1.

Simultaneous Optimization as a Method to Establish Generalized Functional Forms for Empirical Equations of State[1]

R. Span,[2,3] H.-J. Collmann,[2] and W. Wagner[2]

---

[1] Paper presented at the Thirteenth Symposium on Thermophysical Properties, June 22-27, 1997, Boulder, Colorado, U.S.A.

[2] Lehrstuhl für Thermodynamik, Ruhr-Universität Bochum, D-44780 Bochum, Germany

[3] To whom correspondence should be addressed

ABSTRACT

A new algorithm for the optimization of functional forms of empirical equations of state is presented which considers data sets of different substances simultaneously. In this way, functional forms for empirical equations of state can be developed which yield, on average, the best representation of the thermodynamic properties of all substances within larger groups of substances (e.g. "nonpolar" and "polar" substances). The new algorithm is being used to develop a new class of empirical equations of state which meet typical technical requirements on the accuracy of thermodynamic properties with only about ten fittable coefficients. First results for nonpolar fluids are reported.

KEY WORDS: equations of state; development of empirical equations of state; optimization algorithm; nonpolar fluids; methane; argon; oxygen.

1. INTRODUCTION

Since about 1980, the efficiency of highly accurate wide-range equations of state has improved tremendously. These improvements are closely linked to the development of sophisticated optimization algorithms [1-4] which determine the most suitable mathematical form of an equation of state by selecting the best combination of terms[1] from a so-called "bank of terms"; such a bank of terms is an extensive mathematical set up which contains all terms considered to be useful for the description of the problem. In comparison to functional forms established by trial and error, optimized equations of state need about 30% to 50% fewer terms to achieve the same accuracy and they yield better results when extrapolating either beyond the limits of the data set used to set up the equation [5] or into regions where no reliable data have been available.

Unfortunately, extensive data sets are needed to optimize the mathematical structure of an equation of state and when such optimized equations are fitted to data of another substance they often lose their advantages since their functional form is constrained to the substance they were developed for. Therefore, optimized empirical equations of state have so far only been been available for about 20 substances.

The new optimization algorithm presented here has been developed to overcome these limitations. The simultaneous use of data sets of different substances in an optimization

---

[1] Empirical equations of state are usually formulated as a summation over different mathematical expressions with one fittable coefficient each. The word "term" refers to one of these mathematical expressions.

algorithm results in the functional form, which yields on average the best results for all considered substances. If the considered substances are representative for a group of substances (like e.g. the group of "nonpolar" fluids), the resulting formulation can be fitted to limited data sets of other substances out of this group without significantly reducing accuracy. Currently the new simultaneous optimization procedure is used to establish a new class of simple equations of state with about ten fitted coefficients, which are designed to describe broad groups of substances accurately enough for most technical applications. First results from this project are given to illustrate the potential of simultaneous optimization algorithms.

2. THE SIMULTANEOUS OPTIMIZATION ALGORITHM

The simultaneous optimization algorithm presented here is essentially based on the optimization algorithm by Setzmann and Wagner [4]. This algorithm combines deterministic elements from the well known stepwise regression analysis [1], such as adding, deleting and exchanging terms, with elements from the evolutionary optimization method [3], like mutation and optimization of a "population" of equations. In order to shorten the description of the new algorithm, the paper of Setzmann and Wagner [4] is referred to wherever elements of the procedure remained unchanged. The presented paper deals only with the application of simultaneous optimization to empirical equations of state; nevertheless, the new algorithm can be applied to completely different problems as well.

For the optimization algorithm of Setzmann and Wagner [4] and for the new simultaneous optimization algorithm Fig. 1 shows the main steps of the development of an equation of

state. Both with respect to data and to constraints, the construction of the regression matrix is completely identical with the procedure described in [4]. The residuals which are necessary to set up a regression matrix for an equation of state in form of the reduced Helmholtz energy are given e.g. in [6,7]. Examples for typical bank of terms are given in [6-8]. For simultaneous optimization, the variables in the bank of terms should be reduced with the critical temperature T_c and the critical density ρ_c of the corresponding substance in order to make use of a simple corresponding states similarity for the functional form of the equation. Regression matrixes which are set up for simultaneous use have to be based on the same bank of terms since the optimization algorithm identifies terms only by their position in the bank of terms.

Figure 2 shows a flowchart of the optimization algorithm of Setzmann and Wagner as given in [4]. The general structure of this procedure remains unchanged. In step 2, the initial set of equations in the "population" is determined by repeated random selection of terms. In this step, a quality criterion is needed in order to determine the best formulations[2] which form the starting population. In the algorithm of Setzmann and Wagner, this quality criterion is the sum of squares c^2 which can be calculated from the regression matrix considering only those terms from the bank of terms which are part of the current formulation. In the

[2] Here the expression "formulation" refers to the combination of terms which is considered in the current stadium of the optimization process. The best formulation becomes the functional form of an equation of state when the results are written to a file after the optimization procedure is finished.

simultaneous optimization procedure, the corresponding sum of squares c_i^2 has to be calculated from each of the *I* regression matrixes used. Equivalent to c^2 the quality criterion could now be defined as

$$\mathsf{C}^2 = \sum_{i=1}^{I} \mathsf{c}_i^2 . \tag{1}$$

In Eq. (1), extensive data sets of well measured substances which result in high values for c_i^2 have an exaggerated influence on the quality criterion even if the representation of the data is satisfactory. To avoid this problem, in the simultaneous optimization the sums of squares c_i^2 are reduced by reference sums of squares, $\mathsf{c}_{o,i}^2$, resulting from equations of state of the same length which are optimized individually for the corresponding substance. Thus, in the simultaneous optimization the reduced sum of squares of the substance *i*, $\mathsf{c}_i^{*2} = \mathsf{c}_i^2 / \mathsf{c}_{o,i}^2$, becomes one if the current formulation describes the data set as well as the equation of state optimized individually. With these reduced sums of squares, the quality criterion of the simultaneous optimization, C^{*2}, is defined as

$$\mathsf{C}^{*2} = \sum_{i=1}^{I} \mathsf{c}_i^{*2} = \sum_{i=1}^{I} \mathsf{c}_i^2 / \mathsf{c}_{o,i}^2 . \tag{2}$$

The formulations chosen in the initialization process, step 2 in Fig. 2, are those with the minimum values for C^{*2}.

When the starting population is determined, the process continues with step 3, "mutation". In this step, terms from the current formulations are randomly exchanged for other terms from the bank of terms (see [4]). The quality criterion of the "mutant" is determined in the

same way as described above. The old formulation is replaced by its mutant if $C^{*2}_{\text{mutant}} < C^{*2}_{\text{old}}$.

In step 4 the modified regression analysis incorporated in the algorithm of Setzmann and Wagner [4] starts with the selection of the initial terms, which has not been changed in the new procedure. In order to add the most important term in step 6, the quality criterion has to be calculated for all *J* terms from the bank of terms which have not yet been included in the current formulation. When adding the term *j*, the sum of squares c_i^2 resulting from the regression matrix *i* becomes

$$c^2_{i,\text{new},j} = c^2_{i,\text{old}} - \frac{b^2_{Lj,i}}{b_{jj,i}}, \tag{3}$$

where *b* corresponds to an element of the two-dimensional regression matrix and *L* is the bottom row of the matrix (see [4]). If all *I* values $c^2_{i,\text{new},j}$ are determined, the quality criterion for adding the term *j*, $C^{*2}_{\text{new},j}$ can be calculated according to Eq. (2). The term which yields the smallest value $C^{*2}_{\text{new},j}$ is actually inserted into the formulation. To do so, all *I* regression matrixes have to be transformed according to the instructions given in Table I of the paper of Setzmann and Wagner [4].

After adding a term, the mutation procedure (step 7) is repeated and statistical tests are applied to test the significance of the terms in the formulation (Student t test, see [4]) and of the entire formulation (Fisher F test, see [4]). Although these tests are formally valid only for a single data set, the average values of the statistical probabilities calculated from the *I*

considered regression matrixes can be used as criteria for the statistical tests in the simultaneous optimization procedure.

If the Student t test shows that a term in the current formulation yields no significant contribution, this term is removed by transformation of all *I* regression matrixes according to the instructions given in Table I of the paper of Setzmann and Wagner [4]. If the formulation as a whole copes with the F test, the next term is added. If not, it is tested whether the formulation can be improved by exchanging any of the terms in it for any other term from the bank of terms. Therefore, the quality criterion $C^{*2}_{\mathrm{ex},kj}$ has to be determined for every possible exchange of one of the *k* terms in the formulation (index *k*) against one of the *J* terms (index *j*) which are not contained in the formulation. The required values $c^{*2}_{\mathrm{ex},kj,i}$ can be calculated according to Table II in [4]. The exchange with the smallest value for $C^{*2}_{\mathrm{ex},kj}$ is carried out by deleting the term *k* and adding the term *j* (see above), if $C^{*2}_{\mathrm{ex},kj} < C^{*2}_{\mathrm{old}}$. If the step "exchange of a term" is successful it is repeated until C^{*2} cannot be improved further by an exchange of terms. If it is not successful, the F test is applied again to test the significance of the formulation as a whole. The next term is added if the formulation now copes with the F test. If it fails again, no further improvement is possible and the procedure continues with step 5 for the next formulation. When the predetermined number of regression runs has been carried out, the best results replace the poorest formulations in the population (see [4]) and the optimization process continues with step 3 for the next generation until either a predetermined number of generations is finished or all formulations

in the population are the same. The best formulation in the last generation is regarded as the functional form which yields on average the best representation of the I considered data sets. Constraints are dealt with as described in [4], except for the fact that the combination of removing terms and adding constraints has to be applied to all I regression matrixes.

3. FIRST RESULTS

As a first test for the new optimization algorithm simple correlation equations for properties at saturation were developed for groups of nonpolar and polar substances. For this simple application the functional forms resulting from the simultaneous optimization were superior to the functional forms resulting from the algorithm of Setzmann and Wagner [4] with regard to the average reduced sum of squares $\bar{c}^{*2} = C^{*2} / I$, where I is the number of data sets considered. This result was found for both the substances used in the simultaneous optimization and the substances used only for comparison. Thus, the functional forms resulting from simultaneous optimization show the expected advantages when they are fitted to data sets of other substances belonging to the same group.

As a second test, short equations of state were developed for argon, methane and oxygen while at the same time data sets were set up for eleven other non- or almost nonpolar fluids ranging from nitrogen to n-octane. To meet the accuracy requirements of typical technical applications, the following objectives were defined: $\Delta r/r \leq \pm 0.2\%$ to $\pm 0.3\%$ for $p \leq 30$ MPa and $T \leq 473$ K, $\Delta r/r \leq \pm 0.5\%$ for $p \leq 100$ MPa, $\Delta p/p \leq \pm 0.2\%$ to $\pm 0.3\%$ in the extended critical region, and $\Delta y/y \leq \pm 1\%$ to $\pm 2\%$ for caloric properties for $p \leq 100$ MPa, except for the extended critical region.

These objectives correspond roughly to the uncertainty of typical Bender-type [9] equations of state which describe the residual fluid behavior with 19 terms and can be met by individually optimized equations of state with eight to nine terms.

For the residual part of the reduced Helmholtz energy $\phi^{r} = f^{r} / (RT)$ of argon, methane, and oxygen the simultaneous optimization procedure resulted in the functional form

$$\begin{aligned}\phi^{r}(\tau,\delta) = {} & n_1 \delta\tau^{0.375} + n_2 \delta\tau^{0.875} + n_3 \delta^3\tau^{0.375} + n_4 \delta^8\tau^{1.000} + n_5 \delta^2\tau^{0.75} e^{-\delta} \\ & + n_6 \delta^5\tau^{1.50} e^{-\delta} + n_7 \delta\tau^{4} e^{-\delta^2} + n_8 \delta^3\tau^{15} e^{-\delta^3} + n_9 \delta^6\tau^{22} e^{-\delta^4},\end{aligned} \tag{4}$$

where δ is the reduced density ρ/ρ_c, τ is the inverse reduced temperature T_c/T, and R is the gas constant. The substance specific parameters of Eq. (4) are given in Table I. The expressions required to calculate thermodynamic properties from an equation of state formulated in ϕ^{r} can be found, for example, in [6,7]. As an example, Fig. 3 shows the uncertainty of thermal properties for methane calculated from Eq. (4). The uncertainties for argon are slightly lower and those for oxygen are slightly higher especially close to the melting line; the objectives formulated above are basically met for all three substances.

Table II gives a comparison which is based on average reduced sums of squares $\bar{\gamma}^{*2} = \Gamma^{*2} / I$ for the three substances used to develop Eq. (4) and for all nonpolar substances for which the data sets have been completed in the meantime. The reference sums of squares $\gamma_{o,i}^{2}$ result from equations with nine terms individually optimized for each of the fifteen substances. On average, the simultaneously optimized equation yields better results than the equations individually optimized for argon, methane and oxygen both for the substances used to develop the equation ($\bar{\gamma}^{*2}_{1\text{-}3,\mathrm{sim}}$ = 1.49, $\bar{\gamma}^{*2}_{1\text{-}3,\mathrm{arg}}$ = 2.41, $\bar{\gamma}^{*2}_{1\text{-}3,\mathrm{met}}$ = 2.02,

and $\bar{c}^{*2}_{1\text{-}3,oxy}$ = 6.76) and for all substances ($\bar{c}^{*2}_{1\text{-}14,sim}$ = 8.10, $\bar{c}^{*2}_{1\text{-}14,arg}$ = 10.80, $\bar{c}^{*2}_{1\text{-}14,met}$ = 21.66, and $\bar{c}^{*2}_{1\text{-}14,oxy}$ = 11.71). Nevertheless, it is obvious that argon, methane, and oxygen are not representative for the group of nonpolar fluids considered here. Until now, the best functional form with nine terms which has been optimized for a more representative set of five substances has resulted in an average reduced sum of squares of $\bar{c}^{*2}_{1\text{-}14}$= 2.55. A satisfactory description of all substances can be achieved if one more term is added; a preliminary functional form with ten terms has resulted in $\bar{c}^{*2}_{1\text{-}14}$= 1.67. Such functional forms should be suitable at least for all nonpolar pure fluids with an accentric factor of ω < ≈0.4 but the work on this topic has not yet been completed. Fitted to the same data sets, the Bender equation with 19 terms yields $\bar{c}^{*2}_{1\text{-}3,Ben}$ = 3.02 and $\bar{c}^{*2}_{1\text{-}14,Ben}$ = 2.28.

ACKNOWLEDGMENTS

The authors are indebted to the Deutsche Forschungsgemeinschaft for their financial support and to Prof. Dr. R. T Jacobsen and Dr. E. W. Lemmon, whose literature database *BIBLIO* was very helpful for setting up the required data sets.

REFERENCES

[1] W. Wagner, *Fortschr.-Ber. VDI-Z.*, Ser. 3, No. 39 (1974).

[2] K. M. de Reuck and B. Armstrong, *Cryogenics* **25:** 505 (1979).

[3] J. Ewers and W. Wagner, in *Proc. 8th Symp. Thermophys. Prop.*, J. V. Sengers, ed. (American Society of Mechanical Engineers, New York, 1982), pp. 78 - 87.

[4] U. Setzman and W. Wagner, *Int. J. Thermophys.* **10:** 1103 (1989).

[5] R. Span and W. Wagner, accepted for publication in *Int. J. Thermophys.*

[6] U. Setzmann und W. Wagner, *J. Phys. Chem. Ref. Data* **20:** 1061 (1991).

[7] R. Span und W. Wagner, *J. Phys. Chem. Ref. Data* **25:** 1509 (1996).

[8] R. T Jacobsen, R. B. Stewart, and M. Jahangiri, , *J. Phys. Chem. Ref. Data* **15:** 735 (1986).

[9] E. Bender, in *Proc. 5th Symp. Thermophys. Prop.*, C. F. Bonila, ed. (American Society of Mechanical Engineers, New York, 1970), pp. 227 - 235.

Table I. Coefficients and Substance Related Parameters of the Simultaneously Optimized Equations of State, Eq. (4), for Argon, Methane, and Oxygen.

	Argon	Methane	Oxygen
i	n_i	n_i	n_i
1	1.610871306	1.659011608	1.669490380
2	–2.634808041	–2.694111307	–2.711382338
3	0.083409612	0.082946554	0.086374356
4	0.000048431	0.000048633	0.000045263
5	0.202060214	0.211420114	0.229512622
6	–0.038855616	–0.036942268	–0.041004268
7	–0.121023480	–0.129924031	–0.123006016
8	–0.020548449	–0.019663974	–0.021244360
9	0.001466681	0.001226884	0.001396158
T_c	150.687 K	190.564 K	154.595 K
ρ_c	535.6 $kg \cdot m^{-3}$	162.66 $kg \cdot m^{-3}$	436.14 $kg \cdot m^{-3}$
R	0.2081333 $kJ \cdot kg^{-1} \cdot K^{-1}$	0.5182705 $kJ \cdot kg^{-1} \cdot K^{-1}$	0.2598382 $kJ \cdot kg^{-1} \cdot K^{-1}$

Table II. Reduced Sums of Squares c_i^{*2} Resulting from Fitting the Simultaneously Optimized Functional Form (Eq. (4)) and the Functional Forms Individually Optimized for Argon, Methan and Oxygen to Data Sets of all Nonpolar Substances Considered.

		c_i^{*2} for Functional Forms Optimized			
Substance	ω	Simultaneously	For Argon	For Methane	For Oxygen
Argon	–0.002	2.09	1[1)]	2.04	12.99
Methane	0.011	1.00	2.43	1[1)]	6.28
Oxygen	0.022	1.39	3.81	3.03	1[1)]
Average Value $\overline{c}_{1\text{-}3}^{*2}$		**1.49**	2.41	2.02	6.76
Nitrogen	0.037	1.00	1.42	2.44	6.26
Ethylene	0.087	10.84	12.84	27.57	3.05
Ethane	0.099	5.09	7.60	14.10	5.21
Propane	0.153	9.70	13.49	22.15	9.63
Isobutane	0.185	10.23	11.39	14.90	5.28
n-Butane	0.200	17.65	25.85	68.38	22.67
Cyclo-Hexane	0.209	13.05	16.81	34.06	27.95
n-Pentane	0.251	13.83	17.40	27.63	14.67
n-Hexane	0.308	14.26	17.55	33.03	22.68
n-Heptane	0.350	9.21	13.85	23.08	15.85
n-Octane	0.391	4.04	5.74	29.86	10.46
Average Value $\overline{c}_{1\text{-}14}^{*2}$		**8.10**	10.80	21.66	11.71

[1)] Reference equation used to determine c_o^2 for the corresponding substance

FIGURE CAPTIONS

Fig. 1. Main steps of the development of an equation of state using either the well known algorithms for individual opimization or the new algorithm for simultaneous optimization.

Fig. 2. Flowchart of the optimization algorithm of Setzmann and Wagner [4].

Fig. 3. Tolerance diagram for densities calculated from the simultaneously optimized equation of state, Eq. (4), applied to methane.

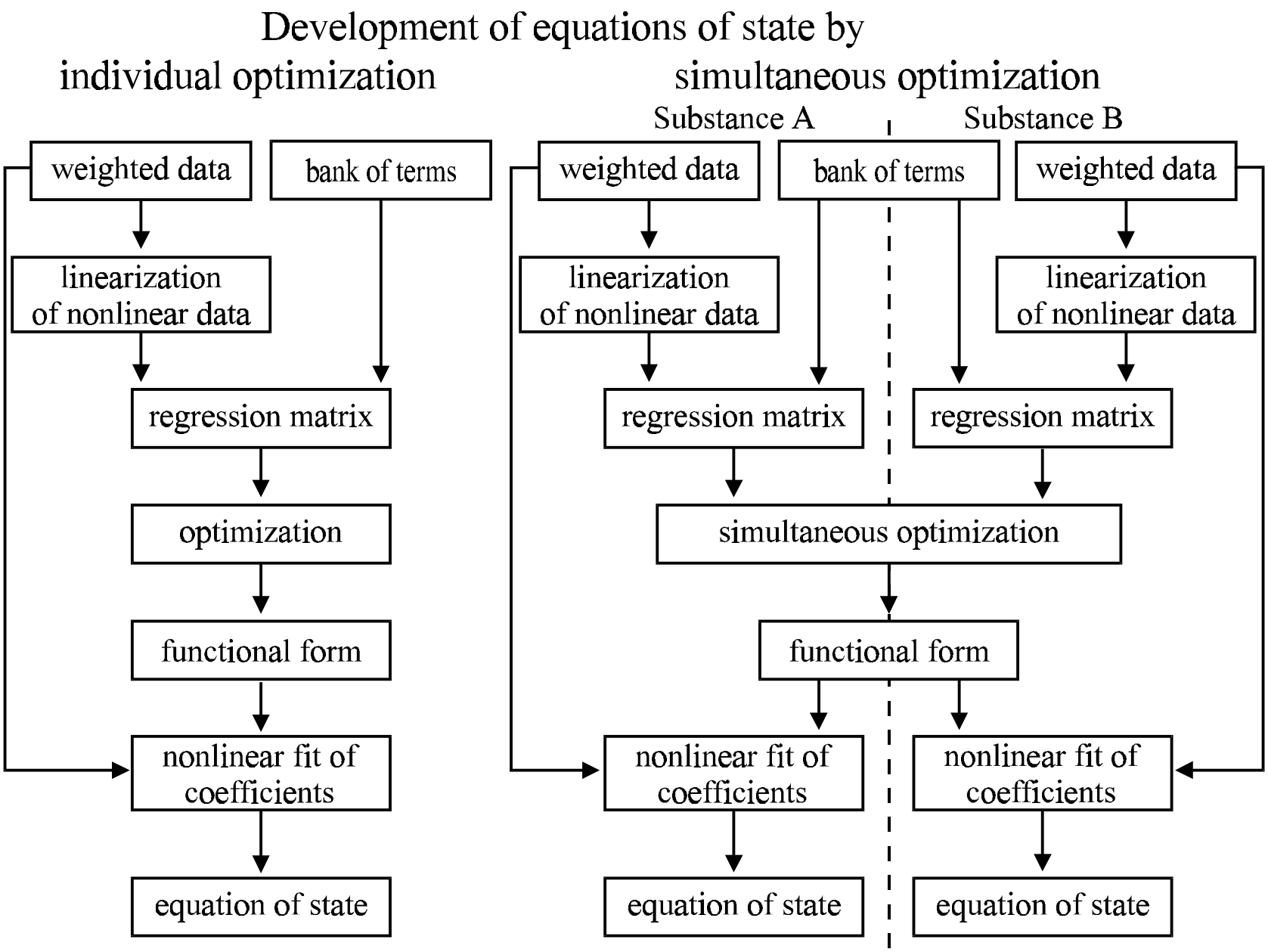

Figure 1
Span et al.
(enlarged by 200%)

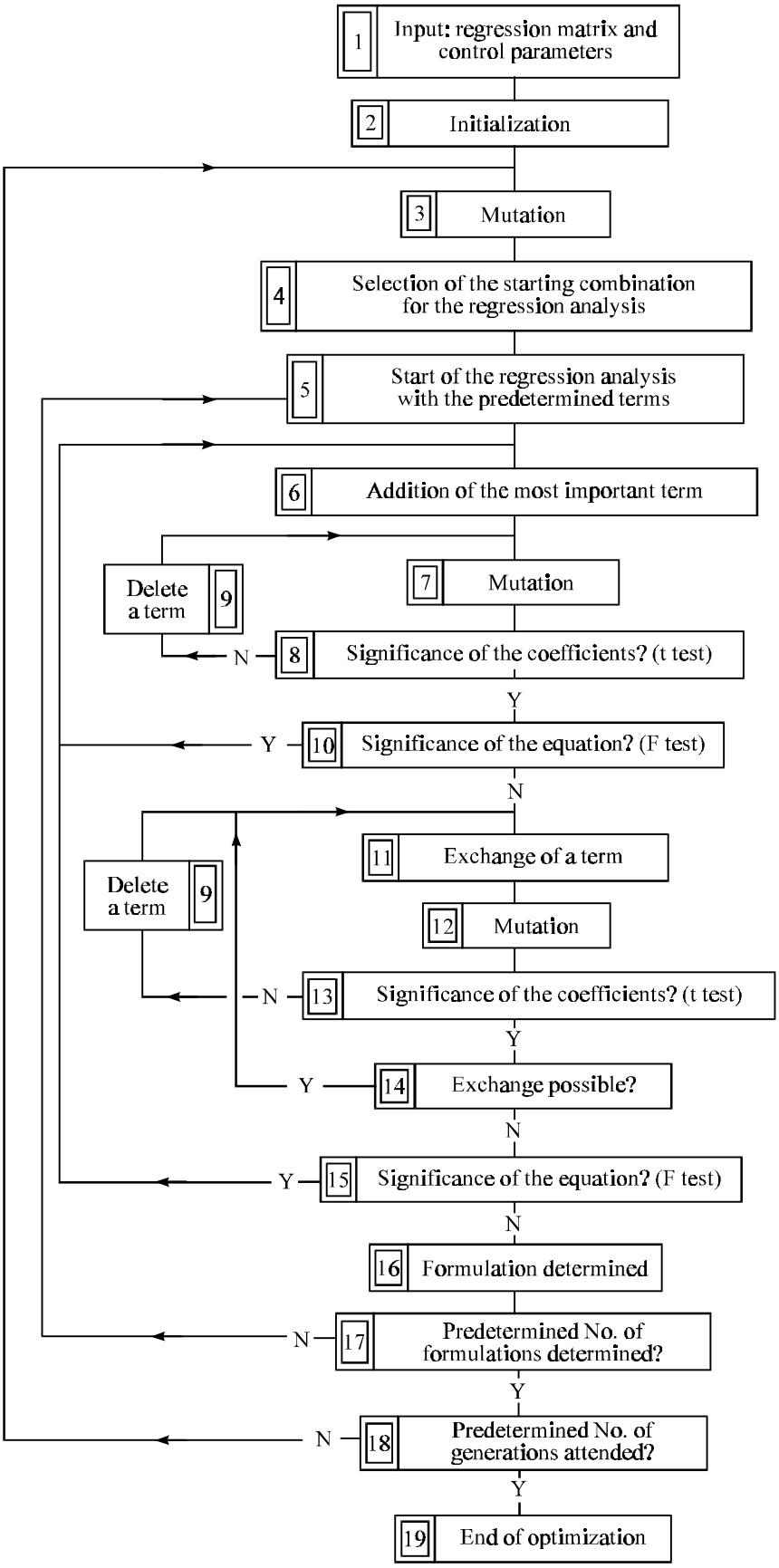

Figure 2
Span et al.

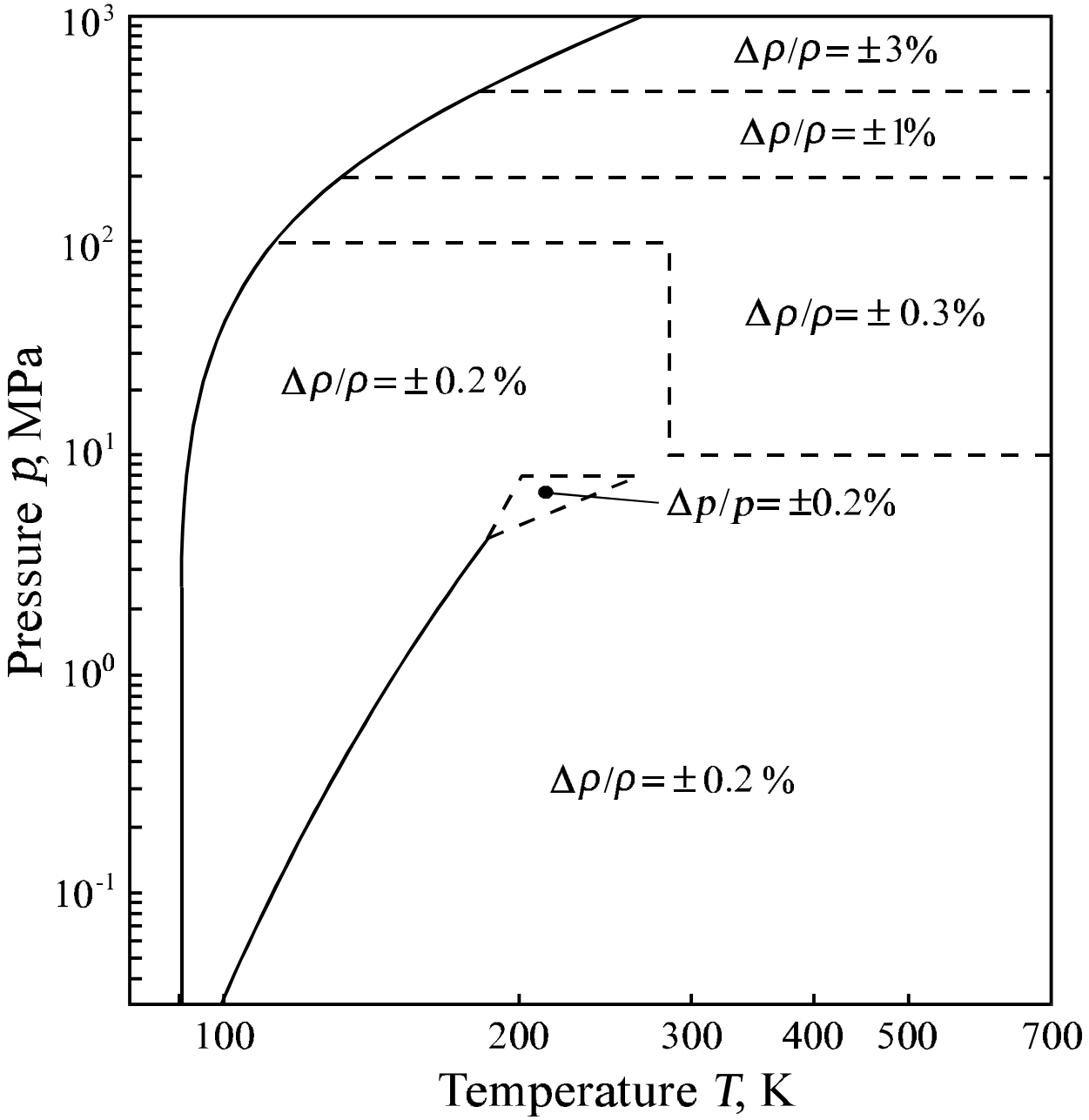

Figure 3
Span et al.
(enlarged by 150%)