34th Annual International Conference of the IEEE EMBS
San Diego, California USA, 28 August - 1 September, 2012

Identification and Quantification of Mixed Air Pollutants Based on Homotopy Method for Gas Sensor Array

Yuning Yang, *Student Member, IEEE,* Andrew J. Mason, *Senior Member, IEEE*
Department of Electrical and Computer Engineering, Michigan State University, East Lansing, MI, USA

***Abstract*—Accurate recognition of air pollutants and estimation of their concentrations are critical for human health and safety monitoring and can be achieved using gas sensor arrays. In this paper, an efficient method based on a homotopy algorithm is presented for the analysis of sensor arrays responding to binary mixtures. The new method models the responses of a gas sensor array as a system of nonlinear equations and provides a globally convergent way to find the solution of the system. Real data measurement for CH_4 and SO_2 are used to model sensor responses. The model is applied to the method for prediction and it shows the prediction results are within 1% variation of true values for both gas models.**

I. Introduction

Exposure to air pollutions has a serious effect on the health of humans and has become the leading cause of global mortality [1]. Toxic gases and volatile compounds (VOCs) in the air are considered the main reason for human diseases, such as cardiovascular or respiratory problems, lung disease and so on. Therefore, it is necessary to both detect and quantify those harmful gases for effective prevention and treatment plans.

An electronic nose (EN) system is an instrument that is used to measure and monitor target gases in the air. An EN usually combines a gas sensor array, an instrumentation circuit and a sensor array signal processing algorithm for detection, identification and quantization of target gases. Ideally, the sensor array is designed and fabricated in such a way that each individual senor reacts exclusively to each individual target gas in a mixture. Hence, the outputs of sensors correspond to the concentrations of relevant gases. However, because there are several air pollutants that need to be monitored simultaneously, it is difficult to build an exclusively sensitive sensor. Most gas sensors generally respond to more than one target analyte. Due to the cross-sensitive nature of gas sensors, an efficient sensor array signal processing algorithm is imperative to identify the species of target gases in the air and quantify their concentrations.

In most cases, data analysis for sensor signals uses a pattern classification technique [2]. The response of the sensor array would generate different patterns or fingerprints for different gases. These fingerprints serve as features and are usually projected to a lower dimensional space to reduce the computational complexity before classification. Such classification algorithm have been implemented in electronic devices with low power, low cost and portability [3-4]. However, classification techniques are only suitable for discriminating individual gases in a mixture and can only identify one unknown sample at a given time. To measure multiple target air pollutants for more general applications, a sensor array signal processing algorithm must be able to analyze multiple components in a gas mixture. The partial least squares (PLS) technique is widely used in gas mixture analysis due to its ability to handle collinear data and makes a meaningful combination from different concentrations [2,5]. The main limitation of this method is that it is only valid in a small concentration range because most sensors exhibit non-linear concentration dependence. In order to deal with non-linear data and improve performance, a non-linear PLS was introduced [6]. This method shows good results for single or binary gases, but it becomes complex and impossible to solve when applied to more target chemicals. Alternatively, an artificial neural network (ANN) for simultaneous classification and concentration estimation has been proposed [7]. This method requires tremendous data for training to build up an accurate ANN model and suffers an over-fitting problem that performs accurately for training data but generates poor prediction for testing data. A more efficient technique using multiple regression method has been used in some applications [8-9]. This technique models responses of a sensor array as a system of nonlinear equations in which the relationship between output measurements and gas concentration variables is expressed by means of regression. By minimizing the error between the real measurements and the estimated outputs based on searching for optimal concentration variables, gas classification and concentration estimation can be performed simultaneously. The main limitation of the minimization process lies in the convergence issue, that is, the algorithm may not converge or may converge to local minima in some cases.

In this paper, we present a new homotopy algorithm for multiple regression models. This algorithm provides a useful approach to find roots of a system of nonlinear equations in a globally convergent way instead of minimizing error by iteration. The homotopy method starts from a system with easy solutions and gradually transforms into the original system. In this way, it overcomes the convergence issue and seeks optimal solutions of the system, hence providing more accurate results for gas classification and concentration estimation than other methods in air pollution monitoring.

II. Gas Sensor Array

Electrochemical sensing provides a promising approach among gas sensing technologies for a low power and low cost application with good selectivity and sensitivity. Ionic liquids (ILs) have been revealed as encouraging sensing materials under room temperature [10].When exposed to an analyte, the sensor impedance will be changed due to reactions of analytes associated with ILs. The impedance can be measured using an electrochemical impedance spectroscopy method. By

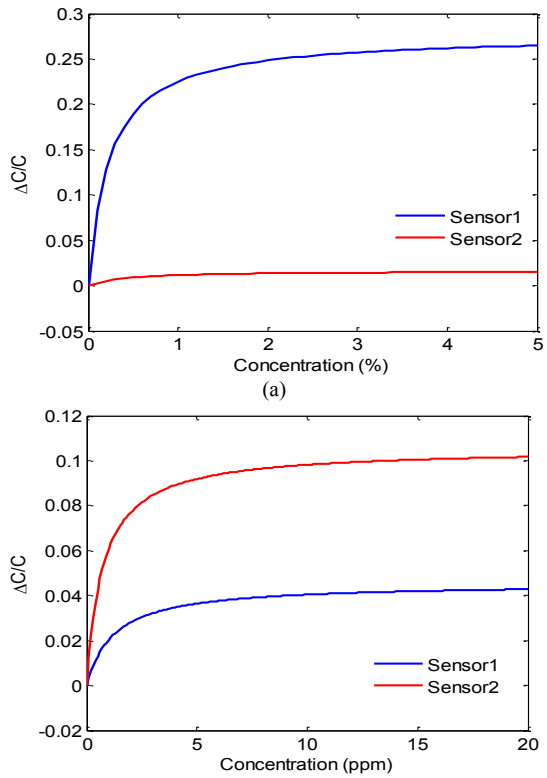

Fig. 1. Normalized sensor responses as a function of concentration for (a) CH_4 and (b) SO_2.

adjusting the DC bias potential to the point of maximum electrochemical response, each sensor element can be tailored to a specific analyte. Although one sensor has high sensitivity to only one analyte, it still has potential sensitivity to secondary analytes. Sensor responses to gases of CH_4 and SO_2 have been shown in Fig. 1, where *C* is the baseline capacitance and ΔC is the change of capacitance. As shown, both sensors have different reactions to target gases. When exposed to the mixture of these two gases, the cross sensitivity will prohibit accurate estimation of each element in the mixture. This demonstrates that an efficient algorithm to remove this cross sensitivity is essential for practical applications.

III. Homotopy Method

As shown in Fig. 1, the output response of a sensor shows different characteristics as a function of concentrations for different gases. To analyze the property of a sensor with respect to a gas, a mathematical model can be described as

$$y_{ij} = f_{ij}(x_j) \tag{1}$$

where y_{ij} is the *i*th sensor output corresponding to the *j*th input gas concentration x_j and f_{ij} represents the model of sensor response. Usually, a polynomial model is employed to fit the curve f_{ij} [6][8-9], expressed by

$$f_{ij}(x_j) = \sum_{k=0}^{K} a_k {x_j}^k \tag{2}$$

where a_k is the coefficient for the *k*th order item. When exposed to a mixed gas, it is assumed that the output response of the *i*th senor y_i is the addition of responses to each individual gas in the mixture, that is

$$y_i = \sum_{j=1}^{N} f_{ij}(x_j) \tag{3}$$

Thus, a system of nonlinear equations is formulated for the sensor array

$$\begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_N \end{bmatrix} = \begin{bmatrix} \sum_{j=1}^{N} f_{1j}(x_j) \\ \sum_{j=1}^{N} f_{2j}(x_j) \\ \vdots \\ \sum_{j=1}^{N} f_{Nj}(x_j) \end{bmatrix} \tag{4}$$

Seeking the roots for the system, the concentration of each individual gas in the mixture can be obtained.

Different methods for solving the system for gas sensor array have been reported in the literature [6,8-9], such as Newton, Levenberg-Marquardt and trust region. The Newton method suffers a critical convergent issue as the starting point is not in a certain neighborhood of solution and thus it is not practical in the real application. The other two methods are based on minimizing the total error between the output measurement y_i and estimated output given x_j, that is

$$e = \sum_{i=1}^{M} \left\| y_i - \sum_{j=1}^{N} f_{ij}(x_j) \right\|^2 \tag{5}$$

they are in the danger of converging to a local minimum of the merit function rather than to a solution of the nonlinear system due to nonsingularity of the *Jacobian* matrix *J(x)* [].

It has been well proven that a homotopy method is a reliable and efficient approach to solve difficult nonlinear polynomial systems [12]. The greatest advantage of this method is that it finds all the roots of the system in a globally convergent way instead of seeking for just one optimal point in a multidimensional space locally. The homotopy method operates in two stages. Firstly, a homotopy map is constructed with a starting system $\boldsymbol{G(X)}$ which has exactly as many regular solutions as the root count in the system $\boldsymbol{F(X)}$ that needs to be solved. This homotopy map can be expressed as

$$\boldsymbol{H}(\boldsymbol{X}, t) = t\boldsymbol{F}(\boldsymbol{X}) + \gamma(1-t)\boldsymbol{G}(\boldsymbol{X}) \tag{6}$$

where $t \in [0,1]$ and the solutions to $\boldsymbol{G}(\boldsymbol{X})$ is trivial to find. Secondly, when *t* moves from 0 to 1, (6) forms a continuous path originating the set of known roots of $\boldsymbol{G}(\boldsymbol{X})$ towards the roots of the target system $\boldsymbol{F}(\boldsymbol{X})$ since that $\boldsymbol{H}(x, 0) = \boldsymbol{G}(\boldsymbol{X})$ and $\boldsymbol{H}(x, 1) = \boldsymbol{F}(\boldsymbol{X})$.

A predictor-corrector path following technique is then used to track the solution paths. The path regards $\boldsymbol{X}$ as a function of *t* as shown in Fig. 2. The path can be divided into a set of discrete points as *t* goes from 0 to 1 at the increments of Δt at each step:

$$t_l = l\Delta t, \quad 0 \le l \le L \tag{7}$$

where $t_L = L\Delta t = 1$. Suppose at the *l*-th step, the solution $\boldsymbol{X_l}$ is already known for t_L, that is

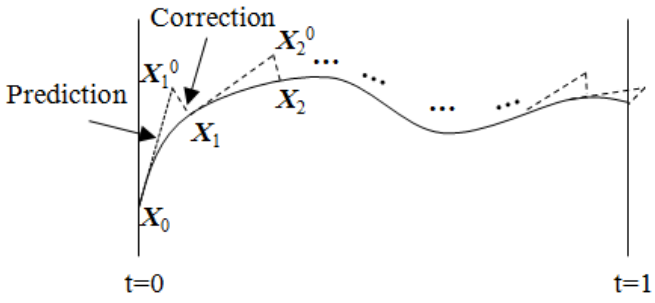

Fig. 2. Illustration of prediction and correction steps in the homotopy path.

$$\boldsymbol{H}(\boldsymbol{X}_l, t_l) = 0 \tag{8}$$

To find the solution $\boldsymbol{X}_{l+1}$ associated with the next step t_{l+1}, a first initial estimation ${\boldsymbol{X}^0}_{l+1}$ is predicted by calculating the tangent direction with respect to t, namely $\frac{d\boldsymbol{X}}{dt}$. Moving $\boldsymbol{X}_l$ in the direction of the slope by the distance of $\frac{d\boldsymbol{X}}{dt}\Delta t$, ${\boldsymbol{X}^0}_{l+1}$ is obtained. As long as Δt is small enough, ${\boldsymbol{X}^0}_{l+1}$ should be in a close neighborhood of path curve. Thus, Newton's method can be applied in the correction phase to locate $\boldsymbol{X}_{l+1}$ in the path.

A more specific and theoretical implementation of gas sensor array algorithm based on homotopy in this paper is summarized below.

1. Construct the starting system $\boldsymbol{G}(\boldsymbol{X})$ such that

$$\boldsymbol{G}(\boldsymbol{X}) = \begin{bmatrix} b_1 {x^K}_1 + c_1 \\ b_2 {x^K}_2 + c_2 \\ \vdots \\ b_1 {x^K}_N + c_N \end{bmatrix} \tag{9}$$

where b and c are random number in complex domain. Notice that the total number of roots of $\boldsymbol{G}(\boldsymbol{X})$ must be exactly the same as the roots of $\boldsymbol{F}(\boldsymbol{X})$.

2. Choose a root from $\boldsymbol{G}(\boldsymbol{X})$ as an initial point $\boldsymbol{X}_0$ from which the homotopy path starts.

3. Calculate the direction slope in prediction stage. Take derivative from (8) with respect to t, such that

$$\frac{\partial}{\partial t}\boldsymbol{H}(\boldsymbol{X}_l, t_l) + \frac{\partial}{\partial \boldsymbol{X}}\boldsymbol{H}(\boldsymbol{X}_l, t_l)\frac{\partial \boldsymbol{X}}{\partial t} = 0 \tag{10}$$

According to (10), the predicted point ${\boldsymbol{X}^0}_{\boldsymbol{l+1}}$ is obtained:

$${\boldsymbol{X}^0}_{l+1} = \boldsymbol{X}_l - \left[\frac{\partial}{\partial \boldsymbol{X}}\boldsymbol{H}(\boldsymbol{X}_l, t_l)\right]^{-1}\frac{\partial}{\partial t}\boldsymbol{H}(\boldsymbol{X}_l, t_l)\Delta t \tag{11}$$

4. Using Newton method and keeping t fixed at t_{l+1}, the actual solution $\boldsymbol{X}_l$ can be realized from ${\boldsymbol{X}^0}_{l+1}$ step by step through iteration in the correction stage. The Newton iteration is given as:

$${\boldsymbol{X}^{m+1}}_{l+1} = {\boldsymbol{X}^m}_{l+1} - \left[\frac{\partial}{\partial \boldsymbol{X}}\boldsymbol{H}({\boldsymbol{X}^{\boldsymbol{m}}}_{l+1}, t_{l+1})\right]^{-1}\boldsymbol{H}({\boldsymbol{X}^m}_{l+1}, t_{l+1}) \tag{12}$$

The iteration stops when the norm of $\boldsymbol{H}({\boldsymbol{X}^m}_{l+1}, t_{l+1})$ is less than a value.

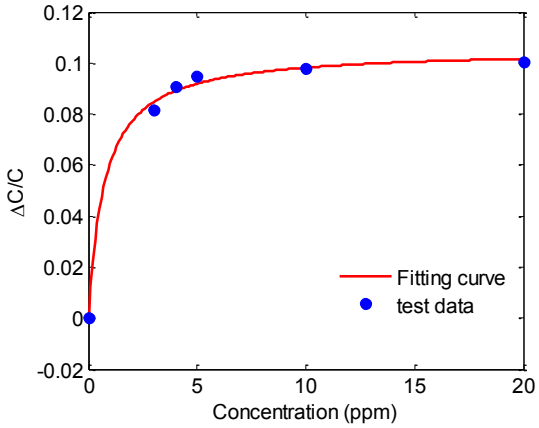

Fig. 3. Experimental SO_2 sensor data and fitting curve of normalized sensor response versus concentration demonstrating that the fitting curve matches the actual data very well.

5. Start from step 3 until t goes to 1 and a root of $\boldsymbol{F}(\boldsymbol{X})$ is found.

6. Start from step 2 choosing another initial point and look for a new solution of $\boldsymbol{F}(\boldsymbol{X})$.

7. After all the solutions of $\boldsymbol{F}(\boldsymbol{X})$ are gained, the one in the concentration range is the final estimation for target gases. If there is a value close to zero in the solution, it means that the target gas does not exist.

The other consideration of the algorithm is the step size of parameter t. If t is too small, it will increase the number of iteration and waste time. If t is moderate, then it is possible that in a certain step, ${\boldsymbol{X}^0}_{l+1}$ is not close enough to the current path. Instead of moving ${\boldsymbol{X}^0}_{l+1}$ back to $\boldsymbol{X}_{l+1}$ in the current path, Newton's method will move it to another ${\boldsymbol{X}'}_{l+1}$ which is the solution in another path. In this case, a root of the system is missing. It is possible that the missing root is the real solution for gas concentrations. To avoid this issue, the parameter t can be made adaptive by tracking the speed of convergence in step 4. If the speed is fast, that means ${\boldsymbol{X}^0}_{l+1}$ is very close to the current path and t can be increased. Otherwise, low speed of convergence reveals that t has to be divided into finer step size in current stage.

IV. Results And Discussion

Sensor responses to CH_4 and SO_2 have been tested and collected in this paper. Other target gases such as NO_2, CO, acetaldehyde and so on are still being tested for the purpose of minimizing the cross sensitivity among those targets. To verify our algorithm, only binary mixture gases will be applied to the proposed method. Based on the collected data, the sensor characteristic to each individual gas can be modeled as

$$f(x) = \frac{p_1 x^2 + p_2 x + p_3}{x + q_1} \tag{13}$$

where p, q are constant coefficients. Fig. 3 shows the fitting curving result for one of the sensor responses. As shown, this model gives smooth representation of the sensor response to various concentrations.

The sensor response to mixtures is assumed to follow the additive law, where the output is the weighted sum of responses of each individual gas in the mixture [8-9]. To estimate the accuracy, the concentration of one analyte is fixed while the other one varies from zero to the maximum value of interest. A measure of accuracy performance is defined by

$$Accuracy = \frac{\|True - Estimation\|}{True} \quad (13)$$

Fig.4 shows the error bar of accuracy at each concentration for CH_4 and SO_2. The algorithm can predict the accuracy within less than 1% variation in the concentration range of interest for both gases. The estimation error of SO_2 is slightly higher than CH_4. This is because both sensors are less sensitive to SO_2 than CH_4, as shown in Fig. 1.

V. Conclusion

In this paper, a gas sensor array signal processing method based on a homotopy algorithm was presented and shows a promising way to determine the individual chemical gas concentrations in a mixture. This method does not rely on the initial estimation to start and converges to the final solution globally. This method is able to both classify the number of gases and quantify their concentrations. For both CH_4 and SO_2 model sensors the prediction results are within 1% variation of true values. These results provide confidence that the method can be expanded to mixtures with more gases.

Acknowledgment

The authors would like to thank Dr. Xianqun Zeng at Oakland University for providing the gas sensor data and Tianran Chen from the Department of Mathematics at Michigan State University for assisting with the algorithm.

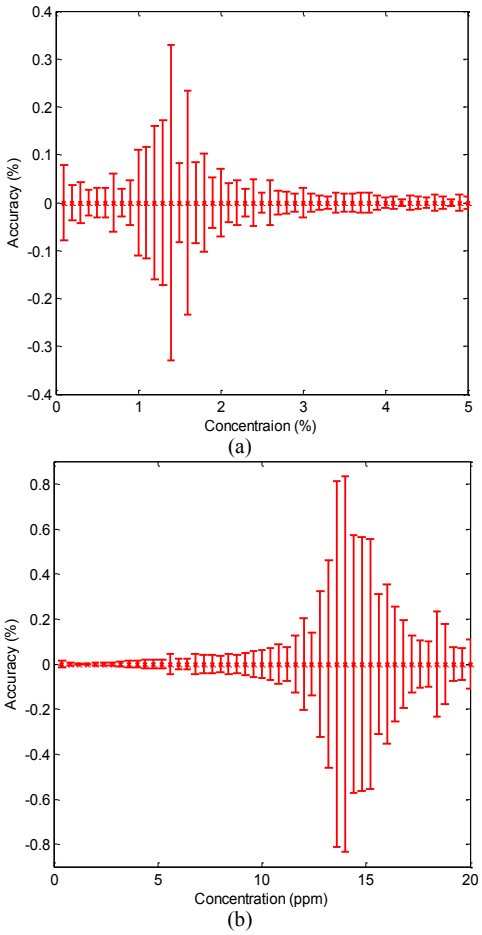

Fig. 4. Estimation accuracy at different concentrations for (a) CH_4 and (b) SO_2.

References

[1] Brook, R. D., "Cardiovascular effects of air pollution." *Clin Sci*, 115, (6), 175-87, 2008.

[2] R. Gutierrez-Osuna, "Pattern analysis for machine olfaction:Areview," *IEEE Sensors J.*, vol. 2, no. 3, pp. 189–202, Jun. 2002.

[3] Kea-Tiong Tang, Shih-Wen Chiu, Meng-Fan Chang, Chih-Cheng Hsieh, and Jyuo-Min Shyu, "A Low-Power Electronic Nose Signal-Processing Chip for a Portable Artificial Olfaction System," *IEEE Trans. Biomed. Circuits and Systems*, vol. 5, no. 4, Aug. 2011.

[4] M. Shi, A. Bermak, S. Chandrasekaran and A. Amira, "A Committee Machine gas identification System Based on Dynamically Reconfigurable FPGA," *IEEE Sensors Journal*, vol. 8, no. 4, pp. 403-414, 2008.

[5] S. Wold, H. Martens, H. Wold, "The multivariate calibration method in chemistry solved by the PLS method," *Proceedings of the Conference on Matrix Pencils*, pp. 286–293, 1983.

[6] H. Zhou, M. L. Homer, A. V. Shevade, and M. A. Ryan, "Nonlinear least-squares based method for identifying and quantifying single and mixed contaminants in air with an electronic nose," *Sensors*, vol. 6, pp. 1–18, 2006.

[7] H. Sundgren, F. Winquist, I. Lukkari, and I. Lundstrom, "Artificial neural networks and gas sensor arrays: Quantification of individual components in a gas mixture," *Meas. Sci. Technol.*, vol. 2, no. 5, pp. 464–469, 1991.

[8] Nayak, M. S.; Dwivedi, R.; Srivastava, S. K, "Application of iteration technique in association with multiple regression method for identification of mixtures of gases using an integrated gas sensor array," *Sensors and Actuators B*: *Chemical* 1994.

[9] H. Dongliang, H. Leung, "Simultaneous Classification and Concentration Estimation for Electronic Nose," *IEEE Sensors Journa*l, 7(5): pp. 825-834, 2007.

[10] Wang, Z., Lin, P., Baker, G. A., Stetter, J., Zeng, X, "Ionic Liquids as Electrolytes for the Development of a Robust Amperometric Oxygen Sensor," *Anal. Chem,* 83, 7066−707, 2011.

[11] Jorge Nocedal, Stephen J. Wright, "Numerical Optimization," Springer. Series in Operations Research, 1999.

[12] Li TY, Sauer T, Yorke JA, "The cheaters homotopy: an efficient procedure for solving systems of polynomial equations," SIAM J Numer Anal 26:1241‑1251, 1989.