

Optimizing Curve Fits for Extreme Physical Properties with High Error

Suyog Soti

Introduction

In measurement laboratories, as the physical properties of materials approach extreme values, the error of the measurement increases to the point where the error could be well above 200%. When the error heightens over acceptable measures, researchers must rely on an equation from data points, collected when the error was acceptable, to calculate the measure of the physical properties. This is more relevant to the chemical engineers who measure properties like vapor pressure at extremely low temperatures and study the behaviors of chemicals at extreme temperatures. When researchers fit a curve to their data, they always run the danger of an unbalanced fit meaning that they will over-fit a curve or their fit will not be able to accurately determine the values derived from the low end equations. An example of this is when a scientist measures the vapor pressure of a substance and fits the data to a curve. The data that the scientist calculates from the fit will not be able to accurately determine the specific heat of that substance at that temperature even if the specific heat is the second derivative of pressure. The specific heat data will be wrong because of overfitting. The only real way for the scientists to take this into account is to curve fit specific heat separately and ignore the data from the pressure curve. The problem to this approach is that the researcher will still be over fitting the specific heat capacity curve, and that the researcher has to do twice as much work. What multi-parameter optimization will do is curve fit all of the derivatives at once to avoid the problem of overfitting and the wasting time. For this project, the focus was on developing the program that does optimizes multi-parameter fits for vapor pressure equations so that one equation fits the curves of saturation pressure, density and specific heat with respect to temperature.

The Base Equations

The base equation for saturation pressure was the equation of state derived from the Clausius-Clapeyron in the form

$$\ln\left(\frac{p_{crit}}{p_{sat}}\right) = \frac{C_1 \times \tau^{b_1} + C_2 \times \tau^{b_2} + C_3 \times \tau^{b_4} + \dots + C_n \times \tau^{b_n}}{1 - \tau} \quad (1)$$

where C can b are different sets of constants, and $\tau = 1 - \frac{T}{T_{crit}}$. The number of terms on the right hand side usually does not exceed 6. This is mainly because once it does, it is generally accepted to be considered over-fitted. Equation 1 is derived from the Gibbs free energy equation ¹.

¹Magee *et al*