# PHYSICAL CHEMISTRY OF AQUEOUS SYSTEMS:
## Meeting the Needs of Industry

Proceedings of
the 12th International Conference on
the Properties of Water and Steam

Edited by

**H. J. White, Jr.**
**J. V. Sengers**
**D. B. Neumann**
**J. C. Bellows**

Organized by:
ASME Research and Technology Committee on
Water and Steam in Thermal Power Systems
and
The International Association for the Properties of Water and Steam

**Begell House**
**New York • Wallingford (U.K.)**

# AN ALGORITHM FOR SETTING UP NUMERICALLY CONSISTENT FORWARD AND BACKWARD EQUATIONS FOR PROCESS MODELLING

**Th. Willkommen*, H.-J. Kretzschmar**, A. Dittmann***

*Technische Universität Dresden
Institut für Thermodynamik und TGA
D 01062 Dresden

**Hochschule Zittau/Görlitz (FH)
Fachgebiet Technische Thermodynamik
D 02755 Zittau

ABSTRACT: The algorithm has been designed for setting up simplified, fast equations for use in energetic process modellings. Simplified numerically consistent equations h=h(T,p), derived from an equation g=g(T,p), and T=T(p,h) for use in power cycle calculations were set up for demonstration purposes.

The algorithm finds numerically consistent equations with a minimized total number of terms automatically and with hardly any subjective influences. Only the corresponding banks of terms have to be established by the equation maker. The algorithm connects the structure optimization method of *Setzmann* and *Wagner* with the simultaneous steady approximation method of *Zschunke*. So the disadvantages of one method can be compensated by the advantages of the other one. Subalgorithms for optimizing the distribution of the input data calculated from a very precise equation of state and for optimizing their weightings have been developed in order to get good equations automatically.

The algorithm can be used for setting up the numerically consistent equation pairs T=T(p,h) and h=h(T,p) as well as T=T(p,s) and s=s(T,p) needed for the IAPWS-project "New Industrial Formulation", too.

## INTRODUCTION

Even though the performance of computers is always improving there is still need to develop faster algorithms for energetic process modelling. This applies mainly to nonstationary processes and real-time modelling. Because the main part of the calculation time of energetic process modellings is taken by the determination of thermodynamic properties this has to be speeded up. One way to do this is to develop simplified equations of state.

The different possibilities for simplification are dependent on the demands of each specific process modelling. Therefore specific equations for each process have to be set up in order to make full use of the possibilities for simplification. Because there are lots of different processes with different demands referring to the equations lots of equations are needed. The algorithm [1] presented in this paper was developed to do this with small manual expenditure.

Because the equations needed for the "New Industrial Formulation" actually also are simplified equations, the algorithm can be used for setting them up, too [2].

## SIMPLIFIED EQUATIONS OF STATE

Five simplifications are possible compared with the complicated, very precise, wide range equations of state such as the IAPS 84 [3] is and the "New Scientific Formulation" will be: Reducing the range of validity, reducing the accuracy, altering the independent variables (T and p instead of T and v), setting up auxiliary "backward" equations and making separate equations for each thermodynamic property.

The IAPS 84 Formulation is valid up to 1500 MPa and 1273.15 K. Obviously such a wide range is not necessary for normal energetic process modellings. For conventional turbine expansions a range up to 873.15 K in temperature and between 6.1 kJ/kg K and 8.7 kJ/kg K in entropy should be wide enought.

It is clear, that the result of process modelling must not be influenced by inaccurate thermodynamic equations of state. However, it is very difficult to determine the influence of the equation's accuracy on process modellings. Therefore mostly equations are made as accurate as possible, i.e. the mean values of measurements are represented by the equations within tolerances equal to the mean deviation of the measurements. Consequently, simplification of equations by reduction of their accuracy is only possible if it is known that the results of the process modellings will not be affected.

Most substances only have (T,v) as independent variables in their equations of state. This makes industrial calculations become very large-scale, because here usually v, h and s as a function of (T,p) are needed and so iterations are necessary. In order to avoid these iterations, equations with (T,p) as independent variables are set up, especially for the commonly used substance water.

Besides (T,p) also (h,p), (s,p), (h,s), (T,s) as well as (h,v) have great importance as independent variables in industrial calculations. Therefore so called "backward" equations with just these independent variables are set up in addition to the (T,p)-equations. In order to avoid unpredictable results in process modellings such backward equations have to be numerically consistent to their corresponding "forward"-(T,p)-equations, i.e. the deviation between forward and backward equations is forced to be 1 or 2 orders of magnitude smaller than the deviation between forward equations and the mean values of measurements.

Canonical equations are often preferred when setting up simplified equations of state. They offer the opportunity of deriving all thermodynamic properties from one equation. On the other hand equations of state derived from a canonical equation are more complicated than equations fitted only for a single property. This has two reasons: First, the mathematical derivations require additional operations and second, each derived equation includes the thermodynamic "information" for all other properties and is complicated accordingly. Because for process modellings there is no need of having all properties derived from one canonical equation, calculation time can be saved by setting up separate equations for each property needed. The algorithm presented is able to set up separate equations as well as equations derived from a canonical equation.

## APPROXIMATION METHODS

### Structure Optimization

Short equations of state with little need for computing time can be set up by so called structure optimization methods only. The aim of these methods is to select that combination of terms (equation structure), from a bank of terms, which is able to fulfil the accuracy requirements with a minimized number of terms. The bank of terms should contain all mathematical functions, which can be considered to be significant for the approximation of the thermophysical dependencies concerned.

At present there are 3 efficient structure optimization methods: the stepwise regression analysis of *Wagner* [4], the evolutionary optimization method of *Ewers* and *Wagner* [5] and the structure optimization method of *Setzmann* and *Wagner* [6]. They have been developed for setting up empirical equations of state from experimantal data and can be applied for approximation of simplified equations from data calculated using a precise equation of state, too. The optimization method of *Setzmann* and *Wagner* is the best with respect to user-friendliness and to the quality of the established equations. Therefore it is preferred for application in the newly developed algorithm.

The structure optimization methods mentioned are based on the least-squares principle. Therefore the regression data have to be weighted in most cases in order to achieve a satisfactorily small maximum deviation and a nearly steady approximation respectively. Because these structure optimization methods are discreet approximation methods as well, distribution of regression data has influence on the quality of the approximated equation. The equation may begin to oscillate between two regression data points, if there are not enough data points. To avoid this, so called "artificial data" are added when using experimental data and supplementary data must be added when approximating data received from a precise equation of state. Such a supplement of the regression data will be called "change of data distribution" within this paper.

Both, the weighting of data and the change of data distribution require much experience by the equation maker. Now, however, the new algorithm includes two sub-algorithms for improvement of the weighting and distribution of the regression data.

### Simultaneous Steady Approximation

The maximum deviation between a base equation or base data and the approximated equation can be controlled directly by the steady approximation method, in contrast to the least square methods. On the other hand there is no structure optimization available for the steady approximation method. Therefore steady approximation only can be used to finally improve the free parameters (coefficients) of equations set up by structure optimization.

Simultaneous approximation of forward and backward equations generally is a non-linear problem. On the other hand the steady approximation is an iterative method, able to fit non-linear parameters. This offers the possibility of extending the steady approximation for simultaneously fitting several equations, which have a relationship to each other.

*Zschunke* has developed an algorithm for simultaneous approximation of forward and backward equations with two independent variables [7]. The deviations between the precise equation of state and the approximated forward equation(s) as well as the deviations between approximated forward and backward equations of one or more equation pairs can be controlled simultaneously. The deviations mentioned will be called "state error" and "numerical consistency error" respectively within this work.

Even if the structure optimization is done with optimized regression data weighting, maximum state error and maximum numerical consistency error can be reduced by about 20% with simultaneous approximation. Therefore, simultaneous steady approximation is necessary in order to find all equation structures able to fulfill the required accuracy. From these equation structures the fastest (shortest) can be chosen later.

## THE NEW ALGORITHM

### Overview

The algorithm is designed for automatically setting up sets of numerically consistent forward and backward equations from given banks of terms with a minimized total number of terms.

Since the length of a backward equation required to fulfil the demanded numerical consistency depends on the structure of the forward equation belonging to it, many different forward equations have to be tested in order to minimize the total number of terms. The structure optimization method of *Setzmann* and *Wagner* is used in connection with newly developed sub-algorithms for changing the weighting and the distribution of the regression data calculated from a precise equation of state for setting up the forward equations. The state error of equations found by structure optimization can be decreased with steady approximation by as much as 40% according to the weighting and distribution of the regression data. Therefore all equations found by structure optimization having a state error less than 1.67 times the allowed maximum deviation are submitted to a steady approximation. Those equations, which reach the demanded accuracy after steady approximation, are used to search for the backward equation(s).

In order to set up backward equation(s) for one forward equation, at first new regression data are calculated from the forward equation. Now structure optimization and simultaneous steady approximation are used for finding the wanted backward equations. Because the numerical consistency error can be reduced with simultaneous approximation by as much as 45%, backward equations need not fulfil the demanded accuracy after structure optimization. The sub-algorithms for changing the weighting and the distribution of the regression data are used here only to find the best equation structure and not to find a lot of equations.

After backward equations for all forward equations meeting the demanded accuracy are set up, those equation sets with the minimum total number of terms can be selected. It's now the task of a programmer to select the equation set with a minimum need of computing time.

### Getting a Variety of Forward Equations

A variety of different forward equation structures can be found by changing the bank of terms, increasing the equation length, changing the weighting or the distribution of the regression data or by forcing the structure optimization to find not only the best, but also other good equation structures.

Changing the bank of terms for setting up the forward equation often is not possible, because one has only one bank of terms leading to good results. If there are more good banks of terms, the new algorithm should be used with each of them.

For a given bank of terms there is a minimum equation length required to fulfil the demanded accuracy. For this minimum equation length at least one, but at most a few, different equation structure(s) can be found. Increasing the equation length results in more different equation structures able to fulfil the demanded accuracy. The minimum equation length is found by increasing the equation length step by step, beginning with an estimated starting value, until the

required accuracy is reached.

It is the task of the developed sub-algorithms for changing the weighting and the distribution of the regression data to find as many equations as possible with a given length fulfilling the demanded accuracy with the help of structure optimization. This task is supported by forcing the structure optimization to keep all equations, it has ever found during the optimization process, and selecting the best ones after it has finished.

**The Subalgorithms**

**Correction of the Weighting.** Normally, the weighted sum of squares $\chi^2$, to be minimized by the structure optimization, is calculated from

$$\chi^2 = \sum_{n=1}^{N} \sum_{i=1}^{I_n} \left( \frac{y^*(\vec{x}_n) - y_{n\,i}}{\sigma_n} \right)^2 \tag{1}$$

where N stands for the number of measured points, $I_n$ for the number of measurements at point n, $y^*$ for the approximated equation, $\vec{x}_n$ for the vector of the independent variables, $y_{n\,i}$ for the measured values and $\sigma_n$ for the total variance of the experimental data at point n. In the case of data being calculated from a precise equation of state there is no variance. However, it is replaced by a weighting factor $W_n = 1/\sigma_n$. Equation (1) now reads:

$$\chi^2 = \sum_{n=1}^{N} \sum_{i=1}^{I_n} W_n^2 \cdot \left( y^*(\vec{x}_n) - y_{n\,i} \right)^2 \tag{1a}$$

Furthermore, there are only $I_n = 2$ values at each point n: $y_{n\,1} = y_n + \frac{1}{3}\Delta y$ and $y_{n\,2} = y_n - \frac{1}{3}\Delta y$. So the sum of squares is calculated now by:

$$\chi^2 = \sum_{n=1}^{N} W_n^2 \cdot \left( y^*(\vec{x}_n) - y_n - \frac{\Delta y_n}{3} \right)^2 + W_n^2 \cdot \left( y^*(\vec{x}_n) - y_n + \frac{\Delta y_n}{3} \right)^2 \tag{2}$$

where $W_n$ can be set to the reciprocal of the allowed tolerance $\Delta y_n$ at point n.

Using (1) and (2) respectively results in the deviation between regression data and approximated equation becoming too big in regions with only a few regression data. This can be compensated by changing the distribution or the weighting of the regression data. The weighting is changed in the new algorithm stepwise in the following way:

In the initial step (k=0) the weighting factor at point n is set to the reciprocal of the allowed tolerance $W_{n\,0} = 1/\Delta y_n$. In all further steps ($k \geq 0$), this value is multiplied by the mean value of the deviations of all preceding steps, divided by the allowed tolerance:

$$W_{n\,k+1} = \frac{1}{\Delta y_n} \cdot \sum_{i=1}^{k} \frac{\left| y_i^*(\vec{x}_n) - y_n \right|}{k \cdot \Delta y_n} \tag{3}$$

Previous calculations have shown that a maximum of 5 steps is necessary to achieve a nearly steady approximation and a small maximum deviation respectively.

**Changing the Distribution of Regression Data.** As described in the section "Structure optimization" there may be oscillations between the regression data points. Within the new algorithm a "sharp oscillation" is defined as a point, whose deviation is more than 15% larger than the deviation of the surrounding regression data points. Such sharp oscillations can hardly be compensated by changing the weighting of the regression data or by steady approximation. Therefore the regression data have to be supplemented at places where sharp oscillations are

appearing, in order to give the structure optimization information about where the approximated equation is to go between the points surrounding the sharp oscillation. The supplemention is carried out in the following way:

Each of the 3 best equations (with respect to their sums of squares), found by the structure optimization, is investigated with respect to whether its iteratively calculated actual maximum deviation is more than 25% greater than the maximum deviation at the regression data points. If an equation does have such a significantly greater actual error maximum, its local error maxima will be determined iteratively. Then, these local error maxima are investigated as to which of them are sharp oscillations. At the positions of these sharp oscillations supplementary data points are added to the regression data that produce the sharp oscillations.

With the supplemented regression data a new structure optimization is done, and again the 3 best equations are investigated as to whether further supplementation is necessary. This cycle is repeated until no further supplementation is necessary. Now the iterative change of data weighting starts. There additional supplementations may prove to be necessary. Because supplementation of data points causes a change of data distribution, and combined with that a change of weighting, the iterative change of data weighting has to be started over after each supplementation. The two interlocked cycles for changing data weighting and distribution continue until 5 iteration steps for data weighting are done in a row without supplementation of regression data.

Changing the weighting and the distribution of the regression data as described enables the structure optimization to provide a nearly steady approximation and approximate equations with small maximum deviation from the basis data respectively. However, the data weighting and distribution optimized for one equation length are not the same for another equation length. Therefore the sub-algorithms for changing the data weighting and distribution have to be run again for each tested equation length.

**Estimation of Minimum Equation Length.** There is a peculiarity when using structure optimization methods for approximation of data calculated from a precise equation of state: Experimental data usually have a nearly normal distributed deviation from the thermophysical truth. In that case, the equation length can be estimated by the statistical Fisher- and Student-tests, included in the structure optimization. Because data calculated from a precise equation of state have no deviation from the precise equation, looked upon as the thermophysical truth, statistical tests will fail. For that reason and in order to avoid numerical problems during the structure optimization (sum of squares becomes negative and coefficients get wrong values), the calculated data now are superimposed with an artificial deviation of ⅓ of the allowed tolerance at each point, i.e. each position is represented by two points with a deviation of ±⅓ from the calculated point.

The superimposion offers the possibility of predicting the minimum possible sum of squares for the structure optimization. If $\Delta y_n$ stands for the allowed tolerance at a point n and N for the number of points, the minimum weighted sum of squares $\chi^2_{min}$ is determined by:

$$\chi^2_{min} = 2 \cdot \sum_{n=1}^{N} W_n^2 \left( \frac{\Delta y_n}{3} \right)^2 \qquad (4)$$

Experiences have shown: If an equation, fitted with use of the algorithm developed, ought to keep the allowed tolerance, the weighted sum of squares of an equation of the same length, but fitted without the algorithm, has to be less than 3 times $\chi^2_{min}$. A modified version of the stepwise regression analysis of *Wagner* is used in order to determine the equation length belonging to a sum of squares 3 times greater than $\chi^2_{min}$. The equation length, found in this way, is used as the estimated minimum equation length for the new algorithm.

## EXAMPLE

The algorithm was applied for setting up simplified equations v=v(T,p), h=h(T,p) and T=T(p,h) for superheated steam in the region between T=273.16 K···873.15 K and s=6.2···8.7 kJ/kg K. The equations for v and h are derived from a canonical equation g=g(T,p). The allowed tolerance of v and h referring to the equation of state of *Saul* and *Wagner* [8], converted into ITS 90, corresponds to the toleances of the International Skeleton Tables IST 85 [9] and the maximum numerical consistency error between h=h(T,p) and T=T(p,h) is less than 0.075 K. With a total number of 22 terms the following is the best equation pair:

An equation g=g(T,p) with 10 terms:

$$\frac{g}{RT} = a_1 - a_2 p_r \frac{1}{(T_r - T_r^{spin} + 0.015)}\left(\frac{1}{T_r} - 0.45\right)^3 + a_3 p_r (T_r - T_r^{spin} + 0.015)\left(\frac{1}{T_r} - 0.45\right)^3 + a_4 p_r \left(\frac{1}{T_r} - 0.45\right)^2$$

$$+ a_5 p_r [\ln(p_r) - 1]\left(\frac{1}{T_r} - 0.45\right)^4 + a_6 p_r + a_7 \ln(p_r) + a_8\left(\frac{1}{T_r} - 0.45\right) + a_9 \ln(T_r) + a_{10} T_r$$

$$\text{with } R = 0.46152 \text{ kJ/kg K}, \; p_r = \frac{p}{22.064 \text{ MPa}}, \; T_r = \frac{T}{647.096 \text{ K}}$$

$$\text{and } T_r^{spin} = \frac{1}{1 - \ln(p_r)[0.173 - 0.011 \ln(p_r) - 0.02 p_r]}$$

and an equation T=T(p,h) with 12 terms:

$$T_r = \sum_{m=1}^{5} b_m p_r^{k_m} h_r^{l_m} + \sum_{m=6}^{11} b_m p_r^{k_m} e^{l_m h_r} + b_{12} h_r$$

$$\text{with } h_r = \frac{h}{2085.1246 \text{ kJ/kg}}$$

$T_r^{spin}$ is an approximation for the reduced gas spinodal temperature. The parameters are:

| m | $a_m$ | $b_m$ | $k_m$ | $l_m$ |
|---|---|---|---|---|
| 1 | -5.3661720201833 | -1.1496248071964 | 0 | -1 |
| 2 | 0.0077061074769843 | 22.776278053539 | 1 | -12 |
| 3 | 1.2789155292727 | -29.233929240872 | 2 | -12 |
| 4 | -1.1548773444235 | 73.357895871987 | 2.5 | -16 |
| 5 | 0.10873781792311 | 10553.501096865 | 2.5 | -32 |
| 6 | -0.016687806471207 | 13.384605319034 | 0 | -5 |
| 7 | 1.0002364850457 | -20.807725012156 | 0.75 | -4 |
| 8 | 6.7787851623073 | 1583836.5752764 | 0.75 | -12 |
| 9 | -3.6181905375738 | 24.99774959016 | 1 | -3 |
| 10 | -0.41039970809357 | -132743279.02303 | 1.5 | -14 |
| 11 | | 0.12785100320748E+35 | 2 | -60 |
| 12 | | 1.1222439584626 | | |

## SUMMARY

The algorithm developed allows setting up simplified, fast equations of state for use in process modellings. Numerically consistent equations with a minimized total number of terms are found with small manual expenditure. The algorithm combines the advantages of structure optimization and simultaneous steady approximation. The banks of terms serving as basis for the approximated equations still have to be given by the equation maker. However, the weighting and the distribution of the regression data are optimized automatically. The number of terms needed to fulfill the accuracy requirements can be estimated before starting the algorithm and is determined exactly within the algorithm. A way was found to overcome numerical problems when approximating data calculated from a precise equation of state.

The application of the algorithm has shown, that the numerical consistency error can be made about one order of magnitude smaller than the state error, without making the backward equation significantly longer than the forward equation.


## REFERENCES


[1] Th. Willkommen: Ein Algorithmus zur Aufstellung numerisch konsistenter Gleichungen für die in Prozeßmodellierungen benötigten thermodynamischen Umkehrfunktionen. VDI-Fortschrittsberichte, Reihe 6 (1995) - in Preparation

[2] J. Trübenbach, H.-J. Kretzschmar, Th. Willkommen, A. Dittmann: Improved Equations for Backward Functions T=f(p,h), T=f(p,s), Ts=f(p) - A Contribution to the IAPWS Project "New Industrial Formulation": Properties of Water and Steam (Proc. of the 12$^{th}$ ICPS)

[3] IAPS: Release on the IAPS Formulation 1984 for the Thermodynamic Properties of Ordinary Water Substance for Scientific and General Use: Report. - Gaithersburg(1984)

[4] W. Wagner: Report PC/T15: IUPAC Thermodynamic Table Project Centre. - London(1977)

[5] J. Ewers, W. Wagner: Eine Methode zur Optimierung der Struktur von Zustandsgleichungen und ihre Anwendung zur Aufstellung einer Fundamentalgleichung für Sauerstoff: VDI-Forschungsheft. Berlin (1982)609. p.27/34

[6] U. Setzmann, W. Wagner: A New Method for Optimizing the Structure of Thermodynamic Correlation Equations: Int. J. Thermophys. - 10(1989)6 - p. 1103/1126

[7] A. Dittmann, T. Zschunke, J. Klinger, H.-J. Kretzschmar: Computer Code for the Generation of Optimized Algorithms to Calculate Thermophysical Properties of Water and Steam: Properties of Water and Steam (Proc. of the 11$^{th}$ ICPS). Hem.Publ.Corp. New York(1990)

[8] A. Saul, W. Wagner: A Fundamental Equation for Water Covering the Range from the Melting Line to 1273 K at Pressures up to 25 000 MPa: J.Phys.Chem.Ref.Data 18(1989)4 p. 1537/1563

[9] IAPWS Release on the Skeleton Tables 1985 for the Thermodynamic Properties of Ordinary Water Substance (Revision 1994). Issued by IAPWS - In Preparation