Applying Firefly Algorithm to Data Fitting for the Van der Waals Equation of State with Bézier Curves

Almudena Campuzano
Oregon State University
Corvallis, OR, USA
University of Cantabria
Santander, Spain
campuzam@oregonstate.edu

Andrés Iglesias
Toho University
Funabashi, Japan
University of Cantabria
Santander, Spain
iglesias@unican.es

Akemi Gálvez
Toho University
Funabashi, Japan
University of Cantabria
Santander, Spain
galveza@unican.es

***Abstract*—The Van der Waals equation is an equation of state that generalizes the ideal gas law. It involves two characteristic curves, called binodal and spinodal curves. They are usually reconstructed through standard polynomial fitting. However, the resulting fitting models are strongly limited in several ways. In this paper, we address this issue through least-squares approximation of the set of 2D points by using free-form Bézier curves. This requires to perform data parameterization in addition to computing the poles of the curves. This is achieved by applying a powerful swarm intelligence method called the firefly algorithm. Our method is applied to real data of a gas. Our results show that the method can reconstruct the characteristic curves with good accuracy. Comparative work shows that our approach outperforms two state-of-the-art methods for this example.**

***Index Terms*—equation of state, Van der Waals equation, data fitting, swarm intelligence, firefly algorithm, Bézier curves.**

I. Introduction

Equations of state (EoS) are mathematical equations describing thermodynamic relationships among pressure P, volume V, and temperature T, for a pure component or a multicomponent system. There are many EoS proposed in the literature. The *ideal gas law* is the most popular and widely known. It works properly for low pressures and high temperatures, but can fail otherwise. Other EoS incorporate different corrections to approximate the behavior of real gases more accurately. The *Van der Waals* (VdW) equation is a popular EoS that generalizes the ideal gas law by taking into account the molecular size and the interaction forces between the molecules through two new parameters, a and b, whose value depends on each particular substance. It is given by:

$$\left[P + a\left(\frac{n}{V}\right)^2\right]\left(\frac{V}{n} - b\right) = R.T \tag{1}$$

with R the universal gas constant, and n the number of moles.

An isotherm curve is obtained by fixing a value of the temperature T in (1). Analyzing the behavior of the substance for different isotherms, we obtain a *phase diagram*, with different single-phase regions separated by phase boundaries, where transitions occur. In particular, the transition between gas and liquid phases is described through the binodal and the spinodal curves. The former describes the conditions for two distinct phases to coexist, while the latter encloses the region for system's instability. A metastable region is defined in between. Interestingly, neither the binodal nor the spinodal curve can be calculated analytically; instead, they are computed numerically by fitting a set of 2D points for different isotherms. The sequence of points for the binodal curve is given by the leftmost roots, the critical point (the end point of the phase equilibrium curve, at which the liquid and gas phases coexist), and the rightmost roots of the isotherms. The spinodal curve is the curve connecting the sequence of points given by the local minimum, the critical point, and the local maximum of the isotherms. Then, polynomial data fitting is applied to obtain the characteristic curves. However, this approach is limited and can be improved through free-form parametric curves [3]. In this paper we focus on Bézier curves. For this case, data parameterization is required in addition to poles computation. This leads to a continuous nonlinear optimization problem unsolvable by mathematical optimization techniques. To address this problem, we apply a powerful swarm intelligence method called firefly algorithm.

II. Problem to be Solved

In this paper, we consider the VdW EoS given by (1). Given a set of fixed temperatures $T_1, T_2, \ldots, T_M$, its graphical representation in the P–V plane is a set of curves called *isotherms*, each corresponding to a certain temperature T_i.

Without loss of generality, we can assume $n = 1$ in Eq. (1) for simplicity. Multiplying it by V^2/P and rearranging terms, we get a cubic polynomial in V:

$$V^3 - \left(b + \frac{RT}{P}\right)V^2 + \frac{a}{P}V - \frac{ab}{P} = 0 \tag{2}$$

which has either one or three real roots. The first case occurs for temperatures T higher than the critical temperature, T_c, a value that is characteristic of each substance. The second case happens for temperatures lower than T_c, when the isotherms oscillate up and down. Both cases are separated by the isotherm for $T = T_c$, when the three real roots merge into a single point (triple root), called the critical point. We focus on the case $T < T_c$, where the isotherms have three real roots, labelled in increasing order as R_1, R_2 and R_3. For a liquid-gas system, the end roots, R_1 and R_3, correspond to the liquid

and the vapor phases, respectively. The second root, R_2, has not physical meaning; it is associated with an unstable molar volume and does not represent any real behavior.

Starting with a temperature $T < T_c$ and increasing it until $T = T_c$ makes R_1 and R_3 move towards each other, until they merge into a single point for $T = T_c$, when the three roots become identical and there is no distinction between a liquid and vapor. The critical values of the VdW EoS of a gas depend only on parameters a and b as [7], [10]: $V_c = 3b$, $P_c = \frac{a}{27b^2}$, $T_c = \frac{8a}{27bR}$. Working with the reduced dimensionless variables $(T_r, P_r, V_r) = \left(\frac{T}{T_c}, \frac{P}{P_c}, \frac{V}{V_c}\right)$, Eq. (2) becomes:

$$V_r^3 - \frac{1}{3}\left(1 + \frac{8T_r}{P_r}\right)V_r^2 + \frac{3}{P_r}V_r - \frac{1}{P_r} = 0 \qquad (3)$$

The stable states of the isotherms for $T < T_c$ are divided into two parts: the first one, on the left, corresponds to the liquid states. The second part, on the right, corresponds to the gas states. They are separated by a set of unstable states. This behavior violates the conditions of stability for thermodynamic equilibrium. A solution to this situation was proposed in [8] (the *Maxwell's equal area rule*): the problem would be fixed if the oscillating part of the isotherm between the diluted liquid and the diluted gas is replaced by a horizontal line. The height of this horizontal line should be taken such that the two regions enclosed by the isotherm curve and the line have the same area.

Now, we generate the list of data points for the binodal curve, $\mathcal{B}$, by considering a set of temperatures $T_1 < T_2 < \ldots T_M < T_c$. For each temperature T_j, $(j = 1, \ldots, M)$ there exists a value of the pressure, say P_j^*, for which the areas of the isotherm curve below and above the horizontal line $P = P_j^*$ are equal. This value is computed through an iterative optimization procedure starting from an initial guess $\tilde{P}_j$. After this step, the corresponding roots of the isotherm, denoted as R_k^j, for $(k = 1, 2, 3)$, can be computed as the intersection of the isotherm curve for T_j and the horizontal line $P = P_j^*$.

The list of data points for the spinodal curve, $\mathcal{S}$, are computed by solving $dP/dV = 0$, and checking the sign of the second derivative, d^2P/dV^2 at the obtained solutions. The values for which $dP/dV > 0$ correspond to unstable states of a substance. This situation occurs for the values of the volume between the local minimum, $\mathbf{l}_j$ and the local maximum, $\mathbf{L}_l$ of the isotherm. Note that in this paper vectors are denoted in bold. Then, $\mathcal{S}$, is given by: $\mathcal{S} = \{\{\mathbf{l}_j\}_j, (1,1), \{\mathbf{L}_j\}_j\}_{j=1,\ldots,M}$.

Once $\mathcal{B}$ and $\mathcal{S}$ are obtained, data fitting is usually applied to reconstruct the binodal and spinodal curves, in the form of least-squares optimization. In this work, we consider φ to be a parametric Bézier curve $\mathbf{\Phi}(\tau)$ of degree η, defined as [3]:

$$\mathbf{\Phi}(\tau) = \sum_{j=0}^{\eta} \mathbf{\Delta}_j \phi_j^\eta(\tau) \qquad (4)$$

where $\mathbf{\Delta}_j$ are vector coefficients called *poles*, $\phi_j^\eta(\tau)$ are the *Bernstein polynomials of index j and degree η* and τ is the *curve parameter*, defined on $[0, 1]$. Given a list of 2D data points $\{\mathbf{\Theta}_\mu\}_{\mu=\mathbf{1},\ldots,\chi}$ of the binodal and spinodal curves, we seek to minimize the error:

$$\Upsilon = \sum_{\mu=1}^{\chi}\left(\mathbf{\Theta}_\mu - \sum_{j=0}^{\eta}\mathbf{\Delta}_j\phi_j^\eta(\tau_\mu)\right)^2 \qquad (5)$$

where we need the value τ_μ associated with each data point $\mathbf{\Theta}_\mu$. Taking the vectors $\mathbf{\Phi}_j = (\phi_j^\eta(\tau_1), \ldots, \phi_j^\eta(\tau_\chi))^T$, where $(.)^T$ means transposition, and $\bar{\mathbf{\Theta}} = (\mathbf{\Theta}_1, \ldots, \mathbf{\Theta}_\chi)$, Eq. (5) can be compacted as: $\mathcal{M}\,\mathcal{D} = \mathcal{R}$ with $\mathcal{M} = \sum_{j=1}^{\chi}\phi_l^\eta(\tau_j)\phi_i^\eta(\tau_j)$, $\mathcal{R} = \sum_{j=1}^{\chi}\mathbf{\Theta}_j\phi_l^\eta(\tau_j)$, and $\mathcal{D} = (\mathbf{\Delta}_0, \ldots, \mathbf{\Delta}_\eta)^T$, for $i, l = 0, \ldots, \eta$. If τ_i are unknowns, the problem becomes a difficult continuous nonlinear optimization problem, unsolvable with classical optimization techniques [2].

```
Initialize the firefly population P^0 = {s_i^0}_{i=1,...,N_P}
while t < T_max do
    Generate υ^t ← UpsilonNew()
    Evaluate s_i^t ← EvaluateFFA(P^t, φ)
    Sort the s_i^t according to the fitness function
    s* ← Best(P^t)
    Compute s^{t+1}                    // eq. (6)
    t ← t + 1
end while
return s*
```

Algorithm 1: Firefly algorithm pseudocode

III. The Firefly Algorithm

The *firefly algorithm* (FFA) is a bio-inspired swarm intelligence algorithm introduced in 2008 to solve optimization problems [12], [13] and applied to data fitting in [5], [6]. The firefly algorithm is based on three particular idealized rules, which account for some of the major flashing characteristics of real fireflies [12]. They are: (1) all fireflies are unisex, so that one firefly will be attracted to other fireflies regardless of their sex; (2) the degree of attractiveness of a firefly is proportional to its brightness, which decreases as the distance from the other firefly increases. For any two flashing fireflies, the less brighter one will move towards the brighter one. If there is not a brighter firefly, it will then move randomly; (3) the brightness or light intensity of a firefly is determined by the value of the objective function of a given problem.

The movement of a firefly i at time $t+1$ which is attracted by a more attractive (i.e., brighter) firefly j is given by:

$$\mathbf{s}_i^{t+1} = \mathbf{s}_i^t + \beta_0 e^{-\gamma(r_{ij}^t)^2}(\mathbf{s}_j^t - \mathbf{s}_i^t) + \upsilon\left(\sigma - \frac{1}{2}\right) \qquad (6)$$

where $\mathbf{s}_i^t$ is the current position of the firefly, the second term accounts for the attractiveness of the firefly to light intensity seen by other fireflies (where r_{ij}^t is the distance to firefly i, β_0 is the initial attractiveness at $t = 0$, and γ is an absorption coefficient at the source that controls the decrease of the light intensity), and the third term is used for the random movement

of a firefly in case there is not a brighter one. The coefficient υ is a randomization parameter, while σ is a random number uniformly distributed on the interval $[0, 1]$.

Algorithm 1 shows the FFA pseudocode. The method considers a population $\mathcal{P}^0$ initialized randomly. The main search is performed within the while loop (lines 2-9): firstly, the function *UpsilonNew*() modifies the value of parameter υ (line 3). Then, the function *EvaluateFFA* evaluates the quality of the solution (line 4). All solutions at time t are sorted according to the fitness function (line 5) and the best value is selected and assigned to the global best $\mathbf{s}^*$ (line 6). The new positions are computed according to the Eq. (6). The process is repeated iteratively. Finally, the best individual at final iteration is selected as the solution of the optimization problem (line 10).

IV. The Method

The input of our problem consists of such parameters a and b and a list of temperatures $T_1 < T_2 < \cdots < T_M$ assumed to be below the corresponding critical temperature, T_c. Our method consists of the following steps:

(1) Compute the critical values V_c, P_c, T_c.

(2) Compute the reduced variables V_r, P_r, T_r.

(3) Compute the isotherms at input temperatures T_j.

(4) For each isotherm of temperature T_j:

(4a) Consider an initial guess $\tilde{P}_j$ and perform optimization to compute the value of P_j^*, according to Maxwell's rule.

(4b) Use P_j^* to compute the roots of (3).

(4c) Obtain the local optima of (3).

The output of steps (4a) and (4b) are the lists $\mathcal{B}$ and $\mathcal{S}$ of data points for the binodal and the spinodal curves, respectively.

(5) Perform data fitting with Bézier curves on $\mathcal{B}$ and $\mathcal{S}$:

(5a) Apply the firefly algorithm to perform data parameterization for the lists $\mathcal{B}$ and $\mathcal{S}$.

(5b) Apply least-squares optimization to compute the poles of the curve. The resulting system of equations can readily be solved by well-known numerical procedures [9].

Step (5a) of this workflow is the most critical part of the method and the most important contribution of this paper. It requires an adequate representation of the free variables of our optimization problem. Our fireflies, $\mathcal{F}_k$, are real-valued vectors of length M corresponding to a parameterization of data points: $\mathcal{F}_k = (\rho_1^k, \rho_2^k, \ldots, \rho_M^k) \subset [0, 1]^M$. They are initialized with uniformly distributed random numbers. The $\{\rho_i^k\}_i$ are sorted in increasing order to replicate the ordered structure of data parameterization. The fitness function corresponds to the least-squares function in (5). We also compute the RMSE (root-mean squared error), given by: $RMSE = \sqrt{\Upsilon/M}$. To determine the curve degree, η, we compute the RMSE for values of η ranging from the minimum $\eta = 2$ to $\eta = 9$. The value minimizing the RMSE is finally selected. Finally, the method depends of the control parameters in Table I.

TABLE I
BAT ALGORITHM PARAMETERS AND THEIR VALUES IN THIS PAPER.

Symbol	*Meaning*	*Range*	*Used Value*
$N_{\mathcal{P}}$	population size	50-300	100
$\mathcal{T}_{max}$	max. number of iterations	10,000–100,000	50,000
β_0	initial attractiveness	$(0, 2)$	1
γ	absorption coefficient	$(0, 1)$	0.5
μ	potential coefficient	$(0, 5)$	2
f_{max}	maximum frequency	$[0, 10]$	1.5
α	randomization parameter	$[0, 1]$	0.5

TABLE III
COMPARATIVE RESULTS OF OUR METHOD WITH OTHER ALTERNATIVE APPROACHES (BEST RESULTS IN BOLD).

Method	*Binodal: best RMSE*	*Spinodal: best RMSE*
Polynomial fitting:	1.480733E−2	1.447021E−1
Multilayer perceptron:	9.098664E−3	6.477843E−3
Our method:	**4.080013E−4**	**5.580561E−4**

V. Experiments and Results

Our method has been applied to the VdW EoS of the Argon (Ar). Its parameter values are [11]: $a = 1.337$ L^2.atm.mol^{-2}, and $b = 0.03201$ L.mol^{-1}. The critical temperature is $T_c = 150.86$ K, with an uncertainty of 0.1 K according to [1], [4].

We applied the steps (1)-(3) of our workflow for the sets of temperatures $\{130,133,135,137,140,142,145,147,148,149,T_c\}$ and$\{128,130,133,135,137,140,142,145,147,148,149,150.2,T_c\}$ for the binodal and the spinodal curves, respectively. To remove the stochastic effects, 30 independent executions are run for each value of η and the 10 worst ones are rejected. Tables II reports our results from $\eta = 2$ to $\eta = 9$ (in rows). The following data are listed (in columns): curve degree, best RMSE, and RMSE mean (20 best runs). We also checked for overfitting (column 4): symbol ✓ indicates that it occurs, symbol $\times$ otherwise. Finally, we also report the stability of the solutions (column 5), from 1 (very low) to 5 (very high) using the diamond $\blacklozenge$ symbol as many times as the value.

We can see that the method works pretty well. Best and mean RMSE are of order 10^{-3} or better for all degrees but $\eta = 2$. Best results are obtained for $\eta = 4$, although $\eta = 3$ and $\eta = 5$ also yield good results. Also, RMSE values from $\eta = 4$ to $\eta = 9$ are generally of the same order, which is consistent with the fact the high-degree functions tend to fit the data points better, at the expense of higher model complexity. Another issue is that too many DOFs might lead to overfitting. We checked this issue found that it happens for models of degree $\eta \geqslant 6$ and $\eta \geqslant 5$ for the binodal and the spinodal curves, respectively. This means that only low degree curves are good predictors for other temperature values. This is confirmed by the stability analysis: the most stable solutions occur for $\eta = 4$ and $\eta = 5$ for both binodal and spinodal curves.

Figure 1 shows the best fitting curves obtained for the binodal (left) and spinodal (right) curves. The original and the reconstructed data points are shown as red filled circles and white upward triangles, respectively. The figure also shows the best Bézier fitting curve, displayed as a blue solid line. Note the excellent matching between original and reconstructed data points, and the very good fitting through the Bézier curve.

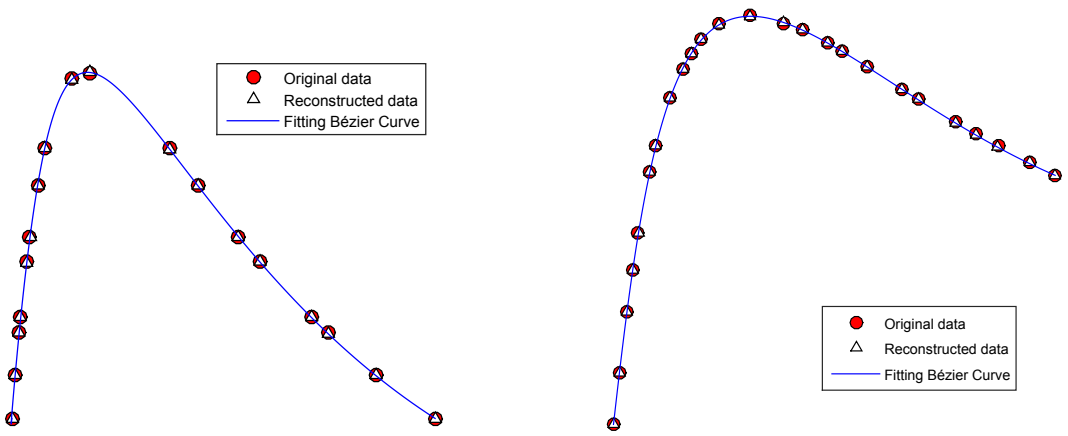

Fig. 1. Original and reconstructed data and best Bézier fitting curve for the binodal (left) and spinodal (right) curves.

TABLE II
COMPUTATIONAL RESULTS FOR THE BINODAL (LEFT) AND THE SPINODAL (RIGHT) CURVES.

Degree	*RMSE (best)*	*RMSE (mean)*	*Overfitting*	*Stability*
$\eta = 2$	4.729115E−1	7.900833E−1	×	◆
$\eta = 3$	1.112042E−3	2.674914E−3	×	◆ ◆ ◆
$\eta = 4$	4.080013E−4	5.649133E−4	×	◆ ◆ ◆ ◆
$\eta = 5$	8.544711E−4	1.637550E−3	×	◆ ◆ ◆ ◆
$\eta = 6$	1.052612E−3	1.467992E−3	✓	◆ ◆ ◆
$\eta = 7$	5.654309E−4	8.965516E−4	✓	◆ ◆
$\eta = 8$	7.273016E−4	9.933128E−4	✓	◆ ◆ ◆
$\eta = 9$	9.015265E−4	1.163426E−3	✓	◆ ◆

Degree	*RMSE (best)*	*RMSE (mean)*	*Overfitting*	*Stability*
$\eta = 2$	6.775012E−2	9.821750E−2	×	◆
$\eta = 3$	2.637961E−3	3.574954E−3	×	◆ ◆ ◆
$\eta = 4$	5.580561E−4	7.083656E−4	×	◆ ◆ ◆ ◆
$\eta = 5$	9.374015E−4	1.063998E−3	✓	◆ ◆ ◆ ◆
$\eta = 6$	7.530443E−4	8.205116E−4	✓	◆ ◆
$\eta = 7$	9.978001E−4	1.140396E−3	✓	◆ ◆ ◆
$\eta = 8$	8.022364E−4	1.079352E−3	✓	◆ ◆
$\eta = 9$	7.895345E−4	9.969872E−4	✓	◆ ◆ ◆

We also compared our method with two state-of-the-art approaches: polynomial curve fitting and a multilayer perceptron (MLP) neural network with 5 neurons in a single hidden layer. Table III reports the comparative results for the best RMSE. Best results are highlighted in bold. Our method outperforms these two state-of-the-art approaches significantly: our results are better by one or two orders of magnitude.

VI. CONCLUSIONS AND FUTURE WORK

In this paper, we introduce a new method to construct the characteristic curves of the VdW EoS by data fitting using free-form Bézier curves. The method is applied to Argon, a noble gas for which the characteristic curves are reconstructed with very good accuracy. Furthermore, it outperforms two state-of-the-art methods by orders of magnitude for the given example. All computations run on a 2.9 GHz. Intel Core i7 processor with 8 GB. of RAM. The source code has been implemented by the authors in *Matlab*, version 2015b. About the CPU times, a single run takes about 18–22 minutes. Our future work includes reducing the CPU times, increasing the accuracy of our method, applying our method to other interesting chemical components and mixtures, and extending this approach to other EoS in the literature.

ACKNOWLEDGMENT

Research supported by the EU Horizon 2020 project PDE-GIR (MSCA grant No 778035) and project #TIN2017-89275-R, Agencia Estatal de Investigación (AEI), Spanish Ministry of Science, Innovation and Universities, and UE FEDER funds.

REFERENCES

[1] Angus, S., Armstrong, B., Gosman, A.L., McCarty, R.D.,Hust, J.G., Vasserman, A.A., Rabinovich, V.A.: *International Thermodynamic Tables of the Fluid State - 1 Argon*, Butterworths, London (1972).
[2] Dierckx, P.: *Curve and Surface Fitting with Splines*. Oxford University Press, Oxford (1993).
[3] Farin, G.: *Curves and surfaces for CAGD (5th ed.)*. Morgan Kaufmann, San Francisco (2002).
[4] Gosman, A.L., McCarty, R.D., Hust, J.G.: Thermodynamic properties of Argon from the triple point to 300 K at pressures to 1000 atmospheres. *Nat. Stand. Ref. Data Ser., Nat. Bur. Stand.*, NSRDS-NBS 27 (1969).
[5] Iglesias, A., Gálvez, A.: Memetic firefly algorithm for data fitting with rational curves. In: *Proc. IEEE CEC'2015*, IEEE Computer Society Press, Los Alamitos, CA, 507–514 (2015).
[6] Iglesias, A., Gálvez, A.: New memetic self-adaptive firefly algorithm for continuous optimisation. *Int. J. of Bio-Insp. Comput.*, **8**(5), 300–317 (2016).
[7] Johnson, D. C.: *Advances in thermodynamics of the van der Waals Fluid*. Morgan & Claypool Publishers, Ames, (2014).
[8] Maxwell, J.C.: On the dynamical evidence of the molecular constitution of bodies". *Nature*, **11**, 357–359 (1875).
[9] Press, W.H., Teukolsky, S.A., Vetterling, W.T., Flannery, B.P.: Numerical Recipes (second ed.). Cambridge University Press, Cambridge (1992).
[10] Smith, J.M., Van Ness, H.C., Abbott, M.M.: *Introduction to chemical engineering thermodynamics*. McGraw-Hill, Boston (2005).
[11] Weast, R.C.: *Handbook of Chemistry and Physics (53rd Edition)*. Chemical Rubber Pub. (1972).
[12] Yang, X.S.: Firefly algorithms for multimodal optimization. *Lectures Notes in Computer Science*, **5792**, 169–178 (2009).
[13] Yang, X.S.: Firefly algorithm, stochastic test functions and design optimisation. *Int. J. of Bio-Insp. Comput.*, **2**(2) 78–84 (2010).