



HAL
open science

On-the-fly spectral unmixing based on Kalman filtering

Hugues Kouakou, José Henrique de M Goulart, Raffaele Vitale, Thomas Oberlin, David Rousseau, Cyril Ruckebusch, Nicolas Dobigeon

► To cite this version:

Hugues Kouakou, José Henrique de M Goulart, Raffaele Vitale, Thomas Oberlin, David Rousseau, et al.. On-the-fly spectral unmixing based on Kalman filtering. *Chemometrics and Intelligent Laboratory Systems*, 2024, 255, pp.105252. 10.1016/j.chemolab.2024.105252 . hal-04774748

HAL Id: hal-04774748

<https://hal.science/hal-04774748v1>

Submitted on 8 Nov 2024

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

On-the-fly spectral unmixing based on Kalman filtering

Hugues Kouakou^a, José Henrique de Morais Goulart^a, Raffaele Vitale^b, Thomas Oberlin^c, David Rousseau^d, Cyril Ruckebusch^b,
Nicolas Dobigeon^a

^aUniversité de Toulouse, IRIT/INP-ENSEEIH, 31071 Toulouse, France

^bUniversité de Lille, CNRS, LASIRE, 59000 Lille, France

^cUniversité de Toulouse, ISAE-SUPAERO, 31400 Toulouse, France

^dUniversité d'Angers, LARIS, UMR IRHS INRAe, 49000 Angers, France

Abstract

This work introduces an on-the-fly (i.e., online) linear spectral unmixing method which is able to sequentially analyze spectral data acquired on a spectrum-by-spectrum basis. After deriving a sequential counterpart of the conventional linear mixing model, the proposed approach recasts the linear unmixing problem into a linear state-space estimation framework. Under Gaussian noise and state models, the estimation of the pure spectra can be efficiently conducted by resorting to Kalman filtering. Interestingly, it is shown that this Kalman filter can operate in a lower-dimensional subspace to lighten the computational burden of the overall unmixing procedure. Experimental results obtained on synthetic and real Raman data sets show that this Kalman filter-based method offers a convenient trade-off between unmixing accuracy and computational efficiency, which is crucial for operating in an on-the-fly setting. The proposed method constitutes a valuable building block for benefiting from acquisition and processing frameworks recently proposed in the microscopy literature, which are motivated by practical issues such as reducing acquisition time and avoiding potential damages being inflicted to photosensitive samples. The code associated with the numerical illustrations reported in this paper is freely available online at <https://github.com/HKouakou/KF-OSU>.

Keywords: spectral unmixing, on-the-fly processing, Kalman filter, essential spectral information

1. Introduction

Spectral (or hyperspectral) imaging is a widely used measurement technique thanks to its ability to provide rich information about sample optical properties at different wavelengths. Applications arise in many domains ranging from pharmaceuticals [1, 2] where it can be used for checking the authenticity of drugs [3] to medicine [4–7], e.g., in order to detect diseases [7]. However, a common issue inherent to hyperspectral imaging is data analysis as mixed spectral pixels most often contain the contributions of several primary components constituting the sample under study, referred to as endmembers or pure spectra (PS) in the literature. Hence, one ubiquitous task associated to hyperspectral data consists in extracting these PS with their respective concentrations for each acquired spectral pixel. This process is called spectral unmixing (SU) [8] or multivariate curve resolution in the field of chemometrics [9].

Since the introduction of spectral mixture analysis in the 1970s, many SU algorithms have emerged. Under a linear mixing model (LMM), estimating the PS and their respective concentrations in each measured spectral pixel amounts to solving a nonnegative matrix factorization (NMF) problem [10]. NMF involves decomposing a matrix into a product of two nonnegative matrices. For spectral data, these two matrices contain the PS profiles and their respective concentrations.

A well-known method for solving the SU problem is multivariate curve resolution-alternating least squares (MCR-ALS) [9, 11], which formulates SU as a bilinear least-squares problem. This algorithm is based on the simple observation that, even though SU is a (jointly) nonconvex problem, it becomes (separately) convex when either the PS or the concentrations are kept fixed. MCR-ALS thus proceeds iteratively, by fixing one of those unknown matrices and solving the (easier) subproblem of minimizing the least-squares criterion only with respect to the other one, until some stopping criterion is met. Each subproblem is solved taking into account specific physically-driven constraints on the PS (e.g., nonnegativity, unimodality) and/or on the concentrations (e.g., positivity, sum-to-one).

Other works have proposed to formulate SU within a Bayesian framework. Such Bayesian methods rely upon postulating some prior distributions for the PS and for the concentrations, which are carefully chosen to incorporate the problem constraints. Then, given the measured observations, the posterior distributions of the unknown quantities (PS and concentrations) are determined. In practice, no closed-form expression of the Bayesian estimators can be derived and computational methods must be implemented. For instance, Markov chain Monte Carlo methods are used to approximate these estimates by empirically averaging samples drawn from the resulting posterior distribution of interest [12–14]. Another family of methods resorts to variational Bayesian inference to approximate this posterior by a simpler, instrumental distribution from

*This work has been supported by the ANR IMAGIN Research Project under grant agreement ANR-21-CE29-0007.

which closed-form expressions of the Bayesian estimates can be derived [15, 16].

Another class of methods, mainly motivated by geometric considerations, has also been developed. These methods can be divided into two groups, according to whether they rely or not on the assumption that at least some of the measurements are not mixed, but rather contain the sought PS. Under the LMM and a sum-to-one constraint on the concentrations, the observed spectra are convex combinations of pure spectra. From a geometric point of view, this means that those spectra lie within a simplex whose vertices are the PS, which are sometimes available among the collected data. A well-known algorithm which makes this assumption is vertex component analysis [17]. This latter algorithm iteratively finds PS by projections onto random axes orthogonal to the subspace spanned by the already extracted PS, so that the next PS is set as the observation with the maximal projection coefficient in absolute value. Other algorithms that rely on the observation of PS are the successive projection algorithm [18], pixel purity index [19] and N-finder [20]. However, this rather strong assumption is not fulfilled in many real-world scenarios. Therefore, other methods seek instead the simplex with the minimum volume enclosing the observations, without assuming the presence of observed PS. Among them, one can cite minimum volume enclosing simplex [21], minimum volume simplex analysis [22], simplex identification via split augmented Lagrangian [23] and minimum volume constrained nonnegative matrix factorization [24].

Regardless of the used algorithm, SU performance obviously depends on the quality of the data acquisition procedure. Acquiring high-quality data comes with its own challenges: often it is quite (or even prohibitively) time-consuming, and involves long exposure to high-power excitation signals, thus potentially inflicting damages on biological samples. To address these challenges, innovative acquisition protocols have been recently developed [25, 26] to target only essential spectral pixels, which are loosely defined as the most dissimilar spectral pixels of an image [27, 28]. In [29], those essential spectral pixels are found by identifying the convex hull of the collected data cloud in a subspace of reduced dimensionality which is determined via principal component analysis (PCA). Yet, PCA requires first acquiring a sufficiently large batch of pixels, which may be somehow limiting. Instead, Coic *et al.* have proposed to extract the essential spectra in the Fourier domain [25], by computing the convex hull within phasor plots [30, 31]. This strategy has shown to be well suited to a spectrum-by-spectrum acquisition approach based on the relevance of the collected signal profiles (i.e. whether they are essential spectra or not). This highlights the increasing need to be able to carry out on-the-fly (i.e., online) unmixing to update PS estimates after each new measurement. However, to the best of the authors' knowledge, none of the above spectral unmixing methods allows processing spectral pixels in an on-the-fly fashion with a reasonable computational cost compatible with operational requirements.

To satisfy this very current need, this work presents an on-the-fly unmixing method based on the Kalman filter (KF) [32]. The estimated PS are updated through the KF each time a new measured observation is available. Aiming to reduce the

computational cost of the estimation algorithm, this KF-based method is implemented in a lower-dimensional subspace. Still, we show how nonnegativity of the estimated PS in the original data space along the algorithm iterations can be ensured by a dedicated strategy, which draws inspiration from the archetypal analysis [33] to formulate this task as a constrained regression problem. Interestingly, the proposed procedure can also be granted with an optional subspace tracking step. This step dynamically accounts for substantial variations of the data subspace occurring, e.g., when a newly acquired measurement is likely composed of an additional PS. Numerical experiments conducted on synthetic and hyperspectral real Raman data sets demonstrate the ability of the proposed method to estimate PS on-the-fly while showing comparable results with state-of-the-art offline unmixing algorithms.

The remainder of the paper is organized as follows. Section 2 presents some related works to the general problem of source separation when the measurements may be delivered or processed sequentially. Section 3 recasts the conventional LMM in an on-the-fly setting, yielding the so-called on-the-fly LMM. The proposed algorithm to estimate the parameter of this on-the-fly LMM is described in Section 4. Experimental protocols relying on synthetic and real data sets are presented in Section 5. Results and discussions are reported in Section 6, while Section 7 concludes the paper.

2. On-the-fly spectral unmixing: related work

The problem of on-the-fly SU tackled in this paper can be framed as a source separation task that should be performed along the acquisition process, i.e., for processing a data stream comprising spectral data that are provided sequentially on a spectrum-by-spectrum basis. Online¹ source separation has in particular been applied with the objective of signal denoising, that is, for removing noise from the signals of interest measured in real-time, e.g., to suppress artefacts originating from eye movements when measuring electroencephalogram signals [34]. Several online algorithms based on independent component analysis have been proposed to solve this problem, see [35, 36]. Independent component analysis relies on the assumption that the measurements are linear combinations of statistically independent sources. Although the linearity hypothesis is relevant in the context of this work, the assumption of statistical independence between the sources (here, the concentrations) is generally not fulfilled.

Source separation has also found applications in Gamma-ray spectroscopy for radionuclide detection [37–39]. The authors in [39] recently proposed a method for performing this task in real time using a series of short-term measurements on the aerosol sample being analyzed. Although their approach is built upon the LMM, it includes a prior imposing that the concentrations of two consecutively processed observations must

¹The term “on-the-fly” used in this paper refers to the ability of updating the PS estimate once a new measurement or a batch of measurements is available. This is rather referred to as “online” in the signal processing literature. Both terms are used indistinctly from now on.

be close, which is not justifiable within our setting. Indeed, in our context the spectral pixels acquired between two successive measurements can be highly dissimilar, resulting in highly variable concentrations.

Online NMF and online dictionary learning (DL) define another class of methods for online source separation, where the basis vectors (that would stand for the PS in the context of SU) are updated after each new observation is recorded. Several online NMF and DL algorithms exist in the literature, including the ones described in [40–43]. These algorithms update the basis vectors from matrices containing information on previously estimated mixing coefficients (i.e., concentrations) and past observations, which are thus sequentially updated and stored. The algorithm proposed in [40] can be interpreted as an online counterpart of the one previously introduced in [44] as it complements the conventional NMF objective function with an additional sparsity-inducing regularization accounting for any deviations (i.e., outlier) from the conventional bilinear model generally adopted to describe spectral mixtures. In [41], the authors depart from the minimization of a standard quadratic cost to derive online instances of Itakura-Saito NMF [45] well suited to analyze audio signals. The seminal work of Mairal *et al.* [42] aims at adaptively learning a basis set from a data stream to perform sparse coding. Their optimization procedure is further improved in [40] to reach a better algorithmic convergence rate based on a robust stochastic approximation of the objective function, at the price of generating random samples over the probabilistic subspace spanned by the observations. Unfortunately, in their canonical implementations, all these methods do not perform any dimensionality reduction of the measurements before processing, which poses a complexity issue when dealing with high-dimensional hyperspectral data. Moreover, the robust and non-quadratic cost functions considered in [41] and [45] significantly deviate from the conventional bilinear model. More interestingly, the computationally lightest instance of online DL introduced in [42] shares some strong similarities with the approach described in the sequel of this paper. We will return to this point later in Section 4.6. However, it turns out that the online DL scheme of [42] performs poorly in the particular context of on-the-fly SU targeted in this work due to an inappropriate update rule, as we will discuss ahead.

Finally, it is worth mentioning that online unmixing has also received attention in recent years to process multi-temporal hyperspectral images acquired over the same scene at different time instants. In this context, online SU aims at estimating the PS as well as the concentrations each time a new image is acquired in order to identify possible variations in terms of PS and concentrations over the scene [46–49]. These methods are clearly not adapted to the context of SU of data delivered according to a spectrum-by-spectrum acquisition protocol, as they require acquiring the entire hyperspectral image at each time step.

3. On-the-fly formulation of LMM

Before delving into the problem statement, the notations for the key quantities used in the core part of the manuscript are

summarized in Table 1. Moreover, in what follows, $\mathbf{1}_K \in \mathbb{R}^{K \times 1}$ denotes a column vector composed of ones, \otimes stands for the Kronecker product, $|\mathcal{S}| = \text{card}(\mathcal{S})$ is the number of elements in the finite set \mathcal{S} .

Table 1: Summary of the symbols used in this article.

Symbol	Description
$K \in \mathbb{N}$	Number of pure spectra (PS)
$L \in \mathbb{N}$	Number of spectral channels
$\tilde{L} \in \mathbb{N}$	Dimension of the lower-dimensional space
$N \in \mathbb{N}$	Number of measurements
$t \in \mathbb{N}$	Time/measurement index
$\mathcal{P} \subset \{1, \dots, N\}$	Set of indices of the archetype spectra
$\mathbf{Y} \in \mathbb{R}^{N \times L}$	Matrix of the N measured spectra
$\mathbf{Y}_{S_t} \in \mathbb{R}_+^{L \times L}$	Matrix of the first t measured spectra
$\mathbf{y}_t \in \mathbb{R}_+^{L \times 1}$	Measured spectrum at time t
$\tilde{\mathbf{y}}_t \in \mathbb{R}_+^{\tilde{L} \times 1}$	Lower-dimensional representation of \mathbf{y}_t
$\mathbf{Y}_{\mathcal{P}} \in \mathbb{R}_+^{ \mathcal{P} \times L}$	Matrix of the archetype spectra
$\tilde{\mathbf{Y}}_{\mathcal{P}} \in \mathbb{R}_+^{ \mathcal{P} \times \tilde{L}}$	Matrix of the lower-dimensional representation of $\mathbf{Y}_{\mathcal{P}}$
$\mathbf{C} \in \mathbb{R}_+^{N \times K}$	Matrix of the N concentration vectors
$\mathbf{C}_{S_t} \in \mathbb{R}_+^{L \times K}$	Matrix of the first t concentration vectors
$\mathbf{c}_t \in \mathbb{R}_+^{K \times 1}$	Vector of the concentration associated with $\tilde{\mathbf{y}}_t$ and \mathbf{y}_t
$\mathbf{H}_t \in \mathbb{R}_+^{L \times LK}$	Matrix representation of the concentration vector \mathbf{c}_t
$\tilde{\mathbf{H}}_t \in \mathbb{R}_+^{\tilde{L} \times LK}$	Matrix representation of the concentration vector \mathbf{c}_t (lower-dimensional space)
$\mathbf{S} \in \mathbb{R}_+^{L \times K}$	Matrix of the PS
$\mathbf{s} \in \mathbb{R}_+^{L \times K \times 1}$	Vector representation of the PS matrix \mathbf{S}
$\hat{\mathbf{S}}_t \in \mathbb{R}_+^{L \times K}$	PS matrix estimate at time t (unconstrained, lower-dimensional space)
$\hat{\mathbf{S}}_t^+ \in \mathbb{R}_+^{L \times K}$	PS matrix estimate at time t (constrained, lower-dimensional space)
$\hat{\mathbf{S}}_t \in \mathbb{R}_+^{L \times K}$	PS matrix estimate at time t (unconstrained, data space)
$\hat{\mathbf{S}}_t^+ \in \mathbb{R}_+^{L \times K}$	PS matrix estimate at time t (constrained, data space)
$\mathbf{S}[k] \in \mathbb{R}_+^{L \times 1}$	k th column of \mathbf{S} corresponding to the k th PS
$\sigma_v^2 \in \mathbb{R}_+$	Process noise variance
$\sigma_e^2 \in \mathbb{R}_+$	Observation noise variance

3.1. Linear mixing model

This work addresses the problem of on-the-fly SU under the LMM, which assumes that each measured spectrum, that is to say each observation, is a linear combination of some unknown PS. The PS and their concentrations in each measured spectrum are supposed nonnegative. Moreover, this work considers also a sum-to-one (closure) constraint imposed on the concentrations. It is worth noting that this constraint can always be satisfied by the data after an appropriate normalization [50]. Assuming that the image to be acquired is composed of N measured spectral pixels, after a standard unfolding procedure, the LMM writes

$$\mathbf{Y} = \mathbf{C}\mathbf{S}^T + \mathbf{E}, \quad (1)$$

where $\mathbf{Y} \in \mathbb{R}_+^{N \times L}$ contains the full set of N L -dimensional spectra, $\mathbf{C} \in \mathbb{R}_+^{N \times K}$ is the concentration matrix where the n th row contains the K concentrations associated to the n th pixel, $\mathbf{S} \in \mathbb{R}_+^{L \times K}$ contains the K unknown pure spectra and $\mathbf{E} \in \mathbb{R}^{N \times L}$ is an additive white Gaussian noise matrix which models measurement noise and/or mismatches with respect to the LMM. The number of spectral channels is denoted L , and the number K of PS is assumed to be known a priori. Moreover, the matrices \mathbf{C} and \mathbf{S} are nonnegative, i.e., they are composed of non-

negative elements, and the sum-to-one constraint is imposed on the concentrations as $\mathbf{C}\mathbf{1}_K = \mathbf{1}_N$.

3.2. On-the-fly linear mixing model

In an on-the-fly microscopy setting, the spectral pixels composing the full scene are acquired sequentially and should be processed on a spectrum-by-spectrum basis. Without loss of generality but to ease the presentation, one can assume that the spectral pixels are arranged in the matrix $\mathbf{Y} = [\mathbf{y}_1, \dots, \mathbf{y}_N]^\top$ according to their order of acquisition, i.e., the vector $\mathbf{y}_t \in \mathbb{R}_+^L$ defining the t th row of \mathbf{Y} corresponds to the t th acquired spectrum ($t = 1, \dots, N$). Given a current estimate $\hat{\mathbf{S}}_{t-1}$ of the PS recovered from the set of $(t-1)$ acquired spectra $\mathbf{Y}_{\leq t-1} = [\mathbf{y}_1, \dots, \mathbf{y}_{t-1}]^\top$, one seeks to update it as soon as a new spectral pixel \mathbf{y}_t is acquired such that $\hat{\mathbf{S}}_t$ tends to the true (but unknown) PS matrix \mathbf{S} . A possible approach would consist in using any offline SU method each time a new spectral pixel \mathbf{y}_t is available. Doing so would give an estimate $\hat{\mathbf{S}}_t$ of the PS based on the whole set of pixels $\mathbf{Y}_{\leq t}$ acquired so far, such that $\mathbf{Y}_{\leq t} \approx \mathbf{C}_{\leq t} \hat{\mathbf{S}}_t^\top$ with $\mathbf{C}_{\leq t} = [\mathbf{c}_1, \dots, \mathbf{c}_t]^\top$ and $\mathbf{c}_t \in \mathbb{R}_+^K$. Yet, this scheme would exploit all available spectra to recompute an estimate at each instant t “from scratch,” which is not suitable for an on-the-fly approach because of the underlying computational burden. This naive strategy will be considered in Section 5 when comparing the proposed method to conventional unmixing algorithms. A question that arises is thus how to efficiently compute the PS estimate at instant t based on the information already available, i.e., exploiting the current estimate at instant $t-1$. One natural way of tackling this challenge is to define the PS matrix \mathbf{S}_t associated with the set of t measurements $\mathbf{Y}_{\leq t}$ as a random (matrix-valued) variable evolving according to $\mathbf{S}_t = f_t(\mathbf{S}_{t-1}) + \mathbf{V}_t$, for some suitable function $f_t(\cdot)$ and an appropriately defined random variable \mathbf{V}_t . According to this formulation, an estimate of \mathbf{S}_{t-1} can be leveraged as a prior information for estimating \mathbf{S}_t , together with assumptions on $f_t(\cdot)$ and on \mathbf{V}_t .

In this paper, we model the evolution of \mathbf{S}_t as a random walk without drift, i.e., $f_t(\cdot)$ is assumed to be the identity function and \mathbf{V}_t is a random matrix with zero-mean entries. This implies that \mathbf{S}_t is distributed around \mathbf{S}_{t-1} . More specifically, one adopts the following on-the-fly formulation of LMM

$$\mathbf{S}_t = \mathbf{S}_{t-1} + \mathbf{V}_t, \quad (2)$$

$$\mathbf{y}_t = \mathbf{S}_t \mathbf{c}_t + \mathbf{e}_t, \quad (3)$$

where $(\mathbf{V}_t)_{ij} \stackrel{\text{i.i.d.}}{\sim} \mathcal{N}(0, \sigma_v^2)$, $\mathbf{e}_t \sim \mathcal{N}(\mathbf{0}, \sigma_e^2 \mathbf{I}_L)$ and \mathbf{I}_L is the identity matrix of size L . The variance σ_v^2 can be interpreted as the uncertainty level associated to the PS estimate \mathbf{S}_t with respect to the previous estimate \mathbf{S}_{t-1} as a new spectrum \mathbf{y}_t arrives. The on-the-fly LMM defined by Eq. (2)–(3) states that the spectrum \mathbf{y}_t depends on the PS matrix \mathbf{S}_t according to the conventional (static) LMM, and this PS matrix evolves according to a recursion. On-the-fly unmixing can then be formulated as the updating of a previous PS estimate $\hat{\mathbf{S}}_{t-1}$ given a new acquired spectral pixel \mathbf{y}_t . This can be achieved within a Bayesian framework by deriving the (conditional) probability density function of \mathbf{S}_t given $\mathbf{y}_1, \dots, \mathbf{y}_t$. This approach is detailed in the next section.

4. The proposed on-the-fly spectral unmixing algorithm

4.1. Sequential estimation of pure spectra

For the sake of simplicity, this subsection assumes that the concentrations $\mathbf{c}_t \in \mathbb{R}_+^K$ are known; their estimation will be addressed later in Section 4.4. Moreover, the proposed on-the-fly LMM is vectorized by stacking the columns of \mathbf{S}_t and \mathbf{V}_t after applying the $\text{vec}(\cdot)$ operator. By using the identity $\text{vec}(\mathbf{S}_t \mathbf{c}_t) = (\mathbf{c}_t^\top \otimes \mathbf{I}_L) \text{vec}(\mathbf{S}_t) = \mathbf{H}_t \mathbf{s}_t$, the on-the-fly LMM defined as in (2) and (3) becomes

$$\mathbf{s}_t = \mathbf{s}_{t-1} + \mathbf{v}_t, \quad (4)$$

$$\mathbf{y}_t = \mathbf{H}_t \mathbf{s}_t + \mathbf{e}_t, \quad (5)$$

where $\mathbf{s}_t \in \mathbb{R}^{LK}$ (resp., \mathbf{v}_t) is the vectorized counterpart of \mathbf{S}_t (resp., \mathbf{V}_t) and $\mathbf{H}_t = \mathbf{c}_t^\top \otimes \mathbf{I}_L \in \mathbb{R}_+^{L \times LK}$. The posterior distribution of the unknown PS \mathbf{s}_t then writes

$$p(\mathbf{s}_t | \mathbf{Y}_{\leq t}) \propto p(\mathbf{y}_t | \mathbf{s}_t) p(\mathbf{s}_t | \mathbf{Y}_{\leq t-1}).$$

Given the Gaussian nature of the noise term \mathbf{v}_t and the error term \mathbf{e}_t , this posterior is also a Gaussian distribution $\mathcal{N}(\boldsymbol{\mu}_t, \boldsymbol{\Sigma}_t)$, thus fully defined by its mean $\boldsymbol{\mu}_t$ and covariance matrix $\boldsymbol{\Sigma}_t$. After each new measurement \mathbf{y}_t is recorded, these parameters $(\boldsymbol{\mu}_t, \boldsymbol{\Sigma}_t)$ can be recursively and explicitly computed from the previous parameters $(\boldsymbol{\mu}_{t-1}, \boldsymbol{\Sigma}_{t-1})$ according to updating rules specified by the Kalman filter (KF), as detailed in Algo. 1. More precisely, the KF updates the conditional mean $\boldsymbol{\mu}_t$ and covariance matrix $\boldsymbol{\Sigma}_t$ at each time instant t following two steps, namely prediction and update. The prediction step calculates the predicted covariance matrix $\boldsymbol{\Sigma}_{t-1/2}$ of the distribution $p(\mathbf{s}_t | \mathbf{Y}_{\leq t-1}) = \mathcal{N}(\mathbf{s}_t; \boldsymbol{\mu}_{t-1}, \boldsymbol{\Sigma}_{t-1/2})$. Once a new spectrum \mathbf{y}_t is available, the mean and covariance matrix are updated. This KF scheme is known to be optimal in the mean-square-error sense [32]. Besides, at each time instant t of the KF (i.e., after each new measurement is collected), it allows a Bayesian estimator to be computed. For example, the minimum mean-square-error estimate boils down to the posterior mean $\boldsymbol{\mu}_t$.

At this stage of the estimation process, two important shortcomings should be discussed. First, the steps of this algorithm can be, in practice, quite costly for spectra typically featuring hundreds or thousands of spectral variables. Thus, it is imperative to perform dimensionality reduction before executing the KF to make it feasible for on-the-fly SU. Second, in its canonical implementation, this KF-based unmixing procedure does not ensure the nonnegativity of the PS. The two next sections show that this estimation procedure can be carefully adapted to tackle these two challenges, i.e., reducing the computational cost via a dimensionality reduction and ensuring the nonnegativity constraint inherent to the PS.

4.2. Kalman filtering in a lower-dimensional subspace

As stated in the introduction, several approaches can be considered in order to perform dimensionality reduction. Certainly, the most popular approach consists in deploying PCA.

²Note that \mathbf{Z}_t is guaranteed to be invertible since it is the sum of a positive semi-definite matrix $\mathbf{H}_t \boldsymbol{\Sigma}_{t-1/2} \mathbf{H}_t^\top$ and a positive definite matrix $\sigma_e^2 \mathbf{I}_L$.

Algorithm 1: KFUpdate

Input : Newly acquired spectrum y_t , posterior mean μ_{t-1} and covariance matrix Σ_{t-1} at time instant $(t-1)$, observation matrix H_t , uncertainty level σ_v^2 , model noise variance σ_e^2

// Predict the prior covariance matrix

1 $\Sigma_{t-1/2} \leftarrow \Sigma_{t-1} + \sigma_v^2 I_{KL}$

// Compute the innovation

2 $r_t \leftarrow y_t - H_t \mu_{t-1}$

// Compute the innovation covariance

3 $Z_t \leftarrow H_t \Sigma_{t-1/2} H_t^\top + \sigma_e^2 I_L$

// Compute the optimal gain²

4 $K_t \leftarrow \Sigma_{t-1/2} H_t^\top Z_t^{-1}$

// Update the posterior mean

5 $\mu_t \leftarrow \mu_{t-1} + K_t r_t$

// Update the posterior covariance matrix

6 $\Sigma_t \leftarrow (I_L - K_t H_t) \Sigma_{t-1/2}$

Output: Posterior mean μ_t and covariance matrix Σ_t at time instant t .

One possible alternative capitalizes on the study conducted in [25], which demonstrated the suitability of the Fourier domain for identifying the essential spectral pixels. In what follows, for the sake of generality, this dimensionality reduction will not be prescribed. Instead, we will refer to a generic linear operator $DR : \mathbb{R}_+^L \rightarrow \mathbb{R}_+^{\tilde{L}}$ which performs dimensionality reduction according to

$$\tilde{y}_t = DR(y_t), \quad (6)$$

where $\tilde{y}_t \in \mathbb{R}_+^{\tilde{L}}$ (with $\tilde{L} \ll L$) denotes the lower-dimensional representation of the pixel spectrum $y_t \in \mathbb{R}_+^L$. It is worth noting that the use of the tilde character $\tilde{\cdot}$ refers to the corresponding quantity expressed in the target lower-dimensional subspace. Thanks to the linearity of the operator $DR(\cdot)$, the on-the-fly LMM (4)-(5) can be rewritten in the lower-dimensional subspace as

$$\tilde{s}_t = \tilde{s}_{t-1} + \tilde{v}_t, \quad (7)$$

$$\tilde{y}_t = \tilde{H}_t \tilde{s}_t + \tilde{e}_t, \quad (8)$$

with $\tilde{H}_t = c_t^\top \otimes I_{\tilde{L}}$. In Algo. 1, which remains unchanged, the KF can then be directly applied to the lower-dimensional representation \tilde{y}_t instead of y_t .

Besides, one valuable feature of KF is its ability of granting Bayesian point estimation with uncertainty quantification. Since now the KF operates in the lower-dimensional subspace induced by the linear operator $DR(\cdot)$, deriving confidence intervals associated with estimation of the lower-dimensional representation of the PS (e.g., expressed in terms of standard deviations) is straightforward in that subspace, as they come as a byproduct provided by KF. Concretely, this amounts to examining the diagonal elements of the error covariance matrix $\tilde{\Sigma}_t$ driving the filtering process. By contrast, translating these quantities into the original data space requires a careful calculation. The proposed methodology is detailed in Appendix C and the resulting confidence measures will be investigated in the numerical experiments (see Section 6).

4.3. Constrained estimation in a lower-dimensional subspace

Enforcing nonnegativity on the PS estimate \hat{S}_t expressed in the original data space would be rather straightforward: it could be achieved by a simple hard-thresholding of the KF output. Unfortunately, performing the sequential estimation in a lower-dimensional subspace makes this constraint more difficult to handle. Indeed, the KF provides an estimate \hat{S}_t of the lower-dimensional representation of the PS while the constraint should apply to the PS estimate \hat{S}_t defined in the original data space. To overcome this issue, the proposed strategy reformulates the constrained estimation problem as a constrained regression inspired by archetypal analysis [33]. More precisely, the nonnegative constrained PS estimate in the data space, herein denoted $\hat{S}_t^{(+)}$, is assumed to result from linear combinations of some so-called archetype spectra $\{y_t, t \in \mathcal{P}\}$ which span the signal subspace. The set $\mathcal{P} \subset \{1, \dots, N\}$ collects the indices of these archetype spectra which are gathered in the matrix $Y_{\mathcal{P}} \in \mathbb{R}^{|\mathcal{P}| \times L}$ with $|\mathcal{P}| = \text{card}\{\mathcal{P}\}$ and $K \leq |\mathcal{P}| \leq N$. Given these archetype spectra, the nonnegative constrained PS estimate in the data space is assumed to be expressed as

$$\hat{S}_t^{(+)} \approx Y_{\mathcal{P}}^\top \hat{R}_t. \quad (9)$$

In (9), the regression matrix \hat{R}_t depends on the unconstrained estimate \hat{S}_t and is defined as the solution of a constrained regression problem leveraging a geometrical interpretation of the LMM. The technical derivations required to its computation are reported in Appendix A.1 and Appendix A.2. The set \mathcal{P} gathering the indices of the archetype spectra can be chosen as the first $P \geq K$ measurements, i.e., $\mathcal{P} = \{1, \dots, P\}$ and $Y_{\mathcal{P}} = Y_{\leq P}$. When the signal subspace is expected to evolve along the acquisition, e.g., when a new component is likely to appear, this set should be updated dynamically. A simple yet efficient procedure for performing such a subspace tracking task is proposed in Appendix B.

Throughout Sections 4.1 to 4.3, the concentrations c_t ($t \geq 1$), or equivalently the observation model matrices \tilde{H}_t , have been assumed known, which is obviously not the case in practice. The next subsection discusses their estimation along the KF iterations (or time instants).

4.4. Estimation of concentrations

Once a spectral pixel y_t is observed at a given instant t , the associated concentration vector c_t , defining the observation matrix \tilde{H}_t , is required to update the PS estimate. Several strategies can be envisaged to infer this quantity. Driven by the constraint of computational efficiency compatible with on-the-fly processing, this work proposes to follow the same strategy advocated in recent works dedicated to online NMF [40, 43] and online DL [42]. More precisely, an estimate of c_t can be computed from the current constrained estimate $\hat{S}_{t-1}^{(+)}$ of the PS defined in (9) by solving the constrained regression problem

$$\min_c \left\| y_t - \hat{S}_{t-1}^{(+)} c \right\|_2^2 \quad \text{s. t.} \quad c \geq \mathbf{0} \quad \text{and} \quad c^\top \mathbf{1}_K = 1. \quad (10)$$

Various efficient off-the-shelf algorithms can be considered to solve this constrained minimization problem, such as the sparse

unmixing by variable splitting and augmented Lagrangian (SUnSAL) algorithm [51].

4.5. Overview of the proposed algorithm

Algorithm 2: Proposed KF-OSU algorithm for on-the-fly spectral unmixing

Input : Number K of PS, first P measured spectra $\mathbf{Y}_\mathcal{P}$, model noise variance σ_e^2 , uncertainty level σ_v^2

Initialization: Constrained PS estimate $\mathbf{S}_p^{(+)}$

- 1 $t \leftarrow P$
// Perform dimensionality reduction (see (6))
- 2 $\tilde{\mathbf{S}}_t^{(+)} \leftarrow \text{DR}(\mathbf{S}_t^{(+)})$
// Initialize the posterior distribution parameters
- 3 $\tilde{\boldsymbol{\mu}}_t \leftarrow \text{vec}(\tilde{\mathbf{S}}_t^{(+)})$
- 4 $\tilde{\boldsymbol{\Sigma}}_t \leftarrow \sigma_v^2 \mathbf{I}_{K\tilde{L}}$
// Update the PS estimate after each acquisition
- 5 **repeat**
- 6 $t \leftarrow t + 1$
// Estimate mixing concentrations \mathbf{c}_t by solving (10)
- 7 $\mathbf{c}_t \leftarrow \text{SUnSAL}(\mathbf{y}_t, \mathbf{S}_{t-1}^{(+)})$
- 8 $\tilde{\mathbf{H}}_t \leftarrow \mathbf{c}_t^\top \otimes \mathbf{I}_{\tilde{L}}$
// Perform dimensionality reduction (see (6))
- 9 $\tilde{\mathbf{y}}_t \leftarrow \text{DR}(\mathbf{y}_t)$
// Update the posterior distribution parameters (Algo. 1)
- 10 $(\tilde{\boldsymbol{\mu}}_t, \tilde{\boldsymbol{\Sigma}}_t) \leftarrow \text{KFUpdate}(\tilde{\mathbf{y}}_t, \tilde{\boldsymbol{\mu}}_{t-1}, \tilde{\boldsymbol{\Sigma}}_{t-1}, \tilde{\mathbf{H}}_t, \sigma_e^2, \sigma_v^2)$
// Update the unconstrained PS estimate in the subspace
- 11 $\tilde{\mathbf{S}}_t \leftarrow \text{unvec}(\tilde{\boldsymbol{\mu}}_t)$
// Optional: update set of archetype spectra (Appendix B)
- 12 $\mathcal{P} \leftarrow \text{SubspaceTracking}(\mathbf{y}_t, t, \mathbf{Y}_\mathcal{P}, \mathcal{P}, \sigma_e^2)$
// Compute the regression matrix by solving (A.4) (Algo. 3)
- 13 $\mathbf{R}_t \leftarrow \text{Regression}(\mathbf{Y}_\mathcal{P}, \tilde{\mathbf{Y}}_\mathcal{P}, \tilde{\mathbf{S}}_t)$
// Compute the constrained PS estimate in the full space
- 14 $\mathbf{S}_t^{(+)} \leftarrow \mathbf{Y}_\mathcal{P} \mathbf{R}_t$
// Perform dimensionality reduction (see (6))
- 15 $\tilde{\mathbf{S}}_t^{(+)} \leftarrow \text{DR}(\mathbf{S}_t^{(+)})$
// Update the posterior mean
- 16 $\tilde{\boldsymbol{\mu}}_t \leftarrow \text{vec}(\tilde{\mathbf{S}}_t^{(+)})$
- 17 **until** end of acquisition;

Output : The estimated PS matrix $\mathbf{S}_t^{(+)}$.

A step-by-step description of the proposed KF-based on-the-fly SU algorithm, termed KF-OSU, is sketched in Algo. 2. To lighten the notations, the symbol $\hat{\cdot}$ for denoting estimates of the various quantities of interest has been omitted, without any ambiguity. The algorithm takes as inputs the number of pure spectra K , the first P measured spectra defining the matrix $\mathbf{Y}_\mathcal{P}$ that can be chosen as the archetype spectra, the observation noise variance σ_e^2 and the uncertainty level σ_v^2 . The constrained PS estimate $\mathbf{S}_p^{(+)}$ can be initialized by applying any offline unmixing algorithm to $\mathbf{Y}_\mathcal{P}$. The parameters of the posterior distri-

butions are also initialized (lines 3-4): the mean $\tilde{\boldsymbol{\mu}}_p$ is initialized as the vectorized form of $\mathbf{S}_p^{(+)}$, and the covariance matrix $\tilde{\boldsymbol{\Sigma}}_p$ as $\sigma_v^2 \mathbf{I}_{K\tilde{L}}$.

Then, the algorithm proceeds iteratively by updating these parameters every time a new spectrum \mathbf{y}_t is acquired. The algorithm estimates the mixing concentrations \mathbf{c}_t by solving the optimization problem (10) via SUnSAL (line 7). The matrix representation $\tilde{\mathbf{H}}_t$ of the concentration vector \mathbf{c}_t is computed (line 8). The new spectrum \mathbf{y}_t is then projected onto the lower-dimensional space using the DR(\cdot) operator (line 9). At the core of the method, KFUpdate updates both the mean and covariance matrix using the new observation (line 10). The unconstrained PS estimate $\tilde{\mathbf{S}}_t$ is then computed by unvectorizing the updated mean (line 11). As an optional step, the set of archetype spectra can be updated according to the SubspaceTracking procedure (line 12). The nonnegative constrained PS estimate is defined as the linear combination of the archetype spectra (line 14) based on the regression matrix estimated using the Regression procedure (line 13). The posterior mean is finally defined after dimensionality reduction (line 15) and vectorization (line 16). The process repeats until the end of acquisition.

The code of the proposed KF-OSU algorithm can be found online at <https://github.com/HKouakou/KF-OSU>.

4.6. Connection to related methods

The core step of the proposed KF-OSU algorithm lies in the update of the PS estimate after each new measurement \mathbf{y}_t is acquired. In this work, this update is performed within a fully Bayesian framework by Kalman filtering (line 10 in Algo. 2). Conceptually, one may think of other updating strategies. This paragraph discusses these alternatives and relates them to the currently implemented KF-based updating rule.

As a preliminary, it should be noticed that the KF-based updating rule fully detailed in Algo. 1 can be compactly rewritten as

$$\mathbf{s}_t^{\text{KF}} = \mathbf{s}_{t-1}^{\text{KF}} + \mathbf{K}_t (\mathbf{y}_t - \mathbf{H}_t \mathbf{s}_{t-1}^{\text{KF}}) \quad (11)$$

where \mathbf{K}_t is the so-called optimal Kalman gain (see line 4 of Algo. 1) resulting from the assumptions of linear-Gaussian models for the state (4) and the observation (5). It mainly consists in first computing a residual $\mathbf{r}_t^{\text{KF}} = \mathbf{y}_t - \mathbf{H}_t \mathbf{s}_{t-1}^{\text{KF}}$, also referred to as innovation within the KF framework. This residual \mathbf{r}_t^{KF} acts as a drift, weighted by the gain matrix \mathbf{K}_t , to correct the previous estimate $\mathbf{s}_{t-1}^{\text{KF}}$.

Then, because of the simple state evolution model (4), it can be shown that there is a clear connection with the updating rule that would result from a recursive least square (RLS) filtering [52]. RLS aims at solving at each time instant t the weighted least-squares minimization problem

$$\min_{\mathbf{s}} \sum_{i=1}^t \lambda^{t-i} \|\mathbf{y}_i - \mathbf{H}_i \mathbf{s}\|_2^2, \quad (12)$$

where $\lambda \in (0, 1)$ is a forgetting factor. This leads to an RLS update rule of the form

$$\mathbf{s}_t^{\text{RLS}} = \mathbf{s}_{t-1}^{\text{RLS}} + \mathbf{G}_t(\lambda) (\mathbf{y}_t - \mathbf{H}_t \mathbf{s}_{t-1}^{\text{RLS}}), \quad (13)$$

where $\mathbf{G}_t(\lambda)$ can be interpreted as a RLS gain. Comparing (11) and (13), it clearly appears that KF and RLS differs by the definition of the gain applied to the correction drift.

Besides, as already discussed in Section 2, on-the-fly SU can also be cast into online NMF or online DL frameworks. In particular, the most popular strategy to perform online DL is certainly the stochastic optimization algorithm introduced by Mairal *et al.* in [42]. In its canonical implementation, updating the dictionary atoms (associated with the PS in the context of SU) is achieved by solving the optimization problem

$$\min_s \sum_{i=1}^t \|\mathbf{y}_i - \mathbf{S}\mathbf{c}_i\|_2^2. \quad (14)$$

Following the vectorization procedure described in Section 4.1, it can be shown that the original DL updating rule to solve (14) and given in [42] can be rewritten as

$$\mathbf{s}_t^{\text{DL}} = \mathbf{s}_{t-1}^{\text{DL}} + \mathbf{D}_t (\mathbf{y}_t - \mathbf{H}_t \mathbf{s}_{t-1}^{\text{DL}}) \quad (15)$$

with

$$\mathbf{D}_t = \left(\left(\sum_{i=1}^t \mathbf{c}_i \mathbf{c}_i^{\text{T}} \right)^{-1} \mathbf{c}_t \right) \otimes \mathbf{I}_L. \quad (16)$$

Again, by interpreting \mathbf{D}_t as a gain, the link between this DL update and the KF update (11) is clear.

Numerical experiments conducted on synthetic data sets empirically showed that replacing the KF update (line 10 in Algo. 2) with one of these two alternative updating rules, (13) or (15), did not improve or even severely degraded the PS estimation. This may be due to an unsuitable definition of the gain matrices in these alternative methods. As a consequence, they will not be considered further in the sequel of this paper.

5. Numerical experiments

This section describes the numerical experiments conducted to assess the performance of the proposed method on synthetic and real Raman data sets. It describes three synthetic data sets and a real data set used in the experiments, introduces two simulation protocols and describes the algorithms against which KF-OSU is compared. In addition to their ability to properly recover the sought PS, these methods will be compared with respect to their computational burden (i.e. algorithmic runtimes).

5.1. Description of the data sets

5.1.1. Synthetic data sets

Data sets SD1. Two hundred data sets, each composed of $N = 4000$ spectra with $L = 300$, have been independently generated from a mixture of $K = 5$ spectra according to the linear mixing model in (1). The profiles of the pure spectra have been generated by combining Gaussian templates with different locations, widths, and amplitudes. Mixture concentrations have been first randomly drawn from a Dirichlet distribution $\mathcal{D}(\alpha)$ guaranteeing their nonnegativity and their sum-to-one constraints. For these data sets, the Dirichlet parameter has been chosen as

$\alpha = [\alpha_1, \dots, \alpha_K]$ with $\alpha_k = 1$ (for all k), i.e., the mixing coefficients are uniformly distributed over the simplex defined by the constraints. Each generated data set contained pure spectra. Finally, the noise term \mathbf{E} was drawn from a white Gaussian distribution where the noise variance σ_e^2 has been adjusted to achieve a signal-to-noise ratio (SNR) of 20dB, as given by

$$\text{SNR}_{\text{dB}} = 10 \log_{10} \frac{\|\mathbf{C}\mathbf{S}^{\text{T}}\|_{\text{F}}^2}{\sigma_e^2 N L}. \quad (17)$$

These data sets containing pure spectra could appear too simplistic to relevantly contribute in assessing the performance of unmixing algorithms. However, such a simple simulation protocol is useful for benchmarking and has been for instance considered in [53, 54]. It is also worth recalling that the focus of this work is online unmixing, in the context of a sequential acquisition on a spectrum-by-spectrum basis. This particular context somehow questions the way in which the difficulty of achieving unmixing is usually discussed. To support this statement, one can refer to the experiments results reported in Section 6.1: even when analyzing these simple data sets, significant differences in unmixing performance will be observed depending on the order of the spectra defining the sequence of acquisition. Moreover, this simple data set has also the merit of setting one of the compared algorithms in a favorable experimental situation. Indeed, one of the geometrical algorithm considered in Section 5.5 specifically relies on a pure pixel assumption.

Data sets SD2. The generation procedure for this second set of data is almost the same as for SD1, with an important difference: all spectra containing more than 75% of one of the primary spectra have been discarded, to eliminate almost pure spectra and to make the unmixing problem more challenging.

Data sets SD3. The data sets SD3 have been generated similarly to SD2 ($N = 4000$, $K = 5$, $L = 300$, no pure spectra), with the key difference that 1000 out of the 4000 spectra in each data set are mixtures of only 4 components. More importantly, these data sets have been rearranged to mimic an acquisition scenario where the 5th component only appears later during the measurement process. Specifically, the mixing concentration of the 5th component is zero in the first 1000 measured spectra, meaning this component appears from the 1001st observation onward. The main rationale behind considering these data sets is to assess the subspace tracking step optionally embedded into the proposed KF-OSU algorithm. The experimental results will show that KF-OSU will be able to provide reliable results once $t \geq 1001$, even when initialized with $K = 5$ and an initial set of incomplete archetype spectra.

5.1.2. Real data set RD

A $101 \times 101 \times 343$ Raman image, depicted in Fig. 1 has been obtained from a mixture of sodium nitrate (NaNO_3), calcium carbonate (CaCO_3) and sodium sulphate (Na_2SO_4) powders. All spectral pixels have been acquired with a LabRAM HR microspectrometer (Horiba France SAS, Palaiseau, France) using a $50\times$ Olympus objective (0.75 NA). Data acquisition has

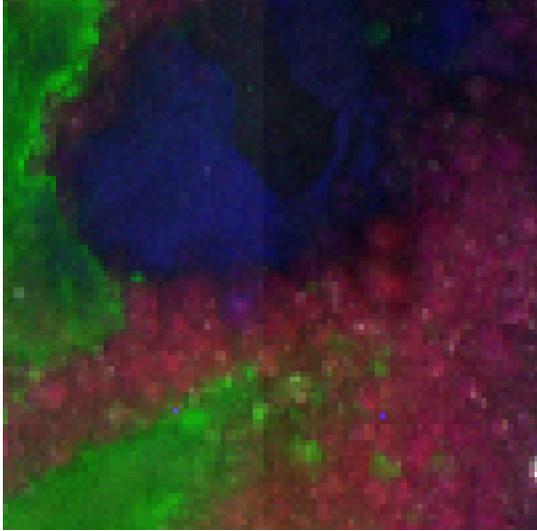


Figure 1: Data set RD: RGB visualization of the real Raman hyperspectral image used in the experiments. The red (R), green (G) and blue (B) channels correspond to Raman shifts 711.8, 622.9 and 723.6 cm^{-1} , respectively.

been performed in the spectral range 497.7–901.2 cm^{-1} with an accumulation time of 5 s per pixel (see [25] for more details on the acquisition conditions). This real data set denoted RD is interesting for assessing the performance of the compared methods as reference spectra, i.e. the spectral signatures of the three components NaNO_3 , CaCO_3 and Na_2SO_4 , are available. After unfolding the acquired hyperspectral image, a $N \times L$ matrix \mathbf{Y} of spectra is obtained with $L = 343$ and $N = 10201$. A batch of 199 copies of this matrix has been generated by randomly permuting the rows. The goal is to evaluate the performance of the proposed KF-OSU method for different initializations and sequences of observed spectra.

5.2. Acquisition protocols

Two distinct acquisition protocols have been simulated. For the first protocol denoted P1, the available spectral pixels are all considered and no preferred measurement sequence has been implemented. The second protocol, denoted P2, is inspired by the principle of essential spectral pixel acquisitions [25, 26]. More precisely, the data sets generated under protocol P2 consist of only essential spectral pixels extracted from the data resulting from protocol P1. These selected pixels are subsequently arranged in a specific order to maximize the spectral diversity (i.e., dissimilarity) between consecutive observed spectra. This order relies on an iterative *peeling* procedure of the data set conducted by identifying successive convex hulls, in a fashion similar to the strategy advocated in [55]. The procedure is fully described in Appendix D. The main motivation behind protocol P2 is to illustrate the impact of the order in which measurements are carried out and then made available for analysis by the on-the-fly unmixing algorithm. In particular, it will help to show that the convergence of KF-OSU can be significantly accelerated by providing the most informative measured spectra at an early stage of the estimation procedure.

5.3. Performance metrics

Four metrics have been taken into account to evaluate the performance of the algorithms under study. The first one is the spectral angle distance (SAD, in degrees), defined by

$$\text{SAD}_k = \arccos \left(\frac{\hat{\mathbf{S}}[k]^\top \mathbf{S}[k]}{\|\hat{\mathbf{S}}[k]\|_2 \|\mathbf{S}[k]\|_2} \right),$$

where $\mathbf{S}[k] \in \mathbb{R}^{L \times 1}$ (resp. $\hat{\mathbf{S}}[k] \in \mathbb{R}^{L \times 1}$) is the k th column of the matrix \mathbf{S} (resp. $\hat{\mathbf{S}}$) associated with the k th true (resp. estimated) PS. A small SAD value indicates a strong similarity between the profiles of the two spectra. To measure the average similarity over all the PS, the average SAD (aSAD) is reported

$$\text{aSAD} = \frac{1}{K} \sum_{k=1}^K \text{SAD}_k.$$

Note that this metric is insensitive to multiplicative factors, i.e. it does not allow to assess whether the estimated PS are scaled versions of the true ones. To capture this information, the root mean square error (RMSE) of the concentrations is additionally calculated as

$$\text{RMSE} = \sqrt{\frac{1}{KN} \|\mathbf{C} - \hat{\mathbf{C}}\|_{\text{F}}^2},$$

where \mathbf{C} (resp. $\hat{\mathbf{C}}$) is the true (resp. estimated) concentration matrix. Finally, to assess the ability of the compared methods to appropriately model the measurements, the spectral reconstruction error (RE), defined as

$$\text{RE} = \frac{\|\mathbf{Y} - \hat{\mathbf{C}}\hat{\mathbf{S}}^\top\|_{\text{F}}}{\|\mathbf{Y}\|_{\text{F}}}$$

is also computed. The matrix \mathbf{Y} contains the observed spectra and $\hat{\mathbf{S}}$ is the estimated PS matrix. In particular, RE is used to compare the performance of the methods when applied to the real data set RD, as the true PS and mixing concentrations are unknown.

It is worth noting that these four figures-of-merits will be computed as functions of the so-called *time index* of the compared unmixing algorithms. By this, one means that these metrics are computed after each new spectral pixel is acquired, in agreement with the considered operational context characterized by a sequential acquisition followed by an on-the-fly processing.

5.4. Details on the KF-OSU parameters

Archetype spectra – For each experiment conducted on the three synthetic data sets SD1–SD3 and the real data set RD, the first $P = 30$ spectra of each data set have been assigned to the matrix \mathbf{Y}_p of the archetype spectra to perform the regression task, i.e., $\mathcal{P} = \{1, \dots, 30\}$ (see (A.4) and line 13 of Algo. 2). The PS have been initialized as the outputs of VCA applied to this set of archetype measurements.

Noise and uncertainty levels – The measurement noise level σ_c^2 has been estimated from these P spectra after denoising using a

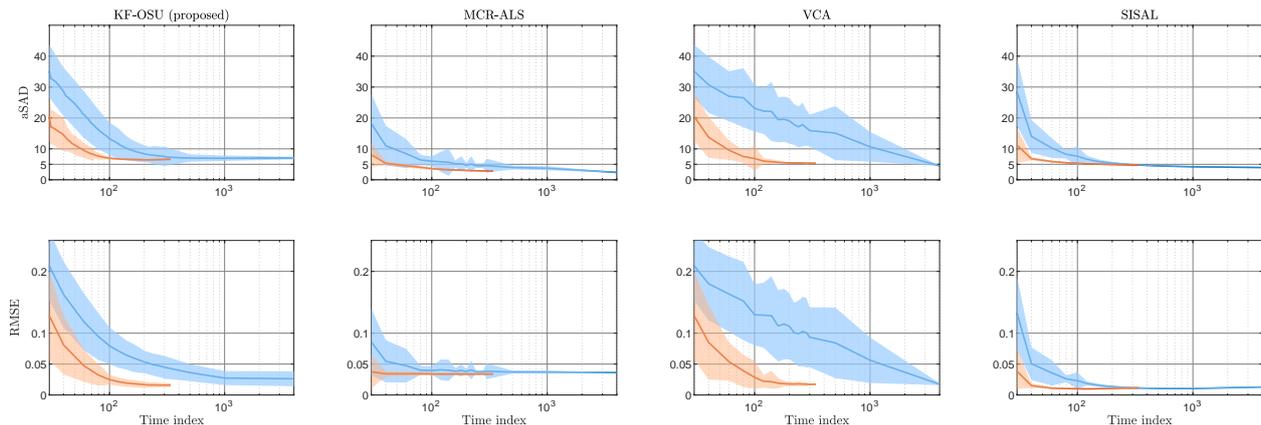


Figure 2: Synthetic data set SD1 – Performance of the compared algorithms in terms of aSAD (1st row) and RMSE (2nd row) as functions of the time index. Blue and orange lines refer to the acquisition protocols P1 and P2, respectively. The results have been averaged over 200 data sets and the shaded areas correspond to one standard deviation.

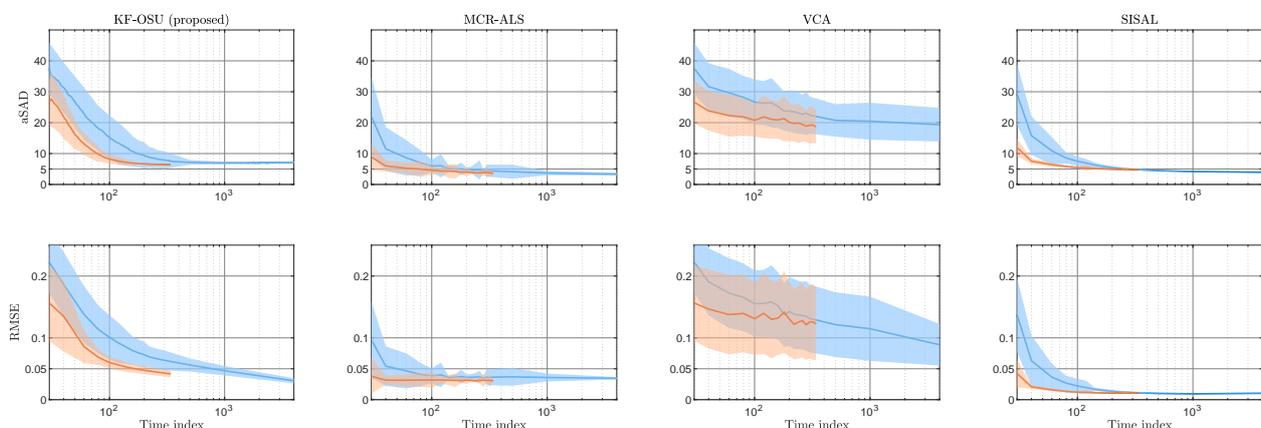


Figure 3: Synthetic data set SD2 – Performance of the compared algorithms in terms of aSAD (1st row) and RMSE (2nd row) as functions of the time index. Blue and orange lines refer to the acquisition protocols P1 and P2, respectively. The results have been averaged over 200 data sets and the shaded areas correspond to one standard deviation.

Savitzky-Golay filter of order 3 with a sliding window of size 5 [56]. To avoid excessive under- or over-estimation of this noise level, the results have been averaged over several portions of the spectra. Formally, each of these P spectra is split into Q segments of size $T = 10$ such that $QT \approx L$. The variance σ_e^2 of the noise is then set as the mean of the median variance calculated on each spectrum. Besides, it has been empirically observed that the uncertainty level σ_v^2 should be greater than or equal to 1; it has been set as $\sigma_v^2 = 1$ for all experiments.

Dimensionality reduction – As stated in Section 4.2, several choices for the operator $\text{DR}(\cdot)$ are possible. Two particular instances will be considered during the experiments. The most conventional solution to perform dimensionality reduction will consist in applying PCA, i.e.,

$$\tilde{\mathbf{y}}_t = \text{DR}(\mathbf{y}_t) = \mathbf{V}^T \mathbf{y}_t, \quad (18)$$

where $\mathbf{V} \in \mathbb{R}^{L \times \tilde{L}}$ is an orthogonal projector onto the estimated principal subspace of dimension \tilde{L} . In addition, another choice adopted in this paper will benefit from the findings reported in [25], which promotes the use of the Fourier coefficients for

identifying essential spectral pixels. More precisely, unless otherwise stated, the dimensionality reduction operator $\text{DR}(\cdot)$ used during the experiments will concatenate the real and imaginary parts of the discrete Fourier transform (DFT) of the input vector for a set of $M = \frac{L}{2}$ frequencies, i.e.,

$$\tilde{\mathbf{y}}_t = \text{DR}(\mathbf{y}_t) = \begin{bmatrix} \text{Re}(\mathbf{F}^T \mathbf{y}_t) \\ \text{Im}(\mathbf{F}^T \mathbf{y}_t) \end{bmatrix}, \quad (19)$$

where $\mathbf{F} \in \mathbb{C}^{L \times M}$ is the truncated DFT matrix. The motivation behind this dimensionality reduction approach, lies in the fact that the DFT can be directly computed on a spectrum-by-spectrum basis, without need of estimating the lower-dimensional subspace from data. Moreover, thanks to the computational efficiency of the fast Fourier transform algorithm, the use of DFT is compatible with real-time data processing required within the context of on-the-fly SU. For these two particular operators (PCA and DFT), the dimension of the lower-dimensional subspace will be estimated from the set \mathbf{Y}_φ of archetypal spectra as the minimum number P such that at least $\eta\%$ energy has been kept, i.e., such that the following cri-

terion is met

$$\sum_{t=1}^P \|\text{DR}(\mathbf{y}_t)\|_2^2 \geq \frac{\eta}{100} \sum_{t=1}^P \|\mathbf{y}_t\|_2^2.$$

Table 2 reports the (range of) values chosen for the parameters σ_e^2 , \tilde{L} and η for each data set.

Table 2: Algorithmic parameter values for each data set.

Data set	DR(\cdot)	\tilde{L}	η (%)	$\hat{\sigma}_e^2$ (estimated)	σ_e^2 (true value)
SD1	DFT	28–32	85–90	8	25
SD2	DFT	28–32	85–90	8	25
	PCA	65–90	97–98		
SD3	DFT	36–40	75–80	5	20
RD	DFT	52–60	85–88	52	-

5.5. Compared methods

The proposed algorithm is compared to three conventional unmixing methods described below:

- MCR-ALS is a popular unmixing algorithm in the chemometrics literature, thanks to its ease of implementation and its ability to produce good estimates of PS [9]. MCR-ALS requires as input the number of PS, the maximum number of iterations (set to 60 in this paper) and an initial estimate of PS chosen as the outputs of PCA.
- VCA is a well-known geometric method widely used by the remote sensing community thanks to its computational efficiency and its ability to identify PS among observations when they have been measured [17]. It is considered in the comparison since it has been used to initialize KF-OSU.
- SISAL is another well-known algorithm widely used by the remote sensing community [23]. Contrary to VCA, it does not assume that PS have been measured. SISAL shows a competitive (often better) processing time than its peers and generally produces good PS estimates. Besides the number of PS, it requires as input the maximum number of constrained quadratic problems to be solved, set to 80 in the experiments.

These algorithms are designed to process data in an offline context, i.e., to identify PS from a whole given set of measured spectra. To compare them to KF-OSU, they have been implemented to satisfy the operational requirements imposed by a sequential acquisition protocol combined with an on-the-fly spectral processing. More precisely, at each time instant t , these three algorithms are always run on the entire set of available acquired spectra $\mathbf{Y}_{\leq t}$. In other words, after each new spectrum \mathbf{y}_t is acquired, MCR-ALS, VCA and SISAL provide a new PS estimate which is computed from all previous acquired spectra

$\mathbf{Y}_{\leq t-1}$ augmented by this newly acquired spectrum \mathbf{y}_t . These three methods are expected to reach good estimation performance. However, it will be shown that their respective computational burden makes them inoperative in the targeted practical setup, imposing an on-the-fly unmixing of data acquired sequentially according to a spectrum-by-spectrum scheme.

6. Results and discussion

6.1. Performance on synthetic data sets SD1 and SD2

Figure 2 shows the performance of the compared methods on the data sets SD1. Overall, all algorithms yield good results, characterized by decreasing aSAD, RMSE and uncertainty (i.e., standard deviation) along the time index. Although the proposed method performs slightly worse than MCR-ALS and SISAL, it presents fairly low aSAD and RMSE values from time index $t = 500$ for protocol P1 (without the selection of essential spectra) and from time index $t = 200$ for protocol P2 (with the selection of essential spectra). This performance discrepancy with respect to the compared methods may be partly explained by the quite low SNR of the investigated data. Since KF-OSU operates in a purely on-the-fly context, a spectrum-by-spectrum processing may be less effective to mitigate the impact of noise. Conversely, the compared methods may benefit from the simultaneous analysis of the whole data sets, since they can infer the signal subspace at each time instant before identifying the PS.

Besides, it can be observed in Fig. 2 that the aSAD and RMSE curves of the four methods reach relatively close final values for the two acquisition protocols. However, it is important to highlight that under the second acquisition protocol P2, the aSAD and RMSE values decrease faster (in comparison with the experiment conducted under protocol P1) while the unmixing improves. This illustrates the benefit of resorting to such a smart acquisition protocol. In particular, one can notice that the curves associated to VCA reach a stationary behavior very late under protocol P1. This results from the fact that pure (or at least essential) spectra may be observed very late under protocol P1, in contrast to P2 which mimics an acquisition sequence maximizing the spectral diversity. This shows once again the relevance of a smart acquisition protocol in the context of on-the-fly unmixing.

All previous findings are corroborated by the outcomes resulting from the analysis of the data sets SD2, as shown in Fig. 3. The performance of KF-OSU, SISAL and MCR-ALS remains broadly the same as when the data sets SD1 are concerned. This demonstrates their ability to provide good PS estimates even in the absence of pure pixels among the data. As expected, VCA shows poorer results, since this geometrical method explicitly assumes that the sought PS belong to the set of measured spectral pixels.

Figures 4 and 5 display the profiles of the pure spectra estimated by the compared methods at time instant $t = 200$, i.e., after the 200th acquisition. They show that the estimated PS signatures are fairly close to the true PS. This demonstrates the ability of KF-OSU to perform on-the-fly unmixing without

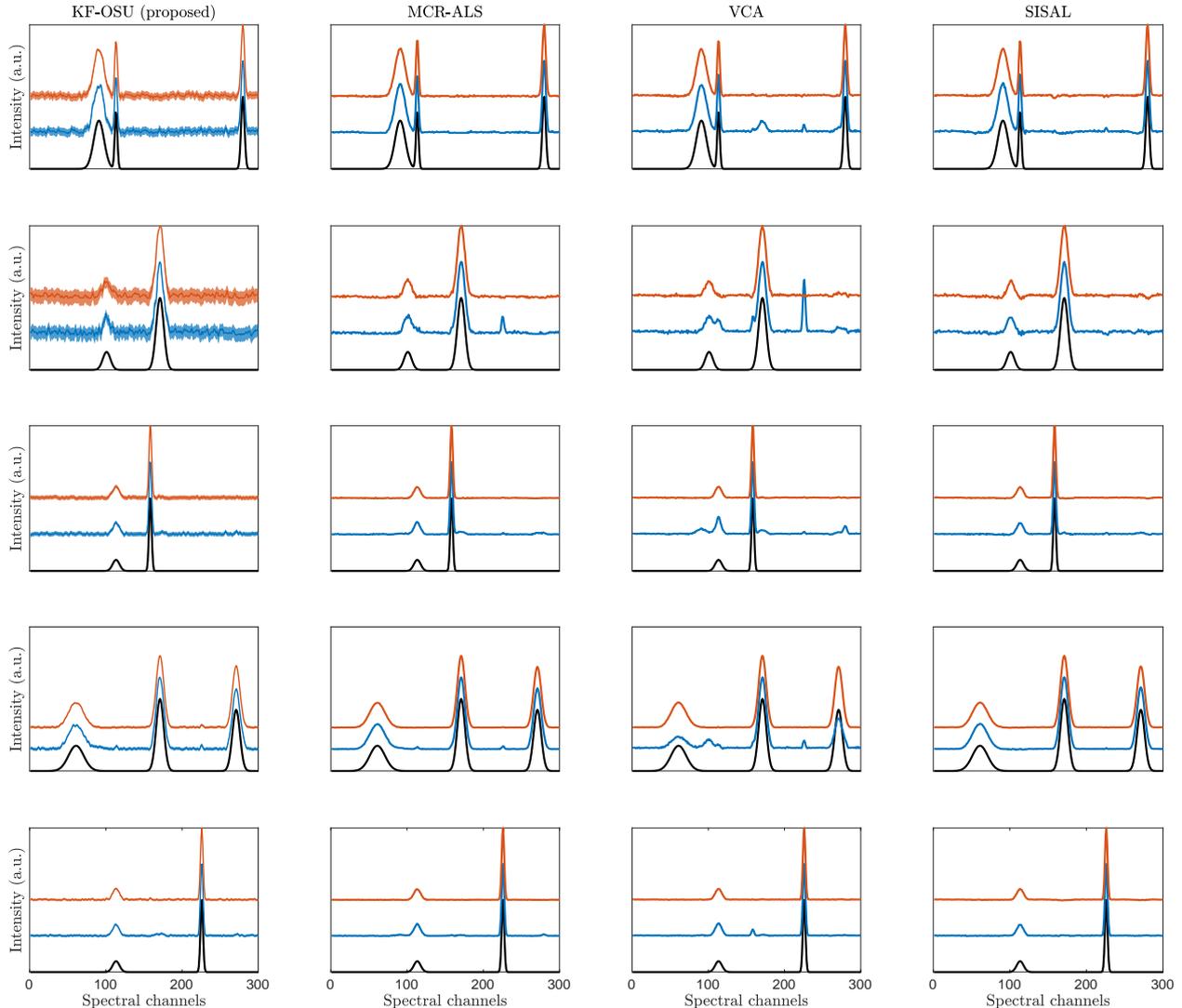


Figure 4: Synthetic data sets SD1 – Spectral profiles of the PS estimated at time index $t = 200$ for acquisition protocols P1 (blue) and P2 (orange). The black curves correspond to the ground-truth spectra. The shaded areas accompanying the results provided by the proposed KF-OSU represent the $\pm 3\sigma$ credible regions.

reprocessing all the previously measured spectra at each new acquisition. It is also worth noting that KF-OSU provides an estimation of the covariance matrix of the obtained estimates as a byproduct, following the strategy detailed in Appendix C. This estimation can be subsequently exploited to grant the estimated spectral with uncertainty quantification in terms of $\pm 3\sigma$ credible regions depicted as shaded areas in Fig. 4 and 5. This is clearly a valuable asset of the proposed algorithm inherited from its intrinsic Bayesian nature.

Finally, Figure 6 presents the performance of KF-OSU on the data sets SD2, when the dimensionality reduction previously operated by DFT was carried out by PCA. Two cases have been considered to demonstrate the relevance of the regression-based procedure to ensure the nonnegativity of the estimated pure spectra. In the first one, no further modification has been made to the KF-OSU algorithm (curves in green). In the second case, benefiting from standard properties of PCA-based dimensionality reduction, the regression-based procedure has been replaced

with a naive back-projection method. In this case, the estimated pure spectra are first back-projected into the original data space after each iteration of the Kalman filtering procedure, then they undergo a crude thresholding to impose nonnegativity (and finally they are projected again onto the lower-dimensional space for the next iteration). Specifically, lines 13–15 of Algo. 1 have been replaced by $\tilde{\mathbf{S}}_t^{(+)} \leftarrow \mathbf{V}^T \max(\mathbf{V}\tilde{\mathbf{S}}_t, 0)$ where $\max(\cdot, 0)$ is a component-wise thresholding operator. As observed, KF-OSU is less effective in the second case, highlighting the importance of enforcing nonnegativity through the regression-based procedure. Moreover, this experiment demonstrates that the DR operator can be easily chosen as PCA, although the latter generally requires a higher dimension P than DFT to achieve comparable performance (see Table 2).

6.2. Performance on the synthetic data sets SD3

Figure 7 presents the results of the KF-OSU algorithm applied to the data sets SD3 which mimic an acquisition scenario

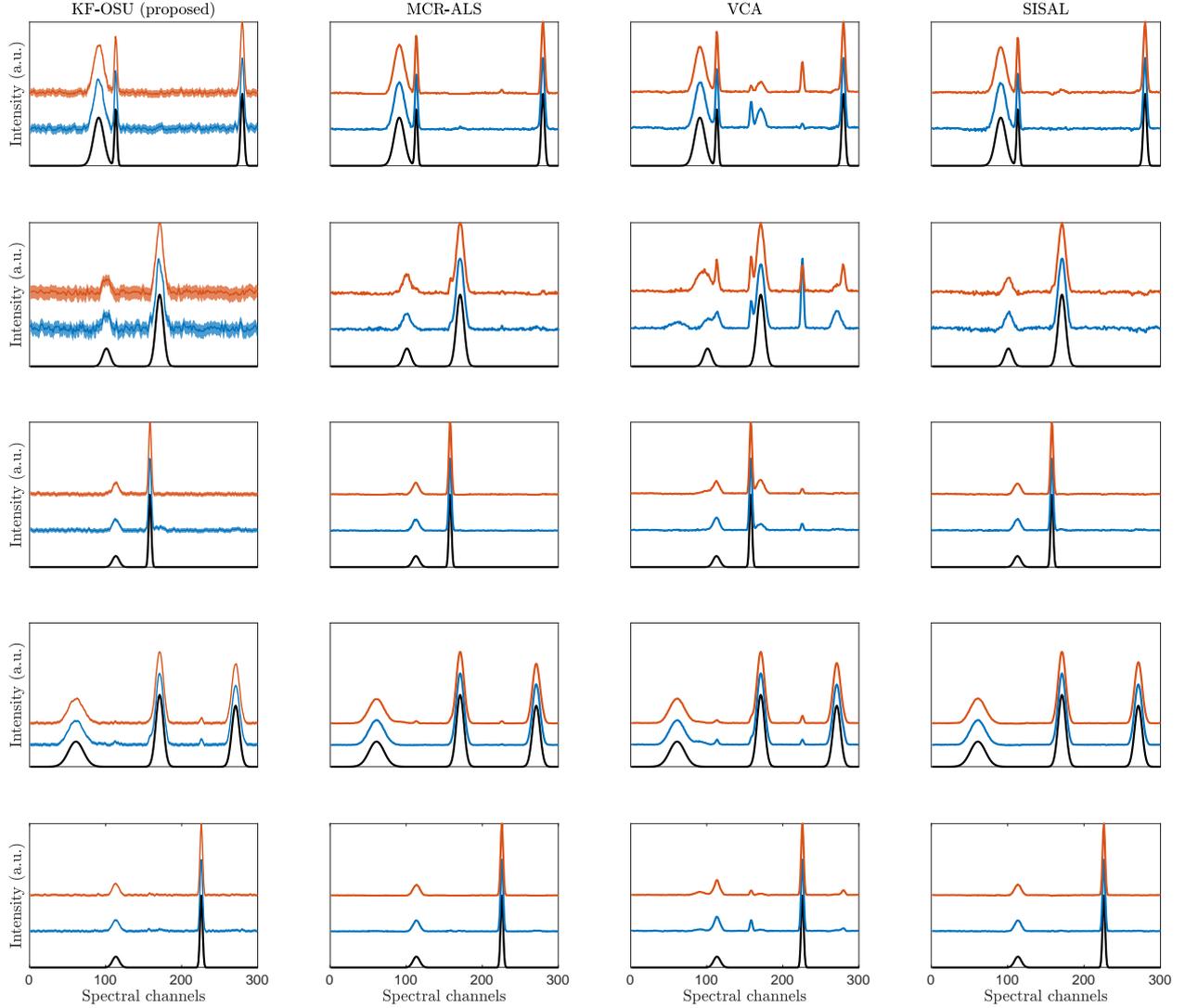


Figure 5: Synthetic data set SD2 – Spectral profiles of the PS estimated at time index $t = 200$ for acquisition protocols P1 (blue) and P2 (orange). The black curves correspond to the ground-truth spectra. The shaded areas accompanying the results provided by the proposed KF-OSU represent the $\pm 3\sigma$ credible regions.

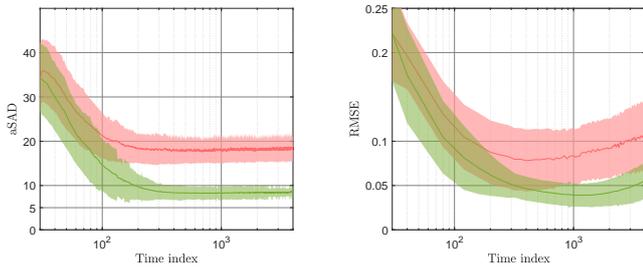


Figure 6: Synthetic data sets SD2 – Performance of KF-OSU with PCA-based dimensionality reduction in terms of aSAD (left) and RMSE (right). Green and red lines refer to KF-OSU with and without the regression-based procedure, respectively. The results have been averaged over 200 data sets and the shaded areas correspond to one standard deviation.

for which a new component is not present in the first 1000 measurements, but appears from the 1001th acquired spectra onward. The green curves represent the proposed KF-OSU algorithm when it is granted with the optional subspace track-

ing step, while the red curves correspond to its canonical version without subspace tracking. As for data sets SD1 and SD2, both versions of the algorithm have been initialized with the true number of PS (i.e., $K = 5$) and a set of archetypal spectra composed of the first $P = 30$ measurements. However, because of the way in which the data sets SD3 have been generated (see Section 5.1.1), these first P spectra do not contain the 5th PS. Therefore the matrix $Y_{\mathcal{P}}$ of archetypal spectra is expected to be unable to explain the measurements acquired after the time instant $t = 1001$. As expected, the two approaches exhibit comparable performance over the first 1000 iterations. Interestingly, when the 5th pure spectrum emerges from the 1001st measurement, only KF-OSU with subspace tracking adapts effectively. This is evidenced by the continuous decrease in the aSAD and RMSE curves until convergence is achieved. Notably, this convergence occurs rapidly after the introduction of the new component.

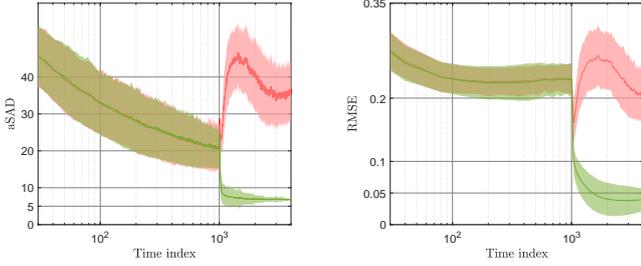


Figure 7: Synthetic data sets SD3 – Performance of KF-OSU in terms of aSAD (left) and RMSE (right). Green and red lines refer to KF-OSU with and without subspace tracking, respectively. The results have been averaged over 200 data sets and the shaded areas correspond to one standard deviation.

6.3. Performance on the real data set RD

Figure 8 depicts the RE as a function of the time index for each of the compared methods. These curves are complemented with a horizontal line corresponding to a so-called *lower bound* which approximates the lowest RE that can be expected given the noise level. This lower bound is defined as the RE with respect to \bar{Y} resulting from a denoising by PCA, i.e.,

$$\text{lower bound} = \frac{\|\mathbf{Y} - \bar{\mathbf{Y}}\|_F}{\|\mathbf{Y}\|_F} \quad (20)$$

with $\bar{\mathbf{Y}} = \mathbf{Y}\mathbf{V}\mathbf{V}^T$ where $\mathbf{V} \in \mathbb{R}^{L \times K}$ is the matrix composed of the first K principal component loadings ($K = 3$). These curves show that all methods behave similarly for each protocol with a clear decreasing of the RE along the time index, i.e., when the number of measured spectra increases. Under protocol P1, MCR-ALS and SISAL reach significantly lower RE (close to the lower bound) when all the spectral pixels are considered to perform unmixing. Under this same protocol, KF-OSU and VCA provide higher RE, even when all spectral pixels have been considered. In general, the RE obtained under protocol P2 are slightly smaller and, above all, they clearly decrease faster along the first 400 observations, with less dispersion around the mean.

It is worth noting that the slightly better results obtained by MCR-ALS and SISAL come with a higher computational burden, since all available spectral pixels are considered at each new measurement. This remains incompatible with the operational constraints imposed by a sequential acquisition protocol combined with an on-the-fly data processing, as it will be discussed in Section 6.4. In particular, they require a larger amount of spectra and, thus, a longer data collection time which, in the context of biological imaging, would significantly increase the risk of damaging the target specimens.

Finally, Fig. 9 shows that the profiles of the PS estimated by the methods at time index $t = 140$ can be easily matched to the reference spectra associated to the main components of the sample. They are globally in good agreement with these spectra that can be roughly considered as ground truth. Note that under the simulated acquisition protocol P1, a higher reconstruction error between Raman shifts 700 and 750 cm^{-1} is observed for the second PS.

Table 3: Average runtimes required by the compared SU algorithms. All the SU methodologies under study were implemented using MATLAB R2022b and run on a laptop with an Intel Core i7-8565U CPU running at 1.80GHz equipped with 8.00GB RAM.

Data set	N	KF-OSU (per obs.)	MCR-ALS (for N obs.)	VCA (for N obs.)	SISAL (for N obs.)
SD1	4000	0.007	1.38	0.04	0.15
SD2	4000	0.007	0.4	0.04	0.12
SD3	4000	0.009	0.4	0.04	0.12
RD	10201	0.005	0.45	0.08	0.25

6.4. Runtime comparison

Table 3 reports the computational times required by the compared algorithms for the processing of the synthetic data sets SD1 – SD3 and the real data set RD. It should be emphasized that KF-OSU updates the PS estimates at each time instant from a single newly available observation, i.e, the last acquired spectrum. Consequently, it has a constant execution time per time index. On the contrary, the other compared methods estimate the PS from the whole set of spectra available at each time instant, which results in computational complexities that increase along the time index. As expected, KF-OSU is shown to be computationally more suitable for on-the-fly SU.

Furthermore, the processing time required by MCR-ALS for the $N = 4000$ spectra of the data sets SD1 is much higher than the time required to analyze the data sets SD2 and SD3. This is certainly due to the absence of PS within the observations, resulting in a slower convergence of the algorithm. A similar observation can be made for SISAL, while KF-OSU takes the same execution time on both data sets.

When considering the real data set RD, KF-OSU shows a lower execution time compared to when it was run on the synthetic data sets SD1–SD3. This can be explained by the number of pure spectra to be estimated ($K = 3$), which is smaller than for the data sets SD1–SD3 ($K = 5$). This is not the case for MCR-ALS, VCA and SISAL, whose computational runtimes required to unmix the real data set RD are higher. In particular, VCA and SISAL execution times double with respect to the simulation experiments. This may be explained by the fact that their computational burden is also highly driven by the difficulty of the SU task when analyzing real data.

7. Conclusion

This work introduced a novel algorithm able to perform spectral unmixing in an on-the-fly setup. This method, named KF-OSU, was specifically designed to be compatible with sequential acquisition protocols delivering measurements on a spectrum-by-spectrum basis. Indeed, KF-OSU was able to update the pure spectra estimates after each individual spectrum was acquired. The evolution of the estimates as well as the spectral mixtures were described by linear Gaussian models, leading to simple and computationally efficient updating rules thanks to a Kalman filtering scheme. To further reduce the

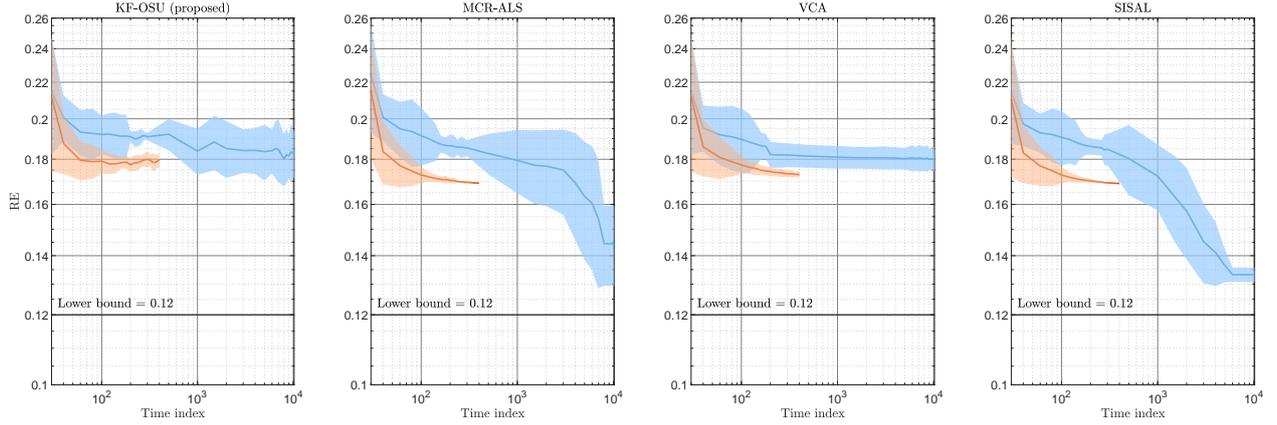


Figure 8: Real data set RD – Performance of the compared algorithms in terms of RE as a function of the time index. Blue and orange lines refer to the acquisition protocols P1 and P2, respectively. The results have been averaged over 200 data sets and the shaded areas correspond to one standard deviation.

computational cost, a dimensionality reduction was performed beforehand. The nonnegativity of the pure spectra estimate was ensured by solving a regression problem relating the pure spectra in the original data space and their representations in a lower-dimensional subspace.

Experiments conducted on synthetic and real Raman data sets showed that the estimation performance of KF-OSU was comparable to those obtained by standard unmixing methods, which were developed to estimate pure spectra from a whole set of measurements. Moreover, an experimental analysis of the runtimes required by the compared methods demonstrated the computational efficiency of KF-OSU, whose cost was found to be in agreement with the requirements imposed by a sequential data acquisition protocol combined with an on-the-fly spectral processing.

This method provides a support for the development of guided image acquisition or smart scanning schemes [25, 26]. Future work could explore extensions of the proposed algorithm to include mechanisms for detecting and correcting potential deviations from nominal behaviors.

Appendix A. Constrained PS estimation as a regression problem

This appendix shows that the constrained PS estimation can be formulated as an easily solvable constrained regression problem. Appendix A.1 explains how the PS estimation can be translated into a regression task. Its efficient resolution is detailed in Appendix A.2.

Appendix A.1. Problem formulation

To ease the presentation, we go back to the non-vectorized notation of the PS estimate at time instant t , i.e., we denote the lower-dimensional estimate recovered by the KF at time instant t as $\hat{\mathbf{S}}_t = \text{unvec}(\hat{\mathbf{s}}_t)$. This current estimate is supposed to be a rather good approximation of the true PS expressed in the lower-dimensional subspace, i.e., for $j \leq t$ it holds

$$\tilde{\mathbf{y}}_j \approx \hat{\mathbf{S}}_t \mathbf{c}_j. \quad (\text{A.1})$$

Reciprocally, akin to an archetypal analysis [33], it is also legitimate to state that the lower-dimensional representations of the PS can be expressed as linear combinations of some lower-dimensional representations of the measured spectral pixels, i.e., there exists a regression matrix $\mathbf{R}_t \in \mathbb{R}^{P \times K}$ such that

$$\hat{\mathbf{S}}_t \approx \tilde{\mathbf{Y}}_\varphi^\top \mathbf{R}_t, \quad (\text{A.2})$$

where the $|\mathcal{P}|$ columns of $\tilde{\mathbf{Y}}_\varphi \in \mathbb{R}^{|\mathcal{P}| \times L}$ contains the lower-dimensional representations $\tilde{\mathbf{y}}_t$ of some archetypal spectra \mathbf{y}_t , $t \in \mathcal{P}$ (with $K \leq |\mathcal{P}| \leq t$). The core idea consists in translating this regression into the original data space while ensuring the nonnegativity of the recovered PS. More precisely, the nonnegative constrained PS estimate in the data space denoted $\hat{\mathbf{S}}_t^{(+)}$ is assumed to be expressed as

$$\hat{\mathbf{S}}_t^{(+)} \approx \mathbf{Y}_\varphi^\top \hat{\mathbf{R}}_t, \quad (\text{A.3})$$

where the regression matrix $\hat{\mathbf{R}}_t$ is defined as the solution of the constrained regression problem

$$\hat{\mathbf{R}}_t = \underset{\mathbf{R}_t}{\text{argmin}} \left\| \tilde{\mathbf{Y}}_\varphi^\top \mathbf{R}_t - \hat{\mathbf{S}}_t \right\|_{\text{F}}^2 \quad \text{s. t.} \quad \mathbf{Y}_\varphi^\top \mathbf{R}_t \geq \mathbf{0}. \quad (\text{A.4})$$

The nonnegativity of the PS estimate defined in (A.3) is directly ensured by the nonnegativity constraint included in the regression problem (A.4). Note that incorporating this constraint into the original measurement space overcomes the non-invertibility of the dimensionality reduction induced by (6). Indeed, it only relies on the straightforward one-to-one mapping between the set of $|\mathcal{P}|$ archetypal spectra expressed in the original space and their representations with respect to the chosen dimensionality reduction operator, namely, \mathbf{Y}_φ and $\tilde{\mathbf{Y}}_\varphi$. We propose to solve the minimization problem in (A.4) by means of the alternating direction method of multipliers (ADMM) [57]. The algorithmic sketch of the regression procedure is detailed in the next section.

Appendix A.2. Resolution of the regression problem

The following derivations provide practical details on the resolution of the regression problem in (A.4). This task is

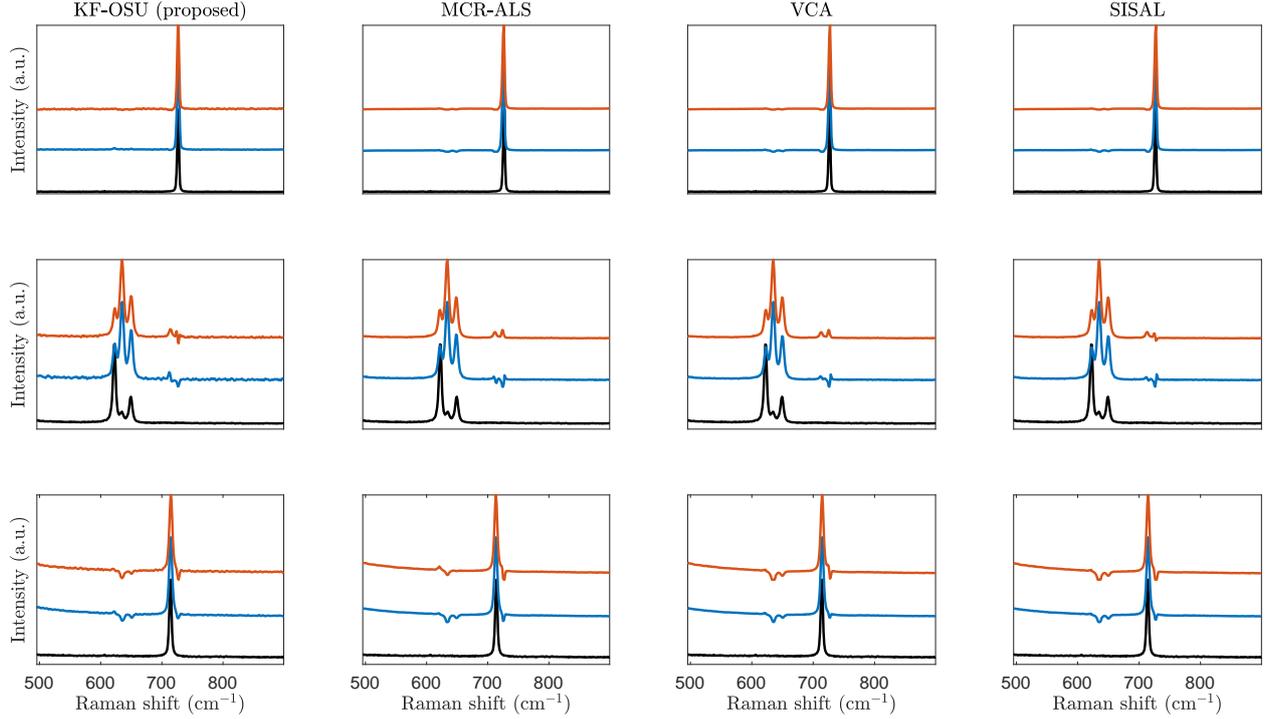


Figure 9: Real data set RD – Spectral profiles of the PS estimated at time index $t = 140$ for acquisition protocols P1 (blue) and P2 (orange). The black curves correspond to the reference spectra.

formulated as the minimization of a quadratic term subject to inequality constraints. After introducing a splitting variable $\mathbf{U} \in \mathbb{R}_+^{L \times K}$, the problem can be equivalently rewritten as

$$\min_{\mathbf{R}, \mathbf{U}} \left\| \tilde{\mathbf{Y}}_\varphi^\top \mathbf{R} - \hat{\mathbf{S}}_t \right\|_F^2 + \iota_{\mathbb{R}_+}(\mathbf{U}) \quad \text{s. t.} \quad \mathbf{U} - \mathbf{Y}_\varphi^\top \mathbf{R} = \mathbf{0}, \quad (\text{A.5})$$

where $\iota_{\mathbb{R}_+}(\cdot)$ is the indicator function such that $\iota_{\mathbb{R}_+}(\mathbf{U}) = 0$ if $\mathbf{U} \geq \mathbf{0}$ and $\iota_{\mathbb{R}_+}(\mathbf{U}) = \infty$ otherwise. The augmented Lagrangian associated to the constrained problem in (A.5) is

$$\mathcal{L}(\mathbf{R}, \mathbf{U}, \boldsymbol{\lambda}) = \left\| \tilde{\mathbf{Y}}_\varphi^\top \mathbf{R} - \hat{\mathbf{S}}_t \right\|_F^2 + \frac{\rho}{2} \left\| \mathbf{U} - \mathbf{Y}_\varphi^\top \mathbf{R} \right\|_F^2 + \langle \boldsymbol{\lambda}, \mathbf{U} - \mathbf{Y}_\varphi^\top \mathbf{R} \rangle + \iota(\mathbf{U})$$

where $\langle \mathbf{X}, \mathbf{Z} \rangle = \text{trace}(\mathbf{X}^\top \mathbf{Z})$ and $\boldsymbol{\lambda}$ is the Lagrange multiplier. Minimizing (A.5) can thus be efficiently achieved using ADMM whose scheme is detailed in Algo. 3. Solutions to lines 3 and 4 are given respectively by

$$\mathbf{R}^{(j+1)} = \left(2\tilde{\mathbf{Y}}_\varphi \tilde{\mathbf{Y}}_\varphi^\top + \rho \mathbf{Y}_\varphi \mathbf{Y}_\varphi^\top \right)^{-1} \times \left(\mathbf{Y}_\varphi \boldsymbol{\lambda}^{(j)} + \rho \mathbf{Y}_\varphi \mathbf{U}^{(j)} + 2\tilde{\mathbf{Y}}_\varphi \hat{\mathbf{S}}_t \right) \quad (\text{A.6})$$

and

$$\mathbf{U}^{(j+1)} = \max\left(\mathbf{0}, \mathbf{Y}_\varphi^\top \mathbf{R}^{(j+1)} - \frac{1}{\rho} \boldsymbol{\lambda}^{(j)}\right). \quad (\text{A.7})$$

It appears that the number $|\mathcal{P}|$ of archetype spectra should be chosen such that the matrix to be inverted in (A.6) is well-conditioned, or at least, nonsingular. For the experiments described in Section 5, the stopping criterion of Algo. 3 is chosen as a maximum of 50 iterations with $\rho = 1$ and $\mathbf{U}^{(0)}$ and $\boldsymbol{\lambda}^{(0)}$ initialized as null matrices.

Algorithm 3: Regression

Input : The regressors \mathbf{Y}_φ in the original data space, their counterparts $\tilde{\mathbf{Y}}_\varphi$ in the lower-dimensional subspace, the PS estimate $\hat{\mathbf{S}}_t$ expressed in the lower-dimensional subspace, the step-size $\rho > 0$

Initialization: $\mathbf{U}^{(0)}, \boldsymbol{\lambda}^{(0)}$

```

1  $j \leftarrow 0$ 
2 repeat
3   // Update the regression matrix (see (A.6))
    $\mathbf{R}^{(j+1)} \in \text{argmin}_{\mathbf{R}} \mathcal{L}(\mathbf{R}, \mathbf{U}^{(j)}, \boldsymbol{\lambda}^{(j)})$ 
4   // Update the splitting variable (see (A.7))
    $\mathbf{U}^{(j+1)} \in \text{argmin}_{\mathbf{U}} \mathcal{L}(\mathbf{R}^{(j+1)}, \mathbf{U}, \boldsymbol{\lambda}^{(j)})$ 
5   // Update the Lagrange multiplier
    $\boldsymbol{\lambda}^{(j+1)} \leftarrow \boldsymbol{\lambda}^{(j)} + \rho(\mathbf{U}^{(j+1)} - \mathbf{Y}_\varphi^\top \mathbf{R}^{(j+1)})$ 
6    $j \leftarrow j + 1$ 
7 until stopping criterion;
Output : The regression matrix  $\mathbf{R}^{(j)}$ .

```

Appendix B. Subspace tracking: updating the set of archetype spectra

Let consider a newly acquired measurement \mathbf{y}_t . Given the LMM in (3), this spectrum can be decomposed as $\mathbf{y}_t = \bar{\mathbf{y}}_t + \mathbf{e}_t$, where $\bar{\mathbf{y}}_t$ is the noise-free spectrum and $(\mathbf{e}_t)_i \stackrel{\text{i.i.d.}}{\sim} \mathcal{N}(0, \sigma_e^2)$ denotes the noise. The question to address is the following: does this new measurement result from the mixture of PS that can be explained by the regression model (9)? This translates

into ensuring that the noise-free component \bar{y}_t belongs to the subspace $\mathbb{P} = \text{span}(\mathbf{Y}_\varphi^\top)$ of dimension $\text{rank}(\mathbf{Y}_\varphi^\top)$ spanned by the archetype spectra in \mathbf{Y}_φ . Inspired by the approach in [58], the proposed solution consists in projecting the measurement \mathbf{y}_t onto \mathbb{P} and evaluating the energy of the residual.

Formally, the orthogonal projector onto \mathbb{P} writes $\mathbf{P} = \mathbf{Y}_\varphi^\top (\mathbf{Y}_\varphi \mathbf{Y}_\varphi^\top)^{-1} \mathbf{Y}_\varphi$. Let $\mathbf{z}_t = \mathbf{y}_t - \mathbf{P}\mathbf{y}_t$ denote the so-called residual, i.e., the projection of \mathbf{y}_t onto the orthogonal complement \mathbb{P}^\perp with dimension $L - \text{rank}(\mathbf{Y}_\varphi^\top)$. If $\bar{\mathbf{y}}_t \in \mathbb{P}$, then $\mathbf{P}\bar{\mathbf{y}}_t = \bar{\mathbf{y}}_t$ and $\mathbf{z}_t = \mathbf{e}_t - \mathbf{P}\mathbf{e}_t$. In other words, as illustrated in Fig. B.10, the residual only consists of the measurement noise projected onto \mathbb{P}^\perp . Given the Gaussian nature of this noise, it follows that the normalized energy of this residual can be statistically described by a χ^2 distribution, i.e.,

$$\frac{1}{\sigma_e^2} \|\mathbf{z}_t\|_2^2 \sim \chi_{L-\text{rank}(\mathbf{Y}_\varphi^\top)}^2. \quad (\text{B.1})$$

This property leads to a binary hypothesis testing in order to check whether $\bar{\mathbf{y}}_t$ lies in \mathbb{P} or not. For a given probability of false alarm p_{FA} , one can decide

$$\begin{cases} \bar{\mathbf{y}}_t \in \mathbb{P} & \text{if } \frac{1}{\sigma_e^2} \|\mathbf{z}_t\|_2^2 \leq F^{-1}(1 - p_{\text{FA}}), \\ \bar{\mathbf{y}}_t \notin \mathbb{P} & \text{otherwise,} \end{cases} \quad (\text{B.2})$$

where $F^{-1}(\cdot)$ denotes the inverse cumulative distribution function of the χ^2 distribution with $L - \text{rank}(\mathbf{Y}_\varphi^\top)$ degrees-of-freedom. Deciding $\bar{\mathbf{y}}_t \notin \mathbb{P}$ means that the current set of archetype spectra is not sufficient to explain the new measurement \mathbf{y}_t and should be complemented with the new measured spectra, i.e., $\mathcal{P} \leftarrow \mathcal{P} \cup \{t\}$ and \mathbf{Y}_φ is updated accordingly. In the numerical experiments reported in Section 6, the probability of false alarm has been set as $p_{\text{FA}} = 0.05$. The overall procedure is sketched in Algo. 4.

Algorithm 4: SubspaceTracking

Input : Newly acquired spectrum \mathbf{y}_t , index t of the new spectrum, archetype spectra \mathbf{Y}_φ , set \mathcal{P} of the indices of archetype spectra, model noise variance σ_e^2

// Project \mathbf{y}_t onto \mathbb{P}^\perp

1 $\mathbf{z}_t = \mathbf{y}_t - \mathbf{Y}_\varphi^\top (\mathbf{Y}_\varphi \mathbf{Y}_\varphi^\top)^{-1} \mathbf{Y}_\varphi \mathbf{y}_t$

// Update the set of indices of archetype spectra

2 **if** $\frac{1}{\sigma_e^2} \|\mathbf{z}_t\|_2^2 \geq F^{-1}(1 - p_{\text{FA}})$ **then**

3 | $\mathcal{P} \leftarrow \mathcal{P} \cup \{t\}$

4 **end if**

Output: Set \mathcal{P} of the indices of archetype spectra.

Appendix C. Estimation of the error covariance matrix in the original space

This section proposes to estimate the posterior covariance matrix $\boldsymbol{\Sigma}_t$ characterizing the PS \mathbf{S}_t in the original space from its counterpart $\tilde{\boldsymbol{\Sigma}}_t$ characterizing the lower-dimensional representation $\tilde{\mathbf{S}}_t$ of the PS and directly provided by the KF. Elaborating

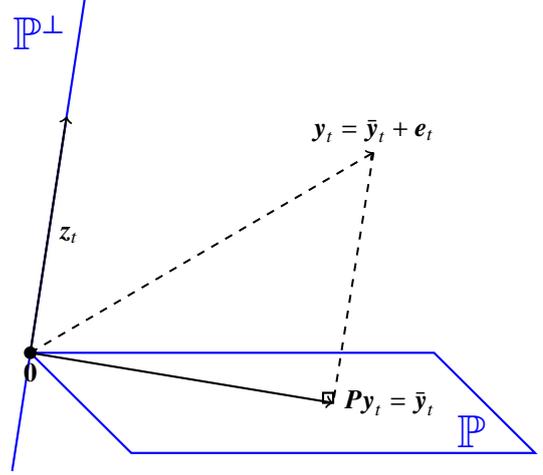


Figure B.10: Geometric illustration of the decomposition of $\mathbf{y}_t = \bar{\mathbf{y}}_t + \mathbf{e}_t$ as the sum of $\mathbf{P}\mathbf{y}_t$ and \mathbf{z}_t , when $\bar{\mathbf{y}}_t \in \mathbb{P}$ ($L = 3$, $\text{rank}(\mathbf{Y}_\varphi^\top) = 2$).

on the regression model (A.2), the matrix of the PS can be written as

$$\mathbf{S}_t = \mathbf{Y}_\varphi^\top \mathbf{R}_t + \mathbf{B}_t. \quad (\text{C.1})$$

where $(\mathbf{B}_t)_{ij} \stackrel{\text{i.i.d.}}{\sim} \mathcal{N}(0, \sigma_e^2)$. The zero mean imposed on noises $(\mathbf{B}_t)_{ij}$ ensures consistency with (A.3). The choice of a variance equal to the observation noise variance, namely σ_e^2 , can be motivated as follows: if a pure spectrum is part of the acquired spectra (or considered as an observable spectrum), it is legitimate to state that it is subject to the measurement noise whose variance is σ_e^2 . Let now assume that the matrix $\tilde{\mathbf{Y}}_\varphi^\top$ of the archetype spectra in the lower-dimensional space is full column rank. Capitalizing on the regression model (A.2), the regression matrix \mathbf{R}_t can be approximated as

$$\mathbf{R}_t \approx \tilde{\mathbf{Y}}_\varphi^{\top+} \tilde{\mathbf{S}}_t. \quad (\text{C.2})$$

where $\tilde{\mathbf{Y}}_\varphi^{\top+} = (\tilde{\mathbf{Y}}_\varphi \tilde{\mathbf{Y}}_\varphi^\top)^{-1} \tilde{\mathbf{Y}}_\varphi$ denotes the (left) pseudo-inverse of \mathbf{Y}_φ^\top . Plugging (C.2) into (C.1) leads to

$$\mathbf{S}_t \approx \mathbf{Y}_\varphi^\top \tilde{\mathbf{Y}}_\varphi^{\top+} \tilde{\mathbf{S}}_t + \mathbf{B}_t. \quad (\text{C.3})$$

For simplicity, instead of characterizing the full covariance matrix $\boldsymbol{\Sigma}_t \in \mathbb{R}^{LK \times LK}$ which describes the intra- and inter-dependence of all pure spectra at all spectral bands, one can restrict the current analysis to characterize the covariance matrices denoted $\boldsymbol{\Sigma}_t[k] = \text{cov}(\mathbf{S}_t[k]) \in \mathbb{R}^{L \times L}$ associated with each pure spectrum $\mathbf{S}_t[k] \in \mathbb{R}^{L \times 1}$, where $\mathbf{S}_t[k]$ refers to the k th column of \mathbf{S}_t ($k \in \{1, \dots, K\}$). From column-wise counterparts of (C.3) with $\mathbf{Y}_\varphi^\top \tilde{\mathbf{Y}}_\varphi^{\top+} = (\tilde{\mathbf{Y}}_\varphi^+ \mathbf{Y}_\varphi)^\top$, and assuming independence of the terms in (C.3) for simplicity, these covariance matrices write

$$\boldsymbol{\Sigma}_t[k] \approx (\tilde{\mathbf{Y}}_\varphi^+ \mathbf{Y}_\varphi)^\top \tilde{\boldsymbol{\Sigma}}_t[k] (\tilde{\mathbf{Y}}_\varphi^+ \mathbf{Y}_\varphi) + \sigma_e^2 \mathbf{I}_L, \quad (\text{C.4})$$

where $\tilde{\boldsymbol{\Sigma}}_t[k] = \text{cov}(\tilde{\mathbf{S}}_t[k])$ can be easily extracted from the covariance matrix $\tilde{\boldsymbol{\Sigma}}_t$ provided as an output by the KF. In particular, the diagonal elements of the matrices $\boldsymbol{\Sigma}_t[k]$ quantify the uncertainties of each estimated pure spectrum along the spectral bands.

Appendix D. Algorithmic sketch of the acquisition protocol P2

The algorithmic sketch of the simulated acquisition protocol P2 is given in Algo. 5. Candidate spectral pixels are first identified by successive convex hull calculations and subsequently stored in the matrix \mathbf{Z} (line 5), following a peeling process (line 7) similar to the one adopted in [55]. These pixels are then partitioned into J classes using K-means and a representative (centroid) of each class is then identified, denoted as \mathbf{m}_j , $j = 1, \dots, J$ (line 9). Finally, essential spectral pixels are selected as the measured spectra which are the closest to the centroids (line 15) and their observation indices are stored in \mathcal{J} (line 16). A permutation $\sigma(\cdot)$ is finally introduced into \mathcal{J} to produce different order sequences (line 18). This introduces a randomization step, by permuting the indices of the previously identified essential spectral pixels. The purpose of this step is to increase the likelihood of selecting spectral pixels that are not too similar in successive measurements. While this randomization does not guarantee dissimilarity every time, it enhances the diversity of the selected pixels compared to a fixed, sequential order. This process contrasts with the acquisition protocol P1, where there is no specific measurement order, and better mimics the acquisition strategies adopted in [25, 26].

For the experiments detailed in Section 5, Protocol P2 sketched in Algo. 5 has been implemented with $N_{\text{ess}} = 340$ for the data sets SD1 and SD2 and $N_{\text{ess}} = 400$ for the real data set RD. The number of clusters used in protocol P2 has been set to $J = 50$ for all data sets.

References

- [1] E. Dinç, Z. C. Ertekin, E. Bükler, Two-way and three-way resolutions of fluorescence excitation-emission dataset for the co-estimation of two pharmaceuticals in a binary mixture, *Chemom. Intell. Lab. Syst.* 239 (2023) 104873.
- [2] S. Wartewig, R. H. Neubert, Pharmaceutical applications of Mid-IR and Raman spectroscopy, *Adv. Drug Deliv. Rev.* 57 (8) (2005) 1144–1170.
- [3] H. Rebiere, M. Martin, C. Ghyselinck, P.-A. Bonnet, C. Brenier, Raman chemical imaging for spectroscopic screening and direct quantification of falsified drugs, *J. Pharm. Biomed. Anal.* 148 (2018) 316–323.
- [4] I. Ten-Doménech, M. Moreno-Torres, J. D. Sanjuan-Herráez, D. Pérez-Guaita, G. Quintás, J. Kuligowski, Analysis of longitudinal metabolomic data using multivariate curve resolution-alternating least squares and pathway analysis, *Chemom. Intell. Lab. Syst.* 232 (2023) 104720.
- [5] C. Bedia, A. Sierra, R. Tauler, Multimodal multisample spectroscopic imaging analysis of tumor tissues using multivariate curve resolution, *Chemom. Intell. Lab. Syst.* 215 (2021) 104366.
- [6] A. Sakudo, Near-infrared spectroscopy for medical applications: Current status and future perspectives, *Clin. Chim. Acta* 455 (2016) 181–188.
- [7] K. Kong, C. Kendall, N. Stone, I. Notingher, Raman spectroscopy for medical diagnostics from in-vitro biofluid assays to in-vivo cancer detection, *Adv. Drug Deliv. Rev.* 89 (2015) 121–134.
- [8] J. M. Bioucas-Dias, A. Plaza, N. Dobigeon, M. Parente, Q. Du, P. Gader, J. Chanussot, Hyperspectral unmixing overview: Geometrical, statistical, and sparse regression-based approaches, *IEEE J. Sel. Topics Appl. Earth Observations Remote Sensing* 5 (2) (2012) 354–379.
- [9] R. Tauler, B. Kowalski, S. Fleming, Multivariate curve resolution applied to spectral data from multiple runs of an industrial process, *Anal. Chem.* 65 (15) (1993) 2040–2047.
- [10] N. Gillis, Nonnegative matrix factorization, SIAM, 2020.
- [11] R. Tauler, Multivariate curve resolution applied to second order data, *Chemom. Intell. Lab. Syst.* 30 (1) (1995) 133–146.

Algorithm 5: Acquisition protocol P2

Input : Data set \mathbf{Y} , number N_{ess} of essential spectra, number J of clusters, number K of pure spectra

// Initialization

- 1 $\mathbf{D} \leftarrow [], \mathbf{Z} \leftarrow [], \tilde{\mathbf{Z}} \leftarrow []$
- 2 $\mathbf{W} \leftarrow \mathbf{Y}$
- // Project the data set onto a K-dimensional subspace*
- 3 $\tilde{\mathbf{W}} \leftarrow \text{PCA}(\mathbf{W}, K)$
- // Progressively extract successive convex hulls*
- 4 **while** $\text{card}(\mathbf{Z}) < N_{\text{ess}}$ **do**
 - // Identify the indices of the pixels on the convex hull*
 - 5 $\mathcal{W} \leftarrow \text{convhull}(\tilde{\mathbf{W}})$
 - // Store the spectra composing the convex hull*
 - 6 $\tilde{\mathbf{Z}} \leftarrow [\tilde{\mathbf{Z}}, \tilde{\mathbf{W}}_{\mathcal{W}}], \mathbf{Z} \leftarrow [\mathbf{Z}, \mathbf{W}_{\mathcal{W}}]$
 - // Peel the data set by removing the convex hull*
 - 7 $\tilde{\mathbf{W}} \leftarrow \tilde{\mathbf{W}} \setminus \tilde{\mathbf{Z}}, \mathbf{W} \leftarrow \mathbf{Y} \setminus \mathbf{Z}$
- 8 **end while**
- // Partition Z into J clusters and identify the centroids*
- 9 $\{\mathbf{m}_1, \dots, \mathbf{m}_J\} \leftarrow \text{K-means}(\mathbf{Z}, J)$
- 10 $n \leftarrow 0$
- 11 $\mathbf{W} \leftarrow \mathbf{Z}$
- 12 **while** $n \leq N_{\text{ess}}$ **do**
 - // Identify J essential spectra (the closest to each centroid)*
 - 13 $\mathcal{J} \leftarrow []$
 - 14 **for** $j = 1, \dots, J$ **do**
 - 15 $k^* \leftarrow \text{argmin}_k \|\mathbf{m}_j - \mathbf{w}_k\|_2^2$
 - 16 $\mathcal{J} \leftarrow \mathcal{J} \cup \{k^*\}$
 - 17 **end for**
 - // Randomly permute the indices in J*
 - 18 $\check{\mathcal{J}} \leftarrow \sigma(\mathcal{J})$
 - // Complement the matrix of essential spectra*
 - 19 $\mathbf{D} \leftarrow [\mathbf{D}, \mathbf{Z}_{\check{\mathcal{J}}}]$
 - // Discard D from Z*
 - 20 $\mathbf{W} \leftarrow \mathbf{Z} \setminus \mathbf{D}$
 - 21 $n \leftarrow n + J$
- 22 **end while**
- 23 $\mathbf{D} \leftarrow \mathbf{D}_{1:N_{\text{ess}}}$

Output: The matrix of essential spectra \mathbf{D} .

- [12] S. Moussaoui, C. Carteret, D. Brie, A. Mohammad-Djafari, Bayesian analysis of spectral mixture data using Markov chain Monte Carlo methods, *Chemom. Intell. Lab. Syst.* 81 (2) (2006) 137–148.
- [13] N. Dobigeon, S. Moussaoui, J.-Y. Tourneret, C. Carteret, Bayesian separation of spectral sources under non-negativity and full additivity constraints, *Signal Process.* 89 (12) (2009) 2657–2669.
- [14] M. Arngren, M. N. Schmidt, J. Larsen, Unmixing of hyperspectral images using Bayesian non-negative matrix factorization with volume prior, *J. Signal Process. Syst.* 65 (2011) 479–496.
- [15] O. Eches, N. Dobigeon, J.-Y. Tourneret, H. Snoussi, Variational methods for spectral unmixing of hyperspectral unmixing, in: *Proc. IEEE Int. Conf. Acoust., Speech, and Signal Processing (ICASSP)*, Prague, Czech Republic, 2011, pp. 957–960.
- [16] O. Hoidn, A. A. Mishra, A. Mehta, Probabilistic mixture model-based spectral unmixing, *Applied Sciences* 14 (11) (2024) 4836.
- [17] J. M. Nascimento, J. M. Bioucas-Dias, Vertex component analysis: A fast algorithm to unmix hyperspectral data, *IEEE Trans. Geosci. Remote Sens.* 43 (4) (2005) 898–910.
- [18] M. C. U. Araújo, T. C. B. Saldanha, R. K. H. Galvao, T. Yoneyama, H. C. Chame, V. Visani, The successive projections algorithm for variable selection in spectroscopic multicomponent analysis, *Chemom. Intell. Lab. Syst.* 57 (2) (2001) 65–73.
- [19] J. W. Boardman, F. A. Kruse, R. O. Green, Mapping target signatures via partial unmixing of AVIRIS data, in: *Summaries of the fifth annual JPL airborne earth science workshop. Volume 1: AVIRIS workshop*, 1995.
- [20] M. E. Winter, N-FINDR: An algorithm for fast autonomous spectral end-member determination in hyperspectral data, in: *Imaging spectrometry V*, Vol. 3753, SPIE, 1999, pp. 266–275.
- [21] T.-H. Chan, C.-Y. Chi, Y.-M. Huang, W.-K. Ma, A convex analysis-based minimum-volume enclosing simplex algorithm for hyperspectral unmixing, *IEEE Trans. Signal Process.* 57 (11) (2009) 4418–4432.
- [22] J. Li, A. Agathos, D. Zaharie, J. M. Bioucas-Dias, A. Plaza, X. Li, Minimum volume simplex analysis: A fast algorithm for linear hyperspectral unmixing, *IEEE Trans. Geosci. Remote Sens.* 53 (9) (2015) 5067–5082.
- [23] J. M. Bioucas-Dias, A variable splitting augmented Lagrangian approach to linear spectral unmixing, in: *Proc. IEEE GRSS Workshop Hyperspectral Image Signal Process.: Evolution in Remote Sens. (WHISPERS)*, IEEE, 2009, pp. 1–4.
- [24] L. Miao, H. Qi, Endmember extraction from highly mixed data using minimum volume constrained nonnegative matrix factorization, *IEEE Trans. Geosci. Remote Sens.* 45 (3) (2007) 765–777.
- [25] L. Coic, R. Vitale, M. Moreau, D. Rousseau, J. H. de Morais Goulart, N. Dobigeon, C. Ruckebusch, Assessment of essential information in the Fourier domain to accelerate Raman hyperspectral microimaging, *Anal. Chem.* 95 (42) (2023) 15497–15504.
- [26] V. Gilet, G. Mabilieu, M. Loumaigne, L. Coic, R. Vitale, T. Oberlin, J. H. de Morais Goulart, N. Dobigeon, C. Ruckebusch, D. Rousseau, Superpixels meet essential spectra for fast Raman hyperspectral microimaging, *Optics Express* 32 (1) (2024) 932–948.
- [27] M. Ghaffari, N. Omidikia, C. Ruckebusch, Essential spectral pixels for multivariate curve resolution of chemical images, *Anal. Chem.* 91 (17) (2019) 10943–10948.
- [28] C. Ruckebusch, R. Vitale, M. Ghaffari, S. Hugelier, N. Omidikia, Perspective on essential information in multivariate curve resolution, *TrAC Trends in Analytical Chemistry* 132 (2020) 116044.
- [29] C. Thureau, K. Kersting, C. Bauckhage, Convex non-negative matrix factorization in the wild, in: *Proc. IEEE Int. Conf. Data Mining (ICDM)*, IEEE, 2009, pp. 523–532.
- [30] L. Scipioni, A. Rossetta, G. Tedeschi, E. Gratton, Phasor S-FLIM: a new paradigm for fast and robust spectral fluorescence lifetime imaging, *Nature methods* 18 (5) (2021) 542–550.
- [31] H. J. Chiang, D. E. Koo, M. Kitano, S. Burkitt, J. R. Unruh, C. Zavaleta, L. A. Trinh, S. E. Fraser, F. Cutrale, HyU: Hybrid unmixing for longitudinal in vivo imaging of low signal-to-noise fluorescence, *Nature Methods* 20 (2) (2023) 248–258.
- [32] R. E. Kalman, A new approach to linear filtering and prediction problems, *Trans. ASME Journal of Basic Engineering* 82 (1960) 35–45.
- [33] A. Cutler, L. Breiman, Archetypal analysis, *Technometrics* 36 (4) (1994) 338–347.
- [34] F. Shayegh, A. Erfanian, Real-time ocular artifacts suppression from EEG signals using an unsupervised adaptive blind source separation, in: *Proc. IEEE Eng. Med. Bio. Soc. (EMBC)*, IEEE, 2006, pp. 5269–5272.
- [35] A. Jafarifarmand, M.-A. Badamchizadeh, S. Khanmohammadi, M. A. Nazari, B. M. Tazehkand, Real-time ocular artifacts removal of EEG data using a hybrid ICA-ANC approach, *Biomed. Signal Process. Control* 31 (2017) 199–210.
- [36] A. Cichocki, R. Unbehauen, Robust neural networks with on-line learning for blind identification and blind separation of sources, *IEEE Trans. Circuits Syst. I: Fundam. Theory Appl.* 43 (11) (1996) 894–906.
- [37] Y. Sepulcre, T. Trigano, Y. Ritov, Sparse regression algorithm for activity estimation in γ spectrometry, *IEEE Trans. Signal Process.* 61 (17) (2013) 4347–4359.
- [38] J. M. Kirkpatrick, B. M. Young, Poisson statistical methods for the analysis of low-count gamma spectra, *IEEE Trans. Nucl. Sci.* 56 (3) (2009) 1278–1282.
- [39] P. Malfrait, J. Bobin, A. de Vismes Ott, Online spectral unmixing in gamma-ray spectrometry, *Appl. Radiat. Isot.* 201 (2023) 111011.
- [40] R. Zhao, V. Y. Tan, Online nonnegative matrix factorization with outliers, *IEEE Trans. Signal Process.* 65 (3) (2016) 555–570.
- [41] A. Lefevre, F. Bach, C. Févotte, Online algorithms for nonnegative matrix factorization with the Itakura-Saito divergence, in: *Proc. IEEE Workshop Applications Signal Process. Audio Acoustics (WASPAA)*, IEEE, 2011, pp. 313–316.
- [42] J. Mairal, F. Bach, J. Ponce, G. Sapiro, Online learning for matrix factorization and sparse coding, *J. Mach. Learn. Res.* 11 (1) (2010).
- [43] N. Guan, D. Tao, Z. Luo, B. Yuan, Online nonnegative matrix factorization with robust stochastic approximation, *IEEE Trans. Neural Netw. Learn. Syst.* 23 (7) (2012) 1087–1099.
- [44] C. Févotte, N. Dobigeon, Nonlinear hyperspectral unmixing with robust nonnegative matrix factorization, *IEEE Trans. Image Processing* 24 (12) (2015) 4810–4819.
- [45] C. Févotte, N. Bertin, J.-L. Durrieu, Nonnegative matrix factorization with the Itakura-Saito divergence: With application to music analysis, *Neural computation* 21 (3) (2009) 793–830.
- [46] S. Henrot, J. Chanussot, C. Jutten, Dynamical spectral unmixing of multi-temporal hyperspectral images, *IEEE Trans. Image Process.* 25 (7) (2016) 3219–3232.
- [47] P.-A. Thouvenin, N. Dobigeon, J.-Y. Tourneret, A hierarchical Bayesian model accounting for endmember variability and abrupt spectral changes to unmix multitemporal hyperspectral images, *IEEE Trans. Computational Imaging* 4 (1) (2018) 32–45.
- [48] P.-A. Thouvenin, N. Dobigeon, J.-Y. Tourneret, Partially asynchronous distributed unmixing of hyperspectral images, *IEEE Trans. Geoscience and Remote Sensing* 57 (4) (2019) 2009–2021.
- [49] L. Nus, S. Miron, D. Brie, An ADMM-based algorithm with minimum dispersion regularization for on-line blind unmixing of hyperspectral images, *Chemom. Intell. Lab. Syst.* 204 (2020) 104090.
- [50] N. Omidikia, S. Beyramysoltan, J. Mohammad Jafari, E. Tavakkoli, M. Akbari Lakeh, M. Alinaghi, M. Ghaffari, S. Khodadadi Karimvand, R. Rajkó, H. Abdollahi, Closure constraint in multivariate curve resolution, *Journal of Chemometrics* 32 (12) (2018) e2975.
- [51] J. M. Bioucas-Dias, M. A. Figueiredo, Alternating direction algorithms for constrained sparse regression: Application to hyperspectral unmixing, in: *Proc. IEEE GRSS Workshop Hyperspectral Image Signal Process.: Evolution in Remote Sens. (WHISPERS)*, IEEE, 2010, pp. 1–4.
- [52] S. S. Haykin, *Adaptive filter theory*, Pearson Education India, 2002.
- [53] I. Chitra Ragupathy, V. Schweikhard, A. Zumbusch, Multivariate analysis of hyperspectral stimulated raman scattering microscopy images, *Journal of Raman Spectroscopy* 52 (9) (2021) 1630–1642.
- [54] S. Hugelier, O. Devos, C. Ruckebusch, Constraining shape smoothness in multivariate curve resolution—alternating least squares, *Journal of Chemometrics* 29 (8) (2015) 448–456.
- [55] M. Ahmad, R. Vitale, M. Cocchi, C. Ruckebusch, Weighted multivariate curve resolution/alternating least squares based on sample relevance, *Journal of Chemometrics* 37 (6) (2023) e3478.
- [56] A. Savitzky, M. J. Golay, Smoothing and differentiation of data by simplified least squares procedures, *Anal. Chem.* 36 (8) (1964) 1627–1639.
- [57] S. Boyd, N. Parikh, E. Chu, B. Peleato, J. Eckstein, et al., Distributed optimization and statistical learning via the alternating direction method of multipliers, *Foundations and Trends® in Machine Learning* 3 (1) (2011) 1–122.
- [58] Y. Altmann, N. Dobigeon, J.-Y. Tourneret, J. C. M. Bermudez, A robust

test for nonlinear mixture detection in hyperspectral images, in: Proc. IEEE Int. Conf. Acoust., Speech, and Signal Processing (ICASSP), Vancouver, Canada, 2013, pp. 2149–2153.