Journal of Chromatography, 550 (1991) 15–25
Elsevier Science Publishers B.V., Amsterdam

CHROMSYMP. 2224

Multiple-criteria optimization

STANLEY N. DEMING
Department of Chemistry, University of Houston, Houston, TX 77204-5641 (USA)

ABSTRACT

In many chromatographic optimization problems, it is unusual to find only one response that must be optimized. Multiple-criteria optimization methods allow the combination of several responses into a single figure-of-merit. Origins are traced to Harringtons' "desirability functions" and Zadeh's "fuzzy sets".

INTRODUCTION

In many chromatographic optimization problems, it is unusual to find only one response that needs to be optimized. Instead, there are usually several responses that must be considered. Additionally, many of the responses must be expressed as intensive properties (*i.e.*, they should not depend on the size or throughput of the system [1]) and must be normalized by one or more factors and/or responses of the system [2]. In industrial work, for example, resolution per unit time is often of greater interest than simply resolution. Similarly, cost is often usefully measured as dollars per separation.

In all of these cases, various ratios, penalties and desirabilities can be used to specify quantitative objective functions [3].

OBJECTIVE FUNCTIONS

As stated by Beveridge and Schechter [2], "The aim of optimization is the selection, out of the multiplicity of potential solutions, of that solution which is the best with respect to some well defined criterion. The choice of this criterion, the objective, is therefore an essential step in any study... In general, economic criteria should be used, although technical forms are common".

An objective function is a mathematical relationship expressing the objective in terms of system factors and/or responses. Objective functions based on overall economic strategies tend to be highly complex [2]. Objective functions based on more restricted technical and quality considerations are usually simpler.

Consider a chromatographic system with two responses: resolution and analysis time. If the system is to be optimized, the question arises, of what the objective of the separation process is. If the separation is to be presented to an academic audience, the objective might be to make the resolution as high as possible and ignore the analysis

0021-9673/91/$03.50

time. However, if the separation is to be presented to industrial process control engineers, the objective might be to minimize the analysis time and not be purists about resolution. The sagacious laboratory manager will recognize that improving both the resolution and the analysis time (within limits) would capture the attention of both audiences.

An objective function could be used to indicate formally just how resolution and analysis time should be combined into a single figure-of-merit to be optimized. For this example, the objective function might be simply the sum of the resolution and some inverse measure of the analysis time, or, if it is desired to emphasize the resolution, the resolution might be weighted twice as much as analysis time. There might also exist target values of either or both responses: minimizing the total deviation from these target values might be the objective.

Considerations such as this illustrate an irony about objective functions: they are highly subjective. To write a proper objective function for this chromatographic example, it would be helpful to have at hand the results of a survey that measures the relative desirabilities of both resolution and analysis time for the intended customers.

OBJECTIVE FUNCTIONS BASED ON RATIOS

Ratios are often used to construct objective functions. Although ratios are simple and attractive, they can lead to unexpected results. This is well illustrated in the valuable paper by Smits *et al.* [4]; the following discussion is based on their paper.

Fig. 1 shows an incomplete liquid chromatographic separation of five inorganic ions. The vertical axis represents detector signal (arbitrary units) which is proportional to the concentration of the ions in the eluent. The horizontal axis represents the time (arbitrary units) after injection of the sample onto the chromatographic column. In many environments, the time required to elute the last ion is important: longer analysis

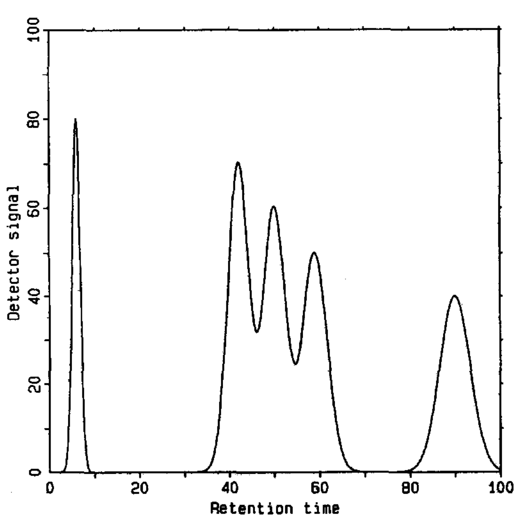


Fig. 1. Incomplete liquid chromatographic separation of five inorganic ions. Computer simulation based on ref. 4.

times mean fewer samples per day; shorter analysis times will increase the daily sample throughput, clearly an economic advantage.

The resolution/analysis time ratio could be chosen for maximization as the objective function. This seems reasonable. As the separation becomes more complete, the quantitative measure in the numerator will become larger and the objective function will become larger. As the analysis time decreases, the denominator will become smaller and again the objective function will become larger. Thus, maximizing the resolution/analysis time ratio should lead to improved separations and shorter analysis times.

Fig. 2 shows the separation that might result from the use of this optimization criterion. The objective function has been increased, but the separation of ions is now worse than when the optimization began. The denominator became small faster than the numerator became small, that is, the analysis time decreased faster than the resolution degraded. While the resolution was going from bad to worse, the analysis time was going from good to better at a faster rate. The net result was a very fast "separation" that was almost totally worthless, even though the objective function ratio (resolution/analysis time) continued to become larger.

Objective functions based on ratios must be used with caution. An alternative is to avoid ratios by basing the optimization on only one of the components (*e.g.*, resolution) and establishing a threshold and penalty function for the other component (*e.g.*, analysis time). Another alternative is to combine multiple responses into a single measure of performance that expresses the desirability of each combination.

OBJECTIVE FUNCTIONS BASED ON PENALTY FUNCTIONS

Practical considerations of sample throughput (*e.g.*, analysis per day) often dictate a maximum permissible analysis time. If an analytical laboratory must carry

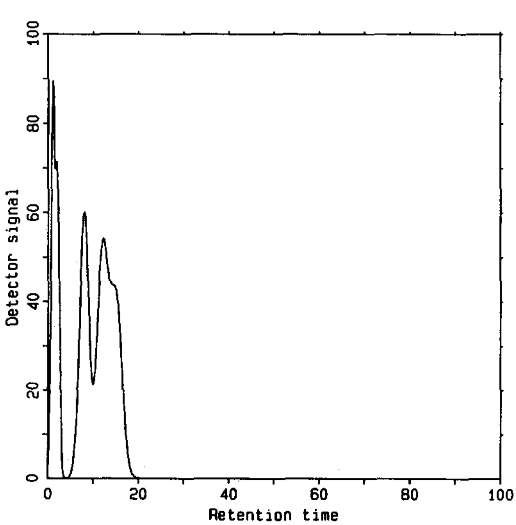


Fig. 2. Results of optimization driven by the maximization of the resolution/analysis time ratio. Computer simulation based on ref. 4.

out fifteen analyses in an 8-h day, then simple calculation suggests a maximum analysis time of *ca.* 30 min [5]. This 30-min maximum analysis time can be considered to be a threshold value: an analysis time of less than 30 min might be desirable but would not be especially beneficial, whereas an analysis time greater than 30 min would be undesirable, perhaps critically undesirable. Thus, an analysis time less than the threshold might not figure in any objective function calculations, but analysis times greater than the threshold should be taken into account: the objective function should be penalized if the analysis time exceeds the 30-min threshold.

Assuming the threshold represents an upper limit, penalty functions can be expressed mathematically is

$$\begin{aligned} p &= 0 && \text{for } y_j \leqslant y_{jt} \\ &= g(y_j - y_{jt}) && \text{for } y_j > y_{jt} \end{aligned} \tag{1}$$

where p is the value of the penalty and y_{jt} represents the threshold value associated with the response y_j. The nature of $g(y_j - y_{jt})$ is subjective but usually follows one of three well defined forms illustrated in Fig. 3. (Similar equations and figures apply to threshold values representing a lower limit.)

The first type of penalty function is an "infinite wall" illustrated at the top of Fig. 3: $g(y_j - y_{jt}) = -\infty$. Thus, violations of the threshold are considered to be infinitely bad. This type of penalty function is usually used for critical responses (those involving safety, for example).

A second type of penalty function is illustrated in the middle of Fig. 3: $g(y_j - y_{jt}) = b_j(y_j - y_{jt})$, where b_j is a slope or proportionality constant expressing the severity of the penalty ($b_j = -\infty$ is equivalent to the "infinite wall"; $b_j = 0$ is equivalent to no penalty). As the response becomes further away from the threshold value, the penalty becomes proportionally more severe. Again, the choice of b_j is often subjective.

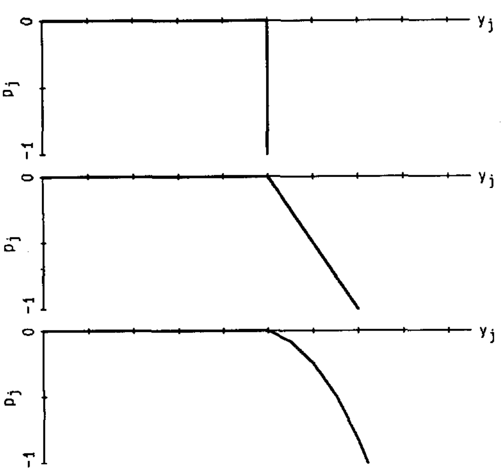


Fig. 3. Possible penalty functions for $y_j > y_{jt}$.

The third type of penalty function is illustrated at the bottom of Fig. 3: $g(y_j - y_{jt}) = b_j(y_j - y_{jt})^n$, where n is usually $\geqslant 2$. This is a power function that expresses the idea that large violations of the threshold value are much more serious than small violations. The choices of both b_j and n are subjective. This is probably the most generally useful type of penalty function and is widely used in many areas.

DESIRABILITY FUNCTIONS

It was suggested earlier that an objective function could be used that might improve both the resolution and the analysis time. But how can the "apples" of resolution and the "oranges" of analysis time be combined? Harrington [6] states the problem well: "In nearly all situations requiring human judgement, one is faced with a multiplicity of measures which must be balanced one against the other, weighted in accordance with their relative importance, compromised where these measures are mutually opposing, and variously manipulated to achieve an optimum judgement... If by some means the several properties could be measured in consistent units, or, even better, could be expressed as numbers on a dimensionless scale, then the arithmetic operations intended to combine these measures becomes feasible". Although Harrington proposed two specific forms for the "desirability function", the concepts are general and can be merged with concepts from Zadeh's field of fuzzy logic [7–11] to yield useful objective functions for optimization.

Lowe [12] proposed a simple procedure for forming desirabilities from multiple responses. If y_{ju} and y_{jd} are measures of the most undesirable and most desirable values, respectively, of a response y_j, and if it is assumed that the desirability increases linearly on going from y_{ju} to y_{jd}, then the desirability contributed by this response is calculated as

$$d_j = 0 \qquad \text{for } y_j < y_{ju}$$

$$d_j = 1 \qquad \text{for } y_j > y_{jd}$$

$$d_j = (y_j - y_{ju})/(y_{jd} - y_{ju}) \qquad \text{for } y_{ju} \leqslant y_j \leqslant y_{jd} \tag{2}$$

where "$<$" and "$>$" represent "worse than" and "better than", respectively. Note that d_j is unitless and ranges from 0 to 1.

The concept is illustrated in Fig. 4. Along the left-hand side at the top of the figure is a desirability axis ranging from 0 (undesirable) to 1 (desirable). Along the bottom of the figure are drawn five response axes, $y_1 - y_5$ (*e.g.*, resolution, separation time, cost). The response axes have undergone zero suppression and scale expansion so that their most undesirable values are aligned vertically with the left-hand side of the figure and their most desirable values are aligned vertically with the right-hand side of the figure.

Running diagonally across Fig. 4 from left to right is a transformation line that maps values of response onto values of desirability. This line is used by reading upward from a given value of response and leftward to the corresponding values of desirability. For example, a response value of $y_5 = 4.0$ corresponds to a desirability value of $d_5 = 0.57$. Similarly, $y_4 = 15$ becomes $d_4 = 0.85$ and $y_1 = 1.2$ becomes $d_1 = 0.15$. These

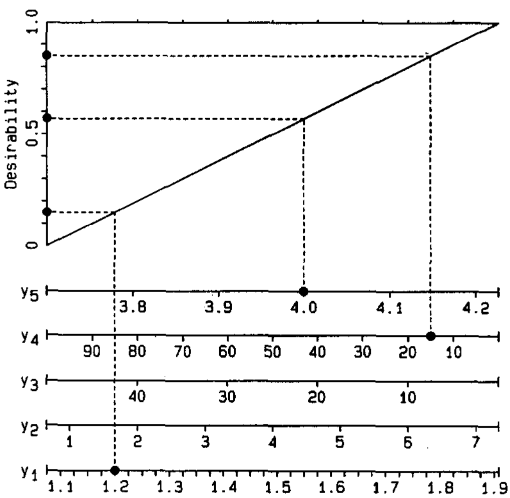


Fig. 4. Desirability as a first-order function of response. Undesirable responses at the left; desirable responses at the right.

results obtained graphically are identical with those obtained using eqn. 2:

$$d_5 = (4.00 - 3.70)/(4.23 - 3.70) = 0.57$$

$$d_4 = (15 - 100)/(0 - 100) = 0.85$$

$$d_1 = (1.2 - 1.075)/(1.908 - 1.075) = 0.15 \tag{3}$$

Responses that lie to the right of the response ranges shown in Fig. 4 would be assigned desirabilities of 1.00; responses to the left of the figure would be assigned desirabilities of 0.00.

Harrington's desirability functions [6] do not assume Lowe's [12] linear (first-order) relationship between response and desirability. Harrington's two-sided desirability function is given by

$$d_j = \exp[-(|y'_j|)^n] \tag{4}$$

where n is a positive number ($0 < n < \infty$, not necessarily integral), y'_j is a linear transform of the response variable, y_j, such that $y'_j = -1$ when y_j is equal to the lower specification limit, y_{j^-}, and $y'_j = +1$ when y_j is equal to the upper specification limit, y_{j^+}, and $|y'_j|$ is the absolute value of y'_j (the use of upper and lower specification limits comes from concerns about product quality). Any particular value of response, y_j, may be transformed into the corresponding y'_j by the relationship

$$y'_j = [y_j - (y_{j^+} + y_{j^-})/2]/[(y_{j^+} - y_{j^-})/2]$$

$$= [2y_j - (y_{j^+} + y_{j^-})]/(y_{j^+} - y_{j^-}) \tag{5}$$

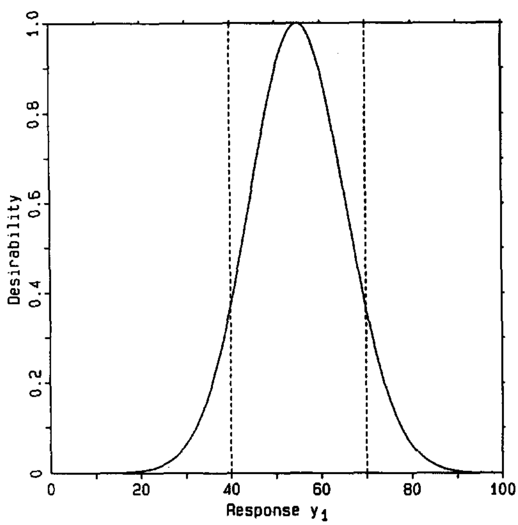


Fig. 5. Harrington's two-sided desirability function for $n = 2$.

which measures the distance of y_j from the midpoint between the upper and lower specification limits, $[(y_{j^+} + y_{j^-})/2]$, in units equal to half the spread between the upper and lower specification limits, $[(y_{j^+} - y_{j^-})/2]$. Fig. 5 illustrates this two-side desirability function for $n = 2$.

For one-sided specification limits a special form of the Gompertz growth curve is used:

$$d_j = \exp\{-[\exp(-y_j')]\} \tag{6}$$

where $y_j' = 0$ at the single spcification limit. The mapping of y_j onto y_j' is accomplished by choosing two ordered pairs of (y_j, d_j) and calculating $y_j' = -\ln[-\ln(d_j)]$. From the resulting ordered pairs of (y_j, y_j'), the straight-line equation

$$y_j' = b_0 + b_1 y_j \tag{7}$$

can be obtained, where b_0 is the intercept and b_1 is the slope. Fig. 6 illustrates this one-sided desirability function for the ordered pairs (40.0, 0.37) and (70.0, 0.90).

These desirability functions are well suited to multiple-criteria optimization work, but many alternative forms are possible. Some of the most useful versions of desirability functions are [free-form] graphical versions such as those shown in Figs. 7 and 8. Derringer and Suich [13] gave examples.

OVERALL DESIRABILITIES

There are many ways in which the individual desirabilities $d_1 - d_n$ can be combined. A simple arithmetic average is one example. However, as Harrington [6] pointed out, in any realistic situation a "basic premise is this—if any one property is so

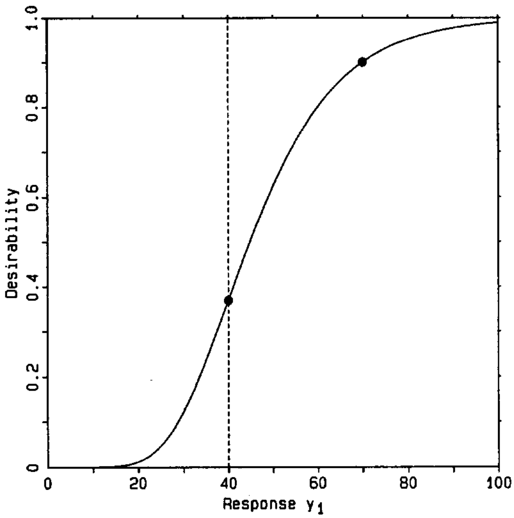
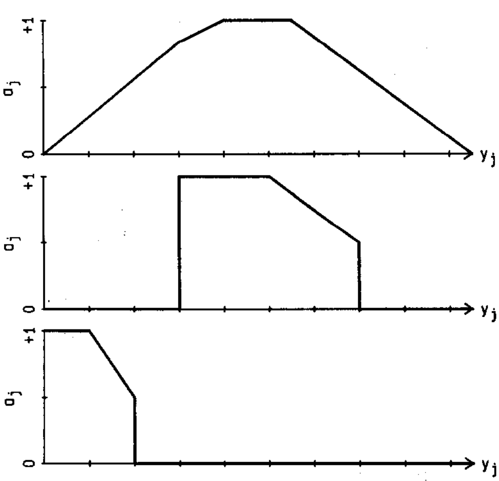


Fig. 6. Harrington's one-sided desirability function for the ordered pairs (40.0, 0.37) and (70.0, 0.90).

poor that the product is not suitable to the application, that product will not be acceptable, *regardless of the remaining properties*... customer reaction to a product is based very largely on the less desirable properties of that product because these are the focus of potential trouble".

The mathematical model analogous to these psychological reactions is the geometric mean of the component d values, or

$$D = (d_1 d_2 \ldots d_n)^{1/n} \tag{8}$$



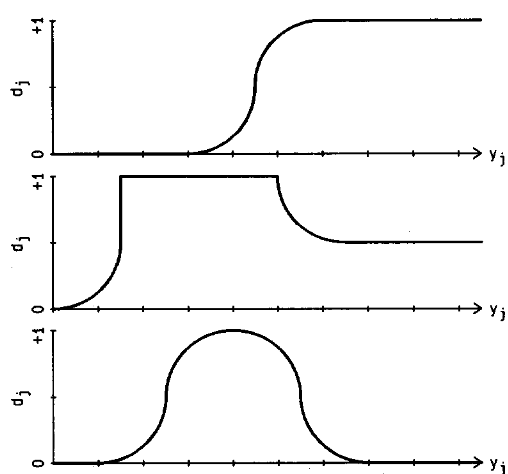


Fig. 7. Free-form desirability functions constructed from straight-line segments.

Fig. 8. Free-form desirability functions with curvature.

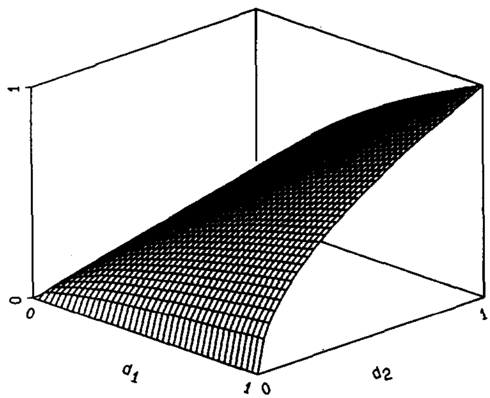


Fig. 9. Illustration of how D (the overall desirability) varies as a function of two d_j values according to eqn. 8.

where D is the overall desirability. It is clear that of any d_j is zero, the associated D will also be zero. Further, D is strongly weighted by the smaller d_j values.

Fig. 9 shows how D varies as a function of two d_j values. The nth root (square root) relationship is clear in this representation. Note again that if either d_1 or d_2 goes to zero, D is zero regardless of the value of the other d.

Fig. 10 shows individual desirabilities, d_1 and d_2, as functions of two responses, y_1 and y_2. Mapping these desirabilities through eqn. 8 gives Fig. 11, which shows how the overall desirability D is affected by the individual responses, y_1 and y_2. Figs. 12 and 13 suggest that more complicated mappings of responses onto desirabilities give rise to more complicated desirability surfaces that might contain multiple optima.

GENERAL COMMENTS

The ultimate mapping would be to show D as a function of the system factors [6], but to do so presumes a knowledge of the relationships between each y_j and all x_1 values. However, because these relationships are not usually known at the beginning of a separation project, such mappings are not usually possible initially.

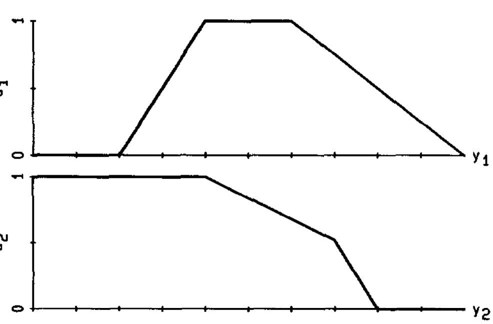


Fig. 10. Individual desirabilities, d_1 and d_2, as functions of two responses, y_1 and y_2.

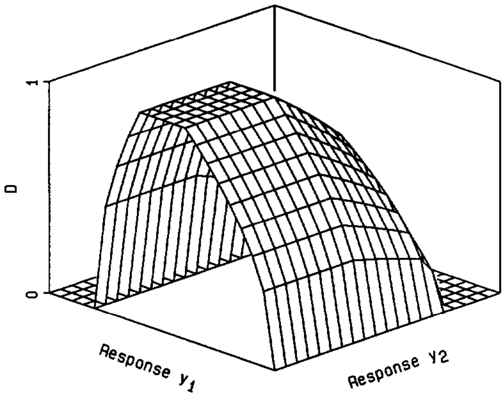


Fig. 11. Overall desirability, D, plotted as a function of the individual responses, y_1 and y_2, mapped through eqn. 8 using the individual desirabilities, d_1 and d_2, shown in Fig. 10.

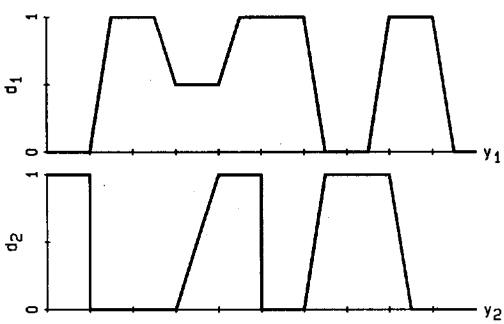


Fig. 12. Polymodal individual desirabilities, d_1 and d_2, as functions of two responses, y_1 and y_2.

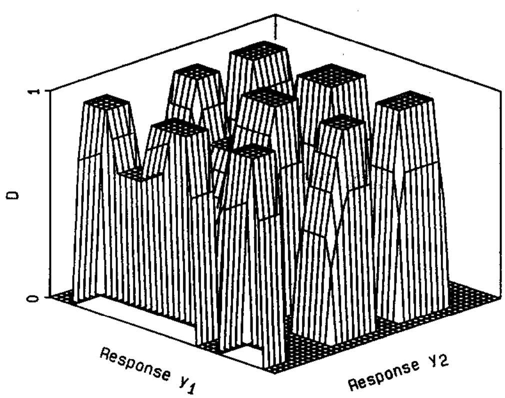


Fig. 13. Overall desirability, D, plotted as a function of the individual responses, y_1 and y_2, mapped through eqn. 8 using the individual desirabilities, d_1 and d_2, shown in Fig. 12.

Desirability functions have been used before in separation science to improve the quality of separations. The work of Glajch and Snyder [14], Laub and Purnell [15], Glajch *et al.* [16], Sachok *et al.* [17], Morgan and Jacques [18], Deming *et al.* [19], Otto and Wegscheider [20,21] and Cela *et al.* [22] may be consulted for examples.

ACKNOWLEDGEMENTS

This work was supported in part by Grant No. 003652-108 from the Applied Technology Program, Texas Higher Education Coordinating Board, Division of Research Programs.

REFERENCES

1 S. N. Deming and S. L. Morgan, *Experimental Design: A Chemometric Approach*, Elsevier, Amsterdam, 1987.
2 G. S. G. Beveridge and R. S. Schechter, *Optimization: Theory and Practice*, McGraw-Hill, New York, 1970.
3 R. E. Steuer, *Multiple Criteria Optimization: Theory, Computation, and Application*, Wiley, New York, 1986.
4 R. Smits, C. Vanroelen and D. L. Massart, *Fresenius' Z. Anal. Chem.*, 273 (1975) 1.
5 S. L. Morgan and S. N. Deming, *J. Chromatogr.*, 112 (1975) 267.
6 E. C. Harrington, Jr., *Ind. Qual. Control*, 21 (1965) 494.
7 L. A. Zadeh, *Inform. Control*, 8 (1965) 338.
8 L. A. Zadeh, *CHEMTECH*, 17 (1987) 340.
9 A. Kandel, *Fuzzy Mathematical Techniques with Applications*, Addison-Wesley, Reading, PA, 1986.
10 A. Kaufmann and M. M. Gupta, *Introduction to Fuzzy Arithmetic: Theory and Applications*, Van Nostrand Reinhold, New York, 1985.
11 C. V. Negoita, *Expert Systems and Fuzzy Systems*, Benjamin/Cummings, Menlo Park, CA, 1985.
12 C. W. Lowe, *Trans. Inst. Chem. Eng.*, 45 (1967) T3.
13 G. Derringer and R. Suich, *J. Qual. Technol.*, 12 (1980) 214.
14 J. L. Glajch and L. R. Snyder, *Computer-Assisted Method Development for High-Performance Liquid Chromatography*, Elsevier, Amsterdam, 1990.
15 R. J. Laub and J. H. Purnell, *J. Chromatogr.*, 112 (1975) 71.
16 J. L. Glajch, J. J. Kirkland, K. M. Squire and J. M. Minor, *J. Chromatogr.*, 199 (1980) 57.
17 B. Sachok, J. J. Stranahan and S. N. Deming, *Anal. Chem.*, 53 (1981) 70.
18 S. L. Morgan and C. A. Jacques, *J. Chromatogr. Sci.*, 16 (1978) 500.
19 S. N. Deming, J. M. Palasota, J. Lee and L. Sun, *J. Chromatogr.*, 485 (1989) 15.
20 M. Otto and W. Wegscheider, *J. Chromatogr.*, 258 (1983) 11.
21 M. Otto, *Chemometr. Intell. Lab. Syst.*, 4 (1988) 101.
22 R. Cela, C. G. Barroso and J. A. Pérez-Bustamante, *J. Chromatogr.*, 485 (1989) 477.