Group Additivity Methods without Group Values

Ilie Fishtik*

Department of Chemical Engineering, Worcester Polytechnic Institute, Worcester, Massachusetts 01609-2280

Received: August 2, 2006; In Final Form: October 6, 2006

The conventional group additivity (GA) formalism may be identically reduced to a stoichiometric and thermochemical analysis of a special class of reactions referred to as GA reactions, that is, reactions that preserve the type and number of groups. Within this approach, the performance (error) of a GA scheme is determined by the stoichiometry and enthalpy changes of the GA reactions. That is, the lower the enthalpy changes of the GA reactions, the better the performance of a GA scheme. Ideally, an exact GA scheme would imply any conceivable GA reaction to be precisely thermoneutral, that is, have a zero enthalpy change. A somewhat surprising result is that, additionally, the performance of GA methods is influenced by a purely stoichiometric factor of GA reactions. These findings do not improve the performance of a given GA scheme. Rather, it is an interpretation that leads to a deeper understanding of the performance of a GA scheme and may be used in designing more accurate GA schemes.

1. Introduction

Group additivity (GA) methods have been used successfully for more than half a century to estimate thermochemical data for species for which experimental data are unavailable.[1−5] Although ab initio methods are becoming increasingly available, the GA methods are still competitive.[6−15] However, there is one aspect of the GA methods that, in our opinion, was overlooked and is causing some confusion. The point is that in many cases the groups happen to be linearly dependent. In such cases, one has to assign arbitrary group values for linearly dependent groups. The numerical arbitrariness of the group values does not cause any mathematical problems. Rather, it diminishes the interpretation capability of GA methods. Thus, recently, Gronert[6] proposed a new group additivity scheme for alkanes, alkenes, and alkyl radicals and used it to rationalize several fine interactions in organic species such as hyperconjugation and the branching effect. Although the model performance for enthalpies of formation is excellent, Wodrich and Schleyer[7] pointed out that *no significance can be attached to the group values obtained from empirical fitting schemes*. Moreover, they showed that when applied to a subset of the species in the Gronert data set, arbitrarily fixing some of the group values, or even removing some of them, results in *mathematically equivalent* models. These results were obtained based on a detailed analysis of various models. A more careful analysis of Gronert's GA scheme reveals, however, that the groups are linearly dependent when applied to this subset of the data and the mathematical equivalence of various models is not surprising at all.

On the basis of the above, it is of interest to consider the significance of the numerical arbitrariness of the group values in GA methods. More specifically, we present a general proof of the independence of the performance of GA methods on the group values as well as an interpretation of this phenomenon. Our approach is based on the analogy between the GA and chemical reaction stoichiometry that has been pointed out by us recently.[16−18] This new interpretation not only proves the independence of the GA methods on the group values and, hence, the physicochemical meaningless of the group values but also provides important insight into the performance of GA methods in general.

2. Theory

Recently, it has been shown that the ordinary least-squares (OLS) method may be modified so as to minimize the residuals (errors) subject to a set of explicit linear relations among residuals.[16] This result reveals an interesting analogy with chemical reaction stoichiometry as well as a new interpretation of the OLS. In particular, the residuals may be explicitly related to various characteristics of special classes of reactions. Below, we present a succinct summary of this approach as applied to GA methods.

Consider a training set of n chemical species B_1, B_2, ..., B_n. Let $\Delta H_{f,i}^{\mathrm{exp}}$ ($i = 1,2,...,n$) be the experimental enthalpies of formation and $\Delta H_{f,i}^{\mathrm{calc}}$ ($i = 1,2,...,n$) the calculated (via the group additvity) enthalpies of formation of the species. It is, thus, convenient to introduce the vectors:

$$\mathbf{B} = (B_1,B_2,...,B_n)^{\mathrm{T}} \tag{1}$$

$$\Delta\mathbf{H}_{\mathbf{f}}^{\mathbf{exp}} = (\Delta H_{f,1}^{\mathrm{exp}},\Delta H_{f,2}^{\mathrm{exp}},...,\Delta H_{f,n}^{\mathrm{exp}})^{\mathrm{T}} \tag{2}$$

$$\Delta\mathbf{H}_{\mathbf{f}}^{\mathbf{calc}} = (\Delta H_{f,1}^{\mathrm{calc}},\Delta H_{f,2}^{\mathrm{calc}},...,\Delta H_{f,14}^{\mathrm{calc}})^{\mathrm{T}} \tag{3}$$

Let further **e** be the error (residual) vector

$$\mathbf{e} = \Delta\mathbf{H}_{\mathbf{f}}^{\mathbf{exp}} - \Delta\mathbf{H}_{\mathbf{f}}^{\mathbf{calc}} = (\Delta H_{f,1}^{\mathrm{exp}} - \Delta H_{f,1}^{\mathrm{calc}}, \Delta H_{f,2}^{\mathrm{exp}} - \Delta H_{f,2}^{\mathrm{calc}},..., \Delta H_{f,14}^{\mathrm{exp}} - \Delta H_{f,14}^{\mathrm{calc}})^{\mathrm{T}} \tag{4}$$

Next, instead of introducing the group value increments and looking for their best fitting we define the so-called GA reactions,[17,18] that is, reactions that conserve the number and type of groups. These reactions are generated based on the group matrix. Mathematical details of the generation of GA reactions may be found in the Appendix as well as in refs 17 and 18. If

---

* E-mail: ifishtik@wpi.edu.

10.1021/jp0649753 CCC: $33.50

the rank of the group matrix is equal to q, then the number of linearly independent GA reactions is equal to $m = n - q$. Let an arbitrary set of m linearly independent GA reactions be

$$\boldsymbol{\rho}: \ \boldsymbol{\nu}\mathbf{B} = \mathbf{0} \tag{5}$$

where

$$\boldsymbol{\rho} = (\rho_1,\rho_2,...,\rho_m)^{\mathrm{T}} \tag{6}$$

$$\boldsymbol{\nu} = \begin{pmatrix} \nu_{11} & \nu_{12} & ... & \nu_{1n} \\ \nu_{21} & \nu_{22} & ... & \nu_{2n} \\ ... & ... & ... & ... \\ \nu_{m1} & \nu_{m2} & ... & \nu_{mn} \end{pmatrix} \tag{7}$$

and $\boldsymbol{\nu}$ is the stoichiometric matrix. Let further $\Delta\mathbf{H}_\rho$ be the vector of the enthalpy changes of the GA reactions $\boldsymbol{\rho}$:

$$\Delta\mathbf{H}_\rho = (\Delta H_1, \Delta H_2,...,\Delta H_m)^{\mathrm{T}} \tag{8}$$

These may be calculated from the experimental enthalpies of formation via[19]

$$\Delta\mathbf{H}_\rho = \boldsymbol{\nu}\ \Delta\mathbf{H}_{\mathrm{f}}^{\mathbf{exp}} \tag{9}$$

The GA reactions have an important property, namely, their calculated enthalpy changes are equal to zero,[18] that is

$$\boldsymbol{\nu}\Delta\mathbf{H}_{\mathbf{f}}^{\mathbf{calc}} = \mathbf{0} \tag{10}$$

Subtracting eq 10 from eq 8 and taking into account eq 4 we obtain

$$\boldsymbol{\nu}\mathbf{e} = \Delta\mathbf{H}_\rho \tag{11}$$

Now, the error vector $\mathbf{e}$ may be evaluated by minimizing the product $\mathbf{e}^{\mathbf{T}}\mathbf{e}$ subject to the constraint given by eq 11. The result is[17]

$$\mathbf{e} = \nu^{\mathbf{T}}(\nu\nu^{\mathbf{T}})^{-\mathbf{1}}\Delta\mathbf{H}_\rho \tag{12}$$

As can be seen, the error vector may be obtained without group increment values!

Within this approach, the enthalpy of formation of a new species, for example, B_{n+1}, is also evaluated based on stoichiometric considerations. The procedure is as follows. From the error vector, the calculated enthalpies of formation of the training set of species may be evaluated

$$\Delta\mathbf{H}_{\mathbf{f}}^{\mathbf{calc}} = \Delta\mathbf{H}_{\mathbf{f}}^{\mathbf{exp}} - \mathbf{e} \tag{15}$$

Let an arbitrary GA reaction involving species B_1, B_2, ..., B_n, B_{n+1} be

$$\sum_{i=1}^{b}\nu_i B_i + \nu_{i+1}B_{i+1} = 0 \tag{16}$$

Because the enthalpy change of a GA reaction expressed via the calculated enthalpies of formation should be equal to zero, that is

$$\sum_{i=1}^{n}\nu_i\ \Delta H_{f,i}^{\mathrm{calc}} + \nu_{i+1}\Delta H_{f,i+1}^{\mathrm{calc}} = 0 \tag{17}$$

we have

$$\Delta H_{f,i+1}^{\mathrm{calc}} = -\frac{1}{\nu_{i+1}}\sum_{i=1}^{n}\nu_i\ \Delta H_{f,i}^{\mathrm{calc}} \tag{18}$$

This result is independent of the choice of the GA reaction.

3. An Example: Gronert's Scheme[6]

For illustration purposes, we consider only the training subset of $n = 14$ species discussed by Gronert.[6] These species are given in Table 1. Gronert's scheme involves seven types of groups, namely, the number of carbon and hydrogen atoms, C−C, and C−H bonds as well as three molecular fragments (interactions), HCH, CCH, and CCC. The respective group matrix is given in Table 2. Its rank is equal to $q = 5$ and, hence, only five out of seven groups are linearly independent. It means that for this data any two groups may be assigned an arbitrary value or, alternatively, could be dropped without affecting the model performance. However, two points should be noted with respect to the Gronert model. First, the full set of species in the Gronert paper[6] gives a group matrix with a rank of $q = 6$. Second, Gronert eliminated two of the group values from the analysis by assigning them values equal to the heats of formation of gaseous carbon (170.6 kcal/mol) and hydrogen atoms (52.1 kcal/mol) while introducing a new parameter EC viewed as a correction term for electron pairing in atomic carbon. Thus, overall Gronert's model is in fact a six-parameter model.

According to the GA reaction approach, the first step in the evaluation of the GA error is the generation of a set of linearly independent GA reactions. This set involving $m = n - q = 14 - 5 = 9$ linearly independent GA reactions is $14 - 5 = 9$. We select these as (see the Appendix for the mathematical details)

$$\rho_1\text{:}\ \ -B_2 + 2B_3 - B_4 = 0 \quad \Delta H_1 = 0.4\ \text{kcal/mol}$$

$$\rho_2\text{:}\ \ -B_1 + 3B_2 - 3B_3 + B_5 = 0 \quad \Delta H_2 = 0.8\ \text{kcal/mol}$$

$$\rho_3\text{:}\ \ -2B_2 + 3B_3 - B_6 = 0 \quad \Delta H_3 = 0.1\ \text{kcal/mol}$$

$$\rho_4\text{:}\ \ -B_1 + 4B_2 - 4B_3 + B_7 = 0 \quad \Delta H_4 = 1.2\ \text{kcal/mol}$$

$$\rho_5\text{:}\ \ -3B_1 + 8B_2 - 6B_3 + B_8 = 0 \quad \Delta H_5 = 3.6\ \text{kcal/mol}$$

$$\rho_6\text{:}\ \ -3B_2 + 4B_3 - B_9 = 0 \quad \Delta H_6 = 0\ \text{kcal/mol}$$

$$\rho_7\text{:}\ \ -6B_2 + 6B_3 - B_{10} = 0 \quad \Delta H_7 = -0.6\ \text{kcal/mol}$$

$$\rho_8\text{:}\ \ -3B_1 + 6B_2 - 3B_3 + B_{11} - 3B_{13} + 2B_{14} = 0 \quad \Delta H_8 = 3.8\ \text{kcal/mol}$$

$$\rho_9\text{:}\ \ -B_1 + 2B_2 - B_3 + B_{12} - 2B_{13} + B_{14} = 0 \quad \Delta H_9 = 1.2\ \text{kcal/mol}$$

The stoichiometric matrix and the enthalpy change vector of

TABLE 1: Training Subset of the 14 Species Used in the Analysis of Gronert's Scheme[6]

		$\Delta H_{f,i}^{\mathrm{exp}}$ (kcal/mol)	e_i (kcal/mol)	$\Delta H_{f,i}^{\mathrm{calc}}$
methane	B_1	−17.9	−0.604	−17.296
ethane	B_2	−20.0	0.441	−20.441
propane	B_3	−25.0	0.319	−25.319
n-butane	B_4	−30.4	−0.203	−30.197
iso-butane	B_5	−32.1	−0.169	−31.931
n-pentane	B_6	−35.1	−0.024	−35.076
iso-pentane	B_7	−36.7	0.109	−36.809
neo-pentane	B_8	−40.1	0.177	−40.277
n-hexane	B_9	−40.0	−0.046	−39.954
cyclohexane	B_{10}	−29.4	−0.130	−29.270
methyl radical	B_{11}	35.0	0.077	34.923
ethyl radical	B_{12}	29.0	−0.097	29.097
isopropyl radical	B_{13}	21.5	−0.039	21.539
tert-butyl radical	B_{14}	12.3	0.056	12.244
propyl radical	B_{15}	23.9	−0.3	24.2

TABLE 2: Group Matrix for Gronert's Scheme[6]

	C	H	CC	CH	HCH	CCH	CCC
B_1	1	4	0	4	6	0	0
B_2	2	6	1	6	6	6	0
B_3	3	8	2	8	7	10	1
B_4	4	10	3	10	8	14	2
B_5	4	10	3	10	9	12	3
B_6	5	12	4	12	9	18	3
B_7	5	12	4	12	10	16	4
B_8	5	12	4	12	12	12	6
B_9	6	14	5	14	10	22	4
B_{10}	6	12	6	12	6	24	6
B_{11}	1	3	0	3	3	0	0
B_{12}	2	5	1	5	4	5	0
B_{13}	3	7	2	7	6	8	1
B_{14}	4	9	3	9	9	9	3
B_{15}	3	7	2	7	5	9	1

this set of linearly independent GA RERs are

$$\boldsymbol{\nu} = \begin{array}{c} \begin{array}{cccccccccccccc} B_1 & B_2 & B_3 & B_4 & B_5 & B_6 & B_7 & B_8 & B_9 & B_{10} & B_{11} & B_{12} & B_{13} & B_{14} \end{array} \\ \left(\begin{array}{rrrrrrrrrrrrrr} 0 & -1 & 2 & -1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ -1 & 3 & -3 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & -2 & 3 & 0 & 0 & -1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ -1 & 4 & -4 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ -3 & 8 & -6 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & -3 & 4 & 0 & 0 & 0 & 0 & 0 & -1 & 0 & 0 & 0 & 0 & 0 \\ 0 & -6 & 6 & 0 & 0 & 0 & 0 & 0 & 0 & -1 & 0 & 0 & 0 & 0 \\ -3 & 6 & -3 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & -3 & 2 \\ -1 & 2 & -1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & -2 & 1 \end{array} \right) \end{array} \begin{array}{l} \\ \rho_1 \\ \rho_2 \\ \rho_3 \\ \rho_4 \\ \rho_5 \\ \rho_6 \\ \rho_7 \\ \rho_8 \\ \rho_9 \end{array}$$

$$\mathbf{\Delta H} = (\Delta H_1, \Delta H_2, ..., \Delta H_9)^{\mathrm{T}}$$

Substituting these into eq 12 and performing the matrix operations gives:

$$e_1 = \frac{1}{4583}(340\Delta H_1 - 554\Delta H_2 + 101\Delta H_3 - 315\Delta H_4 - 743\Delta H_5 - 138\Delta H_6 - 1434\Delta H_7 - 107\Delta H_8 + 107\Delta H_9)$$

$$e_2 = \frac{1}{9166}(1404\Delta H_1 - 778\Delta H_2 + 1118\Delta H_3 - 492\Delta H_4 + 760\Delta H_5 + 832\Delta H_6 - 1716\Delta H_7 + 313\Delta H_8 - 313\Delta H_9)$$

$$e_3 = \frac{1}{4583}(721\Delta H_1 - 393\Delta H_2 + 659\Delta H_3 - 331\Delta H_4 + 325\Delta H_5 + 597\Delta H_6 - 372\Delta H_7 + 164\Delta H_8 - 164\Delta H_9)$$

$$e_4 = \frac{1}{9166}(-7686\Delta H_1 - 794\Delta H_2 + 1518\Delta H_3 - 832\Delta H_4 + 540\Delta H_5 + 1556\Delta H_6 + 228\Delta H_7 + 343\Delta H_8 - 343\Delta H_9)$$

$$e_5 = \frac{1}{9166}(794\Delta H_1 + 8034\Delta H_2 + 802\Delta H_3 - 1140\Delta H_4 - 1816\Delta H_5 + 810\Delta H_6 + 48\Delta H_7 - 169\Delta H_8 + 169\Delta H_9)$$

$$e_6 = \frac{1}{4583}(759\Delta H_1 - 401\Delta H_2 - 3724\Delta H_3 - 501\Delta H_4 + 215\Delta H_5 + 959\Delta H_6 + 600\Delta H_7 + 179\Delta H_8 - 179\Delta H_9)$$

$$e_7 = \frac{1}{4583}(416\Delta H_1 - 570\Delta H_2 + 501\Delta H_3 + 3928\Delta H_4 - 963\Delta H_5 + 586\Delta H_6 + 510\Delta H_7 - 77\Delta H_8 + 77\Delta H_9)$$

$$e_8 = \frac{1}{4583}(-270\Delta H_1 - 908\Delta H_2 - 215\Delta H_3 - 963\Delta H_4 + 1264\Delta H_5 - 160\Delta H_6 + 330\Delta H_7 - 589\Delta H_8 + 589\Delta H_9)$$

$$e_9 = \frac{1}{9166}(1556\Delta H_1 - 810\Delta H_2 + 1918\Delta H_3 - 1172\Delta H_4 + 320\Delta H_5 - 6886\Delta H_6 + 2172\Delta H_7 + 373\Delta H_8 - 373\Delta H_9)$$

$$e_{10} = \frac{1}{4583}(114\Delta H_1 - 24\Delta H_2 + 600\Delta H_3 - 510\Delta H_4 - 330\Delta H_5 + 1086\Delta H_6 - 1667\Delta H_7 + 45\Delta H_8 - 45\Delta H_9)$$

$$e_{11} = \frac{1}{45830}(-1715\Delta H_1 - 845\Delta H_2 - 1790\Delta H_3 - 770\Delta H_4 - 5890\Delta H_5 - 1865\Delta H_6 - 450\Delta H_7 + 12469\Delta H_8 - 17052\Delta H_9)$$

$$e_{12} = \frac{1}{45830}(1715\Delta H_1 + 845\Delta H_2 + 1790\Delta H_3 + 770\Delta H_4 + 5890\Delta H_5 + 1865\Delta H_6 + 450\Delta H_7 - 17052\Delta H_8 + 30801\Delta H_9)$$

$$e_{13} = \frac{1}{45830}(1715\Delta H_1 + 845\Delta H_2 + 1790\Delta H_3 + 770\Delta H_4 + 5890\Delta H_5 + 1865\Delta H_6 + 450\Delta H_7 - 3303\Delta H_8 - 10446\Delta H_9)$$

$$e_{14} = \frac{1}{45830}(-1715\Delta H_1 - 845\Delta H_2 - 1790\Delta H_3 - 770\Delta H_4 - 5890\Delta H_5 - 1865\Delta H_6 - 450\Delta H_7 + 7886\Delta H_8 - 3303\Delta H_9)$$

The numerical values of the errors are presented in Table 1. These errors coincide with those obtained by Gronert[6] as well as Wodrich and Schleyer.[7]

Let us next illustrate the evaluation of the enthalpy of formation of a new species, say, propyl radical (B_{15}). In doing this we generate an arbitrary GA reaction involving species B_{15}, for example

$$B_1 + 2B_3 + 2B_{13} = 3B_2 + B_{14} + B_{15}$$

On the basis of the thermoneutrality condition for this GA reaction

$$-\Delta H_{f,1}^{\mathrm{calc}} - 2\Delta H_{f,3}^{\mathrm{calc}} - 2\Delta H_{f,13}^{\mathrm{calc}} + 3\Delta H_{f,2}^{\mathrm{calc}} + \Delta H_{f,14}^{\mathrm{calc}} + \Delta H_{f,15}^{\mathrm{calc}} = 0$$

we obtain

$$\Delta H_{f,15}^{\mathrm{calc}} = \Delta H_{f,1}^{\mathrm{calc}} - 3\Delta H_{f,2}^{\mathrm{calc}} + 2\Delta H_{f,3}^{\mathrm{calc}} + 2\Delta H_{f,13}^{\mathrm{calc}} - \Delta H_{f,14}^{\mathrm{calc}} = 24.2\ \mathrm{kcal/mol}$$

It may be easily checked that the same value is obtained starting from any other GA reaction. For instance

$$B_1 + 2B_{13} + B_{15} = B_2 + 2B_{12} + B_{14}$$

$$\Delta H_{f,15}^{\mathrm{calc}} = -\Delta H_{f,1}^{\mathrm{calc}} + \Delta H_{f,2}^{\mathrm{calc}} + 2\Delta H_{f,12}^{\mathrm{calc}} - 2\Delta H_{f,13}^{\mathrm{calc}} + \Delta H_{f,14}^{\mathrm{calc}} = 24.2\ \mathrm{kcal/mol}$$

$$3B_1 + 3B_{15} = 3B_2 + 2B_{11} + B_{14}$$

$$\Delta H_{f,15}^{\mathrm{calc}} = \frac{1}{3}(-3\Delta H_{f,1}^{\mathrm{calc}} + 3\Delta H_{f,2}^{\mathrm{calc}} + 2\Delta H_{f,11}^{\mathrm{calc}} + \Delta H_{f,14}^{\mathrm{calc}}) = 24.2\ \mathrm{kcal/mol}$$

This value coincides with the value predicted by the conventional analysis.

4. Interpretation

Equation 12 has been obtained based on an arbitrary selection of a set of linearly independent GA reactions. At first glance it might look like the error vector is also arbitrary. Equation 12, however, has a remarkable property according to which the error vector may be partitioned into a sum of contributions associated with a complete set of stoichiometrically unique GA response reactions (RERs). Let the set of stoichiometrically distinct GA RERs be ρ_j ($j = 1, 2, ..., N$). Then, based on our results reported previously,[17] the group additivity errors of the species may be presented as

$$e_i = \frac{1}{\Delta}\sum_{j=1}^{N} \gamma_j^2 \nu_{ji}\, \Delta H_j;\ i = 1, 2, ..., n \tag{13}$$

$$\Delta = \frac{1}{m}\sum_{j=1}^{N} \gamma_j^2 \left(\sum_{i=1}^{n} \nu_{ji}^2\right) \tag{14}$$

where γ_j^2 is the stoichiometric factor of the GA RER ρ_j, ν_{ji} is the stoichiometric coefficient of species B_i in ρ_j, and ΔH_j is the enthalpy change of ρ_j. Keep in mind that eq 12 is given in terms of m linearly independent GA reactions. Because these m GA reactions are arbitrary, they may be selected from the list of GA RERs. To make the notation consistent we assume, without loss of generality, that the m linearly independent GA reactions in eq 12 are GA RERs and they are the first m from a complete list of N. Because of the stoichiometric uniqueness of the GA RERs, the independence of the error vector on the choice of linearly independent GA RERs is self-evident.

The advantage of eq 13 is that the performance of GA schemes may be rationalized in terms of stoichiometry and enthalpy changes of GA RERs. As can be seen, the error of the group additivity is determined by a product of three terms. Two of them, namely, $\gamma_j^2$and ν_{ji}, are stoichiometric in nature while the third is a thermochemical characteristic of GA RERs, that is, their enthalpy changes ΔH_j. Although it is natural to expect that the performance of GA methods depends on the enthalpy changes and stoichiometric coefficients of the species, the dependence of the GA performance on purely stoichiometric factors of GA RERs is not obvious. The stoichiometric factors of GA RER are complex functions of the system's structure expressed by the group matrix. Mainly, the appearance of the stoichiometric factors, that is, values different from one, is due to the fact that some of the species involved in GA RERs happen to have stoichiometric coefficients equal to zero. As a result, several GA RERs become stoichiometrically identical and, consequently, may be considered as a single GA RER. Furthermore, the RERs formalism, eq 13, requires adding the stoichiometric factors of stoichiometrically identical RERs as γ_j^2. Details of the evaluation of the stoichiometric factors of RERs are given in the Appendix.

On the basis of this interpretation, we conclude that the group additivity error is *an expression of the system's stoichiometry and of the extent to which the GA RERs differ from thermoneutrality*.

The partition of the errors into a unique sum of contributions associated with GA RERs, eq 13, provides valuable insider information concerning the performance of GA methods. Indeed, with a complete list of GA RERs and their stoichiometric factors and enthalpy changes, one can easily determine the GA RERs that are responsible for a poor performance of a given GA scheme. In particular, the partition of the errors into contributions associated with GA RERs is a convenient way to look for outliers.[16]

A complete list of GA RERs for Gronert's GA scheme includes 389 stoichiometrically distinct GA RERs (see the Supporting Information). A quick look at this list immediately reveals several interesting particularities of the Gronert's GA scheme performance. The most surprising of them is the effect of the stoichiometric factor. Thus, the stoichiometric factors of some GA RERs may reach substantial values and, in fact, may dominate the errors of the species. Let us illustrate this statement with the help of an example. Consider the following GA RER

$$\rho_{151}\colon\ -12B_1 + 15B_2 - 3B_9 + 8B_{11} - 12B_{12} + 4B_{14} = 0$$

For this GA RER, we have $\Delta H_{151} = 16.0$ kcal/mol, $\gamma_{151}^2 = 1$ and, hence, according to eq 13, its contribution to B_1 (methane) error is $(-12)(1)(16.0)/\Delta = -192.0/\Delta$ kcal/mol. Now, consider the GA RER

$$\rho_{11}\colon\ -2B_1 + 2B_5 - B_{10} = 0$$

Although the enthalpy change for this GA RER is much smaller, $\Delta H_{11} = 1.0$ kcal/mol, the value of the stoichiometric factor is huge, $\gamma_{11}^2 = 1800$. As a result, its contribution to B_1 (methane) error is substantially higher, that is, $(-2)(1800)(1.0)/\Delta = -3600.0/\Delta$.

5. Discussion and Concluding Remarks

Conventionally, the group additivity methods are based on the evaluation of the group values. These are normally determined by a best fitting of the experimental enthalpies of formation of a training set of species for which accurate thermochemical data are available. Mathematically, this procedure is rigorous. Some problems appear when the groups are linearly dependent. In this case, one can either assign arbitrary group values to linearly dependent groups or, alternatively, remove the linearly dependent groups from the analysis. Clearly, because the selection of the linearly independent groups is arbitrary, the group values lack any physicochemical meaning. Respectively, these group values can not be used to rationalize various physicochemical interactions in molecules.

The approach presented in this work relates the performance of group additivity methods to the stoichiometry and thermochemistry of GA reactions. These reactions are required to preserve the number and type of groups and are expected to have small enthalpy changes, that is, be close to thermoneutral. Thus, a good performing additivity method should result in GA reactions whose enthalpy changes are small, or are close to thermoneutral. The more the GA reactions differ from thermoneutrality, the worse the performance of the GA methods. It is the difference from zero of the enthalpy changes of the GA reactions that points at unbalanced interactions in the molecules, respectively, at a bad choice of groups. Furthermore, this stoichiometric approach allows a partition of the error of the group additivity methods into a sum of contributions associated with a special class of stoichiometrically unique reactions, namely, GA RERs. An analysis of a complete set of GA RERs along with their stoichiometric and thermochemical characteristics provides valuable insight into the performance of the group additivity methods.

Appendix

Generation of GA RERs for Gronert's Scheme.[6] Our starting point is the group matrix, Table 2. After several elementary matrix operations, the group matrix may be presented as

$$
\begin{array}{lccccccc}
B_1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 \\
B_2 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\
B_3 & 0 & 1 & 1 & 0 & 0 & 0 & 0 \\
B_4 & 0 & 1 & 2 & 0 & 0 & 0 & 0 \\
B_5 & 1 & 1 & 3 & 0 & 0 & 0 & 0 \\
B_6 & 0 & 1 & 3 & 0 & 0 & 0 & 0 \\
B_7 & 1 & 1 & 4 & 0 & 0 & 0 & 0 \\
B_8 & 3 & 1 & 6 & 0 & 0 & 0 & 0 \\
B_9 & 0 & 1 & 4 & 0 & 0 & 0 & 0 \\
B_{10} & 0 & 0 & 6 & 0 & 0 & 0 & 0 \\
B_{11} & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\
B_{12} & 0 & 0 & 0 & 1 & 1 & 0 & 0 \\
B_{13} & 1 & 0 & 1 & 1 & 2 & 0 & 0 \\
B_{14} & 3 & 0 & 3 & 1 & 3 & 0 & 0
\end{array}
$$

It is seen that the rank of the group matrix is equal to five. In other words, only five out of seven groups are linearly independent. Consider now a 14 × 6 matrix formed from the first 5 columns and the species column:

$$
\begin{pmatrix}
1 & 1 & 0 & 0 & 0 & B_1 \\
0 & 1 & 0 & 0 & 0 & B_2 \\
0 & 1 & 1 & 0 & 0 & B_3 \\
0 & 1 & 2 & 0 & 0 & B_4 \\
1 & 1 & 3 & 0 & 0 & B_5 \\
0 & 1 & 3 & 0 & 0 & B_6 \\
1 & 1 & 4 & 0 & 0 & B_7 \\
3 & 1 & 6 & 0 & 0 & B_8 \\
0 & 1 & 4 & 0 & 0 & B_9 \\
0 & 0 & 6 & 0 & 0 & B_{10} \\
0 & 0 & 0 & 1 & 0 & B_{11} \\
0 & 0 & 0 & 1 & 1 & B_{12} \\
1 & 0 & 1 & 1 & 2 & B_{13} \\
3 & 0 & 3 & 1 & 3 & B_{14}
\end{pmatrix}
$$

According to the response reactions (RERs) formalism, *every determinant of order six formed from this matrix defines a GA RER*. Hence, the total number of GA RERs should not exceed the number of ways six rows may be selected from a total of fourteen, that is, 14!/6!/8! = 3003. In reality, because of a large number of "zero" GA RERs (the determinant is equal to zero, i.e., all stoichiometric coefficients are equal to zero) as well as repetitions, the number of stoichiometrically distinct GA RERs is smaller. For instance, the determinant formed from the first six rows is obviously equal to zero

$$
\begin{vmatrix}
1 & 1 & 0 & 0 & 0 & B_1 \\
0 & 1 & 0 & 0 & 0 & B_2 \\
0 & 1 & 1 & 0 & 0 & B_3 \\
0 & 1 & 2 & 0 & 0 & B_4 \\
1 & 1 & 3 & 0 & 0 & B_5 \\
0 & 1 & 3 & 0 & 0 & B_6
\end{vmatrix} = 0
$$

This means that the GA RER involving the first six species is, in fact, a "zero" GA RER

$$0B_1 + 0B_2 + 0B_3 + 0B_4 + 0B_5 + 0B_6 = 0$$

The explanation is that in this system there are other, "shorter" GA RERs involving the same species, for example

$$-B_2 + 2B_3 - B_4 = 0$$

$$-B_1 + 3B_2 - 3B_3 + B_5 = 0$$

$$-2B_2 + 3B_3 - B_6 = 0$$

A GA RER involving all of the six species would, therefore, violate the stoichiometric uniqueness of GA RERs. In cases when the number of different from zero stoichiometric coefficients in a GA RER is less than six, we face a situation when several GA RERs are actually stoichiometrically equivalent. For example, the following determinants

$$
\begin{vmatrix}
1 & 1 & 0 & 0 & 0 & B_1 \\
0 & 1 & 3 & 0 & 0 & B_6 \\
1 & 1 & 4 & 0 & 0 & B_7 \\
0 & 0 & 0 & 1 & 0 & B_{11} \\
0 & 0 & 0 & 1 & 1 & B_{12} \\
1 & 0 & 1 & 1 & 2 & B_{13}
\end{vmatrix} = 0B_1 + 4B_6 - 4B_7 + 4B_{11} - 8B_{12} + 4B_{13} = 0
$$

$$
\begin{vmatrix}
0 & 1 & 0 & 0 & 0 & B_2 \\
0 & 1 & 3 & 0 & 0 & B_6 \\
1 & 1 & 4 & 0 & 0 & B_7 \\
0 & 0 & 0 & 1 & 0 & B_{11} \\
0 & 0 & 0 & 1 & 1 & B_{12} \\
1 & 0 & 1 & 1 & 2 & B_{13}
\end{vmatrix} = 0B_2 + 3B_6 - 3B_7 + 3B_{11} - 6B_{12} + 3B_{13} = 0
$$

result in essentially the same GA RERs because the stoichiometric coefficients of B_1 and B_2 are equal to zero.

$$4B_6 - 4B_7 + 4B_{11} - 8B_{12} + 4B_{13} = 0$$

$$3B_6 - 3B_7 + 3B_{11} - 6B_{12} + 3B_{13} = 0$$

All together, there are seven such reactions

$$4B_6 - 4B_7 + 4B_{11} - 8B_{12} + 4B_{13} = 0$$

$$3B_6 - 3B_7 + 3B_{11} - 6B_{12} + 3B_{13} = 0$$

$$2B_6 - 2B_7 + 2B_{11} - 4B_{12} + 2B_{13} = 0$$

$$B_6 - B_7 + B_{11} - 2B_{12} + B_{13} = 0$$

$$B_6 - B_7 + B_{11} - 2B_{12} + B_{13} = 0$$

$$-B_6 + B_7 - B_{11} + 2B_{12} - B_{13} = 0$$

$$-6B_6 + 6B_7 - 6B_{11} + 12B_{12} - 6B_{13} = 0$$

It is convenient to replace these GA RERs by a single one in order to simplify the analysis. In doing this, we accept the convention according to which the stoichiometric coefficients in the GA RERs are selected as the smallest integers and introduce a stoichiometric factor γ for every GA RER

$$\begin{array}{lr}
 & \gamma \\
B_6 - B_7 + B_{11} - 2B_{12} + B_{13} = 0 & 4 \\
B_6 - B_7 + B_{11} - 2B_{12} + B_{13} = 0 & 3 \\
B_6 - B_7 + B_{11} - 2B_{12} + B_{13} = 0 & 2 \\
B_6 - B_7 + B_{11} - 2B_{12} + B_{13} = 0 & 1 \\
B_6 - B_7 + B_{11} - 2B_{12} + B_{13} = 0 & 1 \\
B_6 - B_7 + B_{11} - 2B_{12} + B_{13} = 0 & -1 \\
B_6 - B_7 + B_{11} - 2B_{12} + B_{13} = 0 & -6
\end{array}$$

Now, these GA RERs may be substituted with a single GA RER by introducing an overall stoichiometric factor. Because in eq 13 the stoichiometric factors are squared, the overall stoichiometric factor of a GA RER is equal to the squared sum of stoichiometric factors of stoichiometrically equivalent GA RERs. For instance, the overall stoichiometric coefficient of the above GA RER is equal to $\gamma^2 = 4^2 + 3^2 + 2^2 + 1^2 + 1^2 + (-1)^2 + (-6)^2 = 68$. A complete list of stoichiometrically distinct GA RERs along with their stoichiometric factors and enthalpy changes is given in the Supporting Information.

Supporting Information Available: A complete list of stoichiometrically distinct GA RERs along with their stoichiometric factors and enthalpy changes. This material is available free of charge via the Internet at http://pubs.acs.org.

References and Notes

(1) Benson, S. W.; Buss, J. H. *J. Chem. Phys.* **1958**, *29*, 546.

(2) Benson, S. W. *Thermochemical Kinetics*; Wiley-Interscience: New York, 1976.

(3) Tatevskii, V. M. *The Structure of Molecules*; Khimiya: Leningrad, 1977.

(4) Cohen, N.; Benson, S. W. *Chem. Rev.* **1993**, *93*, 2419.

(5) Benson, S. W.; Cruickshank, F. R.; Golden, D. M.; Haugen, G. R.; O'Neal, H. E.; Rodgers, A. S.; Shaw, R.; Walsh, R. *Chem. Rev.* **1969**, *69*, 269.

(6) Gronert, S. *J. Org. Chem.* **2006**, *71*, 1209.

(7) Wodrich, M. D.; Schleyer, P. v. R. *Org. Lett.* **2006**, *8*, 2135.

(8) Yu, J.; Sumathi, R.; Green, W. H. *J. Phys. Chem. A* **2006**, *110*, 6971.

(9) Hayes, M. Y.; Li, B. Q.; Rabitz, H. *J. Phys. Chem. A* **2006**, *110*, 264.

(10) Sagadeev, E. V.; Safina, Y. G.; Cherkasov, R. A. *Russ. J. Org. Chem.* **2005**, *41*, 1526.

(11) Sabbe, M, K.; Saeys, M.; Reyniers, M. F.; Marin, G. B.; Van Speybroeck, V.; Waroquier, M. *J. Phys. Chem. A*, **2005**, *109*, 7466.

(12) Roganov, G. N.; Pisarev, P. N.; Emel'yanenko, V. N.; Verevkin, S. P. *J. Chem. Eng. Data* **2005**, *50*, 1114.

(13) Yu, J.; Sumathi, R.; Green, W. H., Jr. *J. Am. Chem. Soc.* **2004**, *126*, 12685.

(14) Saraf, S. R.; Rogers, W. J.; Mannan, M. S.; Hall, M. B.; Thomson, L. M. *J. Phys. Chem. A* **2003**, *107*, 1077.

(15) Chen, C.-J.; Wong, D.; Bozzelli, J. W. *J. Phys. Chem. A* **1998**, *102*, 4551.

(16) Fishtik, I.; Datta, R. *J. Chem. Inf. Comput. Sci.* **2003**, *43*, 1259.

(17) Fishtik, I.; Datta, R. *J. Phys. Chem. A* **2003**, *107*, 6698.

(18) Fishtik, I.; Datta, R.; Liebman, J. *J. Phys. Chem. A* **2003**, *107*, 2334.

(19) Sandler, S. I. *Chemical and Engineering Thermodynamics*; Wiley & Sons: New York, 1999.