orthogonalization process.

Let V be a finite-dimensional vector space over a field of characteristic zero, and let f be a symmetric bilinear form on V . Then there is an ordered basis for V in which f is represented by a diagonal matrix.

[®] Abstract published in *Advance ACS Abstracts*, October 1, 1996.

We will now apply this result to the vector space $Q(\sqrt{2}, \sqrt{3})$ over Q . Let $u = a_1 + a_2\sqrt{2} + a_3\sqrt{3} + a_4\sqrt{6}$ and $v = b_1 + b_2\sqrt{2} + b_3\sqrt{3} + b_4\sqrt{6}$ by any two elements of $Q(\sqrt{2}, \sqrt{3})$. Define $f: Q(\sqrt{2}, \sqrt{3}) \times Q(\sqrt{2}, \sqrt{3}) \rightarrow Q$ by

$$f(u, v) = a_1b_1 + 2a_2b_2 + 3a_3b_3 + 6a_4b_4 \quad (2)$$

It is easy to see that f is a symmetric bilinear form and if $u \neq 0$ then $f(u, u) \neq 0$ for any $u \in Q(\sqrt{2}, \sqrt{3})$.

We now will present two examples which illustrate the construction of orthogonal invariants using the method developed in this section.

Example 1. Consider normal alkanes. The connectivity indices ${}^1\chi$ and ${}^2\chi$ are given by

$${}^1\chi = \frac{(n-3)}{2} + \sqrt{2} \quad (3)$$

$${}^2\chi = \frac{(n-4)}{4}\sqrt{2} + 1 \quad (4)$$

where n is the number of vertices (carbon atoms) in the molecular graph.

The relation between ${}^1\chi$ and ${}^2\chi$ is given by

$${}^2\chi = ({}^1\chi/2 - 1)\sqrt{2} \quad (5)$$

On the other hand, since the set $\{{}^1\chi, {}^2\chi\}$ is linearly independent, it is a non-orthogonal basis for $Q(\sqrt{2})$. Using the Gram-Schmidt orthogonalization process we obtain

$${}^1\Omega = {}^1\chi \quad (6)$$

$${}^2\Omega = {}^2\chi - \frac{f({}^1\chi, {}^2\chi){}_1}{f({}^1\chi, {}^1\chi)}\chi \quad (7)$$

where

$$f({}^1\chi, {}^2\chi) = \frac{(2n-7)}{2} \quad (8)$$

and

$$f({}^1\chi, {}^1\chi) = \frac{(n-3)^2}{4} + 2 \quad (9)$$

Consequently,

$${}^2\Omega = -\frac{(n^2 - 7n + 4)}{(n^2 - 6n + 17)} + \frac{(n-3)(n^2 - 7n + 4)}{4(n^2 - 6n + 17)}\sqrt{2} \quad (10)$$

Thus, the set $\{{}^1\Omega, {}^2\Omega\}$ is an orthogonal basis for $Q(\sqrt{2})$.

Example 2. Consider the 2-methylalkanes with n carbon atoms. The following formulas give the first ${}^v\chi$ for those structures

$${}^1\chi = \frac{1}{2}(n-5) + \frac{1}{2}\sqrt{2} + \frac{2}{3}\sqrt{3} + \frac{1}{6}\sqrt{6} \quad (11)$$

$${}^2\chi = \frac{1}{2} + \frac{(n-6)}{4}\sqrt{2} + \frac{1}{2}\sqrt{3} + \frac{1}{3}\sqrt{6} \quad (12)$$

$${}^3\chi = \frac{(n-7)}{4} + \frac{1}{4}\sqrt{2} + \frac{1}{3}\sqrt{3} + \frac{1}{12}\sqrt{6} \quad (13)$$

$${}^4\chi = \frac{1}{4} + \frac{(n-8)}{8}\sqrt{2} + \frac{1}{12}\sqrt{3} + \frac{1}{6}\sqrt{6} \quad (14)$$

Since the set $\{{}^1\chi, {}^2\chi, {}^3\chi, {}^4\chi\}$ is linearly independent, it is a nonorthogonal basis for $Q(\sqrt{2}, \sqrt{3})$. Using the Gram-Schmidt orthogonalization process we obtain

$${}^1\Omega = {}^1\chi \quad (15)$$

$${}^2\Omega = {}^2\chi - \frac{f({}^2\chi, {}^1\chi){}_1}{f({}^1\chi, {}^1\chi)}\chi \quad (16)$$

$${}^3\Omega = {}^3\chi - \frac{f({}^3\chi, {}^2\chi){}_2}{f({}^2\chi, {}^2\chi)}\chi - \frac{f({}^3\chi, {}^1\chi){}_1}{f({}^1\chi, {}^1\chi)}\chi \quad (17)$$

$${}^4\Omega = {}^4\chi - \frac{f({}^4\chi, {}^3\chi){}_3}{f({}^3\chi, {}^3\chi)}\chi - \frac{f({}^4\chi, {}^2\chi){}_2}{f({}^2\chi, {}^2\chi)}\chi - \frac{f({}^4\chi, {}^1\chi){}_1}{f({}^1\chi, {}^1\chi)}\chi \quad (18)$$

where

$$f({}^2\chi, {}^1\chi) = \frac{n}{2} - \frac{17}{12} \quad (19)$$

$$f({}^3\chi, {}^2\chi) = \frac{n}{4} - \frac{23}{24} \quad (20)$$

$$f({}^3\chi, {}^1\chi) = 1 + \frac{(n-7)(n-5)}{8} \quad (21)$$

$$f({}^3\chi, {}^3\chi) = \frac{1}{2} + \frac{(n-7)^2}{16} \quad (22)$$

$$f({}^4\chi, {}^1\chi) = \frac{n}{4} - \frac{31}{24} \quad (1/4 \text{ for } n=7) \quad (23)$$

$$f({}^4\chi, {}^1\chi) = 2 + \frac{(n-5)^2}{4} \quad (24)$$

$$f({}^2\chi, {}^2\chi) = \frac{5}{3} + \frac{(n-6)^2}{8} \quad (n \geq 8) \quad (25)$$

$$f({}^4\chi, {}^3\chi) = \frac{n}{8} - \frac{37}{48} \quad (26)$$

$$f({}^4\chi, {}^2\chi) = \frac{7}{12} + \frac{(n-8)(n-6)}{16} \quad (n \geq 8) \quad (27)$$

Then $\{{}^1\Omega, {}^2\Omega, {}^3\Omega, {}^4\Omega\}$ is an orthogonal basis for $Q(\sqrt{2}, \sqrt{3})$.

4. CONCLUSIONS AND COMMENTS

We have proved a theorem which states that the connectivity indices belong to the vector space $Q(\sqrt{2}, \sqrt{3})$ over Q . This allows us to formalize Randić's idea of orthogonal molecular descriptors.

We note here that there is an important difference between the orthogonalization procedure outlined by Randić, which is based on regression analysis of the intercorrelation of basis invariants, and our approach. Since we consider each molecular structure to be a vector, independent of other structures, we orthogonalize components of a single structure. In contrast, Randić has to consider a set of structures and then orthogonalize all the structures simultaneously. If an additional structure is added to the set considered, the whole process of orthogonalization has to be repeated. In our approach vectors already orthogonalized are not affected when additional structures are added to the set. For example, if in example 1 we orthogonalize ${}^1\chi$ and ${}^2\chi$ using Randić's method we would have ${}^1\Omega = {}^1\chi$ and ${}^2\Omega = {}^2\chi - (a + b){}^1\chi$ where a and b are the coefficients which come from the linear

relationship between ${}^1\chi$ and ${}^2\chi$. Those coefficients are obtained by Randić using linear regression analysis, however, from eq 5 we have those coefficients directly. Thus, in this case ${}^2\Omega = 0$.

It is then seen that while in Randić's procedure the linear relation between two connectivity indices is established by linear regression, which involves the consideration of many molecules of different groups. In our approach we orthogonalize molecules of one type only, for example, n -alkanes, 2-methylalkanes, etc.

To orthogonalize the connectivity indices for the heptane isomers ($n = 7$) it would be necessary to orthogonalize the indices for each molecule, taken one at a time, while in Randić's procedure all molecules (nine isomers) have to be considered in the orthogonalization process.

Even though in this work we have concentrated on the connectivity indices we note that the indices of Wiener, Hosoya, and molecular identification (ID) also belong to the vector space $Q(\sqrt{2}, \sqrt{3})$ over Q . In the case of normal alkanes the ID index is an element of the vector subspace $Q(\sqrt{2})$ over Q . Trinajstić et al.¹¹ showed that the prime ID number belongs to the vector space $L = Q(\sqrt{2}, \sqrt{3}, \sqrt{5}, \sqrt{7}, \sqrt{11}, \sqrt{13}, \sqrt{17}, \sqrt{19}, \sqrt{23})$ over Q and whose dimension is 2^9 , since in the relation $Q \subset Q(\sqrt{2}) \subset Q(\sqrt{2}, \sqrt{3}) \subset \dots \subset L$ each extension is of dimension 2.

ACKNOWLEDGMENT

This work was partially supported by the CDCHT of La Universidad de los Andes (Mérida, Venezuela). O.A. thanks

the organizers of the symposium Frontiers in Mathematical Chemistry (Pacifichem 95), especially Professor M. Randić for the invitation to participate as invited speaker.

REFERENCES AND NOTES

- (1) Randić, M. Orthogonal molecular descriptors. *New J. Chem.* **1991**, 15, 517–525.
- (2) Randić, M. Search for optimal molecular descriptors. *Croat Chem. Acta* **1991**, 64, 43–54.
- (3) Randić, M. Generalized molecular descriptors. *J. Math. Chem.* **1991**, 7, 155–168.
- (4) Randić, M. Correlation of enthalpy of octanes with orthogonal connectivity indices. *J. Mol. Struct. (Theochem)* **1991**, 233, 45–59.
- (5) Lučić, B.; Nikolić, S.; Trinajstić, N.; Juretić, D. The Structure-Property Models can be Improved using the Orthogonalized Descriptors. *J. Chem. Inf. Comput. Sci.* **1995**, 35, 532–538.
- (6) Amić, D.; Davidović-Amić, D.; Trinajstić, N. Calculation of Retention Times of Anthocyanins with Orthogonalized Topological Indices. *J. Chem. Inf. Comput. Sci.* **1995**, 35, 136–139.
- (7) Šoškić, M.; Plavšić, D.; Trinajstić, N. 2-Difluoromethylthio-4,6-bis-(monoalkylamino)-1,3,5-triazines a Inhibitor of Hill Reaction: A QSAR Study with Orthogonalized Descriptors. *J. Chem. Inf. Comput. Sci.* **1996**, 36, 146–150.
- (8) Stewart, I.; **Galois Theory**; Chapman and Hall: New York, 1989.
- (9) Trinajstić, N.; **Chemical Graph Theory**; CRC Press: Boca Raton, FL, 1992.
- (10) Hoffman, K.; Kunze, R.; **Linear Algebra**; Prentice-Hall: NJ, 1971.
- (11) Szymanski, K.; Müller, W. R.; Knop, J. V.; Trinajstić, N. Molecular ID numbers. *Croat. Chem. Acta* **1986**, 59, 719–723.

CI960023Q