

Accelerated Phase Equilibrium Predictions for Subsurface Reservoirs Using Deep Learning Methods

Tao Zhang¹[0000-0001-7216-0423], Yiteng Li¹[0000-0001-8293-279X], and Shuyu Sun¹[0000-0002-3078-864X]

Computational Transport Phenomena Laboratory (CTPL), King Abdullah University of Science and Technology (KAUST), Thuwal 23955-6900, KSA
shuyu.sun@kaust.edu.sa

Abstract. Multiphase fluid flow with complex compositions is an increasingly attractive research topic with more and more attentions paid on related engineering problems, including global warming and green house effect, oil recovery enhancement and subsurface water pollution treatment. Prior to study the flow behaviors and phase transitions in multi-component multiphase flow, the first effort should be focused on the accurate prediction of the total phase numbers existing in the fluid mixture, and then the phase equilibrium status can be determined. In this paper, a novel and fast prediction technique is proposed based on deep learning method. The training data is generated using a selected VT dynamic flash calculation scheme and the network constructions are deeply optimized on the activation functions. Compared to previous machine learning techniques proposed in literatures to accelerate vapor liquid phase equilibrium calculation, the total number of phases existing in the mixture is determined first and other phase equilibrium properties will be estimated then, so that we do not need to ensure that the mixture is in two phase conditions any more. Our method could handle fluid mixtures with complex compositions, with 8 different components in our example and the original data is in a large amount. The analysis on prediction performance of different deep learning models with various neural networks using different activation functions can help future researches selecting the features to construct the neural network for similar engineering problems. Some conclusions and remarks are presented at the end to help readers catch our main contributions and insight the future related researches.

Keywords: deep learning · phase equilibrium · multi-component multiphase flow

1 introduction

Subsurface multiphase fluid flow has attracted increasing attentions from researchers all over world, due to its various applications such as energy industry,

including enhanced oil recovery and unconventional oil and gas development, and environmental production, including greenhouse effect and subsurface water pollution control. In petroleum industry, the study of multi-component multiphase flow is needed in all the exploitation and transportation stages to optimize the oil field development for the maximization of oil recovery [1, 2, 9, 10, 12, 11, 13]. Meanwhile, the simulation of subsurface multi-component multiphase fluid flow is critical to handle a large range of environmental concerned issues, for example, greenhouse gas emission and pollutant disposal. Especially, for the rapid development of CO_2 sequestration technique, phase compositions and phase transition behaviors are urgently needed for the plan designing and optimization to better reduce the CO_2 emission or controllably dispose the environmentally hazardous wastes[3, 4]. In addition, the production of unconventional oil and gas resources has experienced a significant growth in all over the world and successfully becomes an important energy supply, due to the rapid production decline of conventional reservoirs [8]. In order to maximize the oil production from conventional and unconventional reservoir or resolve the aforementioned environment issues, it is great demand of the accurate numerical model of subsurface multiphase fluid systems, as well as the robust and efficient computational algorithm.

A number of studies have investigated the phase equilibrium problems in subsurface reservoirs, which often depend on different sets of given conditions [5, 7]. One conventional phase equilibrium calculation approach is performed under constant chemical compositions, pressure and temperature, which is known as "NPT" flash. The long history of NPT flash has brought a wide range of applications related with phase splitting calculation and stability test, but at the same time some limitations have been found regarding this method. In some specified conditions, the system equilibrium cannot be determined uniquely, or the solution from the flash calculation fails to identify a clear and determined state. For the cubic equation of state, two separate solutions need to be conducted to determine the phase molar volumes of each phase. Furthermore, the root selection procedure has to be considered in the presence of multiple roots, although the middle root is often ignored because which corresponds to a physically meaningless solution. The remaining roots need to be paired to minimize the Gibbs free energy, and the root pairing could be challengeable as there might be two roots for each phase. If the selection of pairing is wrong, the whole procedure will fail with an unstable or metastable solution. Another limitation is that pressure is not always a priori, which makes it inefficient to iteratively solve phase equilibria problems of a different variable specification by the NPT flash. Another approach with priori conditions including constant chemical compositions, molar volume and temperature, namely the NVT flash, has been proposed to handle the above challenges faced by NPT flash. As an alternative, it has shown great potential in compositional multiphase flow simulation in subsurface porous media. However, the problem existing in the application of both the two types is the much CPU time cost in engineering scale. Generally, the temperature-pressure range or temperature-volume range, corresponding to NPT flash and NVT flash respectively, can be so wide that the iterations in each environmen-

tal condition should be reused for many times. Especially for the complex fluid mixture with a large number of components, both the two types of NPT and NVT flash calculations have problems in the quick but reliable phase equilibrium prediction.

To speed up flash calculation, different approaches have been proposed aiming at finding a path faster to estimate the phase equilibrium conditions. In this paper, we will review the general deep learning method and the detailed process, to help readers new in this area get a basic understanding of how this method can be used in the prediction. Comments on recent techniques proposed in deep learning methods have been involved with authors' own opinions, to show the current trend for a better prediction accuracy in engineering need. The performance of neural network models with different activation methods are tested from both the convergence analysis of total loss function and mean estimation errors. Furthermore, the determination of total phase numbers existing in the mixture, which is the key step in the numerical study of compositional multiphase fluid flow in subsurface reservoirs is used as the main test target of our prediction. The performance of different activation functions, on the loss function convergence rate as well as mean absolute and mean relative errors, are compared and suggestions are made on future studies regarding similar problems.

The remainder of this paper is organized as follows. In Section 2, the deep learning method is explained in details, with clear procedures and instructions with author's own comments. Prediction examples are presented in Section 3 to show the robustness and efficiency of the concluded scheme. At the end, we make some conclusions in Section 4.

2 Deep Learning Method

Artificial neural networks (ANNs) are computational models designed to incorporate and extract key features of the original inputs and process data in a manner analogous to neurons in animals central nervous systems (in particular the brain), which are capable of both machine learning, as well as pattern recognition, the former sometimes being called supervised and the latter unsupervised machine learning. The naming convention stems from the fact that in supervised machine learning, unlike the unsupervised fashion, the ANN is presented with the target variable(s) and seeks to find a functional relationship that can be used to predict the target variable(s) from input variable(s) with a desired degree of precision. Deep neural networks usually refer to those artificial neural networks which consist of multiple hidden layers. In this paper a deep fully connected neural network is applied to model the phase equilibrium calculation with data sourced from VT flash calculation are used as input. Following the input layer, a number of fully connected hidden layers, with a certain number of nodes, stack over the other, whose final output is fed into another fully connected connected layer, which is the final output layer. Since we are fitting the compositions of vapor phase and liquid phase in our model, the final output layer contains several

nodes, which could be divided into three sets: one for total phase numbers existing in the mixture, one for vapor phase compositions and the final one for liquid phase compositions. The activation function of this layer is fixed as linear. Naturally the proposed ANN input variables include critical pressure (P_c), critical temperature (T_c), acentric factor (ω) and z of the components comprising the mixture, as well as the temperature T and mole volume c as the environmental condition. Different from previous studies, a much larger range of T and c can be included in our environmental conditions, as it is not a priori in our approach that the mixture should be ensured in a two phase area. On the contrast, our algorithm can decide whether the mixture is in single phase or two phase area, and then process the phase equilibrium estimation accordingly.

The whole package is developed using TFlearn. Trained on a Mac Laptop, which is a common equipment, the training iterations will converge in less than 10 minutes if the source data is with size 151×151 . Formally, for the i -th hidden layer, let \mathbf{a}_i denote the input of the layer, and \mathbf{y}_i to denote the output of the layer. Then we have:

$$\mathbf{y}_i = f_i(\mathbf{W}_i * \mathbf{a}_i + \mathbf{b}_i), \quad (1)$$

where \mathbf{W}_i is the weight; \mathbf{b}_i is the bias; and f_i is the activation functions of the i -th layer. For a network with multiple layers, the output of one hidden layer is the input of the next layer. For example, we can represent the network in Figure ?? as:

$$\mathbf{o} = f_3(\mathbf{W}_3 * f_2(\mathbf{W}_2 * f_1(\mathbf{W}_1 * \mathbf{x}_1 + \mathbf{b}_1) + \mathbf{b}_2) + \mathbf{b}_3), \quad (2)$$

where $\mathbf{o} = (X, T)$; f_1, f_2, f_3 are the activation functions; $\mathbf{W}_1, \mathbf{W}_2, \mathbf{W}_3$ are the weights for each layer; $\mathbf{b}_1, \mathbf{b}_2, \mathbf{b}_3$ are the bias terms of each layer

As shown in Equation (4), the activation function is where the non-linearity and the expressiveness power of deep neural network models comes from. There are numerous activation functions: Rectified linear unit (ReLU), Parametric rectified linear unit (PReLU), TanH, Sigmoid, Softplus, Softsign, Leaky rectified linear unit (Leaky ReLU), Exponential linear unit (ELU), and Scaled exponential linear unit (SELU). Here, we present the formular for two common used activation functions, ReLU and Sigmoid, as an example.

$$f(x) = \begin{cases} 0, & \text{if } x < 0; \\ x, & \text{if } x \geq 0, \end{cases} \quad (3)$$

$$\sigma(x) = \frac{1}{1 + \exp(-x)}. \quad (4)$$

3 Examples

In this section, the predictions of phase total phase numbers existing in a fluid mixture in various complex reservoir environmental conditions are performed based on the 8 components mixture detected in EagleFord Oilfield. Detailed component compositions and the parameters effecting phase equilibrium results

using VT flash calculations based on Peng-Robinson Equation of State are presented in Table. 1. The VT flash calculation approach is selected as described in [8, 6], which is energy stable so as to accelerate the flash calculation for the large scale data. Besides, this scheme is consistent with the first and second thermodynamical laws, which makes the phase equilibrium results more reliable. Unlike machine learning methods reported in previous literatures, the environmental conditions need no more to ensure the mixtures to be in a two phase area, so that the applications of our method can be larger. The number of total phases is the result predicted using our method, and other phase equilibrium properties can be estimated correspondingly.

Using the deep learning approach described in Section 2, a neural network is designed with 5 hidden layers, 100 nodes in each layer, totally 4000 iterations. The performance of using different activation functions are tested in this section, to help provide a suggestion for future researchers. A 101×101 VT flash data source is generated for this mixture, so the total original data for training is 9180, and the tested data is 1021. Namely, the tested porportion in this paper is selected as 10 per cent. Especially, the key effort in phase equilibrium calculation, e.g. the decision of total phase numbers in a mixture, is also processed using our deep learning approach to test its capability to handle this problem.

3.1 Deep Learning Model Training

As explained in Section 2, the model takes the parameters of each components, as well as temperature (T) and mole volume (c) as the input and outputs the predicted value of X and Y in each phase. The key parameters of the model are the weights of each layer, which control what the model outputs given the input. At first, those weights are initialized randomly, which means that the model will output useless values given the inputs. To make the model useful for this problem, we need to train those weight parameters to fit our problem. The difference between the models output and the ground truth is referred as loss. Here, for this regression problem, we use mean square error as the loss function. 10% of the data sources are selected as the test data, where the remaining data are input into the network for the training. The mean absolute error and mean relative error of the tested results from different networks with various activation functions used are listed in Table 2. As indicated from the results in [14], four activation functions are selected from the low error group: 'tanh', 'relu', 'sigmoid' and 'softsign' and one activation function selected from the high error group 'softplus' to test whether their performance will show similar results. It is glad to see from Table 2 that generally the four activations in low error group in [14] will show better estimation errors compared with the high error activation function 'softplus'. It is verified the statement resulted from the binary components cases can be extended to the complex multi-component cases.

The loss function curve decreasing with iterations are also presented to show how the trained outputs are approaching the true value. Results from iteration 200 to iteration 400 are drawn as the loss function in this period represent a clear approaching trend. From Figure 1, it can be referred that the total loss

Table 1. Molar composition and compositional properties for the EagleFord2 oil.

| Component | z_i | $T_{c,i}$ (K) | $P_{c,i}$ (MPa) | $M_{w,i}$ (mol/m ³) | ω_i |
|-----------|---------|---------------|-----------------|---------------------------------|------------|
| C_1 | 0.31231 | 190.72 | 4.6409 | 16.04 | 0.0130 |
| N_2 | 0.00073 | 126.22 | 3.3943 | 28.01 | 0.0400 |
| C_2 | 0.04314 | 305.44 | 4.8842 | 30.07 | 0.0986 |
| C_3 | 0.04148 | 369.89 | 4.2568 | 44.10 | 0.1524 |
| CO_2 | 0.01282 | 304.22 | 7.3864 | 44.01 | 0.2250 |
| iC_4 | 0.01350 | 408.11 | 3.6480 | 58.12 | 0.1848 |
| nC_4 | 0.03382 | 425.22 | 3.7969 | 58.12 | 0.2010 |
| iC_5 | 0.01805 | 460.39 | 3.3336 | 72.15 | 0.2223 |
| nC_5 | 0.02141 | 469.78 | 3.3750 | 72.15 | 0.2539 |
| nC_6 | 0.04623 | 507.89 | 3.0316 | 86.18 | 0.3007 |
| C_{7+} | 0.16297 | 589.17 | 2.7772 | 114.40 | 0.3739 |
| C_{11+} | 0.12004 | 679.78 | 2.1215 | 166.60 | 0.5260 |
| C_{15+} | 0.10044 | 760.22 | 1.6644 | 230.10 | 0.6979 |
| C_{20+} | 0.07306 | 896.78 | 1.0418 | 409.20 | 1.0456 |

Table 2. Estimation error of deep learning model trained with different activation functions

| activation functions | mean absolute error | mean relative error |
|----------------------|---------------------|---------------------|
| softplus | 0.00679 | 81540 |
| tanh | 0.01342 | 0.01455 |
| relu | 0.01378 | 0.01723 |
| sigmoid | 0.01900 | 0.02066 |
| softsign | 0.01398 | 0.01499 |

function using activation function 'softplus' is much larger than that of the other four neural networks. The loss function of 'sigmoid' and 'softsign' show similar lowest value, but they remain almost steady compared to the constant decrease of the network generated using 'tanh' and 'relu'.

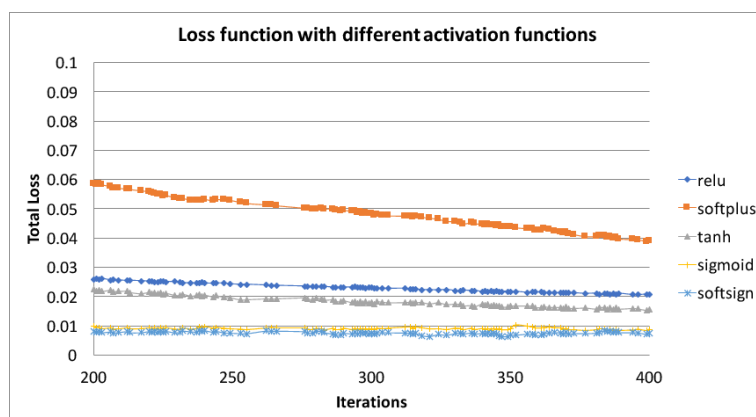


Fig. 1. The neural network to model VLE.

3.2 Phase Number Characterization

The key effort of phase equilibrium calculation is to decide the total number of phases existing in the mixture at certain conditions, which is priori to further prediction of phase behaviors. To test the capability of the deep learning approach to capture this value, a mixture of 8 components detected in EagleFord Oilfield is selected to generate the data source using our proposed VT flash calculation algorithm, with the data size of 101×101 and at the constant $c = 129.9$. As shown in Fig 2, it can be referred that with the temperature increasing from 250K to 850K, this mixture will change from a two phases mixture to a single phase mixture. For all the five activation functions used in constructing the neural network, generally the deep learning result meet well with the original data generated from VT flash calculation, with only a small difference at the phase transition point. It is obvious to see that the results predicted using activation function 'sigmoid' and 'softplus' have the maximum prediction errors, with a larger phase change temperature point and a smaller temperature respectively. Meanwhile, the result of deep learning neural networks generated using activation function 'softsign' meets well with the original data resulted from VT flash data, with only a small difference on the prediction of phase change temperature point. Thus, it can be concluded that the activation function 'softsign' best fits the vapor liquid equilibrium problems, especially on the total phase numbers determination.

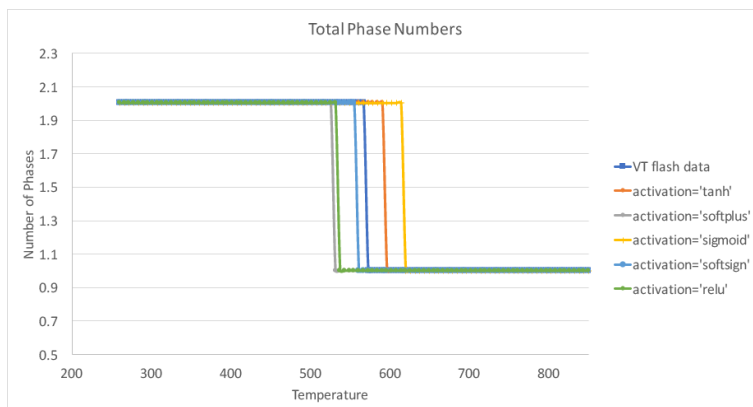


Fig. 2. Number of phases predicted in a mixture with different activation functions used in the neural network at $c = 129.9$

Combined with the error analysis in Section 3.1, it could be referred that activation function 'relu' and 'softsign' can be suggested for future deep learning methods considering to construct the neural networks. To prove this statement, another phase equilibrium prediction is performed with our trained model for constant mole volume set as $c = 249.8$. As shown in Fig. 3, our statement is verified with the fact the results using activation function 'sigmoid' and 'softplus' deviate the most with the original flash calculation data, while the result of activation function selected as 'softsign' totally agree with the original data. The result of activation function 'relu' also meets well with the VT flash result, while the 'tanh' activation function performs not as good in both the two cases.

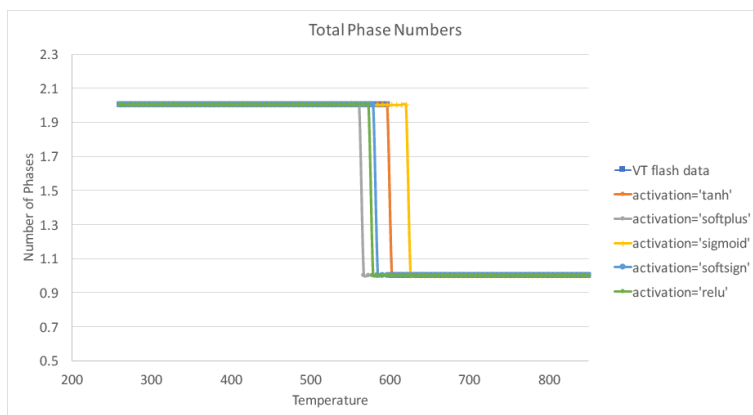


Fig. 3. Number of phases predicted in a mixture with different activation functions used in the neural network at $c = 249.8$

4 Conclusion

The concept and detailed procedures of deep learning method is described in this paper, especially instructions for the construction of deep neural networks used to predict vapor liquid phase equilibrium. Previous methods for phase equilibrium calculations are introduced and analyzed first, and this explains why we use VT flash algorithm for complex multicomponent mixtures to generate the original phase equilibrium data used for the training in our deep learning method. The robustness and efficiency of our designed neural network is verified by the accurate predictions for the complex mixture selected from a realistic oil field, and the comparison of different activation functions is presented from three different viewpoints: loss function convergence rate, mean absolute and relative error and total phase number prediction. Based on our results, it could be concluded that different activations perform with different accuracy in this type of problems, where 'relu' and 'softsign' are suggested to construct the deep neural networks. This statement is similar with the comparison of the performance of activation functions in [14] for binary components. It indicates that this optimized configurations set is capable to handle a large range of phase equilibrium problems, from components, and data sources. As the total phase number can be decided using our method, the priori that the mixture should stay in a two phase area is not necessary any more, which greatly increases the capability and application ranges of deep learning methods in the phase equilibrium predictions. On future studies, a potential direction is to take into consider more mechanisms in reservoirs, like capillary pressure, to see the capability of this network and approach.

References

1. Aziz, K.: Petroleum reservoir simulation. Applied Science, London (1979).
2. Dawson, C., Sun, S., Wheeler, M.F.: Compatible algorithms for coupled flow and transport. *Comput. Methods Appl. Mech. Eng.* 193(23–26), 2565–2580 (2004).
3. El-Amin, M., Sun, S., Salama, A.: Modeling and simulation of nanoparticle transport in multiphase flows in porous media: CO₂ sequestration. *Math. Methods Fluid Dyn. Simul. Giant Oil Gas Reserv.* (2012).
4. Espinoza, D.N., Santamarina, J.C.: Water–CO₂–mineral systems: Interfacial tension, contact angle, and diffusion–Implications to CO₂ geological storage. *Water Resour. Res.* 46(7), W07537(1–10) (2010).
5. Firoozabadi, A.: Thermodynamics of hydrocarbon reservoirs. McGraw-Hill, New York (1999).
6. J. Kou, S. Sun, and X. Wang, Linearly decoupled energy-stable numerical methods for multicomponent two-phase compressible flow, *SIAM J. Numer. Anal.*, 56 (2018), pp. 3219–3248.
7. Jindrová, T., Mikyška, J.: General algorithm for multiphase equilibria calculation at given volume, temperature, and moles. *Fluid Phase Equilib.* 393, 7–25 (2015).
8. Li, Y., Kou, J., Sun, S.: Thermodynamically Stable Two-Phase Equilibrium Calculation of Hydrocarbon Mixtures with Capillary Pressure. *Ind. Eng. Chem. Res.* 57(50), 17276–17288 (2018).

9. Mikyška, J., Firoozabadi, A.: A new thermodynamic function for phasesplitting at constant temperature, moles, and volume. *AIChE J.* 57(7), 1897–1904 (2011).
10. Moortgat, J., Sun, S., Firoozabadi, A.: Compositional modeling of threephase flow with gravity using higherorder finite element methods. *Water Resour. Res.* 47(5) W05511(1–26) (2011).
11. Sun, S., Liu, J.: A locally conservative finite element method based on piecewise constant enrichment of the continuous Galerkin method. *SIAM J. Sci. Comput.* 31(4), 2528–2548 (2009).
12. Wu, Y.S., Qin, G.: A generalized numerical approach for modeling multiphase flow and transport in fractured porous media. *Commun. Comput. Phys.* 6(1), 85–108 (2009).
13. Tao Zhang, Jisheng Kou, Shuyu Sun. *Advances in Geo-Energy Research.* 2017, 1(2) : 124-134
14. Li, Y., Zhang, T., Sun, S., Gao, X. (2018). Accelerating Flash Calculation through Deep Learning Methods. arXiv preprint arXiv:1809.07311.