

Optimal Filters for High-Speed Compressive Detection in Spectroscopy

Gregery T. Buzzard^a and Bradley J. Lucier^{a,b}

^aDepartment of Mathematics, Purdue University, West Lafayette, IN, USA;

^bDepartment of Computer Science, Purdue University, West Lafayette, IN, USA

ABSTRACT

Recent advances allow for the construction of filters with precisely defined frequency response for use in Raman chemical spectroscopy. In this paper we give a probabilistic interpretation of the output of such filters and use this to give an algorithm to design optimal filters to minimize the mean squared error in the estimated photon emission rates for multiple spectra. Experiments using these filters demonstrate that detecting as few as ~ 10 Raman scattered photons in as little time as $\sim 30 \mu\text{s}$ can be sufficient to positively distinguish chemical species. This speed should allow “chemical imaging” of samples.

Keywords: Raman spectroscopy, optimization, binary filters, experiment design, compressive detection, uncertainty quantification

1. INTRODUCTION

Rapid identification and quantification of chemical species in complex mixtures is important in a diverse array of fields, including medicine and biology, pharmacy, counterfeit detection, and national security. A central question in this field is how best to take measurements (subject to specified constraints) to minimize uncertainty in the estimate of a quantity of interest. In this paper, we focus on the selection of optimal filters and measurement times to minimize the error in estimating photon emission rates associated with various chemical species to determine chemical identities and recover concentrations. Although our focus here is on Raman scattering, the underlying ideas apply to optimal measurements more broadly, including light scattering, reflectance, and nonlinear optics-based chemical and materials analysis.

Conventional optical array (e.g., CCD) based spectrometers disperse light of different wavelengths onto N separate detectors in order to measure a spectrum (here N is typically 100 or more). While this approach is advantageous for many applications, it has important drawbacks in the low-signal regime because of the inherent read noise associated with CCD measurements. For example, if we assume that over a fixed time period a given chemical species emits ~ 100 photons distributed over ~ 100 CCD pixels, then the resulting signal at each pixel would be well below the typical CCD read noise of a few counts per pixel. Nevertheless, with full spectral information, multivariate statistical techniques can be employed to address problems requiring component classification, calibration, and hyperspectral imaging. In particular, these techniques have shown much promise in the problem of spectral unmixing: i.e., estimating chemical composition from a combined spectrum using the spectroscopic signatures of various constituents.^{1–6}

In contrast, several previous studies have shown that chemometric techniques could be incorporated directly into the optical measurement process by using either static optical interference filters,⁷ or tunable liquid crystal^{8–10} or micromirror^{11–15} based multivariate optical elements built into the spectrometer hardware. The general principle of these filters is to combine photons of various frequencies in a prescribed way in order to facilitate the estimation of a quantity of interest. The versatility of the tunable multivariate optical element permits these instruments to function as generalized spectrometers, capable of either full spectral acquisition or compressive detection using programmable optical filter functions.

In this paper we consider the problem of selecting optimal filters for the so-called “supervised spectral unmixing problem.” That is, we assume that we have a mixture of a relatively small set of known component

Further author information: G.T.B.: E-mail: buzzard@math.purdue.edu; B.J.L.: E-mail: lucier@math.purdue.edu

species and that the full spectrum of each possible component is known. The goal then is to estimate the concentration of each species in the sample. In fact, we don't estimate concentration directly but rather estimate the rate of photons emission from a given species and assume that concentration can be calculated from this value. We assume that there is only a single photon counting device with no read noise (this is essentially true for high-quality photomultiplier tubes) and that a single filter can modulate and combine photons of various frequencies subject to specified constraints (more precise statements appear later). We look for multiple filters, applied sequentially for variable times, to estimate the emission rates associated with each possible component of the sample. As will be explained, the mean-square error for a given set of filters decays in inverse proportion to the total measurement time, so we assume a unit total time constraint. The objective is to minimize the expected mean square error between estimated and true emission rates.

Our main results are

1. For a number of filters equal to a small number of possible components, the optimal filters for this problem are nearly binary; i.e., apart from a small number of frequencies in each filter, the photons of a given frequency are either all counted or not counted at all.
2. The problem of selecting optimal filters for a small number of components can be decomposed into a low-dimensional nonlinear optimization coupled with a high-dimensional linear programming problem.

The problem of finding optimal filters as described above falls squarely into the subfield of statistics known as *Design of Experiments*, specifically with so-called *linear models*. A mathematical summary of the results in the field as of 1993 is given by Pukelsheim.¹⁶ Because our data are fundamentally photon counts, modeled by Poisson random variables whose variances equal their means, our problem does not fit precisely into the framework he develops, but we adopt his point of view in what follows.

Bioucas-Dias and Plaza⁴ and Keshava⁵ give very general introductions to the problem of supervised spectral unmixing. Work more closely related to ours also uses a Poisson model (with additional error terms) but assumes full spectrum measurements and hence does not consider the choice of optimal filters for multiplex measurements.^{1,17} Another related paper considers the problem of parameter estimation from Poisson observations in the case that there are more parameters than measurements.⁶ This latter paper also includes an informative discussion on other approaches to Poisson estimation.

We note here that the advantages of our compressive detection strategy are primarily restricted to the low signal regime with a relatively small number of component species. A conventional CCD detection scheme may be preferred under high signal conditions (under which read and dark noise do not significantly affect the spectral signal-to-noise ratio) or in cases in which the number of filters required is too large to be practical.

2. PROBLEM FORMULATION

We assume our sample consists of a mixture of n chemical species S_j , $j = 1, \dots, n$; after excitation with a laser, the j th species S_j gives off a stream of photons modeled as a Poisson process, a stream that hits our detector at a rate Λ_j^* photons per second. Here Λ^* is a (column) vector of mean photon emission rates of all chemical species S_j under consideration.

Each photon in the stream of photons given off by the j th species S_j has an energy that falls into one of N bins, and P is a matrix that describes this energy distribution: P_{ij} is the probability that a photon from the photon stream given off by S_j has an energy that falls into the i th energy bin. (In other words, the j th column of P is the spectrum of S_j , normalized to have area 1.) Thus the stream of photons with energies that lie in the i th energy bin from all chemical species hits our detector at an average rate $(P\Lambda^*)_i$, the i th component of the vector $P\Lambda^*$. Hence we assume that the number of photons measured in bin i for time τ is a Poisson random variable with mean and variance both equal to $\tau(P\Lambda^*)_i$. We assume that the columns of P are linearly independent.

When using a charge-coupled device (CCD) to estimate photon rates, one can measure simultaneously the number of photons hitting each of the N receptor bins over the same period of time of length τ . The randomness of the Poisson counts in each bin can be interpreted as noise, called *shot noise* (e.g., [18, §5.5]). But for CCD

receptors, there is further noise associated with each measurement, so-called *read noise*, which can reasonably be modeled as Gaussian with mean 0 and fixed variance σ^2 ; this noise is independent of the number of photons hitting each bin of the detector.

For measurements using optical filtering devices (OFD), we assume that we make m independent measurements, with $m \geq n$, and that in the j th measurement the probability of detecting a photon with energy falling in the i th bin is given by F_{ij} , where $0 \leq F_{ij} \leq 1$. A value of $F_{ij} = 1$ means a complete pass-through to the detector of photons of energy i in the j th experiment, while $F_{ij} = 0$ means a complete blockage of photons with this energy. We assume that in the j th experiment, we note the aggregate count of photons hitting the receptor with all energies for a period of time τ_j ; thus, this count can be modeled as a Poisson random variable with mean $\tau_j \sum_{i,k} F_{ij} P_{ik} \Lambda_k^*$. If we denote by T the matrix with the values τ_i on the diagonal (so $T_{ii} = \tau_i$), then the vector of counts of photons hitting the detector from all measurements consists of independent Poisson random variables with means $TF^T P \Lambda^*$.

The CCD measurement model can be considered a special case of the second model, where $\tau_i = \tau$ for all i and F_{ij} is 1 when $i = j$ and 0 otherwise. The CCD model with these fixed filters F is classical, but we review it briefly later.

The main problem we address is how to design *optimal* filters for OFDs, in the sense of minimizing the expected squared error of our estimate of Λ^* . To formulate this more precisely, let Λ^* , P , F , and $T = \text{diag}(\tau)$ be as above, and let \mathbf{e}_i denote the i th standard basis vector. A measurement using the i th filter for time τ_i is then the inner product of $F\mathbf{e}_i$ and $\tau_i P \Lambda^*$; each such measurement gives a Poisson random variable, \mathbf{x}_i , with mean and variance both equal to $\tau_i (F\mathbf{e}_i)^T P \Lambda^*$, and these \mathbf{x}_i are independent. We assume a total time constraint of T_{total} , so that the sum over τ_i is T_{total} . As will be seen later, for a fixed set of filters, the expected mean square error scales linearly with $1/T_{\text{total}}$, so for the purposes of designing filters, we can assume that $T_{\text{total}} = 1$.

Let $A = F^T P$, so that \mathbf{x} is a vector of m independent random variables, each Poisson, with $E(\mathbf{x}) = T A \Lambda^*$ and $D(\mathbf{x})$, the covariance matrix, equal to $\text{diag}(T A \Lambda^*)$, the diagonal matrix with main diagonal equal to $T A \Lambda^*$. When we discuss CCD measurements, there will be nontrivial read noise, which we model as a vector, \mathbf{z} , of i.i.d. Gaussian random variables with mean 0 and variance σ^2 . Note that replacing \mathbf{x} by $\mathbf{x} + \mathbf{z}$ preserves the mean and adds $\sigma^2 I$ to $D(\mathbf{x})$.

The goal then is to choose measurement filters F and times of exposure T to best estimate Λ^* based on samples, $\hat{\mathbf{x}}$, of \mathbf{x} (or of $\mathbf{x} + \mathbf{z}$). With filters and exposure times fixed, we use the best linear unbiased estimator in the sense of the Gauss–Markov Theorem to estimate Λ^* . The objective function in choosing the filters and times is the expected squared error in the estimate of Λ^* . Without loss of generality, we can write this linear estimation as $\hat{\Lambda} = B T^{-1} \hat{\mathbf{x}}$.

Given this, our problem is

Problem 1: For a given Λ^* and $N \times n$ matrix P of spectra, find m , the number of measurements, an $m \times m$ diagonal matrix, T , and an $N \times m$ filter matrix, F , to minimize

$$E(\|B T^{-1} \hat{\mathbf{x}} - \Lambda^*\|^2) \tag{1}$$

subject to $BA = I$, $A = F^T P$, $0 \leq F_{ij} \leq 1$, T diagonal with $T_{ii} \geq 0$ and $\text{tr}(T) = 1$.

Note that T_{ii} can be 0 only if the i th column of B is 0. Of course, if $m = n$ then the constraint $BA = I$ implies that $B = A^{-1}$. If $m > n$ and A and T are fixed, then B is the best linear unbiased estimator (we give an explicit formula in Theorem 3.3).

An important question in this problem is how the filters depend on the vector Λ^* . If there is no a priori knowledge of Λ^* , then we should perhaps solve a larger problem: find the minimizing F with Λ^* allowed to vary over a compact set. We show later that the F obtained by taking Λ^* equal to the vector of all 1s provides a reasonable bound for any Λ , so we defer questions of optimization over Λ^* to future research.

3. REFORMULATION

First we use standard variance calculations to provide an explicit formula for (1) in terms of the matrices in Problem 1. Recall that \mathbf{e}_i denotes the i th standard basis vector, so that $B\mathbf{e}_i$ is the i th column of B , and $(\mathbf{v})_i$ denotes the i th entry of a vector, \mathbf{v} .

Theorem 3.1. Assume the constraints given in Problem 1. Then

$$E(\|BT^{-1}\hat{\mathbf{x}} - \Lambda^*\|^2) = \sum_{i=1}^m \frac{1}{T_{ii}} \|B\mathbf{e}_i\|^2 \left((A\Lambda^*)_i + \frac{\sigma^2}{T_{ii}} \right). \quad (2)$$

Proof. With A fixed, let $E(B, T)$ denote the left-hand side in (2), and let $\mathbf{y} = BT^{-1}\mathbf{x} - \Lambda^*$. Since $E(\mathbf{x}) = TAA^*$, linearity implies that $E(\mathbf{y}) = 0$. Moreover, the covariance matrix of \mathbf{y} is $D(\mathbf{y}) = BT^{-1}D(\mathbf{x})(BT^{-1})^T$ (see, e.g., [19, Section 2.3.3]), where, as noted above, $D(\mathbf{x})$ is $\text{diag}(TAA^*) + \sigma^2 I$. Then $E(B, T)$ is the expected value of the quadratic form $\mathbf{y}^T \mathbf{y}$, which by [19, Section 2.7.1] is $\text{tr}(D(\mathbf{y}))$, where tr is the usual trace operator.

Since trace is cyclic (e.g., [19, Section 1.3.7]), and both T and $D(\mathbf{x})$ are diagonal, we have

$$E(B, T) = \text{tr}(T^{-1}(\text{diag}(A\Lambda^*) + \sigma^2 T^{-1})B^T B).$$

This gives the sum of the diagonal entries of $B^T B$, weighted by the diagonal entries in $T^{-1}(\text{diag}(A\Lambda^*) + \sigma^2 T^{-1})$. Since the i th diagonal entry of $B^T B$ is $\|B\mathbf{e}_i\|^2$, this proves the theorem. \square

We also consider the choice of optimal T_{ii} when $\sigma^2 = 0$.

Theorem 3.2. For fixed A and B , and $\sigma^2 = 0$, the optimal T_{ii} in Problem 1 are given by

$$T_{ii} = \frac{\|B\mathbf{e}_i\| \sqrt{(A\Lambda^*)_i}}{\sum_{j=1}^m \|B\mathbf{e}_j\| \sqrt{(A\Lambda^*)_j}}. \quad (3)$$

Proof. The optimization over the entries in T is equivalent to minimizing $\sum_i (a_i/t_i)$ subject to $t_i > 0$ and $\sum_i t_i = 1$, where the $a_i > 0$ are fixed. Define vectors $\hat{\mathbf{a}}$ and $\hat{\mathbf{t}}$ with components $\hat{a}_i = \sqrt{a_i}/\sqrt{t_i}$ and $\hat{t}_i = \sqrt{t_i}$. The constraint on the t_i implies that $\|\hat{\mathbf{t}}\| = 1$, and hence the Cauchy-Schwarz inequality implies $|\hat{\mathbf{a}} \cdot \hat{\mathbf{t}}| \leq \|\hat{\mathbf{a}}\|$, with equality exactly when $\hat{\mathbf{t}} = \gamma \hat{\mathbf{a}}$ for some scalar γ . In component form this is

$$\sum_i \sqrt{a_i} \leq \sqrt{\sum_i \frac{a_i}{t_i}}.$$

Minimizing the right hand side is equivalent to minimizing our objective, and the condition $\hat{\mathbf{t}} = \gamma \hat{\mathbf{a}}$ implies $t_i = \gamma \sqrt{a_i}$ for all i . Choosing γ to satisfy $\sum_i t_i = 1$ and rewriting in terms of A , B , and Λ^* gives the result. \square

The previous theorem implies that in the case of no read noise ($\sigma^2 = 0$) we can reformulate Problem 1 by replacing the expected value in (1) with the right hand side of (2) and replacing T_{ii} by the right hand side in (3). This gives

$$E(\|BT^{-1}\hat{\mathbf{x}} - \Lambda^*\|^2) = \left(\sum_{i=1}^m \|B\mathbf{e}_i\| \sqrt{(A\Lambda^*)_i} \right)^2. \quad (4)$$

Instead of minimizing the square, we can equivalently minimize the value itself. This gives the following problem, which is equivalent to Problem 1 when $\sigma^2 = 0$.

Problem 2: For a given Λ^* and $N \times n$ matrix P of spectra, find m , the number of measurements, $N \times m$ filter matrix, F , and $n \times m$ matrix, B , to minimize

$$\sum_{i=1}^m \|B\mathbf{e}_i\| \sqrt{(A\Lambda^*)_i} \quad (5)$$

subject to $BA = I$, $A = F^T P$, $0 \leq F_{ij} \leq 1$.

Applying a simple change of matrices to Problem 1, we give an explicit formula for the linear estimator B . This is the standard Gauss–Markov estimator, but we give a statement and proof for reference.

Theorem 3.3. For fixed, left invertible A , T , and σ^2 , let $\Gamma = \text{diag}(T^{-1}A\Lambda^*) + \sigma^2 T^{-2}$. Then the optimal B in Problem 1 is given by

$$B = (A^T \Gamma^{-1} A)^{-1} A^T \Gamma^{-1}.$$

Proof. Since Γ is diagonal with positive entries on the diagonal, we can let $\tilde{B} = B\Gamma^{1/2}$ and $\tilde{A} = \Gamma^{-1/2}A$. Then the minimization in Problem 1 is equivalent to minimizing the square of the Frobenius norm of \tilde{B} subject to $\tilde{A}^T \tilde{B}^T = I$. Applying the pseudo-inverse of \tilde{A}^T gives the result. \square

Note that if A is invertible, then $B = A^{-1}$. In general, for fixed A , (5) is convex in B , and the constraint $BA = I$ is convex. Hence the optimal B for this A can be found by solving a convex optimization problem.

4. NEARLY BINARY FILTERS

Recall that n is the number of spectra, while each filter has length N (the number of bins in each spectrum), and for our application N is much larger than n .

Theorem 4.1. For any (possibly non-optimal) number of measurements m , the optimal filters obtained as the solution to Problem 2 can be chosen to consist of only 0s and 1s, except for possibly $n - 1$ entries strictly between 0 and 1 in each filter.

Note that the theorem does not imply that the optimal filters must have nonbinary entries, only that they may.

Proof. Note that Problem 2 can be regarded as an optimization over matrices, A , that are realizable as $A = F^T P$, as well as over B , subject to the specified constraints. Note also that there is a maximal scaling of each row of A that minimizes (5) over all nonnegative row scalings. The maximal scaling for a row is found by solving a linear programming problem, and an argument with the KKT conditions for this problem yields the result. Details are presented in a separate paper.²⁰ \square

In a separate paper, we use a version of the Kiefer–Wolfowitz Equivalence Theorem to show that if the number of filters is allowed to increase, then the optimal filters can be chosen to be completely binary.

As noted in the proof of the previous theorem, each candidate A has a maximal row scaling, which may be found by linear programming. As noted above, after scaling, the matrix B may be found by solving a convex optimization problem. These two observations form the basis of the proof of the following theorem.

Theorem 4.2. The constrained optimization in Problem 2 can be formulated as an optimization over $m \times n$ matrices U with unit Euclidean length rows coupled with a linear programming problem of dimension N for each row of U and a convex optimization of dimension $n \times m$ for each U .

The utility of this theorem is most pronounced in the case when m (the number of filters) and n (the number of components) are much smaller than N (the number of bins). Rather than solving a nonlinear optimization problem over a space of dimension mN , we solve a nonlinear optimization problem over a space of dimension $m(n - 1)$ with a standard linear programming problem of dimension N embedded in the generation of candidate matrices.

5. DISCUSSION

The nonnegativity of A implies that if there is a number C such that all $\Lambda_i^* \leq C$ (i.e., if Λ^* is in the compact set $[0, C]^n$), then $(A\Lambda^*)_i \leq C(A\mathbf{1})_i$ for each i . Hence from (4), we have

$$E(\|BT^{-1}\hat{\mathbf{x}} - \Lambda^*\|^2) \leq C \left(\sum_{i=1}^m \|B\mathbf{e}_i\| \sqrt{(A\mathbf{1})_i} \right)^2,$$

where T is determined by Theorem 3.2. Thus by taking the optimal filter matrix, F_0 associated with $\Lambda^* = \mathbf{1}$, along with the corresponding B , we obtain reasonable bounds for any Λ^* with $\Lambda_i^* \leq C$. This estimate is possibly crude, and a better estimate for a given Λ^* can be obtained by calculating (5) using the A and B associated with F_0 .

We note that we can use a similar approach to estimate the error in a single component or a specified subset of components of Λ^* .²¹ That is, in the proof of Theorem 3.1, instead of calculating $E(\mathbf{y}^T \mathbf{y})$ to obtain the estimate in (2), we can instead calculate $E((\mathbf{y})_i^2)$ with essentially the same proof to obtain an estimate on the expected squared error in the i th component of Λ^* (or some specified subset of components).

Theorem 3.3 can be used to recover the generalized least squares approach to estimating Λ^* in the case of conventional CCD data collected in parallel with one measurement per frequency bin. In this case, the matrix A is the same as the distribution matrix, P , and T is T_{total} times the identity matrix, since all bins are measured in parallel for the same time. Applying the theorem with $A = P$ leads directly to generalized least squares estimation of Λ^* : $\hat{\Lambda} = (P^T D^{-1} P)^{-1} P^T D^{-1} \hat{\mathbf{x}}$, where $D = \text{diag}(TP\Lambda^*) + \sigma^2 I$ is the covariance matrix of \mathbf{x} .

In the absence of read noise, this simultaneous measurement approach has lower expected error than the approach presented here due to the fact that the data is collected in parallel, for an effective exposure time of NT_{total} . However, as seen in the numerical results below, when read noise is included, then the filters designed here provide better recovery, even with tighter time constraints.

In a separate paper we show that for a particular set of two (artificial) spectra, having three filters with total exposure time T_{total} can provide more accurate recovery than two filters with the same total exposure time.²⁰ In that paper we also use results from the experiment design literature to give an upper bound on the optimal number of filters.

6. NUMERICAL RESULTS

In this section we describe results obtained with our algorithm for filter design, including both numerical simulation results and experimental results (we are grateful to Dor Ben-Amotz and David Wilcox for the experimental results). Although we have carried out numerical experiments with up to 10 component species, we focus here on the case of 2 chemical components (n-hexane and methylcyclohexane) plus one background spectrum (which arises from the optics of the lab set-up). We use 3 filters, so that $m = n = 3$, in which case $B = A^{-1}$. Although there are various ways of handling the background spectrum, here the background is subtracted from the component spectra, and the resulting ‘‘pure’’ spectra are used for filter selection.

As described in Section 4, for each fixed nonnegative, 3×3 matrix, U , with row vectors of Euclidean length 1, there is a maximal diagonal matrix, R_U , and filter matrix, F_U , so that minimization of (5) over all F can be replaced by a minimization over U , with A taken as $R_U U$ for each such U and $B = A^{-1}$. Hence the full minimization problem (over three filters, each of length N) can be replaced by minimizing over $U \geq 0$ with unit length rows with R_U determined by solving a linear programming problem. That is, the nonlinear part of the optimization now has dimension 6 instead of a dimension of several hundred.

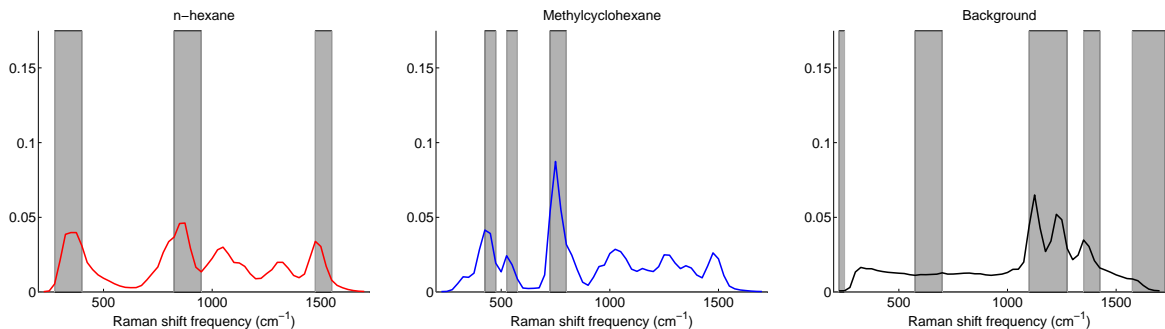


Figure 1. Normalized spectra along with corresponding OB filters. The filters are 1 for frequency shifts in the shaded regions.

The normalized spectra for n-hexane, methylcyclohexane, and background are shown in Figure 1. With these spectra as columns of P , we solved Problem 2 with $\Lambda^* = [1, 1, 1]^T$; the resulting filters (labeled OB filters for optimal binary) are shown as gray bars in that figure. With these spectra and filters, we calculate $A = F^T P$, $B = A^{-1}$ and the optimal T , which in this case are 0.34, 0.36, 0.30 in the order shown in Figure 1.

To simulate the results obtained from an experimental set-up using OB filters, we first fix a choice of Λ . This corresponds to the measured rates to be recovered, but we do not change the filters to reflect this choice of Λ . We then sample from a vector Poisson distribution with vector mean $T A \Lambda$ to obtain $\hat{\mathbf{x}}$. For a given value of T_{total} we take 1000 samples at each time point. For each sample, we estimate Λ by $B T^{-1} \hat{\mathbf{x}}$. This procedure mimics the estimation of Λ based on OB data.

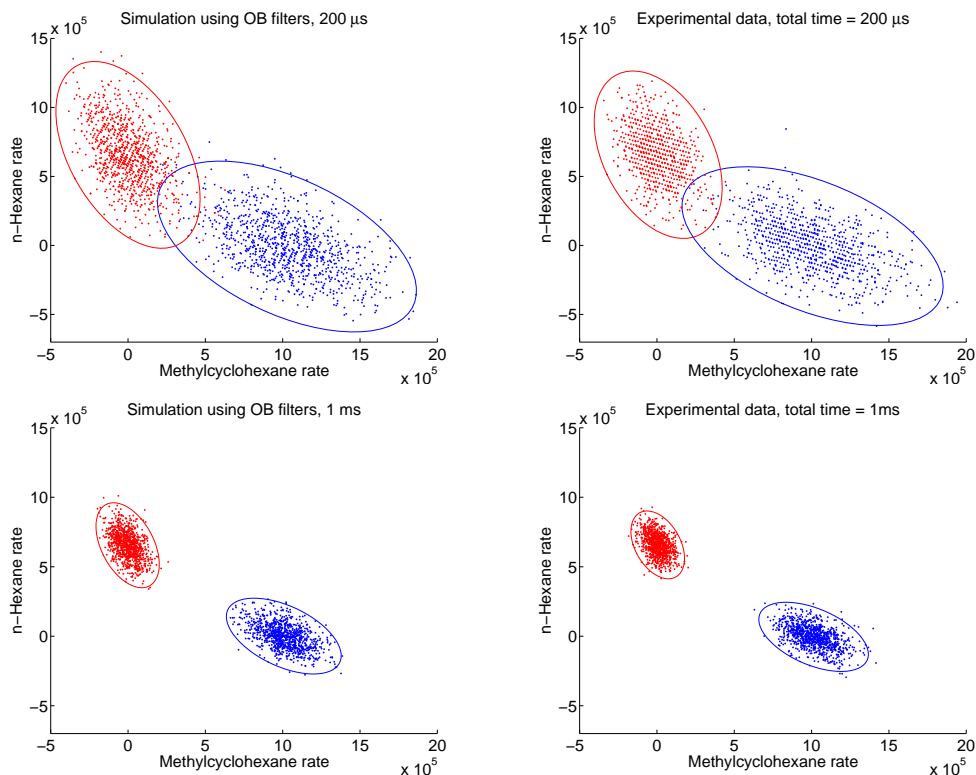


Figure 2. Estimated Λ using OB filters, simulated (left) and experimental (right).

The procedure for simulating results obtained from an experimental set-up using CCD detectors and estimation using generalized least squares is similar. For the same value of Λ , we sample from a vector Poisson distribution with vector mean $T_{\text{Total}} P \Lambda^*$. To include a realistic model of read noise, we used data collected from Dor Ben-Amotz’s lab on the counts obtained using CCD detectors kept completely dark and read after 0.1 seconds. After subtracting the detector-dependent mean from the data for each pixel, we had 134K sample values of read noise. We sampled uniformly at random from this collection to add read noise to the vector with Poisson distribution.

In Figure 2, we compare simulation using OB filters for 200 μs and 1 ms to experimental data collected using these OB filters for the same times. The two point clouds represent the results for two different experiments, one with a pure sample of each chemical. Based on experimental data, we used $\Lambda = (0, 1.01\text{e}6, 3.6\text{e}5)^T$ for pure methylcyclohexane, and $\Lambda = (6.55\text{e}5, 0, 3.6\text{e}5)^T$ for pure n-hexane. Each point in a cloud represents an independent estimate of Λ based on measured data. The ellipse is obtained by calculating the eigenvalue decomposition of the sample covariance matrix for a given cloud and using 3 standard deviations to determine the size of the ellipse. Comparing the figures, we see very good agreement between simulation and data, with somewhat tighter grouping for experimental data than for simulated data. In particular, this comparison validates

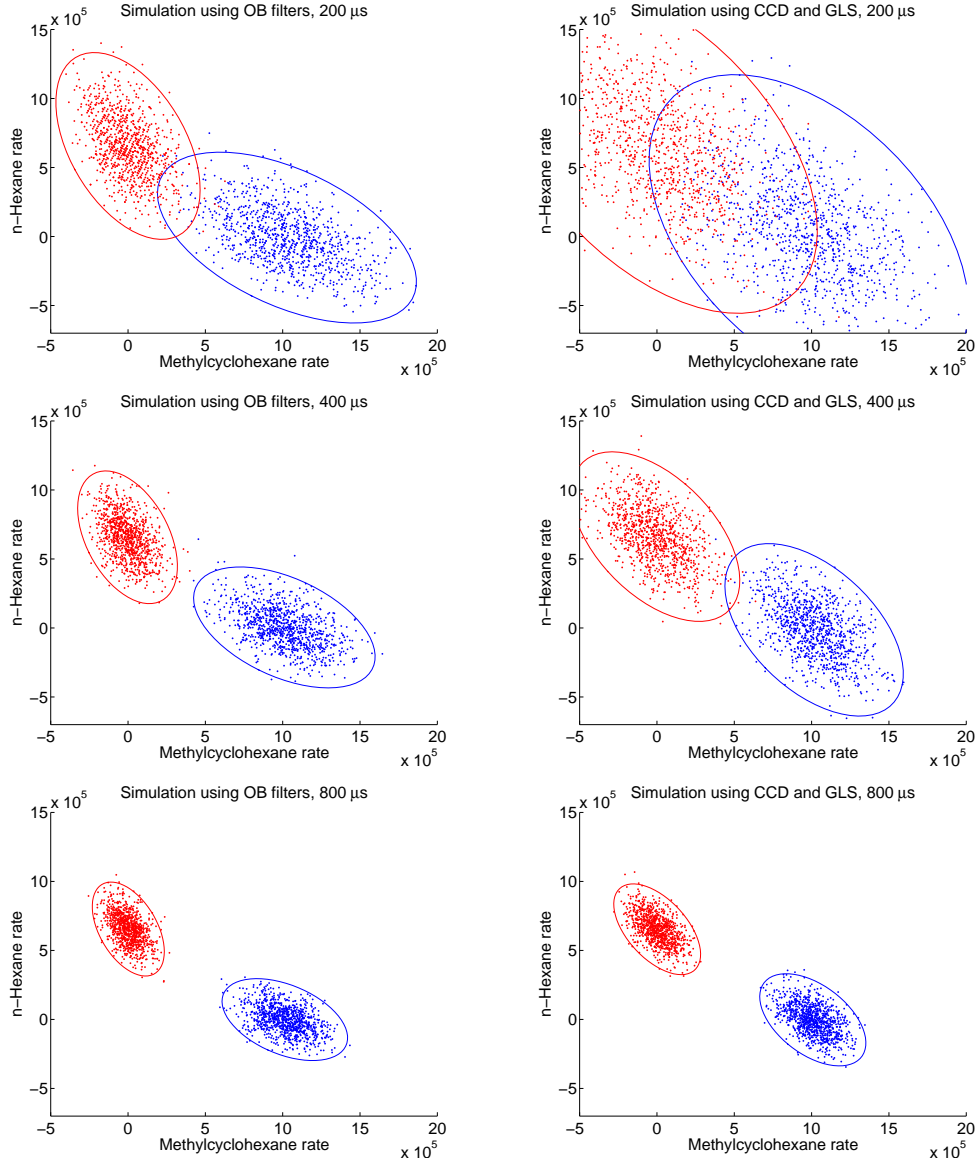


Figure 3. Estimated Λ using simulated OB filters (left) and simulated CCD measurements (right).

our assumption of zero read noise for the detector used with OB filters. Wilcox, et al.,²¹ provide further details and examples about comparison between simulation and experiment.

In Figure 3, we make a similar comparison using simulated OB filters and simulated CCD measurements with generalized least squares estimation. Here we see the pronounced advantage of the OB filters over CCD measurements for small total time. Note that each detector in the CCD array measures for the same time as the total time allotted to all 3 of the OB filters; nevertheless, the read noise inherent in CCD measurements greatly increases the variance of the results obtained with CCD measurements. As seen here, for this example the two methods are essentially equivalent when total time is $800\mu\text{s}$; for greater total time the CCD measurements have the advantage. This is made more precise in Figure 4, which plots the square root of the mean square error of the sampled rates for the two nonbackground spectra divided by the actual rate for the nonzero component rate. This gives an indication of how the uncertainty in the rates decreases with time.

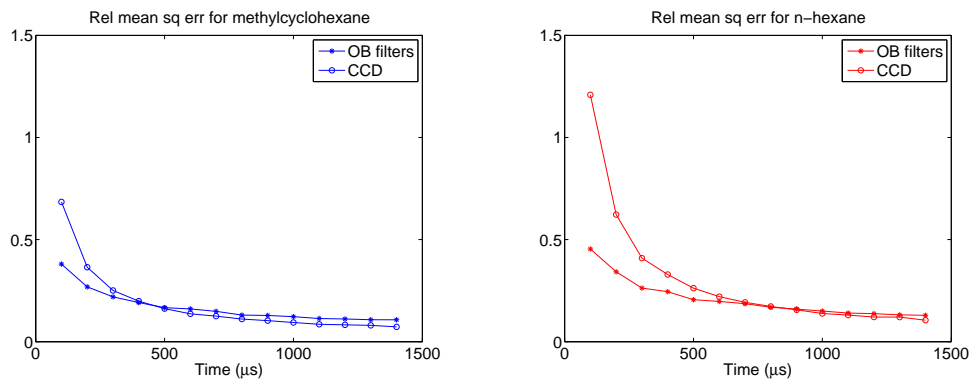


Figure 4. Square root of the mean square error of the sampled rates divided by the actual rate for the nonzero component rate for methylcyclohexane rate from pure methylcyclohexane sample (left) and for n-hexane rate from pure n-hexane (right), using both OB filters and CCD measurements (simulated results in both cases).

7. CONCLUSION

We have presented a model for the measurement of filtered spectroscopic data and developed a method for designing filters that minimizes the expected squared error in the estimation of the vector of photon emission rates (hence concentrations) of the various chemical compounds. We showed numerically that our method works well to recover this vector, even in the case of two similar spectra. For short exposure times and realistic read noise, the method given here for OB data outperforms generalized least squares on Raman CCD spectral data.

ACKNOWLEDGMENTS

We are grateful to Dor Ben-Amotz and his lab group at Purdue for many helpful discussions and for the data used to estimate the read noise for CCD data collection. We are grateful to Bill Studden for many helpful discussions on experiments and for this version of the proof of Theorem 3.2.

This work was supported in part by the Simons Foundation (Awards #209418 and #229816 to BJL).

REFERENCES

- [1] Palkki, R. D. and Lanterman, A. D., "Chemical mixture estimation under a poisson raman spectroscopy model," *Optical Engineering* **49**(11), 113601 (2010).
- [2] Palkki, R. D. and Lanterman, A. D., "Minimum description length approach to detecting chemicals via their raman spectra," *Opt. Eng.* **50**, 083601 (2011).
- [3] Drake, B., Kim, J., Mallick, M., and Park, H., "Supervised raman spectra estimation based on nonnegative rank deficient least squares," in [*2010 13th Conference on Information Fusion (FUSION)*], 1–8 (july 2010).
- [4] Bioucas-Dias, J. M. and Plaza, A., "Hyperspectral unmixing: geometrical, statistical, and sparse regression-based approaches," in [*Image and Signal Processing for Remote Sensing XVI*], *Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series* **7830** (Oct. 2010).
- [5] Keshava, N., "A survey of spectral unmixing algorithms," *Lincoln Laboratory Journal* **14**, 55–78 (2003).
- [6] Harmany, Z., Marcia, R., and Willett, R., "This is spiral-tap: Sparse poisson intensity reconstruction algorithms, theory and practice," *IEEE Transactions on Image Processing* **21**, 1084–1096 (March 2012).
- [7] Nelson, M. P., Aust, J. F., Dobrowolski, J. A., Verly, P. G., and Myrick, M. L., "Multivariate optical computation for predictive spectroscopy," *Analytical Chemistry* **70**(1), 73–82 (1998).
- [8] Davis, B. M., Hemphill, A. J., Maltas, D. C., Zipper, M. A., Wang, P., and Ben-Amotz, D., "Multivariate hyperspectral raman imaging using compressive detection," *Analytical Chemistry* **83**(13), 5086–5092 (2011).
- [9] Sweatt, W. C., Boye, C. A., Gentry, S. M., Descour, M. R., Stallard, B. R., and Grotbeck, C. L., "Isis: an information-efficient spectral imaging system," in [*Imaging Spectrometry IV*], *Proceedings of the Society of Photo-Optical Instrumentation Engineers (SPIE)* **3438**, 98–106 (1998).

- [10] Uzunbajakava, N., de Peinder, P., t Hooft, G. W., and van Gogh, A. T. M., “Low-cost spectroscopy with a variable multivariate optical element,” *Analytical Chemistry* **78**(20), 7302–7308 (2006).
- [11] Smith, Z. J., Strombom, S., and Wachsmann-Hogiu, S., “Multivariate optical computing using a digital micromirror device for fluorescence and raman spectroscopy,” *Optics Express* **19**(18), 16950–16962 (2011).
- [12] Hanley, Q. S., Verveer, P. J., and Jovin, T. M., “Optical sectioning fluorescence spectroscopy in a programmable array microscope,” *Applied Spectroscopy* **52**(6), 783–789 (1998).
- [13] DeVerse, R. A., Hammaker, R. M., Fateley, W. G., Graham, J. A., and Tate, J. D., “Spectrometry and imaging using a digital micromirror array,” *American Laboratory* **30**(21), 112S (1998).
- [14] Quyen, N. T., Da Silva, E., Dao, N. Q., and Jouan, M. D., “New raman spectrometer using a digital micromirror device and a photomultiplier tube detector for rapid on-line industrial analysis. part i: Description of the prototype and preliminary results,” *Applied Spectroscopy* **62**(3), 273–278 (2008).
- [15] Wagner, E. P., Smith, B. W., Madden, S., Winefordner, J. D., and Mignardi, M., “Construction and evaluation of a visible spectrometer using digital micromirror spatial light modulation,” *Appl. Spectrosc.* **49**, 1715–1719 (Nov 1995).
- [16] Pukelsheim, F., [*Optimal design of experiments*], vol. 50 of *Classics in Applied Mathematics*, Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA (2006). Reprint of the 1993 original.
- [17] Drake, B., Kim, J., Mallick, M., and Park, H., “Supervised raman spectra estimation based on nonnegative rank deficient least squares,” in [*2010 13th Conference on Information Fusion (FUSION)*], 1–8 (july 2010).
- [18] Brady, D., [*Optical imaging and spectroscopy*], Wiley: Hoboken, NJ (2009).
- [19] Koch, K.-R., [*Parameter estimation and hypothesis testing in linear models*], Springer-Verlag: Berlin (1999).
- [20] Buzzard, G. and Lucier, B., “Optimal filter design for spectroscopic compressive detection,” (2013). In preparation.
- [21] Wilcox, D. S., Buzzard, G. T., Lucier, B. J., Wang, P., and Ben-Amotz, D., “Photon level chemical classification using digital compressive detection,” *Analytica Chimica Acta* **755**, 17–27 (2012).