Potential energy function from second virial data using sensitivity analysis

Nelson H. T. Lemes, Rita C. O. Sebastião e João P. Braga
Departamento de Química
Universidade Federal de Minas Gerais – UFMG
Belo Horizonte, MG, Brazil
jpbraga@oxigenio.qui.ufmg.br

Abstract. Retrieving the potential energy function from second virial data, using the sensitivity analysis approach, is discussed in this work. A potential energy function, with an initial average error of 92%, in temperature range 100K to 1500K, with respect to a reference potential, was considered as an initial guess. Within the present framework it was possible to produce another potential with an average error of 0.7% and 2.7%, using two regularization methods. Analysis of the sensitivity matrix has shown to be an important step while inverting the data. This preliminar analysis provides important informations about the optimal temperature and coordinate range to be used in the inversion process.

INTRODUCTION

The mathematical formulation of a phenomena is generally expressed as $K(f) = g$, in which g represents an observable quantity, f characterizes the system and K is a nonlinear operator between these two quantities. Obtation of g, from K and f, represents what is called a direct problem, whereas extraction of f, from K and g is the inverse problem [1]. The retrieve of the potential energy function from experimental data, such as from scattering data, spectroscopy and thermodynamics, is particularly important in physical-chemistry.

The above inverse problem is know as an ill posed problem, defined such that one of the three conditions, (a)existence; (b)uniqueness or (c)continuity is not satisfied [2]. In general, none of the above three conditions are satisfied. The inherent experimental error, together with the approximate model used to describe the phenomena, that is, $K(f) \approx g$ makes the solution of the inverse problem to be characterized as an ill posed problem. As consequence, the representation of K will be ill conditioned or rank deficient, which makes traditional methods, such as the Gausssian elimination or the LU decomposition [3], inadequated to handle inverse problems. Instead, more appropriate methods, such as the Tikhonov regularization [1] or the singular value decomposition [3], has to be used to find an adequate solution [1].

The expansion, in terms of density, for the compressibilty factor gives the relation between potential energy function and second virial coefficient, indicating this is a nonlinear ill posed problem. Nevertheless, an equivalent linear problem can be estabilished at the expense of introducing a singularity into the problem. As a consequence, the inverse problem has to be solved in two stages; one to the right and the other to the left of this singularity. Each of these two parts of the potential were obtained by using recurrent neural network [4, 5], by singular value decomposition and by the Tikhonov regularization

[6, 7]. The problem has also been solved in a non linear form using recurrent neural network [8] and on the class of convex-concave functions [9].

In the present work a functional sensitivy analysis approach will be used to obtain the Ar-Ar potential energy function from second virial data. Although the original problem is also transformed in a linear one, no singularity is involved and the whole potential can be obtained in an iterative procedure. The sensitivity matrix, connecting potential energy with second virial coefficient, will be analysed with respect to the temperature and potential energy ranges. This will indicate which region of the potential will be adequated to be inverted in a set of experimental data.

THE SENSITIVITY ANALYSIS APPROACH

Sensitivity analysis approach will be applied to recover the potential energy function from second virial data using the equation of state in the form [10], $\frac{p}{k_B T} = \rho + B(T)\rho^2$, in which $B(T)$ is the second virial coefficient and the other quantities have their usual meaning. The relation between potential energy function, $E_p(R)$ and the second virial coefficient is given by,

$$B(T) = 2\pi \int_0^\infty R^2 (1 - e^{-E_p(R)/k_B T}) dR, \quad (1)$$

The above integral equation will be represent, in general terms, as

$$\int K(x, f(y)) dy = g(x), \quad (2)$$

in which x will stand for the temperature and y for the coordinate. Discretization of (2), at x_j, gives approximately, $\sum_i^n K(x_j, f_i)\Delta R_i = g_j$, with $f_i = f(y_i)$ and $g_j = g(x_j)$. A change in g, with a correspondent change in f, can be written as,

$$\sum_i^n [K(x_j, f_i + \delta f_i) - K(x_j, f_i)]\Delta R_i = \delta g_j \quad (3)$$

Using, $K(f_i + \delta f_i) \approx K(f_i) + J_{ij}\delta f_i$ with $J_{ij} = \frac{\partial K_{ij}}{\partial f_i}$,

$$\sum_i^n J_{ij} \Delta R_i \delta f_i = \delta g_j \quad (4)$$

How the change in g_j will affect the solutions f_i will be measure by the quantity $J_{ij}\Delta R_i$, thereafter denoted by S_{ij}. In an integral form, equation (4) is equivalent to,

$$\int \left(\frac{\delta \mathbf{K}(T,R)}{\delta \mathbf{f}(R)} dR \right) \delta \mathbf{f}(R) = \delta \mathbf{g}(T) \quad (5)$$

or, in matrix notation,

$$\mathbf{S}\delta\mathbf{f} = \delta\mathbf{g} \quad (6)$$

This equation can be used in the following manner:

1. An initial guess, $\mathbf{f}_0$, is given for the unknow.
2. The approximately experimental value, $\mathbf{g}_0$, is computed.
3. The quantity $\delta\mathbf{g}$ is calculated as the difference between the experimental data and $\mathbf{g}_0$.
4. A correction to the initial guess can be evaluated as $\delta\mathbf{f} = \mathbf{S}^{-1}\delta\mathbf{g}$ from which a new value for $\mathbf{f}$ is obtained.

Equation (5) is general and can be applied to any nonlinear problem with a due interpretation of the quantities. For the present problem $\mathbf{f}$ will be the potential energy function, $\mathbf{E}_p$, and $\mathbf{g}$, the second virial data, $\mathbf{B}$.

The original nonlinear problem is, therefore, transformed into a Fredholm integral equation of first order, which will be solved by the singular value decomposition method [3], with two filters factors. The sensitivity matrix is transformed into, $\mathbf{S} = \mathbf{U}\mathbf{\Sigma}\mathbf{V}^T$ in which $\mathbf{U} \in \Re^{m\times m}$ and $\mathbf{V} \in \Re^{n\times n}$ are orthogonal matrices, whereas $\mathbf{\Sigma} \in \Re^{m\times n}$ is a diagonal matrix. The columns of $\mathbf{U}$ and $\mathbf{V}$ will be represented, respectively, by $\mathbf{u}_i$ e $\mathbf{v}_i$, and σ_i the diagonal elements of $\mathbf{\Sigma}$. A general solution of (6) can be written as,

$$\delta\mathbf{E}_p(\lambda) = \sum_{i=1}^n \phi_{i,\lambda} \frac{\mathbf{u}_i^T . \delta\mathbf{B}}{\sigma_i} \mathbf{v}_i \quad (7)$$

with λ a regularization parameter and $\phi_{i,\lambda}$ a filter factor, depending on the singular values and on the regularization parameter. Two special cases are possible for (7), corresponding to two different choices of the filter factor: (a) A filter factor equal to unity before a certain value, k, and zero after this value; and (b) A smooth filter which takes into account all singular values in decreasing order of importance, $\phi_{i,\lambda} = \frac{\sigma_i^2}{\sigma_i^2+\lambda}$, a procedure equivalent to the Tikhonov regularization [1].

RESULTS AND DISCUSSIONS

Important informations on the inversion to be carried out can be obtained by analysing some aspects of $\mathbf{S}(R,T)$, although it is more appropriate to use, instead, the matrix, $\mathbf{S}' = \frac{d\ln(\delta\mathbf{B})}{d\ln(\delta\mathbf{E}_p)}$. Analysis of the $\mathbf{S}'$ matrix will be easier, since it will have information on relative values, rather than on absolute values, as in the $\mathbf{S}$ matrix.

Level curves of the sensitivity matrix, $\mathbf{S}'$ are presented in figure (1). In the region in which the sensitivity matrix assumes larger values it will be more stable to obtain the inverted potential. The opposite is also true. If $\mathbf{S}'$ reaches small values, $E_p(R)$ will be more difficult to be obtained. The appropriate region for the inversion procedure is, therefore, promptly obtained from figures (1).

Simulated data were obtained by carrying the integration of (1) in the coordinate range 1.5Å to 40Å and temperature between 100K to 1500K, using recent potential energy for Ar-Ar system [12]. These data give and error of 4% when compared with the experimental second virial data. However, when analysing the inverted potential, one must be guided by the sensitivity matrix. This analysis gives the optimal region for the coordinate from 3Å to 6Å.

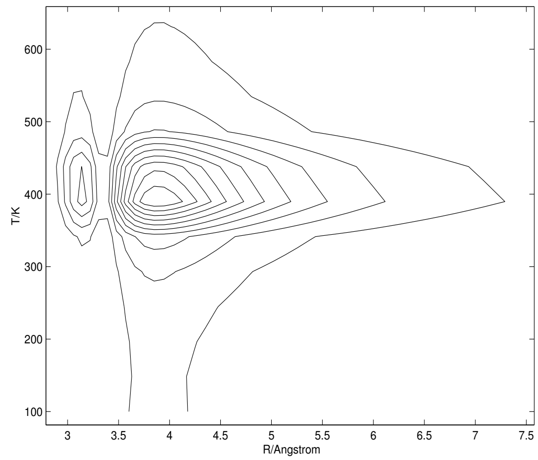

Figure 1: Level curves for the sensitivity matrix.

An initial guess for the potential energy function, which gives an error of 92% for the calculated second virial coefficient, was used to obtain a new potential from equation (7). In figure (2) the computed potential, using the truncated singular value decomposition, is compared against the initial guess and the potential from reference [12]. The new potential gives an error of 0.7% for the calculated second virial coefficient, a value within the experimental error. In figure (3) the inverted potential was computed using the Tikhonov regularization, with 2.7% error in the computed $B(T)$. The discrepancy principle [13] was used to calculated the optimal dimension of the subspaces and the optimal regularization parameter. The computed values are $k = 7$ and $\lambda = 92.367$ Å$^3/eV$, from which the potentials were calculated. Second virial coefficients from Aziz's potential [12], together with the values calculated from the inverted potentials are presented in figure (4).

Although both potential gives basically the same error for the second virial coefficient their quality can be better appreciated by computing the error of these new potentials with respect to the potential of reference [12], taken as exact. In the region from

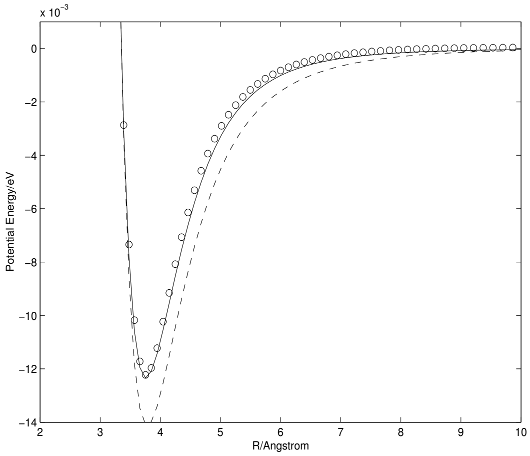

Figure 2: Potential energy function from reference [12] (full lines) initial guess (dashed lines) and inverted potential energy function (o). Results are computed using the singular value decomposition.

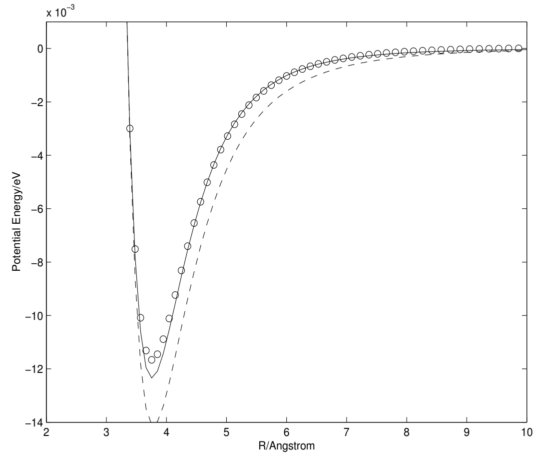

Figure 3: Inverted potential using the Tikhonov regularization. Labels as in previous figure.

1.5 Å to 40 Å the relative error are of the order of 10^3. This is not taken to imply the potentials are not of quality to describe the property being inverted. As inferred by the analysis of the sensitivity matrix the results have to be considered in the region where information can be obtained, that is, between 3Å to 6Å. In this region the potential from figures (2) and (3) gives, respectively, the errors 0.6% and 0.5%. The error in the initial guess, inside this region, is 27%, showing the improvement of the potentials.

CONCLUSION

The functional sensivity approach was used to recover potential energy function from simulated second virial coefficient. Unlike in the Keller and Zumino approach [7] the present functional linearization introduces no singularity in the new integral equation, providing a single procedure, as in the previous nonlinear approach [8], to recover the complete potential.

An optimal region, for the temperature and coordinate, can be inferred from the sensitivy matrix. In this case the optimal coordinate range coincide with the one obtain previously for the same potential [8]. Further insights on the inversion process can be obtained by using the singular value decompostion inversion. For large values of the basis set, and outside the optimal region, it will appear more oscilation in the basis functions. It is clear, therefore, the basis better represents the potential in the optimal region. Since the singular values decays to zero, the basis will be amplified outside the region.

The analysis on the singular value basis $\mathbf{v}_i$ reinforce the previous analysis on the matrix **S**, indicating in which region the inverted potential has to be considered. With an initial guess for the potential energy function, containing an average error of 92% for the second virial data, another potential was recovered, under the above analysis, with an average error of 0.7% for Tikhonov regularization and 2.7% for truncated singular value decomposition.

A general strategy to handle other nonlinear problems is also provide by the sensitivity approach. For example, the determination of the po-

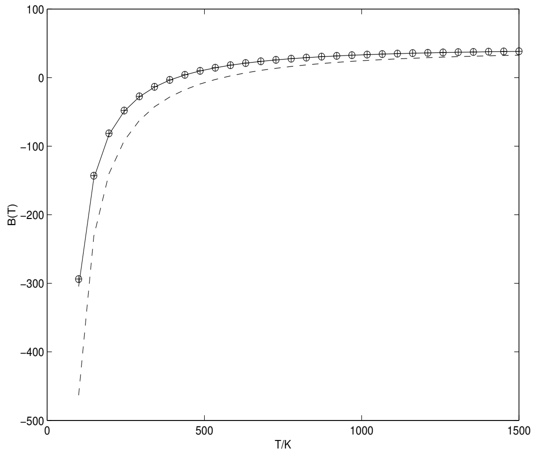

Figure 4: Computed second virial coefficient from reference potential [12] (full line), inverted SVD potential (o), inverted Tikhonov (+) and from the initial guess potential (dashed line).

tential energy from scattering data can be tackled by the present approach. Even in the situation in which the kernel is unknow the sensibility matrix can be constructed by experimental or simulated data.

REFERENCES

1. A. N. Tikhonov and A.V. Goncharsky, Ill-Posed Problem in the Natural Sciences, ed. Mir Publisher, Moscow, 1987.
2. J. Hadamard, *Lectures on Cauchy's Problems in Linear Partial Differential Equations*, Yale University Press, New Haven, 1923.
3. G.H. Golub and C.F. Van Loan, *Matrix Computations*, Johns Hopkins Univ Press, Baltimore, 1996.
4. J.P. Braga, M.B. de Almeida, A.P. Braga, and J.C. Belchior, *Chem. Phys.*, **260**, 347(2000)
5. M.B. de Almeida, A.P. Braga, J.P. Braga, J.C. Belchior and G.F.G. Yared, *Phys. Chem. Chem. Phys.*, **2**, 103(2000).
6. N.H.T. Lemes, J.P. Braga and J.C. Belchior, *Chem. Phys. Lett.*, **296**, 233(1998).
7. J.P. Braga and J.L. Neves, *Phys. Chem. Chem. Phys.*, **3**, 4355(2001).
8. R.C.O. Sebastião, N.H.T. Lemes, L.S. Virtuoso and J.P. Braga, *Chem. Phys. Lett.*, **378**, 406(2003).
9. N.V. Anikeeva, I.V. Kochikov, G.M. Kuramshina, and A.G. Yagola, Numerical Methods and Programming, **4**, 200(2003).
10. J.O. Hirschfelder, C.F. Curtis and R.B. Bird, *Molecular Theory of Gases and Liquids*, John Wiley, New York, 1954.
11. J. H. Dymond and E. B. Smith, *The virial coefficients of pure gases and mixtures: A critical compilation*, Claredon Press, Oxford, 1980.
12. R.A. Aziz, *J. Chem. Phys.*, **99**, 4518(1993).
13. P.C. Hansen, *Rank-Deficient and Discrete Ill-Posed Problems*, Society for Industrial and Applied Mathematics, Philadelphia, 1998.