Study on Sparsity-Aware Signal Processing for Chemicals Detection and Localization

March 2017

Yohei Kawaguchi

Study on Sparsity-Aware Signal Processing for Chemicals Detection and Localization

Graduate School of Systems and Information Engineering

University of Tsukuba

March 2017

Yohei Kawaguchi

Abstract

Beginning with the September 11 attacks of 2001, in recent years, a number or terrorist attacks using improvised explosives such as the November 2015 Paris attacks have occurred all over the world, and the threat of such terrorism has become a serious problem. In addition, the terrorism using toxic chemicals has still been a threat since the Tokyo subway sarin attack in 1995. For preventing terrorist attacks with these hazardous chemicals at train stations, airports, sports stadium, etc., systems for detecting the hazardous chemicals such as improvised explosives are required.

In terms of the size of monitoring areas, there are two types of the detection systems. One is a gate-type which detects chemicals at only an entrance of a high security area (e.g. "walkthrough portal explosives-detection system"), and high accuracy of detection is especially necessary. The other is a large-area-monitoring-type which detects chemicals all over the area. In the large-area-monitoring-type, a function for finding the location of chemicals is required because, if the location of chemicals is known, it will be possible to evacuate people and to capture a suspect. For localization, high speed is especially the most important. We study methods for making the gate-type and the large-area-monitoring-type into practical use through the improvement of the detection-accuracy and the localization-speed respectively.

When considering the improvement of the detection-accuracy for the gate-type and the localization-speed for the large-area-monitoring-type, we are inspired by "sparsity-aware signal processing". Although sparsity-aware signal processing has drawn interest in recent years, it is still rarely applied to chemical signal processing. "Sparsity" means the assumption that the observed signal consists of a small number of basis components. It is known that the sparsity assumption can improve the accuracy of estimation even though the size of the observed signal is small. In this study, methods of sparsity-aware signal

processing are proposed for improving the detection-accuracy and the localization-speed. First, for the gate-type, to improve the detection-accuracy, we improve the separation performance by using the sparsity assumption. Second, for the large-area-monitoring-type, to speed-up localization, we propose a "compressive sensing"-based approach which takes the mixed air from a combination of multiple ducts at each time-frame unlike time-division-sampling, and it switches the combination of active ducts temporally, and it estimates the location of chemicals from time series of observations. The sparsity assumption makes it possible to localize chemicals from a small number of observations.

This dissertation consists of two parts. The first part corresponds to chemicals detection for the gate-type. The second part corresponds to chemicals localization for the large-area-monitoring-type. In the first part, for a "walkthrough portal explosives-detection system", a signal-separation-method using a sparsity assumption is proposed to improve the detection-accuracy. In addition, an independent-component-analysis (ICA)-based acceleration of the signal-separation-method is proposed. Furthermore, for reducing the uncertainty to improve the robustness, a signal-separation-method using an attenuation model is proposed. In the second part, for the large-area-monitoring-type, a compressive sensing-based approach using a sparsity assumption is proposed to speed up localization of chemicals, and, especially, to achieve the robustness to the difference of the number of the positions of chemicals, adaptive Boolean compressive sensing is proposed. In addition, to improve the robustness to estimation errors, an extension of the adaptive Boolean compressive sensing into a "multi-armed bandit" algorithm is proposed. Furthermore, to improve the robustness to change of the location of chemicals, a combination of change-point detection and the adaptive Boolean compressive sensing is proposed.

Acknowledgements

本論文をまとめるにあたり，多くのご支援ならびにご指導を頂きましたことを，心より感謝申し上げます。

指導教員の牧野昭二教授には，数多くの貴重なご指導ならびにご鞭撻を賜りましたことを，深く感謝申し上げます。山田武志准教授には，数多くのご助言を賜りましたことを深く感謝申し上げます。審査委員の福井和広教授，佐久間淳教授，国立情報学研究所の小野順貴准教授には，博士論文をまとめる上での貴重なご助言を頂きました。深く感謝致します。マルチメディア研究室に所属されていた学生各位におかれましては，ゼミでの議論を通じて数多くの貴重なご意見を頂き，深く感謝しております。

株式会社日立製作所の戸上真人氏には，入社以来，研究・会社生活においてご指導ならびにご支援を頂きました。深く感謝致します。額賀信尾氏，東京工科大学の大淵康成教授（当時株式会社日立製作所），永松健司氏には，音声・音響信号処理ユニットの上司として，多大なるご支援と貴重なご助言を頂きました。影広達彦氏，廣池敦氏は，私が音響以外の信号処理に携わるきっかけを与えて下さいました。これが本研究につながる転機となりました。深く感謝致します。永野久志氏には，爆発物探知システムの研究者としての，多大なるご支援とご指導を頂きました。永野氏との議論の結果が本研究に多大な影響を与えています。心より感謝申し上げます。坂入実氏，橋本雄一郎氏，高田安章氏，杉山益之氏，熊野峻氏，西村和茂氏には，爆発物検知や質量分析について数多くのご指導・ご助言を頂きました。研究開発や実証実験などで共に携わらせて頂いたことに大変感謝しています。Shubhranshu Barnwal 氏，長辰彦氏には，当時インターン生として本研究の遂行に対して多大な協力を頂きました。深く感謝致します。金子明人氏には，質量分析における信号処理・パターン認識技術について数多くのご指導・ご助言を頂き，心より感謝しています。故 天野昭雄氏，遠藤隆氏，鴨志田亮太氏，高橋昌史氏，山本正明氏，高島遼一氏，池下林太郎氏には，過去または現在において，同じチームで研究・会社生活を支えて頂きました。深く感謝致します。また，技術開発や製品化など

でご協力頂きました株式会社日立製作所ならびにグループ会社の関係各位に深く感謝致します。

最後に，今日に至るまで私を支えてくれた妻の恵子，ならびに，故郷の両親に感謝致します。

Contents

List of Figures

List of Tables

Chapter 1

Introduction

1.1 Background

Beginning with the September 11 attacks of 2001, in recent years, a number or terrorist attacks using an improvised explosive device (IED) such as the 2005 London bombings, the 2013 Boston Marathon bombing, the November 2015 Paris attacks, 2016 Brussels bombings, etc. have occurred all over the world, and the threat of such terrorism has become a serious problem. IEDs can be created from consumer goods, and information about creating IEDs can be easily accessed via the Internet. In Japan, even a high-school student attacked a train station with an IED in 2002 [3]. In IEDs, explosive substances such as triacetone triperoxide (TATP) are used. TATP has a higher vapor pressure than military explosives [3], and gases emitted from IEDs or traces attached to people or luggage can be detected in principle. Detecting gases from IEDs makes it possible to prevent IED attacks, and countries such as the United States Government states the necessity to counter IED attacks [4]. In addition, the terrorism using toxic chemicals has still been a threat since the Tokyo subway sarin attack in 1995. For such chemical terrorisms, rapid evacuation is important to reduce casualties. For evacuation, it is necessary to detect toxic chemicals. However, in most public areas, detectors for toxic chemicals still have not been applied yet, and countermeasures still have not been taken. For preventing terrorist attacks with these hazardous chemicals at train stations, airports, sports stadium, etc., systems for detecting the hazardous chemicals such as improvised explosives and toxic chemicals are required.

In terms of the size of monitoring areas, there are two types of the detection systems. One is a gate-type which detects chemicals at only an entrance of a high security area (e.g. "walkthrough portal explosives-detection system" [3]), and high accuracy of detection is especially necessary. The other is a large-area-monitoring-type which detects chemicals all over the area. In the large-area-monitoring-type, a function for finding the location of chemicals is required because, if the location of chemicals is known, it will be possible to evacuate people and to capture a suspect. For localization, high speed is especially the most important. We study methods for making the gate-type and the large-area-monitoring-type into practical use through the improvement of the detection-accuracy and the localization-speed respectively.

1.2 Purpose of dissertation

When considering the improvement of the detection-accuracy for the gate-type and the localization-speed for the large-area-monitoring-type, we are inspired by "sparsity-aware signal processing". Although sparsity-aware signal processing has drawn interest in recent years, it is still rarely applied to chemical signal processing. "Sparsity" means the assumption that the observed signal consists of a small number of basis components. It is known that the sparsity assumption can improve the accuracy of estimation even though the size of the observed signal is small. In this study, methods of sparsity-aware signal processing are proposed for improving the detection-accuracy and the localization-speed. First, for the gate-type, to improve the detection-accuracy, we improve the separation performance by using the sparsity assumption. Second, for the large-area-monitoring-type, to speed-up localization, we propose a "compressive sensing"-based approach which takes the mixed air from a combination of multiple ducts at each time-frame unlike time-division-sampling, and it switches the combination of active ducts temporally, and it estimates the location of chemicals from time series of observations. The sparsity assumption makes it possible to localize chemicals from a small number of observations.

Figure 1.1 shows the position of the target of this dissertation. For practical use, both the high accuracy per measurement time and the low sensing-cost must be satisfied. However, traditional approaches could not have satisfied both of them. In this dissertation, we apply sparsity-awareness to solve the problem of traditional approaches.

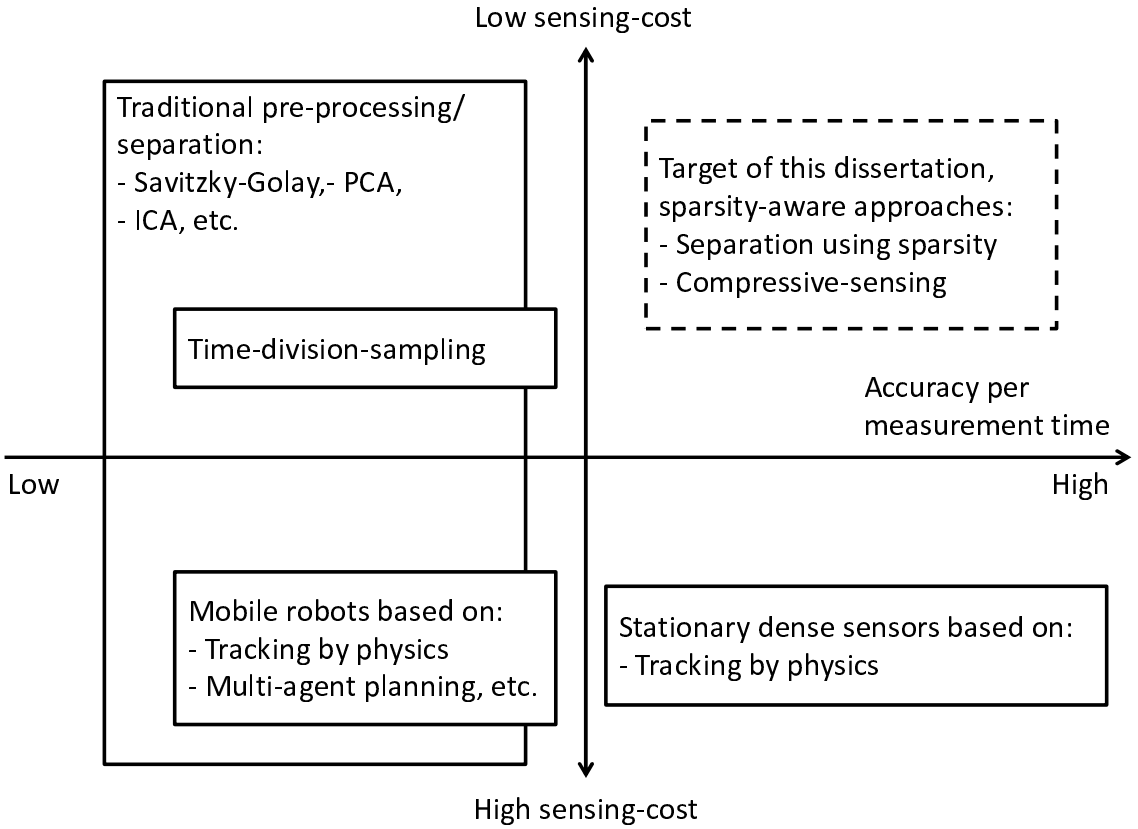


Figure 1.1: Position of the target of this dissertation.

1.3 Overview of dissertation

In Chapter 2, the preliminaries are introduced: first, sensor devices for chemicals detection are explained; second, conventional approaches to signal processing for chemicals detection are explained; third, conventional approaches to chemicals localization are explained; finally, basic theories of sparsity-aware signal processing is explained. The body of this dissertation consists of two parts. The first part (from Chapter 3 to Chapter 5) corresponds to chemicals detection for the gate-type. The second part (from Chapter 6 to Chapter 8) corresponds to chemicals localization for the large-area-monitoring-type. In Chapter 3, for a "walkthrough portal explosives-detection system" [3], a signal-separation-method using a sparsity assumption is proposed to improve the detection-accuracy. In Chapter 4, an independent-component-analysis (ICA)-based ac-

celeration of the signal-separation-method is proposed. In Chapter 5, for reducing the uncertainty to improve the robustness, a signal-separation-method using an attenuation model is proposed. In Chapter 6, for the large-area-monitoring-type, a compressive sensing-based approach using a sparsity assumption is proposed to speed up localization of chemicals, and, especially, to achieve the robustness to the difference of the number of the positions of chemicals, adaptive Boolean compressive sensing is proposed. In Chapter 7, to improve the robustness to estimation errors, an extension of the adaptive Boolean compressive sensing into a "multi-armed bandit" algorithm is proposed. In Chapter 8, to improve the robustness to change of the location of chemicals, a combination of change-point detection and the adaptive Boolean compressive sensing is proposed. Finally, in Chapter 9, the conclusions of this dissertation are shown.

Chapter 2

Preliminaries

2.1 Sensor device

2.1.1 Bulk detection and trace detection

Many kinds of sensors for detecting substances have been developed. There are two types of the sensors, i.e. one is bulk detection, and the other is trace detection [5][6].

Bulk detection includes X-ray imaging, nuclear quadrupole resonance (NQR), and neutron techniques. Bulk detection is used for determining the shape, size, density, and elemental composition of suspicious objects in luggage, letters, packages, etc. Even if the container of a suspicious object is not opened, bulk detection is possible. Therefore, bulk detection is effective for the hazardous materials that do not leak gases or particles such as military explosives. Among the types of bulk detection, X-ray imaging is mostly used for baggage screening at security-checkpoints of most airports. However, bulk detection cannot detect a small amount of substances [5], i.e. if it is divided into small portions, detection is impossible. Also, bulk detection cannot easily identify the type of substances [5][6].

Trace detection is a chemical-analysis method including "mass spectrometry" (MS), etc. It detects the existence of gases or particles of substances directly, and a small amount of substances can be detected. Trace detection is fit to IEDs because gases or particles of IEDs can be detected by it. Trace detection identifies the type of substances, and its selectivity is higher than that of bulk detection. Therefore, the false-alarm-rate of trace detection is less than that of bulk detection. However, it cannot determine the

amount of substances in a suspicious object. As above, the features of bulk detection and trace detection are different. Thus, it is recommended to combine bulk detection and trace detection to improve security [6]. In this study, we focus on trace detection because trace detection is fit to IEDs.

2.1.2 Types of trace detection

Trace detection is classified into three categories: biosensors, optical sensors, electronic/-chemical sensors [7]. In this section, typical technologies of trace detection are explained.

Biosensors include canines (dogs), rats, and bees. Among them, canines are the most typical, and trained canines have been used successfully for mine clearance since World War II [8][9]. Even in the present day, canines are still the most standard for detecting explosives and drugs. A major advantage of canines is that they can move their noses to near a suspicious object actively. However, it is not clearly known what kind of vapor signatures canines rely on [5], and interpretability is low. A research has shown that canines do not rely on vapor signatures from the pure compound, but a combination of solvents, residual substances from the manufacturing process, and degradation by-products [5][8][9]. In addition, canines have the disadvantage that both the costs of bringing up them and the labor costs for leading them are expensive[5].

Optical sensors include "infrared spectroscopy", "terahertz spectroscopy", and "Raman spectroscopy". Based on an interaction between light and matter, these approaches identify the chemical composition of air [7]. These approaches have high selectivity for monitoring from remote distances, whereas they provide good results only in the case of long-path absorption [7]. They have not been widely used for applications such as explosives detection.

"Chemiluminescence" (CL) is one of electronic/chemical-sensing-approaches. CL detects a characteristic emission of radiation from a product by an exothermic reaction. Hexamethylenetriperoxidediamine (HMTD), which is one of explosive substances, can be detected by CL, whereas TATP can not be detected [5]. A major disadvantage of CL is the lack of selectivity [7].

Approaches based on "surface acoustic wave" (SAW) are one class of electronic-/chemical-sensing. SAW sensors are microelectromechanical systems (MEMS) and are

called “electric noses”. SAW sensors rely on the modulation of SAW that occurs when the SAW passes through a film polymer absorbing a gas. Different kinds of film polymers can selectively absorb different gases, and individual sensors can selectively detect different target substances. SAW sensors are thus commonly used as an array of multiple SAW sensors with different polymers, and many kinds of substances can be detected by the array on one chip. Although there are examples of military applications that SAW sensors are used for detecting explosive, biological, and chemical weapons[10], SAW is still in a phase of study.

“Ion mobility spectrometry” (IMS) is one of electronic/chemical-sensing-approaches. IMS ionizes the chemicals and sorts the ions according to their mobility. The detector of IMS counts the number of ions of different mobility, and produces a signal called an “ion mobility spectrum”, which is a record of ions as a function of mobility. Major advantages of IMS are their high response-speeds, their small sizes, and their low cost [5], and there are a number of IMS products. A major disadvantage of IMS is low selectivity, in other words, the resolution of spectra is low, and the false-alarm-rate tends to be high.

“Mass spectrometry” (MS) is one of electronic/chemical-sensing-approaches. MS ionizes the chemicals and sorts the ions according to their mass-to-charge ratios. Sorting is based on the principle that ions of different mass-to-charge ratios have different velocities or resonance frequency in a magnetic field. The detector of MS counts the number of ions of different mass-to-charge ratios, and produces a signal called a “mass spectrum”, which is a record of ions as a function of mass-to-charge ratios. MS has high sensitivity, and it also has higher selectivity than IMS. Historically, MS could not have measured complex mixtures, and it must have been combined with “gas chromatography” (GC), which separates a mixture into different substances according to the strength of adsorption. However, the combination is called GC-MS, and it has the disadvantage that it takes a long time to achieve a signal [5]. To solve this problem, “atmospheric pressure chemical ionization” (APCI) has been proposed [3]. APCI can extract a target-substance roughly by ionization instead of GC, and the combination of APCI and MS can achieve a signal within 1 to 2 seconds after a gas is drawn [5], i.e. the response-speed of the APCI-MS is fast. Historically, two major disadvantages of MS have been its size and cost. However, major progress has been made in downsizing MS devices that reduces

costs [8]. In addition, analysis and detection of explosives by MS has been well studied [6][11]. As above, MS is the most promising approach from various perspectives.

IMS and MS can be used for our system because they can detect explosive substances of IEDs such as TATP and have high response-speed and high sensitivity. In addition, MS has the advantage that it has higher selectivity than IMS, and the spectra of MS have higher resolution than those of IMS. Therefore, in this study, we assume that MS is used for our systems, and we study signal processing methods on the premise of MS. Actually, the spectra of IMS are similar to those of MS, and it can be expected that the results of this study is applicable for IMS.

2.2 Conventional signal processing for chemicals detection

Signal processing for chemicals detection is a part of the area called "chemometrics". Chemometrics is the area of study on extracting information from chemical data, and chemometrics is applied for solving problems of chemistry, biology, medicine, etc. A general procedure of chemometrics is divided into three parts: (1) pre-processing, (2) classification, and (3) quantification. In this section, these techniques are explained.

2.2.1 Pre-processing

Pre-processing includes baseline-removal/noise-reduction/signal-separation and peak-alignment. In most cases, an observed signal is contaminated by noise, and a target component needs to be extracted from the contaminated signal. One of the main problems for chemicals detection is extraction of the target component from the contaminated signal, and baseline-removal/noise-reduction/signal-separation is necessary in most applications.

Baseline-removal removes a "baseline", which is defined as low frequency noise with smoothness. Most baseline-removal approaches consist of two steps: (1) the baseline is estimated from an observed signal based on the assumption of the smoothness, and (2) the fitted function is subtracted from the observed signal. In many cases, for estimation of the baseline, a Savitzky-Golay filter [12] is customarily used (e.g. [13]), and it can achieve

success. Although many researches on baseline-removal have been done [13][14][15], the need for almost automatic methods is still present [16]. However, as for our system, it is not necessary to improve the performance of baseline-removal.

Noise-reduction is defined as reducing high frequency noise in particular here. A smoothing filter is used by noise-reduction in most cases. For smoothing, approaches such as Savitzky-Golay filters and wavelet techniques are used [17][18], and these conventional approaches can achieve success. This topic has been well established, and it is not necessary to develop a new technique.

As above, in the case that the noise has only a higher or lower frequency than the target component, it is comparatively easy to extract the target component. In contrast, if other substances exist with the target substance simultaneously, the target component and the interferences corresponding to the other substances can not be distinguished by frequency, and extraction of the target component is difficult. In this case, signal-separation is applied. Signal-separation is to estimate individual source signals from the mixed signal. In most cases, the observation is modeled by linear mixing:

$$\boldsymbol{X} = \boldsymbol{S}\boldsymbol{C}, \tag{2.1}$$

where $\boldsymbol{X}$ is the observed mixed signal, which is a $M \times T$ matrix, $\boldsymbol{S}$ is the spectral basis components, which is a $M \times K$ matrix, $\boldsymbol{C}$ is the temporal activations, which is a $K \times T$ matrix, M is the number of spectral indices, T is the number of temporal indices, and K is the number of substances. Signal-separation can be formulated as estimation of $\boldsymbol{C}$ from $\boldsymbol{X}$. If all the substances that may be observed are known and the spectral basis components do not change, $\boldsymbol{S}$ can be obtained by using a set of training data in advance. In this case, estimation of $\boldsymbol{C}$ gives a closed-form solution:

$$\hat{\boldsymbol{C}} = \boldsymbol{S}^{+}\boldsymbol{X}. \tag{2.2}$$

Actually, in real cases, not all the substances can be known, i.e. an unknown substance may be observed. In addition, the spectral basis components may change dependently on the condition of a sensor device. Thus, "blind-source-separation", which is applicable to the case that both $\boldsymbol{S}$ and $\boldsymbol{C}$ are unknown, is used. Especially, principal component analysis (PCA) and independent component analysis (ICA) are applied in most researches [17][19]. However these approaches are not always suitable for mass spectrometry. PCA

relies on the orthogonality of the spectral basis components. In mass spectrometry, the spectral basis components of different substances tend to be similar, so the orthogonality of the spectral basis components does not hold. ICA relies on the independence of the temporal activations, so ICA needs a certain time when the temporal activations of different substances change independently. If a sufficient time when the temporal activations change independently cannot be observed, ICA may fail, that is, the independence is not a sufficient constraint.

Peak-alignment is adjusting the spectral axis of the spectrum. Peak-alignment is necessary because the spectral axis tends to be shifted or distorted non-linearly. The shift and distortion are caused by complex chemical mixtures and sensing conditions [20]. For measurement of chemicals with many peaks such as proteomics, peak-alignment can be solved by dynamic time warping (DTW), correlation optimized warping (COW), and similar methods [21]. For our application, chemicals that may be observed have a small number of peaks, and methods like DTW can not work. Conversely, required accuracy of the spectral axis is not high relatively. Thus, for applications such as our systems, the peak-alignment problem can be solved by setting the tolerance of peak positions to a large value instead of DTW.

2.2.2 Classification

Classification includes multiclass-classification as well as detection. Multiclass-classification is important in a typical case of chemometrics, analysis of foods [22]. Multiclass-classification can be applied for investigating the characteristics of foods such as which ingredient is used, where foods are from, etc. Many kinds of pattern-recognition algorithms are applied for chemometrics[22]. Linear discriminant analysis (LDA) is one of the most simple algorithms that assumes that all the classes are Gaussian distributions and the covariance matrices are equal. Quadratic discriminant analysis (QDA) may be used when covariance matrices are different. Partial Least Squares Discriminant Analysis (PLS-DA) is customarily used in chemometrics [23]. It is also known that support vector machines (SVMs) can be applied to many fields [22].

Detection is to determine whether each target substance exists or not, and it is the main function of our systems. In many applications other than chemicals detection for

security, patterns of anomalies are unknown in advance, whereas normal patterns can be known. In these cases, approaches of anomaly-detection/outlier-detection such as the Mahalanobis-distance, Hotelling's T^2, D-Static, the one-class SVM, etc. are widely applied [22]. Anomaly-detection/outlier-detection determines whether an input signal is derived from normal conditions defined by the given normal patterns. In contrast, in chemicals detection for security, spectral patterns of target substances (explosives, toxic chemicals, etc.) can be known in advance, whereas spectral patterns of interferences such as cosmetics are unknown. Therefore, for applications such as our systems, template matching is the most standard approach [24][25], i.e. the input spectrum is matched with known templates in database, a similarity measure is calculated, and an alert will be displayed if the similarity measure is higher than a threshold. As similarity measures, the Euclidean distance, the inner product, the cosine similarity, the Pearson correlation coefficient, etc. are used [22].

2.2.3 Quantification

Quantification is estimation of the amount of a substance from chemical data. Quantification follows a supervised-learning-manner; i.e. in advance, signals need to be recorded on different conditions that a known amount of the pure target substance exists, and a regression function is trained from the combination of the recorded signals and the amount of the substance. Based on the trained regression function, for a new observed signal, the amount of the substance can be estimated. For example, the partial-least-squares (PLS) method is widely used as a regression algorithm similarly to classification. In some cases, according to sensors such as IMS, the variance of the feature value is too large, and quantification is difficult. For our system, accurate quantification is not necessary because an accurately-estimated amount can not be used for capturing a suspect or evacuating people. In this study, we do not treat quantification as a main topic.

2.3 Conventional approaches for chemicals localization

There are two main approaches of chemicals localization. One is to search for chemicals by using an autonomous mobile robot equipped with a sensor [26], and the other is to use stationary sensors to localize a chemical source [27][28].

Approaches of mobile robots have been studied intensively in the research community [26][29][30][31][32][33][34]. The task of mobile robots consists of several tough sub-tasks:

- detecting a target substance in a noisy environment;
- making a robot follow the trail of the substance from the source;
- integrating incomplete information achieved by multiple robots;
- making a strategy such that multiple robots can move cooperatively to achieve as much information as possible.

Also for the hardware, difficult problems must be solved; the size of sensors must be small enough to be equipped in robots; batteries must be designed to provide a long lasting power; everyday maintenance for robots must be done. Although a number of researches on mobile robots have been done, there is rarely an example of practical application. The main reasons are that the cost of robots in itself is high, that the life of batteries is short, and that the cost of maintenance is high [28].

For approaches of stationary sensors, a large number of sensors must be deployed densely arranged [27][28]. On assumption that many sensors can be deployed, Cao *et al.* proposed a localization algorithm based on a distributed least-squares estimation [27]. Nofsinger proposed a localization algorithm based on the multiple hypothesis tracking (MHT) [28]. A greater number of sensor-nodes are needed than the number of sensors for the case of autonomous robots, so the cost of sensors must be much lower, the power consumption must be much lower, and the required maintenance must be much less [28]. However, such sensors can detect only the limited kinds of substances, and they can not be applied for our application.

For both of the above approaches, an enormous sensing-cost is required, so there are few examples of practical use. However, there is another approach for reducing the

sensing cost [35]. This approach uses only a single detector and multiple air-intake-ducts corresponding to the positions all over the area, and all the ducts are connected with the detector. The system can switch the ON/OFF of intake of individual ducts, and at each time-frame, only one duct is selected and is set to ON. The system sequentially selects a duct and takes air from only the selected duct corresponding to the time-frame. Because there is a one-to-one-correspondence between the ducts and the time-frames, The position of the duct corresponding to the time-frame when a target substance is detected can be determined as the position where the substance exists. Hereafter, we call this approach "time-division-sampling". However, time-division-sampling spends a long time to test all the positions, and time-division-sampling is not fit for applications such as evacuation guidance.

2.4 Basic theory of sparsity-aware signal processing

In recent years, sparsity has been a very hot topic of signal processing. The sparsity means the assumption that the source signal is a combination of a small number of basis components in a transform domain. Using the assumption of the sparsity can improve the accuracy and robustness of estimation, and it can also reduce the required number of observations. This approach is called sparsity-aware signal processing. The sparsity-aware signal processing is explained in this section.

We consider a matrix $\boldsymbol{A} \in \mathbb{R}^{n \times m}$ $(n < m)$, and define an underdetermined linear observation system $\boldsymbol{y} = \boldsymbol{A}\boldsymbol{x}$, where $\boldsymbol{A}$ is given, $\boldsymbol{y} \in \mathbb{R}^n$ is an observed vector, and $\boldsymbol{x} \in \mathbb{R}^m$ is an unknown vector. One of the main tasks of signal processing is to achieve $\boldsymbol{x}$ from $\boldsymbol{A}$ and $\boldsymbol{y}$. $\boldsymbol{A}$ and $\boldsymbol{y}$ can be regarded as the degradation process and the degraded signal respectively, and the goal is to reconstruct the source signal $\boldsymbol{x}$. The same problem appears in various fields such as super-resolution [36], image inpainting [37][38], audio inpainting [39], signal compression [40], noise reduction [41], etc. However, there are an infinite number of solutions to the underdetermined system. To find the best one in a sense from an infinite number of solutions, an additional constraint is needed.

A popular way to add a constraint is "regularization". We define the optimization problem:

$$\min_{\boldsymbol{x}} \|\boldsymbol{y} - \boldsymbol{A}\boldsymbol{x}\|_2^2 + J(\boldsymbol{x}), \tag{2.3}$$

where the first term is a fit measure, and $J(\boldsymbol{x})$ is a penalty term on the complexity of $\boldsymbol{x}$. $J(\boldsymbol{x})$ is called a regularization term. In many cases, $J(\boldsymbol{x})$ is a norm of $\boldsymbol{x}$, and a solution with a smaller norm is chosen. The most popular form of $J(\boldsymbol{x})$ is the squared Euclidean norm $\lambda\|\boldsymbol{x}\|_2^2$, and (2.3) is converted to:

$$\min_{\boldsymbol{x}} \|\boldsymbol{y} - \boldsymbol{A}\boldsymbol{x}\|_2^2 + \lambda\|\boldsymbol{x}\|_2^2. \tag{2.4}$$

Equation (2.4) gives the closed-form solution:

$$\hat{\boldsymbol{x}} = \left(\boldsymbol{A}^T\boldsymbol{A} + \lambda\boldsymbol{I}\right)^{-1}\boldsymbol{A}^T\boldsymbol{y}. \tag{2.5}$$

This regularization is well known as L2 regularization [42], Tikhonov regularization [43], or ridge regression [44]. L2 regularization is widely used in various fields, and this is due to its simplicity of the closed-form-algorithm. However, L2 regularization is not always the best way.

If $J(\boldsymbol{x})$ is the ℓ_1-norm $\lambda\|\boldsymbol{x}\|_1$, (2.3) is converted to:

$$\min_{\boldsymbol{x}} \|\boldsymbol{y} - \boldsymbol{A}\boldsymbol{x}\|_2^2 + \lambda\|\boldsymbol{x}\|_1. \tag{2.6}$$

This is called L1 regularization [42] or basis pursuit denoising [45], which can be solved by convex optimization. L1 regularization has a tendency to prefer sparse solutions, i.e. the shortest solution $\boldsymbol{x}$ in ℓ_1-norm tends to have a fewer number of non-zero elements than the shortest in ℓ_2-norm. Sparsity appears in various fields such as audio [46], medical imaging [47], natural images [48], seismic data [49], biological data [50], etc. In many cases, sparse solutions are desired.

A more general case is Lp regularization, which uses ℓ_p-norm ($0 \leq p \leq 1$), has a tendency to prefer sparser solutions. The special case that $J(\boldsymbol{x})$ is the ℓ_0-norm with $\lambda \to \infty$ is the following form:

$$\|\boldsymbol{x}\|_0 \quad \text{s.t.} \quad \boldsymbol{y} = \boldsymbol{A}\boldsymbol{x}, \tag{2.7}$$

where $\|\boldsymbol{x}\|_0$ means the number of non-zero elements of $\boldsymbol{x}$. In a narrow sense, this form is called "compressive sensing" or "compressed sensing". The compressive sensing is a problem of combinatorial search, and the computational complexity of the exhaustive search is exponential in the dimension of $\boldsymbol{x}$, i.e. this problem is NP-hard. Actually, it has been proven that, under some conditions, the unique solution of the compressive

sensing can be calculated by approximation algorithms. For example, if a solution $\boldsymbol{x}$ exists obeying

$$\|\boldsymbol{x}\|_0 < \frac{1}{2}\left(1 + \frac{1}{\mu(\boldsymbol{A})}\right), \tag{2.8}$$

a greedy algorithm called "orthogonal matching pursuit" is guaranteed to find the unique solution exactly, where $\mu(\boldsymbol{A})$ is the mutual coherence of $\boldsymbol{A}$ [51]. In addition, if a solution $\boldsymbol{x}$ exists obeying (2.7), the solution is equal to the unique solution of L1 minimization [52][53]:

$$\|\boldsymbol{x}\|_1 \quad \text{s.t.} \quad \boldsymbol{y} = \boldsymbol{A}\boldsymbol{x}, \tag{2.9}$$

which can be solved by linear programming.

As above, for the undetermined linear system, many researches on guarantees have been done. Actually, it is known that the sparsity-awareness is useful for not only the undetermined linear system but also other problems, such as blind-source-separation [54][55][1], a Boolean system called "group-testing" [2], non-linear systems [56][57], etc. In the following chapters, we study methods based on sparsity-aware signal processing for improving the detection-accuracy and the localization-speed.

Chapter 3

Separation of mass spectra based on probabilistic latent component analysis for explosives detection

3.1 Introduction

The threat of improvised explosive devices (IEDs) has recently become a serious problem for all countries because the recipes for making them are freely available on the Internet. It is therefore necessary to develop technologies for detecting IEDs. Aiming to prevent terrorist attacks with IEDs at train stations, airports, sports stadiums, shopping malls and such places, a "walkthrough portal explosives-detection system" was previously developed. The system consists of a high-throughput ticket gate-type vapor sampler, a high-sensitivity atmospheric-pressure chemical-ionization source, a high-selectivity linear ion-trap mass spectrometer, and an ion-counting detector [3]. In the first step of the detection process, the vapor sampler draws the vapor emitted from traces of substances attached to the body, clothes, and luggage of the person being checked. Next, the ionization source ionizes the molecules in the drawn vapor, and the mass spectrometer separates the ions according to their mass-to-charge ratios (m/z) by means of an electromagnetic field. The detector translates the number of separated ions striking the detector to intensity for each m/z value. The series of the intensities is called the "mass spectrum", and the system observes a time series of the mass spectra continuously.

Hereafter, the time series of the mass spectra is called the "mass spectrogram". Finally, the system detects explosive substances based on whether particular spectral patterns appear in the mass spectrogram.

Signals corresponding to cosmetics, paint, and dust as well as those corresponding to explosive substances are mixed in the mass spectrogram observed in real environments. It is therefore necessary to separate the mass spectrogram into individual substances. The system knows neither the spectral patterns corresponding to individual substances nor the times when individual substances are observed in advance. In other words, the problem known as "blind-source-separation" (BSS) must be solved. In the field of mass spectrometry, algorithms for blind-source-separation, such as principal-component analysis (PCA) [58] and independent-component analysis (ICA) [59][60], are widely used. Multiple substances have spectral peaks at the same m/z value in the mass spectrogram; that is, orthogonality does not hold in general. The separation performance of PCA is degraded because it imposes orthogonality. ICA does not impose orthogonality but statistical independence. ICA, however, fails in many cases because uncertainty in mass spectra separation can not be solved only by statistical independence. Recently, new approaches based on "non-negative matrix factorization" (NMF) [61][62] have been applied to mass spectrometry [63][64]. To solve problems of uncertainty in mass spectra separation, NMF makes use of the fact that all the signals corresponding to individual substances are non-negative. However, there is no study on explosives detection using NMF-based separation of a mass spectrogram.

In regard to NMF, the main approach to further reduce uncertainty is to introduce sparsity constraints. A number of NMF algorithms with sparsity constraints have been proposed [65][1][66][67]. It has been shown that L1 regularization [65] and L1/L2 regularization [1][66] improve the separation performance in audio and image processing by using the constraint that the activation matrix is sparse or that the basis matrix is sparse. Moreover, "probabilistic latent-component analysis" (PLCA) [68][69][70][71], which is a statistical formulation of NMF with the Kullback-Leibler (KL) divergence, has been proposed and applied to musical-signal processing. It is known that sparsity assumptions can be applied to PLCA [70][71]. However, there is no study on whether sparsity assumptions are effective for explosives detection.

In this study, a new method for separating a mass spectrogram for explosives detection-based on PLCA with a sparsity constraint-is proposed. Similarly to NMF, PLCA imposes the constraint of non-negativity, but does not impose orthogonality; therefore, it is applicable to mass spectra. In addition, the separation performance is improved by using a sparsity constraint. The results of experiments on separation with mass-spectra data obtained in a real station demonstrate that the proposed method outperforms existing methods. They also show that the sparsity assumptions improve separation performance. It is also shown that the algorithm accelerated by a graphical processing unit (GPU) can work in real time.

3.2 Problem statement

The input signal is a time series of mass spectra, called a mass spectrogram $x(t, m)$, where t is the time index, and m is the index corresponding to m/z in a mass spectrum. T is the number of the time indices, i.e. $1 \leq t \leq T$. M is the number of the indices corresponding to m/z, i.e. $1 \leq m \leq M$. The set of substances in the air that can be observed by the system consists of explosive substances, interference substances, and a chemical background. Interference substances are defined as those of which peaks rise and fall within seconds, such as cosmetics of the person being checked, and the chemical background is defined as those of which peaks are stationary over a period of minutes, such as paint on building walls, etc. The number of substances that can be observed is given as K. It is assumed that each substance has a particular spectral pattern corresponding to the ions derived from the substance, which is called a spectral basis component. This means that K is also the number of spectral basis components. It is also assumed that spectral basis components are time-invariant over minutes. $x(t, m)$ may be modeled as a linear combination of spectral basis components as follows,

$$x(t, m) = \sum_{k=1}^{K} c(k|t)s(m|k), \tag{3.1}$$

where k $(1 \leq k \leq K)$ is the index of substances, $c(k|t)$ is the intensity corresponding to the amount of the k-th substance at time-point t, and $s(m|k)$ is the m-th element of the spectral basis component corresponding to the k-th substance. $c(k|t)$ is called a temporal activation.

Unknown variables $c(k|t)$ and $s(m|k)$ are estimated from known variables $x(t,m)$. This estimation is a "blind-source-separation" problem. In addition, the following conditions concerning the explosives-detection system are set. First, $s(m|k)$ is non-negative for all m/z values and substances because the mass spectrum represents the number of ions for each m/z value; second, $c(k|t)$ is non-negative for all substances and times because $c(k|t)$ represents the amount of each substance at the time-point t; third, orthogonality among the spectral basis components cannot be assumed because different substances in real environments may have ions with the same m/z value.

3.3 Proposed method

3.3.1 PLCA model

The PLCA model [68][69] consists of a pair of probabilistic trials:

1. At each time t, the probability that the k-th substance is selected is $P(k|t)$.

2. For the selected substance k, the m-th index corresponding to m/z is selected from probability distribution $P(m|k)$, and a positive constant value, Δ, is added to the spectral bin (t,m) corresponding to $x(t,m)$.

It is assumed that $x(t,m)$ is generated by the iteration of this selection-addition process for each time-point t. It is therefore possible to formulate the probability distribution that $x(t,m)$ is generated as follows:

$$P(x(t,m)\forall t,m) = \prod_{t=1}^{T}\prod_{m=1}^{M}\left\{\sum_{k=1}^{K}P(k|t)P(m|k)\right\}^{x(t,m)}, \tag{3.2}$$

which leads to the log-likelihood

$$\begin{aligned}&\log P(x(t,m)\forall t,m)\\ &\quad= \sum_{t=1}^{T}\sum_{m=1}^{M}x(t,m)\log\sum_{k=1}^{K}P(k|t)P(m|k).\end{aligned} \tag{3.3}$$

With the PLCA model, (3.1) can be interpreted as trials following a probability distribution (3.3). $P(k|t)$ corresponds to $c(k|t)$ in (3.1), and $P(k|t)$ is called the probabilistic temporal activation. Moreover, $P(m|k)$ corresponds to $s(m|k)$ in (3.1), and $P(m|k)$ is

called the probabilistic spectral basis component. Therefore, first, the estimation process calculates $P(k|t)$ and $P(m|k)$ that maximizes (3.3). Next, it calculates the maximum-likelihood estimates of $c(k|t)$ and $s(m|k)$ from these probabilistic distributions. However, it cannot estimate both $P(k|t)$ and $P(m|k)$ directly because $P(k|t)$ is needed to estimate $P(m|k)$. The expectation-maximization (EM) algorithm for maximum-likelihood estimation with missing data can be employed. Therefore, EM-based estimation algorithm (Algorithm 1) can be obtained. After the iteration process, the estimate $\hat{c}(k|t)$ of $c_(k|t)$ can be calculated from (3.5), and $\hat{s}(m|k)$ of $s(m|k)$ can also be estimated from (3.7). The PLCA-based method can utilize the non-negativity constraint similarly to NMF because both $P(k|t)$ and $P(m|k)$ are probability variables that are always non-negative. Moreover, the PLCA-based method does not make use of orthogonality, thus we can estimate the spectral basis components that are not orthogonal to each other.

The statistical relationship between PLCA and the mass spectrometry-based explosives detection is described as follows. There are three cost functions commonly used in NMF, namely, the Euclidean distance, KL divergence, and Itakura-Saito divergence. These versions are special cases of NMF with β-divergence [72][73]. PLCA is a statistical formulation of NMF with KL-divergence (KL-NMF). In the case of NMF with KL-divergence, the cost function equals the minus-log-likelihood under the assumption that the input signal is generated from a Poisson distribution [73]. On the other hand, it is commonly known that the variation of intensities of mass spectra follows a Poisson distribution [74]. Because both of them follow a Poisson distribution, it can be considered that PLCA is valid for mass spectrometry-based explosives detection. However, the problem concerning the scale factor of the Poisson distribution [75] must be noted. In regard to the developed system, the scale factor of the Poisson distribution corresponds to the sensitivity of the mass spectrometer. After the sensitivity of the mass spectrometer is tuned adequately in advance, the sensitivity does not change largely. In this study, the problem concerning the scale factor can therefore be neglected.

3.3.2 Solving the uncertainty problem by using a sparsity assumption

The non-negativity constraint of NMF or PLCA cannot completely solve the uncertainty problem. That is, the uncertainty can degrade the accuracy of $P(k|t)$ and that of $P(m|k)$. To obtain the solution accurately, a sparsity assumption can be applied to a mass spectrogram obtained by the explosives-detection system is introduced as follows:

Sparsity of activation: It is assumed that only a few substances k are active at the same time. This assumption is derived from two facts. The first is that people can walk through the gate of the explosives detection system one by one. The second is that when a person passes through the detector, peaks in the mass spectrogram rise rapidly and attenuate within seconds. Due to these facts, the sparsity of activation, $P(k|t)$, can be assumed.

As mentioned in Section 3.1, there are many approaches to formulate the sparsity assumption for NMF and PLCA. For example, to represent the sparsity of activation, L1 regularization [65], L1/L2 regularization [1][66], Dirichlet prior[76], and Gamma-Poisson Model[77] can be applied to NMF. Here, to represent the sparsity constraint, an entropic prior [70], which is widely used in PLCA, is used. From an industrial viewpoint, it is important that an algorithm is easy to be implemented for applying it to a product. The sparsity constraint by the entropic prior is easier to be implemented than other approaches. We therefore choose the entropic prior. The objective function is defined by adding the term corresponding to the "entropic prior" to (3.3) as follows:

$$\begin{aligned} & J\left(\{P(k|t)\}, \{P(m|k)\}\right) \\ &= \sum_{t=1}^{T}\sum_{m=1}^{M} x(t,m) \log \sum_{k=1}^{K} P(k|t)P(m|k) \\ &\quad -\alpha \sum_{t=1}^{T} H(\{P(k|t)\}_k), \end{aligned} \tag{3.9}$$

where α is the parameter of the sparsity of $P(k|t)$, and $H(\{P(k|t)\}_k)$ represents the entropy of $P(k|t)$, i.e. $H(\{P(k|t)\}_k) = -\sum_k P(k|t) \log P(k|t)$. In (3.9), the second term corresponds to the sparsity of $P(k|t)$. Because the objective function is transformed from (3.3) to (3.9), one step of **Algorithm 1** is also transformed to a new step. First, (3.6)

is transformed by the sparsity of $P(k|t)$ as follows:

$$P(k|t) = g(\beta, \{\hat{c}(k|t)\}_k), \tag{3.10}$$

where $\beta \geq 1.0$ is the sparsity parameter, and $g(\beta, \{\gamma_k\}_k)$ is the sparsity-controlling function proposed by Grindlay and Ellis [78]: $g(\beta, \{\gamma_k\}_k) = \frac{{\gamma_k}^\beta}{\sum_k {\gamma_k}^\beta}$. The sparsity-controlling function proposed by Grindlay and Ellis [78] is widely used. The derivation of the function has not been proven, whereas it is empirically known that the function is effective. β can be tuned manually by checking the separation performance for a test sample. If the detection performance can be evaluated in a real environment, β can be set such that the detection performance is maximized.

It is necessary to pay attention to that the sparsity assumption does not hold for the chemical background. The signal corresponding to the chemical background is not sparse because the chemical background is stationary. Thus, the sparsity constraint may have a bad effect on the separation performance. In order not to apply the sparsity constraint to the chemical background, it is required to distinguish the index corresponding to the chemical background from those corresponding to the other substances, and it is required to apply the sparsity constraint only to the other substances. Here, the chemical background is assigned to the first index of substances, i.e. $k = 1$. To assign the signal corresponding to the chemical background to $k = 1$, the feature that the chemical background always exist can be used, i.e. $P(k = 1|t)$ is always larger than $P(k \neq 1|t)$. By using this feature, $P(k = 1|t)$ is set to a higher value than $P(k \neq 1|t)$ in (3.10) as follows:

$$P(k|t) = \begin{cases} \frac{1}{1+\sum_{k' \neq 1} g(\beta, \{\hat{c}(k'|t)\}_k)} & \text{if } k = 1, \\ \frac{g(\beta, \hat{c}(k|t))}{1+\sum_{k' \neq 1} g(\beta, \{\hat{c}(k'|t)\}_k)} & \text{otherwise.} \end{cases} \tag{3.11}$$

Because $P(k = 1|t)$ is the highest value at all the time-points, the spectral bins that are active at all the time-points are mapped to the first substance $k = 1$. This approach makes the algorithm robust to noise from the chemical background. Hereafter, this heuristic approach is called "background mapping". Converting the steps as above gives an algorithm (Algorithm 2), for estimating $P(k|t)$ and $P(m|k)$. After the iteration process, finally, the estimate $\hat{c}(k|t)$ of $c(k|t)$ can be calculated from (3.5), and the estimate $\hat{s}(m|k)$ of $s(m|k)$ can be calculated from (3.7).

The convergence of the proposed method is explained in the following. For the NMF

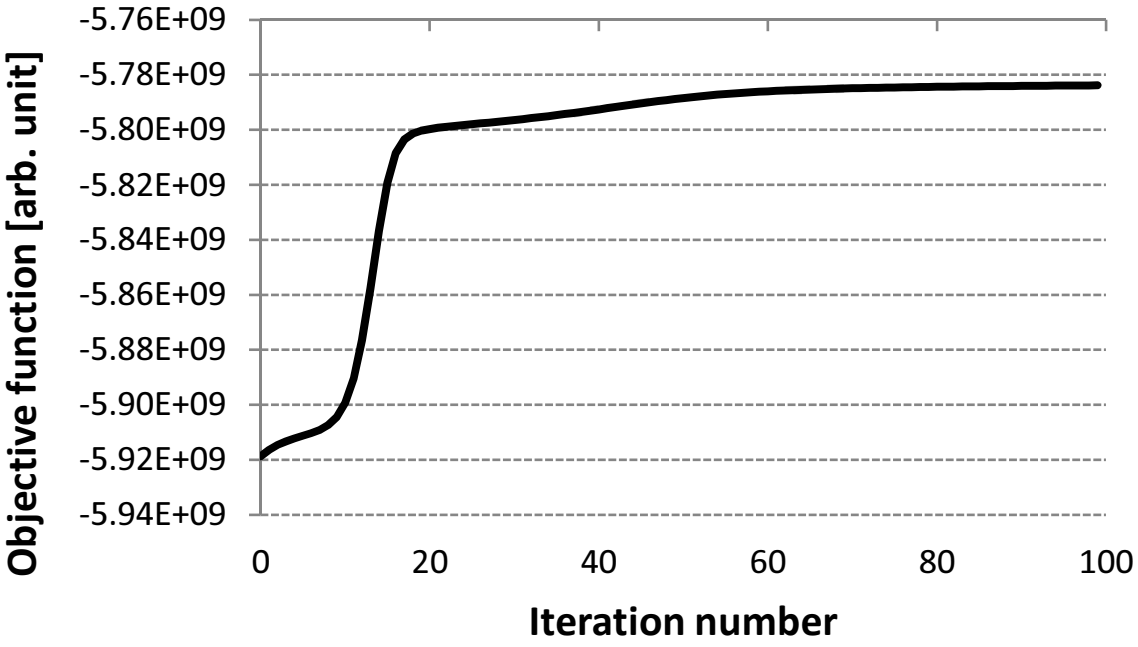


Figure 3.1: Convergence of the objective function (3.9).

with KL-divergence, Lee and Seung proved the monotonic convergence of the objective function [62]. However, in general, it is hard to prove the convergence of unknown parameters. Badeau *et al.* analyzed the convergence properties of unknown parameters [79]. For the supervised NMF, Badeau *et al.* proved the convergence of unknown parameters [79], whereas for the unsupervised NMF, the convergence of unknown parameters was not proven. Also for Algorithm 2, it is difficult to prove the convergence of $P(k|t)$ and $P(m|k)$. For practical use, the convergence of the objective function of Algorithm 2 has been experimentally checked. Here, an example of the convergence of the objective function (3.9) is shown in Fig. 3.1.

The sensitivity of Algorithm 2 to the sparsity parameter is explained as follows. This sensitivity is not an essential problem in regard to the developed system for the following reason. It is considered that the sparsity of activation depends mainly on pedestrian traffic. There is therefore a probability that the sparsity parameter will be changed by pedestrian traffic. Pedestrian traffic can be measured easily because the walk-through portal explosives detection system [3] has an optical sensor for detecting people passing through it. Consequently, if the sparsity parameter corresponding to the pedestrian traffic is tuned in advance, explosive substances can be separated by selecting the tuned

parameter according to on the pedestrian traffic. (The sensitivity of the algorithm to the sparsity parameter will be empirically verified in future work.)

3.3.3 Methods for applying the proposed algorithm to the explosives-detection system

In this section, how to apply the proposed separation algorithm to the explosives-detection system is described. Fig. 3.2 shows the process block diagram of an example of the explosives-detection system using the proposed separation algorithm. The system obtains a new mass spectrum from the mass spectrometer every sampling interval. Each time when the system obtains the new mass spectrum, the system runs in the following steps:

1. The new mass spectrum is appended to the past spectrogram buffered in a memory, and the input signal for separation, $x(t, m)$, is generated.

2. By the proposed separation algorithm, $x(t, m)$ is separated into individual substances, and the estimated basis component $\hat{s}(m|k)$ and the estimated activation $\hat{c}(k|t)$ are obtained for each substance.

3. For each k-th substance, the separated spectrum, $\boldsymbol{y}_k = [y(1, k), \cdots, y(M, k)]^T$, is calculated by $y(m, k) = \hat{c}(k|T)\hat{s}(m|k)$.

4. For each v-th explosive substance registered in a database, $\boldsymbol{y}_k$ is compared with the v-th template spectrum, $\boldsymbol{d}_v = [d(1, v), \cdots, d(M, v)]^T$, and the score corresponding to the amount of the v-th explosive substance in $\boldsymbol{y}_k$, $a(v, k)$, is calculated. A simple method for calculating $a(v, k)$ is the inner product, i.e. $a(v, k) = \boldsymbol{d}_v^T \boldsymbol{y}_k$. However, actually, the m/z-axis may shift due to the condition of the mass spectrometer. Thus, to improve the robustness to the shift of the m/z-axis, $a(v, k)$ is calculated as follows:

$$a(v, k) = \max_{\delta \in [-\Delta, \Delta]} \sum_m d(m, v) y(m + \delta, k), \tag{3.12}$$

where Δ is the tolerance range for the shift, and δ is an integer between $-\Delta$ and Δ.

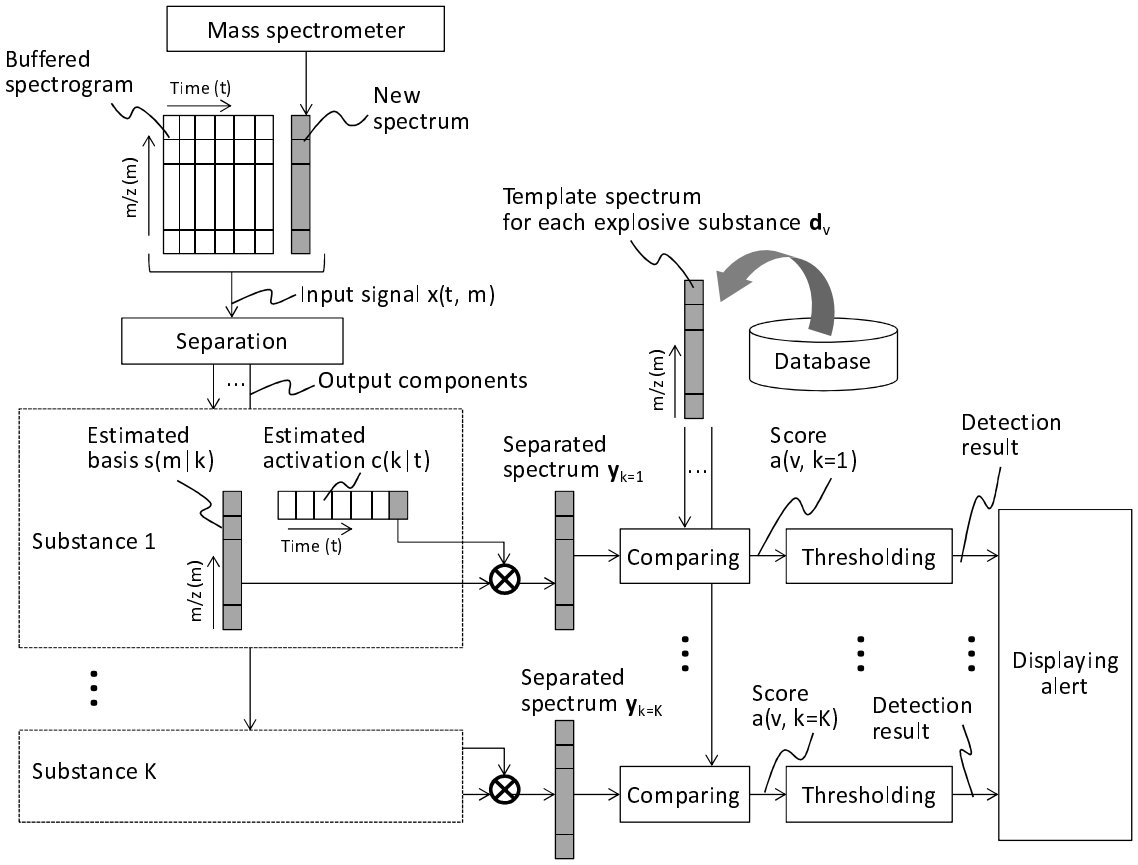


Figure 3.2: Process block diagram of explosives-detection using the proposed algorithm.

5. If $a(v, k)$ is larger than a threshold for any v and k, then an alert is displayed as the detection result.

As above, the proposed separation algorithm works as an online algorithm, i.e. the proposed algorithm uses only past data, Thus, the system can detect explosive substances in real time in principle.

The calculation time is also important for real-time detection. The computational complexity of the proposed separation algorithm is $O(KTML)$, where L is the number of iterations. As shown from the computational complexity, if T or M is larger, the algorithm can not work in real time by central processing unit (CPU) computation. To solve this problem, it is possible to utilize graphical processing units (GPU), which are installed on a number of personal computers today. In the proposed algorithm (Algorithm 2), three parts can be performed in parallel by GPU computation as follows:

1. (3.4) in **E step** can be calculated in parallel for each m and t.

2. (3.5) (3.11) in **M step** can be calculated in parallel for each t.

3. (3.7) in **M step** can be calculated in parallel for each m.

By parallel computing, the algorithm can be accelerated and can work in real time.

3.4 Experimental results

The performance of the proposed method was experimentally evaluated by using the walk-through portal explosives-detection system [3] to record an input mass spectrogram. We previously developed a prototype system as supported by The Ministry of Education, Culture, Sports, Science and Technology, Japan (2007 to 2009). A model of the system is shown in Fig. 3.3. To measure the chemical background of a real environment, the device was used to record a mass spectra recorded in a real station. 3500 mass spectra obtained from several people passing through the detector for about five minutes from the whole recorded data; i.e. $T = 3500$, were used, and the number of the m/z points, M, was 512. The input mass spectra are shown in Fig. 3.4 (a), and the chromatogram (time profile) corresponding to around m/z 59 is shown in Fig. 3.4 (b). The chemical-background components have stationary peaks at m/z 59, m/z 62, and m/z 75 (Fig. 3.4 (a)). In this experimental evaluation, an experimenter passed through the system carrying substance 1 (m/z 59) four times in the first half of the time, and with substance 2 (m/z 59, m/z 62, m/z 76 and m/z 77) five times in the second half of the time. As shown in Fig. 3.4 (b), the fourth peak of substance 1 ($t = 1600$) is small and at the same level as those of when substance 2 was passed (e.g. $t = 1950$). From Fig. 3.4, $P(m|k)$ for substance (k) is not orthogonal to $P(m|k)$s for the other substances. The number of substances (K) in the estimation process was set at eight. The maximum number of iterations was 100. The sparsity parameter β was 1.01. This sparsity parameter with fully manual tuning such that they maximize the separation performance was found.

The estimates of $P(m|k)$ and $P(k|t)$ are shown in Fig. 3.5 and 3.6. As shown in Fig. 3.5, all correct main peaks for the chemical background, substance 1, and substance 2 were estimated. As shown in Fig. 3.6, the peaks exist at the correct times when substance 1 and substance 2 were passed. In particular, for substance 1, the fourth peak of $P(k|t)$ is obviously higher than that of $P(k|t)$ in the second half of the time (Fig. 3.6 (b)).

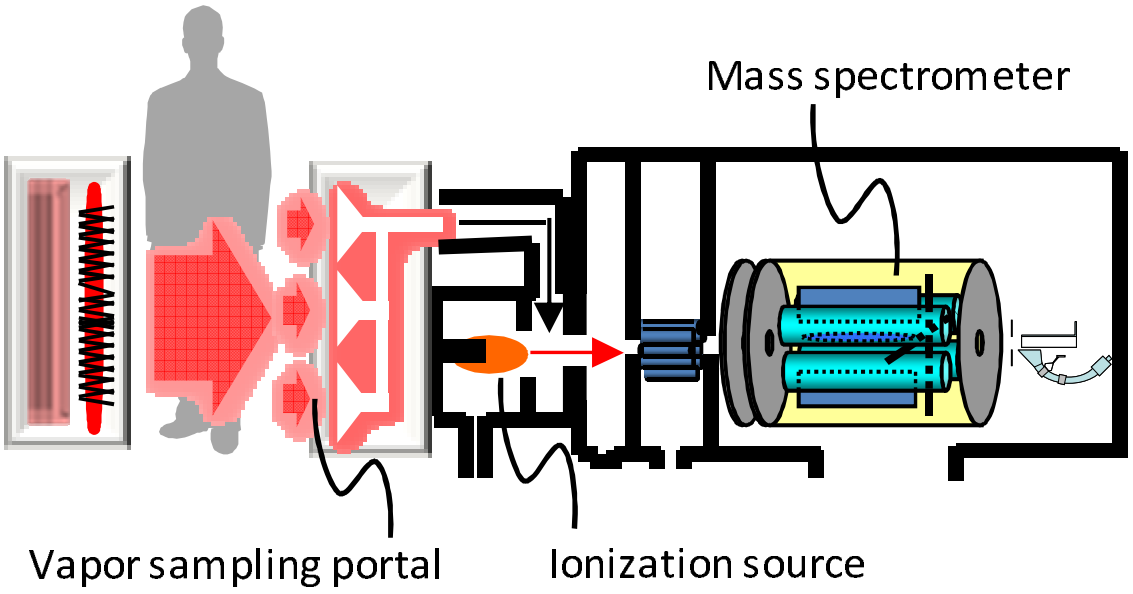


Figure 3.3: Explosives detector.

The results shown in Figs 3.5 and 3.6 indicate that the proposed algorithm can separate explosive substances precisely. However, substance 2 was separated into two different basis components $k = 3$ and 4. The side lobes of the spectral basis component $P(m|k=3)$ (Fig. 3.5 (c)) are broader than those of $P(m|k=4)$ (Fig. 3.5 (d)). This feature infers that the component for $k = 3$ corresponds to a saturation state known as the "space charge effect" [74], and the component for $k = 4$ corresponds to a non-saturation state. A partially enlarged view of the probabilistic temporal activation of Figs. 3.6 (c) and (d) is shown in Fig. 3.7. As shown in Fig. 3.7, the component for $k = 4$ rose immediately after the person passed through the system, the peak of the component for $k = 3$ rose largely, and the component of for $k = 4$ became active with time. These results indicate that when the component of for $k = 3$ was active, the amount of substance 2 was large. This finding strengthens the assumption that the component for $k = 3$ corresponds to the saturation state. In this experimental evaluation, the saturation actually occurred in the mass spectrometer because the amount of substance 2 was too large. The algorithm described in this chapter cannot handle temporal structure of spectral basis components because each basis component is modeled by only one spectral pattern. An method for modeling temporal structure is described in Chapter 5.

The separation performance of the proposed method was compared with that of

existing methods (i.e. PCA, ICA, and PLCA) without the sparsity constraint. For performance evaluation of source separation, in general, it is desired to measure the error between the ground truth and the separated signal if the ground truth can be obtained. However, in this experimental evaluation, the ground truth can not be prepared because, even in laboratory environments, not only explosive substances but also interference substances and a chemical background exist in the air. It is therefore required to use another method for evaluating the separation performance. Here, it is possible to use the fact that, conversely, the explosive substances did not exist in the real station when the experimenter did not pass through the system. Thus, if the separation performance is high, the intensity of the separated signal will be small at the time when the experimenter did not pass through the system. Considering this feature, we evaluated the separation performance by using the Signal-to-Noise Ratio (SNR) defined by

$$\mathrm{SNR}_{k,i,j} = 10 \log_{10} \frac{\max_{t \in \mathcal{A}_{k,i}} |\hat{c}(k|t)_j|}{\sqrt{\frac{1}{|\mathcal{N}_k|} \sum_{t \in \mathcal{N}_k} |\hat{c}(k|t)_j|^2}} \ [\mathrm{dB}] \,, \tag{3.13}$$

where the numerator in the logarithm is the intensity of the separated signal at the time when the k-th substance is passed through the device, the denominator in the logarithm is the root-mean-square of the intensity of the separated signal at the time when the k-th substance is not passed through the device, $\mathcal{A}_{k,i}$ is the i-th short-time range when the k-th substance is passed through the device, $\mathcal{N}_k$ is the time range when the k-th substance is not passed through the device; i.e. $\mathcal{N}_{k=2}$ is $[2000, 3500]$, and $\mathcal{N}_{k=3}$ is $[0, 1500]$, j is the index of executions of the estimation process, and the number of the execution times is 50, i.e. $1 \leq j \leq 50$. (In each execution, all the unknown parameters were initialized by different random values.) If the separation performance is high, the denominator in the logarithm of (3.13) is small, and then (3.13) is large. In addition, SNR_k is defined as an ensemble mean over i and j. In the case of the arithmetic mean, SNR is prone to an excessively high value even if there is a low $\mathrm{SNR}_{k,i,j}$. Therefore, the arithmetic mean is not fit to measure the robustness. In contrast, the harmonic mean is sensitive to the case that $\mathrm{SNR}_{k,i,j}$ is low. Therefore, the harmonic mean is fit to measure the robustness. SNR_k is thus defined as the harmonic mean of $\mathrm{SNR}_{k,i,j}$:

$$\begin{aligned} \mathrm{SNR}_k \ &= \ \#\text{passing trials} \times \#\text{process executions} \\ &\quad \times \left\{ \sum_{i,j} \frac{1}{\mathrm{SNR}_{k,i,j}} \right\}^{-1} . \end{aligned} \tag{3.14}$$

SNR_k is shown in Fig. 3.8. "PLCA" means the original PLCA (Algorithm 1). "Background mapping" means Algorithm 2 without the sparsity constraint ($\beta = 1$). "Sparse" means Algorithm 2 with the sparsity constraint ($\beta = 1.01$). In Fig. 3.8, the separation performance of the proposed algorithm ("PLCA", "background mapping", and "sparse") is higher than that of PCA and ICA. In relation to the results, the spectral basis components for substance 1 estimated by PCA and ICA are shown in Fig. 3.9. Figs. 3.5 (b) and (c) are compared with Figs. 3.9 (a) and (b). PCA and ICA estimate the spectral basis components without the constraint of non-negativity, so the estimates of the spectral basis components have both a negative peak and a positive peak. In contrast, the proposed method can estimate the spectral basis components accurately by using the constraint of non-negativity. From this figure, it is considered that the reasons for the high performance of the proposed algorithm are the constraint of non-negativity and no assumption about orthogonality. In addition, in Fig. 3.8, the separation performance of background mapping is higher than that of the original PLCA. The results of this comparison indicate that background mapping improves robustness to noise from the chemical background. Moreover, to evaluate the effect of the sparsity constraint, PLCA with the sparsity constraint ("sparse") was compared with the other algorithms. The results of this comparison (Fig. 3.8) indicate that the sparsity constraint improves the separation performance.

Fig. 3.10 shows the separation performance of the proposed algorithm and that of the conventional NMF with L1/L2 regularization [1]. As shown in Fig. 3.10, in L1/L2 regularization, the separation performance was improved by tuning the sparseness parameter. These results indicate that L1/L2 regularization is effective. However, the performance of the L1/L2 regularization was not higher than that of the proposed algorithm. The L1/L2 regularization applies the sparsity constraint to all the substances equally without distinguishing the chemical background from the other substances, whereas the proposed algorithm can apply the sparsity constraint only to the substances except the chemical background. It can be considered that the reason why the separation performance of the proposed algorithm was higher than the L1/L2 regularization. These results indicate that the sparsity constraint of the proposed algorithm is effective for improve the performance although it is easy to be implemented.

Table 3.1: The calculation time for CPU and GPU. The real-time factors (RTF) were calculated in the assumption that the sampling interval is 0.5 seconds. (Under the condition that the sampling interval is 1.0 seconds, the RTFs are equivalent to the columns of the calculation time [sec].)

	CPU		GPU	
T	Calculation time [sec]	RTF	Calculation time [sec]	RTF
50	0.92	1.84	0.21	0.42
500	4.24	8.48	0.27	0.54
1500	11.63	23.26	0.42	0.84

Finally, the calculation time described in Section 3.3.3 was evaluated. Computation was performed on a desktop personal computer running Windows 7. The computer had an Intel Core i7 980 CPU with 3.33GHz, 12GB of RAM, and an NVIDIA GeForce GTX580 GPU with 512 cores. CPU computation was performed using a single computation thread. For CPU computation, the code of the algorithm was implemented by using the C language. For GPU computation, the code was implemented by using the CUDA. M was 512, K was 8, and the number of iterations was fixed to 100. The calculation time was measured for $T = 50$, 500, and 1500. Table 3.1 shows the results of the calculation time. The real-time factors (RTF) were calculated in the assumption that the sampling interval is 0.5 seconds. (The sampling interval is set typically between 0.5 seconds and 1.0 seconds.) As shown in Table 3.1, the RTFs were lower than 1 in the case of GPU computation. These results indicate that the algorithm accelerated by GPU computation works in real time. Particularly, if the sampling interval is 0.5 seconds, the algorithm can be performed in real time by using 3000-second mass spectrogram ($T = 1500$) as the input signal.

3.5 Conclusion

A new algorithm for separating a mass spectrogram into individual substances was proposed for explosives detection. The proposed algorithm is based on PLCA. By using PLCA, the algorithm imposes the constraint of non-negativity and does not impose

orthogonality. In addition, to estimate the components more accurately, the sparsity assumptions were applied to PLCA. In this chapter, industrial application of the algorithm into the explosives-detection system was shown. In an experimental evaluation of the algorithm using data obtained from a real environment, it was shown that the proposed algorithm outperforms conventional ones. Moreover, the evaluation results show that the sparsity constraint improves separation performance. It was also shown that the algorithm accelerated by a graphical processing unit (GPU) can work in real time.

However, the results also show that one basis component tends to be split into multiple basis components mistakenly, because the proposed method does not model the temporal structure of spectral basis components. In Chapter 5, an extended version of the proposed method for modeling temporal structure of a mass spectrogram is explained.

Algorithm 1 PLCA-based algorithm

1. Initialization process

Set all the unknown parameters to random values.

2. Iteration process

Iterate the following E step and M step until convergence

or a maximum number of iterations is reached.

E step:

Compute $P(k|t,m)$ as follows:

$$P(k|t,m) = \frac{P(k|t)P(m|k)}{\sum_{k'} P(k'|t)P(m|k')}. \tag{3.4}$$

M step:

Compute $\hat{c}(k|t)$ as follows:

$$\hat{c}(k|t) = \sum_{m} x(t,m)P(k|t,m), \tag{3.5}$$

Compute $P(k|t)$ as follows:

$$P(k|t) = \frac{\hat{c}(k|t)}{\sum_{k} \hat{c}(k|t)}, \tag{3.6}$$

Compute $\hat{s}(m|k)$ as follows:

$$\hat{s}(m|k) = \sum_{t} x(t,m)P(k|t,m), \tag{3.7}$$

Compute $P(m|k)$ as follows:

$$P(m|k) = \frac{\hat{s}(m|k)}{\sum_{m} \hat{s}(m|k)}. \tag{3.8}$$

Algorithm 2 PLCA-based algorithm with the sparsity constraint

1. Initialization process

Set all the unknown parameters to random values.

2. Iteration process

Iterate the following E step and M step until convergence

or a maximum number of iterations is reached.

E step:

Compute (3.4).

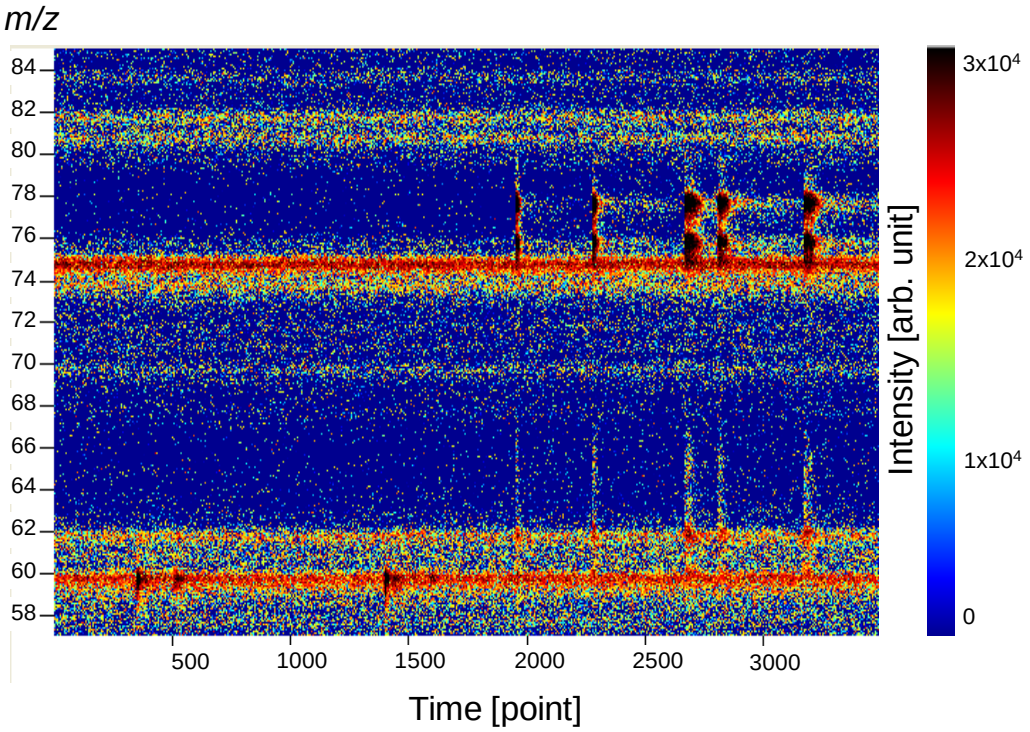
M step:

Compute (3.5).

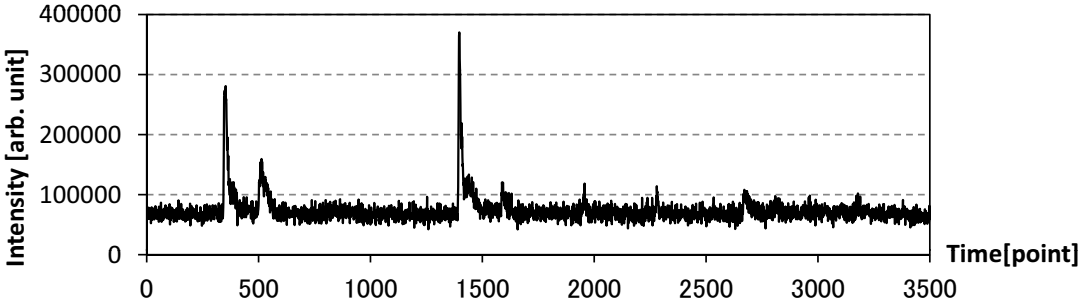
Compute (3.11).

Compute (3.7).

Compute (3.8).

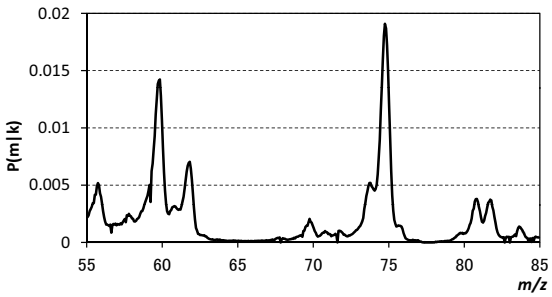


(a) Mass spectra $x(t, m)$. X and Y axis show t and m/z.

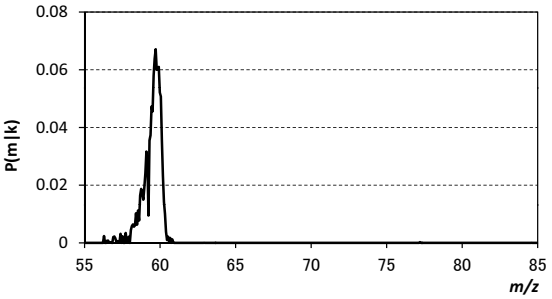


(b) Chromatogram (time profile) for around m/z 59. X and Y axis show t and the intensity $I(t) = \sum_{m \in [m/z\ 58,\ m/z\ 60]} x(t, m)$.

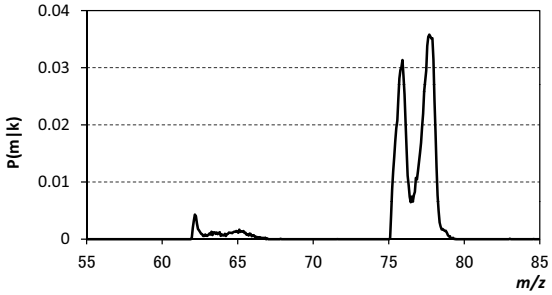
Figure 3.4: Input signal.



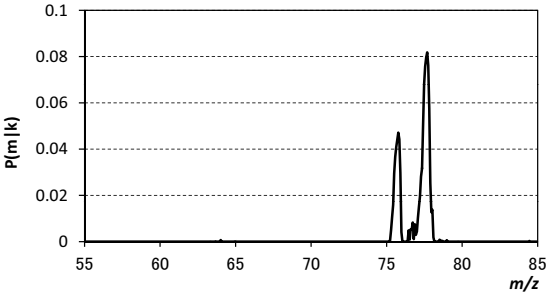
(a) $k = 1$ (Chemical background)



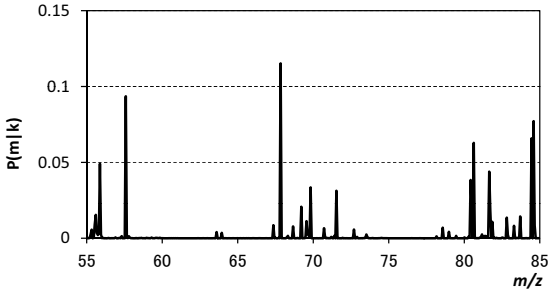
(b) $k = 2$ (Substance 1)



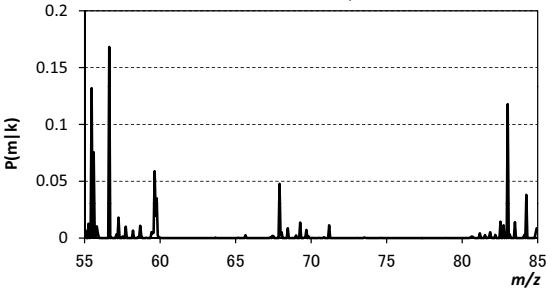
(c) $k = 3$ (First component of substance 2)



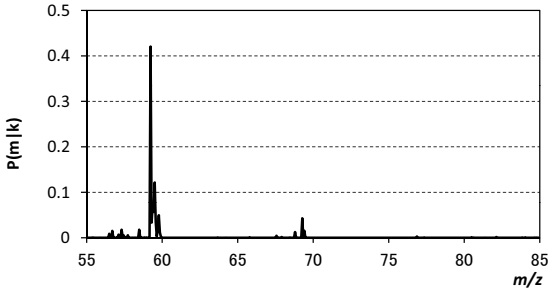
(d) $k = 4$ (Second component of substance 2)



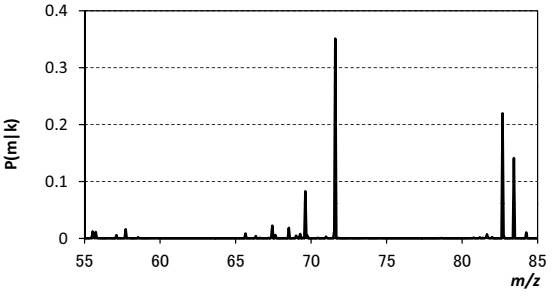
(e) $k = 5$ (Dummy)



(f) $k = 6$ (Dummy)

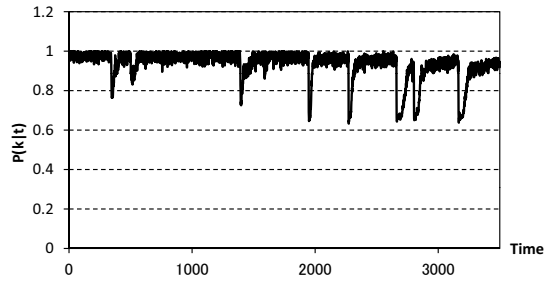


(g) $k = 7$ (Dummy)

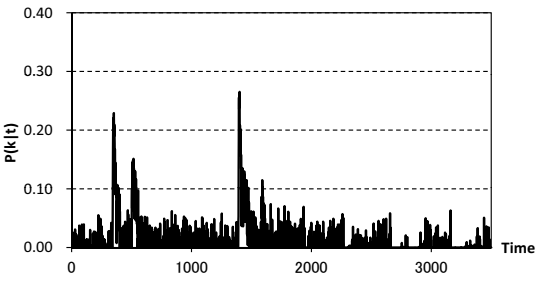


(h) $k = 8$ (Dummy)

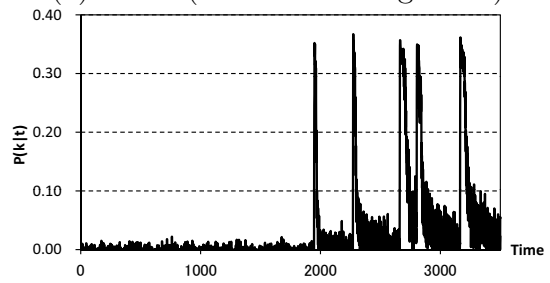
Figure 3.5: Estimates of probabilistic spectral basis components $P(m|k)$ for each k. X and Y show m/z and $P(m|k)$.



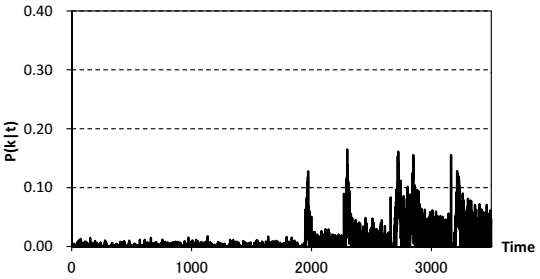
(a) $k = 1$ (Chemical background)



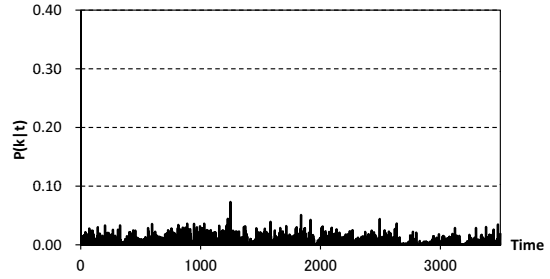
(b) $k = 2$ (Substance 1)



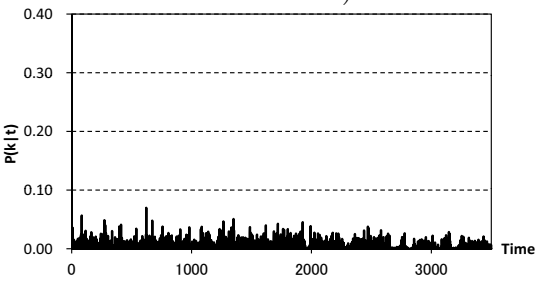
(c) $k = 3$ (First component of substance 2)



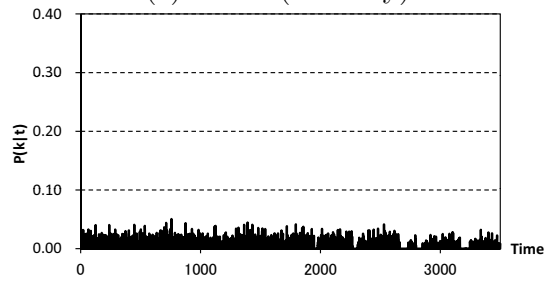
(d) $k = 4$ (Second component of substance 2)



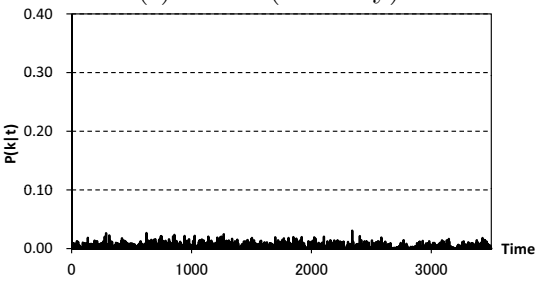
(e) $k = 5$ (Dummy)



(f) $k = 6$ (Dummy)



(g) $k = 7$ (Dummy)



(h) $k = 8$ (Dummy)

Figure 3.6: Estimates of probabilistic temporal activation $P(k|t)$ for each k. X and Y show t and $P(k|t)$.

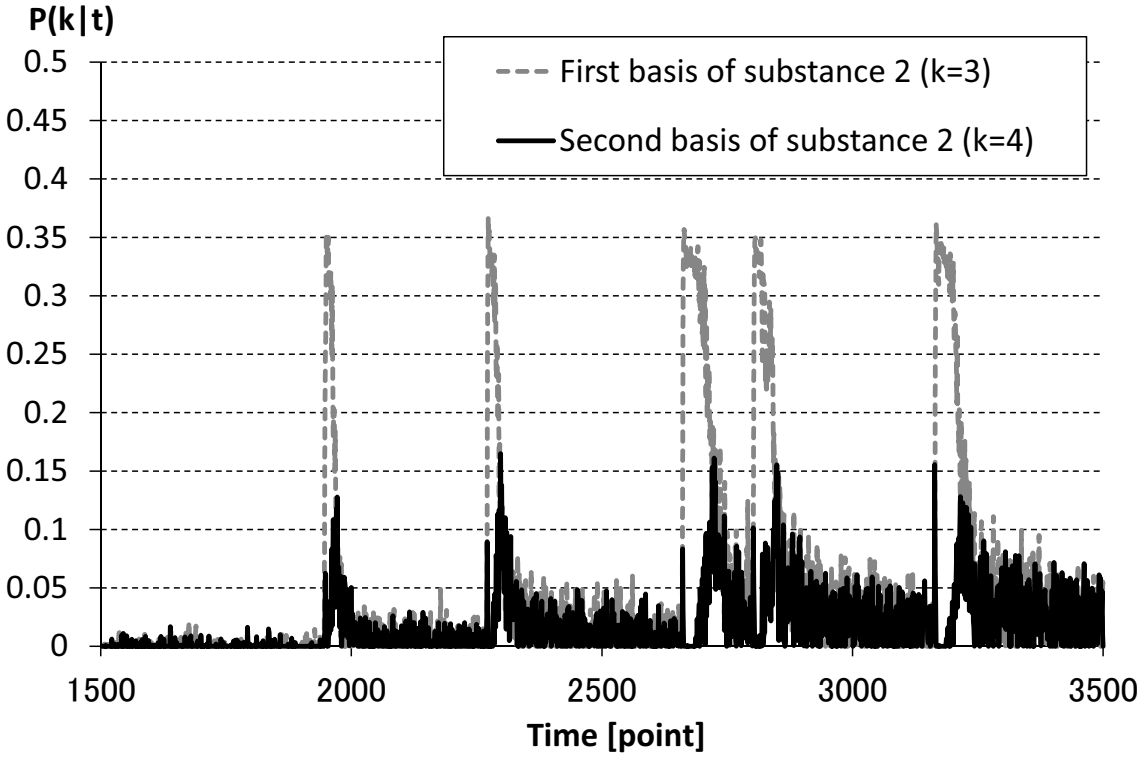
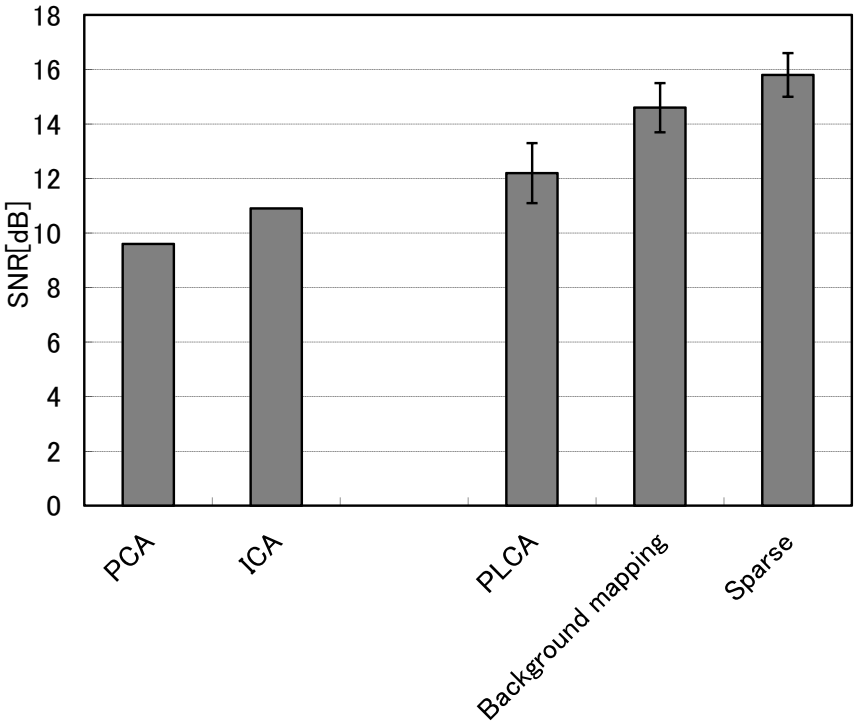
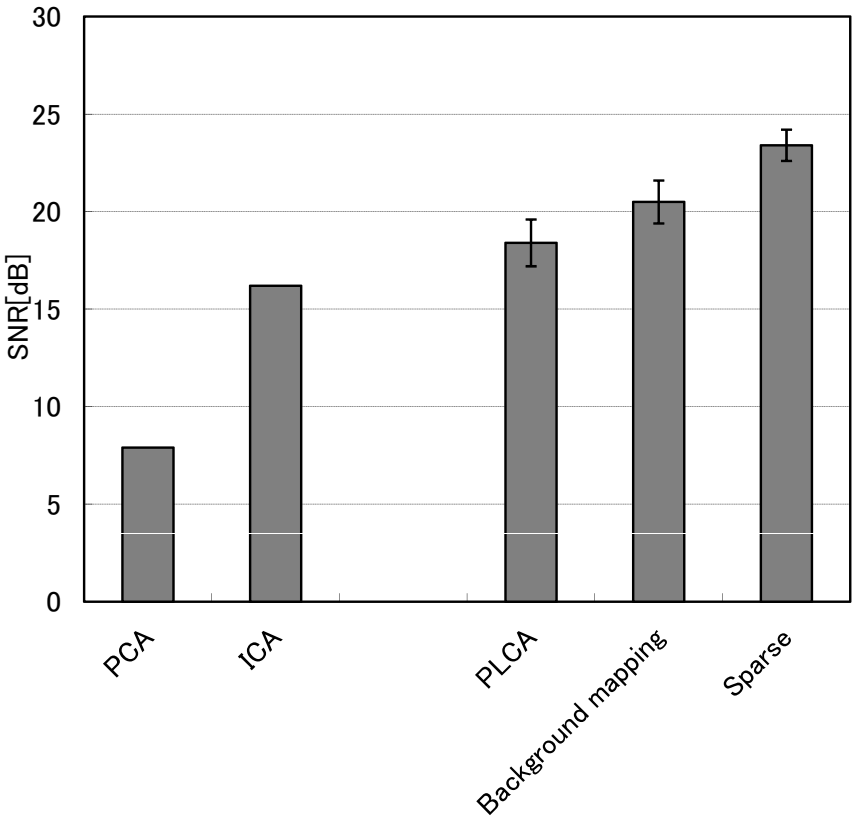


Figure 3.7: Partially enlarged view of probabilistic temporal activation $P(k|t)$ of the two components of substance 2 ($k = 3$ and $k = 4$).

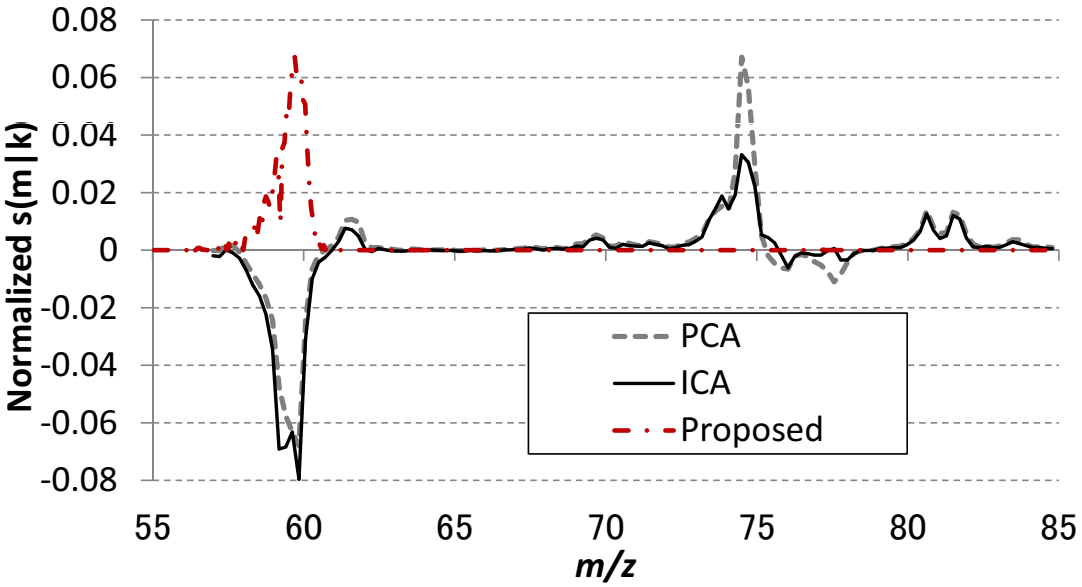


(a) Substance 1 ($k = 2$)

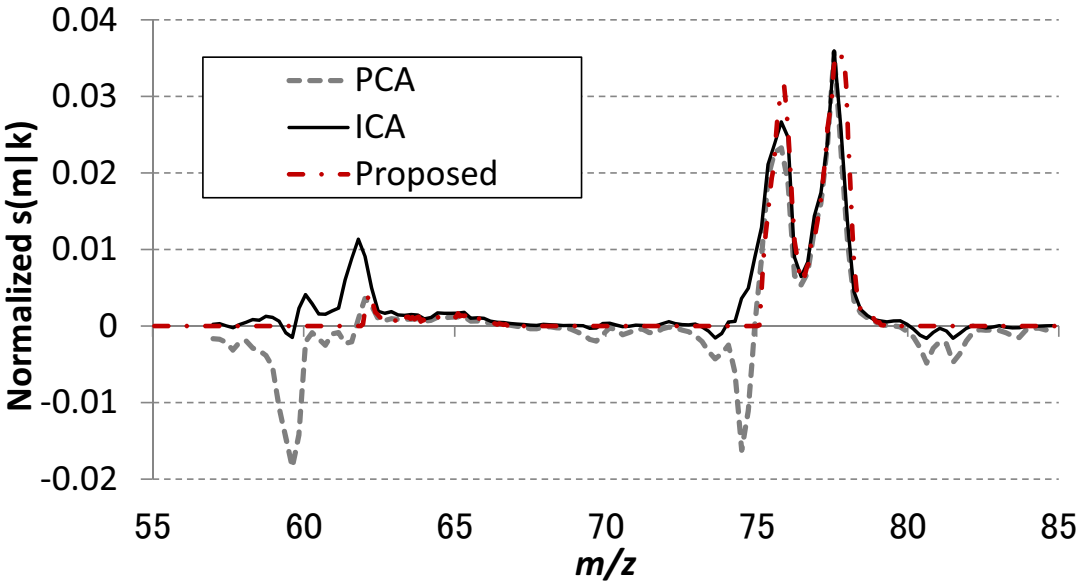


(b) Substance 2 ($k = 3$)

Figure 3.8: SNR [dB] for each condition. Error bars represent 95% confidence intervals.



(a) Substance 1



(b) Substance 2

Figure 3.9: Estimates of the spectral basis components, $\hat{s}(m|k)$, of PCA, ICA, and the proposed algorithm (Algorithm 2). The X-axis shows m/z, and the Y-axis shows the estimated $s(m|k)$. $s(m|k)$ is normalized by the maximum value.

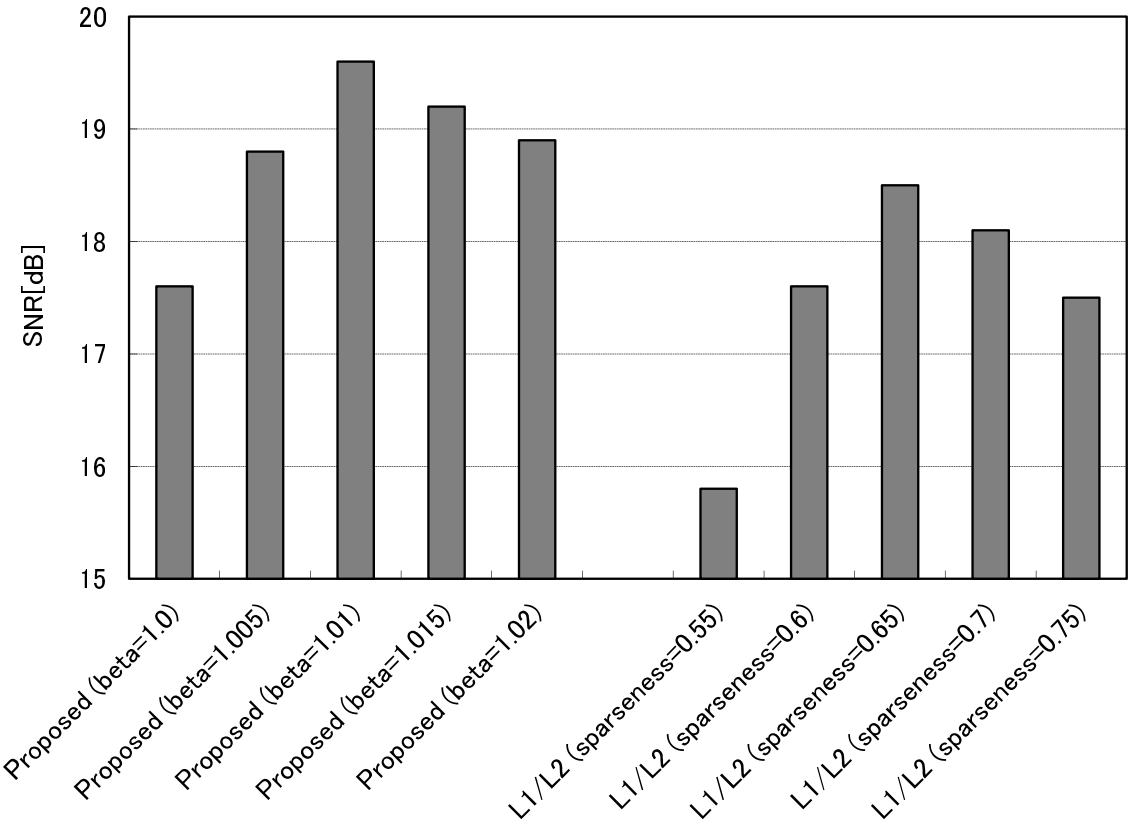


Figure 3.10: SNR [dB] for the proposed algorithm (Algorithm 2) and the conventional NMF with L1/L2 regularization [1]. "Proposed" means the proposed algorithm, and "beta" means β. "L1/L2" means the NMF with L1/L2 regularization, and "Sparseness" means the sparseness measure in [1] for the temporal activation. The SNR was the average of $\mathrm{SNR}_{k=2}$ and $\mathrm{SNR}_{k=3}$.

Chapter 4

ICA-based acceleration of probabilistic latent component analysis for mass spectrometry-based explosives detection

4.1 Introduction

In Chapter 3, an algorithm for separating a mass spectrogram into individual substances was proposed for explosives detection. However, the convergence of the method is slow, and the total calculation time is long. As shown in Chapter 3, for GPU, the method runs in real time, whereas it can not run in real time for CPU, and the execution environments are limited. In this chapter, we propose an acceleration method for PLCA. We focus on that ICA can stably obtain a solution near the correct solution, and its speed is fast. Thus, the proposed method makes use of ICA in the initialization process of PLCA. Experimental results indicate that the convergence of the proposed method is accelerated, and total calculation time is decreased.

4.2 Problem statement

The problem statement is the same as Chapter 3. The input signal is the time series of mass spectra $x(t, m)$, where t is the index of a time, and m is the index of m/z. T is the number of the time index, and M is the number of the index of m/z. $x(t, m)$ is modeled as follows,

$$x(t, m) = \sum_k c(k|t)s(m|k), \tag{4.1}$$

where k is the index of a compound basis, K is the number of the kinds of the compounds in the air, $c(k|t)$ is the intensity of the k-th compound in the time index t, and $s(m|k)$ is the time-invariant spectral basis component of the k-th compound. The task is to estimate the unknown variables $c(k|t)$ and $s(m|k)$ from the known variables $x(t, m)$, i.e. this is a blind-source-separation problem.

4.3 Proposed method

We assume that the reason why the speed of convergence is slow is that the initial solution is not adequate. As shown in Chapter 3, in mass spectra separation domain, the correct solutions have a tendency to be a sparse solution. For example, temporal activations tend to be sparse on the time axis, and spectral basis components also tend to be sparse on the m/z axis. However, the solution initialized by random values tends to be dense and far from the correct solution. In addition, there is a small probability that an incorrect solution is obtained by the random initialization because the result of PLCA can be highly dependent on the initial values. Thus, it is necessary to develop an initialization method that finds a solution near the correct solution and does not use randomness.

We focus on ICA for initialization. Similarly to PLCA, ICA is a blind-source-separation method, so that ICA is available for initialization. ICA does not impose non-negativity or orthogonality but independence to the solution. The independence is assumed also in PLCA, and the solution of ICA is near that of PLCA. Although the orthogonality does not hold in mass spectrometry, PCA and SVD impose orthogonality, and so the solution of ICA tends to be closer to the correct solution than PCA and SVD. In addition, ICA does not use randomness, so ICA does not suffer from the initial-value-

sensitivity. Moreover, fast algorithms of ICA are commonly known, for example, Fast ICA [80], the Natural Gradient algorithm [81], and the Auxiliary-function-based ICA [82]. So that, by comparison with PLCA, the calculation time of ICA is extremely short. Thus, by initializing the unknown parameters by ICA and reducing the number of iterations of PLCA, we aim to shorten the total calculation time. By converting PLCA, we can achieve ICA-PLCA algorithm (Algorithm 2). Similarly to the method proposed in Chapter 3, in order to concentrate the stationary chemical background on the first basis, i.e. k = 1, we set $P(m|k=1)$ to the uniform distribution in (4.6), and set $P(k=1|t)$ to the higher value than $P(k \neq 1|t)$ in (4.7).

The calculation complexity of PLCA is $O(LTKM)$, where L is the number of iterations. In contrast the calculation complexity of the above initialization process is $O(LTK^2)$. Therefore the initialization process is faster than PLCA in the case of $K < M$. The proposed method makes use of this feature, and it can reduce the total calculation time by increasing the number of iterations of the initialization process and decreasing that of PLCA.

4.4 Experimental results

By using the mass spectrogram of the experiment of Chapter 3, the performance of the proposed algorithm was experimentally evaluated. We compared the proposed method (ICA-PLCA) with the PLCA-based method of Chapter 3. In PLCA, all the unknown parameters were initialized by random values. On each condition, the estimation process was run 20 times. We set the number of bases K in the estimation process at eight, and β was 1.02. The estimation process was run in C# on a PC with an Intel Core i7 3.3GHz CPU and 12GB of RAM. The measurements were $\mathrm{SNR}_{k,i,j}$ as follows:

$$\mathrm{SNR}_{k,i,j} = 10 \log_{10} \frac{\max_{t \in \mathcal{A}_{k,i}} |\hat{c}(k|t)_j|}{\sqrt{\frac{1}{|\mathcal{N}_k|} \sum_{t \in \mathcal{N}_k} |\hat{c}(k|t)_j|^2}} \ [\mathrm{dB}]\,, \tag{4.8}$$

where $\mathcal{A}_{k,i}$ was the area around the i-th time when the k-th compound is passed through the device, $\mathcal{N}_k$ is the non-active time area; i.e. $\mathcal{N}_{k=2}$ was $[2000, 3500]$, and $\mathcal{N}_{k=3}$ was $[0, 1500]$, and j is the index of executions. Next, we defined SNR as an ensemble mean over k i, and j. In the case of the arithmetic mean, a peak $\mathrm{SNR}_{k,i,j}$ of which will be extremely high tends to cause SNR to be higher excessively. In order to make much

Algorithm 3 ICA-PLCA

In PLCA, replace the initialization process with the following equations:

1. By the whitening matrix $\boldsymbol{W}$, prewhiten $\boldsymbol{x}(t) = [x(t,1), \cdots, x(t,M)]^T$ and reduce the number of dimensions:

$$\boldsymbol{z}(t) = [z(t,1), \cdots, z(t,K)]^T = \boldsymbol{W}\boldsymbol{x}(t). \tag{4.2}$$

2. Compute the separated signals:

$$\boldsymbol{y}(t) = [y(t,1), \cdots, y(t,K)]^T = \boldsymbol{V}\boldsymbol{z}(t). \tag{4.3}$$

3. Compute the natural gradient;

$$\boldsymbol{V} = \boldsymbol{V} + \eta \left[\boldsymbol{I} - \frac{1}{T} \tanh(\boldsymbol{y}(t))\boldsymbol{y}(t) \right] \boldsymbol{V}. \tag{4.4}$$

4. Return to 2. until convergence.

5. Convert $\boldsymbol{V}$ into a basis matrix $\boldsymbol{S}$ on the m/z space:

$$\boldsymbol{S} = \boldsymbol{V}\boldsymbol{W}. \tag{4.5}$$

6. By normalizing $\boldsymbol{S}$, initialize $P(m|k)$:

$$P(m|k) = \begin{cases} \frac{1}{M} & \text{if } k = 1, \\ \frac{|S_{m,k-1}|}{\sum_m |S_{m,k-1}|} & \text{otherwise.} \end{cases} \tag{4.6}$$

7. By normalizing $\boldsymbol{y}(t)$, initialize $P(k|t)$:

$$P(k|t) = \begin{cases} \frac{1}{1+\sum_k' {y'}_k(t)} & \text{if } k = 1, \\ \frac{{y'}_{k-1}(t)}{1+\sum_k' {y'}_k(t)} & \text{otherwise.} \end{cases} \tag{4.7}$$

where $y'_k(t) = \frac{|y_k(t)|}{\sum_k |y_k(t)|}$.

8. Run the iteration process of PLCA with the sparsity constraint (Algorighm 2).

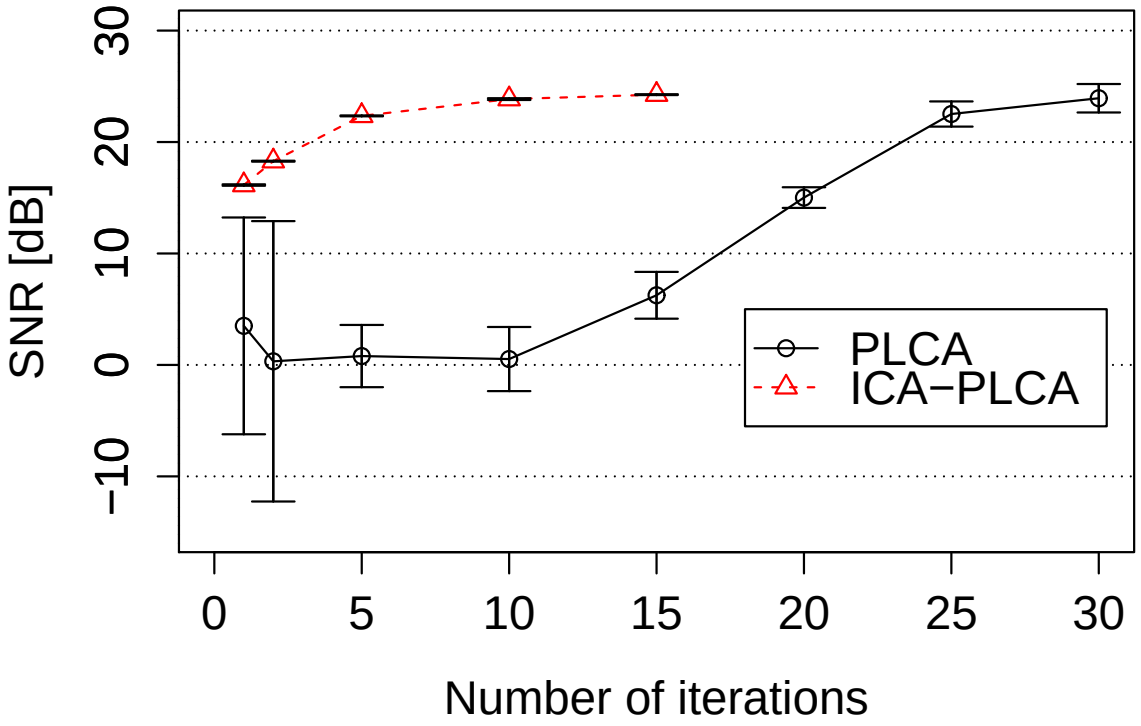


Figure 4.1: SNR for each method. X and Y show the number of iterations and SNR [dB]. Error bars represent 95% confidence intervals.

account of worse $\mathrm{SNR}_{k,i,j}$, we defined SNR as the harmonic mean of $\mathrm{SNR}_{k,i,j}$ over k i, and j:

$$\mathrm{SNR} = \left\{ \sum_{k=2,3} \sum_{i,j} \frac{1}{\mathrm{SNR}_{k,i,j}} \right\}^{-1} \tag{4.9}$$

As Fig. 4.1 shows, the larger the number of iterations, mostly the higher the performance. SNR of ICA-PLCA converged to about 22 dB at about 10 iterations. However, in the cases that the range of the number of iterations was 1 to 10, SNR of PLCA was about 0 dB. The convergence of PLCA was much slower than that of ICA-PLCA, and SNR of ICA-PLCA converged at about 30 iterations. These results indicate that the performance of ICA-PLCA with 10 iterations is comparable to that of PLCA with 30 iterations. In contrast, as Fig. 4.2 shows, the calculation time of PLCA with 30 iterations was much longer than that of ICA-PLCA with 10 iterations. This indicates that the total calculation time can be reduced to less than 1/2 without loss of performance by reducing the number of iterations of PLCA. Thus, the proposed method can reduce the total calculation time by using ICA. Referring back to the experiment of Chapter

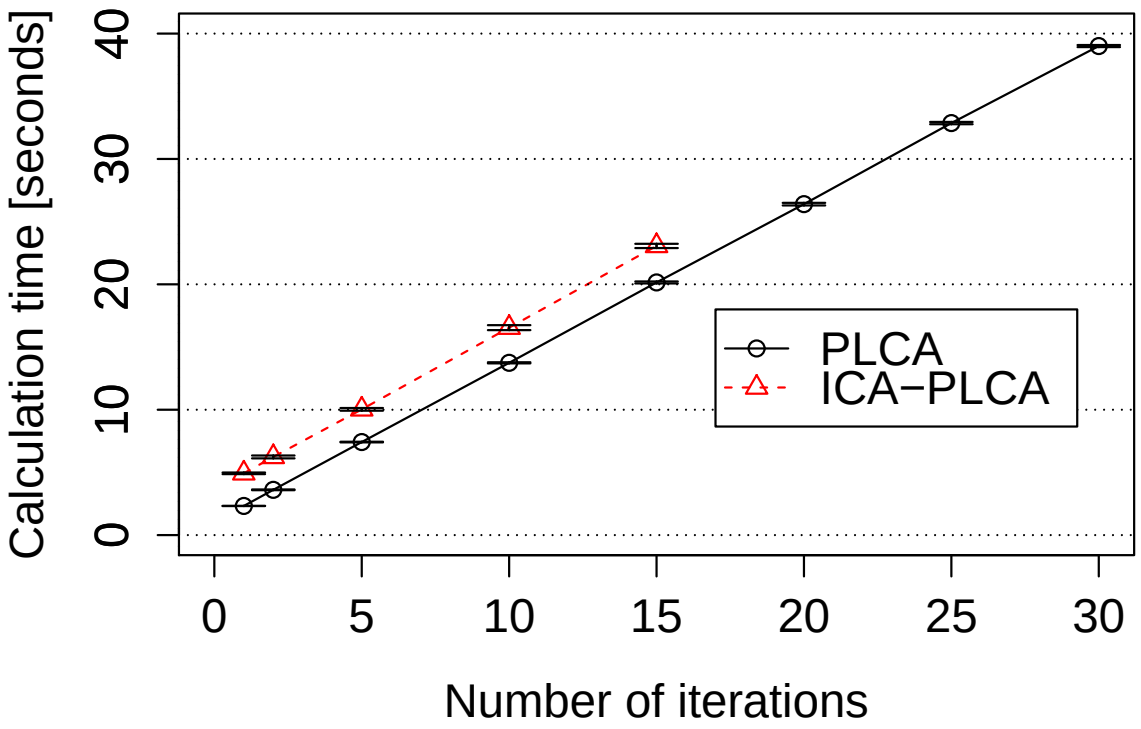


Figure 4.2: The calculation time for each method. X and Y show the number of iterations and the calculation time [second]. Error bars represent 95% confidence intervals.

3, the RTF of CPU was less than 2.0 under the condition of $T = 50$ (c.f. Table 3.1). Combined with the fact that the total calculation time can be reduced to less than 1/2, it is indicated that the proposed method can work in real time even on CPU.

4.5 Relation to prior work

As far as we know, in the domain of mass spectrometry, there are no researches on acceleration of PLCA because, so far, it has not been necessary that signal separation is executed in real-time in the domain of mass spectrometry.

There are several approaches on improving the initialization process of NMF such as PCA/SVD [83][84][85] and fuzzy clustering [86]. Although the orthogonality does not hold in mass spectrometry, PCA and SVD impose orthogonality, and so PCA and SVD are not fit for initialization in mass spectrometry. Fuzzy clustering uses random values, and it suffers from the initial-value-sensitivity. From these points of view, the proposed method is superior to conventional approaches.

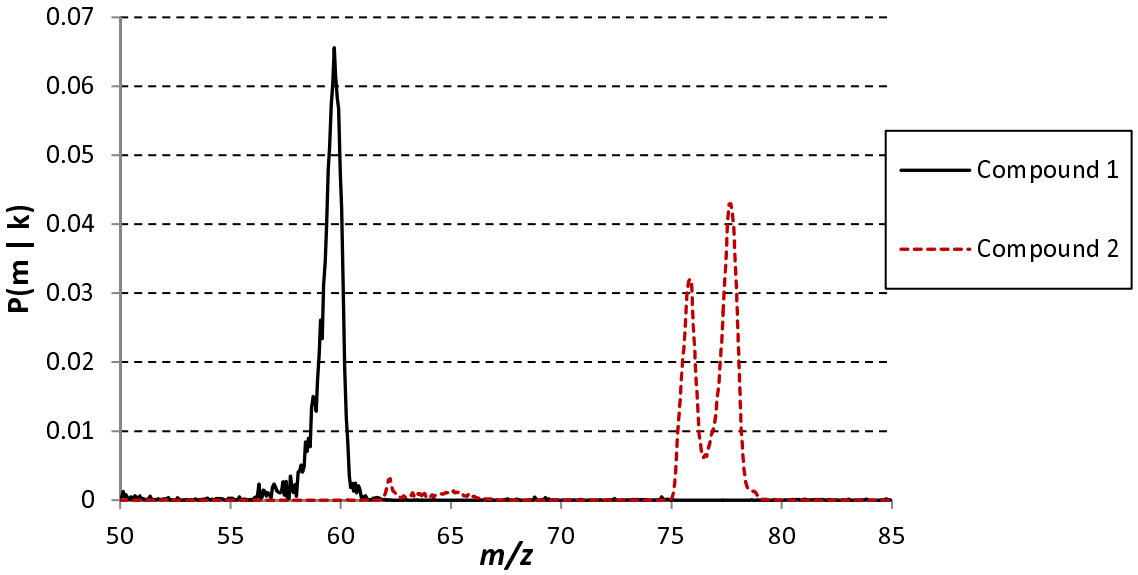


Figure 4.3: Estimates of the probabilistic spectral basis components $P(m|k)$. X and Y show m/z and $P(m|k)$.

4.6 Conclusion

We proposed a new method to separate mass spectra into components of each chemical compound for explosives detection. In order to speed up the conventional method based on PLCA, the proposed method makes use of ICA in the initialization process. In the experiment using the data in a real environment, it was shown that the proposed method can reduce the total calculation time. Referring back to the experiment of Chapter 3, the RTF of CPU was less than 2.0 under the condition of $T = 50$ (c.f. Table 3.1). Combined with the fact that the total calculation time can be reduced to less than 1/2, it is indicated that the proposed method can work in real time even on CPU.

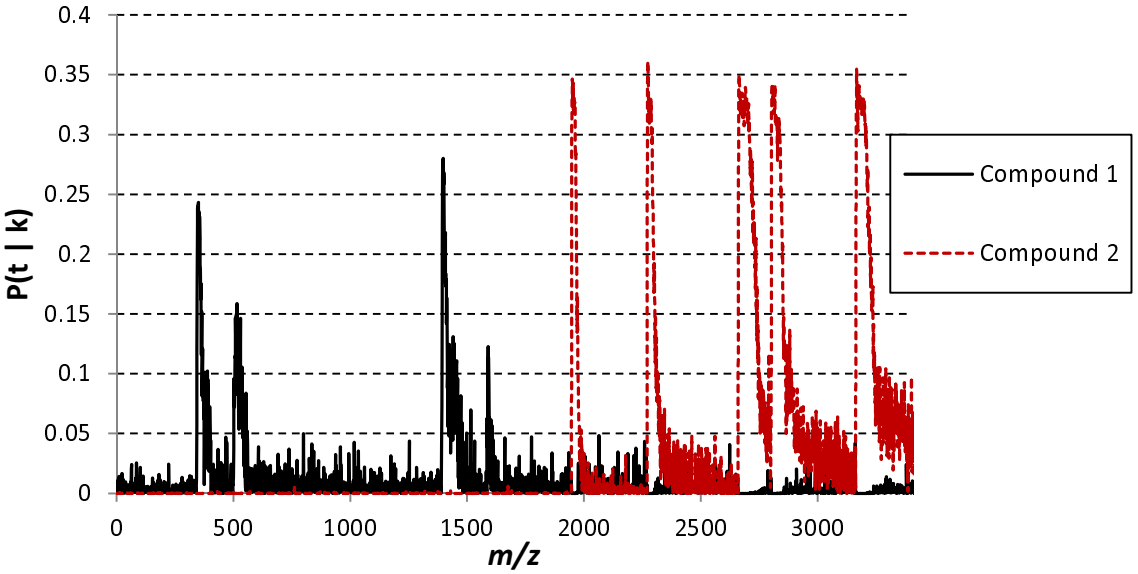


Figure 4.4: Estimates of the probabilistic activations $P(k|t)$. X and Y show t and $P(k|t)$.

Chapter 5

Mass spectra separation for explosives detection by using an attenuation model

5.1 Introduction

In the previous chapters, it was required for a separation algorithm to work with online in real time, and the PLCA-based online algorithm was proposed. However, in some cases, the time delay for detection is allowed, and it is not required for the algorithm to work with online. In this chapter, a separation algorithm for such cases is proposed. The PLCA-based algorithm has the problem that there is a high probability that uncertainty still remains. To reducing the uncertainty, mass spectra separation based on shift-invariant PLCA (SIPLCA) making use of temporal correlation of the mass spectrogram is proposed. In addition, to prevent overfitting, the temporal correlation is modeled with a function representing attenuation by focusing on the fact that the amount of a substance is attenuated continuously and slowly with time. Results of an experimental evaluation of the algorithm with data obtained in a real railway station demonstrate that the proposed algorithm outperforms the PLCA-based algorithm and the simple SIPLCA-based one. Also, an evaluation of the detection performance of explosives detection is demonstrated, and results of the evaluation indicate that the proposed separation algorithm can improve the detection performance.

5.2 Problem statement

Most parts of the problem statement are the same as Chapter 3. The system observes a time series of mass spectra, called a mass spectrogram $x(t, m)$, where t is defined as the time index, and m is defined as the index corresponding to m/z in a mass spectrum. T is defined as the number of the time indices, and M is defined as the number of the indices corresponding to m/z. The set of substances in the air consists of explosive substances, interference substances, and a chemical background. The number of substances that can be observed is set to K. $x(t, m)$ can be modeled as a linear combination of spectral basis components as follows,

$$x(t, m) = \sum_{k=1}^{K} c(k|t)s(m|k), \tag{5.1}$$

where k $(1 \leq k \leq K)$ is defined as the index of substances, $c(k|t)$ is defined as the intensity corresponding to the amount of the k-th substance at the time-point t, and $s(m|k)$ is defined as the m-th element of the spectral basis component corresponding to the k-th substance. $c(k|t)$ is called a temporal activation. The problem is to estimate unknown variables $c(k|t)$ and $s(m|k)$ from known variables $x(t, m)$.

The difference of the problem statement is that it is not required for the separation process to work with online; i.e. the algorithm is allowed to use the signal obtained in the future for estimation. The reason why offline separation is permitted is described in Section 5.4.3.

5.3 PLCA-based mass spectra separation

In this section, the conventional algorithm is clarified. The conventional algorithm is the PLCA-based one, which is proposed in Chapter 3. In PLCA [68], $x(t, m)$ is modeled with the following probability distribution:

$$x(t, m) \propto P(t, m) = P(t) \sum_{k=1}^{K} P(k|t)P(m|k) \tag{5.2}$$

The conventional algorithm can estimate the unknown parameters $P(k|t)$ and $P(m|k)$ from the input signal $x(t, m)$. $P(k|t)$ corresponds to $c(k|t)$ in (5.1), and $P(k|t)$ is called a probabilistic temporal activation. $P(m|k)$ corresponds to $s(m|k)$ in (5.1), and $P(m|k)$

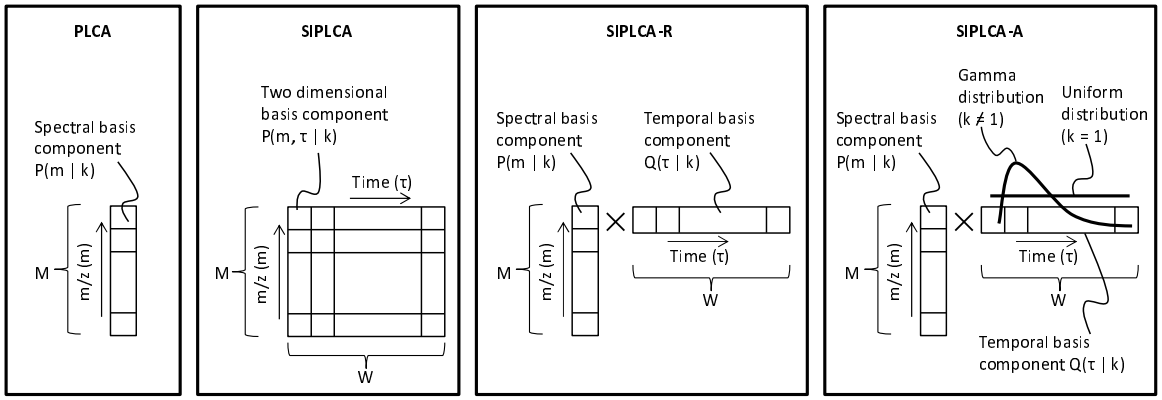


Figure 5.1: Basis components of each algorithm (the PLCA-based algorithm, SIPLCA, SIPLCA-R, and SIPLCA-A).

is called a probabilistic spectral basis component. PLCA is a statistical formulation of NMF with the KL divergence. PLCA is thus suitable for separating a mass spectrogram because all the elements of temporal activations and spectral basis components are non-negative. In addition, the conventional algorithm makes use of sparsity of temporal activations to reduce uncertainty. The conventional algorithm uses the entropic prior proposed by Grindlay and Ellis [78] for modeling the sparsity. The entropic prior is effective for improving the separation performance although it is easy to be implemented. By using the entropic prior, the conventional algorithm can estimate $P(k|t)$ and $P(m|k)$ accurately in many cases. The algorithm, however, fails in some cases because uncertainty can not be resolved only by non-negativity and sparsity constraints.

5.4 Proposed algorithm

5.4.1 Shift-invariant PLCA-based mass spectra separation

To reduce the uncertainty, shift-invariant PLCA (SIPLCA) [87] is introduced into the PLCA-based algorithm in this chapter. In the PLCA-based algorithm, the basis component of one substance is regarded as a one-dimensional probability distribution $P(m|k)$, whereas in SIPLCA, it is regarded as a two-dimensional probability distribution $P(m, \tau|k)$ as Fig. 5.1 shows. Here, $\tau = 1, \cdots, W$ is defined as the time index in the basis component $P(m, \tau|k)$, where W is the frame size of the basis component. With the two-dimensional

basis components, SIPLCA can make use of the fact that the mass spectrogram is correlated temporally. In SIPLCA, it is assumed that the input signal $x(t,m)$ is generated by convolving $P(m,\tau|k)$ over time as follows:

$$\begin{aligned} & x(t,m) \propto P(t,m) \\ = & \sum_k P(k) \sum_\tau P(m,\tau|k) P(t-\tau|k) \qquad (5.3) \\ = & \sum_\tau P(t-\tau) \sum_k P(m,\tau|k) P(k|t-\tau) \qquad (5.4) \end{aligned}$$

As the above equation shows, (5.3) and (5.4) are equivalent. Smaragdis [87] uses the formulation of (5.3). Equation (5.4) can be also used instead of (5.3). In this chapter, (5.4) is used to keep consistency between the conventional PLCA-based algorithm and the proposed algorithm.

From (5.4), an algorithm for estimating $P(k|t)$ and $P(m,\tau|k)$ is obtained as the Algorithm 4 shows. Similarly to the PLCA-based algorithm, $P(k|t)$ for $k=1$ is set to a higher value than $P(k|t)$ for all $k \neq 1$ in (5.7) to map the signal of the stationary chemical background to the first substance, i.e. $k=1$. In the case of $W=1$, the SIPLCA-based algorithm is equivalent to the PLCA-based algorithm.

In the conventional PLCA-based algorithm, the number of the unknown variables $P(k|t)$ and $P(m|k)$ is $(KT+MK)$, whereas in the SIPLCA-based algorithm, the number of the unknown variables $P(k|t)$ and $P(m,\tau|k)$, is $(KT+MWK)$. However, the SIPLCA-based algorithm is prone to overfitting because it has more unknown parameters than the PLCA-based algorithm.

5.4.2 Attenuation model

To prevent overfitting, a model for the temporal correlation is introduced into the SIPLCA-based algorithm in this subsection. When a person with an explosive substance passes through the detector, ions of the substance will be measured within two or three seconds, and the peaks of the basis component corresponding to the substance will rise rapidly in the mass spectrogram. After that, the intensity of the peaks decreases continuously and slowly. In this way, the temporal profile of $P(m,\tau|k)$ varies largely depending on τ, whereas the spectral pattern of $P(m,\tau|k)$ does not vary largely depending on τ. Therefore, for the explosives-detection system, $P(m,\tau|k)$ can be decomposed

into a spectral basis component $P(m|k)$ and a temporal basis component $Q(\tau|k)$, which are mutually independent. The temporal basis component is represented as $Q(\tau|k)$ to distinguish the temporal basis component $Q(\tau|k)$ with the temporal activation $P(k|t)$. $Q(\tau|k)$ is invariant for t, whereas $P(k|t)$ varies depending on t. $P(m,\tau|k)$ is therefore converted to the following equation:

$$P(m,\tau|k) = P(m|k)Q(\tau|k). \tag{5.11}$$

Using (5.11), the estimation process of $P(m,\tau|k)$ in the SIPLCA-based algorithm is decomposed into the estimation process of $P(m|k)$ and that of $Q(\tau|k)$. Here, the new separation algorithm based on this derivation is called "SIPLCA-R" (Algorithm 5). As Fig. 5.1 shows, the basis component of SIPLCA-R can be interpreted as a rank-one matrix, which is the multiplication of $P(m|k)$ and $Q(\tau|k)$.

To enhance robustness, a constraint is imposed on the temporal basis component $Q(\tau|k)$. We focus on the fact that $Q(\tau|k)$ rises rapidly first, and then it decreases continuously and slowly with time after passing through the detector. To represent such temporal structure of attenuation, $Q(\tau|k)$ for each k is modeled by a gamma distribution having parameters different depending on k. $\mathcal{G}(\tau;\theta_k,\phi_k) = \frac{{\phi_k}^{\theta_k}}{\Gamma(\theta_k)}\tau^{\theta_k-1}e^{-\phi_k\tau}$, where $\Gamma(\theta_k)$ is the gamma function, and θ_k and ϕ_k are the parameters of the gamma distribution for k. It is known that θ_k can be estimated by the following update rule [88]:

$$\theta_k\text{new} = \theta_k - \frac{\log(\theta_k) - \Psi(\theta_k) - \log(\overline{Q(\tau|k)}/\widetilde{Q(\tau|k)})}{1/\theta_k - \Psi'(\theta_k)}, \tag{5.14}$$

where $\overline{Q(\tau|k)} = \frac{1}{W}\sum_\tau Q(\tau|k)$, $\widetilde{Q(\tau|k)} = (\prod_\tau Q(\tau|k))^{1/W}$, $\Psi(\cdot)$ is the digamma function, and $\Psi'(\cdot)$ is the trigamma function. ϕ_k can be estimated by

$$\phi_k = \theta_k/\overline{Q(\tau|k)}. \tag{5.15}$$

Therefore, Algorithm 6 can be achieved, and the new separation algorithm based on this derivation is called "SIPLCA-A" here. In (5.16), only for k = 1, $Q(\tau|k)$ is set to the uniform distribution to represent the fact that the intensity of the chemical background does not change in a short time. As Fig. 5.1 shows, the basis component of SIPLCA-A can be interpreted as the multiplication of $P(m|k)$ and $Q(\tau|k)$, where $Q(\tau|k)$ is approximated by the gamma distribution.

It is necessary to pay attention to that the proposed algorithm is offline and that the time delay occurs. In (5.6) of the proposed algorithm, $x(t+\tau, m)$, which is obtained in the future, is used for calculating $\hat{c}(k|t)$, which is past activation. For example, if the frame size of the basis component, W, is 20, the separated signals can be calculated at 19 $(= W - 1)$ time-points after the input signal corresponding to the same time-point is observed. A method for solving the problem of the time delay is explained in Section 5.4.3.

5.4.3 Method for applying the proposed algorithm to the explosives-detection system

In this section, methods for applying the proposed separation algorithm to the explosives-detection system are described.

First, a method for integrating the proposed separation algorithm with the detection process is described. By the proposed separation algorithm, for each k-th substance, the estimated basis $\hat{s}(m|k)$ and the estimated activation $\hat{c}(k|T-W+1)$ are calculated, and the separated spectrum, $\boldsymbol{y}_k = [y(1,k), \cdots, y(M,k)]^T$, is calculated by $y(m,k) = \hat{c}(k|T-W+1)\hat{s}(m|k)$. Because the proposed algorithm is a blind-source-separation approach, it can not be known which k-th separated spectrum corresponds to an explosive substance. Therefore, for each k-th separated spectrum $\boldsymbol{y}_k$, the detection process is performed equally. The detection process compares $\boldsymbol{y}_k$ with the template spectrum for each v-th explosive substance, $\boldsymbol{d}_v = [d(1,v), \cdots, d(M,v)]^T$, and the score corresponding to the v-th explosive substance is calculated. As described in Section 5.1, in the explosive-detection system, the m/z-axis may shift due to the condition of the mass spectrometer. To solve the problem of the shift of the m/z-axis, $a(v,k)$ is calculated as follows:

$$a(v,k) = \max_{\delta \in [-\Delta, \Delta]} \sum_m d(m,v) y(m+\delta, k), \tag{5.17}$$

where Δ is the tolerance range for the shift, and δ is an integer between $-\Delta$ and Δ. If $a(v,k)$ is larger than a threshold for any v and k, then an alert is displayed as the detection result.

Next, a method for solving the problem of the time delay is described. As described in Section 5.4.2, the proposed algorithm is offline, and the time delay occurs. To solve

the problem of the delay for detection, it is required to integrate another sub-system for tracking passengers, e.g. a video surveillance system. We have proposed a system for tracking passengers by using distributed surveillance cameras [89]. The tracking system observes the images of passengers passing through the explosives-detection system, and the tracking system searches for the images observed by the distributed surveillance cameras by using the images of passengers as query, and the tracking system performs tracking passengers after passing through the explosives-detection system. Therefore, it can be considered that the problem of the time delay due to the proposed algorithm can be solved by integrating the tracking system with the explosives-detection system.

5.5 Experimental results

5.5.1 Evaluation of separation performance

The input mass spectrogram is the same as that of the experiment of Chapter 3. The separation performance of the proposed algorithm (SIPLCA-A) was compared with that of the PLCA-based algorithm, SIPLCA, and SIPLCA-R. In the case of $W = 1$, SIPLCA, SIPLCA-R, and SIPLCA-A are equivalent to the PLCA-based algorithm. In the estimation process, all the unknown parameters were initialized by random values. Under each condition, the estimation process was run twenty times. The number of substances (K) in the estimation process was set at eight. The sparsity parameter β was 1.01. Separation performance was measured by $\mathrm{SNR}_{k,i,j}$ as follows:

$$\mathrm{SNR}_{k,i,j} = 10 \log_{10} \frac{\max_{t \in \mathcal{A}_{k,i}} |\hat{c}(k|t)_j|}{\sqrt{\frac{1}{|\mathcal{N}_k|} \sum_{t \in \mathcal{N}_k} |\hat{c}(k|t)_j|^2}} \ [\mathrm{dB}]\,, \tag{5.18}$$

where $\mathcal{A}_{k,i}$ is the temporal span around the i-th time when the k-th substance is passed through the system, i.e. $\mathcal{N}_{k=2}$ is $[2000, 3500]$, and $\mathcal{N}_{k=3}$ is $[0, 1500]$, and j $(1 \leq j \leq 20)$ is the index of executions. In addition, SNR is defined as an ensemble mean over k, i, and j. In the case of the arithmetic mean, SNR is prone to an excessively high value even if there is a low $\mathrm{SNR}_{k,i,j}$. Therefore, the arithmetic mean is not fit to measure the robustness. In contrast, the harmonic mean is sensitive to the case that $\mathrm{SNR}_{k,i,j}$ is low. Therefore, the harmonic mean is fit to measure the robustness. SNR is thus defined as

the harmonic mean of $\mathrm{SNR}_{k,i,j}$:

$$\begin{aligned}\mathrm{SNR} &= 2 \times \#\text{passing trials} \times \#\text{process executions} \\ &\times \left\{ \sum_{k=2,3} \sum_{i,j} \frac{1}{\mathrm{SNR}_{k,i,j}} \right\}^{-1}. \end{aligned} \tag{5.19}$$

Fig. 5.2 shows estimates of the temporal activations $P(k|t)$ and estimates of the spectral basis components $P(m|k)$ respectively. As shown in Fig. 5.2 (a), the peaks exist at the correct times when substance 1 and substance 2 were passed. As shown in Fig. 5.2 (b), the correct main peaks for substance 1 and substance 2 were estimated. The results shown in Fig. 5.2 indicate that the proposed algorithm can separate explosive substances precisely. SNR is shown in Fig. 5.3. As shown in Fig. 5.3, the longer the frame size W is, mostly the higher the separation performance is. The separation performances of SIPLCA, SIPLCA-R, and SIPLCA-A for $W = 20$ were higher than that of the PLCA-based algorithm (corresponding to $W = 1$). The results shown in Fig. 5.3 indicate that the two-dimensional basis component is effective for making use of the temporal correlation of the mass spectrogram. For $W = 1$ to 5, the separation performances of each version are not significantly different, whereas for $W = 20$, SNR of SIPLCA-A is higher than those of the other versions at about 4dB. These results indicate that SIPLCA and SIPLCA-R are prone to overfitting, and that SIPLCA-A can prevent overfitting successfully by using the attenuation model. In relation to the results, the temporal basis components for substance 1 and substance 2 estimated by SIPLCA-R are shown in Fig. 5.4. In this experiment, substance 1 was split into two basis components ($k = 2$ and $k = 4$), and substance 2 was also split into two basis components ($k = 3$ and $k = 5$). As shown in Fig. 5.4, the temporal basis components estimated by SIPLCA-R did not follow the attenuation model described in Section 5.4.2, i.e. it is clear that the temporal basis components were not estimated precisely. The results shown in Fig. 5.4 indicate that SIPLCA-R suffered from overfitting.

5.5.2 Evaluation of detection performance

The performance of detection using the proposed separation algorithm was experimentally evaluated. In this experiment, both "negative data" and "positive data" are required. The negative data are signals containing only noise components, and the positive

data are signals containing explosive components and noise components. For obtaining the negative data, first, signal in a real environment was recorded during a day when consenting volunteers passed through the system. Then, the recorded signal was segmented into 351 signals corresponding to individual subjects by using a infrared passenger sensor. The time range of each segmented signals is between 20 seconds before passing and 10 seconds after passing, i.e. the length of each segmented signal is 30 seconds ($T = 45$). We assumed that the segmented signals do not contain explosive components and that the segmented signals contain only noise components, and the 351 segmented signals were used as the negative data. To obtain the positive data, first, 18 signals containing explosive components were recorded by measuring 0.05μg particles of an explosive substance 18 times in a laboratory environment. In this experiment, a lot of explosive signals can not be recorded. To obtain 1000 explosive signals, the Monte Carlo method was used based on the 18 recorded explosive signals. For each trial, a signal was selected from the 18 recorded explosive signals, and the simulated signal was obtained multiplying the selected signal by a random variable generated from a log-normal distribution, which had been fit to the samples of the maximum intensity of the recorded explosive signals in advance. Then, for each simulated explosive signal, a signal was selected from the negative data at random, and the simulated explosive signal was mixed with the selected negative signal. The 1000 mixed signals were regarded as the positive data. In this experiment, the number of the explosive substances registered in the database was 1, i.e. only the template spectrum for $v = 1$ was used. The explosive substance was detected by checking whether $a(v = 1, k)$ calculated by (5.17) is larger than a threshold value or not. If $a(v = 1, k)$ is larger than the threshold, the input signal is recognized as positive. If not so, the input signal is recognized as negative. Δ was set to the integer corresponding to 0.5 atomic mass units. The detection performance was evaluated by the Receiver Operating Characteristic (ROC) curves. The ROC curves evaluate the false alarm rate and the misdetection rate simultaneously. The false alarm rate is defined by:

$$\text{False alarm rate} = \frac{\#\text{the negative data recognized as positive}}{\#\text{all the negative data}}, \tag{5.20}$$

The misdetection rate is defined by:

$$\begin{aligned} &\text{Misdetection rate} \\ &= \frac{\#\text{the positive data recognized as negative}}{\#\text{all the positive data}}, \end{aligned} \tag{5.21}$$

The detection performance of the proposed algorithm (SIPLCA-A) was compared with that of the case without separation and that of the PLCA-based algorithm. In both SIPLCA-A and the PLCA-based algorithm, T was 45, M was 512, K was 8, and β was 1.01. W was set to 5 in SIPLCA-A. Fig. 5.5 shows the ROC curves for each algorithm. The detection performance of the PLCA-based algorithm was higher than that of the case without separation. Moreover, the detection performance of SIPLCA-A was higher than that of the case the PLCA-based algorithm. These results in Fig. 5.5 indicate that SIPLCA-A is effective for improving the detection performance of the explosives-detection system.

5.6 Relation to prior work

The proposed algorithm is strongly related to non-negative matrix factor deconvolution (NMFD) [90]. NMFD is a convolutional version of NMF and is capable of separating the signal into two-dimensional components with temporal structure [90]. In the point that temporal structure is modeled by two-dimensional components, NMFD is similar to the SIPLCA-based algorithm, and it can be considered that NMFD also suffers from overfitting similarly to the SIPLCA-based algorithm. However, SIPLCA-A has two different points from NMFD for preventing overfitting. The first is that two-dimensional components are decomposed to a spectral basis component and a temporal basis component as SIPLCA-R, and the second is that the constraint of gamma distributions as SIPLCA-A is imposed on temporal basis components.

Some researches on methods using "temporal continuity" have been done [91] [92]. Virtanen [91] proposed a method finding a solution such that the differences between adjacent temporal activations are small. Bertin *et al.* [92] proposed a method finding a solution such that temporal activations are generated by an inverse-gamma-Markov-chain. These conventional approaches impose constraints only on a relationship between adjacent temporal activations, and it can be considered that temporal activations may

change more freely than SIPLCA-A. In musical-signal processing, temporal envelopes may change largely, and it is expected that these approaches are suitable. In contrast, mass spectrometers can be modeled by a linear time-invariant system, and it can be considered that the mass spectrogram is generated by convolution. SIPLCA-A is therefore more reasonable than the conventional approaches.

5.7 Conclusion

A new algorithm for separating a mass spectrogram into individual substance was proposed for explosives detection. To resolve the uncertainty of the conventional algorithm, the proposed algorithm is based on SIPLCA utilizing the temporal correlation of the mass spectrogram. Moreover, to prevent overfitting, the temporal correlation is modeled with the Gamma distribution for utilizing the fact that the amount of a substance is attenuated continuously and slowly with time. In an experimental evaluation of the algorithm using data obtained from a real environment, it was shown that the proposed algorithm (SIPLCA-A) outperforms the conventional PLCA-based one and the other SIPLCA-based versions (SIPLCA and SIPLCA-R). It was also shown that the detection performance is improved by the proposed separation algorithm.

Algorithm 4 SIPLCA-based algorithm (SIPLCA)

1. Initialization process

Set all the unknown parameters to random values.

2. Iteration process

Iterate the following E step and M step until convergence

or a maximum number of iterations is reached.

E step:

Compute $P(k, \tau|t, m)$ as follows:

$$P(k, \tau|t, m) = \frac{P(t-\tau)P(k|t-\tau)P(m, \tau|k)}{\sum_{k',\tau'} P(t-\tau')P(k'|t-\tau')P(m, \tau'|k')}. \tag{5.5}$$

M step:

Compute $\hat{c}(k|t)$ as follows:

$$\hat{c}(k|t) = \sum_{m,\tau} x(t+\tau, m)P(k, \tau|t+\tau, m). \tag{5.6}$$

Compute $P(k|t)$ as follows:

$$P(k|t) = \begin{cases} \frac{1}{1+\sum_{k'\neq 1} g(\beta, \{\hat{c}(k'|t)\}_k)} & \text{if } k = 1, \\ \frac{g(\beta_{\mathfrak{a}}, \hat{c}(k|t))}{1+\sum_{k'\neq 1} g(\beta, \{\hat{c}(k'|t)\}_k)} & \text{otherwise,} \end{cases} \tag{5.7}$$

where $g(\beta, \{\gamma_i\}_i)$ is the entropic prior of Grindlay and Ellis [78]:

$$g(\beta, \{\gamma_i\}_i) = \frac{{\gamma_i}^\beta}{\sum_i {\gamma_i}^\beta}. \tag{5.8}$$

Compute $P(m, \tau|k)$ as follows:

$$P(m, \tau|k) = \frac{\sum_t x(t, m)P(k, \tau|t, m)}{\sum_{m,\tau,t} x(t, m)P(k, \tau|t, m)}. \tag{5.9}$$

Compute $P(t)$ as follows:

$$P(t) = \frac{\sum_{k,\tau,m} x(t+\tau, m)P(k, \tau|t+\tau, m)}{\sum_{t,k,\tau,m} x(t+\tau, m)P(k, \tau|t+\tau, m)}. \tag{5.10}$$

Algorithm 5 SIPLCA-based algorithm with a rank-one model (SIPLCA-R)

1. Initialization process

Set all the unknown parameters to random values.

2. Iteration process

Iterate the following E step and M step until convergence

or a maximum number of iterations is reached.

E step:

Compute (5.5).

M step:

Compute (5.6).

Compute (5.7).

Compute $P(m|k)$ as follows:

$$P(m|k) = \frac{\sum_{t,\tau} x(t,m)P(k,\tau|t,m)}{\sum_{m,t,\tau} x(t,m)P(k,\tau|t,m)}. \tag{5.12}$$

Compute $Q(\tau|k)$ as follows:

$$Q(\tau|k) = \frac{\sum_{t,m} x(t,m)P(k,\tau|t,m)}{\sum_{\tau',t,m} x(t,m)P(k,\tau'|t,m)}. \tag{5.13}$$

Compute (5.11).

Compute (5.10).

Algorithm 6 SIPLCA-based algorithm with an attenuation model

1. Initialization process

Set all the unknown parameters to random values.

2. Iteration process

Iterate the following E step and M step until convergence

or a maximum number of iterations is reached.

E step:

Compute (5.5).

M step:

Compute (5.6).

Compute (5.7).

Compute (5.12).

Compute (5.13).

Estimate θ_k for each k by computing (5.14) until convergence.

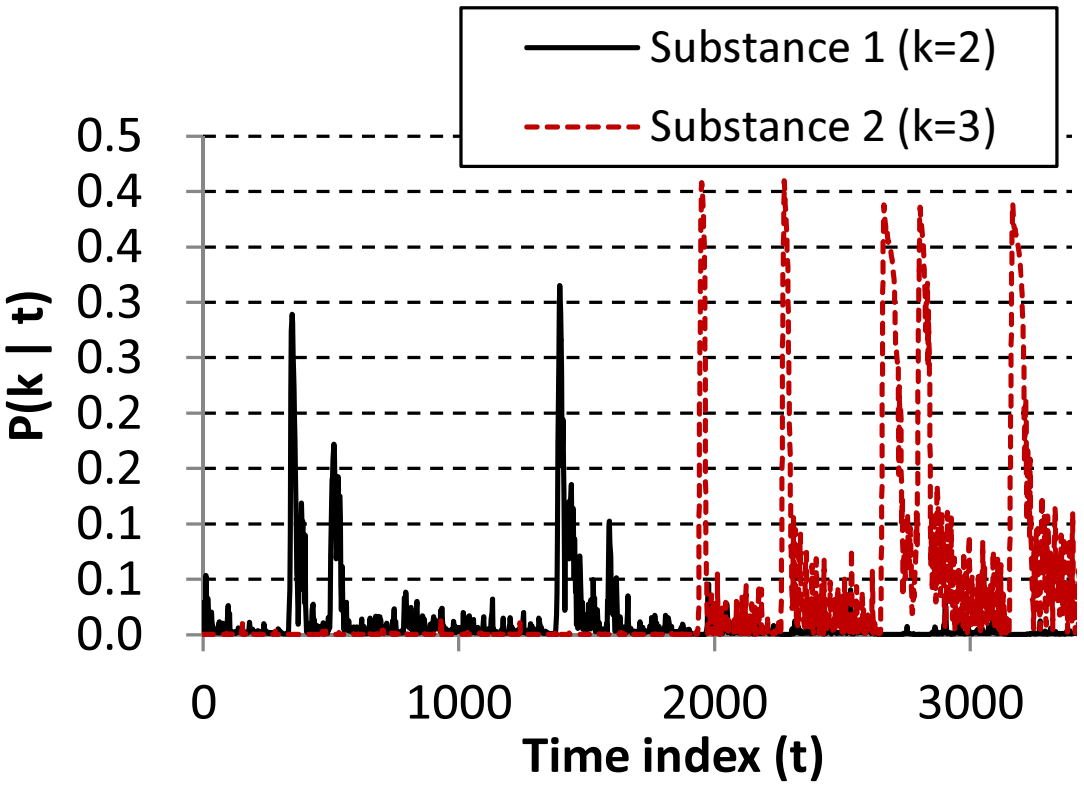
Estimate ϕ_k for each k by computing (5.15).

Update $Q(\tau|k)$ as follows:

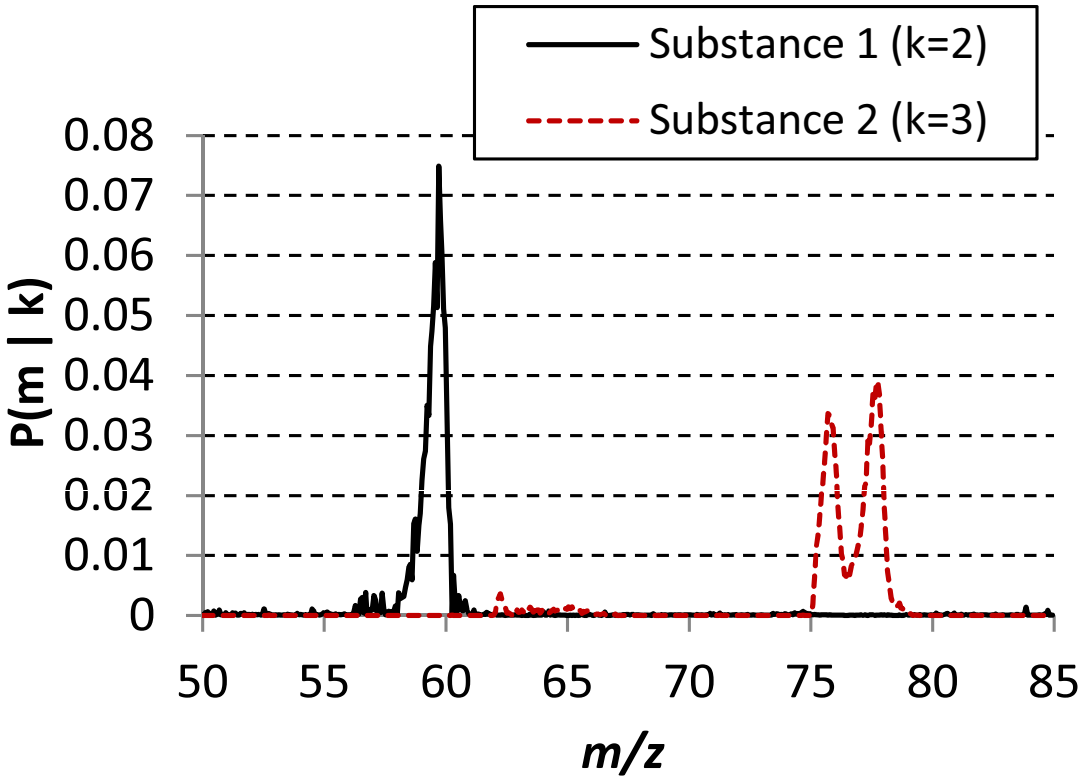
$$Q(\tau|k) = \begin{cases} \frac{1}{W} & \text{if } k = 1, \\ \mathcal{G}(\tau; \theta_k, \phi_k) & \text{otherwise.} \end{cases} \tag{5.16}$$

Compute (5.11).

Compute (5.10).



(a) Temporal activations $P(k|t)$.



(b) Spectral basis components $P(m|k)$.

Figure 5.2: Estimates for substance 1 ($k = 2$, black) and substance 2 ($k = 3$, red).

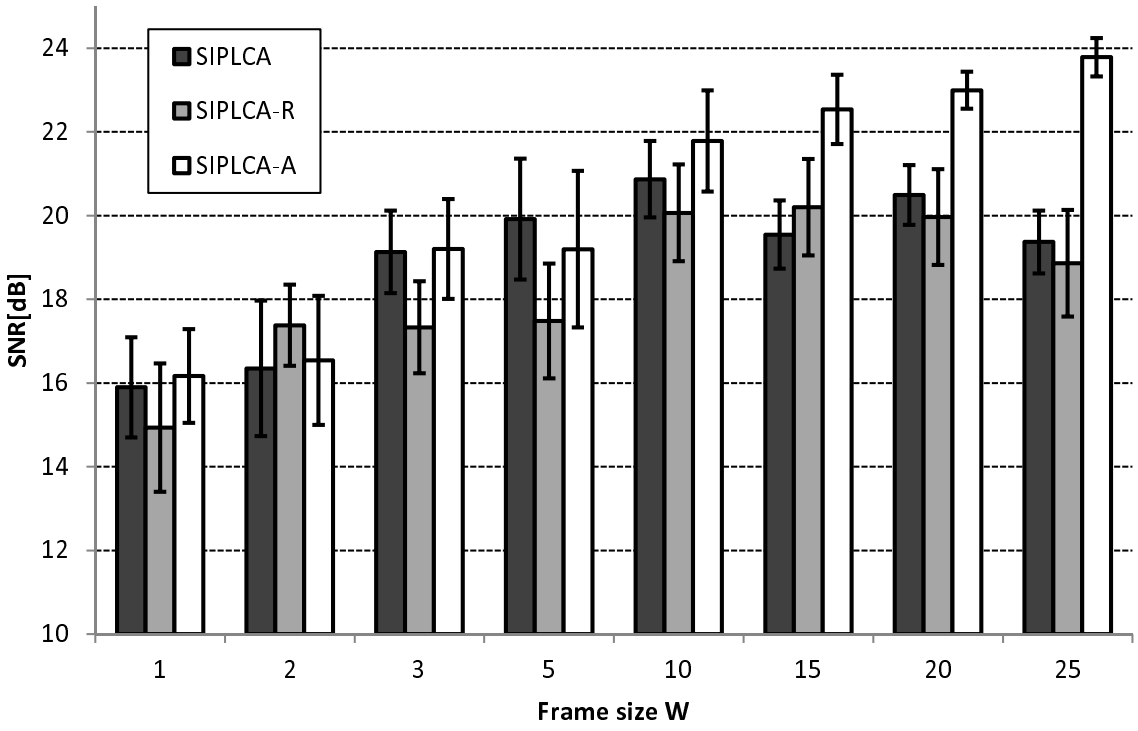
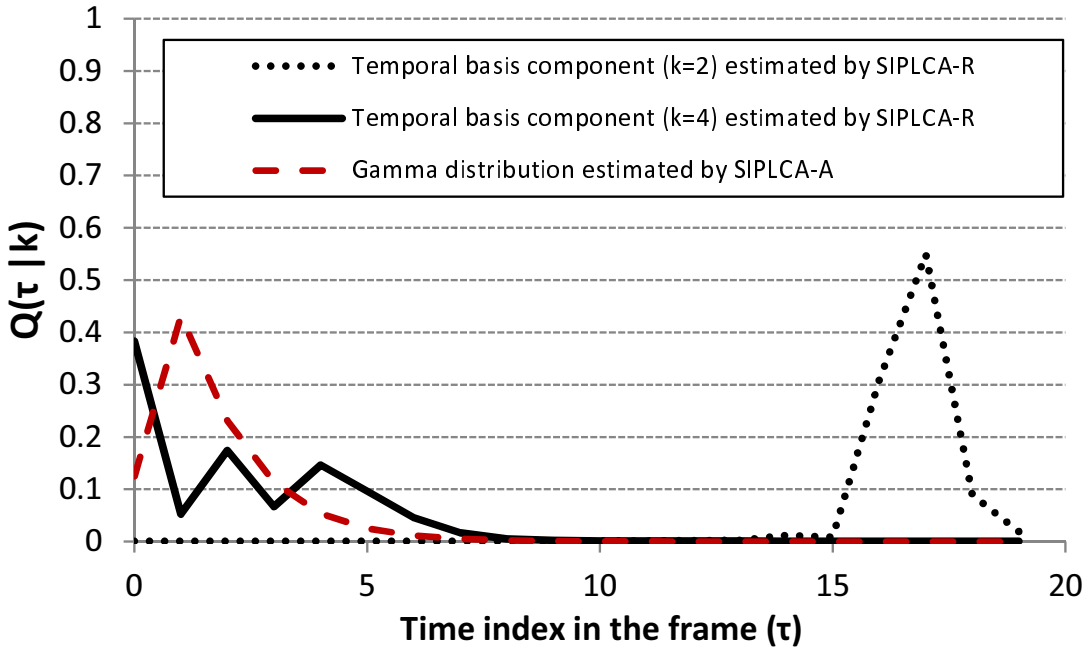
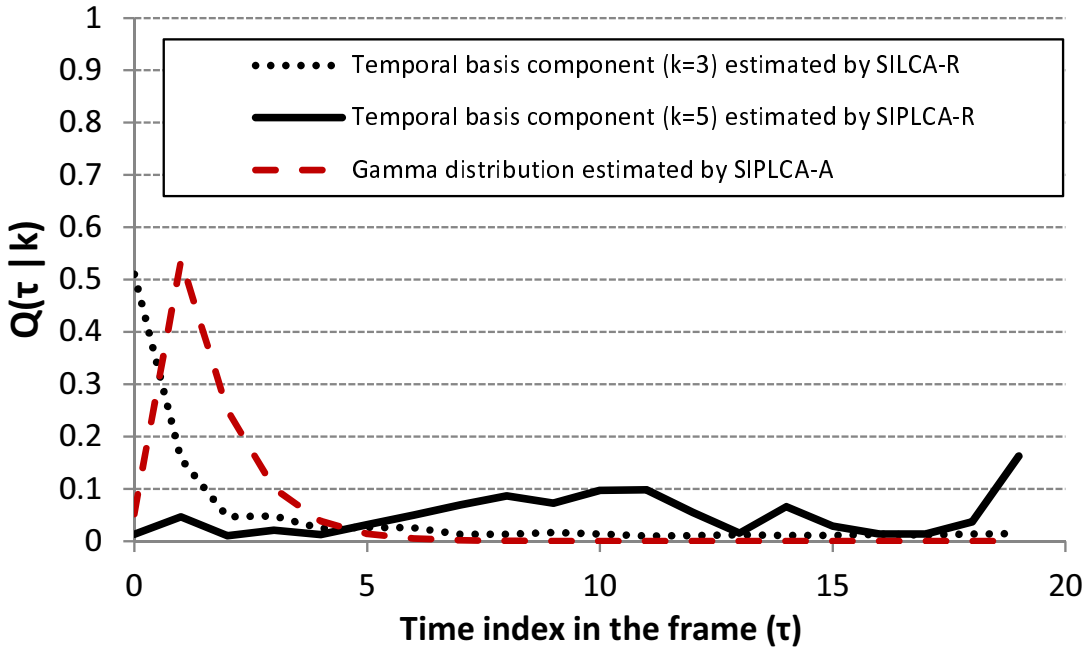


Figure 5.3: SNR for each algorithm. X and Y show the frame size of the basis component W and SNR [dB]. Error bars represent 95% confidence intervals.



(a) Substance 1.



(b) Substance 2.

Figure 5.4: Splitted temporal basis components estimated by SIPLCA-R for $W = 20$. X and Y show τ and $Q(\tau|k)$. The broken red lines mean the temporal basis components by SIPLCA-A.

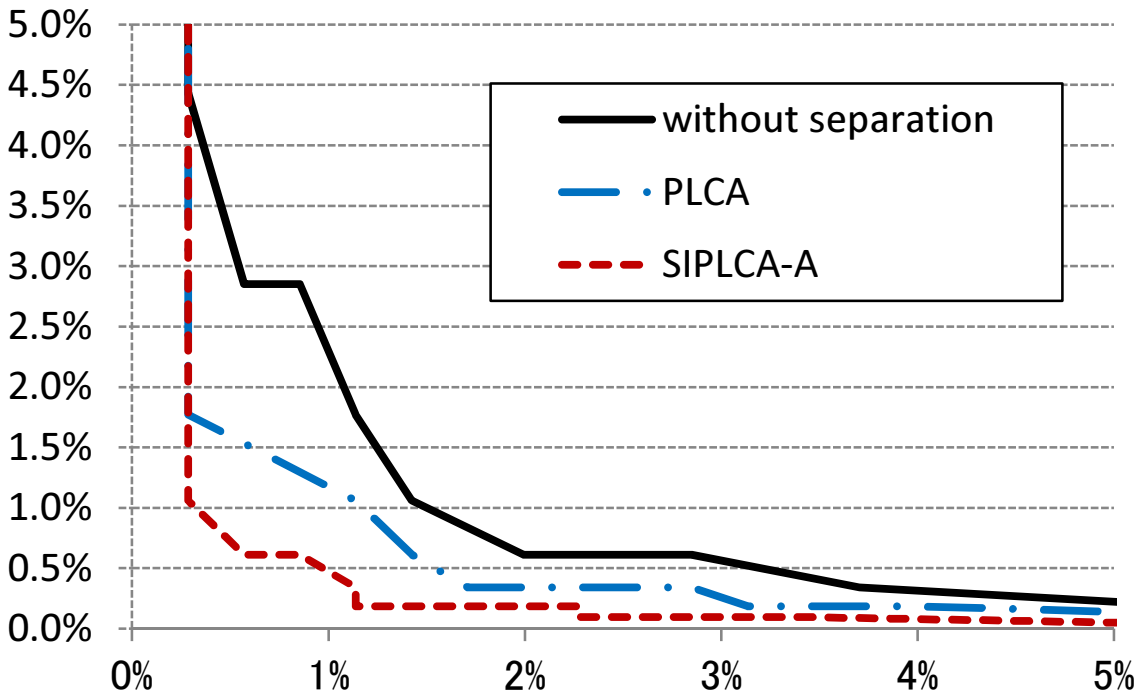


Figure 5.5: ROC curve for each algorithm. X and Y show the false alarm rate and the misdetection rate. "without separation" means the case without separation, "PLCA" means the method of Chapter 3 (Algorithm 2), and "SIPLCA-A" means the proposed algorithm (Algorithm 6).

Chapter 6

Adaptive Boolean compressive sensing for large-area-monitoring

6.1 Introduction

In previous chapters, signal processing techniques for a single explosive detector were proposed; however, a single explosive detector can monitor only limited spaces such as an entrance of a high security area. For monitoring all over the area, there is a strong need to provide a large-area-monitoring-type system. In the large-area-monitoring-type, a function for finding the location of chemicals is required because, if the location of chemicals is known, it will be possible to evacuate people and to capture a suspect. For localization, high speed is especially the most important. In this chapter, an approach for high-speed chemicals localization is proposed.

There are two main approaches of chemicals localization. One is to search for chemicals by using an autonomous mobile robot equipped with a sensor [29], and the other is to estimate the location of chemicals by using a large number of densely arranged sensors [27]. For both of them, a large amount of sensing cost is required, so there are few examples of practical use. However, there is another approach for reducing the sensing cost [35]. This approach uses only a single detector and multiple air-intake-ducts corresponding to the positions all over the area, and all the ducts are connected with the detector. The system can switch the ON/OFF of intake of individual ducts, and at each time-frame, only one duct is selected and is set to ON. The system sequentially

selects a duct and takes air from only the selected duct corresponding to the time-frame. Because there is a one to one correspondence between the ducts and the time-frames, The position of the duct corresponding to the time-frame when a target substance is detected can be determined as the position the substance exists at. We call this approach "time-division-sampling". However, time-division-sampling spends a long time to test all the positions, and time-division-sampling is not fit for applications such as evacuation guidance.

To shorten the measurement time of time-division-sampling, we have proposed a "compressive sensing"-based approach [93]. This approach uses a hardware structure, which are similar to time-division-sampling, consists of a single detector and the multiple air-intake-ducts corresponding to the positions all over the area. This approach takes the mixed air from a combination of multiple ducts at each time-frame unlike time-division-sampling, and it switches the combination of active ducts temporally, and it estimates the location of chemicals from time series of observations. Using the sparsity assumption that a target substance exists at only a small number of positions, it becomes possible to localize chemicals from a small number of observations, i.e. the measurement time can be shortened. Actually, such a hardware structure can not have sufficient quantitativeness, and we assume that it can observe only existence or absence of explosives, i.e. positive/negative (1/0) by one observation. The localization result is also the vector of existence, the elements of which are positive/negative. This formulation is a Boolean version of compressive sensing, which is called group-testing.

Group-testing is a well-known approach for discovering a sparse subset of positive elements in a large set of elements by using a small number of tests. In group-testing, each test consists of three processing steps: (1) selecting elements for a pool on the basis of a certain method, (2) mixing the selected elements into the pool, and (3) observing a single Boolean result by testing the pool. When the proportion of positive elements is small, a small number of the tests on the mixed pool are sufficient to detect the positive elements; that is, all the elements need not be tested directly. Group-testing dates back to the work of Dorfman [94] in 1943, during the Second World War. Dorfman developed this approach in order to test blood for detecting sick soldiers. Nowadays, it is commonly known that group-testing has a lot of applications such as blood screening,

deoxyribonucleic acid screening, and anomaly detection in computer networks [95].

Traditionally, group-testing has been regarded as a combinatorial problem. As for this problem, many researches about the upper and lower bounds on the number of tests required to find all the positive elements have been done. A set of information-theoretic bounds for group-testing with random mixing was established by Malyutov [96, 97], Atia and Saligrama [98], Sejdinovic and Johnson [99], and Aldridge *et al.* [100]. In addition, several tractable approximation algorithms, such as one based on belief propagation [99] and one based on matching pursuit [101], have been proposed.

In recent years, group-testing has drawn interest from the active research area of compressive sensing. Compressive sensing solves a kind of underdetermined linear equation, namely, $\boldsymbol{y} = \boldsymbol{A}\boldsymbol{x}$, where $\boldsymbol{x}$ is an unknown high-dimensional vector to be estimated, $\boldsymbol{A}$ is a given mixing matrix, and $\boldsymbol{y}$ is a given low-dimensional observed vector. The problem with compressive sensing is similar to that with group-testing from the viewpoint that both of them are underdetermined problems such that an unknown high-dimensional vector is decoded from an observed low-dimensional vector. However, while compressive sensing is defined in a real vector space, group-testing is defined in a Boolean vector space. To improve the performance of group-testing by using compressive sensing, Malioutov and Malyutov [2] proposed a method for converting group-testing into compressive sensing through linear-programming relaxation. As for this conversion method, ℓ_1 minimization imposes the sparsity constraint to the solution and solves the uncertainty of the underdetermined problem. It thus outperforms other methods (i.e. the method based on belief propagation [99], the method based on matching pursuit [101], etc.). However, the method based on linear-programming relaxation is defined in non-adaptive group-testing, which cannot choose the pool adaptively based on observation data. In particular, the optimal size of the pool depends on the number of positive elements, and the number of positive elements is unknown; therefore, in the case that Malioutov's method is applied, a larger number of tests are required when the pool-size is not optimal.

To reduce the number of tests of Malioutov's method, a method for adaptive group-testing is proposed here. The proposed method controls the pool-size for each test. The criterion of the control is the expected information gain that can be calculated from the ℓ_0 norm of the estimated solution. Simulation results indicate that the proposed

method outperforms the conventional method even under the condition that the number of positive elements is varied and the number of positive elements is unknown.

6.2 Problem statement

To state the problem, first, the following notation is fixed. N is the number of elements, of which a subset of size K is positive. Non-positive elements are called "negative". $x_n = 1$ indicates that the n-th element is positive, and $x_n = 0$ indicates that the n-th element is negative. For convenience, $\boldsymbol{x} = [x_1, x_2, \cdots, x_N]^T$ is written. T tests, where $T < N$, are then performed. As explained above, in each test, g elements are selected from all the elements, and they are mixed into the same pool. g is called the pool-size. This selection is defined by a mixing matrix, $\boldsymbol{A}$, which is a $T \times N$ binary matrix. The element of the t-th row and the n-th column of $\boldsymbol{A}$ is given as $a_{t,n}$, where $a_{t,n} = 1$ indicates that the n-th element is mixed into the pool of the t-th test, and $a_{t,n} = 0$ indicates that the n-th element is not mixed into the pool of the t-th test. The observed signal of each test, t, is a single Boolean value, $y_t \in \{0, 1\}$. y_t is obtained by taking the Boolean sum (the OR operation) of $\{x_n | a_{t,n} = 1\}$. Thus, the observation model is represented by

$$y_t = \bigvee_{n=1}^{N} (a_{t,n} \wedge x_n) . \tag{6.1}$$

For convenience, $\boldsymbol{y} = [y_1, y_2, \cdots, y_T]^T$ is written. The vector notation

$$\boldsymbol{y} = \boldsymbol{A}\boldsymbol{x} \tag{6.2}$$

is used in the following.

The problem of group-testing is to estimate unknown vector $\boldsymbol{x}$ from given $\boldsymbol{A}$ and $\boldsymbol{y}$. In addition, the noise of the observation is considered. The noise includes both the false positive and the false negative. The former represents the case that $y_t = 1$ even when the Boolean sum (the OR operation) of $\{x_n | a_{t,n} = 1\}$ is 0. The latter represents the case that $y_t = 0$ even when the Boolean sum (the OR operation) of $\{x_n | a_{t,n} = 1\}$ is 1. The observation with noise is represented by

$$\boldsymbol{y} = (\boldsymbol{A} \vee \boldsymbol{x}) \otimes \boldsymbol{v}, \tag{6.3}$$

where $\boldsymbol{v}$ is the Boolean vector of errors, and $\otimes$ means the XOR operation.

A number of works have studied the design of $\boldsymbol{A}$ [95]. For example, K-separating and K-disjunct are well-known properties of $\boldsymbol{A}$. When these properties hold, $\boldsymbol{x}$ can be recovered exactly. However, such design is often unsuitable for practical situations because it assumes that the exact number of the positive elements (K) is known before group-testing. Moreover, if all T tests cannot be carried out, the performance of the method will not be guaranteed [100]. Therefore, in many works, the simple-random-sampling-design and the Bernoulli-random-design is used for design of $\boldsymbol{A}$. In the simple-random-sampling-design, each element of $\boldsymbol{A}$ is randomly generated subject to the constraint $\sum_n a_{t,n} = g$ for all t. In the Bernoulli-random-design, each element of $\boldsymbol{A}$ is randomly generated with a probability p; that is, $a_{t,n}$ is 1 with probability p, and $a_{t,n}$ is 0 with probability $1-p$. The simple-random-sampling-design is used in this study.

There are two types of group-testing, i.e. one is called non-adaptive group-testing, and the other is adaptive group-testing. In non-adaptive group-testing, $\boldsymbol{A}$ is designed in advance and can not be changed dependently on test results $\boldsymbol{y}$. In contrast, adaptive group-testing can control $\boldsymbol{A}$ adaptively to $\boldsymbol{y}$. One of the problems of non-adaptive group-testing is that, although the number of positive elements K is unknown, the optimal pool-size largely depends on K. To solve this problem, the present study thus focuses on adaptive group-testing. In addition, because the simple-random-sampling-design is assumed, control of g is focused on. Therefore, the next mixing vector after the T-th test $\boldsymbol{a}_{T+1}$ is determined by the following random trial:

$$\Pr\left[\boldsymbol{a}_{T+1}\right] = \begin{cases} 1/\begin{pmatrix} N \\ g \end{pmatrix} & \text{if } \|\boldsymbol{a}_{T+1}\|_0 = g, \\ 0 & \text{otherwise,} \end{cases} \tag{6.4}$$

where g is determined by

$$g = F_T\left(\boldsymbol{A}, \boldsymbol{y}\right), \tag{6.5}$$

and F_T is a function of $\boldsymbol{A}$ and $\boldsymbol{y}$. In Section 6.4, a new function F_T is proposed.

6.3 Boolean compressive sensing for group-testing

6.3.1 Compressive sensing

Malioutov and Malyutov [2] proposed a conversion of group-testing into compressive sensing through a linear-programming relaxation. This conventional method is the basis of our method, which is explained in this section.

Many works on compressive sensing have been reported [102]. In this study, a sparse signal, $\boldsymbol{x} \in \mathbb{R}^N$, is assumed, and it is estimated from M measurements $\boldsymbol{y} \in \mathbb{R}^T$ by using a random measurement matrix $\boldsymbol{A}$, where $M < N$. Compressive sensing, namely, decoding $\boldsymbol{x}$, uses the following ℓ_0 minimization:

$$\min_{\boldsymbol{x}} \|\boldsymbol{x}\|_0 \quad \text{subject to} \quad \boldsymbol{y} = \boldsymbol{A}\boldsymbol{x}. \tag{6.6}$$

However, Eq. (6.6) is a NP-hard problem, which cannot be solved practically. Candes *et al.* [102] proved that if certain conditions hold, $\boldsymbol{x}$ can be decoded exactly by the following ℓ_1 minimization:

$$\min_{\boldsymbol{x}} \|\boldsymbol{x}\|_1 \quad \text{subject to} \quad \boldsymbol{y} = \boldsymbol{A}\boldsymbol{x}. \tag{6.7}$$

Since ℓ_1 minimization is a simple linear-programming problem, a number of practicable algorithms can be used to solve it.

6.3.2 Noise-free case

Equation (6.2) is similar to constraint equation (6.7). However, it is not a linear equation in a real vector space but a Boolean equation. It is shown in [2] that (6.2) can be replaced with a closely related linear formulation: $\mathbf{1} \leq \boldsymbol{A}_{\mathcal{I}}\boldsymbol{x}$, and $\mathbf{0} = \boldsymbol{A}_{\mathcal{J}}\boldsymbol{x}$, where $\mathcal{I} = \{t|y_t = 1\}$ is the set of the tests that obtain positive results, and $\mathcal{J} = \{t|y_t = 0\}$ is the set of the tests that obtain negative results. A linear-programming formulation similar to Eq. (6.7) is therefore given as

$$\begin{aligned} &\min_{\boldsymbol{x}} \left\{ \sum_n x_n \right\} \\ &\text{subject to} \quad \mathbf{0} \leq \boldsymbol{x} \leq \mathbf{1}, \\ &\boldsymbol{A}_{\mathcal{I}}\boldsymbol{x} \geq \mathbf{1}, \quad \boldsymbol{A}_{\mathcal{J}}\boldsymbol{x} = \mathbf{0} \end{aligned} \tag{6.8}$$

6.3.3 Noisy case

Because (6.8) can not represent noisy cases, the performance of the method is degraded in noisy cases. One version of [2]'s method thus covers the noisy case by adding slack variables as follows:

$$\begin{aligned} &\min_{\boldsymbol{x},\boldsymbol{\xi}} \left\{ \sum_n x_n + \alpha \sum_t \xi_t \right\} \\ \text{subject to} \quad &\mathbf{0} \leq \boldsymbol{x} \leq \mathbf{1}, \quad \mathbf{0} \leq \boldsymbol{\xi}_{\mathcal{I}} \leq \mathbf{1}, \quad \mathbf{0} \leq \boldsymbol{\xi}_{\mathcal{J}}, \\ &\boldsymbol{A}_{\mathcal{I}}\boldsymbol{x} + \boldsymbol{\xi}_{\mathcal{I}} \geq \mathbf{1}, \quad \boldsymbol{A}_{\mathcal{J}}\boldsymbol{x} = \boldsymbol{\xi}_{\mathcal{J}}, \end{aligned} \tag{6.9}$$

where $\boldsymbol{\xi} = [\xi_1, \cdots, \xi_T]$ is the vector composed of the slack variables for preventing degradation in the case that the test-results $\boldsymbol{y}$ include noise, and α is the regularization parameter that balances the noise tolerance and the sparsity of the solution.

6.4 Proposed method

The proposed method for controlling g is described as follows. Expected information gain of the next $(T+1)$-th test is introduced as

$$I_{T+1}(g \mid \|\boldsymbol{x}\|_0 = K) = \gamma I_{\mathrm{NEG}} + (1-\gamma) I_{\mathrm{POS}}, \tag{6.10}$$

where $I_{T+1}(g \mid \|\boldsymbol{x}\|_0 = K)$ is the expected information gain corresponding to the pool-size g given $\|\boldsymbol{x}\|_0 = K$, γ is the probability that the result of the $(T+1)$-th test is negative, I_{NEG} is the information gain of the negative test, and I_{POS} is the information gain of the positive test. The negative test means that all the elements of the pool are negative, so γ is given by

$$\gamma = \frac{\begin{pmatrix} N-K \\ g \end{pmatrix}}{\begin{pmatrix} N \\ g \end{pmatrix}}. \tag{6.11}$$

The negative test gives the information that all the elements of the pool are negative, so I_{NEG} is the sum of the current entropy of the g elements of the pool; therefore, I_{NEG} is given by

$$I_{\mathrm{NEG}} = g\left\{-r\log r - (1-r)\log(1-r)\right\}, \tag{6.12}$$

where $r = K/N$ is the probability that each element is positive. The positive test gives the information that there is at least one positive element in the pool; therefore, I_{POS} is given as

$$I_{\mathrm{POS}} = \left\{ -r^g \log r^g - (1 - r^g) \log(1 - r^g) \right\}. \tag{6.13}$$

The temporary estimate of $\boldsymbol{x}$, $\hat{\boldsymbol{x}}$, is obtained by using T tests, K can be substituted for $\|\hat{\boldsymbol{x}}\|_0$ in (6.10), and g can be optimized by maximizing $I_{T+1}(g \mid \|\boldsymbol{x}\|_0 = K)$. However, $\hat{\boldsymbol{x}}$ may include an estimation error because $\hat{\boldsymbol{x}}$ is only a temporary result based on a small number of tests. The control of g is degraded by the estimation error; therefore, the objective function (6.10) is revised in consideration of the estimation error as follows:

$$\begin{aligned} \bar{I}_{T+1}(g \mid \|\boldsymbol{x}\|_0 = K) \quad = \quad & \sum_{\{K' | K' = K - a + b\}} I_{T+1}(g \mid \|\boldsymbol{x}\|_0 = K) \\ & \times \left(\begin{array}{c} K' \\ a \end{array} \right) \epsilon^a (1 - \epsilon)^{K' - a} \\ & \times \left(\begin{array}{c} N - K' \\ b \end{array} \right) \epsilon^b (1 - \epsilon)^{N - K' - b}, \end{aligned} \tag{6.14}$$

where ϵ is the probability of the estimation error, a is the number of the false-positive elements, and b is the number of the false-negative elements. g can be optimized by maximizing (6.14). This revision can be regarded as smoothing the information gain, and it is expected that robustness to the estimation error is achieved.

The convergence of the above-described adaptation of the case of no noise is discussed as follows. $\hat{\boldsymbol{x}}_T$ is defined as the estimates of $\boldsymbol{x}$ by using T tests. $\hat{\boldsymbol{x}}_{T+1}$ is defined as the estimates of $\boldsymbol{x}$ by using $(T + 1)$ tests. $\hat{\boldsymbol{x}}_T$ is given by Eq. (6.8). When the $(T + 1)$-th test is positive, $\hat{\boldsymbol{x}}_{T+1}$ is given by

$$\begin{aligned} \hat{\boldsymbol{x}}_{T+1} = \arg\min_{\boldsymbol{x}} & \left\{ \sum_n x_n \right\} \\ \text{subject to} \quad & \mathbf{0} \leq \boldsymbol{x} \leq \mathbf{1}, \\ & \boldsymbol{A}_{\mathcal{I}} \boldsymbol{x} \geq \mathbf{1}, \quad \boldsymbol{A}_{\mathcal{J}} \boldsymbol{x} = \mathbf{0}, \\ & {\boldsymbol{a}_{\boldsymbol{T+1}}}^T \boldsymbol{x} \geq 1. \end{aligned} \tag{6.15}$$

(6.15) is the version of (6.8) with an additional constraint (${\boldsymbol{a}_{T+1}}^T \boldsymbol{x} \geq 1$). By the additional constraint, $\|\hat{\boldsymbol{x}}_{T+1}\|_0$ may increase from $\|\hat{\boldsymbol{x}}_T\|_0$ by 1 or 0. This leads to

$\|\hat{\boldsymbol{x}}_T\|_0 \leq \|\hat{\boldsymbol{x}}_{T+1}\|_0$. When the $(T+1)$-th test is negative, $\hat{\boldsymbol{x}}_{T+1}$ is given by

$$\begin{aligned} \hat{\boldsymbol{x}}_{T+1} &= \arg\min_{\boldsymbol{x}} \left\{ \sum_n x_n \right\} \\ &\text{subject to} \quad \mathbf{0} \leq \boldsymbol{x} \leq \mathbf{1}, \\ &\boldsymbol{A}_{\mathcal{I}}\boldsymbol{x} \geq \mathbf{1}, \quad \boldsymbol{A}_{\mathcal{J}}\boldsymbol{x} = \mathbf{0}, \\ &{\boldsymbol{a}_{T+1}}^T \boldsymbol{x} = 0. \end{aligned} \tag{6.16}$$

$\|\hat{\boldsymbol{x}}_{T+1}\|_0$ does not change from $\|\hat{\boldsymbol{x}}_T\|_0$ because $\|\hat{\boldsymbol{x}}_T\|_0$ has been already minimized at the time of the T-th test. From the above, for all T, $\|\hat{\boldsymbol{x}}_T\|_0 \leq \|\hat{\boldsymbol{x}}_{T+1}\|_0$; i.e. $\|\hat{\boldsymbol{x}}_T\|_0$ weakly monotonically increases as T increases. In addition, it is obvious that $\|\hat{\boldsymbol{x}}\|_0 \leq \|\boldsymbol{x}\|_0$ because $\|\hat{\boldsymbol{x}}\|_0$ is minimized under the constraints that also holds for $\boldsymbol{x}$, so $\|\boldsymbol{x}\|_0$ is an upperbound of $\|\hat{\boldsymbol{x}}_T\|_0$. Therefore, $\|\hat{\boldsymbol{x}}_T\|_0$ moves to $\|\boldsymbol{x}\|_0$, and the adaptation of the pool-size converges as T increases.

6.5 Experimental results

The performance of the proposed method was evaluated by simulation. In particular, the averaged probability of correct estimation was computed over 100 trials as a function of T, for $N = 150$. N elements were generated independently for each trial. In this simulation, $\hat{\boldsymbol{x}} = \boldsymbol{x}$ was considered to be the correct case. The proposed method was compared with the non-adaptive conventional method [2] . To evaluate the robustness to the difference in the number of positive elements, K, the simulation was conducted for two cases: $K = 2$ and $K = 6$. The optimal pool-size for $K = 2$, i.e. $g = 50$, that for $K = 4$, i.e. $g = 30$, and that for $K = 6$, i.e. $g = 21$, were calculated by simulation in advance and were used in the non-adaptive conventional method. The original version of the information gain, (6.10), and the revised version, (6.14), were then compared. α was set to 1.0.

First, the performance of the proposed method in the case of no noise was computed. Figure 6.1 shows the probability of exact recovery in the case of $K = 2$, and Figure 6.2 shows that in the case of $K = 6$. NON-ADAPT means the non-adaptive conventional method [2], ADAPT means the proposed method maximizing (6.10), and REVISED-ADAPT means the proposed method maximizing (6.14). In both cases, the proposed

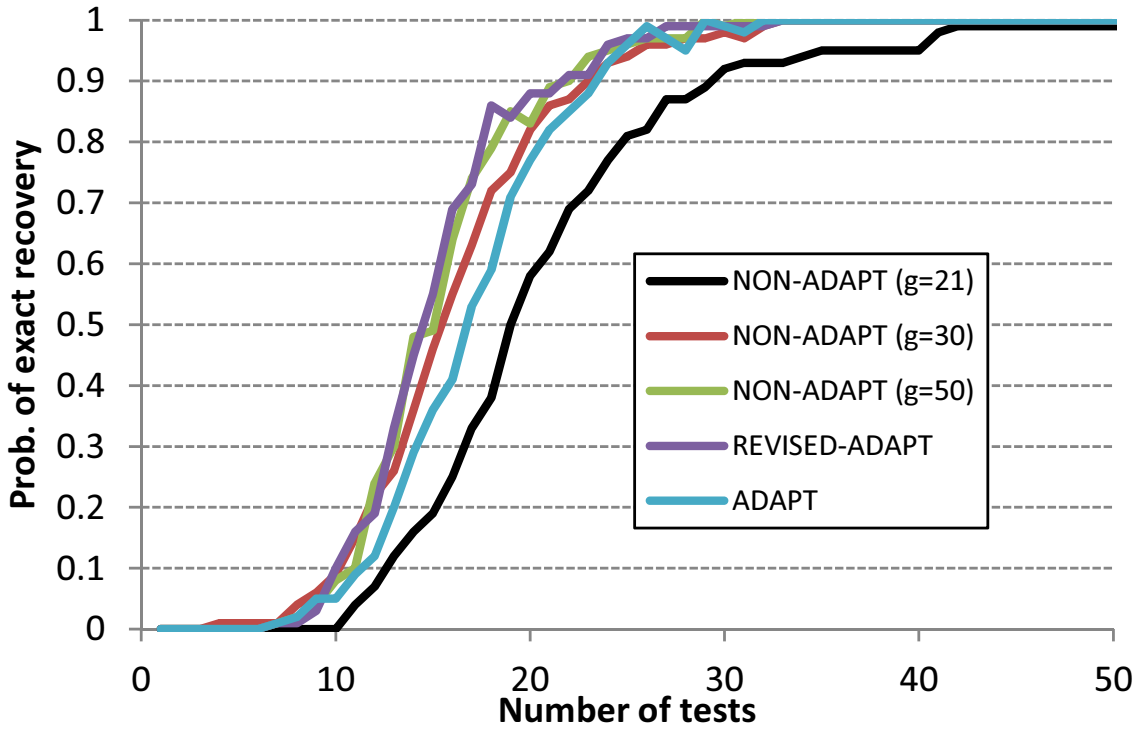


Figure 6.1: Probability of exact recovery in noiseless case as a function of number of tests, T. $N = 150$, $K = 2$.

method, namely, REVISED-ADAPT, is better than the non-adaptive method in the worst cases, and the performance of the proposed method is near the level of that using the optimal pool-size. These results indicate that the proposed method can effectively control pool-size. Figure 6.2 shows that the performance of the ADAPT is low. This result indicates that ADAPT is degraded by the estimation error.

The performance of the proposed method in the noisy case was simulated next. In the simulation, noise with i.i.d 5% probability of flipping each bit of $\boldsymbol{y}$ was added. Figure 6.3 shows the probability of exact recovery in the case of $K = 2$, and Figure 6.4 shows that in the case of $K = 6$. NON-ADAPT means the non-adaptive conventional method [2], ADAPT means the proposed method maximizing (6.10), and REVISED-ADAPT means the proposed method maximizing (6.14). According to these results, the proposed method (REVISED-ADAPT) is better than the non-adaptive method in the worst case, and the performance of the proposed method is near the level of that using the optimal pool-size. These results indicate that the proposed method can effectively control the pool-size even under noisy conditions.

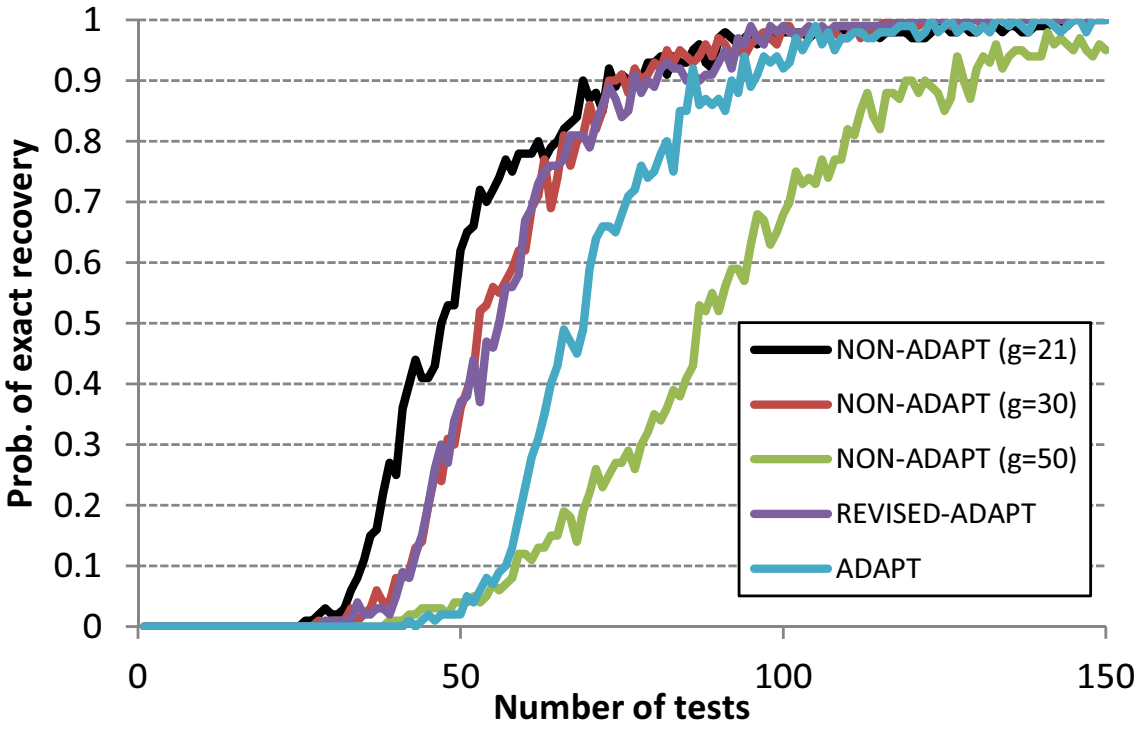


Figure 6.2: Probability of exact recovery in noiseless case as a function of number of tests, T. $N = 150$, $K = 6$.

6.6 Conclusion

A new method for solving the adaptive group-testing problem was proposed. The proposed method controls pool-size adaptively by using information gain calculated from the ℓ_0 norm of the estimated solution. Moreover, to improve the robustness to the estimation error, smoothing of the information gain in consideration of the estimation error is applied. An experimental simulation showed that the proposed method outperforms the conventional method even when the number of positive elements is varied and the number of positive elements is unknown.

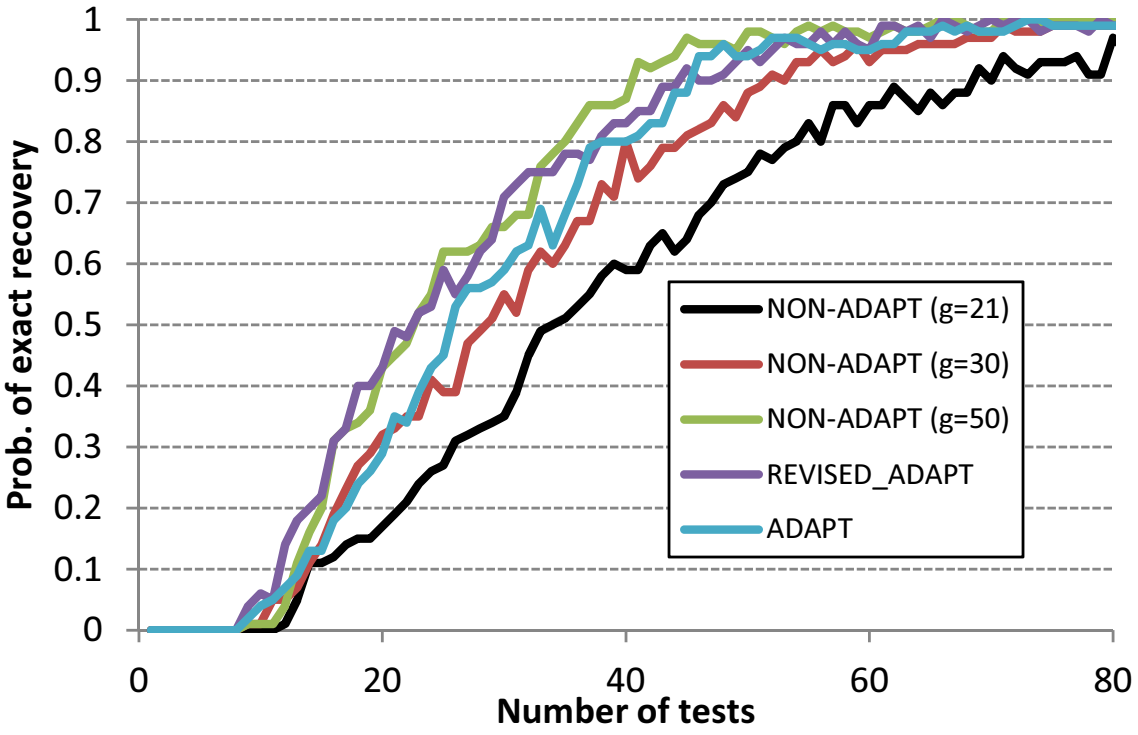


Figure 6.3: Probability of exact recovery in noisy case as a function of number of tests, T. $N = 150$, $K = 2$, and 5% noise was added.

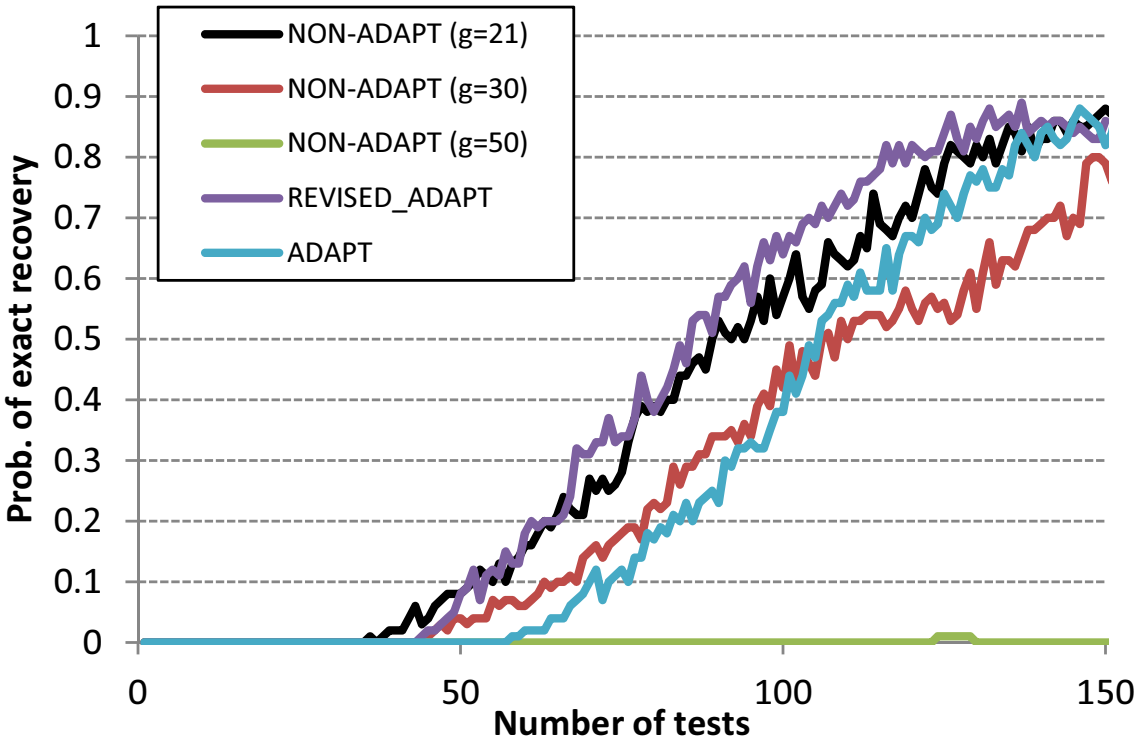


Figure 6.4: Probability of exact recovery in noisy case as a function of number of tests T. $N = 150$, $K = 6$, and 5% noise was added.

Chapter 7

Adaptive Boolean compressive sensing by using multi-armed bandit

7.1 Introduction

In Chapter 6, an algorithm of adaptive Boolean compressive sensing is introduced. The method proposed in Chapter 6 optimizes the pool-size by maximizing the expected information gain, and this optimization can be regarded as a greedy algorithm. The greedy algorithm has no guarantee of convergence that the group-size becomes close to the optimal one after sufficient tests because the temporary estimate may include an estimation error. To solve this problem, an extended version of the method is proposed here. Based on the multi-armed bandit, the proposed method controls the pool-size adaptively. The information gain of the greedy method explained in Chapter 6 is rewritten as the reward of the multi-armed bandit, and the multi-armed bandit is introduced into adaptive Boolean compressive sensing. Experimental results indicate that the correct rate of exact recovery of the proposed method converges to 1 fast without prior knowledge about the number of positive elements and that the proposed method outperforms the non-adaptive method [2] and the conventional greedy method explained in Chapter 6 in the case that the number of positive elements is large.

7.2 Problem statement

The problem statement is the same as Chapter 6. N is the number of elements, of which a subset of size K is positive. The N elements are unknown and are written as $\boldsymbol{x} = [x_1, x_2, \cdots, x_N]^T$. $x_n = 1$ indicates that the n-th element is positive, and $x_n = 0$ indicates that the n-th element is negative. T tests, where $T < N$, are then performed. The mixing matrix $\boldsymbol{A}$ is a $T \times N$ binary matrix. The element of the t-th row and the n-th column of $\boldsymbol{A}$ is given as $a_{t,n}$, where $a_{t,n} = 1$ indicates that the n-th element is mixed into the pool of the t-th test, and $a_{t,n} = 0$ indicates that the n-th element is not mixed into the pool of the t-th test. The observed signal of each test, t, is a single Boolean value, $y_t \in \{0, 1\}$. y_t is obtained by taking the Boolean sum of $\{x_n | a_{t,n} = 1\}$. For convenience, $\boldsymbol{y} = [y_1, y_2, \cdots, y_T]^T$ is written. The vector notation

$$\boldsymbol{y} = \boldsymbol{A}\boldsymbol{x} \tag{7.1}$$

is used in the following. The observation with noise is represented by

$$\boldsymbol{y} = (\boldsymbol{A} \vee \boldsymbol{x}) \otimes \boldsymbol{v}, \tag{7.2}$$

where $\boldsymbol{v}$ is the Boolean vector of errors, and $\otimes$ means the XOR operation. The problem of group-testing is to estimate unknown vector $\boldsymbol{x}$ from given $\boldsymbol{A}$ and $\boldsymbol{y}$.

Similarly to Chapter 6, the simple-random-sampling-design is used for design of $\boldsymbol{A}$. Non-adaptive group-testing such as the conventional Malioutov's method [2] has the problem that the optimal pool-size largely depends on K although K is unknown. To solve this problem, the present study thus focuses on adaptive group-testing. In addition, because the simple-random-sampling-design is assumed, control of g is focused on. Therefore, the next mixing vector after the T-th test $\boldsymbol{a}_{T+1}$ is determined by the following random trial:

$$\Pr\left[\boldsymbol{a}_{T+1}\right] = \begin{cases} 1/\begin{pmatrix} N \\ g \end{pmatrix} & \text{if } \|\boldsymbol{a}_{T+1}\|_0 = g, \\ 0 & \text{otherwise,} \end{cases} \tag{7.3}$$

where g is determined by

$$g = F_T\left(\boldsymbol{A}, \boldsymbol{y}\right), \tag{7.4}$$

and F_T is a function of $\boldsymbol{A}$ and $\boldsymbol{y}$. In Section 7.4, a new function F_T is proposed.

7.3 Boolean compressive sensing for group-testing

Malioutov and Malyutov [2] proposed a conversion of group-testing into compressive sensing through a linear-programming relaxation. This method is the basis of our method, which is explained in this section.

Equation (7.1) is not a linear equation in a real vector space but a Boolean equation. However, it is shown in [2] that (7.1) can be replaced with a closely related linear formulation: $\mathbf{1} \leq \boldsymbol{A}_{\mathcal{I}}\boldsymbol{x}$, and $\mathbf{0} = \boldsymbol{A}_{\mathcal{J}}\boldsymbol{x}$, where $\mathcal{I} = \{t|y_t = 1\}$ is the set of the tests that obtain positive results, and $\mathcal{J} = \{t|y_t = 0\}$ is the set of the tests that obtain negative results. A linear-programming formulation is therefore given as

$$\begin{aligned} &\min_{\boldsymbol{x},\boldsymbol{\xi}} \left\{ \sum_n x_n + \alpha \sum_t \xi_t \right\} \\ \text{subject to} \quad &\mathbf{0} \leq \boldsymbol{x} \leq \mathbf{1}, \quad \mathbf{0} \leq \boldsymbol{\xi}_{\mathcal{I}} \leq \mathbf{1}, \quad \mathbf{0} \leq \boldsymbol{\xi}_{\mathcal{J}}, \\ &\boldsymbol{A}_{\mathcal{I}}\boldsymbol{x} + \boldsymbol{\xi}_{\mathcal{I}} \geq \mathbf{1}, \quad \boldsymbol{A}_{\mathcal{J}}\boldsymbol{x} = \boldsymbol{\xi}_{\mathcal{J}}, \end{aligned} \tag{7.5}$$

where $\boldsymbol{\xi} = [\xi_1, \cdots, \xi_T]$ is the vector composed of the slack variables for preventing degradation in the case that the test-results $\boldsymbol{y}$ include noise, and α is the regularization parameter that balances the noise tolerance and the sparsity of the solution.

7.4 Proposed method

The proposed method for controlling the pool-size g in adaptive Boolean compressive sensing is described as follows.

Here, similarly to Chapter 6, expected information gain of the next $(T+1)$-th test is introduced as

$$I_{T+1}(g \mid \|\boldsymbol{x}\|_0 = K) = \gamma I_{\mathrm{NEG}} + (1-\gamma) I_{\mathrm{POS}}, \tag{7.6}$$

where $I_{T+1}(g \mid \|\boldsymbol{x}\|_0 = K)$ is the expected information gain corresponding to the pool-size g given $\|\boldsymbol{x}\|_0 = K$, γ is the probability that the result of the $(T+1)$-th test is negative, I_{NEG} is the information gain of the negative test, and I_{POS} is the information gain of the positive test. The negative test means that all the elements of the pool are negative,

so γ is given by

$$\gamma = \frac{\begin{pmatrix} N-K \\ g \end{pmatrix}}{\begin{pmatrix} N \\ g \end{pmatrix}}. \tag{7.7}$$

The negative test gives the information that all the elements of the pool are negative, so I_{NEG} is the sum of the current entropy of the g elements of the pool; therefore, I_{NEG} is given by

$$I_{\mathrm{NEG}} = g\left\{-r\log r - (1-r)\log(1-r)\right\}, \tag{7.8}$$

where $r = K/N$ is the probability that each element is positive. The positive test gives the information that there is at least one positive element in the pool; therefore, I_{POS} is given as

$$I_{\mathrm{POS}} = \left\{-r^g\log r^g - (1-r^g)\log(1-r^g)\right\}. \tag{7.9}$$

The temporary estimate of $\boldsymbol{x}$, $\hat{\boldsymbol{x}}$, is obtained by using T tests, and the greedy method of Chapter 6 optimizes g by maximizing $I_{T+1}(g \mid \|\boldsymbol{x}\|_0 = K)$ in (7.6) based on $\|\hat{\boldsymbol{x}}\|_0$. However, $\hat{\boldsymbol{x}}$ may include an estimation error because $\hat{\boldsymbol{x}}$ is only a temporary result based on a small number of tests. The greedy method for controlling g is degraded by the estimation error, so the greedy method has no guarantee that g becomes close to the optimal pool-size after sufficient tests. To solve the problem, the multi-armed bandit approach is introduced into the greedy method here. The multi-armed bandit was introduced by Robbins [103]. The multi-armed bandit is a method for solving the trade-off between to gain new knowledge by exploring an environment and to exploit a current reliable knowledge [104]. There are several approaches for the multi-armed bandit. One of the approaches that have a guarantee that the selection of actions converges to the optimal one after sufficient plays is Upper-Confidence-Bounds (UCB) algorithm [105]. Here, the multi-armed bandit problem is defined by a random variable $R(t) \in [0,1]$ for $t \geq 1$, where $R(t)$ is called a "reward" and is yielded from the i-th machine selected at each t-th play. $R(t)$ at each play is independent and identically distributed following an unknown expected value μ_i. UCB algorithm selects the next machine to play based on the sequence of past plays and obtained rewards. At $(T+1)$-th play, UCB algorithm

selects the i-th machine such that the following function $f_i(T+1)$ is maximized:

$$f_i(T+1) = \bar{\mu}_i(T+1) + c\sqrt{\frac{2\log T}{T_i}}, \tag{7.10}$$

where $\bar{\mu}_i(T+1)$ is the average of $R(t)$ obtained from the i-th machine, and T_i is the number of times that the i-th machine has been selected, and c is a constant positive value. Then the regret after T-th play is defined by

$$\mathbb{E}\left[\mu^* T - \sum_{t}^{T} R(t)\right], \tag{7.11}$$

where $\mu^* = \max_i \mu_i$. Auer *et al.* [105] showed that the regret at the T-th play is bounded by:

$$8\left[\sum_{i:\mu_i<\mu^*} \frac{\log T}{\Delta_i}\right] + \left(1 + \frac{\pi^2}{3}\right)\sum_i \Delta_i, \tag{7.12}$$

where $\Delta_i = \mu^* - \mu_i$. Here, to introduce UCB algorithm into the conventional pool-size control method, the reward $R(t)$ is replaced by the information gain, and the index of machine i is interpreted as the pool-size g. Equation (2.3) is converted to:

$$f_g(T+1) = \bar{\mu}_g(T+1) + c\sqrt{\frac{2\log T}{T_g}}, \tag{7.13}$$

where $\bar{\mu}_g(T+1)$ is the average of the information gain obtained by the pool-size g, and T_g is the number of times that the pool-size g has been selected. $\bar{\mu}_g(T+1)$ is calculated by

$$\bar{\mu}_g(T+1) = \frac{1}{T_g}\left(|\mathcal{I}_g| I_{\mathrm{POS}} + |\mathcal{J}_g| I_{\mathrm{NEG}}\right), \tag{7.14}$$

where $\mathcal{I}_g$ is the set of the tests that obtain positive results by the pool-size g, and $\mathcal{J}_g$ is the set of the tests that obtain negative results by the pool-size g, and so $|\mathcal{I}_g| + |\mathcal{J}_g| = T_g$. Here, to make the range of reward-values to $[0, 1]$, I_{POS} and I_{NEG} are normalized by maximum. In addition, to accelerate the convergence, the predicted information gain $\hat{\mu}_g(T+1)$ is introduced:

$$\hat{\mu}_g(T+1) = \sum_K p(\|\boldsymbol{x}\|_0 = K \mid \|\hat{\boldsymbol{x}}\|_0) I_{T+1}(g \mid \|\boldsymbol{x}\|_0 = K), \tag{7.15}$$

where $p(\|\boldsymbol{x}\|_0 = K \mid \|\hat{\boldsymbol{x}}\|_0)$ is the conditional probability of $\|\boldsymbol{x}\|_0 = K$ given the estimated number of the positive elements $\|\hat{\boldsymbol{x}}\|_0$. Here, to make the range of reward-values to $[0, 1]$,

$I_{T+1}(g \mid \|\boldsymbol{x}\|_0 = K)$ is normalized by maximum. $p(\|\boldsymbol{x}\|_0 = K \mid \|\hat{\boldsymbol{x}}\|_0)$ is modeled by the following probability distribution function of the binomial distribution:

$$\begin{aligned} & p\left(\|\boldsymbol{x}\|_0 = K \mid \|\hat{\boldsymbol{x}}\|_0\right) \\ = & \left(\begin{array}{c} N \\ \|\hat{\boldsymbol{x}}\|_0 \end{array} \right) \left[\frac{K}{N}\right]^{\|\hat{\boldsymbol{x}}\|_0} \left[1 - \frac{K}{N}\right]^{N - \|\hat{\boldsymbol{x}}\|_0}. \end{aligned} \tag{7.16}$$

Then, $f_g(T+1)$ is redefined as:

$$f_g(T+1) = (1-\beta^T)\bar{\mu}_g(T+1) + \beta^T \hat{\mu}_g(T+1) + c\sqrt{\frac{2\log T}{T_g}}, \tag{7.17}$$

where β is a constant parameter ($0 < \beta < 1$). The proposed method selects the pool-size g such that (7.17) is maximized. In (7.17), when T is small, the first term of the right hand side is near 0, and the second term and the third term are dominant. Especially, the third term has a tendency to prefer the pool-size whose T_g is small and explores various pool-sizes. In the middle phase, the second term predicts the optimal pool-size and accelerates the convergence. As T increases, the second term becomes near 0, and the first term and the third term are dominant, that is, (7.17) becomes close to the original UCB. Therefore, as (7.12), the regret grows at least logarithmically, and the optimal pool-size is achieved at $T \to \infty$. From the above, it is expected that the UCB-based proposed method has fast convergence and the guarantee that the group-size becomes close to the optimal one after sufficient tests.

7.5 Experimental results

We evaluated the performance of the proposed method by simulation. We computed the averaged probability of the correct estimation over 100 trials as a function of T, for $N = 150$. The N elements were generated independently for each trial. In this experiment, we considered the case of $\hat{\boldsymbol{x}} = \boldsymbol{x}$ as correct. We compared the non-adaptive method [2], the conventional greedy maximization of (7.6) (proposed in Chapter 6), and the UCB-based proposed method. In order to evaluate the robustness to the difference of the number of the positive elements K, we conducted the experiment in the case of $K = 2$ and that of $K = 6$. Also, we calculated the optimal pool-size for $K = 2$, i.e. $g = 50$, that for $K = 4$, i.e. $g = 30$, and that for $K = 6$, i.e. $g = 21$ by simulation. In

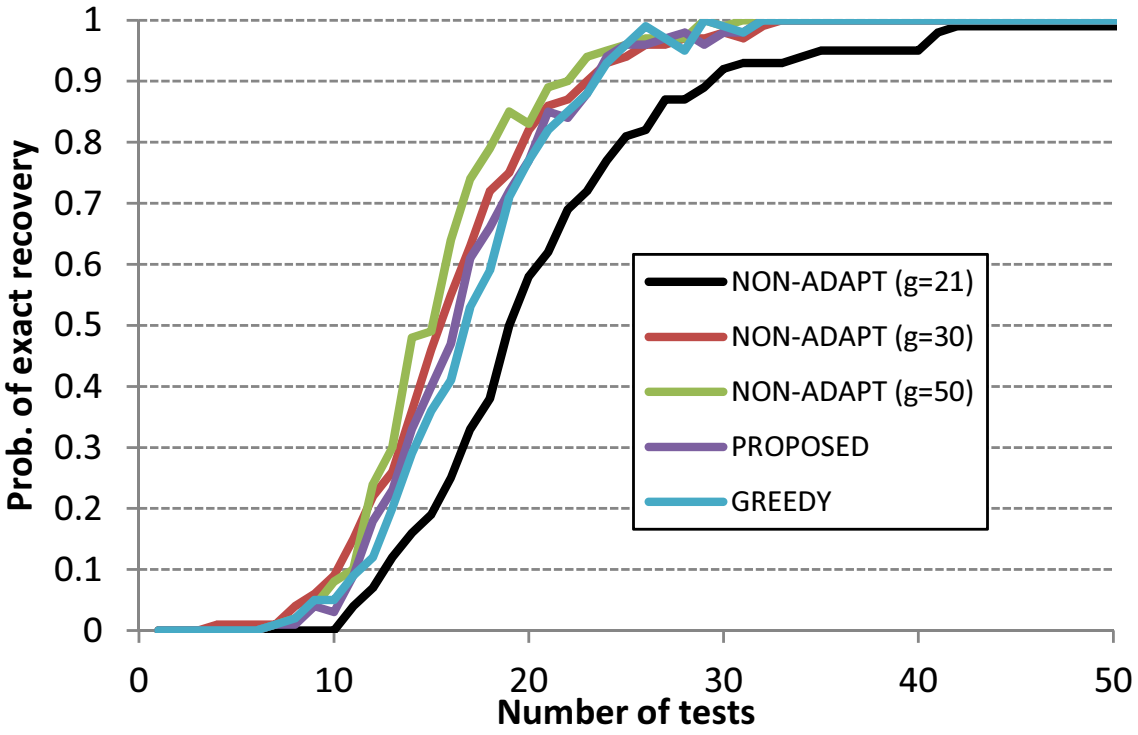


Figure 7.1: Probability of exact recovery in the noiseless case as a function of the number of tests T. NON-ADAPT means the non-adaptive method [2], PROPOSED means the UCB-based proposed method, and GREEDY means the conventional greedy maximization of (7.6) proposed in Chapter 6. $N = 150$, $K = 2$.

the non-adaptive method, these fixed optimal pool-sizes were used. In the conventional greedy method and the proposed method, the adaptively-determined pool-size was used. α was set to 1.0.

First, we computed the performance in the case of no noise. Figure 7.1 shows the case of $K = 2$, and Fig. 7.2 shows the case of $K = 6$. In both cases, the convergence of the proposed method was faster than that of the worst cases of the non-adaptive method, and the correct rate after convergence was 1. The convergence speed of the proposed method was on the same level with the optimal pool-size. Also, as Fig. 7.2 shows, in the case that K was 6, the convergence of the proposed method was faster than that of the conventional greedy method. The results in Fig. 7.2 indicate that the exploration of the multi-armed bandit works well in the case that K is large. These results indicate that the control of the pool-size of the proposed method is effective.

Second, we computed the performance in the noisy case. We added noise with i.i.d

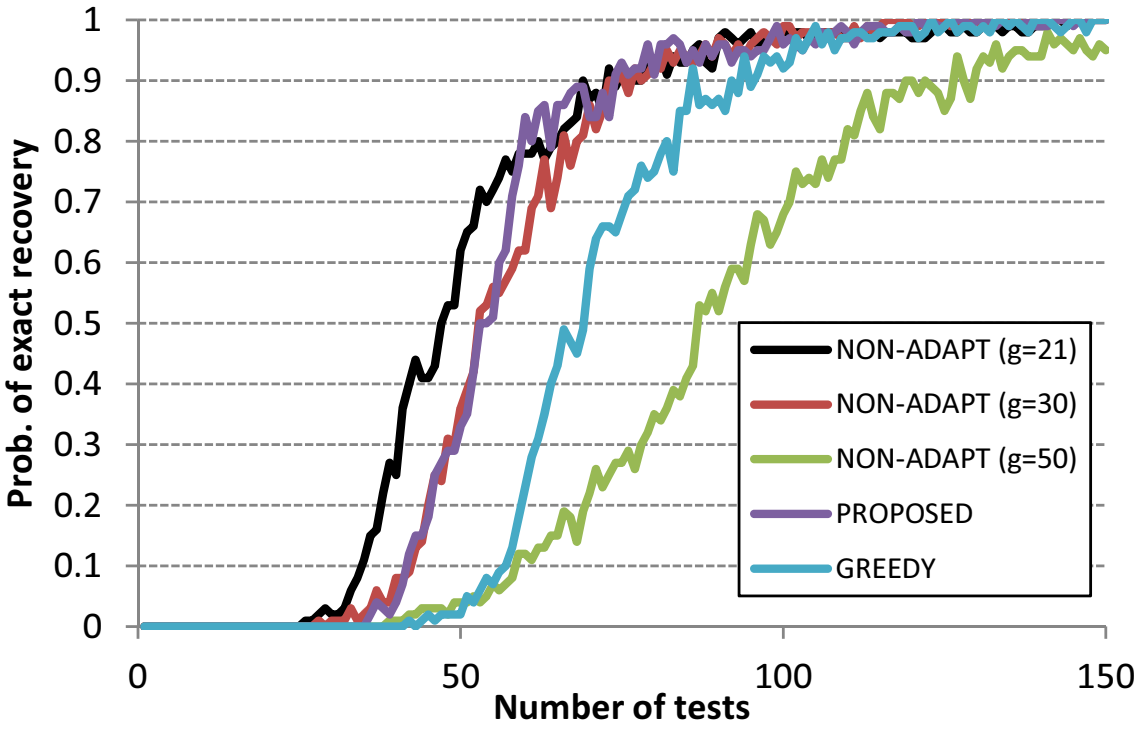


Figure 7.2: Probability of exact recovery in the noiseless case as a function of the number of tests T. NON-ADAPT means the non-adaptive method [2], PROPOSED means the UCB-based proposed method, and GREEDY means the conventional greedy maximization of (7.6) proposed in Chapter 6. $N = 150$, $K = 6$.

5% probability of flipping each bit of $\boldsymbol{y}$. Figure 7.3 shows the case of $K = 2$, and Fig. 7.4 shows the case of $K = 6$. Also, in these results, the convergence of the proposed method was faster than that of the worst cases of the non-adaptive method, and the correct rate after convergence was 1. The convergence speed of the proposed method was on the same level with the optimal pool-size. Also, as Fig. 7.4 shows, in the case that K was 6, the convergence of the proposed method was faster than that of the conventional greedy method. The results in Fig. 7.4 indicate that the exploration of the multi-armed bandit works well in the case that K is large under noisy conditions. These results indicate that the control of the pool-size of the proposed method is effective even under noisy conditions.

Finally, we checked the number of false positives and that of false negatives. The false positives are defined as the elements that are actually negative but were identified as positive. The false negatives are defined as the elements that are actually positive

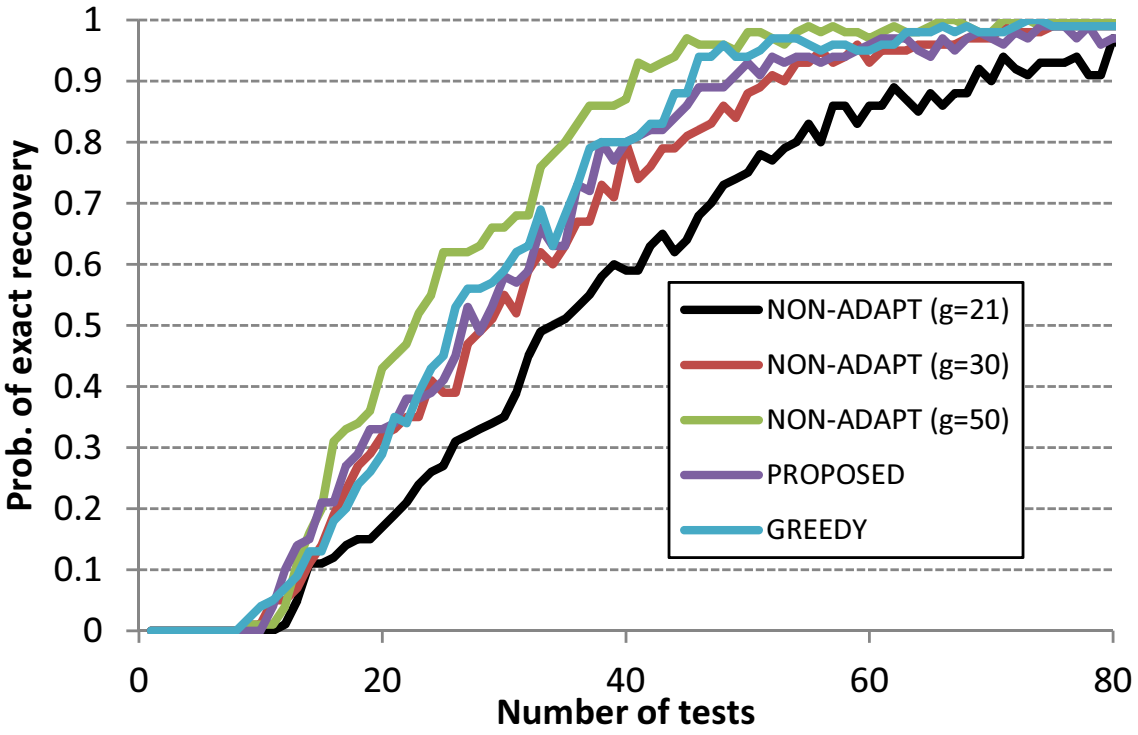


Figure 7.3: Probability of exact recovery in the noisy case as a function of the number of tests T. NON-ADAPT means the non-adaptive method [2], PROPOSED means the UCB-based proposed method, and GREEDY means the conventional greedy maximization of (7.6) proposed in Chapter 6. $N = 150$, $K = 2$.

but were identified as negative. In this experiment, the number of positive elements K was set to 6 and noise with i.i.d 5% was added. Figure 7.5 shows the average number of false positives, and Fig. 7.6 shows the average number of false negatives. In the case of the UCB-based method, both false positives and false negatives were reduced faster than the greedy one. These results indicate that the UCB-based method has effectiveness for both false positives and false negatives.

7.6 Conclusion

A new method for solving adaptive Boolean compressive sensing was proposed. To achieve the guarantee that the group-size becomes close to the optimal one after sufficient tests, based on the multi-armed bandit, the proposed method controls the pool-size adaptively. The information gain of the conventional greedy method was rewritten as

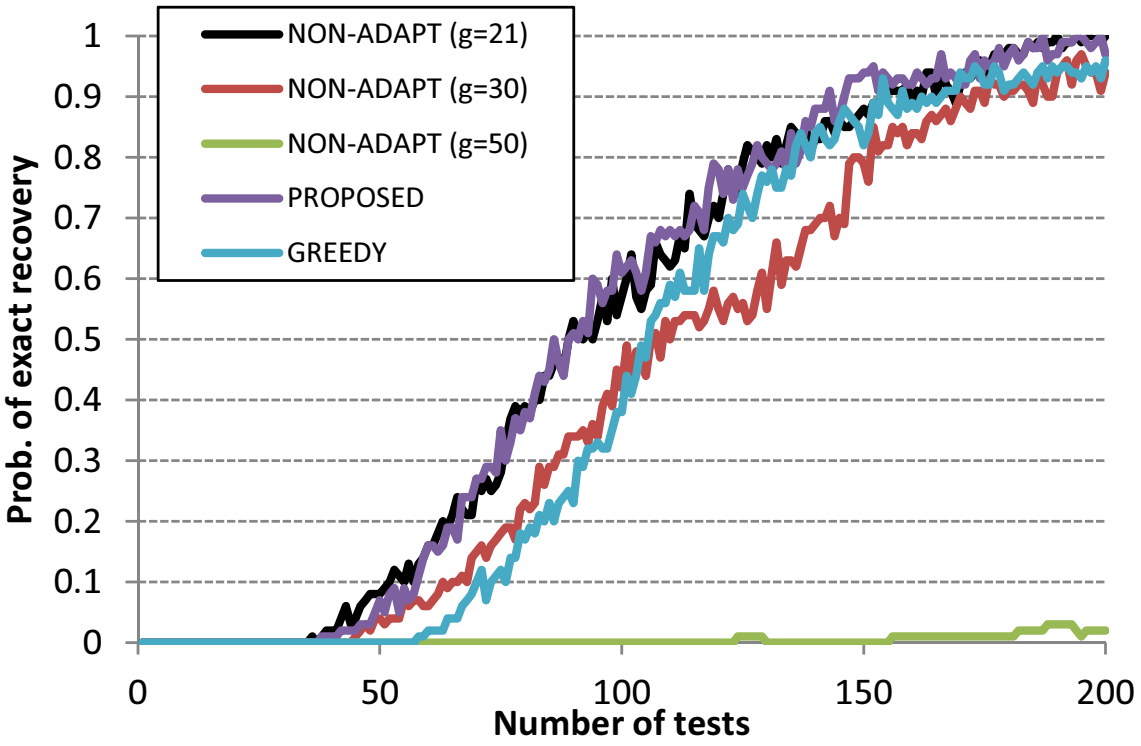


Figure 7.4: Probability of exact recovery in the noisy case as a function of the number of tests T. NON-ADAPT means the non-adaptive method [2], PROPOSED means the UCB-based proposed method, and GREEDY means the conventional greedy maximization of (7.6) proposed in Chapter 6. $N = 150$, $K = 6$.

the reward of the multi-armed bandit, and the multi-armed bandit was introduced into adaptive Boolean compressive sensing. In an experimental evaluation of the method, it was shown that the correct rate of exact recovery of the proposed method converges to 1 fast without prior knowledge about the number of positive elements and that the proposed method outperforms the non-adaptive method and the conventional greedy method in the case that the number of positive elements is large.

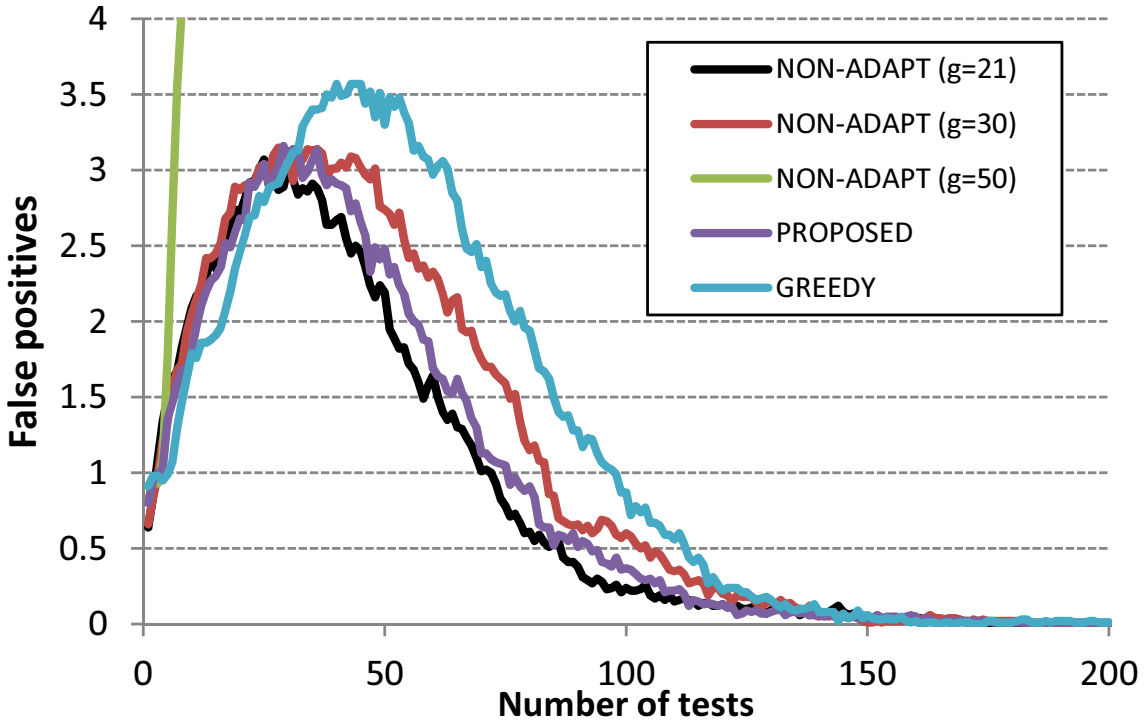


Figure 7.5: Average number of false positives as a function of the number of tests T. NON-ADAPT means the non-adaptive method [2], PROPOSED means the UCB-based proposed method, and GREEDY means the conventional greedy maximization of (7.6) proposed in Chapter 6. $N = 150$, $K = 6$.

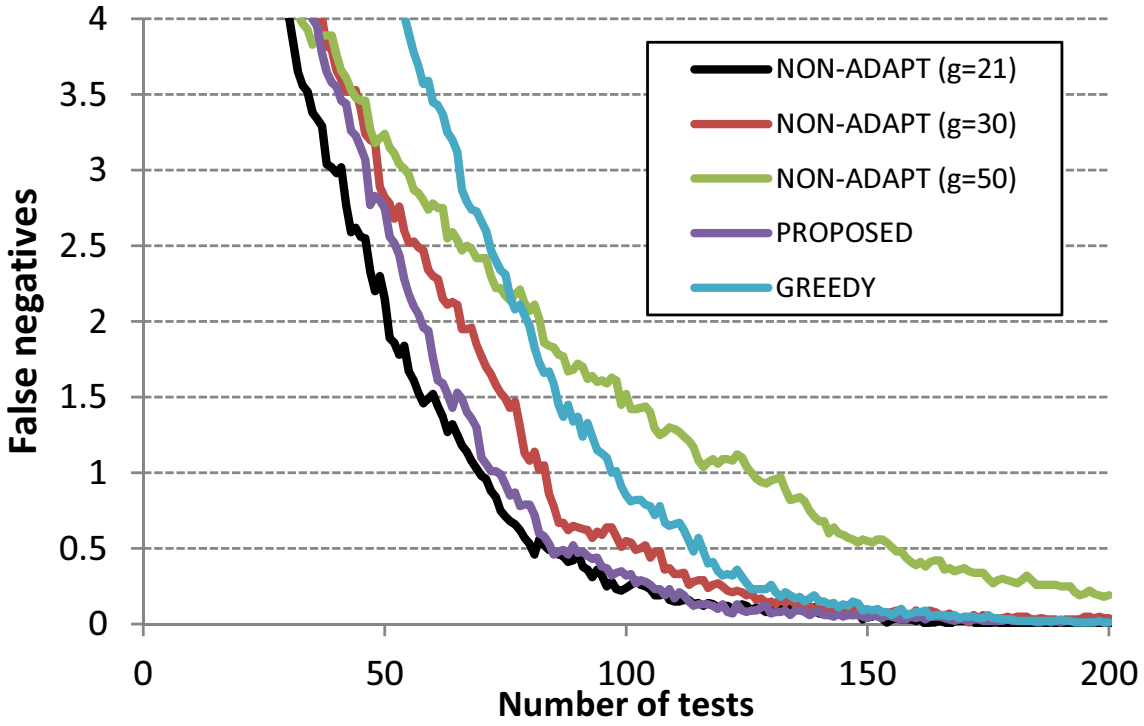


Figure 7.6: Average number of false negatives as a function of the number of tests T. NON-ADAPT means the non-adaptive method [2], PROPOSED means the UCB-based proposed method, and GREEDY means the conventional greedy maximization of (7.6) proposed in Chapter 6. $N = 150$, $K = 6$.

Chapter 8

Improvement of robustness to change of positive elements in Boolean compressive sensing

8.1 Introduction

In the previous chapters, adaptive Boolean compressive sensing was proposed for achieving the robustness to the difference of the number of positive elements. However, both these adaptive approaches and the conventional non-adaptive ones assume that positive elements does not change in the middle of tests, and the estimation performance is degraded in the case that positive elements change because the results of the tests before a change-point are inconsistent with those of the tests after the change-point. Also, in the application for location estimation of substances, the locations of substances change in the middle of tests.

To improve the robustness to change of positive elements, a method for group-testing is proposed here. The proposed method detects the latest change-point of positive elements, and it finds positive elements by using only the results of the tests after the change-point. To detect the change-point, the proposed method makes use of the fact that the distribution of the results depends on the number of positive elements. Experimental simulation indicates that the proposed method outperforms the conventional method [2] on the condition that positive elements change in the middle of tests.

8.2 Problem statement

Most parts of the problem statement are the same as Chapter 6. N is the number of elements, of which a subset of size K is positive. The N elements are unknown and are written as $\boldsymbol{x} = [x_1, x_2, \cdots, x_N]^T$. $x_n = 1$ indicates that the n-th element is positive, and $x_n = 0$ indicates that the n-th element is negative. T tests, where $T < N$, are then performed. The mixing matrix $\boldsymbol{A}$ is a $T \times N$ binary matrix. The element of the t-th row and the n-th column of $\boldsymbol{A}$ is given as $a_{t,n}$, where $a_{t,n} = 1$ indicates that the n-th element is mixed into the pool of the t-th test, and $a_{t,n} = 0$ indicates that the n-th element is not mixed into the pool of the t-th test. The observed signal of each test, t, is a single Boolean value, $y_t \in \{0, 1\}$. y_t is obtained by taking the Boolean sum of $\{x_n | a_{t,n} = 1\}$. For convenience, $\boldsymbol{y} = [y_1, y_2, \cdots, y_T]^T$ is written. The vector notation

$$\boldsymbol{y} = \boldsymbol{A}\boldsymbol{x} \tag{8.1}$$

is used in the following. The observation with noise is represented by

$$\boldsymbol{y} = (\boldsymbol{A} \vee \boldsymbol{x}) \otimes \boldsymbol{v}, \tag{8.2}$$

where $\boldsymbol{v}$ is the Boolean vector of errors, and $\otimes$ means the XOR operation. The problem of group-testing is to estimate unknown vector $\boldsymbol{x}$ from given $\boldsymbol{A}$ and $\boldsymbol{y}$. Similarly to Chapter 6, the simple-random-sampling-design is used for design of $\boldsymbol{A}$.

In a number of applications such as location estimation of substances, the unknown vector $\boldsymbol{x}$ may change in the middle of tests, particularly, $\boldsymbol{x}$ may change at an occasional time-point, in other words, a "change-point". One of the problems of group-testing is that, on the condition that the unknown vector $\boldsymbol{x}$ changes at the change-point, the elements of $\boldsymbol{y}$ corresponding to the tests before the change-point are inconsistent with those corresponding to the tests after the change-point, and $\boldsymbol{x}$ can not be estimated accurately. The present study thus focuses on an method of the improvement of robustness to change of $\boldsymbol{x}$. Here, we assume that all the tests have a particular order relation of the time t, and $\boldsymbol{x}$ can be rewritten as $\boldsymbol{x}(t)$ considering change of $\boldsymbol{x}(t)$. The task is to estimate the current unknown vector $\boldsymbol{x}(T)$ from $\boldsymbol{A}$ and $\boldsymbol{y}$.

8.3 Proposed method

The conventional method assumes that $\boldsymbol{x}$ does not change by time, so the current $\boldsymbol{x}$, i.e. $\boldsymbol{x}(T)$, can not be estimated accurately. To improve the robustness to change of $\boldsymbol{x}(t)$, the proposed method detects the latest change-point $c(T)$, and estimates $\boldsymbol{x}(T)$ using only the results of only the tests after $c(T)$. Here, $c(T)$ is defined as t such that $\boldsymbol{x}(t-1) \neq \boldsymbol{x}(t)$, and $\boldsymbol{x}(t) = \boldsymbol{x}(t+1) = \cdots = \boldsymbol{x}(T)$. To detect the change-point $c(T)$, the proposed method performs a likelihood ratio test as follows:

$$\frac{P(H_1|y_t, \cdots, y_T)}{P(H_0|y_t, \cdots, y_T)} = \frac{1 - P(H_0|y_t, \cdots, y_T)}{P(H_0|y_t, \cdots, y_T)} > \theta, \tag{8.3}$$

where H_0 is the hypothesis that there is no change-point in the tests of $\tau = t \cdots T$, H_1 is the hypothesis that there is a change-point in the tests of $\tau = t \cdots T$, and θ is a threshold parameter. $P(H_0|y_t, \cdots, y_T)$ can be rewritten as:

$$\begin{aligned} & P(H_0|y_t, \cdots, y_T) \\ \propto \; & P(H_0, y_t, \cdots, y_T) \\ = \; & P(\boldsymbol{x}(t) = \cdots = \boldsymbol{x}(T), y_t, ..., y_T) \\ \leq \; & P(\|\boldsymbol{x}(t)\|_0 = \cdots = \|\boldsymbol{x}(T)\|_0, y_t, ..., y_T). \end{aligned} \tag{8.4}$$

In the application of location estimation of substances, not only the combination of positive elements but also the number of positive elements changes at most change-points because change of $\boldsymbol{x}$ is caused by substance diffusion. Therefore, we approximate (8.4) as follows:

$$\begin{aligned} & P(\boldsymbol{x}(t) = \cdots = \boldsymbol{x}(T), y_t, ..., y_T) \\ \approx \; & P(\|\boldsymbol{x}(t)\|_0 = \cdots = \|\boldsymbol{x}(T)\|_0, y_t, ..., y_T) \end{aligned} \tag{8.5}$$

Based on the approximation of (8.5), the proposed method can make use of the fact that the distribution of $\boldsymbol{y}$ depends on $\|\boldsymbol{x}(t)\|_0$. (8.5) is converted as follows:

$$\begin{aligned} & P(\|\boldsymbol{x}(t)\|_0 = \cdots = \|\boldsymbol{x}(T)\|_0, y_t, ..., y_T) \\ = & \sum_K P(\|\boldsymbol{x}(t)\|_0 = \cdots = \|\boldsymbol{x}(T)\|_0 = K, y_t, ..., y_T) \\ = & \sum_K \prod_{\tau=t}^{T} P(\|\boldsymbol{x}(\tau)\|_0 = K) \\ & \times \prod_{\tau=t}^{T} P(y_\tau \mid \|\boldsymbol{x}(\tau)\|_0 = K). \end{aligned} \tag{8.6}$$

Here, we can assume that $P(\|\boldsymbol{x}(\tau)\|_0 = K)$ is a sparse prior distribution, for example, the Poisson distribution $P(\|\boldsymbol{x}(\tau)\|_0 = K) = \frac{\lambda^K e^{-\lambda}}{K!}$, where λ is the parameter of the distribution. Furthermore, $P(y_\tau \mid \|\boldsymbol{x}(\tau)\|_0 = K)$ can be estimated by summation of y_τ, $Y(\tau) = \sum_{f=\tau-F}^{\tau+F} y_f$, which is the sufficient statistics of the Bernoulli distribution, as follows:

$$\begin{aligned} & P(y_\tau \mid \|\boldsymbol{x}(\tau)\|_0 = K) \\ = & \left\{ 1 - \left(1 - \frac{K}{N} \frac{\sum_{n=1}^{N} a_{\tau n}}{N} \right)^N \right\}^{Y(\tau)} \\ & \times \left\{ \left(1 - \frac{K}{N} \frac{\sum_{n=1}^{N} a_{\tau n}}{N} \right)^N \right\}^{2F+1-Y(\tau)} \end{aligned}, \tag{8.7}$$

where F is the frame size for summation of y_τ. Thus, each time the result of the test is obtained, the proposed method calculates (8.7), (8.6), and (8.4), and it evaluates (8.3) of each t, and it regards the latest test satisfying (8.3) as the latest change-point $c(T)$.

Finally, we rewrite the linear-programming formulation (6.9) as follows:

$$\begin{aligned} & \min_{\boldsymbol{x}, \boldsymbol{\xi}} \left\{ \sum_n x_n + \alpha \sum_{t=c(T)}^{T} \xi_t \right\} \\ \text{subject to} \quad & \mathbf{0} \leq \boldsymbol{x} \leq \mathbf{1}, \quad \mathbf{0} \leq \boldsymbol{\xi}_{\mathcal{I}_c} \leq \mathbf{1}, \quad \mathbf{0} \leq \boldsymbol{\xi}_{\mathcal{J}_c}, \\ & \boldsymbol{A}_{\mathcal{I}_c} \boldsymbol{x} + \boldsymbol{\xi}_{\mathcal{I}_c} \geq \mathbf{1}, \quad \boldsymbol{A}_{\mathcal{J}_c} \boldsymbol{x} = \boldsymbol{\xi}_{\mathcal{J}_c}, \end{aligned} \tag{8.8}$$

where c is the latest change-point, and $\mathcal{I}_c = \{t | y_t = 1, c(T) < t\}$ is the set of the tests that obtain positive results after $c(T)$, and $\mathcal{J}_c = \{t | y_t = 0, c(T) < t\}$ is the set of the tests that obtain negative resultss after $c(T)$. By calculating (8.8), the proposed method estimate the $\boldsymbol{x}(T)$.

8.4 Experimental results

The performance of the proposed method was evaluated by simulation. In particular, the averaged probability of correct estimation was computed over 100 trials as a function of T, for $N = 150$. N elements were generated independently for each trial. In this simulation, the probability p of the Bernoulli random design of $\boldsymbol{A}$ was 0.333, noise with i.i.d 3% probability of flipping each bit of $\boldsymbol{y}$ was added, λ was 1.5, F was 15, and α was 1.0. The proposed method was compared with the conventional method [2]. To evaluate the robustness to change of the unknown vector $\boldsymbol{x}(t)$, the simulation was conducted for three cases:

Case1 $\|\boldsymbol{x}(t)\|_0$ changed from 0 to 2.

Case2 $\|\boldsymbol{x}(t)\|_0$ changed from 1 to 4.

Case3 $\|\boldsymbol{x}(t)\|_0$ changed from 4 to 1.

In all the cases, the change-point, c was 100. As for the conventional method, the latest 20, 50, 100, and all the T tests were used. As for the proposed method, at each test, $c(T)$ was estimated, and the latest $(T - c(T))$ tests were used.

The performance of the proposed method in the case of no noise was computed. Figure 8.1 shows the probability of exact recovery in **Case1**, Figure 8.2 shows that in **Case2**, and Figure 8.3 shows that in **Case3**. "20 TESTS", "50 TESTS", "100 TESTS", and "ALL TESTS" mean respectively the case that the latest 20 tests were used, the case that the latest 50 tests were used, the case that the latest 100 tests were used, and the case that all the tests were used in the conventional method [2]. "PROPOSED" means the proposed method.

First, these results show that the probability of exact recovery of "ALL TESTS" did not increase along with the increase of number of tests after the change-point $c = 100$. This indicates that the case that all the tests were used in the conventional method is not robust to change of positive elements. In all the cases, the probability of exact recovery of the proposed method, i.e. "PROPOSED", converged to 1. In contrast, that of "20 TESTS" in **Case1**, that of "50 TESTS" in **Case2**, and that of "50 TESTS" in **Case3** converged to a value lower than 1. These results represent that the fixed numbers of the

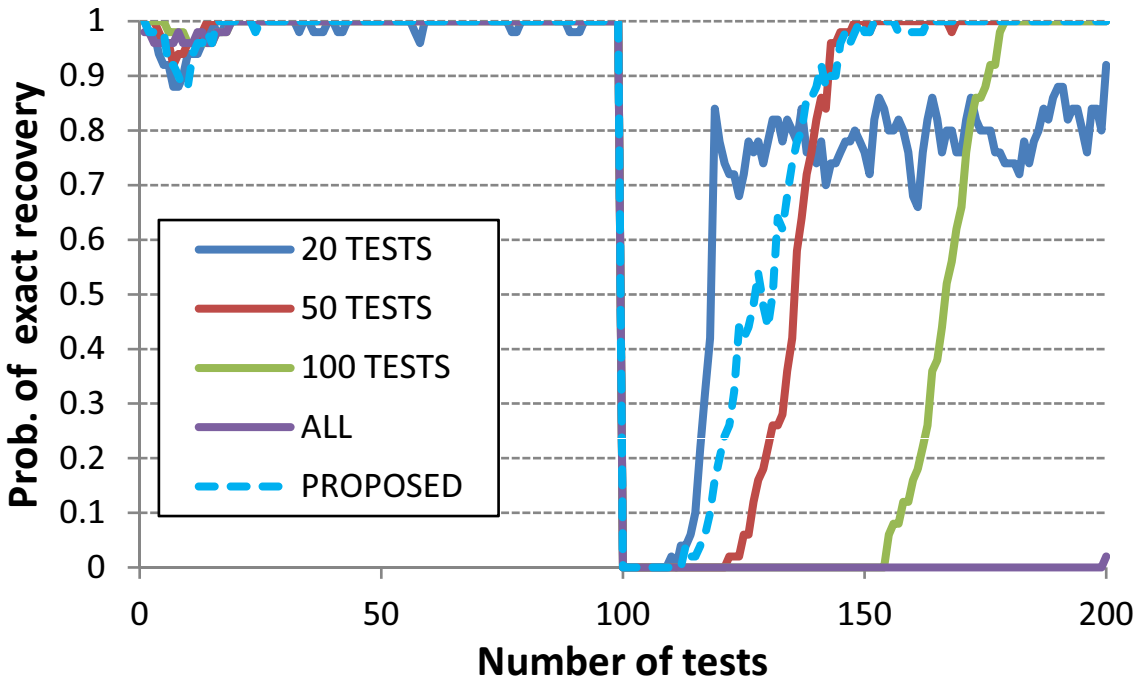


Figure 8.1: Probability of exact recovery as a function of number of tests, T, in the case that $\|\boldsymbol{x}(t)\|_0$ changed from 0 to 2. $N = 150$, the change-point c was 100, and 3% noise was added.

tests that were used were too small in these cases. In all the cases, the probability of exact recovery of "100 TESTS" converged to 1, however, the speed of the convergence was slower than that of the proposed method. These results represent that the fixed number of the tests that were used was too large in these cases. Thus, it is indicated that the proposed method is robust to change of positive elements.

8.5 Conclusion

A new method for solving the group-testing problem is proposed. To improve the robustness to the condition that positive elements change in the middle tests, the proposed method detects the latest change-point of positive elements, and it finds positive elements by using only the results of the tests after the change-point. To detect the change-point, the proposed method makes use of the fact that the distribution of the results depends on the number of positive elements. An experimental simulation showed that the proposed method outperforms the conventional method on the condition that positive elements

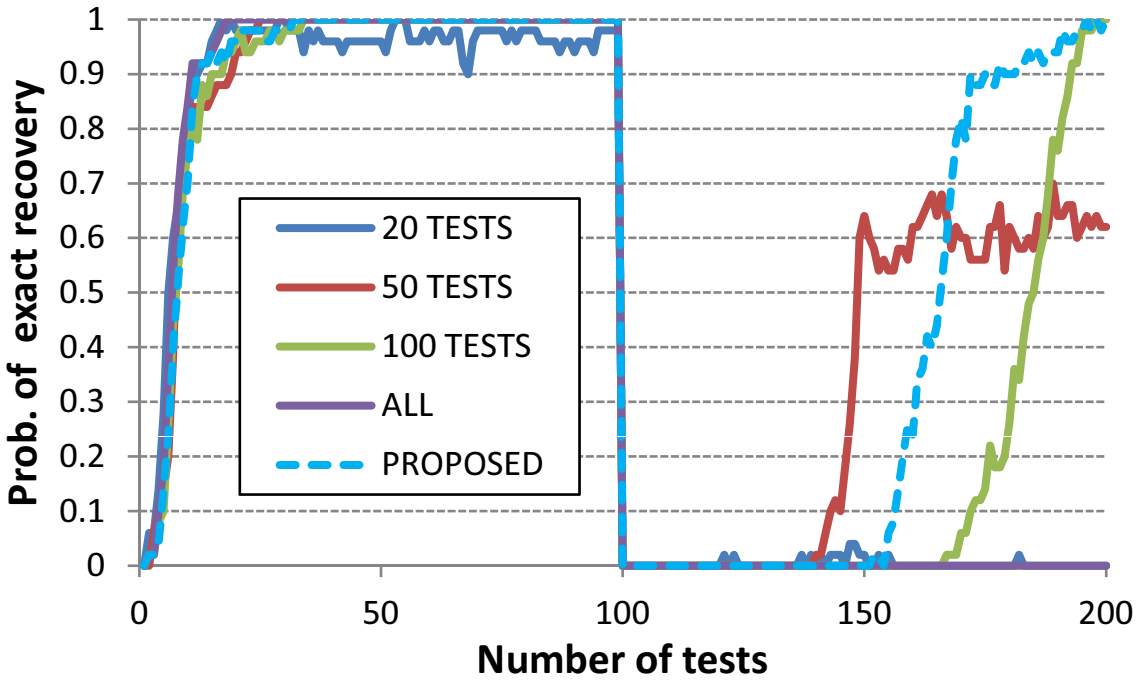


Figure 8.2: Probability of exact recovery as a function of number of tests, T, in the case that $\|\boldsymbol{x}(t)\|_0$ changed from 1 to 4. $N = 150$, the change-point c was 100, and 3% noise was added.

change in the middle tests.

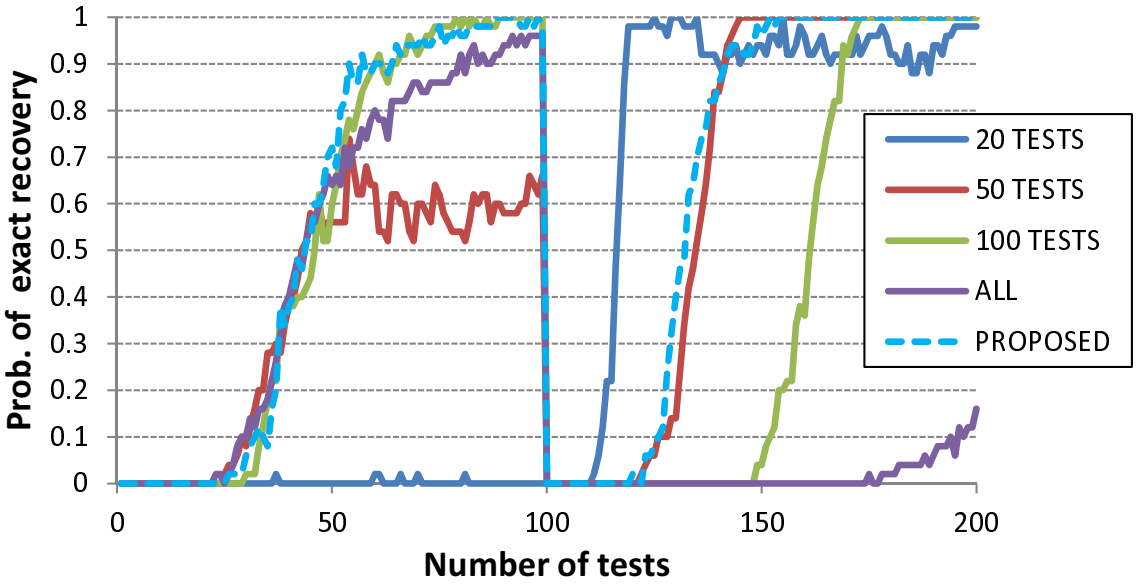


Figure 8.3: Probability of exact recovery as a function of number of tests, T, in the case that $\|\boldsymbol{x}(t)\|_0$ changed from 4 to 1. $N = 150$, the change-point c was 100, and 3% noise was added.

Chapter 9

Conclusions

9.1 Summary

In this study, for improving the detection-accuracy of the gate-type-system and the localization-speed of the large-area-monitoring-type-system, methods of sparsity-aware chemical signal processing were developed. In Chapter 3, for a walkthrough portal explosives-detection system, a signal-separation-method based on PLCA with a sparsity assumption was proposed. In an experimental evaluation, it was shown that the separation performance is high and that the proposed method can work in real time on GPU. In Chapter 4, an ICA-based acceleration of the signal-separation-method was proposed, and experimental results indicated that the proposed method can work in real time even on CPU. In Chapter 5, for reducing the uncertainty to improve the robustness, a signal-separation-method using an attenuation model was proposed. In an experimental evaluation, it was shown that the proposed method improves not only the robustness of separation but also the detection-accuracy. In Chapter 6, for the large-area-monitoring-type, a compressive sensing-based approach using a sparsity assumption was proposed to speed up localization of chemicals, and, especially, to achieve the robustness to the difference of the number of the positions of chemicals, adaptive Boolean compressive sensing was proposed. In Chapter 7, to improve the robustness to estimation errors, an extension of the adaptive Boolean compressive sensing into a multi-armed bandit algorithm was proposed, and the effectiveness was shown in simulation results. In Chapter 8, to improve the robustness to change of the location of chemicals, a combination of

change-point detection and the adaptive Boolean compressive sensing was proposed, and the effectiveness was shown in simulation results.

9.2 Remained problems and future works

In this dissertation, methods for improving the detection-accuracy of the gate-type-chemicals-detection-system were proposed. As explained in Chapter 2, signal processing for chemicals detection consists of pre-processing including separation, classification including detection, and quantification, and we focused on the improvement of separation in pre-processing because the noise problem is the most significant for practical use in real environments. However, not only in pre-processing, but also in detection and quantification, there is still room for improvement. There is a probability that an improvement of detection or quantification will further boost detection-accuracy.

In this dissertation, methods for accelerating the localization-speed of large-area-monitoring-type-system were proposed. Toward the practical use of localization of chemicals, things remain to be done. First, the group-size is determined by the proposed methods, whereas which samples should be selected can not be determined, and the proposed methods are based on a simple random sampling. If a way in which samples should be selected is found, the localization-speed will be more accelerated. Therefore, a way in which samples should be selected should be studied in the future. Second, in this dissertation, the measurement was modeled as group-testing, and the noise was also modeled as a random bit-flip independent on the mixing matrix like the conventional group-testing. However, in real cases, it can be predicted that the noise-level will depend on the number of ON-ducts, i.e. the group-size because the larger the number of ON-ducts, the lower the concentration of a target substance drawn into the detector. Thus, in the future, more precise modeling of the measurement needs to be studied. Finally, it is the most important to evaluate the proposed approach in a real environment in the future.

9.3 Concluding remarks

In recent years, the threat of such terrorism has become a serious problem. In this study, for preventing terrorist attacks with hazardous chemicals at train stations, airports, sports stadium, etc., signal processing techniques for systems that detects the hazardous chemicals were proposed. Especially, we studied methods for making the gate-type and the large-area-monitoring-type into practical use through the improvement of the detection-accuracy and the localization-speed respectively. This study was inspired by successes of sparsity-aware signal processing in many other areas than chemometrics, and we proposed methods of sparsity-aware signal processing for improving the detection-accuracy and the localization-speed. In this dissertation, first, to improve the detection-accuracy of the gate-type-system, a signal-separation-method based on a sparsity assumption was proposed. We showed that the proposed method improves not only the separation-performance but also the detection-accuracy. Also, it was shown that an online version of the proposed method can run in real time. Second, to speed-up localization of the large-area-monitoring-type-system, we proposed a compressive sensing-based approach that reduces the number of observations by using the sparsity assumption. Especially, to achieve the robustness to the difference of the number of the positions of chemicals, an idea of adaptive Boolean compressive sensing was proposed. In addition, to improve the robustness to estimation errors, a method based on the multi-armed bandit was applied; and to improve the robustness to change of the location of chemicals, a combination of change-point detection and the proposed method was applied; and simulation results showed that each proposed method has effectiveness.

Bibliography

[1] P.O. Hoyer, "Non-negative matrix factorization with sparseness constraints," *Journal of Machine Learning Research*, vol. 5, no. 7, pp. 1457–1469, 2004.

[2] D. Malioutov and M. Malyutov, "Boolean compressed sensing: LP relaxation for group testing," in *Proc. The 37th IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, 2012, pp. 3305–3308.

[3] Y. Takada, Y. Suzuki, H. Nagano, M. Sugiyama, E. Nakajima, M. Sugaya, Y. Hashimoto, and M. Sakairi, "High-throughput walkthrough detection portal as a measure for counter terrorism: Design of a vapor sampler for detecting triacetone triperoxide vapor by atmospheric-pressure chemical-ionization ion-trap mass spectrometry," *IEEE Sensors Journal*, vol. 12, no. 6, pp. 1673–1680, June 2012.

[4] The White House, "Countering improvised explosive devices," https://www.whitehouse.gov/sites/default/files/docs/cied_1.pdf, Feb. 2013.

[5] Japan Explosives Society, Ed., *Bakuhatsubutsu tanchi handobukku (Handbook on explosives detection)*, Maruzen, 2010.

[6] Y. Takada, H. Nagano, Y. Kawaguchi, H. Kashima, M. Sugaya, K. Terada, Y. Hashimoto, and M. Sakairi, "Automated trace-explosives detection for passenger and baggage screening," *IEEE Sensors Journal*, vol. 16, no. 5, pp. 1119–1129, Mar. 2016.

[7] Z. Bielecki, J. Janucki, A. Kawalec, J. Mikołajczyk, N. Pałka, M. Pasternak, T. Pustelny, T. Stacewicz, and J. Wojtas, "Sensors and systems for the detection of explosive devices - An overview," *Metrology and Measurement Systems*, vol. 19, no. 1, pp. 3–28, Mar. 2012.

[8] M. Tourné, "Developments in explosives characterization and detection," *Journal of Forensic Research*, vol. S12, no. 002, 2013.

[9] T. Osborn, W.A. Burns, J. Green, and S.W. Reeve, "An optical nose approach to explosive detection: One strategy for optically based sensing," *Spectroscopy*, vol. 26, no. 1, pp. 34–45, Jan. 2011.

[10] A.D. Wilson and M. Baietto, "Applications and advances in electronic-nose technologies," *Sensors*, vol. 9, no. 7, pp. 5099–5148, June 2009.

[11] S.A. McLuckey, D.E. Goeringer, K.G. Asano, G. Vaidyanathan, and Jr. J.L. Stephenson, "High explosives vapor detection by glow discharge-ion trap mass spectrometry," *Rapid Communications in Mass Spectrometry*, vol. 10, no. 3, pp. 287–298, 1996.

[12] A. Savitzky and M.J.E. Golay, "Smoothing and differentiation of data by simplified least squares procedures," *Analytical Chemistry*, vol. 36, no. 8, pp. 1627–1639, July 1964.

[13] V. Mazet, C. Carteret, D. Brie, J. Idier, and B. Humbert, "Background removal from spectra by designing and minimising a non-quadratic cost function," *Chemometrics and Intelligent Laboratory Systems*, vol. 76, no. 2, pp. 121–133, Apr. 2005.

[14] A. Antoniadis, J. Bigot, and S. Lambert-Lacroix, "Peaks detection and alignment for mass spectrometry data," *Journal of the French Statistical Society*, vol. 151, no. 1, pp. 17–37, Dec. 2010.

[15] W. Ilewicz, M. Kowalczuk, M. Niezabitowski, D. Buchczik, and A. Gałuszka, "Comparison of baseline estimation algorithms for chromatographic signals," in *The 20th International Conference on Methods and Models in Automation and Robotics (MMAR)*, 2015, pp. 925–930.

[16] L. Duval, L.T. Duarte, and C. Jutten, "An overview of signal processing issues in chemical sensing," in *Proc. The 38th IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, 2013, pp. 8742–9746.

[17] T.A. Lee, L.M. Headley, and J.K. Hardy, "Noise reduction of gas chromatography/mass spectrometry data using principal component analysis," *Analytical Chemistry*, vol. 63, no. 4, pp. 357–360, Feb. 1991.

[18] V.J. Barclay, R.F. Bonner, and I.P. Hamilton, "Application of wavelet transforms to experimental spectra: Smoothing, denoising, and data set compression," *Analytical Chemistry*, vol. 69, no. 1, pp. 78–90, Jan. 1997.

[19] A.A. Kardamakis, A. Mouchtaris, and N. Pasadakis, "Linear predictive spectral coding and independent component analysis in identifying gasoline constituents using infrared spectroscopy," *Chemometrics and Intelligent Laboratory Systems*, vol. 89, no. 1, pp. 51–58, Oct. 2007.

[20] J.M. Wells, W.R. Plass, G.E. Patterson, Z. Ouyang, E.R. Badman, and R.G. Cooks, "Chemical mass shifts in ion trap mass spectrometry: Experiments and simulations," *Analytical Chemistry*, vol. 71, no. 16, pp. 3405–3415, July 1999.

[21] R. Smith, D. Ventura, and J.T. Prince, "Lc-ms alignment in theory and practice: A comprehensive algorithmic review," *Briefings in Bioinformatics*, vol. 16, no. 1, pp. 104–117, Jan. 2015.

[22] R.G. Brereton, *Chemometrics for pattern recognition*, Wiley, 2009 edition, 2009.

[23] P.S. Gromski, H. Muhamadali, D.I. Ellis, Y. Xu, E. Correa, M.L. Turner, and R. Goodacre, "A tutorial review: Metabolomics and partial least squares-discriminant analysis - a marriage of convenience or a shotgun wedding," *Analytica Chimica Acta*, vol. 879, no. 1, pp. 10–23, Feb. 2015.

[24] S. Singh and M. Singh, "Explosives detection systems (EDS) for aviation security," *Signal Processing*, vol. 83, no. 1, pp. 31–55, Jan. 2003.

[25] M. Axelsson, O. Friman, I. Johansson, M. Nordberg, and H. Östmark, "Detection and classification of explosive substances in multi-spectral image sequences using linear subspace matching," in *Proc. The 38th IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, 2013, pp. 3492–3496.

[26] H. Ishida, Y. Wada, and H. Matsukura, "Chemical sensing in robotic applications: A review," *IEEE Sensors Journal*, vol. 12, no. 11, pp. 3163–3173, Nov. 2012.

[27] M.L. Cao, Q.H. Meng, M. Zeng, B. Sun, W. Li, and C.J. Ding, "Distributed least-squares estimation of a remote chemical source via convex combination in wireless sensor networks," *Sensors*, vol. 14, no. 7, pp. 11444–11466, June 2014.

[28] G.T. Nofsinger, *Tracking based plume detection*, PhD dissertation, Dartmouth College, 2006.

[29] G. Kowadlo and R.A. Russell, "Using naïve physics for odor localization in a cluttered indoor environment," *Autonomous Robots*, vol. 20, no. 3, pp. 215–230, June 2006.

[30] A.T. Hayes, A. Martinoli, and R.M. Goodman, "Distributed odor source localization," *IEEE Sensors Journal*, vol. 2, no. 3, pp. 260–271, June 2002.

[31] H. Ishida, G. Nakayama, T. Nakamoto, and T. Moriizumi, "Controlling a gas/odor plume-tracking robot based on transient responses of gas sensors," *IEEE Sensors Journal*, vol. 5, no. 3, pp. 537–545, June 2005.

[32] A.J. Lilienthal, A. Loutfi, and T. Duckett, "Airborne chemical sensing with mobile robots," *Sensors*, vol. 6, no. 11, pp. 1616–1678, Nov. 2006.

[33] A. Loutfi, S. Coradeschi, A.J. Lilienthal, and J. Gonzalez, "Gas distribution mapping of multiple odour sources using a mobile robot," *Robotica*, vol. 27, no. 2, pp. 311–319, Mar. 2009.

[34] V.P. Sesé, *Signal processing approaches to the detection and localization of gas chemical sources using partially selective sensors*, PhD dissertation, Universitat de Barcelona, 2011.

[35] Bruker, "Protecting mass transit system," https://www.bruker.com/applications/homelandsecurity/critical-infrastructure-protection/subway-protection.html.

[36] J. Yang, J. Wright, T.S. Huang, and Y. Ma, "Image super-resolution via sparse representation," *IEEE Transactions on Image Processing*, vol. 19, no. 11, pp. 2861–2873, Nov. 2010.

[37] J. Sulam and M. Elad, "Large inpainting of face images with trainlets," *IEEE Signal Processing Letters*, vol. 23, no. 12, pp. 1839–1843, Dec. 2016.

[38] M. Bertalmio, L. Vese, G. Sapiro, and S. Osher, "Simultaneous structure and texture image inpainting," *IEEE Transactions on Image Processing*, vol. 12, no. 8, pp. 882–889, Aug. 2003.

[39] A. Adler, V. Emiya, M.G. Jafari, M. Elad, R. Gribonval, and M.D. Plumbley, "Audio inpainting," *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 20, no. 3, pp. 922–932, Mar. 2012.

[40] A. Shashua and A. Levin, "Linear image coding for regression and classification using the tensor-rank principle," in *Proc. IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR)*, 2001, pp. I–42–I–49.

[41] J.-L. Starck, E.J. Candes, and D.L. Donoho, "The curvelet transform for image denoising," *IEEE Transactions on Image Processing*, vol. 11, no. 6, pp. 670–684, June 2002.

[42] A.Y. Ng, "Feature selection, L1 vs. L2 regularization, and rotational invariance," in *Proc. The 21st International Conference on Machine Learning (ICML)*, 2004, pp. 78–85.

[43] M. Vauhkonen, D. Vadasz, P.A. Karjalainen, E. Somersalo, and J.P. Kaipio, "Tikhonov regularization and prior information in electrical impedance tomography," *IEEE Transactions on Medical Imaging*, vol. 17, no. 2, pp. 285–293, Apr. 1998.

[44] Y. Zhou, S.-C. Huang, and M. Bergsneider, "Linear ridge regression with spatial constraint for generation of parametric images in dynamic positron emission tomography studies," *IEEE Transactions on Nuclear Science*, vol. 48, no. 1, pp. 125–130, Feb. 2001.

[45] S.S. Chen, D.L Donoho, and M.A. Saunders, "Atomic decomposition by basis pursuit," *SIAM Review*, vol. 43, no. 1, pp. 129–159, Jan. 2001.

[46] L. Daudet, "Audio sparse decompositions in parallel," *IEEE Signal Processing Magazine*, vol. 27, no. 2, pp. 90–96, Mar. 2010.

[47] M. Zhang, C. Desrosiers, Q. Qu, F. Guo, and C. Zhang, "Medical image super-resolution with non-local embedding sparse representation and improved ibp," in *Proc. The 41st IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, 2016, pp. 888–892.

[48] L.U. Perrinet and J.A. Bednar, "Sparse coding of natural images using a prior on edge co-occurences," in *Proc. The 23rd European Signal Processing Conference (EUSIPCO)*, 2015, pp. 2231–2235.

[49] A.Y. Aravkin, T. van Leeuwen, and N. Tu, "Sparse seismic imaging using variable projection," in *Proc. The 38th IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, 2013, pp. 2065–2069.

[50] Y. Li and A. Ngom, "Fast sparse representation approaches for the classification of high-dimensional biological data," in *Proc. The 12th IEEE International Conference on BioInformatics and BioEngineering (BIBE)*, 2012, pp. 306–311.

[51] J.A. Tropp, "Greed is good: Algorithmic results for sparse approximation," *IEEE Transactions on Information Theory*, vol. 50, no. 10, pp. 2231–2242, 2004.

[52] D.L. Donoho and M. Elad, "Optimally sparse representation in general (nonorthogonal) dictionaries via l1 minimization," *PNAS*, vol. 100, no. 5, pp. 2197–2202, 2003.

[53] M. Elad, "Optimized projections for compressed sensing," *IEEE Transactions on Signal Processing*, vol. 55, no. 12, pp. 5695–5702, 2007.

[54] Y. Li, A. Cichocki, and S. Amari, "Sparse component analysis for blind source separation with less sensors than sources," in *Proc. The 4th International Conference on Independent Component Analysis and Signal Separation (ICA)*, 2003, pp. 89–94.

[55] F. Georgiev, F. Theis, and A. Cichocki, "Blind source separation and sparse component analysis of overcomplete mixtures," in *Proc. The 29th IEEE International*

Conference on Acoustics, Speech, and Signal Processing (ICASSP), 2004, pp. V–493–V–496.

[56] W. Xu, M. Wang, J. F. Cai, and A. Tang, "Sparse error correction from nonlinear measurements with applications in bad data detection for power networks," *IEEE Transactions on Signal Processing*, vol. 61, no. 24, pp. 6175–6187, Dec. 2013.

[57] Z. Yang, Z. Wang, H. Liu, Y.C. Eldar, and T. Zhang, "Sparse nonlinear regression: Parameter estimation and asymptotic inference under nonconvexity," in *arXiv:1511.04514*, 2015.

[58] Y.R. Lau, L. Weng, K. Ng, and C. Chan, "Time-of-flight-secondary ion mass spectrometry and principal component analysis: Determination of structures of lamellar surfaces," *Analytical Chemistry*, vol. 82, pp. 2661–2667, 2010.

[59] M. Heikkinen, A. Sarpola, H. Hellman, J. Ramo, and Y. Hiltunen, "Independent component analysis to mass spectra of aluminium sulphate," *World Academy of Science, Engineering and Technology*, vol. 26, pp. 173–177, 2007.

[60] D. Mantini, F. Petrucci, P.D. Boccio, D. Pieragostino, M.D. Nicola, A. Lugaresi, G. Federici, P. Sacchetta, C.D. Ilio, and A. Urbani, "Independent component analysis for the extraction of reliable protein signal profiles from MALDI-TOF mass spectra," *Bioinfomatics*, vol. 24, no. 1, pp. 63–70, June 2008.

[61] P. Paatero and U. Tapper, "Positive matrix factorization: A non-negative factor model with optimal utilization of error estimates of data values," *Environmetrics*, vol. 5, no. 2, pp. 111–126, 1994.

[62] D.D. Lee and H.S. Seung, "Algorithms for non-negative matrix factorization," in *Proc. Advances in Neural Information Processing Systems (NIPS)*, 2001, pp. 556–562.

[63] P.W. Siy, R.A. Moffitt, R.M. Parry, Y. Chen, Y. Liu, M.C. Sullards, A.H. Merrill, and M.D. Wang, "Matrix factorization techniques for analysis of imaging mass spectrometry data," in *Proc. The 8th IEEE International Conference on BioInformatics and BioEngineering (BIBE)*, 2008, pp. 1–6.

[64] R. Dubroca, C. Junot, and A. Souloumiac, "Weighted NMF for high-resolution mass spectrometry analysis," in *Proc. The 20th European Signal Processing Conference (EUSIPCO)*, 2012, pp. 1806–1810.

[65] J.F. Gemmeke, T. Virtanen, and A. Hurmalainen, "Exemplar-based sparse representations for noise robust automatic speech recognition," *IEEE Transactions on Audio, Speech and Language Processing*, vol. 19, no. 7, pp. 2067–2080, 2011.

[66] D. Gong, X. Zhao, and Q. Yang, "Sparse non-negative pattern learning for image representation," in *Proc. The 15th IEEE International Conference on Image Processing (ICIP)*, 2008, pp. 981–984.

[67] S.K. Tjoa and K.J.R Liu, "Multiplicative update rules for nonnegative matrix factorization with co-occurrence constraints," in *Proc. The 35th IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, 2010.

[68] P. Smaragdis, B. Raj, and M. Shashanka, "A probabilistic latent variable model for acoustic modeling," in *Proc. NIPS Workshop on Advances in Models for Acoustic Processing*, 2006.

[69] B. Raj, R. Singh, M. Shashanka, and P. Smaragdis, "Bandwidth expansion with a Pólya urn model," in *Proc. The 32nd IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, 2007, pp. IV–597–IV–600.

[70] M. Shashanka, B. Raj, and P. Smaragdis, "Sparse overcomplete latent variable decomposition of counts data," in *Proc. Advances in neural information processing systems (NIPS)*, 2007, pp. 1313–1320.

[71] P. Smaragdis, M. Shashanka, B. Raj, and G.J. Mysore, "Probabilistic factorization of non-negative data with entropic co-occurrence constraints," in *Proc. The 8th International Conference on Independent Component Analysis and Signal Separation (ICA)*, 2009, pp. 330–337.

[72] M. Nakano, H. Kameoka, J. Le Roux, Y. Kitano, N. Ono, and S. Sagayama, "Convergence-guaranteed multiplicative algorithms for nonnegative matrix factorization with β-divergence," in *Proc. IEEE International Workshop on Machine Learning for Signal Processing (MLSP)*, 2010, pp. 283–288.

[73] C. Févotte, N. Bertin, and J-L. Durrieu, "Nonnegative matrix factorization with the Itakura-Saito divergence: With application to music analysis," *Neural Computation*, vol. 21, no. 3, pp. 793–830, 2009.

[74] R.K. Boyd, C. Basic, and R.A. Bethem, *Trace quantitative analysis by mass spectrometry*, Wiley, 2008 edition, 2008.

[75] M.D. Hoffman, "Poisson-uniform nonnegative matrix factorization," in *Proc. The 37th IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, 2012, pp. 5361–5364.

[76] R.J. Weiss and J.P. Bello, "Unsupervised discovery of temporal structure in music," *IEEE Journal of Selected Topics in Signal Processing*, vol. 5, no. 6, pp. 1240–1251, 2011.

[77] O. Dikmen and C. Févotte, "Maximum marginal likelihood estimation for nonnegative dictionary learning in the Gamma-Poisson model," *IEEE Transactions on Signal processing*, vol. 60, no. 10, pp. 5163–5175, 2012.

[78] G. Grindlay and D.P.W. Ellis, "A probabilistic subspace model for multi-instrument polyphonic transcription," in *Proc. The 11th International Society for Music Information Retrieval Conference (ISMIR)*, 2010, pp. 21–26.

[79] R. Badeau, N. Bertin, and E. Vincent, "Stability analysis of multiplicative update algorithms and application to non-negative matrix factorization," *IEEE Transactions on Neural networks*, vol. 21, no. 12, pp. 1869–1881, 2010.

[80] A. Hyvärinen, "Fast and robust fixed-point algorithms for independent component analysis," *IEEE Transactions on Neural Networks*, vol. 10, no. 3, pp. 626–634, May 1999.

[81] S. Amari, S.C. Douglas, A. Cichocki, and H.H. Yang, "Multichannel blind deconvolution and equalization using the natural gradient," in *Proc. The 1st IEEE Signal Processing Workshop on Signal Processing Advances in Wireless Communications (SPARC)*, 1997, pp. 101–104.

[82] N. Ono and S. Miyabe, "Auxiliary-function-based independent component analysis for super-Gaussian sources," in *Proc. The 9th International Conference on Latent Variable Analysis and Signal Separation (LVA/ICA)*, 2010, pp. 165–172.

[83] L. Zhao, G. Zhuang, and X. Xu, "Facial expression recognition based on PCA and NMF," in *Proc. The 7th World Congress on Intelligent Control and Automation (WCICA)*, 2008, pp. 6826–6829.

[84] Y. Kim and S. Choi, "A method of initialization for nonnegative matrix factorization," in *Proc. The 32nd IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, 2007, pp. II–537–II–540.

[85] C. Boutsidisa and E. Gallopoulosb, "SVD based initialization: Ahead start for nonnegative matrix factorization," *Pattern Recognition*, vol. 41, no. 4, pp. 1350–1362, 2008.

[86] M. Rezaei, R. Boostani, and M. Rezaei, "An efficient initialization method for nonnegative matrix factorization," *Journal of Applied Sciences*, vol. 11, no. 2, pp. 354–359, 1999.

[87] P. Smaragdis, B. Raj, and M. Shashanka, "Shift-invariant probabilistic latent component analysis," *MERL Technical Report TR2007-009*, 2007.

[88] S.C. Choia and R. Wette, "Maximum likelihood estimation of the parameters of the gamma distribution and their bias," *Technometrics*, vol. 11, no. 4, pp. 683–690, Nov. 1969.

[89] Y. Kawaguchi, H. Nagano, D. Matsubara, T. Kagehiro, Y. Takada, Y. Hashimoto, and A. Hiroike, "Real-time person tracking by using walk-through type explosives detector (in japanese)," in *Proc. The 25th Annual Conference of the Japanese Society for Articial Intelligence*, 2011, pp. 1I1–1in.

[90] P. Smaragdis, "Non-negative matrix factor deconvolution; extraction of multiple sound sources from monophonic inputs," in *Proc. The 5th International Conference on Independent Component Analysis and Signal Separation (ICA)*, 2004, pp. 494–499.

[91] T. Virtanen, "Monaural sound source separation by nonnegative matrix factorization with temporal continuity and sparseness criteria," *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 15, no. 3, pp. 1066–1074, Mar. 2007.

[92] N. Bertin, R. Badeau, and E. Vincent, "Enforcing harmonicity and smoothness in bayesian non-negative matrix factorization applied to polyphonic music transcription," *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 18, no. 3, pp. 538–549, Mar. 2010.

[93] T. Kagehiro, K. Yoneji, H. Kiyomizu, Y. Watanabe, Y. Kawaguchi, Z. Li, H. Nagano, and Y. Matsuda, "Traceable physical security systems for a safe and secure society," *Hitachi Review*, vol. 63, no. 5, pp. 254–258, 2014.

[94] R. Dorfman, "The detection of defective members of large populations," *Annals of Mathematical Statistics*, vol. 14, no. 4, pp. 436–440, Dec. 1943.

[95] D.Z. Du and F.K. Hwang, *Pooling designs and nonadaptive group testing: Important tools for DNA sequencing*, World Scientific, 2006.

[96] M.B. Malyutov, "On planning of screening experiments," in *Proc. 1975 IEEE-USSR Workshop on Information Theory*, 1976, pp. 144–147.

[97] M.B. Malyutov, "The separating property of random matrices," *Matematicheskie Zametki*, vol. 23, pp. 155–167, 1978.

[98] G. Atia and V. Saligrama, "Boolean compressed sensing and noisy group testing," *IEEE Transactions on Information Theory*, vol. 58, no. 3, pp. 1880–1901, 2012.

[99] D. Sejdinovic and O. Johnson, "Note on noisy group testing: Asymptotic bounds and belief propagation reconstruction," in *Proc. The 48th Annual Allerton Conference on Communication, Control, and Computing (Allerton)*, 2010, pp. 998–1003.

[100] M. Aldridge, L. Baldassini, and O. Johnson, "Group testing algorithms: Bounds and simulations," in *arXiv:1306.6438*, 2013.

[101] C.L. Chan, P.K. Che, S. Jaggi, and V. Saligrama, "Nonadaptive probabilistic group testing with noisy measurements: Near-optimal bounds with efficient algorithms," in *arXiv:1107.4540*, 2011.

[102] E. Candes and M.B. Wakin, "An introduction to compressive sampling," *IEEE Signal Processing Magazine*, vol. 24, no. 4, pp. 118–121, 2007.

[103] H. Robbins, "Some aspects of the sequential design of experiments," *Bulletin of the American Mathematical Society*, vol. 58, no. 5, pp. 527–535, 1952.

[104] V. Kuleshov and D. Precup, "Algorithms for the multi-armed bandit problem," *Journal of Machine Learning Research*, pp. 1–48, 2000.

[105] P. Auer, N. Cesa-Bianchi, and P. Fischer, "Finite-time analysis of the multiarmed bandit problem," *Machine Learning*, vol. 47, no. 2-3, pp. 235–256, May 2002.

List of publications

Journal papers

1. Y. Kawaguchi, M. Togami, H. Nagano, Y. Hashimoto, M. Sugiyama, and Y. Takada, "Separation of mass spectra based on probabilistic latent component analysis for explosives detection," IEICE Transactions on Fundamentals of Electronics, Communications and Computer Sciences, Vol.E98-A, No.9, pp. 1888-1897, Sep. 2015.

2. Y. Kawaguchi, M. Togami, H. Nagano, Y. Hashimoto, M. Sugiyama, and Y. Takada, "Mass spectra separation for explosives detection by using an attenuation model," IEICE Transactions on Fundamentals of Electronics, Communications and Computer Sciences, Vol.E98-A, No.9, pp. 1898-1905, Sep. 2015.

Peer-reviewed conference papers

1. Y. Kawaguchi and M. Togami, "Adaptive Boolean compressive sensing by sequential pool-design," in Proc. The Asia-Pacific Signal and Information Processing Association Annual Summit and Conference (APSIPA ASC), 2016, pp. 1-5.

2. Y. Kawaguchi and M. Togami, "Adaptive Boolean compressive sensing by using multi-armed bandit," in Proc. The 2016 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), 2016, pp. 1197-1201.

3. Y. Kawaguchi, T. Osa, H. Nagano, and M. Togami, "Improvement of robustness to change of positive elements in Boolean compressive sensing," in Proc. The 23rd European Signal Processing Conference (EUSIPCO), 2015, pp. 1197-1201.

4. Y. Kawaguchi, T. Osa, S. Barnwal, H. Nagano, and M. Togami, "Information-based pool size control of Boolean compressive sensing for adaptive group testing," in Proc. The 22nd European Signal Processing Conference (EUSIPCO), 2014, pp. 2280-2284.

5. Y. Kawaguchi, M. Togami, H. Nagano, Y. Hashimoto, M. Sugiyama, and Y. Takada, "Mass spectra separation for explosives detection by using PLCA with an attenuation model," in Proc. The 21st European Signal Processing Conference (EUSIPCO), 2013, pp. 1-5.

6. Y. Kawaguchi, M. Togami, H. Nagano, Y. Hashimoto, M. Sugiyama, and Y. Takada, "ICA-based acceleration of probabilistic latent component analysis for mass spectrometry-based explosives detection," in Proc. The 2013 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), 2013, pp. 2795-2799.

7. Y. Kawaguchi and M. Togami, "Conversation clustering based on PLCA using within-cluster sparsity constraints," in Proc. The 20th European Signal Processing Conference (EUSIPCO), 2012, pp. 619-623.

8. Y. Kawaguchi, M. Togami, H. Nagano, Y. Hashimoto, M. Sugiyama, and Y. Takada, "Mass spectra separation for explosives detection by using probabilistic latent component analysis," in Proc. The 2012 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), 2012, pp. 1665-1668.

9. Y. Kawaguchi, M. Togami, and Y. Obuchi, "Turn taking-based conversation detection by using DOA estimation," in Proc. The 11th Annual Conference of the International Speech Communication Association (INTERSPEECH 2010), 2010, pp. 3134-3137.

10. Y. Kawaguchi and M. Togami, "Soft masking based adaptation for time-frequency beamformers under reverberant and background noise environments," in Proc. The 18th European Signal Processing Conference (EUSIPCO 2010), 2010, pp. 736-740.

Academic activities

1. 電子情報通信学会信号処理研究専門委員（2014 年度から現在まで）