On the structure-dependence of the largest eigenvalue of the distance matrix of an alkane

Ivan Gutman
Faculty of Science, University of Kragujevac, P.O. Box 60, YU -34000 Kragujevac, Yugoslavia

and

Mihai Medeleanu
Department of Organic Chemistry, Polytechnic Institute, Timisoara, Romania

Received 8 August 1997

The distance matrix of an alkane has a unique positive eigenvalue $\lambda_1$, which has recently been applied as a topological index and used for ordering alkane isomers with respect to their branching and for calculating their boiling points. We have now determined the main structural features of the alkane molecule, on which $\lambda_1$ depends. These are the number of vertices of the molecular graph ($n$) and the sum of the squares of the distances between all pairs of vertices ($S$). A good linear correlation between $\lambda_1$ and $(nS)^{1/2}$ is shown to exist. Lower and upper bounds of $\lambda_1$ are deduced, both depending solely on $n$ and $S$. Because $S = 2WW - W$, there exists a connection between $\lambda_1$, the Wiener index ($W$) and the hyper-Wiener index ($WW$).

In 1973 Lovasz and Pelikan proved that among trees with a fixed number of vertices, the largest eigenvalue of the adjacency matrix is minimal for the path-graph and maximal for the star-graph. This finding implies that the largest eigenvalue of the adjacency matrix of the molecular graph of an alkane reflects the extent of branching of the molecular carbon-atom skeleton. Consequently, this eigenvalue has been often considered and applied as a structure-descriptor and topological index. The theory of the largest eigenvalue of the adjacency matrix is nowadays well elaborated; for details and further references see the review in ref. 6.

It has been recently proposed to use the largest eigenvalue $\Lambda$ of the distance matrix as another topological index sensitive to molecular branching. It has been shown that $\Lambda$ can be successfully used to infer about the extent of branching of alkanes. The correlation between $\Lambda$ and the boiling points (BP) of alkanes was investigated in detail and it was found that the expressions,

$$BP = 56.518 \ln \left(\frac{n\Lambda}{10}\right) - 36.0298$$

and

$$BP = 18.593 \left(\frac{\Lambda}{n}\right) + 24.150n - 123.99$$

reproduce the boiling points (at normal pressure) with standard errors of only $8.9^\circ C$ and $7.1^\circ C$, and have correlation coefficients as high as 0.981 and 0.989, respectively. A still better (hitherto unpublished) expression of the same kind is:

$$BP = 14.37 \sqrt{\Lambda} + 107.28 \sqrt{n} - 251.76$$

for which the standard error is $5.8^\circ C$ and the correlation coefficient 0.992.

The aim of this paper is to shed some light on the dependence of $\Lambda$ on molecular structure. Our analysis is much facilitated by means of the following algebraic properties of the eigenvalues of the distance matrix.

Let $T$ be a tree (= a connected acyclic graph) on $n$ vertices, $n > 1$ and let $D = \|d_{ij}\|$ be its distance matrix. Let $\Lambda_1, \Lambda_2, \ldots, \Lambda_n$ be the eigenvalues of $D$. Then

$$\Lambda_1 > 0 \quad \& \quad \Lambda_i < 0 \quad \text{for } i = 2, \ldots, n \quad \ldots(1)$$

$$\det D = \Lambda_1 \cdot \Lambda_2 \cdot \cdots \cdot \Lambda_n = (-1)^{n-1}(n - 1)2^{n-2} \ldots(2)$$
In other words, \( \Lambda \) is the unique positive eigenvalue of the distance matrix, all other eigenvalues being negative-valued. Further, the determinant of the distance matrix depends only on the number of vertices, and is fully independent of any other structural feature of the underlying graph. In what follows we shall need also the following two elementary properties of the eigenvalues of the distance matrix. Because its diagonal elements are zero, one has

\[
\Lambda_1 + \Lambda_2 + \cdots + \Lambda_n = 0 \quad \cdots (3)
\]

According to a well-known property of matrix eigenvalues,

\[
\Lambda_1^2 + \Lambda_2^2 + \cdots + \Lambda_n^2 = \sum_{i=1}^{n} \sum_{j=1}^{n} d_{ij} = 2 \sum_{r<s}^n d_{rs}
\]

which by introducing the function \( S \),

\[
S = \sum_{r<s}^n d_{rs}^2
\]

can be written as

\[
\Lambda_1^2 + \Lambda_2^2 + \cdots + \Lambda_n^2 = 2S \quad \cdots (5)
\]

Note that \( S \) is just the sum of the squares of the distances between all pairs of vertices of the molecular graph.

Some auxiliary identities

In what follows, we denote the absolute value of the \( i \)-th eigenvalue of the distance matrix by \( \lambda_i \). Because of Eq. (1) all the numbers \( \lambda_1, \lambda_2, \ldots, \lambda_n \) are positive-valued (i.e none of them is equal to zero), and, in particular,

\[
\lambda_1 = \Lambda \quad \& \quad \lambda_i = -\Lambda_i \quad \text{for} \quad i = 2, \ldots, n
\]

Then Eq. (5) can be rewritten as

\[
\sum_{i=1}^{n} \lambda_i^2 = 2S \quad \cdots (6)
\]

whereas from Eq. (3) we get

\[
\Lambda = -\sum_{i=2}^{n} \Lambda_i = \sum_{i=2}^{n} \lambda_i = \sum_{i=1}^{n} \lambda_i - \Lambda
\]

and thus

\[
\Lambda = \frac{1}{2} \sum_{i=1}^{n} \lambda_i \quad \cdots (7)
\]

Consider now an auxiliary quantity \( Q \), defined as

\[
Q = \sum_{i<j}^{n} (\lambda_i - \lambda_j)^2 \quad \cdots (8)
\]

Taking into account identities (6) and (7), by direct expansion of the right-hand side of Eq. (8) we arrive at the relation,

\[
Q = 2nS - 4\Lambda^2 \quad \cdots (9)
\]

From (1) it immediately follows that \( Q = 0 \) for \( n = 2 \), and that \( Q > 0 \) for \( n \geq 3 \). Thus, for alkanes with at least three carbon atoms,

\[
2nS - 4\Lambda^2 > 0
\]

i.e.

\[
\Lambda < \sqrt{\frac{1}{2} n S} \quad \cdots (10)
\]

The simple upper bound (10) for the topological index \( \Lambda \) suggests that a linear correlation between \( \Lambda \) and the expression \((nS)^{1/2}\) might exist. This hypothesis is tested in a subsequent section.

The bound (10) can be somewhat improved using Kober's estimate of the difference between the arithmetic and geometric means. By this we are able to obtain both an upper and a lower bound for \( \Lambda \).

Bounds for the topological index \( \Lambda \)

Consider positive-valued real numbers \( \mu_1, \mu_2, \ldots, \mu_n \) whose arithmetic and geometric means are \( a(\mu) \) and \( g(\mu) \), respectively:

\[
a(\mu) = \frac{1}{n} \sum_{i=1}^{n} \mu_i \quad ; \quad g(\mu) = \left( \prod_{i=1}^{n} \mu_i \right)^{1/n}
\]

It is well-known that the arithmetic mean exceeds the geometric mean,

\[
a(\mu) \geq g(\mu) \quad \text{i.e.} \quad a(\mu) - g(\mu) \geq 0
\]
with equality if and only if $\mu_i = \mu_j = \ldots = \mu_n$. Kober$^{13}$ established the following remarkable improvement of the above relation:

$$\frac{1}{n(n-1)} \sum_{i<j} (\sqrt{\mu_i} - \sqrt{\mu_j})^2 \leq a(\mu) - g(\mu) \leq \frac{1}{n} \sum_{i<j} (\sqrt{\mu_i} - \sqrt{\mu_j})^2. \quad (11)$$

Now, because formula (11) is satisfied by any positive-valued number, it will hold also if we set $\mu_i = \lambda^2$, for $i = 1, 2, \ldots, n$. For this choice, by taking into account Eqs (8) and (9), we get

$$\sum_{i<j} (\sqrt{\mu_i} - \sqrt{\mu_j})^2 \equiv Q = 2nS - 4\Lambda^2$$

whereas from Eqs (6) and (2) follow

$$a(\mu) \equiv \frac{2S}{n}$$

and

$$g(\mu) \equiv (|\text{det} D|)^{2/n} = (n-1)^{2(n-2)/n}$$

respectively.

Substituting the above relations back into Eq. (11) we obtain after a lengthy, yet elementary calculation:

$$\sqrt{\frac{1}{2} S + n(n-1) \left( \frac{n-1}{4} \right)^{2/n} } < \Lambda < \sqrt{\frac{n-1}{2} S + n \left( \frac{n-1}{4} \right)^{2/n} } \quad \ldots (12)$$

The bounds (12) hold for alkanes with at least three carbons atoms. They are necessarily strict (i.e., equality can never occur). Their quality is seen from the data given in Table 1.

The data given in Table 1 show that the lower bound is much closer to the true value of $\Lambda$ than the upper bound. Both bounds exhibit an excellent linear correlation with $\Lambda$. For instance, in the class of $C_9$-alkanes the approximations

$$\Lambda = 0.7995 + 0.8681 \text{ (lower bound)}$$

$$\Lambda = 4.0541 + 0.3168 \text{ (upper bound)}$$

Table 1—Exact value, lower bound and upper bound for the topological index $\Lambda$ of some alkanes

<table>
<thead>
<tr>
<th>Molecule*</th>
<th>Exact $\Lambda$</th>
<th>Lower bound</th>
<th>Upper bound</th>
</tr>
</thead>
<tbody>
<tr>
<td>C3</td>
<td>3.1273</td>
<td>2.2731</td>
<td>3.7269</td>
</tr>
<tr>
<td>C4</td>
<td>5.5129</td>
<td>5.1623</td>
<td>7.9664</td>
</tr>
<tr>
<td>C5</td>
<td>8.3666</td>
<td>8.2882</td>
<td>14.3178</td>
</tr>
<tr>
<td>C6</td>
<td>12.1093</td>
<td>11.7182</td>
<td>23.0535</td>
</tr>
<tr>
<td>22MM-C5</td>
<td>13.0698</td>
<td>12.4563</td>
<td>25.6098</td>
</tr>
<tr>
<td>24MM-C5</td>
<td>14.1768</td>
<td>13.4595</td>
<td>28.4932</td>
</tr>
<tr>
<td>2233MMMMM-C4</td>
<td>14.9373</td>
<td>14.1566</td>
<td>31.0032</td>
</tr>
<tr>
<td>2233MMM-C5</td>
<td>16.0683</td>
<td>14.9803</td>
<td>33.6036</td>
</tr>
<tr>
<td>2233MMM-C5</td>
<td>16.3152</td>
<td>15.2121</td>
<td>34.3249</td>
</tr>
<tr>
<td>C7</td>
<td>16.6254</td>
<td>15.5935</td>
<td>34.4073</td>
</tr>
<tr>
<td>3E-3M-C5</td>
<td>16.6705</td>
<td>15.3756</td>
<td>34.8310</td>
</tr>
<tr>
<td>2244MMM-C5</td>
<td>17.0338</td>
<td>15.8874</td>
<td>36.4033</td>
</tr>
<tr>
<td>3E-2M-C5</td>
<td>17.4187</td>
<td>15.9815</td>
<td>36.6906</td>
</tr>
<tr>
<td>333M-C6</td>
<td>17.4426</td>
<td>16.1062</td>
<td>37.0702</td>
</tr>
<tr>
<td>344M-C6</td>
<td>17.6759</td>
<td>16.2607</td>
<td>37.5393</td>
</tr>
<tr>
<td>222M-C6</td>
<td>18.1815</td>
<td>16.7454</td>
<td>39.0026</td>
</tr>
<tr>
<td>244M-C6</td>
<td>18.3964</td>
<td>16.9531</td>
<td>39.6258</td>
</tr>
<tr>
<td>222M-C6</td>
<td>18.4133</td>
<td>17.0707</td>
<td>39.9775</td>
</tr>
<tr>
<td>3E-C6</td>
<td>18.7788</td>
<td>17.1000</td>
<td>40.0650</td>
</tr>
<tr>
<td>2233MMMMM-C5</td>
<td>18.8440</td>
<td>17.3208</td>
<td>41.6953</td>
</tr>
<tr>
<td>25M-C6</td>
<td>19.1115</td>
<td>17.6751</td>
<td>41.7756</td>
</tr>
<tr>
<td>4M-C7</td>
<td>19.5420</td>
<td>17.8720</td>
<td>42.3580</td>
</tr>
<tr>
<td>3M-C7</td>
<td>19.7628</td>
<td>18.1221</td>
<td>43.0953</td>
</tr>
<tr>
<td>2M-C7</td>
<td>20.4792</td>
<td>18.8523</td>
<td>45.2350</td>
</tr>
<tr>
<td>2244MMM-C6</td>
<td>21.1839</td>
<td>19.2351</td>
<td>47.9427</td>
</tr>
<tr>
<td>C8</td>
<td>21.8364</td>
<td>20.0102</td>
<td>48.5922</td>
</tr>
<tr>
<td>22344MMM-C5</td>
<td>22.8477</td>
<td>20.6603</td>
<td>53.8587</td>
</tr>
<tr>
<td>345M-C7</td>
<td>25.9693</td>
<td>23.0401</td>
<td>61.9416</td>
</tr>
<tr>
<td>C9</td>
<td>27.7422</td>
<td>24.9798</td>
<td>65.8065</td>
</tr>
<tr>
<td>C10</td>
<td>34.3429</td>
<td>30.5098</td>
<td>86.2367</td>
</tr>
<tr>
<td>C11</td>
<td>41.6384</td>
<td>36.6052</td>
<td>110.0590</td>
</tr>
<tr>
<td>C12</td>
<td>49.6287</td>
<td>43.2698</td>
<td>137.4416</td>
</tr>
<tr>
<td>C13</td>
<td>58.3189</td>
<td>50.5047</td>
<td>168.5449</td>
</tr>
<tr>
<td>C14</td>
<td>67.6939</td>
<td>58.3127</td>
<td>203.5229</td>
</tr>
</tbody>
</table>

*for instance: C8 = n-octane, 2233MMM-C5 = 2,3,3-trimethylpentane, 3E-3M-C5 = 3-ethyl-3-methylpentane, etc.
are obeyed, with correlation coefficients of 0.9975 and 0.9999, respectively. For other sets of alkanes, equations with very similar slopes and comparable accuracy are obtained.

Observe that both the lower and the upper bounds depend only on the simple topological parameters $n$ and $S$. The former reflects the size-dependency of $A$, and the latter its branching-dependency. Whatever the actual mathematical form of the size-dependency of $A$ might be, the inequalities (12) indicate that it is not a simple function. On the other hand, inequalities (12) strongly suggest that $A$ is linearly proportional to the square root of $S$. Besides $n$ and $S$ seem to be the main factors determining the numerical value of $A$. These conclusions are supported by the numerical studies outlined in the subsequent section.

An approximate formula for the topological index $A$

Already the upper and/or lower bound (12) can be used to approximate the topological index $A$, namely, by means of formulae of the type (13) and (14). However, the algebraic expression occurring in (12) are too complicated to be attractive for practical usage. We found that a much simpler formula, namely $\sqrt{nS}$, may serve the same purpose. The choice of $\sqrt{nS}$ has, of course, been motivated by the inequal-

Numerical testing revealed that a remarkably good linear correlation exists between $A$ and $\sqrt{nS}$. This is illustrated in Figs 1 & 2.

Figure 1 shows the relation between $A$ and $\sqrt{nS}$ for the set of all alkanes with $n \leq 14$ (3323 molecules). By least-squares fitting we arrive at the formula:

$$A \approx A \sqrt{nS} + B$$

$$A = 0.2027 \pm 0.0002;$$

$$B = 7.339 \pm 0.040 ... \ (15)$$

which reproduces $A$ with a standard deviation of only $s = 0.49$, having a correlation coefficient $R = 0.9981$. Because in the case of the lower alkanes, $n \leq 7$, a marked deviation from linearity is observed (see Fig. 1), we examined separately the $C_8 - C_{14}$-alkanes (3302 molecules), in which case — of course — the quality of the approximation (15) is even better: $s = 0.40, R = 0.9986, A = 0.2006 \pm 0.002, B = 7.753 \pm 0.034$.

As could be anticipated from the discussion in the preceding section, in sets of isomeric alkanes ($n = \text{const}$) the linearity between $A$ and $\sqrt{nS}$ is almost perfect. Thus for the $C_{14}$-set (1858 molecules) we get:

$$A \approx (0.2007 \pm 0.003) \sqrt{nS} + (7.606 \pm 0.069)$$

with $s = 0.34$ and $R = 0.997$. Similarly, for the $C_{13}$-set
(802 molecules) we get (see Fig. 2):
\[ \Lambda \approx (0.2073 \pm 0.005) \sqrt{n S} + (6.799 \pm 0.092) \]
with \( S = 0.29 \) and \( R = 0.997 \).

Except for the few lower alkanes, Eq. (15) and its refined versions are capable of reproducing \( \Lambda \) with an error of about 1\% or less. This is a fairly sufficient accuracy for any presently imaginable chemical application of this topological index.

**Discussion**

At this point we wish to emphasize that the parameter \( S \), playing an important role in the present analysis, is closely related to two structure-descriptors that have been extensively investigated in the chemical literature. These are the classical Wiener index (introduced in 1947)\(^{14}\) and its extension—the hyper-Wiener index (conceived in 1993)\(^{15}\). The Wiener index \( W \) is just the sum of the distances between all pairs of vertices (for review see ref. 8):

\[ W = \sum_{r<s} d_{rs} . \]

The definition of the hyper-Wiener index \( WW \) is somewhat less straightforward (for details see ref 15-17). It has been shown\(^{16}\) that the hyper-Wiener index of alkanes satisfies the identity:

\[ WW = \frac{1}{2} \sum_{r<s} [d_{rs}^2 + d_{rs}] . \]

Bearing the above in mind and taking into account Eq. (4) we readily conclude that

\[ S = 2 WW - W . \] \ ...(16)

Equation: (16) implies that any result on \( \Lambda \) which has been stated above in terms of the quality \( S \) can be reformulated as a result relating \( \Lambda \) with the well-known Wiener and hyper-Wiener topological indices. Thus our analysis revealed the existence of a non-trivial and previously not anticipated connection between the topological indices \( \Lambda, W \) and \( WW \).

**Acknowledgement**

The authors thanks Dr. V. Gogonea (University of Erlangen, Germany) for a computer program for generating trees.

**References**