ACTIVE ANALYSIS OF CHEMICAL MIXTURES
WITH MULTI-MODAL SPARSE NON-NEGATIVE LEAST SQUARES

Jin Huang and Ricardo Gutierrez-Osuna

Department of Computer Science and Engineering, Texas A&M University
{tonmey,rgutier}@cse.tamu.edu

ABSTRACT

New sensor technologies such as Fabry-Pérot interferometers (FPI) offer low-cost and portable alternatives to traditional infrared absorption spectroscopy for chemical analysis. However, with FPIs the absorption spectrum has to be measured one wavelength at a time. In this work, we propose an active-sensing framework to select a subset of wavelengths that best separates the specific components of a chemical mixture. Compared to passive feature-selection approaches, in which the subset is selected offline, active sensing selects the next feature on-the-fly based on previous measurements so as to reduce uncertainty. We propose a novel multi-modal non-negative least squares method (MM-NNLS) to solve the underlying linear system, which has multiple near-optimal solutions. We tested the framework on mixture problems of up to 10 components from a library of 100 chemicals. MM-NNLS can solve complex mixtures using only a small number of measurements, and outperforms passive approaches in terms of sensing efficiency and stability.

Index Terms—Active sensing, tunable sensors, multi-modal optimization, chemical mixture analysis.

1 INTRODUCTION

Infrared (IR) spectrometry and wavelength-selection algorithms have been widely used for chemical identification and quantitative analysis [1]. New technologies such as Fabry–Pérot interferometry (FPI) offer low-cost and more portable options [2]. However, unlike traditional techniques such as Fourier Transform Infrared Spectroscopy (FTIR), which can measure the complete absorption spectrum with thousands of spectral lines within a second , FPI devices can only measure one spectral line at a time. Thus, obtaining the whole spectrum with FPIs is time consuming because it requires scanning all wavelengths. Doing so is also unnecessary: since every chemical has its own spectrum signature, only a subset of the spectral lines is needed to identify each particular sample. Thus, by extracting only the signature components of the spectrum, we can significantly shorten the sensing process and reduce computational complexity without sacrificing much accuracy.

Conventional approaches such as feature-selection [3, 4] can be used to identify a handful of relevant wavelengths. These approaches can be considered "passive" since the selected features are obtained offline to maximize discrimination among all chemicals in the library. As such, passive wavelength selection approaches have two major drawbacks. First, finding the optimal feature subset is computationally intense and, in the case of mixture analysis, infeasible since the number of possible mixtures grows factorially with the size of the chemical library. Second, finding a feature subset that discriminates among all chemicals is inefficient since some of the chemicals in the library can be easily excluded with a handful of observations.

Given these drawbacks, this paper presents an "active" approach to wavelength selection in chemical mixture analysis. In active sensing, features are selected sequentially based on information obtained from all previous measurement. Each new measurement either solves the problem or helps rule out unfeasible solutions and thus reduce the searching space for the next sensing step. For tunable sensors such as FPI devices, which acquire information sequentially, this active-sensing approach is a more efficient solution, computationally and time-wise.

2 RELATED WORK

Active perception has had a long history in computer vision [5] and robotics [6, 7]. However, only a handful of groups have applied these ideas in the domain of chemical sensing and machine olfaction. In one of the earliest studies, Nakamoto et al. [8] developed an algorithm for active odor blending, where the goal was to reproduce an odor blend by creating a mixture from its individual components. The algorithm adjusted the mixture ratio so that the response of a gas-sensor array to the mixture matched the response to the odor blend. Priebe et al. [9] developed a statistical pattern-recognition method for active sensing. The approach builds a decision tree to partition feature space in a hierarchical fashion: nodes close to the root select features that can cluster examples regardless of class labels, whereas nodes at the leaves select features that discriminate examples from different classes. The authors showed that the approach can reduce misclassification rates by half, while requiring only 20% of the features to make any individual classification. Dinakarababu et al. [10] proposed an adaptive device that uses a digital micro-mirror as a tunable spectral filter to multiplex certain spectral bands. The system thus measures the projection of the spectral density onto a set of basis vectors, rather than measuring each spectral line directly. The basis vector changes over time based on previous information to maximize discrimination of different chemicals.

Contribution and relation to prior work: Over the last few years, we have applied active sensing principles to various chemical identification and estimation problems. In an initial study [11], we developed an active classification method to identify individual chemicals at fixed concentration using metal-oxide sensors. Later, we reformulated the active-sensing approach to be able to identify single chemicals as well as their concentrations using an FPI device [12]. In this paper, we address a more challenging problem of identifying multiple components in a chemical mixture and their respective concentrations. To solve this problem, we take advantage of Beer's law , which states that IR absorption is linear in concentration and mixture components. If the complete IR spectrum was available, the linear problem could be solved using non-negative least squares. However, with active sensing the problem becomes significantly more challenging. First, with active sensing the number of wavelengths measured is

generally much smaller than the number of chemicals in the library; as a result, the linear system is under-determined. Second, the number of chemical mixture combinations grows factorially with the number of chemicals in the library, so it is computationally impractical to evaluate all the solutions. To overcome these difficulties, this paper proposes a multi-modal[1] non-negative least squares (MM-NNLS) method with sparsity regularization constraints. Instead of trying to test all the possible solutions, the MM-NNLS generates multiple sparse solutions. Providing alternative solutions not only offers users flexibility as a solver, but also guides the feature-selection process towards wavelengths that can further tell apart these multiple alternative solutions.

3   METHODS

3.1 Problem statement

Consider the problem of quantifying a gas sample containing a mixture of chemicals from a library $\Gamma = \{C_1, C_2, \ldots, C_{N_c}\}$ with N_c distinct chemicals. Each chemical C_i in the library is characterized by a spectrum $b_i = [b_i^1, b_i^2, \ldots, b_i^l]^T$ representing the IR absorption at l distinct wavelengths $\Lambda = \{\lambda_1, \lambda_2, \ldots, \lambda_l\}$. Thus, we can represent this library as a $l \times N_c$ column matrix A, where each column represents the absorption spectrum s_i of the corresponding chemical C_i: $A = [s_{C_1}, s_{C_2}, \ldots s_{C_{N_c}}]$. According to Beer's law, the measured IR absorbance is linear to its concentration. Therefore, the absorption spectrum of the chemical mixture can be represented as an $l \times 1$ column vector $b = Ax$ where x is an $N_c \times 1$ column vector with each element being the concentration of each chemical in the library. Because the measurements represent absorption, all the elements in the matrix A, observation b and concentration x should be non-negative. Denoting by ϵ the sensor noise, the problem is to solve the linear system:

$$Ax = b + \epsilon, x > 0 \tag{1}$$

In active sensing, though, we can only measure one wavelength at a time. As a result, the entire matrix A is not available but only the rows for those wavelengths that have been measured. Denote by $F = \{\lambda_{F_1}, \lambda_{F_2}, \ldots, \lambda_{F_n}\}$ the set of measured wavelengths, where λ_{F_i} is the λ_{F_i}-th wavelength in the spectrum at sensing step i. Then, using Matlab notation[2] the actual matrix A becomes $A' = A(F, :)$, and the actual sensor observation becomes $b' = b(F)$. Starting from an empty matrix (zero observations), A'grows by adding one row at a time, as each new wavelength is measured. The goal is to find a sequence of features λ_F, and use the corresponding observation $b(F)$ and the spectral library $A(F, :)$ to estimate the mixture components and their concentration x.

3.2 Proposed solution

Because the system is underdetermined (there are far more chemicals in the library than measured wavelengths), we prevent overfitting by imposing a sparsity constraint on the solution x. Namely, we seek to minimize the number of non-zero elements in x while satisfying the equality constraint $A'x = b'$

---

[1] In this work, the term multi-modal refers to the fact that the optimization can generate multiple solutions (multiple modes); refer to [13] for a review of multi-modal optimization techniques.

[2] In Matlab notation, the expression $A(:, i)$ denotes the i-th column of A, whereas $A(j, :)$ denotes the j-th column of A.

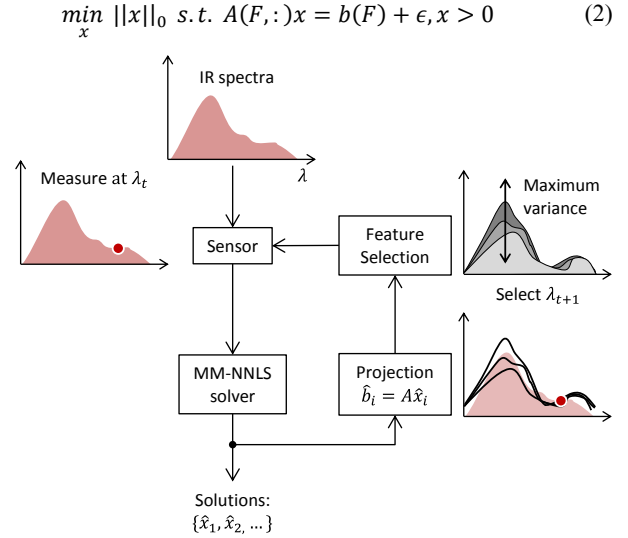$$\min_x \ ||x||_0 \ s.t. \ A(F, :)x = b(F) + \epsilon, x > 0 \tag{2}$$

Fig. 1. Building blocks of the active sensing framework

To solve this minimization problem, our method operates in a two-phase cycle. In the first phase, we make a measurement at a particular wavelength (step 1) and combine it with all previous measurements to solve the underlying linear system in eq. (2) (step 2). Because this linear system generally contains multiple near-optimal solutions, we use a solver that can generate multiple possible solutions $\{\hat{x}_1, \hat{x}_2, \ldots\}$. In the second phase, we project each of these solutions back to observation space (step 3), and then identify the wavelength that can best distinguish among these projected observations (step 4). As shown in Fig. **1**, the whole process can be illustrated as a closed-loop control system.

Step 1: Sensing. In the first step, the algorithm takes a new measurement. In the absence of prior knowledge about the mixture, we randomly select the initial wavelength to start the sensing process. Assuming that the sensor signal is contaminated with additive Gaussian noise (of equal variance across wavelengths) we obtain observation o_i:

$$o_i = b_i + \epsilon, where \ \epsilon \sim \mathcal{N}(0, \sigma^2) \tag{3}$$

Step 2: multi-modal non-linear least squares (MM-NNLS). After making each new observation, we solve the linear system in eq. (2). Since conventional NNLS solvers generate only one solution, we propose a new multi-modal non-negative least squares (MM-NNLS) that is capable of generating multiple near-optimal solutions, each having a unique combination of non-zero elements. This allows us to explore several promising regions in parallel and avoids early convergence to a sub-optimal solution.

MM-NLS maintains a queue of the best N_Q candidate solutions found so far, sorted by their projected l^2 norm error $||A\hat{x} - b||_2$. For each of these candidates, MM-NNLS expands it to the N_{xplore} directions of highest directional gradient; this ensures the solutions remain sparse because only one chemical is added to the solution at a time. The pseudo code of the algorithm is described in **Table** 1. The algorithm consists of an outer loop and an inner loop. The outer loop runs the heuristic search, which explores the candidates incrementally. The inner loop (sub-procedure NNLS_Step) performs a local searching using a simplex[3]-style NNLS solver

---

[3] The simplex algorithm searches the solution on the vertices of a polytope that models the underlying linear problem.
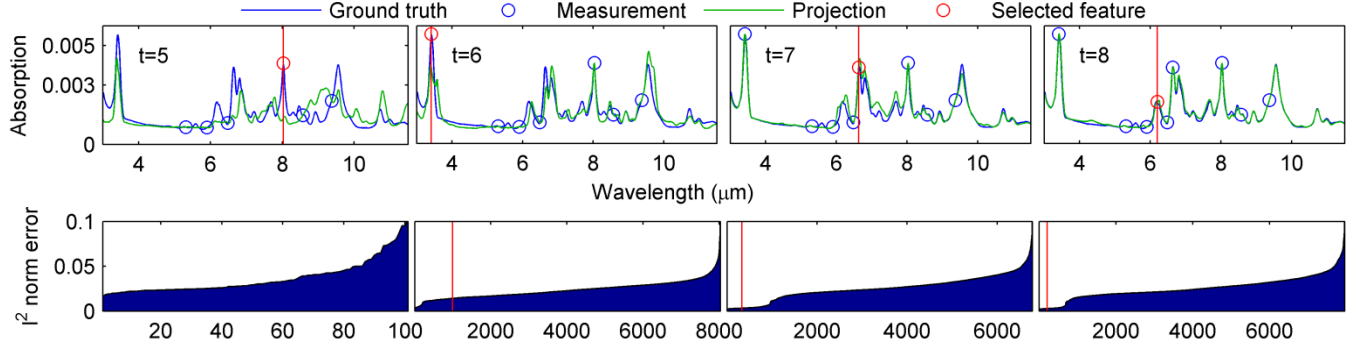
Fig. 2 Test case for the active sensing framework. **(a)** Wavelength selected at each step in the process (red bars), previous measurements (blue circle), projected spectra (green) and ground truth (blue). **(b)** Error ranking of all candidate solutions; the red vertical line marks the ranking of the correct solution.

Table 1: PSEUDO CODE FOR MM-NNLS

```
procedure Q = MM_NNLS(A, b, MaxIter)
Initialization:
- Initialize solution to x₀ = [0,...,0]
- Create initial queue of solutions Q = φ
- Add initial solution to Q = {x₀}
for i = 1: N_maxIter
- Calculate l² norm error for all candidates
- Eᵢ = {e = ||Ax − b||₂ ∀ x ∈ Q}
- Stop no improvement:
- if min(Eᵢ)/min(Eᵢ₋₁) < ϵ; return; endif
- Sort Q in ascending order (lowest error
   E first)
- Explore the first N_Q candidates in Q
   for j = 1 : N_Q
        Qⱼ = NNLS_Step(A, b, Q)
   endfor
- Add all new solutions to the queue Q
endfor

sub-procedure Q = NNLS_Step(A, b, x₀)
- Compute gradient ∇F = Aᵀ(Ax₀ − b)
- Let p₀ be the index of non-zero entries in x₀
- Sort ∇F in decreasing order ∇F = [∇f₁, ∇f₂ ... ∇f_N]
- Explore best N_xplore directions to the corners
for i = 1 : N_xplore
- Add component corresponding to ∇fᵢ, pᵢ = p₀ ∪ i
- Compute pseudo-inverse solution z = A(:, pᵢ)⁺b
if ∃ zⱼ < 0, ∀j ∈ pᵢ (solution is not feasible)
     Project the solution back to the polytope:

     xᵢ = minⱼ { x₀ʲ/(x₀ʲ − zᵢʲ) : zⱼ ≤ 0, j ∈ pᵢ } × (z − xᵢ) + x₀

else xᵢ = z; endif
endfor
- Collect all the solutions: Q = [x₁ ... x_Nxplore]
```

(see [14]; pp. 160). Unlike the conventional simplex algorithm, which uses a swapping operation to locate the extreme points, our approach uses the pseudo-inverse solution for exact line searching, and then projects the solution (if unfeasible) back to the polytope.

Step 3: Projection. The MM-NNLS algorithm generates multiple solutions $\hat{X} = \{\hat{x}_1, ..., \hat{x}_{N_Q}\}$, each solution $\hat{x}_i$ consisting of the estimated concentration for each chemical in the library . Since the ground truth X is unknown, the goodness of each solution can only be measured in observation space. For this reason, we project each solution $\hat{x}_i$ back to observation space:

$$\hat{B} = \{\hat{b}_i \,|\, \hat{b}_i = A\hat{x}_i, i = 1, ..., N_Q\} \qquad (4)$$

.**Step 4: Feature selection.** In a final step, we select the wavelength that provides maximum discrimination among these projected spectra $\hat{B}$. In prior work [12], we used Bayes Risk to measure the information content of each feature. To reduce computational complexity, in this work we use a more efficient measure: the variance of each wavelength across the N_Q projected spectra of all candidate solutions. The higher the variance, the better that wavelength can further discriminate among all candidate solutions $\hat{X}$.

$$\lambda_{\text{fsi}} = \max_{\lambda_i}\{\text{Var}(A(\lambda_i, :) \times \hat{X})\} \qquad (5)$$

4   EXPERIMENTAL RESULTS

We validated the active sensing algorithm on a dataset of IR spectra from the NIST WebBook library [15], which contains high resolution FTIR spectra in the range $3 - 21\,\mu m$. From this dataset, we randomly picked 100 spectra with resolution higher than 1,000 spectral lines over the full spectral range. To simulate the resolution of the FPI sensor, we downsampled and interpolated these spectra to 660 spectral lines over the range $3 - 11.5\,\mu m$ and normalize them so that sum all each spectrum is 1. We also added Gaussian noise on the spectra with a fixed standard deviation of 3×10^{-5} or about 2% of the average absorption across all chemicals in the spectral library. Given that we assume the mixture is sparse, we only simulate mixtures of at most 10 chemicals.

First, we illustrate the active sensing process on a problem where the target is a mixture of three chemicals. We set parameters $N_Q = 30$ (queue size) and $N_{xplore} = 100$ (fan out), so the algorithm considers a maximum of 3,000 candidate solutions at each step. The sequence of sensing actions chosen by the algorithm is shown in Fig. 3(a). Due to space limit, we only show four steps of the process (from step 5 to step 8). It actually takes 10 measurements to correctly identify the chemical. As shown in the figure, the active sensing algorithm identifies unique spectral peaks in the unknown sample. Fig. 2(b) shows the l^2 norm error of the candidate solutions at each step; the correct solution is not found until step 6 (red vertical line), it appears in the ranking, and its rank
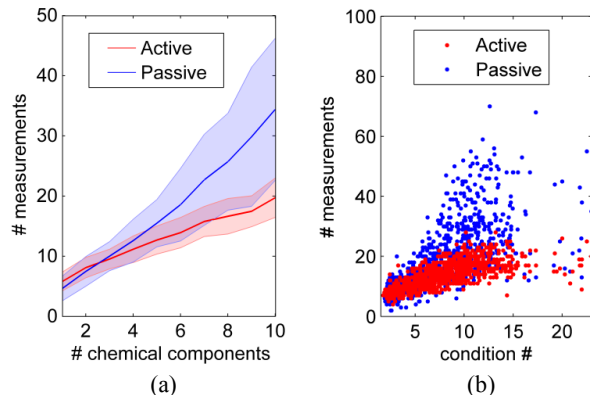
Fig. 3. (a) Number of nonzero components in the mixture vs. number of measurements acquired to identify the mixture. The blue and red centerlines are the average number of measurements for passive and active approach, respectively; the shadow bands are their respective spreads. **(b)** Condition number vs. number of observation acquired to identify the chemical.

progressively improves as more and more features are measured, until it reaches the top of the queue.

Having illustrated the active sensing process on a case study, we now put the framework under a more complete test. For comparison purposes, we use a passive approach –Sequential Forward Selection (SFS)[16] as a baseline. We train SFS using the same evaluation function: variance of the projected spectra at each wavelength. Since it is unfeasible to enumerate all possible mixtures (10^{100}), we train SFS to maximize discrimination among the 100 individual chemicals. To ensure a fair comparison, we use the same stopping criteria for both algorithms: the sensing process stop once the projected l^2 error on the whole spectrum ($||Ax - b||_2$) is smaller than a certain threshold (10^{-3}); this large threshold value was chosen to ensure both methods converge in all the trials.

We compared both methods on chemical mixture problems from 1 to 10 components. We randomly selected the components, and, for fair comparison, test both approaches on the same mixture. We run this trial 100 times on mixtures with same number of components. So there are 100x10 times trails in total. Each trial the components of the mixture are randomly selected again. Results are shown in Fig. 3(a). With the exception of small mixtures (1-2 components), where SFS outperforms active sensing by a small margin (1.5 measurements), active sensing converges to the optimal solution with fewer measurements than SFS.

The number of components is not the only determinant of the complexity of a solution: the mixture itself could be ill-conditioned, i.e. there exist multiple near-optimal solutions that can approximate the target spectrum well. Denoting by Xs the index of the components in the solution, we measure its complexity as the condition number[4] of matrix $A(:,Xs)$. Since condition number is a continuous measure, we can plot out all the solutions found by either method. Fig. 3(b) shows the condition number vs. the number of measurement required. For a given number of measurements, the condition number for the MM-NNLS solutions is much lower than that for SFS. Likewise, for a given condition number, MM-NNLS finds the optimal solution with fewer measurements than SFS. These results indicate that MM-NNLS provide solutions that are numerically more stable than those provided by SFS.

---

[4] Large condition numbers indicate unstable solutions.

5 DISCUSSION

In this paper, we have presented an active sensing framework for the problem of estimating the components of a chemical mixture using tunable IR sensors (Fabry-Pérot interferometers). This represents a challenging computational problem not only because the possible solutions are combinatorial but also because the underlying linear system is normally ill-conditioned: chemical components in a spectral library are notoriously correlated with each other. As an example, Lo et al. [17] showed that a library with 3,169 chemicals can be compressed down to 12% of its size.

To condition the problem, our approach assumes that the solutions are sparse, i.e. among all the possible solutions; we only consider those containing a few components. Sparse linear least squares methods have been extensively used in the compressive sensing community. The most common solver is the l^1 regularized (such as l^1-magic [18]) or directly l^0 regularized (such as [19]) linear least squares method based on convex optimization. These approaches, however, only generate one solution. To acquire multiple solutions, a general solution is provided by memetic algorithms; see [20] for a detailed review. A memetic algorithm randomly assigns multiple seeds and lets them converge to local optima using convex optimization. However, combining memetic algorithms with convex optimization to obtain multiple near-optimal solutions is non-trivial. First, convex-optimization techniques provide very limited control (through scaling and the smoothing parameters) of where the algorithm can converge to, so it is hard to seed the algorithm properly and efficiently. Second, convex optimization techniques are in general computationally expensive, to where seeding and running multiple searching can become prohibitively slow. The multi-modal non-negative least squares (MM-NNLS) method proposed in this paper addresses these issues. First, we replace the convex optimization stage with an NNLS solver [14] that has the efficiency of the simplex algorithm. And second, we replace the random seeding strategy of memetic algorithms with a heuristic search that increases the complexity of the solution one component at a time. Hence, MM-NNLS takes advantage of the computational efficiency of simplex-based NNLS solvers and combines it with a parallel heuristic search to generate multiple close-to-optimal solutions.

Our current implementation of the MM-NNLS solver provides no control over the diversity of the generated solutions. When solutions become very close to each other, they project to extremely similar spectra. As a result, these projections offer very little information to guide feature selection. In these cases, many of these solutions can be clustered into a few meaningful representatives. One way to address this issue is to introduce diversity metrics (e.g., distances between solutions) into the heuristic search. Encouraging diversity can also improve speed because it prevents the algorithm from having evaluate loads of similar solutions.

Finally, work in our group is underway to validate the MM-NNLS algorithm on real chemical mixture problems with FPI-based sensors. Though we have achieved promising experimental results on single-chemical problems [12], most chemical stimuli in the real world are mixtures, so the proposed framework for active multicomponent analysis would significantly expand the applications of FPI devices.

6 ACKNOWLEDGEMENTS

7   REFERENCES

[1]   W. Kaye, "Near-infrared spectroscopy," *Spectrochimica Acta,* vol. 6, pp. 257-E2, 1954.

[2]   *Variable Color Infrared Detectors for IR Sensors* Available: http://www.infratec-infrared.com/products_variable-color.htm

[3]   W. Wu, S. C. Rutan, A. Baldovin, and D. L. Massart, "Feature selection using the Kalman filter for classification of multivariate data," *Analytica Chimica Acta,* vol. 335, pp. 11-22, 1996.

[4]   W. Wu, B. Walczak, D. L. Massart, K. A. Prebble, and I. R. Last, "Spectral transformation and wavelength selection in near-infrared spectra classification," *Analytica Chimica Acta,* vol. 315, pp. 243-255, 1995.

[5]   J. Aloimonos, I. Weiss, and A. Bandyopadhyay, "Active vision," *International Journal of Computer Vision,* vol. 1, pp. 333-356, 1988.

[6]   D. Fox, W. Burgard, and S. Thrun, "Active Markov localization for mobile robots," *Robotics and Autonomous Systems,* vol. 25 pp. 195 - 207, 1998.

[7]   N. Roy, W. Burgard, and S. T. D. Fox, "Coastal navigation-mobile robot navigation with uncertainty in dynamic environments," in *IEEE International Conference on Robotics and Automation*, 1999, pp. 35-40.

[8]   T. Nakamoto, H. Matsushita, and N. Okazaki, "Improvement of optimization algorithm in active gas/odor sensing system," *Sensors and Actuators A-Physical,* vol. 50, pp. 191-196, 1995.

[9]   C. E. Priebe, D. J. Marchette, and D. M. Healy, "Integrated sensing and processing decision trees," *IEEE Transactions on Pattern Analysis and Machine Intelligence,* vol. 26, pp. 699-708, 2004.

[10]   D. V. Dinakarababu, D. R. Golish, and M. E. Gehm, "Adaptive feature specific spectroscopy for rapid chemical identification," *Optics Express,* vol. 19, pp. 4595-4610, 2011.

[11]   R. Gosangi and R. Gutierrez-Osuna, "Active Temperature Programming for Metal-Oxide Chemoresistors," *IEEE Sensors Journal,* vol. 10, pp. 1075 -1082, 2010.

[12]   J. Huang, R. Gosangi, and R. Gutierrez-Osuna, "Active Concentration-Independent Chemical Identification With a Tunable Infrared Sensor," *IEEE Sensors Journal,* vol. 12, pp. 3135-3142, 2012.

[13]   S. Das, S. Maity, B.-Y. Qu, and P. N. Suganthan, "Real-parameter evolutionary multimodal optimization — A survey of the state-of-the-art," *Swarm and Evolutionary Computation,* vol. 1, pp. 71-88, 2011.

[14]   C. L. Lawson and R. J. Hanson, "Linear least squares with linear inequality constraints," in *Solving least squares problems*, R. J. Hanson: Prentice-Hall, 1974, p. 161.

[15]   I. T. Jolliffe, *Principal Component Analysis*: Springer, 2002.

[16]   K. Fukunaga, "Feature Subset Selection," in *Introduction to Statistical Pattern Recognition, Second Edition (Computer Science & Scientific Computing)*: Academic Press, 1990, pp. 489-502.

[17]   S.-C. Lo and C. W. Brown, "Infrared Spectral Search for Mixtures in Large-Size Libraries," *Applied Spectroscopy,* vol. 45, pp. 1628-1632, 1991.

[18]   E. Candes and J. Romberg, "l1-magic: Recovery of sparse signals via convex programming," User's Guide for L1-magic tool. URL: www.acm.caltech.edu/l1magic/downloads/l1magic.pdf, 2005.

[19]   H. Mohimani, M. Babaie-Zadeh, and C. Jutten, "A Fast Approach for Overcomplete Sparse Decomposition Based on Smoothed L0 norm," *IEEE Transactions on Signal Processing,* vol. 57, pp. 289-301, 2009.

[20]   F. Neri and C. Cotta, "Memetic algorithms and memetic computing optimization: A literature review," *Swarm and Evolutionary Computation,* vol. 2, pp. 1-14, 2012.