Sensors and Actuators B 208 (2015) 245–257

Contents lists available at ScienceDirect

Sensors and Actuators B: Chemical

journal homepage: www.elsevier.com/locate/snb

# Active wavelength selection for mixture analysis with tunable infrared detectors

Jin Huang*, Ricardo Gutierrez-Osuna

*Department of Computer Science and Engineering, Texas A&M University, United States*

ARTICLE INFO

*Article history:*
Received 31 May 2014
Received in revised form
11 September 2014
Accepted 21 October 2014
Available online 6 November 2014

*Keywords:*
Active sensing
Tunable sensors
Wavelength selection
Multi-modal optimization
Model selection
Chemical mixture analysis

ABSTRACT

This article presents an active wavelength selection algorithm for multicomponent analysis with tunable infrared sensors. Traditional techniques for wavelength selection operate off-line; as a result, the resulting feature subset is fixed and only optimal for the specific mixtures and noise levels in the training set. To address this limitation, the proposed algorithm interleaves the wavelength-selection and sensing steps so that the feature subset adapts to information from previous measurements. At each point in the process, the algorithm maintains a pool of candidate solutions (i.e., mixtures) consistent with all past measurements, then selects the wavelength that maximizes discrimination across the pool. The algorithm uses a weighting function based on Akaike information criterion to promote parsimonious solutions and balance exploration vs. exploitation strategies. The algorithm is validated experimentally on binary mixture problems with a tunable infrared detector (Fabry-Perot interferometer), and its performance on higher-order mixtures characterized in simulation with a large spectral library. Active wavelength selection outperforms passive strategies, particularly at low signal-to-noise and foreground-background ratios, and when mixture components are similar, in which case the problem becomes ill conditioned.

© 2014 Elsevier B.V. All rights reserved.

## 1. Introduction

Compact tunable chemical sensors based on Fabry-Perot interferometry have become available in recent years [1,2], offering the prospect of low-cost, portable embedded spectroscopy for chemical identification and quantitative analysis. However, compared to traditional spectroscopy such as Fourier transform infrared spectroscopy (FTIR), these compact tunable sensors generally have lower sensitivity and resolution, which leads to higher sensor noise and greater collinearity. Wavelength selection can help alleviate these problems; by removing uninformative wavelengths, it has been shown both theoretically [3] and experimentally [4–8] that prediction accuracy can be improved. Additionally, since tunable sensors can only scan one spectral line at a time, wavelength selection can significantly speed up the sensing process by avoiding non-informative wavelengths.

Given the combinatorial complexity of the wavelength selection problem, an efficient searching algorithm is crucial to make the process computationally tractable. Several approaches have been proposed in the literature, including various randomized algorithms such as genetic algorithms [8], simulated annealing [7], colony optimization [9], as well as greedy strategies. A noteworthy greedy search technique is the successive projection method of Araújo et al. [5], which extracts wavelengths that minimize collinearity using the sequential orthogonal projections of the Gram–Schmidt procedure. To further reduce the search space, a common technique is to group wavelengths into individual non-overlapping windows, as in the changeable size moving window scheme of Du et al. [10].

Notwithstanding the effectiveness of these wavelength selection algorithms, their performance is limited by the fact that wavelengths are selected offline using a subset of all the possible mixtures to which the device may later be exposed. For mixture analysis problems of even moderate size (e.g., tens of potential chemicals), and in the absence of prior knowledge of the most likely components (and possibly their relative concentrations), the search will generally produce wavelength subsets that are either highly redundant or too specific to the particular mixtures in the training set.

To address the limitations of these methods, this article presents an adaptive algorithm that interleaves the wavelength-selection and sensing processes. Given a set of previous measurements (i.e., absorption at specific wavelengths), the algorithm generates a pool of candidate solutions, each solution representing a vector of concentrations across all the chemicals in the library, then selects as its

---

* Corresponding author. Tel.: +1 5108593583.
*E-mail addresses:* tonmey@cse.tamu.edu (J. Huang), rgutier@cse.tamu.edu (R. Gutierrez-Osuna).

http://dx.doi.org/10.1016/j.snb.2014.10.094
0925-4005/© 2014 Elsevier B.V. All rights reserved.

next wavelength the one that maximizes discrimination among the candidate solutions. In this fashion, the algorithm can be viewed as adaptively generating a training set of chemical mixtures for the wavelength-selection process. As the sensing process continues, the training set becomes closer to the test sample, and the selected wavelengths grow more relevant. A weighting function over candidate solutions according to their fitness (i.e., consistency with the measured wavelengths) can then be used to bias the algorithm towards exploration (e.g., during the initial stages) or exploitation (e.g., to promote convergence in the final steps).

We validated the approach through a series of experiments with real and synthetic data. First, we established proof-of-concept on real data from an FPI prototype exposed to binary mixtures (background/foreground) randomly selected from a small library of volatile organic compounds. To make the problem challenging, the foreground component was set to have a concentration 1–2 orders of magnitude below that of the background component; at these levels, repeated sampling at the informative wavelengths for the weaker foreground is necessary before it can be detected. Then, we characterized the approach through simulations of binary mixtures with different degrees of numerical ill-conditioning (i.e., increasingly similar mixture components). Finally, we extended the simulation to complex mixture problems containing up to 15 chemical components from a library of 500 analytes. In all three sets of experiments, a passive algorithm (sequential forward selection) was used as a baseline for comparison purposes. Our results indicate that the adaptive strategy outperforms the passive strategy systematically, particularly in the presence of noise or numerical ill-conditioning.

The remaining sections of this manuscript are organized as follows. Section 2 reviews related work in the areas of active sensing, adaptive sampling and sparse signal recovery. Section 3 describes the proposed active sensing framework, including a strategy to balance exploration and exploitation criteria. Section 4 presents an experimental evaluation of our framework on two-chemical mixture problems with a Fabry-Perot interferometer. Section 5 provides a thorough evaluation of the framework on high-order mixtures using synthetic spectra from a library of 500 chemicals. The article concludes with a discussion of results and directions for future work.

## 2. Related work

Inspired by the theory of 'active perception' [11,12], which states that organisms actively probe their environments to enhance their ability to extract information, the concept of active sensing was originally proposed during the 1980s in the robotics and vision community [13]. In a classic study, Aloimonos et al. [14] proposed an active vision framework to adjust camera geometric parameters (positions, rotation, etc.) to solve 3D reconstruction problems. The authors showed that, by using an active strategy, an otherwise ill-posed problem became well-posed, which dramatically improved the algorithm's problem solving efficiency. Other problems in robotics and computer vision soon adopted the idea of active sensing, including modeling of facial expressions with temporal information [15], multi-target detection and tracking [16], and robot navigation [17], localization [18] and simultaneously map building and localization [19]. These results show that active sensing works exceptionally well compared to passive methods, especially when the observations are noisy or the problem dynamics must be considered.

Interestingly, a similar concept (adaptive sampling) had already been proposed twenty years earlier in the optimization community. In early work, Kushner and Mockus proposed a stochastic method for function minimization [20,21], a method to be later known as Bayesian optimization [22]. The approach samples the objective function sparsely, then uses a Gaussian process [23] to estimate the objective function and the variance of its estimate at all other locations in sample space. These can then be used to guide the selection process, either by further sampling at the predicted highest/lowest locations to converge to a solution (exploiting) or by sampling areas of high variance to improve the estimation accuracy (exploring).

More recently, adaptive sampling techniques have been developed for sparse signal recovery, an area known as compressed sensing [24]. Compressed sensing uses $\ell_1$ norm regularization to encourage sparsity in the solution of an under-determined linear system. Compressed sensing algorithms generally used random selection of the variables to be measured, but recent work by Haupt et al. [25] has shown that the accuracy of the reconstruction can be improved by use of adaptive sampling, particularly in the case of weak signals. Their algorithm assigns each feature an importance measure that is proportional to the value of that feature and diminishes exponentially over the number of times that feature is sampled. At the beginning, all features are assigned uniform importance; as some features are sampled more frequently, their importance diminishes, allowing other features with stronger values to be sampled. The authors evaluated the method on a telescope star image recovery problem. The proposed adaptive sampling scheme outperformed a non-adaptive scheme, requiring fewer samples to achieve higher star-detection rates.

Adaptive sampling concepts have also been developed in the machine learning literature, where they are referred to as active learning. In contrast with active sensing, where the goal is to select an optimal sequence of features for each test case, the goal of active learning is to select training samples to improve the learning of decision boundaries. In a theoretical study, Castro and Nowak [26] compared the mini–max bounds[1] of adaptive and passive classification methods, and showed that adaptive methods have superior error reduction rates with reasonable complexity. These findings are particularly relevant in chemical sensing applications, where collecting or labeling new samples can be laborious. As such, active learning can alleviate this problem by minimizing the required number of samples without reducing the performance. Motivated by this issue, Lomasky et al. [27] proposed an "active class selection" method for the problem of discriminating vapors with an array of chemical sensor. Their method generated the next $n$ training instances according to the instability of the class boundaries, the latter being measured by the number of test instances whose classification results change upon inclusion of the previous set of $n$ training samples. In related work, Rodriguez et al. [28] developed an active sampling method for sensor array calibration that selected not only the classes of the vapors to be measured but also their concentrations. The authors modeled the preference for a particular concentration $c$ with the pseudo-distribution $P(c) \propto e^{-kc}$, where parameter $k$ can be used to favor low concentrations ($k > 0$) or high concentrations ($k < 0$). Given a sequence of calibration batches $B_1$, $B_2$, ..., $B_n$ and their respective $k$ parameter sequence $k_1$, $k_2$, ..., $k_n$, the algorithm selects the value $k_{n+1}$ that provides the lowest cross-validation error on batch $B_{n+1}$.

Active sensing techniques have rarely been used in the context of chemical sensing. One notable exception is the work of Dinakarababu et al. [29], who proposed an optical implementation of active-sensing principles for rapid identification of chemicals. The authors developed a digital micro-mirror device capable of multiplexing certain spectral bands and directing them onto a photo-detector. In this way, the system was able to measure the

---

[1] The mini–max bound estimates the best possible error reduction rate (mini) under a worst difficulty scenario (max), where difficulty is measured by the dimensionality of the classification problem and the noise level of the measurements.

projection of the incoming spectral density onto a set of basis vectors, rather than measure the spectral density directly. The basis vectors consisted of the eigenvectors of a covariance matrix probabilistically weighted by the likelihood of different classes based on previous measurements.

While the sparse recovery problem shares some similarities with our multicomponent chemical analysis problem, the latter has a unique characteristic. The objective of signal recovery is to minimize the error between the measurements and the recovered signal; in contrast, in chemical component analysis one seeks to minimize the error between the measurements and the solutions of the underlying linear system that give rise to the IR absorption spectra. This requires solving the linear system explicitly, and validating its solutions not only by the error with respect to the measured absorptions (error in the observation space) but also the error in the estimated concentrations (error in the solution space).

Our early investigations of active sensing focused on temperature-modulated metal oxide sensors (MOX). In [30] we studied the problem of discriminating $M$ chemicals at a fixed concentration with a single MOX sensor. For this purpose, we used a partially observable Markov decision process (POMDP) combined with a myopic policy that selected sensing actions based on the expected reduction in Bayesian risk. The method was later extended to binary mixtures [31] and arrays of MOX sensors [32]. More recently, our work has focused on mid-IR spectrometry with Fabry-Perot Interferometers which, unlike MOX sensors, have a linear response to concentration. In [33] we proposed a method for concentration-independent chemical identification. This new approach used nonnegative matrix factorization [34] to create concentration-independent absorption profiles for each chemical, and linear least squares to fit sensor observations to the response profiles. Later, we extended the framework to the problem of analyzing chemical mixtures [35], which requires an efficient algorithm to search among a factorial combination of mixtures. For this purpose, we developed a multi-modal non-negative linear least squares method and established proof-of-concept on simulated data [35]. The work proposed here significantly extends our earlier solution [35] in three main respects. First, we propose a weighting scheme based on the Akaike information criterion (AIC) to balance exploration and exploitation strategies. Second, and more importantly, we evaluate the approach experimentally on a prototype Fabry-Perot Interferometer. Finally, we perform a thorough evaluation of the approach using higher-order mixtures and increasing degrees of ill-conditioning.

## 3. Methods

### 3.1. Problem definition

Consider the problem of estimating the concentration $\boldsymbol{x}$ for the analytes in a chemical mixture given the mixture's infrared absorption spectrum $\boldsymbol{b}$. The number of constituents in the mixture $M$ and their identity are unspecified (i.e., must also be determined), but they can be assumed to belong to a library of $C$ distinct chemicals with known spectra. Denoting the spectral library by an $L \times C$ column matrix $\boldsymbol{A}$, where each column is the $L$-dimensional absorption spectrum of each chemical, the problem can be stated as solving the linear system:

$$\boldsymbol{A}\boldsymbol{x} + \delta = [\boldsymbol{A}\,\boldsymbol{1}]\begin{bmatrix}\boldsymbol{x}\\ \delta\end{bmatrix} = \boldsymbol{b} \tag{1}$$

where $\delta$ is an offset term that accounts for any bias errors. Assuming that samples only absorb energy ($\boldsymbol{A}, \boldsymbol{b} \geq 0$) and since concentrations must be non-negative ($\boldsymbol{x} \geq 0$), the problem consists of solving a partially constrained linear system:

$$\begin{aligned}&[\boldsymbol{A}\,\boldsymbol{1}]\begin{bmatrix}\boldsymbol{x}\\ \delta\end{bmatrix} = \boldsymbol{b}\\ &s.t.\ \boldsymbol{x} \geq \boldsymbol{0} \text{ and } \delta \in \mathbb{R}\\ &\text{where } \boldsymbol{A}, \boldsymbol{b} \geq 0\end{aligned} \tag{2}$$

For active sensing with a tunable sensor, the $L$-dimensional spectrum must be measured one wavelength at a time. As a result, the entire matrix $\boldsymbol{A}$ is not available but only the rows for those wavelengths that have already been measured. Let us denote by $\boldsymbol{F} = \{\lambda_{F_1}, \lambda_{F_2}, \ldots, \lambda_{F_N}\}$ the set of measured wavelengths, where $\lambda_{F_i}$ is the $F_i$th wavelength in the spectrum at sensing step $i$. Then, using Matlab notation,[2] the actual matrix $\boldsymbol{A}$ becomes $\boldsymbol{A}(\boldsymbol{F}, :)$ and the actual sensor observation becomes $\boldsymbol{b}(\boldsymbol{F})$. Starting from an empty matrix (zero observations), $\boldsymbol{A}(\boldsymbol{F}, :)$ grows by adding one row at a time, as each new wavelength is measured. The goal is to find a short ($N \ll L$) sequence of wavelengths $\boldsymbol{\lambda_F}$ such that the corresponding observation $\boldsymbol{b}(\boldsymbol{F})$ and the spectral library $\boldsymbol{A}(\boldsymbol{F}, :)$ resolve the mixture components and their concentration $\boldsymbol{x}$.

$$\begin{aligned}&[\boldsymbol{A}(\boldsymbol{F}, :)\,\boldsymbol{1}]\begin{bmatrix}\boldsymbol{x}\\ \delta\end{bmatrix} = \boldsymbol{b}(\boldsymbol{F})\\ &\quad s.t.\ \boldsymbol{x} > \boldsymbol{0} \text{ and } \delta \in \mathbb{R}\\ &\qquad \text{where } \boldsymbol{A}, \boldsymbol{b} \geq \boldsymbol{0}\end{aligned} \tag{3}$$

Since Eq. (3) is under-determined for $N < C$ (i.e., there are more chemicals in the mixture library than measured wavelengths), the problem is only solvable when the solution $\boldsymbol{x}$ is assumed to be sparse ($M \ll C$). As will be later described, in this paper we assume that the mixture contains at most 15 chemical components out of 500 chemicals in the library ($M \leq 15$, $C = 500$).

### 3.2. Overview of the approach

To solve the constrained linear system in Eq. (3) the proposed algorithm operates in two broad strokes: at each sensing step, it generates a pool of sparse solutions based on previous measurements, then selects the next wavelength to maximize separability among the solutions in the pool. The approach is illustrated in Fig. 1. With the arrival of the $i$th measurement at wavelength $\lambda_i$, the algorithm generates a number of candidate solutions $\{\boldsymbol{x}_1, \boldsymbol{x}_2 \ldots \boldsymbol{x}_{\boldsymbol{K}}\}$ through a non-negative least squares solver (see Section 3.3); the use of multiple solutions is needed given the instability of the problem. Each solution (a vector containing the concentration of the $C$ chemicals in the library) is then transformed into the estimated full spectrum $\boldsymbol{b}_i = \boldsymbol{A}\boldsymbol{x}_i$, and each spectrum is weighted according to its fitness based on the Akaike information criterion (AIC) to prevent overfitting (see Section 3.4). The wavelength $\lambda'$ with largest weighted variance $\sigma_{\boldsymbol{W}}$ across the solution pool (each solution weighted by its fitness) is then chosen as the next measurement, and the process is repeated. Though other selection criteria may be used (e.g., maximize correlation, goodness of fit, mutual information [36]), weighted variance $\sigma_{\boldsymbol{W}}$ is simple to compute and is an approximation of the misclassification risk of the candidates (see Appendix A). The individual steps are described in more detail in the following subsections.

---

[2] The expression $\boldsymbol{A}(:,i)$ denotes the $i$th column of $\boldsymbol{A}$, whereas $\boldsymbol{A}(i:,)$ denotes the $j$th row of $\boldsymbol{A}$.

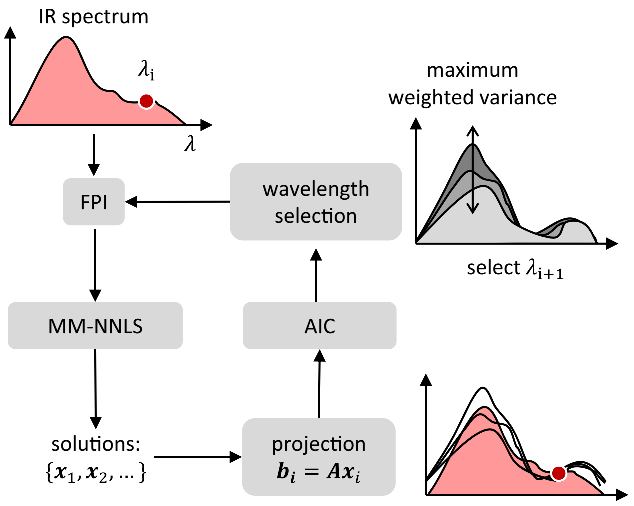


**Fig. 1.** Overview of the multimodal wavelength selection approach; FPI: Fabry-Perot Interferometer; MM-NNLS: multi-modal non-negative least squares; AIC: Akaike information criterion.

### 3.3. Generating candidate solutions: multi-modal non-negative least squares

The most critical component of the active sensing framework is the multivariate solver for the underlying linear system in Eq. (3). Our implementation uses a non-negative least squares (NNLS) algorithm [37] based on matching pursuit, a numerical technique that finds the best matching projections onto an over-complete dictionary [38]. To promote sparsity, the NNLS algorithm starts with an all-zero solution vector $\boldsymbol{x} = [0\,0\dots0]^T$, and adds one nonzero entry to $\boldsymbol{x}$ at a time, each entry representing a chemical component; namely, the algorithm computes the gradient term $V\boldsymbol{x} = \boldsymbol{A^T}(\ \boldsymbol{b} - \ \boldsymbol{Ax})f$ for each chemical in the library, and adds the one whose gradient is largest. Once a new entry has been selected, the pseudo-inverse solution[3] is used to find the minimum–minus the constraints. At this point, if any of the non-zero elements is negative, the solution $\boldsymbol{x'}$ is shifted back to the closest point in the feasible area (along any one axis) using the iteration:

$$\begin{aligned} &\text{while}\ \exists i | \boldsymbol{x}'(i) < 0 \\ &\quad \boldsymbol{x}' \leftarrow \boldsymbol{x}' - \alpha(\boldsymbol{x}' - \boldsymbol{x}) \\ &\quad \text{where}\ \alpha = \min\left\{\frac{\boldsymbol{x}'(j)}{\boldsymbol{x}'(j) - \boldsymbol{x}(j)} : j \in \boldsymbol{P}'\right\} \\ &\text{end while} \end{aligned} \tag{4}$$

where $\boldsymbol{P}$ and $\boldsymbol{P}'$ are the set of indices of nonzero entries in $\boldsymbol{x}$ and $\boldsymbol{x}'$, respectively, $\boldsymbol{x}$ is the previous feasible solution, i.e., before the non-zero entry was added, and $\boldsymbol{x}'(i)$ represents the estimated concentration for the $i$th chemical in the library. The iteration in Eq. (4) is repeated until all entries in $\boldsymbol{x}'$ are non-negative.

The process of estimating the local minimum (via the pseudo-inverse) and adjusting to the non-negative constraint[4] is repeated until no more negative entries exist. If the error term over measured wavelengths $\hat{\varepsilon} = \boldsymbol{b}(\boldsymbol{F}) - \boldsymbol{A}(\boldsymbol{F}, :)\boldsymbol{x}'_2$ is large, a new entry is added to the solution vector, and the NNLS process is repeated; otherwise, the process is terminated. A flowchart of the overall process is illustrated in Fig. 2a.

#### 3.3.1. Multi-modality: tracking multiple solutions

The NNLS algorithm generates a single solution, which is problematic for two reasons. First, the underlying linear problem is often ill-defined since only a few wavelengths $\boldsymbol{F}$ are measured; this is particularly severe at the beginning of the sensing process. Second, the process of wavelength selection requires a training set of mixtures from which to rate individual wavelengths. To address these issues, our approach wraps the NNLS algorithm around a multi-modal loop to generate multiple solutions. In contrast with heuristic multi-modal optimization techniques, which use random seeds for the search [39], our algorithm takes advantage of the gradient information ($V_{\boldsymbol{x}} = \boldsymbol{A^T}(\ \boldsymbol{b} - \ \boldsymbol{Ax})$) and the closed-form solution of ordinary linear least squares (i.e., the pseudo-inverse), which significantly improve computational efficiency. Starting with an all-zero solution vector $\boldsymbol{x}$, the algorithm selects the top $M$ entries (according to their gradient $\nabla_{\boldsymbol{x}}$), and transforms each into a feasible solution, as described earlier – see Eq. (4). The result is a set of $M$ solutions with one non-zero entry (i.e., a single chemical). The best $N$ of these $M$ solutions according to their sum squared error ($\varepsilon = \boldsymbol{b}(\boldsymbol{F}) - \boldsymbol{A}(\boldsymbol{F}, :)\boldsymbol{x}'_2$) are saved to the candidate pool, and used as seeds for the next step to yield $N \times M$ solutions with 2 non-zero entries (i.e., two chemicals). The best $N$ of those according to their error $\varepsilon$ are added to the pool, and used as seeds for the next step (solutions with 3 non-zero entries). The process continues until the candidate pool contains a solution whose error $\varepsilon$ is below a pre-specified threshold; duplicate solutions, which may occur due to the parallel searches, are removed in a final step. Parameters $M$ and $N$ are set by the user depending on the computing resources available. A flowchart of the resulting MM-NNLS algorithm is shown in Fig. 2b.

### 3.4. Wavelength selection

The MM-NNLS algorithm returns a pool of candidate solutions, where each solution represents a mixture by its concentration vector across all chemicals in the library. Thus, each concentration vector $\boldsymbol{x}_i$ can be used to reconstruct the full spectrum of the corresponding mixture**:**

$$\boldsymbol{b_i} = \boldsymbol{Ax_i} \tag{5}$$

It is this collection of spectra that can be used as a "training set" for wavelength selection. Caution must be exercised, however, because these spectra are synthesized from solutions $\boldsymbol{x}_i$ that were obtained by fitting a small number of absorption wavelengths $\boldsymbol{b}(\boldsymbol{F})$ – see Eq. (3). As a result, there is the distinct possibility that some of the solutions overfit the noisy measurements $\boldsymbol{b}(\ \boldsymbol{F})$. This is particularly problematic at low concentrations, where measurement noise can dominate the sensor response, which may lead the MM-NNLS algorithm to include additional entries (i.e., chemicals) in the solutions.

We illustrate this problem with an example. For this purpose, we randomly selected a binary mixture from a spectral library containing 500 chemicals, then randomly sampled absorption at 20 wavelengths, and added 1% white noise to each of the 20 measurements – refer to Section 5 for details on the spectral library. Then, we allowed NNLS to generate a number of solutions. Results are illustrated in Fig. 3b, with solutions ranked according to the *sampled error* ($\hat{\varepsilon} = \boldsymbol{b}(\boldsymbol{F}) - \boldsymbol{A}(\boldsymbol{F}, :)\boldsymbol{x}_2$), as well as the *true error* ($\varepsilon = \ \boldsymbol{b} - \ \boldsymbol{Ax}_2$) with respect to the noise-free ground-truth full spectrum (assumed known in this case). Fig. 3a shows the solution $\boldsymbol{x}_1$ with the lowest sampled error $\hat{\varepsilon}_1$, and the solution $\boldsymbol{x}_2$ with the lowest true error $\hat{\varepsilon}_1$.

---

[3] Using Matlab notation, the pseudo-inverse solution can be computed as $[x'^T \delta]^T = [\boldsymbol{A}\left(:, \boldsymbol{P}'\right) 1]\backslash\boldsymbol{b}$.

[4] Note that the offset δ is not included in the adjusted term $\boldsymbol{x}$ since δ has no constraints. Instead, δ is re-estimated together with the adjusted term $\boldsymbol{x}'$ (i.e., through the pseudo-inverse) after the constraint adjustment. Using the pseudo-inverse ensures that the offset is local-optimal –this is in contrast to setting δ to the minimum observed value in $\boldsymbol{x}'$, a common heuristic in NNLS.

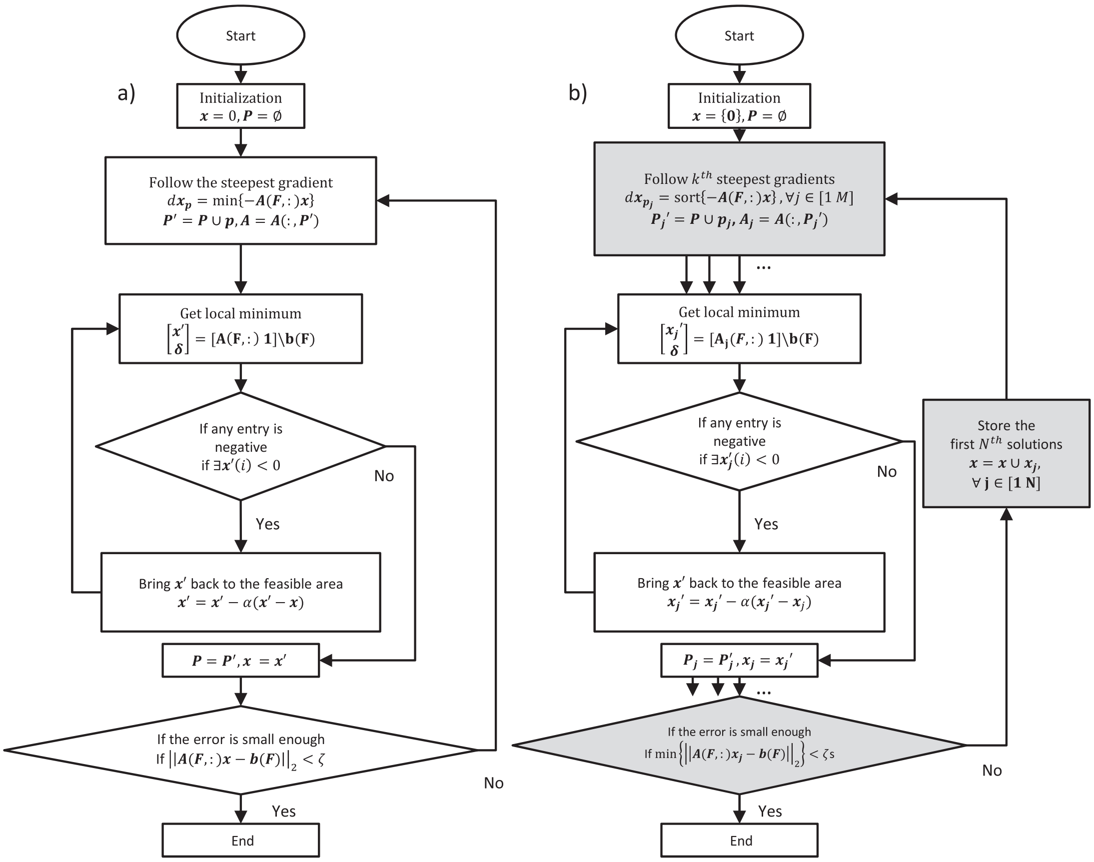


**Fig. 2.** Flowchart of the non-negative least-squares (NNLS) algorithm (a) and multi-modal NNLS algorithm (b). The shaded blocks in (b) highlight differences between both algorithms.

Even though solution $\boldsymbol{x}_1$ has lower sampled error ($\hat{\varepsilon}_1 < \hat{\varepsilon}_2$), it overfits the noisy measurements $\boldsymbol{b}(\boldsymbol{F})$ by using 31 non-zero entries, as shown in Fig. 3c; notice also how solution $\boldsymbol{x}_1$ deviates quite significantly from ground truth at λ = 8.87, 9.07 µm. In contrast, solution $\boldsymbol{x}_2$ has only 4 non-zero entries and has smaller true error ($\varepsilon_2 < \varepsilon_1$) but unfortunately ranks in position #89 according to the sampled error (the only error that can be measured in practice).

The above example illustrates how, in the presence of noise, the solution with the smallest sampled error can overfit the measurements by using extra complexity (i.e., number of chemicals in

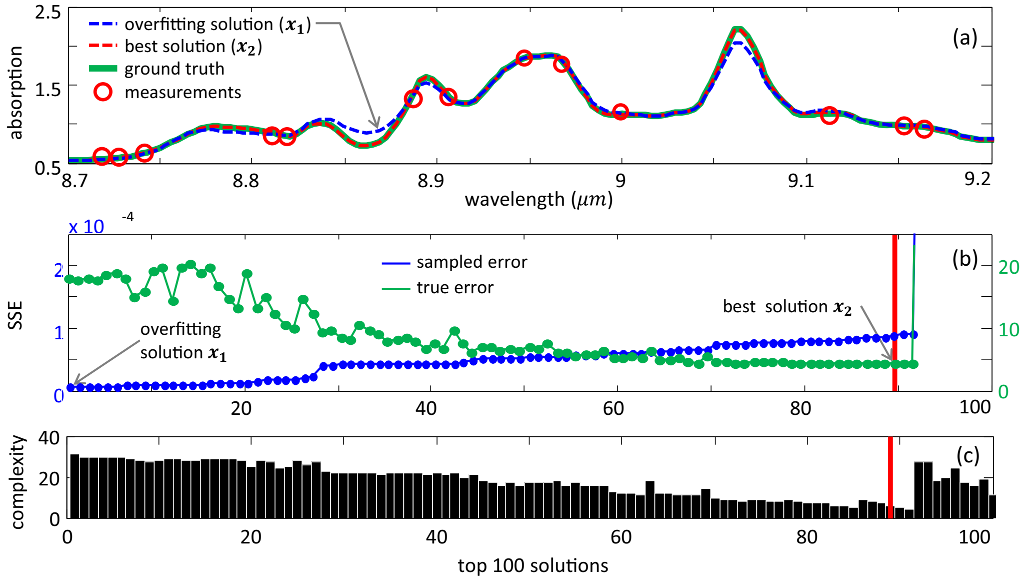


**Fig. 3.** (a) Projected spectra of the solutions ranked by the sampled error ($b(F) - A(F,:)x_2'$) and the true error over the full spectrum ($b - Ax_2'$); only the range λ = 8.7 – 9.2 µm is shown for illustration purposes. (b) The top 100 solutions according to the sampled error, and the corresponding true error. (c) Complexity of each model (i.e., number of components in the mixture) for the top 100 solutions.

the mixture). To address this issue, we rank solutions based on the Akaike information criterion (AIC), an information-theoretic measure that takes into account both error and parsimony [40]. The AIC score can be computed as:

$$AIC_i = -2\log L_i + 2V_i \tag{6}$$

where $L_i$ is the likelihood for candidate model $i$, and $V_i$ is its number of free parameters (i.e., number of components in the mixture). For linear regression, assuming the sensor noise at each measurement is uncorrelated with the sensor noise of previous measurements, the formula can be further simplified as [41]:

$$AIC_i = n \times \log\left(\frac{\hat{\varepsilon}_i}{n}\right) + 2V_i \tag{7}$$

where $\hat{\varepsilon}_i$ is the number of measured wavelengths, and $\hat{\varepsilon}_i$ is the sum-squared error for model $i$, i.e., $\hat{\varepsilon}_{\boldsymbol{i}} = \boldsymbol{b}(\boldsymbol{F_i}) - \boldsymbol{A}(\boldsymbol{F_i}, :)\boldsymbol{x}_{\boldsymbol{i}2}$. It is these AIC scores that we use to rank candidate solutions, following conversion into likelihoods [40]:

$$w_i = e^{-1/2\Delta(AIC_i)} \tag{8}$$

where $\Delta(AIC_i)$ is the difference in AIC scores between model $i$ and the best candidate: $\Delta(AIC_i) = AIC_i - \min\limits_{\forall i} AIC_i$. We then normalize the weights to ensure they add up to one:

$$w_i \leftarrow \frac{w_i}{\sum_{\forall i} w_i} \tag{9}$$

### 3.4.1. Selection criterion

A number of traditional selection criteria (i.e., maximize correlation, goodness of fit, mutual information [36]) can be used at this point to determine the next wavelength to be sampled. In prior work [35] we used variance as a measure of uncertainty, choosing as the next measurement the wavelength having the highest variance across spectra in the solution pool $\left\{\boldsymbol{b_i}|\forall i\right\}$:

$$\lambda_{opt} = \max_{\lambda_{j|\forall j \in [1N]}} \{\sigma^2\{\boldsymbol{b_i}(\lambda_j)|\forall i\}\} \tag{10}$$

where $\lambda_{opt}$ is the next wavelength to be sampled, $\boldsymbol{b_i}(\lambda_j)$ is the (estimated) absorption at wavelength $\lambda_j$ for mixture solution $i$, and $\sigma^2$ is the variance operator. As shown in Appendix A, the variance at each wavelength is an approximation of the misclassification risk. Thus, by selecting the wavelength with largest variance the algorithm can be viewed as minimizing the risk of choosing the wrong candidate.

However, this approach treats all candidate solutions equally, regardless of their fitness. To address this issue, we then weigh each candidate solution as:

$$\lambda_{opt} = \max_{\lambda_{j|\forall j \in [1N]}} \{\sigma^2_{\boldsymbol{W}}\{\boldsymbol{b_i}(\lambda_j)|\forall i\}\} \tag{11}$$

where $\sigma^2_{\boldsymbol{W}}$ is the weighted variance, which can be calculated as:

$$\sigma^2_{\boldsymbol{W}}(\boldsymbol{x}) = \frac{1}{1 - \sum_i w_i^2} \sum_{i=1}^{N} w_i (x_i - \mu_{\boldsymbol{W}})^2 \tag{12}$$

and $\mu_{\boldsymbol{W}}$ is the weighted mean, calculated as:

$$\mu_{\boldsymbol{W}} = \sum_{i=1}^{N} w_i x_i \tag{13}$$

### 3.4.2. Balancing exploitation and exploration

During the initial stages, when only a few measurements are available, the sampling process can be dominated by a few solutions. Whenever this happens, the algorithm invests a large number of measurements to investigate a narrow region of the spectrum (i.e., to discriminate among the few early solutions) rather than explore the global structure of the spectrum in search of new candidates. This often leads to premature convergence. To guard against this problem, our implementation includes a parameter $\varsigma^2$ that can be used to adjust the spread of the AIC weights:

$$w_i = e^{-\frac{\Delta(AIC_i)}{2\varsigma^2}} \tag{14}$$

Selecting $\varsigma^2$ is non-trivial, as its value reflects on the credibility of the estimated solutions. If $\varsigma^2$ is small, the solutions will be weighted aggressively, and active sensing runs the risk of premature convergence to a suboptimal solution. In contrast, if $\varsigma^2$ is large, the weights become uniform regardless of how different they were before normalization, which may lead to irrelevant features being introduced. Thus, the weighting scheme controls how many solutions to consider for the next measurement, a trade-off commonly referred to as the exploration–exploitation dilemma [42].

To balance this exploration–exploitation dilemma, we propose an entropy-guided method that adjusts the offset parameter $\varsigma^2$ such that the entropy of the weight landscape $H(\boldsymbol{W}_O)$ remains constant. Assuming $l$ candidate solutions, the highest entropy ($\log(l)$) is achieved when the weights are uniformly distributed, whereas the lowest entropy[5] (0) is obtained when only one of the $l$ solutions has a non-zero weight. Finally, we select a value between these two extremes:

$$\begin{aligned} \dot{H} &= (\sup\{H(\boldsymbol{W})\} - \inf\{H(\boldsymbol{W})\})\alpha + \inf\{H(\boldsymbol{W})\} \\ &= (\log(l) - 0)\alpha + 0 = \alpha \log(l) \end{aligned} \tag{15}$$

where the multiplier $a$ ($0 \le \alpha \le 1$) controls the balance between exploitation and exploration: i.e., $\alpha = 0$ leads to extreme exploitation whereas $\alpha = 1$ leads to extreme exploration. In the studies reported here, we use $\alpha = 0.5$ to balance exploration and exploitation. Once the desired entropy $\dot{H}$ has been fixed, the last step is to find the corresponding parameter $\varsigma^2$. Since entropy grows monotonically as $\varsigma^2$ increases, this can be easily done with a continuous linear binary search; see Algorithm 1.

```
Algorithm 1: bSearch
Input: $f(*)$, $\varsigma^2_{min}$, $\varsigma^2_{max}$, $\dot{H}$
Output: $\varsigma^2$
if $\varsigma^2_{max} \ge \varsigma^2_{min}$
    if $\dot{H} < f(\varsigma^2_{min})$
        return $\varsigma^2_{min}$
    $\varsigma^2 = \left(\dot{H} - f(\varsigma^2_{min})\right)\frac{(\varsigma^2_{max} - \varsigma^2_{min})}{f(\varsigma^2_{max}) - f(\varsigma^2_{min})} + \varsigma^2_{min}$
    if $\varsigma^2_{max} - \varsigma^2_{min} > 10^{-2}$ // Resolution of the search
        if $|f(\varsigma^2) - \dot{H}| < 10^{-2}$
            return $\varsigma^2$;
        elseif $f(\varsigma^2) < \dot{H}$
            return $bSearch(f(*), \varsigma^2, \varsigma^2_{max}, \dot{H})$
        elseif $f(\varsigma^2) > \dot{H}$
            return $bSearch(f(*), \varsigma^2_{min}, \varsigma^2, \dot{H})$
return $\emptyset$
```

## 4. Validation on experimental data

### 4.1. Experimental setup

As a first step, we validated the active wavelength selection algorithm experimentally on a tunable IR sensor prototype. A typical configuration for a tunable IR sensor consists of a Fabry-Perot interferometer (FPI) atop a broadband IR pyroelectric detector [43]. An FPI contains two partially transmitting movable mirrors facing each other at a distance $d$, forming an optical gap characterized by the

---

[5] $\sup\{H(W)\} = -\sum_{i=1}^{l} \frac{1}{l} \log\left(\frac{1}{l}\right) = \log(l) \inf\{H(W)\} = -1 \log(l) = 0.$

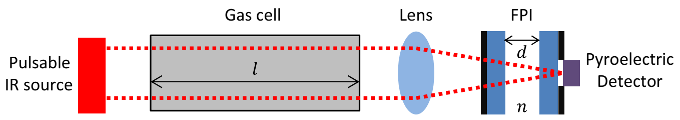


**Fig. 4.** Basic setup for an FPI: the optical path starts with the collimated IR radiation (pulsed); the IR radiation passes through the sample in the gas cell, where it is absorbed at different wavelengths depending on the sample; the focusing lens collects the transmitted radiation onto the FPI sensor, which can be tuned to sample one individual wavelength at a time.

refractive index $n$; see Fig. 4. As incoming radiation bounces back and forth between the mirrors, only a small portion is transmitted; if these reflections are out of phase, a destructive interference occurs. Thus, only radiation satisfying the resonance condition $\lambda_m = 2nd\cos\beta/m$ can be transmitted (where $m$ can be any positive integer and $\beta$ is the angle of incidence). By changing the distance $d$ between the two mirrors, one can selectively transmit certain wavelengths, and the device becomes a tunable IR sensor.

As illustrated in Fig. 4, to measure IR absorption we place the analyte sample in a gas cell, illuminate the sample with a broadband IR source, and measure the energy $I$ transmitted through the cell. The absorbed energy is then calculated with respect to the energy $I_0$ of a reference sample (air) sampled beforehand:

$$\alpha = -\ln\left(\frac{I}{I_0}\right) \tag{16}$$

Based on Beer's law, the absorption $\alpha$ is proportional to the length of the gas cell $l$ and the analyte concentration $c$:

$$\alpha = \varepsilon lc \tag{17}$$

where $\varepsilon$ is the molar extinction coefficient of the sample. With the cell of fixed length, absorption then becomes a linear function of the analyte's concentration.

For the experiments described here, we used a long-wave FPI sensor (LFP-80105, Infratec, Inc) with 107 tunings (absorption lines) in the range $8-10.5\ \mu\text{m}$, coupled with a collimated broadband IR source (INTX 20-1000-R; Intex Inc.) modulated at 10 Hz and 50% duty cycle. We mounted a 10 cm gas cell (66001-10A; Specac Inc.) with ZnSe window (602L08; Specac, Inc.) and a ZnSe focusing lens (LA7542-F, Thorlabs Inc.). The FPI, IR source and sample cell were mounted onto an opto-mechanics fixture (Thorlabs Inc.) to ensure precise alignment. The FPI device was controlled using Matlab$^{\text{TM}}$ through a USB evaluation board provided by the vendor.

The sample delivery system is illustrated in Fig. 5. Vapors from the headspace of 30 mm glass vials are delivered using negative pressure with a pump connected downstream from the sample cell. The pump is modulated at 0.125 Hz with 20% duty cycle to avoid exhausting the headspace and keep the sample concentration relatively stable. Two diluters (1010 precision gas diluter, Custom Sensor Solutions, Inc.) independently mix the foreground and background sample vapors with dry air. Since water and carbon dioxide have major peaks outside of the sensor's range, air has a negligible contribution to the spectrum.

Eight different volatile commercial chemicals that show absorption peaks in the range of our sensor ($8-10.5\ \mu\text{m}$) were used for the experiments; see Table 1. Of those, acetone was chosen as the strong background because it has the strongest absorption peak of all. The remaining seven chemicals were randomly chosen as the weak foreground. Experiments were conducted in a laboratory environment at a temperature of 22.2 °C and standard atmospheric pressure of 760 mmHg.

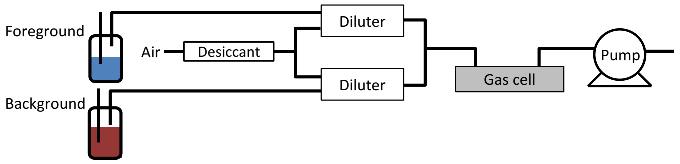


**Fig. 5.** Schematic diagram of the headspace vapor sampling system.

**Table 1**
List of chemicals used in the experiments, and their major components.

| Chemical | Components |
|---|---|
| Propanol | Propanol |
| Acetone (background) | Acetone |
| Ethyl alcohol | Ethyl alcohol |
| Isopropyl alcohol | Isopropyl alcohol |
| Tert-Butyl alcohol | Tert-butyl alcohol |
| Air | Air and sensor drift |
| Denatured alcohol | Ethyl alcohol, methanol |
| Brush cleaner | Raffinates, acetone, methanol |
| Lacquer thinner | Toluene, methanol, hexane, light aliphatic naphtha |

To illustrate the signal-to-noise ratio that is achievable with the FPI, Fig. 6 shows the full spectrum (107 spectral lines) for a sample of isopropyl alcohol at 2.5% dilution in dry air. Sensor measurements are shown as circles. Passing the full measured spectrum to the NNLS solver in Eq. (2) with $\boldsymbol{A}_{107\times 9}$ (8 chemicals plus offset) returns a solution vector $[\boldsymbol{x}^T\delta]^T$. Multiplying this solution by the library matrix yields the smooth spectrum $\boldsymbol{b} = [\boldsymbol{A}\,\boldsymbol{1}][\boldsymbol{x}^T\delta]^T$ and reconstructed baseline $\delta$ shown in the figure. Despite the high level of noise present in the measurements (notice how most measurements are negative, an indication of emitter drift), the solver is able to recover a non-negative baseline and a solution whose spectrum matches that of the ground truth. To estimate the signal-to-noise ratio, we calculate the ratio between the median predicted absorption and the root mean square (RMS) of the error:

$$\text{SNR} = \frac{\text{median}\{\boldsymbol{b}\} - \delta}{\sqrt{\frac{1}{N}\sum_{N}^{j=1}\left(\boldsymbol{b}(j) - \boldsymbol{b}'(j)\right)^2}} \tag{18}$$

where $\boldsymbol{b}$ is the reconstructed spectrum, $\boldsymbol{b}'$ is the actual measurement and $\delta$ is the offset; notice that the absorption is adjusted by the offset $\delta$ to represent the actual absorption. The resulting SNR of 17.4% illustrates the degree of noise under which the active wavelength selection algorithm has to operate.

### 4.2. Experiment 1: test case

In a first experiment, we illustrate the performance of the active wavelength selection algorithm on a two-chemical mixture problem containing acetone at 2.5% dilution as background, and isopropyl alcohol at 5% as foreground. Fig. 7a shows the full spectra of the background, foreground and the final mixture; circles represent the actual measurements that took place during the sensing process. The background chemical shows a major peak at 8.3 μm, while the foreground has a minor peak at 8.8 μm. Fig. 7d shows the rank of the correct solution as iterations progress; the correct solution is added to the pool at the 5th measurement, and is confirmed (ranked #1) at the 20th measurement.

Fig. 7c shows the total number of solutions considered by the algorithm as the iterations progress, whereas Fig. 7b shows the distribution[6] of selected wavelengths prior to and after the 20th measurement. We observe a typical two-stage pattern emerging from the active sensing process: at first, the algorithm performs a broad sampling of absorption peaks for both chemicals (the exploration stage), then performs a focused search on spectral details and smaller peaks to confirm the identity of the weaker chemical

---

[6] This distribution was obtained by applying a Gaussian kernel with 0.1 μm standard deviation to smooth the discrete distribution of measurements.

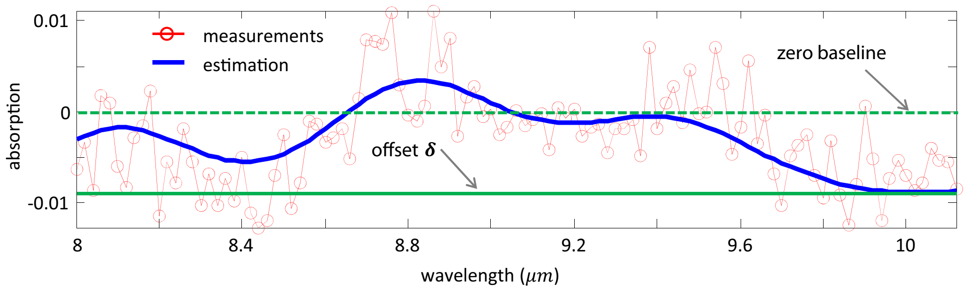


**Fig. 6.** Spectrum for a sample of 2.5% isopropyl alcohol: actual measurements (red circles), estimated solution (blue), zero baseline (green dash line) and estimated offset (green solid line).

(the exploitation stage). The most selected wavelength is around 8.8 μm, which is consistent with an absorption peak for the weak foreground chemical. It is important to note that this shift from exploration to exploitation is not programmed but rather an emerging behavior of the algorithm, driven by the lower SNR from the weak foreground contributing to most of the uncertainty, which the algorithm seeks to minimize.

### 4.3. Experiment 2: active vs. passive

In a second experiment, we compared the active wavelength selection algorithm against a “passive” baseline algorithm based on sequential forward selection [44]. The passive algorithm selects a *fixed* subset of wavelengths that best represents the average absorption spectrum across all chemicals in the library: $\bar{\boldsymbol{b}} = \frac{1}{\boldsymbol{C}}\sum_{\boldsymbol{i}=1}^{\boldsymbol{C}} \boldsymbol{A}(:, \boldsymbol{i})$. The passive algorithm works as follows:

- The first wavelength $\lambda_1$ is selected (deterministically) as the one with the highest variance in absorption across all chemicals in the library: $\lambda_1 = \max_{\lambda_{j|\forall j \in [1N]}} \{\sigma^2\{\boldsymbol{A}(:, \boldsymbol{i})|\forall i\}\}$.
- To select the second wavelength $\lambda_2$, the passive algorithm estimates the concentration of each individual spectrum in the library to fit that first measurement $\bar{\boldsymbol{b}}(\lambda_1)$; doing so exposes variance at other wavelengths but the first one–which can be fitted with zero error. The second wavelength is then selected as the one with the highest weighted variance across all newly fitted one-chemical spectra.
- To select the third wavelength $\lambda_3$, the passive algorithm randomly generates 10,000 two-chemical mixtures to fit the two measurements $\bar{\boldsymbol{b}}([\lambda_1\lambda_2])^{\boldsymbol{T}}$, and selects the wavelength with the highest weighted variance across the 10,000 fitted mixture spectra.
- The process is repeated until the desired number of wavelengths has been selected: to select the $(n+1)$th wavelength, the passive algorithm randomly generates 10,000 $n$-chemical mixtures to form a full-rank linear system to fit all previous measurements. This ensures that neighboring wavelengths, which are generally correlated to those already selected, are not selected before the whole range of the spectrum has been sampled at least once. This idea of decorrelating observations is common in passive wavelength selection methods such as successive orthogonal projection [5].

In contrast with the passive algorithm outlined above, our active algorithm requires no training. To ensure a fair comparison, both methods used the same evaluation function and solver, and were stopped after 20 sensing steps. To measure performance, we computed the rank of the correct solution among those returned by the MM-NNLS solver, averaged over all tests cases. The lower the average rank, the better the method, i.e., a rank of one indicates that the solver has placed the correct solution as the first one in its pool. As a supplementary measure we also used the classification rate, measured as the percentage of trials where the correct solution was ranked as #1.

To perform the comparison, we randomly selected 5 chemicals as foregrounds and diluted them at multiple levels while keeping the background acetone at 100% concentration. Each sample was tested 5 times at dilutions ratios of 1/50, 1/33, 1/20, 1/10 and 1/5, for a total of 125 tests samples (5 chemicals × 5 dilutions × 5 replicates). Fig. 8 shows the average classification rate (1 if the correct solution is ranked as #1; 0 otherwise) and the average ranking of the correct solution achieved by both methods. There is no significant difference at dilution ratios above 1/20; at such concentrations the problem becomes trivial, and both approaches can find the correct solution with only 4–5 measures. At low concentrations, however, active sensing outperforms its passive counterpart

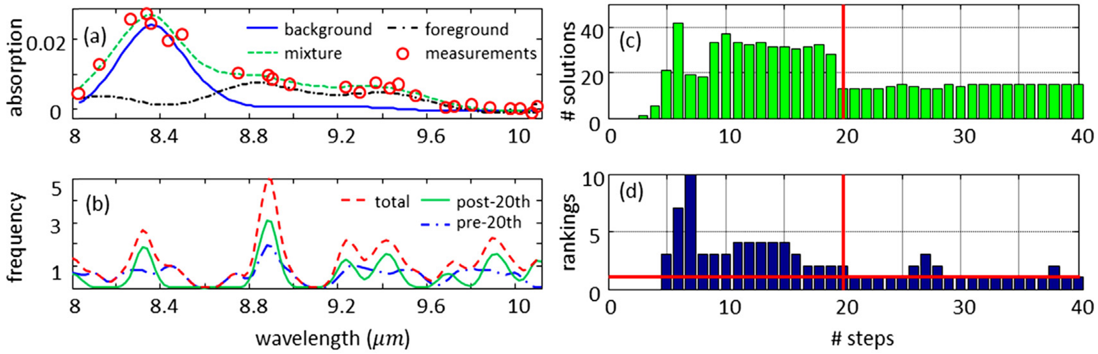


**Fig. 7.** Test case with acetone as background and isopropyl alcohol as foreground. (a) Background, foreground and the mixture; (b) Sampling frequency distribution before and after confirming the ground truth (20-th iteration); (c) Total number of solutions generated, 20-th iteration (vertical red line); (d) Ranking of the correct solution, ranking #1 (horizontal red line).

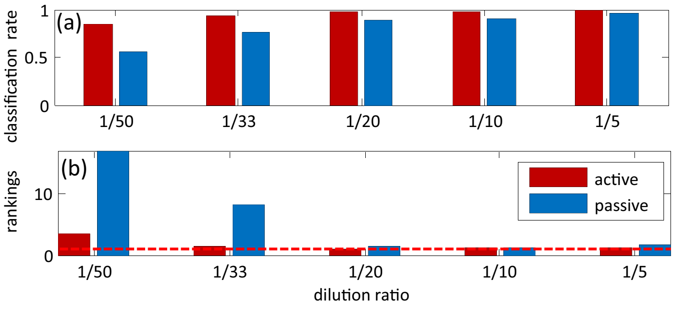


**Fig. 8.** (a) Average classification rate for the active and passive wavelength selection algorithms as a function of the foreground dilution ratio. (b) Average ranking of the correct solution as a function of the foreground dilution ratio; the dashed line represents a ranking of one, indicating that the correct solution was found.

in terms of classification rate, and more significantly when considering the average ranking of the correct solution. At the lowest dilution ratio (1/50), active sensing ranks the correct solution as #3 on average, whereas the passive algorithm ranks it at #16. This is largely because active sensing samples the most informative wavelengths repeatedly, avoiding the introduction of new irrelevant wavelengths. This results in a much more compact feature set and, as a consequence, fewer distortions due to noise are introduced to the solver.

### *4.4. Experiment 3: analyzing the exploration-exploitation tradeoff*

In a third experiment, we evaluated the effect of the entropy setting for the AIC weightings described in Section 3.4.2. For this purpose, we randomly picked one chemical five times as the foreground (out of seven chemicals) while keeping the background fixed (acetone). To make the problem more challenging, the foreground/background ratio was set to 1/20 (i.e., background twenty times stronger than foreground). For each of the five foreground cases, we ran experiments with five different AIC weighting entropies of 0.1, 0.3, 0.5, 0.7 and 0.9; settings of 0 (converge to the first solution found) and 1 (never converge) were not considered since they lead to trivial strategies. Each setting was tested ten times for each foreground, for a total of 250 experiments (5 chemicals × 5 dilutions × 10 replicates), or 50 experiments for each entropy setting. From these experiments, we then counted the number of tests for which the correct solution appeared in the candidate pool ($N_{FIND}$) and the number of tests for which the correct solution ranked as #1 in the pool ($N_{CONF}$). Denoting the total number of tests by $N_{TOT}$, we then calculated three measures:

(1) The discovery rate, measured as the proportion of times that the correct solution is included in the pool ($N_{FIND}/N_{TOT}$),
(2) The resolution rate, measured as the number of times the correct solution is confirmed given that it was included in the pool ($N_{CONF}/N_{FIND}$), and
(3) The confirmation rate, measured as the proportion of times the correct solution is selected ($N_{CONF}/N_{TOT}$)

Results are shown in Fig. 9. When the algorithm uses a higher explorative setting, the discovery rate in Fig. 9a increases, but at the cost of reducing the confirmation rate in Fig. 9b. The final classification rate shows an asymmetric inverted U curve, suggesting that a tradeoff between exploitation and exploration may be found at an entropy setting around 0.5. Interestingly, as shown in Fig. 9c, too much exploitation appears to be more dangerous than too much exploration. In our case, extreme exploitation leads the algorithm to stop gathering information prematurely, eliminating any chance of discovering the correct solution; in contrast, extreme exploration will tend to evaluate the whole spectrum, with repeated sampling to compensate for noise, allowing the algorithm to slowly converge to the correct solution.

## 5. Validation on synthetic data

To provide a more thorough evaluation than what can be afforded experimentally, we also analyzed the active wavelength selection algorithm on a large dataset of synthetic IR spectra. The dataset consisted of FTIR spectra (660 spectral lines) from 500 chemicals in the NIST Webbook infrared absorption spectrum database [45]. To simulate the spectral resolution of FPIs, we convolved the FTIR spectra with a Gaussian filter of 0.1 $\mu$m spread, and added white noise (details included in Section 5.1) to each individual wavelength. Each spectrum was normalized to sum up to one. For the subsequent experiments, we compared the proposed active wavelength selection algorithm against the passive algorithm described in Section 4.3. In all cases, we allowed the algorithms to sample each wavelength multiple times.

### *5.1. Binary mixtures*

In a first experiment, we tested the algorithm on a similar two-chemical mixture problem as in Section 4. However, instead of using a fixed background, both foreground and background were randomly chosen from the library. We then evaluated the algorithm at increasing levels of difficulty by adding Gaussian noise with standard deviation from 0% to 35% of the median value of the complete absorption spectrum library. We also evaluated the algorithm as a function of the degree of collinearity between the foreground and background analytes, measured as the condition number of the column matrix containing the spectra of the two chemical components in the target mixture; the higher the condition number, the more collinear the two chemical are. Fig. 10(e–h) illustrates pairs of spectra at different condition numbers: for lowest condition number[7] the two spectra are nearly orthogonal, whereas for the highest condition number all major peaks from both chemicals overlap.

To measure performance, we considered the number of iterations (wavelength measurements) required for the algorithm to converge to the correct solution, with convergence strictly defined as the correct solution being ranked as #1 among all solutions *and* being 10 times more likely than the second most likely solution. Results are shown in Fig. 10(a–d). At low noise levels, there are no significant differences between both algorithms. As noise levels increase, performance degrades for both algorithms. The effects of noise are considerably amplified by collinearity: when the two spectra are very dissimilar (condition number close to 1) noise has minimal impact, whereas for similar spectra (condition number of 10), the number of required steps increases significantly with noise. The active algorithm consistently outperforms its passive counterpart in all cases.

### *5.2. Higher-order mixtures*

In a second and final experiment, we tested the algorithms on higher-order mixture problems containing up to 15 chemical components. In this case, the noise level was fixed at 1% of the median value of all absorption spectra in the library. As the difficulty of a mixture problem can vary dramatically (i.e., a badly conditioned two chemical mixture can become unsolvable),

---

[7] The lowest condition number for any two pairs of chemicals in our library is 1.2.

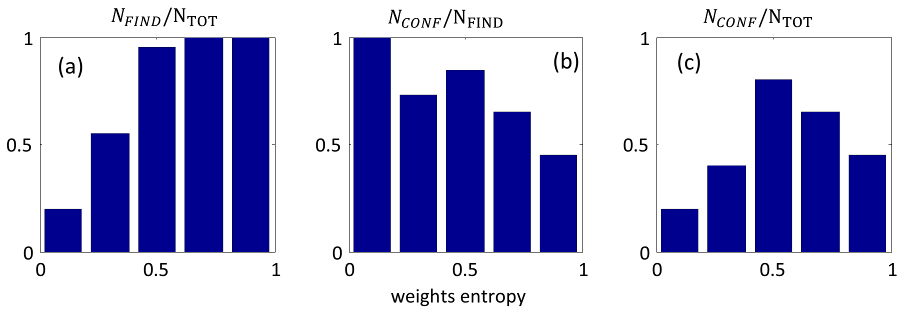


**Fig. 9.** (a) Discovery rate, (b) resolution rate and (c) confirmation rate. The entropy controls the balance between exploitation (entropy being zero) and exploration (entropy being one).

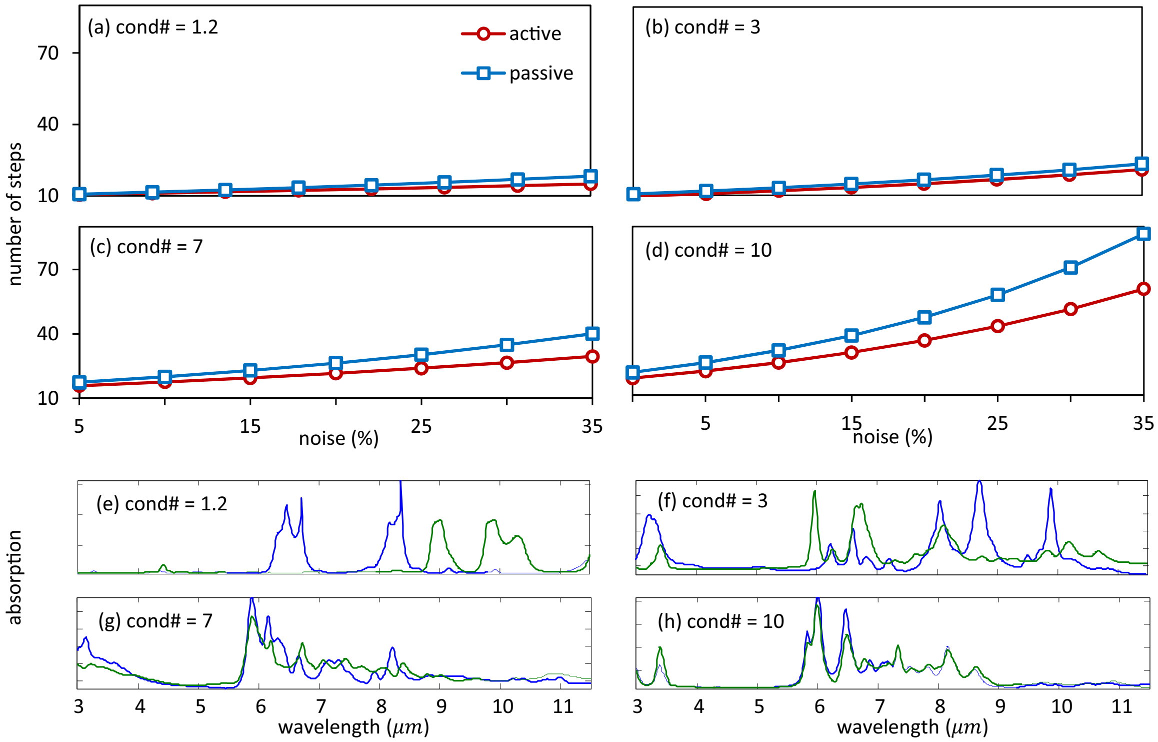


**Fig. 10.** (a–d) Number of steps needed to converge to the correct solution. (e–h) The corresponding foreground and background for each condition number; spectra were normalized to sum up to one.

we designed a mixture construction policy so that the chosen problems would not be arbitrarily easy or hard. For this purpose, instead of condition numbers we used a random wavelength-selection algorithm to rate the difficulty of one hundred randomly selected 15-chemical mixtures, and selected five mixtures that could be correctly classified 1 – 10% of the times using a maximum of 200 randomly selected measurements; this ensured that the highest-order mixture problems were solvable but non-trivial. For each of these five 15-chemical mixtures, we sequentially removed one component at a time to form chemical mixtures of lower order; this process ensured a graded transition in problem complexity from hard to easy. For each of the resulting 45 mixtures (15 × 5), we evaluated the active and passive algorithms 40 times, each time with different added noise, for a total of 3000 cases.

Results are shown in Fig. 11 in terms of the number of measurements needed for the correct solution to be ranked as #1, up to a maximum of 200 measurements. Since the noise level is low (1%), there is no significant difference between active and passive algorithm for problems with up to four chemicals. With five or more

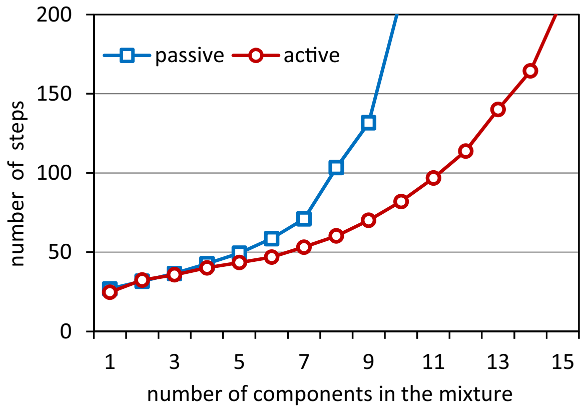


**Fig. 11.** The number of steps used to converge to the correct solution with mixture problems up to 15 chemicals.

chemicals, the active algorithm gradually outperforms the passive algorithm. As expected, the number of measurements needed grows exponentially for both algorithms with the number of chemicals, but the active algorithm can solve a significantly more complex problem than its passive counterpart for a fixed sensing budget. As an example, given 100 measurements the active algorithm can solve an 11-chemical mixture problem whereas the passive algorithm can only solve an 8-chemical problem at best. Likewise, to solve a 9-chemical problem the active algorithm requires 70 measurements on average, whereas the passive algorithm requires 130 measurements.

## 6. Conclusions and discussion

We have proposed an active wavelength selection algorithm for mixture analysis with tunable chemical sensors. The algorithm uses a multi-modal solver to maintain a pool of likely candidate solutions based on previous measurements, then selects its next wavelength as the one which maximizes discrimination among all the candidates in the pool. To address the ill-conditioned nature of the problem, the algorithm promotes sparse solutions with two complementary strategies. First, the algorithm adds mixture components to the candidate solutions in an incremental fashion, from single analytes, to binary mixtures, to ternary mixtures, and so on. Second, the algorithm promotes sparse candidates by means of a weighting function based on the Akaike information criterion. To prevent the search from converging prematurely, the algorithm also uses an entropy-guided normalization method that rebalances the AIC weights such that the strongest candidate solutions do not dominate the wavelength-selection process during the early stages.

The algorithm is first validated experimentally on binary mixture problems with a Fabry-Perot Interferometer. Our results show that active wavelength selection outperforms its passive counterpart, particularly at low concentrations and low foreground-background ratios. We also characterized the algorithm on synthetic data at increasing levels of ill-conditioning and higher-order mixtures, and compared it against a passive algorithm. Active wavelength selection provides higher and more stable performance than passive selection, and more importantly, shows higher tolerance to noise and collinearity. Compared against passive wavelength-selection techniques, which require retraining if additional chemicals are added to (or removed from) the library, active wavelength selection can also be trivially adapted to problems of varying library sizes.

Correlation between neighboring wavelengths can make the library matrix $A(F, :)$ close to singular. In practice, however, the system rarely selects neighboring wavelengths before the underlying linear system reaches full rank: once observations have been made at certain wavelengths, the NNLS solver will fit the candidate models at those measurements with zero error because the system is under-determined. As a result, variance at those wavelengths will be minimized, and so will be the variance at neighboring wavelengths, significantly reducing the chances that they will get selected at the next iteration. It is not until the linear system becomes full rank that the algorithm begins to sample neighboring frequencies to average out noise.

## 7. Future work

Heavier penalties on model complexity may be applied by replacing the AIC in our current implementation with the Bayesian information criterion (BIC) [46]. In contrast with the AIC, which uses the penalty term $2V$, BIC penalizes model complexity according to $V\log(n)$, where $n$ is the number of measurements. As a result, BIC places increasing penalty as additional measurements are taken, making it more useful when the sampling becomes dense. Alternatively, both scores may be used in combination to obtain a stronger indicator of fitness (i.e., when both criteria agree).

Our framework is not limited to sequential sensing with tunable devices, and may be used with conventional spectroscopic instruments. In this case, one would measure the complete spectrum, and use active wavelength selection to select a few measured wavelengths (discarding the rest) to improve the numerical conditioning of the solution pool. The process (measure full spectrum, then actively select wavelengths) would be repeated until the best solution in the pool dominates other solutions, e.g., its AIC score is N times larger than that of the second solution. Such an approach would be particularly useful when the sample has unknown identity, since the framework automates the process of accurately identifying and quantifying the sample without the need for additional training.

Work is underway to extend the active wavelength selection framework to the problem of tracking chemical mixtures, i.e., when concentrations in a chemical mixture are constantly changing. For portable and tunable devices such as FPI-based sensor, this opens the door for a wide range of real-world applications.

## Acknowledgement

This work was supported by NSF under award #1002028.

## Appendix A.

In order to reduce computational costs, the variance of each wavelength across all candidate observations in Eq. (10) is used as an approximation of the misclassification risk. Assume the binary classification problem illustrated in Fig. 12. Using Bayesian decision theory [47], its misclassification risk $\mathscr{R}$ can be calculated as:

$$\mathscr{R} = C_{12}P_{12} + C_{21}P_{21} \tag{19}$$

where $C_{ij}$ is the cost of wrongly assigning a sample to class $j$ when $i$ is the correct class, $P_{ij}$ is the probability of such misclassification: $P_{ij} = \int_{\Gamma_{ij}} p(x)dx$, and $\Gamma_{ij}$ represents the region where such misclassifications may occur (see Fig. 12). Assume that the noise in observation space is independent and normally distributed, and that both costs are equal ($C_{12} = C_{21} = C$). Since the two distributions are symmetric relative to the classification boundary, the misclassification risk is monotonically related to the distance between the two means ($\mu_1, \mu_2$):

$$\mathscr{R} = C_{12}P_{12} + C_{21}P_{21} = \frac{2C}{\sigma\sqrt{2\pi}} \int_{\Gamma_{12}} e^{-\frac{(x-\mu_1)^2}{2\sigma^2}} dx \propto -\Phi\left(\frac{|\mu_1 - \mu_2|}{2}\right) \tag{20}$$

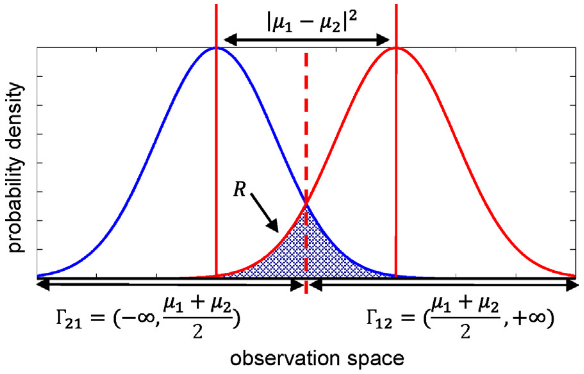


**Fig. 12.** Misclassification risk of a binary classification problem.

where $\Phi(*)$ is a monotonically increasing function. Intuitively, this means that the further the two Gaussian means are, the easier the binary classification problem is.

In our case we have $M$ candidates, so the problem becomes one of $M-ary$ classification with total misclassification risk given by:

$$\mathfrak{R} = \sum_{\forall i,j,i\neq j} C \times P_{ij} \propto -\sum_{\forall i,j,i\neq j} \Phi\left(\frac{|\mu_i-\mu_j|}{2}\right) \tag{21}$$

This computation is expensive when $M$ is large, as is our case. However, since $\Phi(*)$ is monotonic, there also exists a monotonic function $G(*) = \Phi(\sqrt{*}/2)$ such that

$$\Phi\left(\frac{|\mu_i-\mu_j|}{2}\right) = \Phi\left(\frac{\sqrt{|\mu_i-\mu_j|^2}}{2}\right) = G(|\mu_i-\mu_j|^2) \tag{22}$$

Using the 1st order Taylor approximation: $G(z) \approx G(0) + G'(0) \times z$, we have:

$$\begin{aligned} R &\propto -\sum_{\forall i,j,i\neq j} \Phi\left(\frac{|\mu_i-\mu_j|}{2}\right) \\ &\approx -\left(G(0) + G'(0) \times \sum_{\forall i,j,i\neq j} |\mu_i-\mu_j|^2\right) \propto -\sigma^2(\mu) \end{aligned} \tag{23}$$

where $\sigma^2(\mu)$ is the variance defined in Eq. (10). Thus, by selecting the wavelength with maximum variance we minimise the misclassification risk.

## References

[1] H. Ding, J. Liang, J. Cui, X. Wu, A novel fiber Fabry-Perot filter based mixed-gas sensing system, Sens. Actuators B: Chem. 138 (2009) 154–159.

[2] N. Neumann, M. Ebermann, S. Kurth, K. Hiller, Novel MWIR microspectrometer based on a tunable detector, in: SPIE MOEMS and Miniaturized System VIII, 2009.

[3] C.H. Spiegelman, M.J. McShane, M.J. Goetz, M. Motamedi, Q.L. Yue, G.L. Gote, Theoretical justification of wavelength selection in PLS calibration: development of a new algorithm, Anal. Chem. 70 (1998) 35–44.

[4] B. Hemmateenejad, R. Ghavamia, R. Mirib, M. Shamsipurc, Net analyte signal-based simultaneous determination of antazoline and naphazoline using wavelength region selection by experimental design-neural networks, Talanta 67 (2006) 1222–1229.

[5] M.C.U. Araújo, T.C.B. Saldanha, R.K.H. Galvão1, T. Yoneyama, H.C. Chame, V. Visani, The successive projections algorithm for variable selection in spectroscopic multicomponent analysis, Chemometr. Intell. Lab. Syst. 57 (2001) 65–73.

[6] B. Hemmateenejad, M. Akhond, F. Samari, A comparative study between PCR and PLS in simultaneous spectrophotometric determination of diphenylamine, aniline, and phenol: Effect of wavelength selection, Spectrochim. Acta A 67 (2007) 958–965.

[7] H. Swierengaa, F. Wülfertb, O.E. de Noordc, A.P. de Weijerd, A.K. Smildeb, L.M.C. Buydens, Development of robust calibration models in near infra-red spectrometric applications, Anal. Chim. Acta 311 (2000) 121–135.

[8] R. Leardi, Genetic algorithms in chemometrics and chemistry: a review, J. Chemometr. 15 (2001) 559–569.

[9] M. Shamsipur, V. Zare-Shahabadi, B. Hemmateenejad, M. Akhond, Ant colony optimisation: a powerful tool for wavelength selection, J. Chemometr. 20 (2006) 146–157.

[10] Y.P. Du, Y.Z. Liang, J.H. Jiang, R.J. Berry, Y. Ozaki, Spectral regions selection to improve prediction ability of PLS models by changeable size moving window partial least squares and searching combination moving window partial least squares, Anal. Chim. Acta 501 (2004) 183–191.

[11] J.J. Gibson, The Ecological Approach to Visual Perception, 1st ed., Lawrence Erlbaum Associates, New Jersey, 1986.

[12] J.J. Gibson, Observations on active touch, Psychol. Rev. 69 (1962) 477–491.

[13] R. Bajcsy, Active perception, Proc. IEEE 76 (1988) 966–1005.

[14] J. Aloimonos, I. Weiss, A. Bandyopadhyay, Active vision, Int. J. Comput. Vis. 1 (1988) 333–356.

[15] Z. Yongmian, J. Qiang, Active and dynamic information fusion for facial expression understanding from image sequences, IEEE Transactions on Pattern Analysis and Machine Intelligence 27 (2005) 699–714.

[16] T.H. Chung, V. Gupta, J.W. Burdick, R.M. Murray, On a decentralized active sensing strategy using mobile sensor platforms in a network, in: 43rd IEEE Conference on Decision and Control (CDC 2004), vol. 2, 2004, pp. 1914–1919.

[17] N. Roy, W. Burgard, 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.

[18] D. Fox, W. Burgard, S. Thrun, Active Markov localization for mobile robots, Robot. Auton. Syst. 25 (1998) 195–207.

[19] J.J. Leonard, H.F. Durrant-Whyte, Simultaneous map building and localization for an autonomous mobile robot, in: IEEE/RSJ International Workshop on Intelligent Robots and Systems '91. 'Intelligence for Mechanical Systems, 1991, pp. 1442–1447.

[20] H.J. Kushner, A new method of locating the maximum point of an arbitrary multipeak curve in the presence of noise, J. Basic Eng. 86 (1964) 97–106.

[21] J. Močkus, On Bayesian methods for seeking the extremum, in: Optimization Techniques IFIP Technical Conference, 1975, pp. 400–404.

[22] J. Mockus, Application of Bayesian approach to numerical methods of global and stochastic optimization, J. Glob. Optim. 4 (1994) 347–365.

[23] C.E. Rasmussen, C.K.I. Williams, Gaussian Processes for Machine Learning, The MIT Press, 2006.

[24] E.J. Candes, T. Tao, Near-optimal signal recovery from random projections: universal encoding strategies, IEEE Trans. Inf. Theory 52 (2006) 5406–5425.

[25] J. Haupt, R. Castro, R. Nowak, Distilled sensing: selective sampling for sparse signal recovery, in: 12th International Conference on Artificial Intelligence and Statistics (AISTATS), 2009.

[26] R.M. Castro, R.D. Nowak, Minimax bounds for active learning, IEEE Trans. Inf. Theory 54 (2008) 2339–2353.

[27] R. Lomasky, C.E. Brodley, M. Aernecke, D. Walt, M. Friedl, in: J. Kok, J. Koronacki, R. Mantaras, S. Matwin, D. Mladenič, A. Skowron (Eds.), Active Class Selection, In Machine Learning: ECML 2007, vol. 4701, Springer, Berlin, Heidelberg, 2007, pp. 640–647.

[28] I. Rodriguez-Lujan, J. Fonollosa, A. Vergara, M. Homer, R. Huerta, On the calibration of sensor arrays for pattern recognition using the minimal number of experiments, Chemometr. Intell. Lab. Syst. 130 (2014) 123–134.

[29] D.V. Dinakarababu, D.R. Golish, M.E. Gehm, Adaptive feature specific spectroscopy for rapid chemical identification, Opt. Express 19 (2011) 4595–4610.

[30] R. Gosangi, R. Gutierrez-Osuna, Active temperature programming for metal-oxide chemoresistors, IEEE Sens. J. 10 (2010) 1075–1082.

[31] R. Gosangi, R. Gutierrez-Osuna, Active temperature modulation of metal-oxide sensors for quantitative analysis of gas mixtures, Sens. Actuators B: Chem. 185 (2013) 201–210.

[32] R. Gosangi, R. Gutierrez-Osuna, Active classification with arrays of tunable chemical sensors, Chemometr. Intell. Lab. Syst. 132 (2014) 91–102.

[33] J. Huang, R. Gosangi, R. Gutierrez-Osuna, Active concentration-independent chemical identification with a tunable infrared sensor, IEEE Sens. J. 12 (2012) 3135–3142.

[34] D.D. Lee, H.S. Seung, Learning the parts of objects by non-negative matrix factorization, Nature 401 (1999) 788–791.

[35] J. Huang, R. Gutierrez-Osuna, Active analysis of chemical mixtures with multimodal sparse non-negative least squares, in: Presented at the International Conference on Acoustics, Speech, and Signal Processing (ICASSP), 2013.

[36] I. Guyon, A. Elisseeff, An introduction to variable and feature selection, J. Mach. Learn. Res. 3 (2003) 1157–1182.

[37] C.L. Lawson, R.J. Hanson, Solving Least Squares Problems, Prentice-Hall Inc., 1974.

[38] G.S. Mallat, Z. Zhang, Matching pursuits with time–frequency dictionaries, IEEE Trans. Signal Process. 41 (1993) 3415–3997.

[39] S. Das, S. Maity, B. Qu, P.N. Suganthan, Real-parameter evolutionary multimodal optimization – a survey of the state-of-the-art, Swarm Evol. Comput. 1 (2011) 71–88.

[40] E.J. Wagenmakers, S. Farrell, AIC model selection using Akaike weights, Psychon. Bull. Rev. 11 (2004) 192–196.

[41] K.P. Burnham, D.R. Anderson, Model Selection and Multimodel Inference: A Practical Information-Theoretic Approach, Springer, 2002, pp. 63.

[42] S.W. Wilson, Explore/exploit strategies in autonomy, in: Proceedings of the 4th International Conference on Simulation of Adaptive Behavior From Animals to Animats 4: Proceedings of the 4th International Conference on Simulation of Adaptive Behavior, 1996, pp. 325–332.

[43] N. Neumann, S. Kurth, K. Hiller, M. Ebermann, Tunable infrared detector with integrated micromachined Fabry-Perot filter, J. Micro/Nanolithogr. MEMS MOEMS 7 (2008) 021004-l–121004-l.

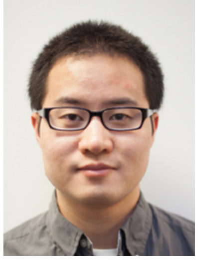
[44] A.W. Whitney, A direct method of nonparametric measurement selection, IEEE Trans. Comput. 100 (1971) 1100–1103.

[45] P.M. Chu, F.R. Guenther, G.C. Rhoderick, W.J. Lafferty, Qunatitative infrared database, in: P.J. Linstrom, W.G. Mallard (Eds.), NIST Chemistry WebBook, NIST Standard Reference Database Number 69, National Institute of Standards and Technology, Gaithersburg MD, 2010.

[46] G. Schwarz, Estimating the dimension of a model, Ann. Stat. 6 (1978) 461–464.

[47] H.V. Poor, An Introduction to Signal Detection and Estimation, Springer, 1994, pp. 5–9.

## Biographies



**Jin Huang** received the B.E. degree in control science and engineering from Huazhong University of Science and Technology (HUST) in 2009. He is currently working towards the Ph.D. degree at Texas A&M University, College Station. His research interests include pattern recognition, active sensing, and chemical sensors.

**Ricardo Gutierrez-Osuna** (M'00) received the B.S. degree in electrical engineering from the Polytechnic University of Madrid, Madrid, Spain, in 1992, and the M.S. and Ph.D. degrees in computer engineering from North Carolina State University, Raleigh, in 1995 and 1998, respectively. From 1998 to 2002, he served on the faculty at Wright State University. He is currently a Professor of Computer Engineering at Texas A&M University. His research interests include pattern recognition, neuromorphic computation, chemical sensor arrays, and audio-visual speech processing.