

Adaptive spectroscopy for rapid chemical identification

Dineshbabu V. Dinakarababu^a and Michael E. Gehm^{a,b}

^aDepartment of Electrical and Computer Engineering, University of Arizona, Tucson, AZ;

^bCollege of Optical Sciences, University of Arizona, Tucson, AZ

ABSTRACT

Spectroscopic chemical identification is fundamentally a classification task where sensor measurements are compared to a library of known compounds with the hope of determining an unambiguous match. When the measurement signal-to-noise ratio (SNR) is very low (e.g. from short exposure times, weak analyte signatures, etc.), classification can become very challenging, requiring a multiple-measurement framework such as sequential hypothesis testing, and dramatically extending the time required to classify the sample. There are a wide variety of defense, security, and medical applications where rapid identification is essential, and hence such delays are disastrous. In this paper, we discuss an approach for adaptive spectroscopic detection where the introduction of a tunable spectral filter enables the system to measure the projection of the sample spectrum along arbitrary bases in the spectral domain. The net effect is a significant reduction in time-to-decision in low SNR cases. We describe the general operation of such an instrument, present results from initial simulations, and report on our experimental progress.

Keywords: spectroscopy, adaptive spectroscopy, adaptive optics, chemical detection, sequential hypothesis testing

1. INTRODUCTION

Chemical identification with the aid of spectroscopic techniques has a wide range of defense^{1,2}, security³ and medical^{4,5} applications. One of the most important drawback of such chemical sensing techniques arises when the measurement signal-to-noise ratio (SNR) is very low. A variety of real-world scenarios (limited exposure time, low analyte concentrations, etc.) frequently result in critical classification tasks that must operate in this SNR regime. Under these circumstances, generating an accurate decision requires multiple measurements, thereby significantly increasing the required time-to-identification.

In this manuscript, we discuss a detection scheme based on a new type of *adaptive spectrometer* that reconfigures its internal optical structure based on the results of prior measurements. The proposed system is a *feature-specific* spectrometer that measures the projection of the incoming spectral density onto a set of *feature vectors*, rather than sampling the spectral density directly. This general approach has been explored previously in imaging.^{6,7} Here we explore the extension to a version where the nature of the feature vectors can change over time in response to information gained from the compound under test. We call this device an *adaptive, feature-specific spectrometer* (AFSS).

Below, we present the general theory behind the AFSS as well as initial simulation results comparing the performance of the AFSS to a traditional spectrometer. We show that, in the low-SNR limit, the AFSS achieves a time-to-decision that is several orders of magnitude shorter than a traditional system—providing a dramatic advantage for a host of critical chemical detection situations. Subsequent sections detail our current efforts to develop a working experimental prototype.

2. THEORY

We state the chemical detection problem as follows: Given a *sample under test*, find the closest match in a known *spectral library*. We represent the spectrum of the sample under test as the column vector s and the spectral library as the $p \times r$ matrix S , with the columns of S being the r *known spectra* in the library.

Corresponding address: gehm@ece.arizona.edu

2.1 MEASUREMENT MODELS

Before we can discuss the introduction of adaptivity into the system, we must first develop a detection framework for use with both the traditional and feature-specific measurement paradigms. To do this, we require a measurement model for the two system architectures.

2.1.1 TRADITIONAL SPECTROMETER

A *traditional spectrometer* samples the incoming spectrum s in the p spectral channels corresponding to the channels in the spectral library S . The resulting measurement can be represented by the column vector

$$m_t = s + n_t, \quad (1)$$

where n_t is a length- p column vector representing the noise contribution from the sampling in each of the spectral channels.

2.1.2 FEATURE-SPECIFIC SPECTROMETER

As discussed briefly above, the proposed adaptive spectrometer is more fully referred to as an adaptive feature-specific spectrometer (AFSS). In a feature-specific spectrometer, the spectral channels are not measured individually, but are instead measured in one or more weighted sums that are equivalent to forming the inner product(s) of the spectrum s with a set of *feature vectors* (the inner product(s) representing the *projection* of the signal vector onto an feature vector). If we represent a set of q analysis vectors as the $q \times p$ *projection matrix* P , then the result is a length- q measurement vector:

$$m_p = Ps + n_p, \quad (2)$$

where n_p is now a length- q column vector representing the noise contribution from the formation of each of the projections.

In general, feature-specific instruments have improved performance compared to traditional instruments as a result of the increase in measurement signal-to-noise ratio (SNR)^{6,7}—each measurement in a feature-specific instruments involves multiple signal elements (spectral channels in our case) and only a single noise contribution.

2.2 DETECTION/CLASSIFICATION

Given a measurement vector m_t or m_p , how does one determine the best match in the spectral library S ? The straightforward approach is through the application of a matched filter. For the traditional spectrometer, the measured signal is tested with matched filters based on the r spectra in S , and the largest result is taken as indicating the “match”. A similar approach can be formulated with the feature-specific spectrometer, but we must first transform S into a form that represents how the spectra appear in the given projections:

$$S_p = PS. \quad (3)$$

The measurement vector m_p is then tested with matched filters based on the r “signatures” in S_p . The largest result is taken as indicating the “match”.

2.2.1 SEQUENTIAL HYPOTHESIS TESTING

The matched-filter approach works well when the signal-to-noise ratio (SNR) is large. However, as SNR degrades, it becomes increasingly likely that the matched filter approach will misidentify the spectrum. A possible solution to this problem is to modify the decision framework so that it becomes possible to consider the results of a *measurement sequence* rather than a single measurement. The simplest formulation uses the standard Neyman-Pearson or Bayesian decision criteria⁸⁻¹⁰ on a fixed measurement sequence. This formulation, however, has two significant shortcomings. By deciding the length of the measurement sequence *a priori* we can either force a classification when we still lack sufficient information (the “need more data” scenario) or we can continue acquiring data long after we have sufficient information to make the classification (the “too much data” scenario). There are many situations (medical, security, etc.) where either scenario counts as a failure.

There exists a decision framework that constantly monitors the quality of the information and decides, after each measurement, whether the system has reached a point where classification will meet our desired tolerances for misses and false alarms. This framework, *sequential hypothesis testing* (SHT) was first formulated by Wald¹¹ for decisions between two competing hypothesis, and was extended to the multiple hypothesis case by Armitage.¹²

Two-hypotheses SHT is formulated as follows. After each measurement, we wish to know the probability associated with each of the two hypotheses (H_0 and H_1) given the total set of k measurements taken to that point ($\{m\}_k$). We can write these conditional probabilities as

$$\Pr(H_0|\{m\}_k), \quad \Pr(H_1|\{m\}_k). \quad (4)$$

Using Bayes' theorem, we can write these conditional probabilities in terms of likelihoods

$$\Pr(H_0|\{m\}_k) = \frac{L_{0,k} \Pr(H_0)}{\Pr(\{m\}_k)}, \quad \Pr(H_1|\{m\}_k) = \frac{L_{1,k} \Pr(H_1)}{\Pr(\{m\}_k)}, \quad (5)$$

where $L_{i,k} = \Pr(\{m\}_k|H_i)$ is the likelihood of the i th hypothesis based on the entire dataset $\{m\}_k$.

If we form the ratio of the probabilities

$$\Lambda_k = \frac{\Pr(H_0|\{m\}_k)}{\Pr(H_1|\{m\}_k)} = \frac{L_{0,k} H_0}{L_{1,k} H_1} = L_{01,k} \frac{H_0}{H_1}, \quad (6)$$

we see that this can be expressed in terms of the *likelihood ratio* $L_{01,k} = L_{0,k}/L_{1,k}$ multiplied by the ratio of the prior probabilities for the two hypotheses. We then note that we can expand the i th likelihood as

$$L_{i,k} = \Pr(H_i|m) \Pr(H_i|\{m\}_{k-1}), \quad (7)$$

where m is now the result of *just the k th measurement*. In this form we see that

$$\Lambda_k = L_{01} L_{01,k-1} \frac{H_0}{H_1} = L_{01} \Lambda_{k-1}. \quad (8)$$

Here L_{01} represents the likelihood ratio of *just the k th measurement*. Thus we have an update procedure, whereby we can update the probability ratio Λ_k based on the likelihood ratio of the most recent measurement and the previous value of the probability ratio Λ_{k-1} .

Clearly, when Λ gets very large, we wish to decide in favor of hypothesis H_0 and when Λ is very small, we wish to decide in favor of hypothesis H_1 . As before we can use Neyman-Pearson or Bayesian methods to define *decision thresholds* Θ_0 and Θ_1 based on our desired probabilities of type I and II errors (false positive and miss, respectively). Given these thresholds, the SHT framework computes Λ after each measurement. The outcome after that measurement is determined as follows:

Decide " H_0 " if $\Lambda > \Theta_0$
 Decide " H_1 " if $\Lambda < \Theta_1$
 Otherwise, make no decision. Make another measurement.

2.2.2 EXTENSION TO MULTIPLE HYPOTHESES

Armitage¹² provided a straightforward extension of the SHT framework to the w -hypothesis case ($w > 2$). We define

$$\Lambda_{i,j;k} = L_{ij,m} \Lambda_{i,j;k-1}. \quad (9)$$

Here Λ is no longer a scalar, but instead a $w \times w$ *matrix* which captures all pairwise Λ 's calculated as in the two-hypothesis case. The i, j th element of Λ is then the probability ratio $\Pr(H_i|\{m\}_k)/\Pr(H_j|\{m\}_k)$. We decide for a hypothesis when every element in a row (other than the diagonal element) is larger than the corresponding Θ_i (thus ruling in *favor of H_i* with respect to all the other hypotheses). By construction, $\Lambda_{i,j;k} = 1/\Lambda_{j,i;k}$, thus ensuring that only one hypothesis can meet this row-wise test at a time. Thus the outcome after any measurement is determined as follows:

Decide " H_i " if $\Lambda_{i,j;k} > \Theta_i \forall j \neq i$
 Otherwise, make no decision. Make another measurement.

2.3 ADAPTIVITY

Previously, we noted how use of a feature-specific instrument can improve system performance given the resulting increase in measurement SNR. Precisely how features are selected was not discussed, but we could imagine many *ad hoc* methods based on various decompositions of the spectra in the spectral library. When we introduce the SHT framework, we gain an iterative measurement/decision approach that works with either a traditional or feature-specific spectrometer.

In the case of the feature-specific spectrometer, however, application of the SHT framework raises an interesting point: as we proceed through the SHT process towards a decision, we are continually refining our knowledge about the relative probabilities of the various hypotheses. A straightforward feature-specific spectrometer utilizes a constant set of features that are not informed by any of the knowledge gained during the SHT process. A better approach would be to adaptively modify the features based on the information contained in all of the data collected in prior measurements. In that way, the system will evolve to features that are *maximally discriminatory* for the particular hypotheses still in serious contention. If implemented properly, the result can only be a faster time-to-decision.

To understand the adaptive approach we utilize, some further discussion of the feature-specific methodology is in order. We can view in the input spectrum s as an p -dimensional signal vector. The measurements of a traditional spectrometer are the projections of this signal onto the orthogonal basis associated with the different spectral channels. A feature-specific spectrometer, by contrast, measures the projections onto a set of q arbitrary basis vectors in this p -dimensional space. Ideally, these vectors are chosen such that the projections are highly discriminatory between the hypotheses.

A reasonable choice for a non-adaptive feature-specific spectrometer would be to utilize a set of features based on some number of the *principal components* of the spectra in the library. By definition, the first principal component of a set of signals is the direction that captures the greatest proportion of the variance in the signals¹³ (hence is the *most discriminatory* direction for a single projection). Likewise, the second principal component is the direction orthogonal to the first principal component that captures the greatest proportion of the remaining variance in the signal, and so on. Note that a scheme based on the principal components is necessarily *ad hoc*, as the maximally-discriminating nature of the principal components is only with regard to a single measurement. Given a total number k of measurements, it is possible that there exist sets of k features that are more discriminatory than any set of k features based on principal components. However, this global optimization is not obviously tractable, so we fall back on an *ad hoc* strategy based on principal components.

The extension to an adaptive system is then clear. Rather than working with the principal components of the spectral library, we wish to determine the principal components of some *probabilistically-weighted* library based on our current probability estimates for each of the hypotheses. Thus, our features will be highly discriminatory between the *hypotheses we believe to be in serious contention*.

The principal components are normally defined as the eigenvectors of the signal covariance matrix. If we define the b th spectrum in the library as S_b , then the mean spectrum \bar{S} in the library is

$$\bar{S} = \frac{1}{r} \sum_{b=1}^r S_b. \quad (10)$$

We can define the signal covariance matrix as

$$C = \sum_{b=1}^r (S_b - \bar{S})(S_b - \bar{S})^T, \quad (11)$$

and the q dominant eigenvectors of C would then be a reasonable set of features *given equiprobable hypotheses*. To incorporate the probability estimates generated during the SHT process, we must first convert from probability ratios Λ to probabilities. If we take the j th column of the Λ matrix, the numbers represent the probability ratios $\Pr(H_i|\{m\}_k)/\Pr(H_j|\{m\}_k)$. Summing these terms and normalizing the sum to one is sufficient to determine the denominator, and hence the value of the numerator for each term—producing the desired probability estimate for each hypothesis. If we define the probability estimate associated with the b th spectrum after k measurements

as $\Pr(H_b|\{m\}_k)$, then we can write the *probability-weighted covariance matrix* (also called the *inter-class scatter matrix*) Q_k as

$$Q_k = \sum_{b=1}^r \Pr(H_b|\{m\}_k) (S_b - \bar{S})(S_b - \bar{S})^T, \quad (12)$$

with \bar{S} now redefined as the *probabilistically-weighted signal mean*

$$\bar{S} = \sum_{b=1}^r \Pr(H_b|\{m\}_k) S_b. \quad (13)$$

For the $k + 1$ th measurement, we then use the q dominant eigenvectors of Q_k as our feature basis. A block diagram showing the combination of SHT with adaptive projective measurement is shown in Fig. 1.

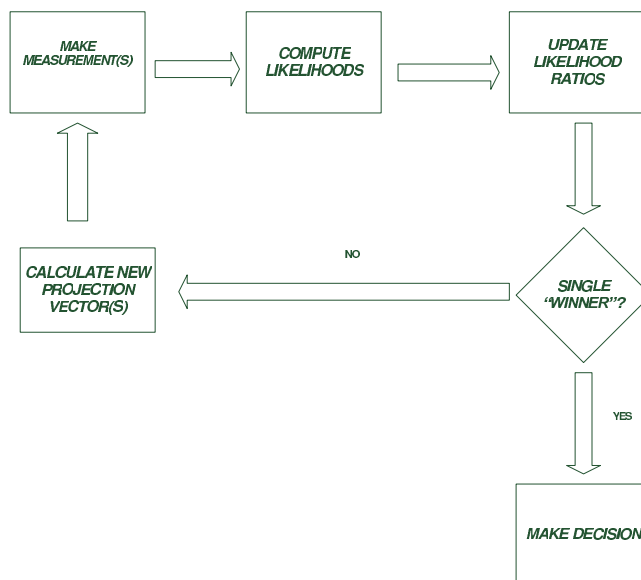


Figure 1. A detailed block diagram showing the flow of the decision making process that applies SHT in an instrument that utilizes adaptively-generated projections.

3. SIMULATION RESULTS

The sections above describe the basic operational concept of an AFSS. To investigate the relative performance of this type of system, we ran a number of comparative simulations. We had at our disposal a master spectral library containing the Raman spectra of several hundred pharmaceuticals sampled in 1300 spectral channels.

Each simulation run began by selecting five spectra at random from the master library. These spectra then formed the spectral library S for that particular run (in other words, we are simulating performance on the $w = 5$ -class problem). We wish to compare the performance of the systems at various SNR values, so we need a workable definition of that quantity. We define the *class separation* of the library σ_l as the mean of the per-channel standard deviation of the spectra in the library. A simulation can then be run at a desired $\text{SNR} = \sigma_l/\sigma_n$ by using an AWGN noise contribution of standard deviation σ_n .

For a given library/SNR combination, one of the spectra in the library was selected and used as the input to simulations of both a traditional spectrometer and an AFSS. The traditional spectrometer sampled all 1300 spectral channels in parallel, using the measurement model described in Sec. 2.1.1. The decision procedure was based on multiple-hypothesis SHT, with decision thresholds set so that there was a 1% chance of a miss or a false alarm. The number of acquisitions required before reaching a decision was recorded.

For the AFSS, a projection matrix based on the q largest principal components of the inter-class scatter matrix was used. A particular detail of the AFSS approach is that the principal component methodology can sometimes result in features that contain both positive and negative values. As a physical AFSS system will work with optical intensity, creation of negative weights is not physically possible. Thus, during the simulation, vectors that had both positive and negative quantities were decomposed into two feature vectors: one representing the positive weights and one representing the negative weights. The simulation measured both projections (using positive weights), and synthesized the desired projection by subtracting the results. This decomposition was accurately tracked in the count of required acquisitions. Given the set of feature vectors, the results were generated according to the measurement model in Sec. 2.1.2. After each acquisition, the probability estimates were revised, and new feature vectors were calculated. The decision process was again multiple-hypothesis SHT, with identical decision thresholds.

Obviously, the relative performance of the two instruments on a given simulation run will be highly influenced by the particular nature of the spectral library. Thus, we repeated the simulation at each desired SNR level 500 times (generating a new random spectral library from the master library each time), and report the *mean* time to decision for the two system types. The results are shown in Fig. 2. At low-SNR, the AFSS clearly comes to a decision more rapidly than the traditional instrument.

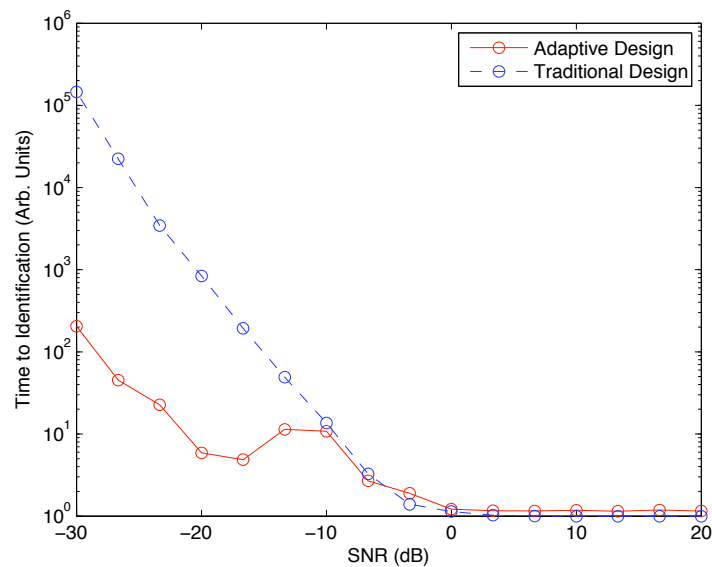


Figure 2. Performance on the $w = 5$ -class problem for both a traditional and an AFSS. Each data point represents the mean of 500 unique simulations. As expected the adaptive approach produces superior results in low-SNR situations. This specific simulation represents the simplest possible AFSS case: $q = 1$ and greyscale features were approximated by posterizing the feature with the *signum* function—all positive weights were converted to a value of 1 and all negative weights were converted to a value of -1.

We are currently performing additional simulations to explore how the relative performance of the systems scales with the size of the spectral library (i.e. with the number of classes in the decision problem).

4. EXPERIMENTAL IMPLEMENTATION

As discussed above, initial simulation results have matched our expectations with regard to the relative performance of the two system types. We have begun construction of an experimental AFSS prototype to validate the simulation results. Obviously, the key challenge in constructing a prototype is developing the adaptive spectral filter that implements the desired spectral projections. We have settled on a system design that is based on a traditional, dispersive spectrometer. Light entering the system strikes a dispersive element that spatially spreads

the light according to its spectral content. In the traditional instrument, this spatial distribution is then sampled by a detector array and the spatial variation of intensity is directly related to the spectral intensity of the source.

For the adaptive instrument, we replace the detector array by a digital micromirror device (DMD). The spatial variation of the mirror orientation can be controlled to selectively combine certain spectral bands onto a single detector element, while diverting the other spectral bands to a beam dump. Grayscale weighting of the bands is possible by dithering the mirror orientation and changing the effective duty cycle. Schematics of both the traditional dispersive instrument and the modified adaptive version are shown in Fig. 3.

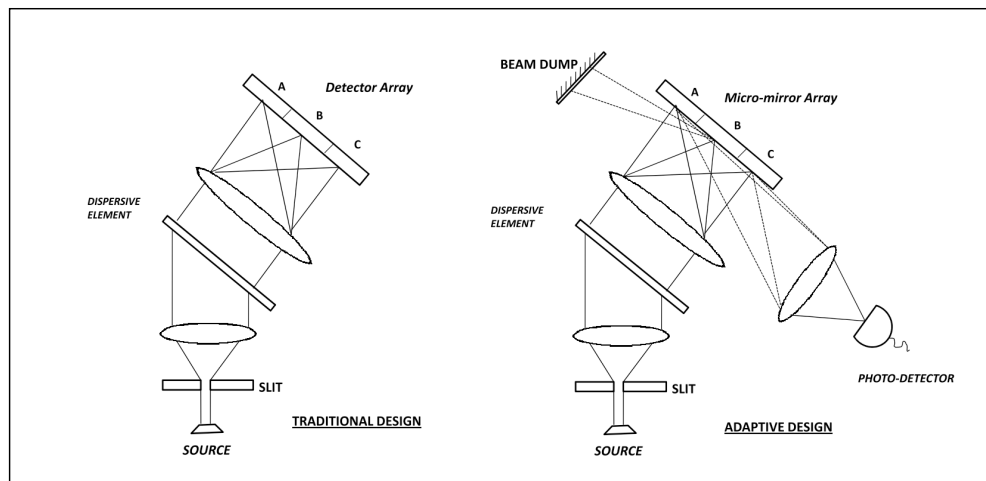


Figure 3. (Left) General layout of a traditional dispersive spectrometer. A dispersive element spatially separates spectral channels, causing them to fall on adjacent photodetectors in a detector array. The spatial variation of intensity is then directly related to the spectral density of the source. (Right) Schematic of the proposed AFSS. The detector array is replaced with a DMD, allowing individual spectral channels to either be combined on a single photodetector (channels A and C) or redirected to a beam dump (channel B).

Our experimental efforts are ongoing. We hope to have experimental results in time for presentation at the conference.

5. CONCLUSION AND FUTURE WORK

In summary, we have presented a new spectrometer architecture that is optimized for the chemical detection task. The system is a feature-specific spectrometer (thereby gaining the improvement in measurement SNR typical of such systems) that utilizes features that adaptively change in time in response to growing knowledge about the identity of the compound under test (thereby reaching a classification decision more rapidly).

The feature vectors are currently selected in an *ad hoc* (yet reasonable) approach based on principal components of the inter-class scatter matrix (which incorporates the changing probabilistic estimates for the different classification outcomes). We are currently investigating methods for deriving truly optimal feature vectors. Simulations support our intuition, demonstrating reductions in the time-to detection of several orders of magnitude over a traditional spectrometer.

We are currently constructing an experimental prototype to validate the simulation results. The proposed architecture utilizes a DMM array to allow active mixing of various spectral channels onto a single photodetector—thereby allowing projections onto arbitrary feature vectors.

REFERENCES

- [1] Sun, Y. and Ong, K., [*Detection technologies for chemical warfare agents and toxic vapors*], CRC Press (2005).

- [2] Pearman, W. and Fountain, A., "Classification of chemical and biological warfare agent simulants by surface-enhanced Raman spectroscopy and multivariate statistical techniques," *Applied spectroscopy* **60**(4), 356–365 (2006).
- [3] Liu, H., Chen, Y., Bastiaans, G., and Zhang, X., "Detection and identification of explosive RDX by THz diffuse reflection spectroscopy," *Optics Express* **14**(1), 415–423 (2006).
- [4] Maquelin, K., van Vreeswijk, T., Endtz, H., Smith, B., Bennett, R., Bruining, H., and Puppels, G., "Raman spectroscopic method for identification of clinically relevant microorganisms growing on solid culture medium," *Anal. Chem* **72**(1), 12–19 (2000).
- [5] Maquelin, K., Kirschner, C., Choo-Smith, L., Van Den Braak, N., Endtz, H., Naumann, D., and Puppels, G., "Identification of medically relevant microorganisms by vibrational spectroscopy," *Journal of microbiological methods* **51**(3), 255–271 (2002).
- [6] Pal, H., Ganotra, D., and Neifeld, M., "Face recognition by using feature-specific imaging," *Applied Optics* **44**(18), 3784–3794 (2005).
- [7] Shankar, M., "Feature-Specific Imaging," *Appl. Opt* **42**, 3379–3389 (2003).
- [8] Kay, S., [*Fundamentals of Statistical signal processing, Volume 2: Detection theory*], Prentice Hall PTR (1998).
- [9] Helstrom, C., [*Statistical theory of signal detection*], Pergamon (1968).
- [10] Wickens, T., [*Elementary signal detection theory*], Oxford University Press, USA (2002).
- [11] Wald, A., [*Sequential analysis*], Wiley New York (1947).
- [12] Armitage, P., "Sequential analysis with more than two alternative hypotheses, and its relation to discriminant function analysis," *Journal of the Royal Statistical Society. Series B (Methodological)* , 137–144 (1950).
- [13] Jolliffe, I., [*Principal component analysis*], Springer New York (2002).