

Additive and Nonadditive Effects of Two Factors on a Response Function

Otto Exner

Institute of Organic Chemistry and Biochemistry, Academy of Sciences of the Czech Republic,
166 10 Praha 6, Czech Republic

Received April 10, 1995*

General dependence of a response function on two factors can be always approximated by a linear and additive empirical relationship. When this is not sufficiently precise, a nonadditive relationship is preferred which is nevertheless linear in each variable separately. Properties of this equation are shown in examples, possible interpretation of the parameters is discussed, and the characteristic φ is proposed, measuring quantitatively the deviation from additivity. For distinguishing additive and nonadditive behavior of a certain response function, several statistical methods were tested, and a method was described which avoids any external parameters and is more efficient than a method proposed recently. There is good evidence that nonadditive relationships of this type can be used much more extensively than it has been usual in the literature. Extension to a function of three factors was briefly examined.

INTRODUCTION

If a given property—the response function y_{ij} —depends on two factors, this dependence can be expressed in the first approximation as a sum,^{1,2} sometimes even additivity is understood in the same meaning as independency.^{2,3} The simplest and most general representation of additivity is eq 1. When the factors can be expressed by any external parameters, denoted here σ in the broadest sense, the additive relationship has the form of eq 2a:⁴ ρ are proportionality factors, and the residuals ϵ_{ij} are assumed to be the values of a random variable.

$$y_{ij} - y_{iq} = y_{pj} - y_{pq} \quad (1)$$

$$y_{ij} = y_{00} + \rho_1 \sigma_i + \rho_2 \sigma_j + \epsilon_{ij} \quad (2a)$$

$$y_{ij} = y_{00} + \rho_1 \sigma_i^{\text{in}} + \rho_2 \sigma_j^{\text{in}} + \epsilon_{ij} \quad (2b)$$

$$y_{ij} = c_i + d_j + \epsilon_{ij} \quad (2c)$$

When external parameters σ are not at disposal and the data matrix $|y_{ij}|$ is sufficiently large, one line and one column can be used to define *ad hoc* constants σ , also called *de novo* or internal constants.⁵ In this way, this line and column are given more weight than the others. A more exact statistical procedure⁶ is based on the reduced eq 2c: the task is calculating all values c_i and d_j at once from the data matrix to meet the least-squares condition of eq 3.

$$\sum_{ij} \epsilon_{ij}^2 = \min \quad (3)$$

In whatever form, the additive relationship was found satisfactory in numerous examples, particularly when the changes of y_{ij} are small. A well-known example is the effect of remote substituents in the Hammett equation.^{4,7} Sometimes, the additivity was even anticipated like a dogma, and no other relationship was taken into consideration.⁸ Nevertheless, with the increasing number of data and their improved accuracy, the additive relationship has appeared

insufficient. As the next approximation, eq 4 was recommended,¹ which is linear in each independent variable separately. Either external parameters,⁹ eq 4a, or internal parameters,⁵ eq 4b, can be used. The reduced form,^{6,10} eq 4c, can be obtained from eq 4a by the substitution $\sigma_i = (c_i - \rho_2)/\rho_{12}$ and $\sigma_j = d_j - \rho_1/\rho_{12}$.

$$y_{ij} = y_{00} + \rho_1 \sigma_i + \rho_2 \sigma_j + \rho_{12} \sigma_i \sigma_j + \epsilon_{ij} \quad (4a)$$

$$y_{ij} = y_{00} + \rho_1 \sigma_i^{\text{in}} + \rho_2 \sigma_j^{\text{in}} + \rho_{12} \sigma_i^{\text{in}} \sigma_j^{\text{in}} + \epsilon_{ij} \quad (4b)$$

$$y_{ij} = y^0 + c_i d_j + \epsilon_{ij} \quad (4c)$$

The constant y^0 in eq 4c is independent of possible transformations of the parameters c_i and d_j and could possess a physical meaning: it was called the isoparametric¹² or magic¹³ point. Much attention was also given to the fact that for certain values of one independent variable, the response function is constant and independent of the second variable. This occurs for $c_i = 0$ or $d_j = 0$ in eq 4c, or for $\sigma_i = -\rho_2/\rho_{12}$ or $\sigma_j = -\rho_1/\rho_{12}$ in eq 4a (the isoparametric values^{11,12}).

The validity of nonadditive relationships was proven several times,^{12,13,15c} particularly for the effects of two adjacent substituents.^{6,10,15a,b} In the field of Hammett equation, the deviations from additivity are weaker, and an evidence is more difficult to obtain.^{5,14,16,17} A statistical proof that eq 4a is more accurate than eq 2a means testing the null hypothesis H_0 : $\rho_{12} = 0$. When the correlation with σ is not close, the effect of the cross-term may be masked by the noise, and a decision becomes difficult. For this reason, Kevill and D'Souza⁵ suggested recently a treatment with internal σ . It proceeds in three steps. Firstly, one line and one column of the matrix are selected to define the internal constants σ^{in} : $\sigma_i^{\text{in}} = y_{i0} - y_{00}$ and $\sigma_j^{\text{in}} = y_{0j} - y_{00}$. Next, multiple linear regressions are carried out on all y_{ij} values (including those used for the definition of σ_{in}), according to eq 2b, on the one hand, and eq 4b, on the other hand. Ultimately, the two regressions are compared: qualitatively according to the correlation coefficients and semiquantitatively from the coefficient ρ_{12} and its standard deviation. The procedure has thus several defects:

* Abstract published in *Advance ACS Abstracts*, October 1, 1995.

1. All data are not given the same weight.

2. The degrees of freedom in the regression are mistaken. Even when the standard deviation of the fit was not calculated, the correlation coefficient lead to an overestimation of the improvement of the fit.

3. Certain incomplete matrices cannot be simply processed.

All the defects are absent when eqs 2c and 4c are compared as suggested earlier.⁶ A correct statistical test is the main objective of this note. Three procedures will be compared and tested by means of the F-test: original regression¹⁷ upon external constants σ , Kevill–D'Souza improved procedure,⁵ and our least-squares treatment.⁶ In addition, some properties of eq 4c and its interpretation will be shortly discussed on the basis of new examples.

RESULTS AND DISCUSSION

Statistical Tests of Additive and Nonadditive Relationships. As a testing sample we used the same data as Kevill and D'Souza⁵ (Table 1). These are arranged^{17a} in 4×3 matrices; six such matrices are available for six different solvents. Data processing was carried out here using three procedures.

Procedure A. Standard multiple linear regression with constants σ were carried out according to eqs 2a and 4a, respectively, in the same way as in the original work.^{17a} The same values of σ were used. The improvement of eq 4a against eq 2a (inclusion of the cross-term) was evaluated by F-tests with 1 and 8 degrees of freedom, separately for each matrix.

Procedure B. The recommendation of Kevill and D'Souza⁵ was followed exactly. Internal constant σ^{in} were calculated with respect to the substituent hydrogen: $\sigma_i^{\text{in}} = \log k_{i,H} - \log k_{H,H}$; $\sigma_j^{\text{in}} = \log k_{H,j} - \log k_{H,H}$. Multiple linear regressions proceeded according to eqs 2b and 4b, respectively. The standard deviations which appeared on the output of the regression program are related to 9 or 8 degrees of freedom, respectively, and had to be recalculated to the correct value $f = 6$ or 5. The F-test has then 1 and 5 degrees of freedom. (Values of f must be equal as in the next procedure when the degrees of freedom consumed in the definition of σ^{in} are taken into account.)

Procedure C. In eq 2c all parameters c_i and d_j are calculated at once in order to meet the least-squares condition, eq 3. Similarly, all c_i and d_j as well as y^0 are calculated in eq 4c. The programs were reported;⁶ some details are mentioned in the Appendix. The F-tests have 1 and 5 degrees of freedom as in the procedure B.

The results are listed in Table 1. If we observe first the standard deviations, an evident improvement in the procedure B against A is due to the standard values of constant σ which have been determined^{4,18} for the whole body of experimental data and may fit in a particular case only with some inaccuracy. A smaller improvement of the procedure C against B is due only to a more rigorous statistical treatment which distributes the residual errors equally on all data and does not give any special significance to the substituent H (as was given in⁵). The small improvement in the standard deviations is nevertheless quite significant in the F-tests (Table 1, middle part). In spite of less degrees of freedom f , the F-tests in the procedure B are more significant than in A: more cases are decided in favor of the nonadditive

Table 1. Statistical Comparison of Three Procedures for Evaluating the Deviations from Additivity^a

Standard Deviations of the Fit, 10 ⁴ log units						
A. external σ , additive, eq 2a	675	680	398	510	544	800
nonadditive, eq 4a	584	627	366	324	485	761
B. internal σ , ^b additive eq 2b	554	507	517	557	448	884
nonadditive, eq 4b	364	307	502	302	314	822
C. de novo, additive, eq 2c	539	486	397	529	411	612
nonadditive, eq 4c	333	267	348	215	241	466
experimental uncertainty				128		
F-Tests						
A. $F_{1,8}$	4.02	2.58	2.64	14.29	3.32	1.94
confidence level	0.10			0.01		
B. $F_{1,5}$	8.90	11.36	1.36	15.41	7.21	1.94
confidence level	0.05	0.025		0.025	0.05	
C. $F_{1,5}$	10.72	14.87	2.81	31.32	12.45	5.34
confidence level	0.025	0.025		0.005	0.025	0.10
Cross-terms						
A. $-\rho_{ij}$	0.31	0.27	0.16	0.32	0.23	0.28
	± 0.15	± 0.17	± 0.10	± 0.09	± 0.13	± 0.20
B. ρ_{ij}	0.29	0.22	0.11	0.29	0.18	0.20
	± 0.08	± 0.05	± 0.08	± 0.06	± 0.05	± 0.11
C. $-y^0$	3.32	3.81	7.45	3.21	4.71	3.53
φ (eq 6)	0.17	0.14	0.074	0.17	0.12	0.14

^a Tested on six series from reaction kinetics: logarithms of rate constants of solvolyses of 5-substituted-indan-2-yl 4-substituted-benzenesulfonates, experimental data ref 17a. ^b Procedure recommended in ref 5.

relationship and with a higher confidence. In the procedure C there is an additional improvement, while f is unchanged. Alternative statistical testing,^{5,14} based on the value of the cross-term coefficient, ρ_{12} , and on its standard deviation, is less efficient (Table 1, bottom) although it generally parallels the results from the F-tests. The values of y^0 (last line) are given here without any standard deviation since their distribution is far from normal, see the next section.

In conclusion, the results of testing are convincing, although the dimensions of the test sets are rather small. Note also that even the smallest residual standard deviation is still significantly greater than the assumed^{17a} experimental uncertainty. Hence we test actually the fit of an empirical relationship, and the test is not disturbed by the experimental error. Evidently, the statistically rigorous test based on eqs 2c, 4c, and 3 (our procedure C) is the method of choice whenever the task is to decide between additive and nonadditive behavior. As it was stressed,^{5,6} this kind of processing does not allow the controlling factors to be revealed: for this purpose equations of the type of eqs 2a or 4a with various kinds of constants σ must be used.

Some Properties of the Nonadditive Relationship. In the preceding examples the differences between additive and nonadditive processing were relatively small, partly because the data matrices were not extensive. Some experience with this treatment is shown better on more practical examples in which the data matrix is larger, but some elements are usually lacking. In addition, the data need not be quite homogeneous and not equally reliable. We choose several new examples, concerning the acid-base properties of simple compounds in the gas phase¹⁹ and in dimethyl sulfoxide²⁰ and processed them according to eqs 2c and 4c (Table 2). All the data matrices are symmetrical with the same substituents in the two positions, X and Y. This affects calculation of the percentage filling of the matrix (Table 2, fourth column). For some calculation details see Appendix.

Table 2. Correlations According to a Nonadditive Relationship, Eq 4a

reaction	matrix dimension	N^a	matrix filling	standard deviations ^b			F-test ^c	y^0 b,d	φ eq 6
				s_0^e	eq 2c	eq 4c			
1. acidity of XCH_2Y , gas ^f	8×8	23	64%	109	29.1	16.2	34.4 (<0.001)	1205 (1067,1268)	0.39
2. basicity of $\text{XYC}=\text{O}$, gas ^g	12×12	39	50%	49.8	7.95	8.37		<-10 000	
3. basicity of $\text{XYC}=\text{O}$, gas ^h	6×6	19	90%	33.1	9.41	3.64	75.7 (<0.001)	871 (858, 895)	0.49
4. acidity of XCH_2Y , DMSO ⁱ	9×9	30	67%	67.0	21.1	14.0	28.0 (<0.001)	-49 (-152, +6)	0.41
5. acidity of XCH_2Y , DMSO ^j	9×9	28	62%	68.7	16.1	6.91	84.8 (<0.001)	-79 (-135, -45)	0.36

^a Number of independent data. ^b In kJ mol^{-1} . ^c Confidence level in parentheses. ^d Confidence interval in parentheses corresponds to $\alpha = 0.05$. ^e Standard deviation of all the values from their mean value. ^f Substituents X, Y = H, F, CH_3 , C_6H_5 , CN, COCH_3 , COC_6H_5 , and COCF_3 ; experimental data refs 19a and 8c. ^g Substituents X, Y = H, CH_3 , C_2H_5 , $\text{CH}(\text{CH}_3)_2$, $\text{C}(\text{CH}_3)_3$, C_6H_5 , $\text{c-C}_3\text{H}_5$, CH_2F , CF_3 , OH, OCH_3 , and OC_2H_5 ; experimental data ref 19b. ^h Substituents restricted to H, CH_3 , C_2H_5 , $\text{CH}(\text{CH}_3)_2$, C_6H_5 , OCH_3 . ⁱ Substituents X, Y = H, C_6H_5 , CN, COCH_3 , COC_6H_5 , OC_6H_5 , SC_6H_5 , $\text{SO}_2\text{C}_6\text{H}_5$, SO_2CF_3 ; experimental data ref 20. ^j Two compounds omitted: $\text{CH}_2(\text{COC}_6\text{H}_5)_2$ and $\text{CH}_3\text{COCH}_2\text{SO}_2\text{C}_6\text{H}_5$.

In all series of Table 2, eq 4c is evidently the method of choice. The nonadditive behavior was proven with an extraordinary reliability. Note that in a recent paper^{8c} in which some experimental data for the series 1 are given, the additivity was considered as the single scheme possible, and the data were treated only in terms of deviations from additivity. However, proceeding from eq 2c to eq 4c brings a very significant improvement at the cost of only one degree of freedom.

In the series 2, the lacking items are relatively numerous and distributed irregularly with the result that there are only three items in certain lines or columns. This causes generally numerical instability: the iterations proceed slowly, and in the limiting case no convergence is reached. This happened also in the series 2. No value of y^0 was found, and eq 4c was not better than 2c. The picture is changed dramatically when the data matrix has been radically reduced as in the series 3. This situation is typical for many examples: more significant results are obtained with a smaller but better filled data matrix. In the series 4 two items were eliminated as outliers to give series 5. The fit was improved significantly (SD reduced to one half), but the main conclusion has not been changed. Eliminating outliers is generally not easy in data processing of this type. When the data matrix is irregularly filled, the fitting error may be spread over the other data, and its actual source need not be discovered. Hence one cannot rely only on the values of residuals, and a graphical representation like the one in Figure 1 may be useful. In this plot, the data of each matrix line are plotted against the parameter d_j . A family of straight lines is obtained, intersecting in a point with the ordinate y^0 . Definite identification of an outlier must be based on comparison of the whole calculation with and without this particular item.

In Figure 1 also the significance and reliability of the parameter y^0 can be demonstrated. When y^0 is given a predetermined value a , eq 4c can be solved with this constraint, and the standard deviation obtained can be plotted against a (Figure 1 at the bottom). The curve is strongly unsymmetrical. The fit is much worsened when the point of intersection should approach the experimental points, whereas it is less sensitive to a displacement in the opposite direction. The limiting value, s_{add} for $y^0 \rightarrow \infty$, represents the case when the straight lines are parallel and the additive relationship, eq 2c, holds.²¹ The confidence interval of y^0 for a chosen α can be approximately found by calculating SD for the critical value F^α and finding the two points on the curve. The intervals given in Table 2 have been calculated in this way for $\alpha = 0.05$. Evidently the interval may be large when the straight lines intersect at small angles.

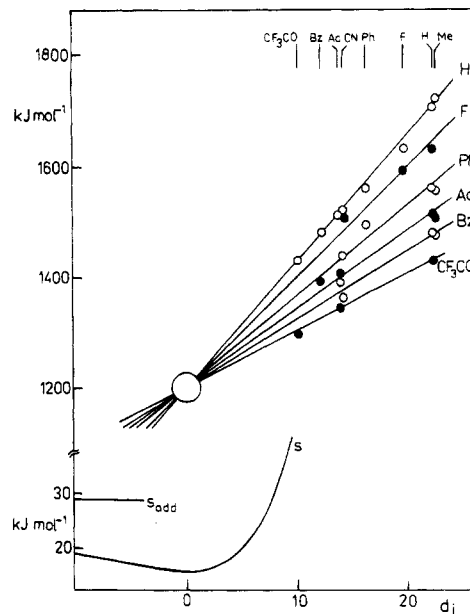


Figure 1. Correlation with eq 4c: ΔG° of the ionization of C-acids XCH_2Y in the gas phase (reaction series 1 in Table 2) plotted vs the constants d_j of the group Y. At the bottom, dependence of the residual standard deviation on the assumed fixed value of y^0 (i.e., on the possible position of the point of intersection).

Nevertheless, Figure 1 proves that the magic point evidently exists and is not a mathematical artifact.

Interpretation of Nonadditive Relationships. Equation 4a contains three parameters which can be given a physical meaning. The two proportionality constants, ρ_1 and ρ_2 , are often understood in the same sense as ρ in the simple Hammett equation or in eq 2a: as sensitivity to substituent effects.^{4,5,17} However, sensitivity to σ_i is in fact given by $(\rho_1 + \rho_{12}\sigma_i)$ and is reduced to ρ_1 only for $\sigma_i = 0$. The cross-interaction coefficient ρ_{12} is usually understood as a measure of interaction or, say, as the degree of nonadditivity.^{5,16,17} This is evident particularly when the cross term is small. However, for comparison of diverse reactions or even different response functions y , a characteristic would be desirable, invariable both to the units of y and definition of σ . In the simple Hammett equation and in eq 2a, the proportionality constants ρ have dimension of $y\sigma^{-1}$ and do not depend on the choice of the standard substituent ($\sigma_H = 0$). For a given quantity y (for instance logarithm of a rate constant) two constants ρ can be compared: their ratio is invariant. This is not the case with eq 4a. Here only ρ_{12} can be compared in relative values, but ρ_1 and ρ_2 depend on the choice of a standard substituent. One could argue that the choice $\sigma_H = 0$ in the Hammett equation is natural or

even the only reasonable possibility. However, this need not hold for all similar equations: let us remember for instance the standard solvent. As a better measure of interaction, the parameter α was suggested¹² and recommended⁵ as defined in eq 5. However, even this parameter depends on the standard substituent and is in this respect worse than simple ρ_{12} .

$$\alpha = \rho_{12}/\rho_1\rho_2 \quad (5)$$

When we proceed to eq 4c, the dependence on any external parameter is lost and there remains only the parameter y^0 , independent of scaling the variables c_i and d_j but dependent on the units of y . For this reason it was believed^{6,10} that it should have a physical meaning of any limiting value of y (magic point¹³). However, it is often rather far from the experimentally accessible values and determined with a considerable inaccuracy. One can even doubt whether the linear dependence would remain linear up to the point of intersection. In such cases y^0 is merely an artificial product of linear extrapolation of a function which is only approximately linear.

When we want to find a measure of nonadditive behavior in eq 4c, we can consider the distance of the point of intersection from the center of gravity of all points, $(y^0 - \bar{y})$. Even this value has the dimension of y . As a dimensionless parameter, independent of units of y , the parameter φ can be suggested, eq 6, where s_y is the standard deviation of all y_{ij} from their mean value. The values of φ in Tables 1 and 2 are remarkably similar for similar series, while in Table 2 they are evidently much larger. In geometrical terms, φ is related to the tangents of the mean angle of the straight lines, see Figure 1. (This is valid only for a symmetrical matrix, for an unsymmetrical matrix there are two different graphical representations.)

$$\varphi = s_y/|y^0 - \bar{y}| \quad (6)$$

An interpretation of modest kind could be satisfied with the position of y^0 , i.e., with the sign of $(y^0 - \bar{y})$. For instance in the case of reaction rates, y^0 can lie either on the side of low or of high rates.²² For the acidities of compounds XCH_2Y in the gas phase, in DMSO (Table 2), in water,⁶ also for the acidity^{15b} or XNH_2 , y^0 is always on the side of stronger acidity. For the basicity of XCO_2Y (Table 2) it is on the side of stronger basicity, for XNH_2 of weaker basicity.⁶ All cases have in common that y^0 is on the opposite side than the point for $X, Y = H$. In other terms, the cumulative substituent effects are always weaker than additive. Within the framework of eq 4a, similar considerations were connected with the sign^{5,12,16} of α or simply with the combination of signs²³ of $\rho_1, \rho_2, \rho_{12}$. This is not quite exact since y^0 and ρ_{12} are connected by eq 7. Hence $(y^0 - \bar{y})$ would equal $-1/\alpha$ only if $\bar{y} = y_{00}$. (This is approximately valid in many cases.)

$$y^0 = y_{00} - \rho_1\rho_2/\rho_{12} \quad (7)$$

Extension to Three Independent Variables. The first steps have been made in developing theory of correlation equations with three variables.^{17b,24,25} The additive forms, eq 8a or 8b, present no new problems compared to eq 2a or 2c, respectively.

$$y_{ijk} = y_{000} + \rho_1\sigma_i + \rho_2\sigma_j + \rho_3\sigma_k + \epsilon_{ijk} \quad (8a)$$

$$y_{ijk} = b_i + c_j + d_k + \epsilon_{ijk} \quad (8b)$$

A complete, nonadditive equation, linear in each variable separately, has the form of eq 9.

$$y_{ijk} = y_{000} + \rho_1\sigma_i + \rho_2\sigma_j + \rho_3\sigma_k + \rho_{12}\sigma_i\sigma_j + \rho_{13}\sigma_i\sigma_k + \rho_{23}\sigma_j\sigma_k + \rho_{123}\sigma_i\sigma_j\sigma_k + \epsilon_{ijk} \quad (9)$$

There are no isoparametric values of one variable which would make y insensitive to other variables.^{24a} There are only combinations of two variables,^{24a} as defined for instance by eq 10 for σ_i and σ_j , which make y insensitive to the third variable (here σ_k).

$$\rho_3 + \rho_{13}\sigma_i + \rho_{23}\sigma_j + \rho_{123}\sigma_i\sigma_j = 0 \quad (10)$$

On the other hand, critical values of another type exist (e.g., $\sigma_i = -\rho_{23}/\rho_{123}$) which reduce eq 9 to an additive relationship between the two remaining variables (here σ_j and σ_k).

Equation 9 requires systematic experimental data, arranged in a three-dimensional matrix. Almost all known examples are from the work of the Korean school and concern nucleophilic reactions.^{17b,26} Two kinds of statistical analysis were carried out. The original authors treated the data within the framework of eq 4a, after dividing them into subsets, always with one constant substituent.^{17b,26} This approach proceeds actually in steps, and a part of the information is lost. Therefore, it leads to some incorrect claims, see for instance Table 3, note *e*. The second analysis^{24b} was based on eq 9 in two forms, either with external or internal σ constants. However, the statistical tests were inefficient. Comparison was carried out only with the additive relationship, eq 8a, and with eq 9 in which the last, third-order term was omitted. The fit was estimated only from the correlation coefficients, and the statistical significance of each term separately was not tested.²⁷

In this paper the more reliable descendent regression was applied. Multiple linear regression was carried out according to eq 9, the least significant term was eliminated, and the procedure was repeated. After each step, significance of the eliminated term was examined by an F-test on the residual standard deviations. Much telling is also a simple plot (not shown) of these standard deviations against the number of terms which reveals usually a marked break when the optimum number of terms has been attained. The results obtained on several examples are listed in Table 3. The additive scheme was found satisfactory only in one case; on the other hand, the full equation, eq 9, was never necessary. Generally, three to six terms were significant, and their number depends on the significance level and may differ even in quite similar series. Comparison with previous statistical analyses^{17b,24b,26} may be shown on the series 1. The coefficient ρ_{XYZ} was estimated to -0.53 with an unknown reliability^{17b} or^{24b} to -0.52 ± 0.24 ; with internal σ^{in} $\rho_{XYZ}^{in} = -0.93 \pm 0.37$ was obtained.^{24b} We calculated values from -0.52 ± 0.26 (from eq 9) to -0.19 ± 0.17 (from an equation with four terms), always statistically insignificant. As a result, reduction to two terms (with σ_X and σ_Z) was recommended from the first analysis,^{17b} while the second retained all seven terms.^{24b} However, with two terms the fit is extremely imprecise, with seven terms it is not better

Table 3. Correlations of Three-Dimensional Data According to Additive and Nonadditive Relationships

reaction	matrix dimensions	standard deviations with the number of terms ^a					
		2	3 (eq 8a)	4	5	6	7 (eq 9)
1. $\text{XC}_6\text{H}_4\text{NH}_2 + \text{YC}_6\text{H}_8\text{OSO}_2\text{C}_6\text{H}_4\text{Z}^b$	$4 \times 3 \times 4$	1346 ^c	397	396	384	377	381
2. $\text{XC}_6\text{H}_4\text{NH}_2 + \text{YC}_6\text{H}_4\text{CH}(\text{CH}_3)\text{OSO}_2\text{C}_6\text{H}_4\text{Z}^c$	$4 \times 4 \times 4$	1053	762	350	322	307^f	309
3. $\text{XC}_6\text{H}_4\text{NH}_2 + \text{YC}_6\text{H}_4\text{CH}(\text{CH}_3)\text{OSO}_2\text{C}_6\text{H}_4\text{Z}^c$	$4 \times 4 \times 4$	1096	782	359	318^g	313	316
4. $\text{XC}_6\text{H}_4\text{NH}_2 + \text{YC}_6\text{H}_4\text{CH}(\text{CH}_3)\text{OSO}_2\text{C}_6\text{H}_4\text{Z}^c$	$4 \times 4 \times 4$	1189	841	470	415^h	412	416
5. $\text{XC}_6\text{H}_4\text{NH}_2 + \text{YC}_6\text{H}_4\text{COCH}_2\text{OSO}_2\text{C}_6\text{H}_4\text{Z}^d$	$4 \times 3 \times 4$	3367	1213	445	407^h	392	396
6. $\text{XC}_6\text{H}_4\text{NH}_2 + \text{YC}_6\text{H}_4\text{COCH}_2\text{OSO}_2\text{C}_6\text{H}_4\text{Z}^d$	$4 \times 3 \times 4$	3269	956	368	341^h	328	332
7. $\text{XC}_6\text{H}_4\text{NH}_2 + \text{YC}_6\text{H}_4\text{COCH}_2\text{OSO}_2\text{C}_6\text{H}_4\text{Z}^d$	$4 \times 3 \times 4$	3202	1184	424ⁱ	402	395	400

^a In 10^4 log units: when all terms are significant at $\alpha = 0.01$, SD is printed in bold (in all cases but one they are significant even at $\alpha = 0.005$), when at $\alpha = 0.05$, they are in italics. ^b Data from ref 17b. ^c Data from ref 26a: reaction series 2–4 differ only in composition of the solvent. ^d Data from ref 26b: series 5–7 differ only in the composition of the solvent. ^e Using the terms with σ_X and σ_Z as in ref 17b. ^f Significant are terms with $\sigma_X\sigma_Z$, $\sigma_X\sigma_Y$, and $\sigma_Y\sigma_Z$. ^g Significant are terms with $\sigma_X\sigma_Z$ and $\sigma_X\sigma_Y$. ^h Significant are terms with $\sigma_Y\sigma_Z$ and $\sigma_X\sigma_Z$. ⁱ Significant is the term with $\sigma_Y\sigma_Z$.

than with three (Table 3). The observed numerical instability²⁸ is due mainly to a low sensitivity to the effects of relatively remote substituents. We agree fully with Kevill and D'Souza^{24b} that the accuracy and extent of present data is not sufficient for evaluating the interaction terms. Further problems are connected with the interpretation of the number of necessary terms, on the one hand, and of the parameters ρ , on the other hand. The latter depend on the units of y and on the definition of σ in a much more complex way than in the case of two variables in the preceding section.

The above problems could be avoided if the redundant terms were removed from eq 9 by a convenient coordinate transformation, similarly as eq 4a was transformed into 4c. Analysis revealed²⁵ that any single form does not exist. When all ρ 's and also certain of their functions are all different from zero, eq 11 is obtained. When this is not the case, degenerate forms arise. Eight such forms were described:²⁵ some of them rather curious, and one possibility was omitted (eq 12c). Here four degenerate forms are reproduced, eqs 12a–d, the ultimate reduction leads to the additive relationship, eq 8b. In some forms, isoparametric values are possible.²⁵

$$y_{ijk} = b_i + c_j + d_k + \omega b_i c_j d_k + \epsilon_{ijk} \quad (11)$$

$$y_{ijk} = y^0 + b_i c_j d_k + \epsilon_{ijk} \quad (12a)$$

$$y_{ijk} = y^0 + b_i c_j + b_i d_k + c_j d_k + \epsilon_{ijk} \quad (12b)$$

$$y_{ijk} = y^0 + b_i(c_j + d_k) + \epsilon_{ijk} \quad (12c)$$

$$y_{ijk} = b_i c_j + d_k + \epsilon_{ijk} \quad (12d)$$

An analysis of experimental data would require to try subsequently eqs 8b, 11, 12a–d (or still further forms²⁵): the equation giving the best fit should be revealed by statistical testing (a nontrivial problem). To date, there are few data sets sufficiently extensive,^{16,17b,24a,26} but in all the substituent effects are relatively weak (see Table 3). Consequently, several possible types of the correlation equation have not yet been encountered in practice, and the whole theory appears oversophisticated. Equation 11 is the most general and probably also the most common type, but the physical meaning of the parameters y^0 and ω is not clear. Only in the limiting case, when the cross term is small, it can be understood as a small correction to the otherwise additive relationship and the coefficient ω as a measure of nonadditivity. Equations 12c and 12d were called semiad-

ditive²⁵ since certain pairs of variables behave additively and the other pairs do not. These relationships can be easily decomposed into relations between two variables, additive and nonadditive.

CONCLUSIONS

Undoubtedly, additivity of two factors is the fundamental simplest scheme, valid in every case as first approximation—in the same sense as every curve can be approximated by a straight line. On the other hand, the additive scheme has been used too frequently, even in the cases when it was not sufficiently precise and could be improved readily. As the second approximation, eq 4c is recommended here to be tested first. In many examples, the fit was improved in a quite convincing manner at the cost of only one degree of freedom. The treatment within the framework of eq 4c is also the method of choice whenever the additive and nonadditive behavior is to be distinguished. It is more effective than the procedure recommended recently.⁵ The parameter y^0 —the magic point—is independent of scaling the variables but need not possess an immediate physical meaning, sometimes it is only a product of a formal extrapolation. Generally, deviations from additivity can be measured by the dimensionless parameter φ , eq 6, independent of units and scaling.

An extension to three independent variables seems to be not useful at present: it also brings great difficulties. Several nonadditive equations are possible, for instance eqs 12a–e, but the results obtainable are hardly worth the enormous experimental effort since also the possible interpretation is obscure.

APPENDIX

Numerical Calculation of the Parameters. All the parameters c_i and d_j of eq 2c are defined unambiguously by the least-squares condition, eq 3. One obtains the normal equations, eqs 13 and 14 in which the weights w_{ij} may equal 1 or 0, according to whether the pertinent y_{ij} is known or not. (Of course, w_{ij} can be given variable values according to the accuracy of the pertinent measurement, but this is not recommended unless the unequal accuracy is proven by rigorous tests.)

$$c_i \sum_{ij} w_{ij} + \sum_{ij} w_{ij} d_j = \sum_{ij} y_{ij} \quad (13)$$

$$d_j \sum_i w_{ij} + \sum_{ij} w_{ij} c_i = \sum_{ij} y_{ij} \quad (14)$$

In the case of a complete data matrix, there is an explicit solution, eqs 15 and 16, where C is an arbitrary constant which may be added to each c_i when it is simultaneously subtracted from each d_j .

$$c_i = \sum_j y_{ij}/q + C \quad (15)$$

$$d_j = \sum_i y_{ij}/p - \sum_{ij} y_{ij}/pq - C \quad (16)$$

In the case of an incomplete data matrix, iterations are possible, calculating alternately all c_i from d_j using eq 13 and all d_j from c_i using eq 14. The convergence is rapid in all known examples.⁶ The arbitrary constant C can be added even in this case. Therefore, the residual sum of squares has $f = N - p - q + 1$ degrees of freedom.

Similarly, the parameters c_i , d_j , and y^0 of eq 4c are determined by the normal equations, eqs 17–19.

$$c_i \sum_j w_{ij} d_j^2 + y^0 \sum_j w_{ij} d_j = \sum_j w_{ij} d_j y_{ij} \quad (17)$$

$$d_j \sum_i w_{ij} c_i^2 + y^0 \sum_i w_{ij} c_i = \sum_i w_{ij} c_i y_{ij} \quad (18)$$

$$\sum_{ij} w_{ij} c_i d_j + N y^0 = \sum_{ij} w_{ij} y_{ij} \quad (19)$$

When y^0 is given any value, the pertinent values of c_i and d_j can be obtained by iterations using eqs 17 and 18. The convergence is reasonably fast. However, improving the value of y^0 is a tedious procedure. Cyclic iterations, using successively eqs 17, 18, and 19 to obtain c_i , d_j , and y^0 , respectively, are unsatisfactory. Particularly, when the data matrix is filled irregularly, convergence is slow, sometimes even one value of y^0 can be reached twice. A more sophisticated equation suggested⁶ instead of eq 19 has not helped too much. Here we used a kind of accelerated iterations where the value of y^0 was always corrected by a ten times greater increment than would be calculated from eq 19 (see eq 20). Note that an irregularly filled matrix may cause problems even in this procedure.^{15c} We conclude that most reliable is a somewhat primitive method: For a given value of y^0 , the values of c_i and d_j are calculated by iterations, and the standard deviation of the fit is evaluated with $f = N - p - q$ degrees of freedom. This is repeated for variable values of y^0 , and the optimum value of y^0 is searched for which SD is at minimum.

$$y_{\text{new}}^0 = (\sum_{ij} w_{ij} y_{ij} - \sum_{ij} w_{ij} c_i d_j) / (10 - 9 y_{\text{old}}^0) \quad (20)$$

No matter how c_i and d_j have been calculated, their values are never unambiguous. Each c_i can be multiplied by an arbitrary constant C , provided that each d_j is divided by the same constant.

Certain data matrices are inherently symmetrical, i.e., $y_{ij} = y_{ji}$, see for instance Table 2 where the substituents X and Y are the same. The number of independent data is then reduced: it cannot be more than $N = p(p + 1)/2$ for a complete matrix. For the parameters of eqs 2c or 4c the

constraint $c_i = d_i$ must hold. This is best incorporated into the iteration program after each step. The constant C is cancelled out both in the case of eqs 2c and 4c, so that the values of c_i are obtained unambiguously. The standard deviations are to be calculated with $f = N - p$ degrees of freedom in the case of eq 2c, and with $f = N - p - 1$ in the case of eq 4c. As a good approximation, one can use the common program also for a symmetrical matrix when each nondiagonal element is input twice and the constraint $c_i = d_i$ is realized additionally. The only shortcoming is then that the diagonal elements are given half weight. This is usually not important and may be fully corrected in the values w_{ij} when these are used.

REFERENCES AND NOTES

- (1) Exner, O. *Correlation Analysis of Chemical Data*; Plenum Press: New York, 1988; Chapters 3.1 and 6.3.
- (2) "Wie denn dem menschlichen Geiste keine andere Form, unter welcher die gemeinschaftliche Wirkung verschiedener Elemente stattfinden kann, vorstellbar ist, als die der Summe", Ostwald, W. Über die Affinitätsgrößen organischer Säuren und ihre Beziehungen zur Zusammensetzung und Konstitution derselben. *Z. Physik. Chem.* **1889**, *3*, 369–422.
- (3) Hückel, W. Die Chemische Bindung. Kritische Betrachtung der Systematik, der Ausdruckweisen und der formelmässigen Darstellung. *J. Prakt. Chem.* **1957**, *277*, 105–208.
- (4) Jaffé, H. H. A Reexamination of the Hammett Equation. *Chem. Rev.* **1953**, *53*, 191–261.
- (5) Kevill, D. N.; D'Souza, M. J. Improved Procedures for Determining Second-derivative Cross-Interaction Coefficients and for Establishing a Relationship between the Values and Reaction Type. *J. Chem. Soc., Perkin Trans. 2* **1994**, 945–948.
- (6) Exner, O. The Non-additive Substituent Effects of Two Substituents Adjacent to the Reaction Centre. *Collect. Czech. Chem. Commun.* **1976**, *41*, 1516–1540.
- (7) (a) Shorter, J.; Stubbs, F. J. The Additive Effect of Substituents on the Strength of Benzoic Acids. *J. Chem. Soc.* **1949**, 1180–1183. (b) Kalfus, K.; Kroupa, J.; Večera, M.; Exner, O. Additivity of Substituent Effects in *meta*- and *para*-Substituted Benzoic Acids. *Collect. Czech. Chem. Commun.* **1975**, *40*, 3009–3019.
- (8) (a) Seth-Paul, W. A.; Van Duyse, A. A new Substituent Constant Derived from Carbonyl Stretching Frequencies of Simple R'R''CO Molecules. *Spectrochim. Acta* **1972**, *28A*, 211–234. (b) Pascual, C.; Meier, J.; Simon, W. Regel zur Abschätzung der chemischen Verschiebungen von Protonen an einer Doppelbindung. *Helv. Chim. Acta* **1966**, *49*, 164–168. (c) Koppel, I. A.; Taft, R. W.; Anvia, F.; Zhu, S.-Z.; Hu, L.-Q.; Sung, K.-S.; DesMarteau, D. D.; Yagupol'skii, L. M.; Yagupol'skii, Yu. L.; Ignat'ev, N. V.; Kondratenko, N. V.; Volkonskii, A. Yu.; Vlasov, V. M.; Notario, R.; Maria, P. C. The Gas-Phase Acidities of Very Strong Neutral Brønsted Acids. *J. Am. Chem. Soc.* **1994**, *116*, 3047–3057.
- (9) Equation 4a was anticipated in some qualitative statements, e.g., Skrabal, A. Reaktionsgeschwindigkeit-Temperatur-Studien. I. Die Grösse der Temperaturabhängigkeit der Reaktionsgeschwindigkeit. *Monatsh. Chem.* **1916**, *37*, 495–534. In a mathematical form, it appeared probably for the first time in Willpate-Steinert, L.; Fierens, P. J. C.; Hannaert, H. Quelques remarques relatives aux équations linéaires de Hammett et de Winstein et Grunwald. *Bull. Soc. Chim. Belg.* **1955**, *64*, 628–634. It was mathematically analyzed by Miller, ref. 11, and later, refs 12 and 14.
- (10) Eq 4c was used probably first by Exner, O.; Horák, M. Infrarotspektroskopie von Carbonsäurederivaten. *Chem. Listy* **1958**, *52*, 1451–1460; *Collect. Czech. Chem. Commun.* **1959**, *24*, 968–977. It was analyzed in ref 6, 13.
- (11) Miller, S. I. Multiple Variation in Structure-Reactivity Correlations. *J. Am. Chem. Soc.* **1959**, *81*, 101–106.
- (12) Palm, V. A.; Istomin, B. I. Isoparametric Relations. Generalization of Isokinetic and Analogous Relations. *Reakts. Sposonost. Org. Soedin. (Tartu)* **1969**, *6*, 427–441.
- (13) Abraham, M. H.; Grellier, P. L.; Prior, D. V.; Morris, J. J.; Taylor, P. J. Hydrogen Bonding. Part 10. A Scale of Solute Hydrogen-bond Basicity using log K Values for Complexation in Tetrachloromethane. *J. Chem. Soc., Perkin Trans. 2* **1990**, 521–529.
- (14) Dubois, J.-E.; Ruasse, M.-F.; Argille, A. Theoretical and Experimental Evaluation of IFER for MSE (Interactive Free Energy Relationship for Multiple-Substituent Effects). Mechanistic Significance of the Reaction Constant and Cross-Interaction Constant. *J. Am. Chem. Soc.* **1984**, *106*, 4840–4845.

- (15) (a) Exner, O. Characteristic Vibrations of the Sulfonyl Group. *Collect. Czech. Chem. Commun.* **1963**, *28*, 935–941. (b) Exner, O.; Janák, P. The Acidity of N-Acids. Substituent Effects on the Dissociation Constants of Imides and Sulphonamides. *Collect. Czech. Chem. Commun.* **1975**, *40*, 2510–2523. (c) Exner, O.; Dahn, H.; Péchy, P. Correlation of Oxygen-17 Substituent Induced Chemical Shifts: *meta*- and *para*-Substituted Methyl Benzoates. *Magn. Reson. Chem.* **1992**, *30*, 381–386.
- (16) Lee, I. Characterization of Transition States for Reactions in Solution by Cross-interaction Constants. *Chem. Soc. Rev.* **1990**, *19*, 317–333.
- (17) (a) Lee, I.; Lee, Y. S.; Lee, B.-S.; Lee, H. W. Solvolysis Mechanism of Indan-2-yl Arenesulfonates. *J. Chem. Soc., Perkin Trans. 2* **1993**, 1441–1445. (b) Lee, I.; Lee, Y. S.; Huh, C.; Lee, H. W.; Lee, B. C. Nucleophilic Substitution Reactions of Indan-2-yl Arenesulfonates with Anilines in Methanol. *J. Chem. Soc., Perkin Trans. 2* **1993**, 2415–2418. (c) Lee, I.; Lee, W. H.; Lee, H. W.; Bentley, T. W.: Nucleophilic Substitution Reaction of 1-Phenylethyl Chlorides in Methanol. *J. Chem. Soc., Perkin Trans. 2* **1993**, 141–146.
- (18) Exner, O. A Critical Comparison of Substituent Constants. In *Correlation Analysis in Chemistry*; Chapman, N. B., Shorter, J., Eds.; Plenum Press: New York, 1978; pp 439–540.
- (19) (a) Lias, S. G.; Bartmess, J. E.; Liebman, J. F.; Holmes, J. L.; Levin, R. D.; Mallard, G. W. Gas-phase Ion and Neutral Thermochemistry. *J. Phys. Chem. Ref. Data* **1988**, *17*, Suppl. 1, 1–872. (b) Lias, S. G.; Liebman, J. F.; Levin, R. D. Evaluated Gas Phase Basicities and Proton Affinities of Molecules. *J. Phys. Chem. Ref. Data* **1984**, *13*, 695–808.
- (20) Bordwell, F. G. Equilibrium Acidities in Dimethyl Sulfoxide Solution. *Acc. Chem. Res.* **1988**, *21*, 456–463.
- (21) Similar distribution was found for the isokinetic temperature which can be also understood as abscissa of the common point of intersection of several regression lines: Exner, O. Statistics of the Enthalpy-Entropy Relationship. I. The Special Case. *Collect. Czech. Chem. Commun.* **1972**, *37*, 1425–1444.
- (22) Exner, O. The Reactivity-Selectivity Principle: Redefinition and Reexamination on an Experimental Basis. *J. Chem. Soc., Perkin Trans. 2* **1993**, 973–979.
- (23) Exner, O. Zur Frage der Additivität in der Gleichungen nach Hammett und Taft. *Collect. Czech. Chem. Commun.* **1960**, *25*, 1044–1051 (note added in proof).
- (24) (a) Istomin, B. I.; Pivovarov, S. A.; Selivanov, V. F.; Gidasov, B. V.; Istomina, S. N. The Isoparametrical Effect and Multiple Cross Correlations. *Org. Reactivity (Tartu)* **1975**, *12*, 287–298. (b) Kevill, D. N.; D'Souza, M. J. The Relevance of Third-derivative Cross-interaction Coefficients in Hammett-type Treatments of Nucleophilic Substitution Reactions. *J. Chem. Soc., Perkin Trans. 2* **1994**, 2427–2430.
- (25) Exner, O. Additive and Non-additive Relationships with Three Variables. *Org. Reactivity (Tartu)* **1977**, *14*, 353–362.
- (26) (a) Lee, I.; Kim, H. Y.; Kang, H. K.; Lee, H. V. Nucleophilic Substitution Reactions of 1-Phenylethyl Benzenesulfonates with Anilines in Methanol-Acetonitrile. *J. Org. Chem.* **1988**, *53*, 2678–2683. (b) Lee, I.; Shim, C. S.; Chung, S. Y.; Lee, H. W. Nucleophilic Substitution Reactions of Phenacyl Benzenesulfonates with Anilines in Methanol-Acetonitrile Mixtures. *J. Chem. Soc., Perkin Trans. 2* **1988**, 975–981.
- (27) Low efficiency of the procedure is seen in ref 24b clearly from the values of correlation coefficients, sometimes differing very little for the regressions with three, six, or seven terms, respectively. Note that correlation coefficient does not reflect the degrees of freedom.
- (28) A systematic error was suspected (ref 24b) to occur in the data of ref 17b. When six items were deliberately adjusted by 8–12% of the k value, the results changed markedly, the value $\rho_{XYZ} = -0.19 \pm 0.25$ was then similar to ours with four terms (maybe by chance). The correction of 8–12% corresponds to 0.033–0.049 in $\log k$, significantly greater than the experimental uncertainty of 0.013 (ref 17a). Probably the experimental error was somewhat underestimated; in any case this is a further proof of the numerical instability.

CI950035W