Compressed Raman method combining classification and estimation of spectra with optimized binary filters

Timothée Justel, Frédéric Galland, Antoine Roueff

Compressed Raman method combining classification and estimation of spectra with optimized binary filters

TIMOTHÉE JUSTEL[1], FRÉDÉRIC GALLAND[1,*], AND ANTOINE ROUEFF[1]

[1] Aix Marseille Univ, CNRS, Centrale Marseille, Institut Fresnel, Marseille, France
[*] Corresponding author: frederic.galland@fresnel.fr

Compiled January 14, 2022

Compressed Raman methods allow classification between known chemical species with only few measurements performed through binary filters. We propose a methodology for binary filter optimization, in which filters are modified at each pixel so that classification can still be performed pixel by pixel with few measurements acquired in parallel, while being able to reconstruct a full spectrum when combining measurements from several pixels. This approach is robust to intensity variations between pixels. It relies on a generalized Bhattacharyya bound and on the Cramér-Rao bound to tailor filters with optimized performance.

http://dx.doi.org/10.1364/ao.XX.XXXXXX

Raman spectroscopy is commonly used to classify chemical species. However, there is no need to acquire the entire spectrum to discriminate between species. Supervised methods can leverage a priori information about the spectra to select only some combinations of spectral components that are relevant for classification. In order to obtain high speed classification systems, compressed Raman approaches have been proposed [1–11]. With these approaches only a few measurements on photon-noise-limited sensors are performed after filtering of the Raman radiation with binary filters designed to achieve high classification performance between the different known species. In particular the approach proposed in [8] enables to carry out supervised classification with measurements performed in parallel. Moreover, the number of measurements can be smaller than the number of species to discriminate. Nevertheless, since all these supervised compressed classification approaches have been designed to discriminate between the reference species, they do not question the validity of the model. Thus, if some calibration issues occur or if an unexpected species is present in the sample, these approaches may lead to erroneous classification results and will not detect the problem.

In this letter, we propose a new compressed Raman strategy. With this strategy, rather than always using the same binary filters optimized for classification, the binary filters are modified at each pixel so that the combination of measurements made on a sufficient number of pixels allows the estimation of a full Raman spectrum. This new strategy opens the possibility to question the validity of the a priori or to re-calibrate the system during acquisition. We design a new binary filter selection strategy that ensure the reliability of the classification performed on each pixel and maximize the accuracy of spectrum estimation. Our methodology relies on a generalized Bhattacharyya bound (GBB) to control the probability of classification error [7, 8] and on the Cramér-Rao bound (CRB) to quantify the accuracy of spectrum estimation.

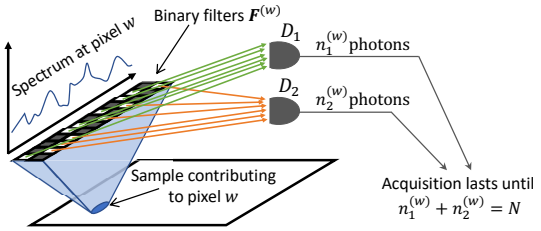

Fig. 1. Schematic diagram of the acquisition system for $P = 2$.

It is considered without loss of generality that acquisitions are performed at different pixels (see Fig. 1). An alternative would for example be to perform successively several acquisitions on the same pixel. On each pixel the Raman scattering radiation is filtered by a set of P orthogonal binary filters (OBF), i.e. binary filters without frequency overlap. This enables to perform the P measurements in parallel [5, 12]. The set of P filters used at pixel w is denoted by $\boldsymbol{F}^{(w)} = (\boldsymbol{F}_1^{(w)}, \ldots, \boldsymbol{F}_P^{(w)})$ where $\boldsymbol{F}_m^{(w)} = (F_m^{(w)}(1), \ldots, F_m^{(w)}(K))^T$ is the m^{th} filter, K the number of frequency bins and $.^T$ the transpose operation. Since filters are OBF, for each frequency bin $k \in [\![1, K]\!]$, $F_m^{(w)}(k) = 0$ or 1 and $F_m^{(w)}(k)\, F_p^{(w)}(k) = 0\ \forall m \neq p$. Each filter is associated with a photon-counting detector D_m and the number of photons received on D_m through filter $\boldsymbol{F}_m^{(w)}$ is denoted by $n_m^{(w)}$. Following [8], rather than fixing the measuring time, acquisition is stopped on each pixel when the total number of detected photons reaches a fixed value N (i.e. $\sum_{m=1}^{P} n_m^{(w)} = N$). Let $\boldsymbol{S} = (S_1, \ldots, S_K)^T$ denotes the normalized Raman spectrum of the measured radiation (i.e. $\sum_{k=1}^{K} S_k = 1$). S_k is the average proportion of photons in the frequency range $[\nu_k - \delta_\nu/2, \nu_k + \delta_\nu/2]$ and it is assumed that $\nu_{k+1} - \nu_k > \delta_\nu$ (i.e. non-overlapping frequency bins). Let $\gamma^{(w)}$ be the intensity at pixel w of the Raman radiation before filtering and $\tau^{(w)}$ the corresponding acquisition time. The mean number

of photons detected by D_m is then

$$\langle n_m^{(w)} \rangle = \tau^{(w)} \gamma^{(w)} \sum_{k=1}^{K} F_m^{(w)}(k)\, S_k = \tau^{(w)} \gamma^{(w)} \boldsymbol{F}_m^{(w)} \cdot \boldsymbol{S} \quad (1)$$

where $\cdot$ denotes the dot product and $\langle\ \rangle$ the statistical mean. Since the P measurements are performed in parallel, if a photon emitted by a species with a spectrum $\boldsymbol{S}$ is detected on the pixel w, the probability that it is detected by D_m is $p_m^{(w)} = \langle n_m^{(w)} \rangle / \sum_{p=1}^{P} \langle n_p^{(w)} \rangle = \boldsymbol{F}_m^{(w)} \cdot \boldsymbol{S} / \sum_{p=1}^{P} (\boldsymbol{F}_p^{(w)} \cdot \boldsymbol{S})$. Furthermore, we constrain the OBF sets to be lossless (LOBF), i.e. $\sum_{m=1}^{P} F_m^{(w)}(k) = 1\ \forall k \in [\![1, K]\!]$, so that every photon reaching the filters is detected [12]. This guarantees that the average acquisition time required to detect the N expected photons is finite no matter the species (see Appendix A). Moreover, with LOBF, $p_m^{(w)}$ simplifies into

$$p_m^{(w)} = \boldsymbol{F}_m^{(w)} \cdot \boldsymbol{S} = \sum_{k=1}^{K} F_m^{(w)}(k)\, S_k. \quad (2)$$

Since N is fixed, the vector $\boldsymbol{n}^{(w)} = (n_1^{(w)}, \ldots, n_P^{(w)})^T$ that corresponds to the number of detected photons follows a multinomial distribution depending only on the spectrum $\boldsymbol{S}$ and the LOBF set $\boldsymbol{F}^{(w)}$ through the parameters $p_m^{(w)}$ and on N, i.e.

$$\mathcal{P}(\boldsymbol{n}^{(w)} \mid \boldsymbol{S}, w) = N! \prod_{m=1}^{P} \frac{\left(p_m^{(w)}\right)^{n_m^{(w)}}}{n_m^{(w)}!}. \quad (3)$$

This distribution is independent of $\tau^{(w)}$ and $\gamma^{(w)}$. Therefore, classification and estimation performance are also independent of these quantities, and in particular of intensity variations between pixels.

Let us assume that one aims to discriminate between M species on each pixel w. Following [8], knowing the reference spectrum $\boldsymbol{S}^{(u)}$ associated with each of the M species (with $u \in [\![1, M]\!]$), it is possible to perform classification from $\boldsymbol{n}^{(w)}$ on each pixel. Assuming that all species are equally likely to appear (i.e. $\mathcal{P}(u) = 1/M\ \forall u$), the classifier minimizing the probability of classification error is [8]

$$\hat{u}_{\text{opt}}^{(w)} = \underset{u}{\text{argmax}} \left[\sum_{m=1}^{P} n_m^{(w)} \log p_m^{(w,u)} \right], \quad (4)$$

where $p_m^{(w,u)}$ stands for the probability $p_m^{(w)}$ in Eq. (2) for the spectrum $\boldsymbol{S}^{(u)}$. As shown in [8] the probability of classification error $\mathcal{P}(\epsilon^{(w)})$ is upper-bounded by a generalized Bhattacharyya bound (GBB) $\mathcal{B}^{(w)}$ easier to interpret and to compute, i.e.

$$\mathcal{P}(\epsilon^{(w)}) \leq \mathcal{B}^{(w)} = \frac{1}{M} \sum_{u} \sum_{v>u} \left(\sum_{m=1}^{P} \sqrt{p_m^{(w,u)} p_m^{(w,v)}} \right)^N. \quad (5)$$

It is then possible to tailor $\boldsymbol{F}^{(w)}$ to minimize $\mathcal{B}^{(w)}$ hence an upper bound on the probability of error [7, 8].

Contrary to classification between M species that can be performed at each pixel using only P measurements (even with $P < M$), the estimation of an unknown spectrum composed of $K > P$ frequency bins cannot be performed using measurements acquired with a single OBF set (without using an a priori information about the spectrum). Nevertheless, such an estimation can be carried out by combining measurements made with different OBF sets, as done for example using couples of Hadamard filters [1]. Let $\boldsymbol{n}^{(1)}, \ldots, \boldsymbol{n}^{(W)}$ be independent measurements made on W pixels of a species with spectrum $\boldsymbol{S}$ obtained using the LOBF sets $\boldsymbol{F}^{(1)}, \ldots, \boldsymbol{F}^{(W)}$. The spectrum $\boldsymbol{S}$ can be estimated from these measurements because their distributions only depend on $\boldsymbol{S}$, on the LOBF sets and on N (see Eq. (2) and Eq. (3)). Moreover, since their distributions are independent from the radiation intensity $\gamma^{(w)}$, this strategy is robust to the unknown variations of intensity from one pixel to another. A standard approach to perform an accurate estimation consists in looking for minimum variance estimators among unbiased estimators. The Cramér-Rao bound (CRB) gives a lower bound over the variance of any unbiased estimator [13]. Let σ_k^2 be the variance of an unbiased estimator of S_k. Note that since $\boldsymbol{S}$ is normalized there are only $K-1$ unknown parameters to estimate. Let now $\mathbf{Y} = \boldsymbol{I}_F^{-1}$ be the CRB matrix where $\boldsymbol{I}_F$ is the Fisher information matrix (FIM) defined as $[\boldsymbol{I}_F]_{k,l} = -\langle \partial_{S_k} \partial_{S_l} \ell \rangle$ with $l, k \in [\![1, K-1]\!]$ and $\ell = \sum_{w=1}^{W} \log \mathcal{P}(\boldsymbol{n}^{(w)} | \boldsymbol{S}, w)$ the log likelihood of the W measurements. Then,

$$\sigma_k^2 \geq [\mathbf{Y}]_{k,k} = [\boldsymbol{I}_F^{-1}]_{k,k} = \text{CRB}_k \quad \forall k \in [\![1, K-1]\!]. \quad (6)$$

Using Eq. (2) and Eq. (3), as shown in Appendix B,

$$\boldsymbol{I}_F = N\, \boldsymbol{G}\, \boldsymbol{Q}\, \boldsymbol{G}^T = N\, (\boldsymbol{G}\sqrt{\boldsymbol{Q}})(\boldsymbol{G}\sqrt{\boldsymbol{Q}})^T, \quad (7)$$

where the matrix $\boldsymbol{G} \in \{-1, 0, 1\}^{(K-1) \times WP}$ is defined by

$$\boldsymbol{G} = (\boldsymbol{\delta F}^{(1)}, \boldsymbol{\delta F}^{(2)}, \ldots, \boldsymbol{\delta F}^{(W)}), \quad (8)$$

with $[\boldsymbol{\delta F}^{(w)}]_{k,m} = F_m^{(w)}(k) - F_m^{(w)}(K)$ for $k \in [\![1, K-1]\!]$ and $m \in [\![1, P]\!]$, and $\boldsymbol{Q} \in {\mathbb{R}_+}^{WP \times WP}$ is diagonal with coefficients

$$[\boldsymbol{Q}]_{i, i=P(w-1)+m} = 1/p_m^{(w)}. \quad (9)$$

The CRB being a function of the LOBF sets, the trace of the CRB matrix $\text{tr}\, \mathbf{Y} = \sum_{k=1}^{K-1} \text{CRB}_k$ can be used to select the sets. Indeed, minimizing $\text{tr}\, \mathbf{Y}$ is equivalent to minimize the mean squared error (MSE) of any unbiased and efficient estimator of $\boldsymbol{S}$. To compute the CRB and therefore to quantify the estimation precision independently of the choice of the estimator, it is necessary to select LOBF sets such that $\boldsymbol{I}_F$ is invertible. Eq. (7) shows that $\boldsymbol{I}_F$ is the Gram matrix of the rows of $\boldsymbol{G}\sqrt{\boldsymbol{Q}}$. Thus, $\boldsymbol{I}_F$ is invertible if and only if the rank of $\boldsymbol{G}$ is equal to $K-1$. Since LOBF sets are used, $\sum_{m=1}^{P} [\boldsymbol{\delta F}^{(w)}]_{k,m} = 0\ \forall w, k$ and the maximum number of linearly independent columns in $\boldsymbol{G}$ is $W(P-1)$. Consequently, the minimum number of LOBF sets necessary for $\boldsymbol{I}_F$ to be invertible is $W_{\text{inf}} = \left\lceil \frac{K-1}{P-1} \right\rceil$, where $\lceil x \rceil$ denotes the smallest integer greater or equal to x.

One possibility to minimize the probability of classification error is to select the LOBF set that minimizes the GBB and apply it on each pixel. However, this does not allow to estimate the spectrum. We need to accept a loss in classification performance at each pixel in order to be able to make an accurate spectrum estimation from several pixels. We quantify classification performance with the GBB and spectrum estimation accuracy with the trace of the CRB matrix. Based on these two criteria, several approaches can be envisaged to optimize the LOBF sets. In the procedure proposed in this paper, we first optimize V randomly initialized LOBF sets to minimize their GBB, which results in a pool of LOBF sets able to perform an accurate classification. In this objective we use the same algorithm as in [7, 8], to the exception that it does not perform a fixed number of iterations

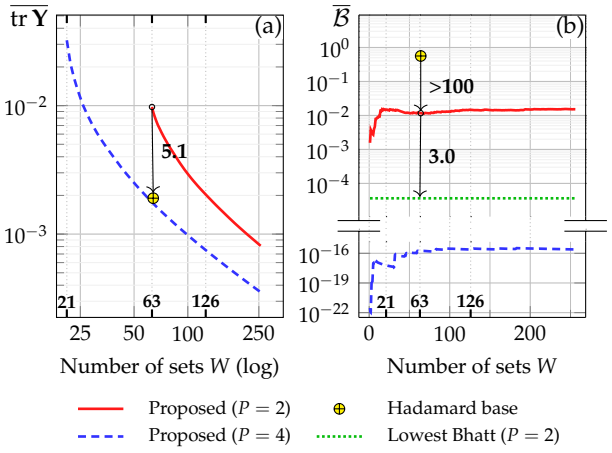

Fig. 2. $\overline{\mathrm{tr}\,\mathbf{Y}}$ (left) and $\overline{\mathcal{B}}$ (right) function of W. Computation done for $P = 2$ and $P = 4$ filters generated with $M = 4$ spectra having $K = 64$ frequency bins and for $N = 500$ photons per pixel.

but stops when a local minimum of $\mathcal{B}$ is reached. The objective is then to select from this pool the W LOBF sets minimizing the trace of the CRB matrix. To this end, the algorithm selects the LOBF sets one by one, determining at each step the set leading to the highest decrease of $\mathrm{tr}\,\mathbf{Y}$.

Since the CRB depends on the unknown spectrum to estimate we choose to select our sets with the CRB of the flat spectrum (i.e. $S^{\mathrm{flat}}(k) = 1/K\ \forall k$). Other solutions could be envisaged such as using the average of $\mathrm{tr}\,\mathbf{Y}$ over the reference spectra or using other a priori. Moreover, during the first steps of the selection, $\boldsymbol{I}_F$ is not invertible. Thus, we have empirically chosen that the algorithm selects the LOBF set leading to the highest increase in the rank of $\boldsymbol{I}_F$ and, in case of ambiguity, the set minimizing the trace of the pseudo-inverse of $\boldsymbol{I}_F$. Furthermore, the GBB optimization algorithm used to generate the initial pool tends to converge to LOBF sets that are similar. We need a sufficient diversity of sets so that $\boldsymbol{I}_F$ is invertible. One solution is to roll back the last L iterations of the GBB optimization and add the corresponding sets to the pool. Another solution is to set $V \gg W_{\mathrm{inf}}$. Throughout all the following $L = 5$ and $V = 2000$. Typically, the computation time of the whole procedure is two minutes for M=4, P=2, K=64 and W=126 with a standard computer, but this optimization can be performed off-line, i.e. before starting the acquisitions.

We illustrate the interest of the proposed approach on an example with $M = 4$ species with spectra composed of $K = 64$ frequency bins, for $N = 500$ photons detected per pixel. These spectra are randomly generated as $S_k^{(u)} = X_k^{(u)} / \sum_{l=1}^{K} X_l^{(u)}$, where $X_k^{(u)}$ are exponential random variables. This process draws uniformly distributed samples from the simplex of normalized spectra. The evolution of the performance of classification and spectrum estimation are plotted as a function of W in Fig.2. More precisely, rather than plotting $\mathrm{tr}\,\mathbf{Y}$ for a given spectrum we plot on Fig.2.a the empirical mean of $\mathrm{tr}\,\mathbf{Y}$ over 10^3 uniformly drawn normalized spectra and denote it $\overline{\mathrm{tr}\,\mathbf{Y}}$. We also plot on Fig.2.b the mean Bhattacharyya bound over the W LOBF sets, i.e. $\overline{\mathcal{B}} = \sum_{w=1}^{W} \mathcal{B}^{(w)}/W$.

We first analyse the performance of the proposed approach with $P = 2$ filters (see Fig.2, solid red lines). Fig.2.a shows that the FIM is invertible as soon as W reaches $W_{\mathrm{inf}} = 63$ sets. As suggested by Eq. (7), $\overline{\mathrm{tr}\,\mathbf{Y}}$ is a decreasing function of W. At first, $\overline{\mathrm{tr}\,\mathbf{Y}}$ decreases faster than $1/W$ and then slows down until it reaches an asymptote in $1/W$. For $W = 63$ $\overline{\mathrm{tr}\,\mathbf{Y}}$ is equal to 9.8×10^{-3}, whereas for $W = 126$ it is equal to 2.0×10^{-3}. On the other hand, Fig.2.b shows that $\overline{\mathcal{B}}$ stays between 1.5×10^{-2} and 1.6×10^{-3}.

As a comparison we plot the classification performance of the LOBF set of the optimized pool with the lowest GBB (see Fig.2.b, dotted green line), which corresponds to the approach proposed in [8]. As expected $\overline{\mathcal{B}}$ is smaller than with the proposed approach but it is not possible to calculate $\overline{\mathrm{tr}\,\mathbf{Y}}$ because the FIM is not invertible. Note that classification and estimation performance can in general be improved by increasing the total number of photons detected per pixel, N. Indeed, according to Eq. (7), $\overline{\mathrm{tr}\,\mathbf{Y}}$ is proportional to $1/N$ while $\overline{\mathcal{B}}$ varies asymptotically as a power of N. For $W = 63$ the sets generated using our approach would need 3 times more photons to perform as well in classification as the set with the lowest GBB (see Fig.2.b, bottom black arrow).

Another category of LOBF sets with $P = 2$ that is widely used for spectrum estimation is derived from the Hadamard S-Matrix [1, 2, 14]. The performance of such sets with $W = 63$ are shown in Fig.2 (yellow discs). The value of $\overline{\mathrm{tr}\,\mathbf{Y}}$ obtained with these Hadamard sets is lower than with the proposed approach, which confirms the interest of using them for estimation. In fact, it would require 5.1 times more photons to our sets to have the same estimation performance (see Fig.2.a, black arrow). On the other hand, the GBB corresponding to these sets is greater and they need over 100 times more photons to reach the same $\overline{\mathcal{B}}$ as our sets for $P = 2$ (see Fig.2.b, top black arrow). The proposed approach therefore allows to find a trade-off between classification and spectrum estimation performance.

To improve performance, rather than increasing N, one can increase the number of filters per set, P. Going from $P = 2$ to $P = 4$ filters without changing N (see Fig.2, blue dashed lines) improves both classification and estimation performance. For $W = 63$ and $P = 4$ the proposed sets reach the performance of the Hadamard sets in terms of $\overline{\mathrm{tr}\,\mathbf{Y}}$ while the mean GBB is under 5×10^{-16}, which guarantees a very small probability of classification error. Furthermore, the FIM is invertible as soon as $W = 21$ when $P = 4$ instead of $W = 63$ when $P = 2$.

Once N and the LOBF sets have been chosen, one can use them for classification and spectrum estimation. Fig.3 displays the classification and estimation performance of the LOBF sets built with our method for $P = 2$ and $W = 126$ for the experiment shown in Fig.2. The probability of classification error with the classifier of Eq. (4) has been evaluated using a Monte-Carlo simulation with 10^5 samples per pixel for each of the M spectra. The variance of the pseudo inverse (PI) estimator when estimating the spectrum shown in Fig.2 (top) has been evaluated using a Monte-Carlo simulation with 10^4 samples per pixel. All samples were generated following the multinomial distribution of Eq. (3). The empirical probabilities of error per pixel and empirical variances per frequency bin resulting of the simulations are denoted respectively by $\widehat{\mathcal{P}}(\epsilon^{(w)})$ and $\widehat{\sigma}_k^2$. Eq. (2) shows that the PI estimator of the spectrum $\boldsymbol{S}$ is $\widehat{\boldsymbol{S}} = (\boldsymbol{H}\boldsymbol{H}^T)^{-1}\boldsymbol{H}\,\boldsymbol{v}$, with $\boldsymbol{v}$ the concatenation of the W vectors $\frac{1}{N}\boldsymbol{n}^{(w)}$ and $\boldsymbol{H} = (\boldsymbol{F}^{(1)}, \ldots, \boldsymbol{F}^{(W)})$ the concatenation of the W sets of filters. The PI estimator is unbiased. Indeed $\langle\widehat{\boldsymbol{S}}\rangle = (\boldsymbol{H}\boldsymbol{H}^T)^{-1}\boldsymbol{H}\langle\boldsymbol{v}\rangle$ and since $\langle\boldsymbol{v}\rangle = \boldsymbol{H}^T\boldsymbol{S}$, $\langle\widehat{\boldsymbol{S}}\rangle = \boldsymbol{S}$. Moreover in this simulation, its variance is close to the CRB. Note that the PI estimator does not guarantee the positivity nor the sum-to-one. The proposed OBF selection algorithm leads to $\widehat{\mathcal{P}}(\epsilon^{(w)})$ between 3.3×10^{-5} and 2.2×10^{-2} and $\widehat{\sigma}_k^2$ between

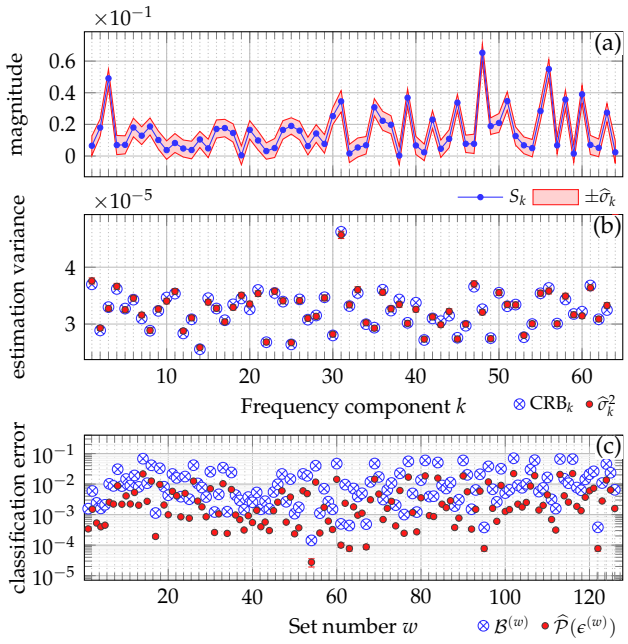

Fig. 3. Monte-Carlo simulation. Top : Spectrum $\boldsymbol{S}$ to be estimated (in blue) along with the interval of confidence of the PI estimator at $\pm\widehat{\sigma}_k$ (in red). Middle : CRB_k (in blue) along with the corresponding empirical variances $\widehat{\sigma}_k^2$ (in red). Bottom : Bhattacharyya bounds $\mathcal{B}^{(w)}$ (in blue) along with the corresponding empirical probabilities of error $\widehat{\mathcal{P}}(\epsilon^{(w)})$ (in red). Error bars are $\pm$ the corresponding empirical standard deviation in all graphs. Simulation comprises $W = 126$ optimized sets of $P = 2$ filters and $N = 500$ photons per pixel.

2.6×10^{-5} and 5.1×10^{-5}.

In conclusion, we propose a methodology that enables to perform classification between known species on each pixel with a number of filters smaller than the number of species to discriminate, while being able to estimate a full spectrum combining a sufficient number of measurements made with different sets of filters. The LOBF sets built with our method achieve a trade-off between the GBB and the CRB, enabling to perform both reliable classification and accurate estimation. It allows one to determine a priori the minimum number N of photons per pixel required to achieve the desired performance. However, note that if the measured number of photons is different from the one initially expected, both classification and estimation can still be performed. One obvious perspective to this work is to implement this approach on a real optical setup. This would enable to update on-the-fly the reference spectra used for classification or to perform anomaly detection. Of course, since spectrum estimation requires to combine several pixels, a straightforward estimation in a sliding window will generate errors when the window includes pixels of different species. To solve this issue, segmentation strategies or robust estimation methods have to be developed. A careful study on the way to distribute the W OBF sets on the imaging plane also has to be carried out.

A. LOSSLESS ORTHOGONAL BINARY FILTERS (LOBF)

Let $\tau_N^{(w)}$ be the acquisition time that has been necessary to detect N photons at pixel w. Since $\sum_{m=1}^{P} n_m^{(w)} = N$, Eq. (1), leads to $\langle \tau_N^{(w)} \rangle = \frac{N}{\gamma^{(w)} (\sum_{m=1}^{P} \boldsymbol{F}_m^{(w)}) \cdot \boldsymbol{S}}$. For OBF with loss, $\sum_{m=1}^{P} F_m^{(w)}(k) = 0$ for at least one bin k. Thus there exist spectra for which $(\sum_{m=1}^{P} \boldsymbol{F}_m^{(w)}) \cdot \boldsymbol{S} = 0$, and then $\langle \tau_N^{(w)} \rangle \to \infty$. With lossless OBF, $\sum_{m=1}^{P} F_m^{(w)}(k) = 1 \ \forall k$ and since $\sum_{k=1}^{K} S_k = 1$, $\langle \tau_N^{(w)} \rangle = N/\gamma^{(w)}$, which is finite for any $\boldsymbol{S}$ provided $\gamma^{(w)} > 0$. Moreover, it allows a direct estimation of $\gamma^{(w)}$ by measuring the acquisition time.

B. FISHER INFORMATION MATRIX FOR LOBF SETS

Eq. (3) shows that the log-likelihood of W measures $\boldsymbol{n}^{(1)}, \ldots, \boldsymbol{n}^{(W)}$ of a spectrum $\boldsymbol{S}$ is:

$$\ell = \sum_{w=1}^{W} \left[\log N! + \sum_{m=1}^{P} n_m^{(w)} \log p_m^{(w)} - \log(n_m^{(w)}!) \right]. \qquad \textbf{(10)}$$

Since $S_K = 1 - \sum_{k=1}^{K-1} S_k$, then $\partial_{S_k} p_m^{(w)} = F_m^{(w)}(k) - F_m^{(w)}(K) = \delta F_m^{(w)}(k)$ for $k \in [\![1, K-1]\!]$ and thus

$$\frac{\partial \ell}{\partial S_k} = \sum_{w=1}^{W} \sum_{m=1}^{P} \frac{\delta F_m^{(w)}(k)\ n_m^{(w)}}{p_m^{(w)}} \ \forall k \in [\![1, K-1]\!]$$

$$\frac{\partial^2 \ell}{\partial S_k \partial S_l} = -\sum_{w=1}^{W} \sum_{m=1}^{P} \frac{\delta F_m^{(w)}(k)\ n_m^{(w)}\ \delta F_m^{(w)}(l)}{(p_m^{(w)})^2} \ \forall k, l \in [\![1, K-1]\!].$$

Since $\langle n_m^{(w)} \rangle = N p_m^{(w)}$ and $[\boldsymbol{I}_F]_{k,l} = -\langle \partial_{S_k} \partial_{S_l} \ell \rangle$, $\boldsymbol{I}_F = N\, \boldsymbol{G}\, \boldsymbol{Q}\, \boldsymbol{G}^T$, with $\boldsymbol{G}$ and $\boldsymbol{Q}$ from Eq. (8) and Eq. (9).

Acknowledgments. The authors are grateful to Philippe Réfrégier for his advice and guidance and to Camille Scotté and Hervé Rigneault for fruitfull discussions.

Disclosures. The authors declare no conflicts of interest.

Data availability. Data underlying the results presented in this paper are not publicly available at this time but may be obtained from the authors upon reasonable request.

REFERENCES

1. R. A. DeVerse, R. M. Hammaker, and W. G. Fateley, Appl. Spectrosc. **54**, 1751 (2000).
2. B. M. Davis, A. J. Hemphill, D. Cebeci Maltaş, M. A. Zipper, P. Wang, and D. Ben-Amotz, Anal. chemistry **83**, 5086 (2011).
3. D. Wilcox, G. Buzzard, B. Lucier, P. Wang, and D. Ben-Amotz, Anal. chimica acta **755**, 17 (2012).
4. D. Wilcox, G. Buzzard, B. Lucier, O. Rehrauer, P. Wang, and D. Ben-Amotz, The Analyst **138** (2013).
5. P. Réfrégier, C. Scotté, H. B. de Aguiar, H. Rigneault, and F. Galland, J. Opt. Soc. Am. A **35**, 125 (2018).
6. C. Scotté, H. B. de Aguiar, D. Marguet, E. M. Green, P. Bouzy, S. Vergnole, C. P. Winlove, N. Stone, and H. Rigneault, Anal. chemistry **90**, 7197 (2018).
7. P. Réfrégier and F. Galland, Opt. Lett. **44**, 2228 (2019).
8. P. Réfrégier, E. Chevallier, and F. Galland, Optics Letters **44**, 5836 (2019).
9. C. Scotté, S. Sivankutty, P. Stockton, R. A. Bartels, and H. Rigneault, Opt. letters **44**, 1936 (2019).
10. C. Scotté, S. Sivankutty, R. A. Bartels, and H. Rigneault, Opt. Lett. **45**, 5567 (2020).
11. D. Cebeci, B. R. Mankani, and D. Ben-Amotz, J. Imaging **5**, 1 (2019).
12. O. G. Rehrauer, V. C. Dinh, B. Mankani, G. Buzzard, B. Lucier, and D. Ben-Amotz, Appl. Spectrosc. **72**, 69 (2018).
13. P. H. Garthwaite, I. T. Jolliffe, and B. Jones, *Statistical Inference* (Wiley, 2002), 2nd ed.
14. P. Berto, C. Scotté, F. Galland, H. Rigneault, and H. B. de Aguiar, Opt. letters **42**, 1696 (2017).

FULL REFERENCES

1. R. A. DeVerse, R. M. Hammaker, and W. G. Fateley, "Realization of the hadamard multiplex advantage using a programmable optical mask in a dispersive flat-field near-infrared spectrometer," Appl. Spectrosc. **54**, 1751–1758 (2000).
2. B. M. Davis, A. J. Hemphill, D. Cebeci Maltaş, M. A. Zipper, P. Wang, and D. Ben-Amotz, "Multivariate hyperspectral raman imaging using compressive detection," Anal. chemistry **83**, 5086–5092 (2011).
3. D. Wilcox, G. Buzzard, B. Lucier, P. Wang, and D. Ben-Amotz, "Photon level chemical classification using digital compressive detection," Anal. chimica acta **755**, 17–27 (2012).
4. D. Wilcox, G. Buzzard, B. Lucier, O. Rehrauer, P. Wang, and D. Ben-Amotz, "Digital compressive chemical quantitation and hyperspectral imaging," The Analyst **138** (2013).
5. P. Réfrégier, C. Scotté, H. B. de Aguiar, H. Rigneault, and F. Galland, "Precision of proportion estimation with binary compressed raman spectrum," J. Opt. Soc. Am. A **35**, 125–134 (2018).
6. C. Scotté, H. B. de Aguiar, D. Marguet, E. M. Green, P. Bouzy, S. Vergnole, C. P. Winlove, N. Stone, and H. Rigneault, "Assessment of compressive raman versus hyperspectral raman for microcalcification chemical imaging," Anal. chemistry **90**, 7197–7203 (2018).
7. P. Réfrégier and F. Galland, "Bhattacharyya bound for raman spectrum classification with a couple of binary filters," Opt. Lett. **44**, 2228–2231 (2019).
8. P. Réfrégier, E. Chevallier, and F. Galland, "Compressed Raman classification method with upper-bounded error probability," Optics Letters **44**, 5836 (2019).
9. C. Scotté, S. Sivankutty, P. Stockton, R. A. Bartels, and H. Rigneault, "Compressive raman imaging with spatial frequency modulated illumination," Opt. letters **44**, 1936–1939 (2019).
10. C. Scotté, S. Sivankutty, R. A. Bartels, and H. Rigneault, "Line-scan compressive raman imaging with spatiospectral encoding," Opt. Lett. **45**, 5567–5570 (2020).
11. D. Cebeci, B. R. Mankani, and D. Ben-Amotz, "Recent trends in compressive raman spectroscopy using dmd-based binary detection," J. Imaging **5**, 1 (2019).
12. O. G. Rehrauer, V. C. Dinh, B. Mankani, G. Buzzard, B. Lucier, and D. Ben-Amotz, "Binary complementary filters for compressive raman spectroscopy," Appl. Spectrosc. **72**, 69 – 78 (2018).
13. P. H. Garthwaite, I. T. Jolliffe, and B. Jones, *Statistical Inference* (Wiley, 2002), 2nd ed.
14. P. Berto, C. Scotté, F. Galland, H. Rigneault, and H. B. de Aguiar, "Programmable single-pixel-based broadband stimulated raman scattering," Opt. letters **42**, 1696–1699 (2017).