AN OVERVIEW OF SIGNAL PROCESSING ISSUES IN CHEMICAL SENSING

Laurent Duval1, Leonardo T. Duarte2, and Christian Jutten3

1 IFP Energies nouvelles
Technology Division
Rueil-Malmaison, France
laurent.duval@ifpen.fr

2 University of Campinas (UNICAMP)
School of Applied Sciences
Limeira, Brazil
leonardo.duarte@fca.unicamp.br

3 University of Grenoble
and Institut Universitaire de France
GIPSA-lab (UMR 5216), Grenoble, France
christian.jutten@gipsa-lab.grenoble-inp.fr

ABSTRACT

This tutorial paper1 aims at summarizing some problems, ranging from analytical chemistry to novel chemical sensors, that can be addressed with classical or advanced methods of signal and image processing. We gather them under the denomination of “chemical sensing”. It is meant to introduce the special session “Signal Processing for Chemical Sensing” with a large overview of issues which have been and remain to be addressed in this application domain, including chemical analysis leading to PARAFAC/tensor methods, hyper spectral imaging, ion-sensitive sensors, artificial nose, chromatography, mass spectrometry, etc. For enlarging and illustrating the points of view of this tutorial, the invited papers of the session consider other applications (NMR, Raman spectroscopy, recognition of explosive compounds, etc.) addressed by various methods, e.g. source separation, Bayesian, and exploiting typical chemical signal priors like positivity, linearity, unit-concentration or sparsity.

Index Terms— Chemical analysis, Chemical sensors, Gas chromatography, Signal processing algorithms, Spectroscopy

1. INTRODUCTION: AIMS AND MOTIVATIONS

With the advent of more affordable, higher resolution or novel data acquisition techniques, chemical analysis has been using, progressively, more advanced signal and image processing tools.

Since analytical chemistry (AC) has numerous applications in forensics, bioinformatics, clinical, environmental and material analysis, investigating how signal processing (SP) methods, thereafter encompassing image analysis as well, can be used for solving analytical chemistry problems will be interesting in many application domains. Indeed, both specialties (AC and SP) share very similar values of best practice in carrying out identifications and comprehensive characterizations, althetby of chemical samples for AC or of numerical data for SP.

1This work has been partly supported by the European project ERC-2012-AdG-320684-CHESS.

For instance, the chemical analyst approach to performing an analysis, resorting from different preparation steps to different analytical techniques (Section 2.1), resembles the manner employed in traditional signal or image analysis. As a consequence, a better interaction between both communities is possible and desirable.

Interactions between SP and AC communities would be useful in providing new types of data and constraints and in solving AC related issues [1]. Conversely, interactions can also be beneficial for the SP community, with opportunities in less known tools. As a first example, it is clear that the well known PARAFAC approach [2] played an important role in SP: it has been (and still is) a source of inspiration for source separation methods and other representations of complex multi-way data based on tensor decomposition. A second example is the Savitsky-Golay filter, whose original work [3] is one of the most cited papers in analytical chemistry [4]. The design of these filters make them shape-preserving smoothers, better suited to denoising empirical data comprised of sum of round-shaped peaks than standard averaging or frequency designed filters. Interestingly, even if it falls in the category of least-squares, polynomial interpolating filters, it is barely present in signal processing textbooks and rarely known from signal processing specialists [5]. The recent tutorial paper [6] might renew the interest of the community on this specific topic.

We similarly aim at bringing less known chemical sensing...
issues and references to the signal processing community. The paper, obviously far from exhaustive, provides a selection of key contributions to the field of analytical chemistry, whose modus operandi bears some similarities with those in signal processing, as described in Section 2, along with a description of the main types of data encountered and some of the needs in routine chemical analysis. Section 3 forms the core of the tutorial. It first reminds prior seminal works, followed by a decomposition of the main issues in SP-related topics. Some conclusions are provided in Section 4.

2. ANALYTICAL CHEMISTRY AND CHEMICAL SIGNALS: SCOPE AND NEEDS

2.1. Scope, definitions and field of interest

Analytical chemistry resorts to qualitative (detection: what compound is present?) and quantitative (estimation: how much of it?) analyses, to study physical and chemical properties of compounds and mixtures of natural or artificial materials. It relies on many chemico-physical interactions of atoms and molecules. Its specificity, with respect to routine chemical analysis, resides in the continuous improvement of analytical methods, experimental designs and chemometrics [7], i.e. “the art of extracting chemically relevant information from data produced in chemical experiments”. The latter essentially borrows methods from multivariate analysis and statistics.

A typical one-dimensional chemical signal is depicted in Figure 1-left. The observed amplitude at each location is generally considered as related to the proportion of a certain component. Since some uncertainty and variability exist, the proportion of an elementary component is generally distributed around an average location on the ordinal axis, so as to form a “peak”. Different peak parametric models [8, p. 97 sq.], for instance Gaussian or Lorentzian, have been developed to address different types of observed separation processes or analyzed components. The ordinal variable is not restricted to time or space. It represents a physico-chemical property which realizes the separation between elementary components, e.g. boiling point (temperature), migration (molecular mass), sensitivity to electro-magnetic fields (mass-to-charge ratio), etc. When considering additional instrumental drift and disturbances, the resulting chemical signal, often termed “spectrum”, is composed, at first order, of a linear combination of a sum of peaks of different amplitudes (more of less overlapped, or “co-eluted”), a slow-varying, sometimes monotone, baseline (or background) representing the lower limit of peak amplitude quantification, and noise. Hence, the most simple model is a linear mixture. Globally, those signals somehow differ from SP standards, as they rarely exhibit either jumps, step edges or oscillatory behavior. Hence, they deserve a set of analysis tools that drift away from usual derivative-based contour detectors, frequency-domain filters or multi-scale detectors. From a signal processing perspective, they often enjoy additional useful properties. For instance, when considering concentrations of analytes in a mixture, elementary spectra should be non-negative and have unit-concentration, taking into account the stoichiometric constants of balanced chemical equations (conservation of mass, charge or atoms). Recently, sparsity constraints on chemical species in a reaction have come into play.

Among the most common separation techniques, one distinguishes for instance: gas, liquid or ion chromatography; gel electrophoresis (2D-PAGE, 2D-DIGE); diode array detectors; ultraviolet (UV), Infrared (IR, near and far: NIR, FIR), Raman or Nuclear Magnetic Resonance (NMR) spectroscopy, X-ray diffraction or absorption, mass spectrometry (MS); ion-selective electrodes (ISE), etc. [9, 10]. Each of the aforementioned techniques deserves a closer inspection to devise chemically sound constraints, which may be employed in restoration problems. Analytical chemistry also resorts to image acquisition and processing, e.g. with scanning or transmission electron microscopy (SEM, TEM). We refer to [9, 11] for details.

2.2. On needs and trends in chemical signal analysis

Due to the need of routine chemical analysis and testing for quantities of data from high-throughput instruments, it is very important that processing techniques have a limited number of parameters, with semi-automatic or at least intuitive determination. Since chemical analyses are often made relative, the repeatability of signal processing is a very important feature. Despite the simplistic linear model described above, analytical methods possess different specificities and pose distinct challenges. As the studied chemical compounds steadily become more complex, their separation into elementary components is often difficult with a single technique. Even with instrumental resolution increases (e.g., the capability to output finer peaks, with respect to the full width at half maximum), the need for a separation based on two or more chemical properties (e.g. boiling point and electronic structure) has emerged. This has given birth to hyphenated techniques, combining some of the aforementioned techniques in pair, triple, etc. For instance, two-dimensional or comprehensive chromatography (GC×GC, Fig. 1-right) [12, 13] generates a two-dimensional signal with the above features. The resulting images are far different from the standard cartoon/texture model, and promote innovative methods [14]. Hyphenation may be extended to higher dimensions [15], providing an enhancement of resolution at the costs of more drastic data management problems.

Despite the variety of techniques, AC methods include common concepts of separation, detection, identification and
quantification (here of atomic, molecular, and ionic species). Such broad concepts are cores in signal and image processing as well, albethey with different meanings. Due to the close relationship between both disciplines, we choose to decompose chemical sensing issues in SP related fields, better suited to the target audience.

3. SIGNAL PROCESSING ISSUES, BY FIELDS

3.1. Historical mentions and early works

According to [4], “the dawn of the computer-controlled analytical instrument can be traced to” the Savitzky-Golay paper [3]. Gottschalk [16] relates analytical chemistry to information theory and considers the materials analyzed as more generic “systems”. The late professor J. B. Phillips, who fathered the comprehensive chromatography evoked in 2.1, considers that “It is no longer possible to understand the chemistry without considering signal processing [...] as a whole” [17]. This paper may have been overlooked, or at least underdiced. Other insightful considerations on signal processing interplay with analytical chemistry may be found in [8, 18].

3.2. An overview of major fields of application

Signal processing problems are recurrent in many stages of chemical sensing, from low-level applications, such as data acquisition and compression, to high-level tasks, such as feature extraction and identification. A brief look at the literature shows that, in addition to considering classical signal processing methods, research in chemical sensing has been also resorted to advanced techniques.

In the sequel, we aim at providing more information on some problems in chemical sensing that, in our opinion, could be of interest to the signal processing community. Of course, our survey is by no means exhaustive, due to the large number of works on this subject, and many relevant references are missing in our short survey. Still, we believe that the selected topics, in addition to the invited papers of the special session, may provide a nice view for researchers interested in joining this exciting area.

3.3. Acquisition and compression related problems

Data acquisition is a fundamental problem in chemistry. While classical techniques can be considered for some chemical data, there are several situations in which acquisition is a demanding step. This is the case in SEM images, for which reducing the acquisition time is a crucial need. In [19], the author proposes an approach based on smoothing that reduces the acquisition time by about one-third, yet ensuring a good signal-to-noise ratio. Efforts on acquisition methods have also been conducted for different chemical analyzes, involving for instance sampling of parametrized peaks [20], adapted to peak-like, non harmonic signals, or detector modulation [21], akin to a hybrid between multiplexing and time-frequency representations.

Another task that has been focus of many works is data compression. The need for compression of chemical data arises in techniques for which one must store large datasets that are used as reference. Among compression techniques, the ones based on the wavelet transform are the most adopted solutions in chemistry [22, 23] — these methods were applied in ion mobility spectrometry (IMS) sensors, MS, IR and NMR spectroscopy.

Finally, it worth mentioning the recent works on compressive sensing (CS) for chemical data. Briefly speaking, CS can be seen as a paradigm in which acquisition and compression are conducted at the same time. By exploiting the fact that the desired signal is sparse in a given domain — and by relying on a sort of random acquisition — CS methods are able to reconstruct the desired signal from a number of samples lower than the one predicted by the Nyquist-Shannon theorem [24].

In analytical chemistry, CS methods have been used, for instance, in NMR spectroscopy [25, 26, 27]. In these works, the application of CS methods provides relevant gains in terms of acquisition time.

3.4. Background estimation and filtering

In the basic signal formation model given in Section 2.1, two disturbances affect the desired signal considered as a linear combination of elementary peaks: an analytical background or baseline, accounting for slow-varying instrumental perturbations, and noise. Both should be remediated without harming peak shapes. In [28], the baseline is defined as “the portion of a detector record resulting from only eluant or carrier gas emerging from the column”. Broader definitions exits, encompassing more deterministic components such as temperature fluctuations or even small peaks that cannot be easily distinguished from a notional or arbitrary zero level [29], serving as a reference for peak properties quantification (height, area). Removing a slow-varying, potentially monotone, trend (Fig. 1-left) in a signal is generally an ill-posed posed problem, despite its apparent simplicity. The need for almost automatic methods is still present, after many attempts using least-square fits, wavelet preprocessing, robust or asymmetric error regression or factor analysis methods [30, 31, 32, 33, 34].

Given the importance of Savitzky and Golay filters, denoising and filtering have inspired many works in analytical chemistry, for instance [35, 36]. We refer to [11, vol. 2] for a broad overview of both background removal and filtering, while [37] provide a focus on the use of wavelet transforms.

3.5. Sensor array processing and signal separation

In analytical chemistry, a major issue is the lack of selectivity of some sensors. Large efforts have been undertaken to de-
velop new materials providing more selective sensors. However, despite the good results provided by this approach, it usually leads to expensive solutions. An interesting alternative to overcome this problem is to resort to sensing mechanisms based on diversity. For instance, instead of considering only one sensor, data can be acquired by an array composed of sensors responding differently to a given analyte. Then, the acquired data can be treated by signal processing methods with the aim of retrieving the desired information. This approach, which will be referred to as Smart Sensor Arrays (SSA), is usually adopted in the estimation of ionic concentration and gas analysis [38].

Since the sensors within the array are not necessarily selective in the SSA approach, the acquired signals correspond to mixtures of the desired signals, e.g., the temporal evolution of the activities of different ions. Therefore, the problem here is to retrieve the original signals (sources) from a set of mixtures of these sources. If one has access to a set of training (or calibration) data, this problem can be solved by multivariate regression methods. On the other hand, if only the mixtures are available, one can use blind source separation methods (BSS) [39]. An interesting feature of blind (or unsupervised) methods in chemistry is the possibility of avoiding (or simplifying) calibration stages, which are usually time demanding.

BSS methods have been providing interesting results in SSAs composed of potentiometric sensors used for measuring ionic activities. The main challenge in this application is due to the fact that potentiometric sensors are, as a rule, nonlinear devices [40]. As a result, the mixing processes are nonlinear, thus requiring advanced BSS methods. For instance, in [41], nonlinear BSS techniques based on Independent Component Analysis (ICA) were proposed for dealing with the mixing model that stems from potentiometric sensors. In [42], ICA methods were also applied to process the data acquired by arrays composed of tin oxide gas sensors, whose resulting mixing model is nonlinear, too. Despite the good results provided by ICA in these applications, it has a major limitation: ICA is based on the source statistical independence. Unfortunately, many chemical sources are clearly dependent. Consequently, other priors on sources have to be used, such as positivity or sparsity.

In order to deal with dependent sources in potentiometric arrays, [43] proposed a Bayesian BSS method. Since this approach is not based on source statistical independence and can easily take into account positivity constraints as statistical priors, it may provide good results even when the sources are highly correlated — this fact was illustrated in an actual dataset acquired by an array of ion-selective electrodes

Beyond SSA processing, source separation methods have been successfully used for solving many other issues in chemical engineering, e.g., separation of molecules in mass spectrometry, or spectral unmixing in hyperspectral imaging, etc.

3The dataset used in this work is publicly available [44].

3.6. Matching, identification, and learning

Higher-level data processing tasks, such as matching and pattern recognition, are also common in chemical applications. A typical example in this context is the problem of peak matching in chromatography. The output of this laboratory technique comprises a set of chromatograms, which in turn are composed of peaks. Ideally, different material samples containing the same components with different proportions should display peaks at the same positions on the ordinal axis. However, due to several experiments issues, one may observe shifts between these peaks, which may be potentially harmful for subsequent analyzes. Most of solutions to peak alignment are based on time warping (see [46] for a comparison between three matching methods).

Another example of chemical systems using high-level data processing can be found in the electronic noses [47] and tongues [48]. These systems are mainly adopted to perform automatic recognition of chemical elements. They are based on several signal processing stages. First, the signals acquired by the array are pre-processed with the aim of mitigating impairments such as sensor drift. Second, feature extraction is performed for (1) reducing the dimensionality of the data, and (2) extracting the relevant information that will feed the automatic classifier. Feature extraction methods used in electronic noses and tongues include principal component analysis (PCA) and self-organized maps (SOM) [49]. The last step is the classification, which can be carried out by artificial neural networks such as the multilayer perceptron, support vector machines, and Bayesian classifiers [47].

4. CONCLUSIONS

Since the aim of the present work is to introduce the special session entitled “Signal Processing for Chemical Sensing”, our major goal is to bring some relevant problems to the signal processing community. Of course, possible interaction between the two domains is very wide, and the paper cannot be exhaustive. Additional topics include e.g. statistical inference for simulating large scale modeler dynamics, or estimation of chemical concentration in bioreactors, etc. Conversely, in [50], authors suggest using chemical systems for doing signal processing tasks. Due to limited space, additional references are available [51].

5. REFERENCES
