Towards Optimal Classifier of Spectroscopy Data

David D. Pokrajac, Poopalasingam Sivakumar, Yuri Markushin, Daniela Milovic, Mukti Rana, *Member, IEEE,* Gary Holness, Jinjie Liu, Noureddine Melikechi

***Abstract*—Laser spectroscopy can produce vast amounts of data, anticipating needs for automatization of tasks such as classification and discrimination of spectra. Using the apparatus of statistical theory of detection, we develop the optimal classifier for spectroscopy data for a linear model of an echelle spectrograph system. We validate model assumptions through statistical analysis of "dark signal" and laser-breakdown induced spectra of standardized NIST glass. The experimental results suggest that the quadratic classifier may provide optimal performance if the spectroscopy signal and noise can be considered Gaussian.**

***Index Terms*—Spectroscopy; echelle; laser-induced breakdown spectroscopy; optimal classifier; statistical learning theory.**

I. Introduction

Laser spectroscopy [e.g., 1] has reached stage where it can produce vast amounts of data, anticipating needs for automatization of tasks such as classification and discrimination of spectra. Automatic classification of spectroscopy data is a scientific and technical field where chemical molecules, compounds and mixtures are distinguished based on their spectral signatures by means of computer algorithms. Automatic classification has been attempted on various spectroscopy techniques: magnetic resonance [2], Fourier Transform infrared spectroscopy [3], Raman spectroscopy [4] and laser-induced breakdown data [5—9]. The utilized methods usually involve linear models (e.g., linear discriminant analysis [2, 3]) on amplitudes of some spectral components, selected by means of feature selection machine learning algorithms. Other publications describe utilization of principal component analysis of spectral components to reduce data dimensionality, followed by an instance based machine learning algorithm, that provides a linear or non-linear model [4, 7—10]. While these approaches may perform well in practice, they are ad-hoc and lack theoretical justification; more specifically, there is no assurance of their optimality from the point of view of statistical theory of detection.

Here, we concentrate on spectroscopy data from echelle spectrograph with an Intensified Charge Coupled Device (ICCD) sensors (iStar, Andor Technology, DH734-18F 03) [11, 12] and establish the optimal classifier for this type of data. Then, we utilize experimental data to validate assumptions leading to the model.

David D. Pokrajac, Poopalasingam Sivakumar, Yuri Markushin, Mukti Rana, Gary Holness, Jinjie Liu and Noureddine Melikechi are with The Optical Science Center for Applied Research, Delaware State University, 1200 N DuPont Hwy, Dover, DE, 19901, USA (e-mails: {dpokrajac, PSivakumar, yMarkushin, mrana, gholness, jliu, nmelikechi}@desu.edu).

Daniela Milovic is with School of Electrical Engineering, University of Nis, Aleksandra Medvedeva 14, 18000 Nis, Serbia (e-mail: dachavuk@gmail.com).

II. Methodology

A. Model of Spectroscopy System

The conceptual schema of the Andor Mechelle 5000 spectrograph system based on the echelle grating [11, 12] is shown in Fig.1a. The goal of the system is to measure spectrum $s_i(\lambda)$, $\lambda \in [\lambda_{min}, \lambda_{max}]$ of a light source. The light from a light source passes through diffraction grating, which creates the high dispersion of the wavelength into several different directions. Due to diffraction and interference [12, 13], the phenomena of spectral lines widening occurs, see Fig. 2. The spectral lines widening can be modeled through the following convolution:

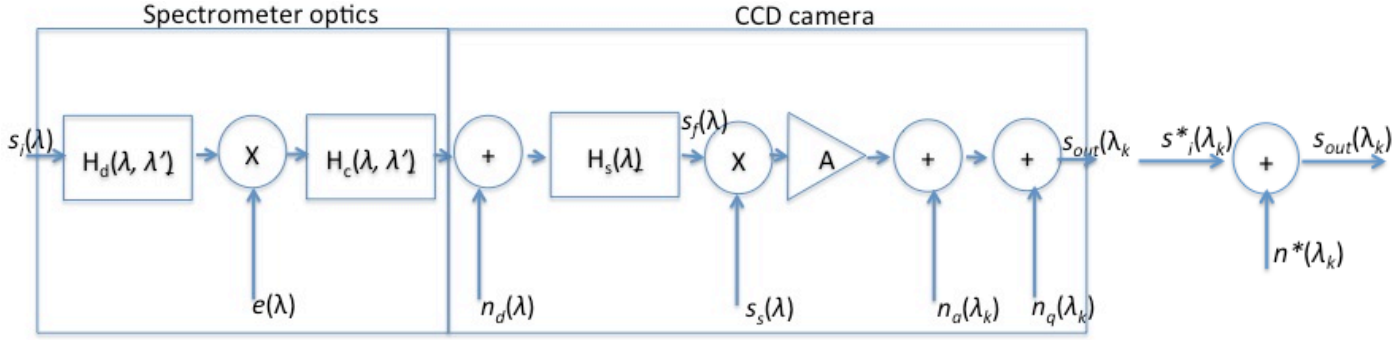
$$s_d(\lambda) = \int_{\lambda_{min}}^{\lambda_{max}} H_d(\lambda, \lambda') s_i(\lambda') d\lambda', \quad (1)$$

where $H_d(\lambda, \lambda')$ is a wavelength-dependent impulse response of the system.

The intensity of the measured signal is proportional to the echelle efficiency [13, 14] $e(\lambda)$ that is wavelength-dependent. Note that in the echelle spectrometer, high-order diffraction orders are utilized, and the measurements in each order appear as one linear pattern on the detector. The uneven distribution of orders may lead to closely stacking-up orders and cross-talk ("ghost lines") [11, 13]. We model cross-talk with a linear system with pulse response $H_c(\lambda, \lambda')$.

The light is converted into electric signal in a CCD sensor, where the number of electrons at each pixel is proportional to the intensity of the incident light at the pixel. In a CCD sensor, three types of noises exist based on the intensity of photo signal present on CCD pixel [15]. These three noises are: read-out noise (at low light intensities), shot noise (at medium intensities) and fixed pattern noise (at high intensities). The shot noise is a combination of photon noise and dark noise. Photon noise comes from random variation of photon flux from the light source while the dark noise is created because of the thermal generation of carriers. Fixed pattern noise exists because of the variation of charge created in individual pixels of CCD for photon signal input. Considering the laser signal as medium intensity, the dominating noise source for this case is shot noise, which comes mainly from dark current as the device was operating at room temperature. We assume that component of dark current noise is $n_d(\lambda)$ [16]. In the CCD sensor, the signal gets discretized in space (corresponding to discrete wavelengths λ_k) which we model with a low-pass filter $H_s(\lambda)$ followed by multiplication with a Dirac pulse trail $s_s(\lambda) = \sum_{k=1}^{K} \delta(\lambda - \lambda_k)$ [e.g., 17]. The pixel voltages get amplified (A) and quantized. The amplifier introduces the amplifier noise $n_a(\lambda_k)$. The quantization adds quantization noise $n_q(\lambda_k)$. The output of the system is therefore the signal $s_{out}(\lambda_k)$ discretized in the wavelength domain. Due to the linearity of the observed system, it can be simplified as shown in Fig. 1b. The output signal $s_{out}(\lambda_k)$ consists of the equivalent input signal $s^*_i(\lambda_k)$ and additive equivalent noise $n^*(\lambda_k)$.

Note that the input signal $s_i(\lambda_k)$ is proportional to the number of photons with energy hc/λ_k and hence has a Poisson distribution [18]. Under assumption that $s_i(\lambda_i)$ and $s_i(\lambda_j)$ are independent for $\lambda_i \neq \lambda_j$, since the sum of independent Poisson variables has Poisson distribution [19], $s^*_i(\lambda_k)$ has the Poisson distribution which can be approximated as Gaussian when its mean is large enough [20].
Dark current noise $n_d(\lambda)$ is here modeled as Gaussian [21]. The read-out noise $n_a(\lambda_k)$ consists of thermal (Johnson) noise and $1/f$ –noise (flicker) noise and can also be modeled as Gaussian [22]. Quantization noise $n_q(\lambda_k)$, on the other hand, has uniform distribution (if the quantizer is not overloaded) and is not correlated with the discretized signal value [23]. We assume that the number of quantization levels is large enough so that the influence of $n_q(\lambda_k)$ is small and that $n^*(\lambda_k)$ can also be modeled as Gaussian.
It is known [24] that dark current noise in CCD detectors is spatially uncorrelated (leading to $E(n_d(\lambda_i)n_d(\lambda_i)) = 0, \lambda_i \neq \lambda_j$) (here, E denotes expectation). We assume that the independence of the noise applies to *all* components of $n^*(\lambda_k)$, i.e., $E(n^*(\lambda_i)n^*(\lambda_i)) = 0, \lambda_i \neq \lambda_j$.



a) b)

Fig. 1. a) Block diagram of spectrograph; b) Simplified block diagram

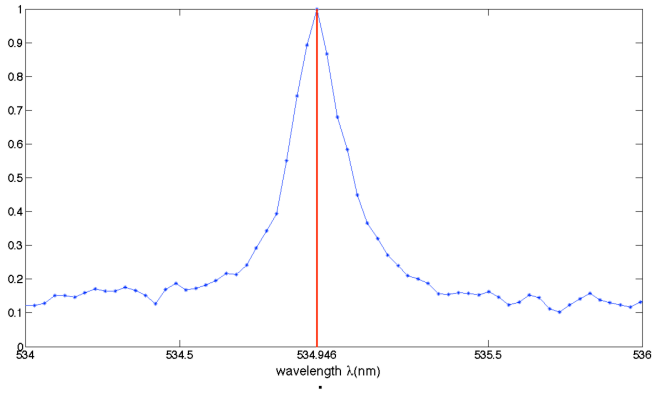


Fig. 2. Appearance of spectral line of 534.946nm of NIST standard reference wafer 612. Shown are effects of spectral line widening due to interference and diffraction at diffraction grating.

B. Optimal Classifier of Spectroscopy Data

The goal of classification is to distinguish between two hypotheses:

$$H_1: s_{out}(\lambda_k) = s^*_{i,1}(\lambda_k) + n^*(\lambda_k), k = 1, \dots K$$
$$H_2: s_{out}(\lambda_k) = s^*_{i,2}(\lambda_k) + n^*(\lambda_k), k = 1, \dots, K$$

based on observed values of $s_{out}(\lambda_k), k = 1, \dots, K$.

Following the discussion in Section II.A, we assume that $s^*_{i,1}(\lambda_k), s^*_{i,2}(\lambda_k)$ and $n^*(\lambda_k)$ are Gaussian. Since the sum of two Gaussian variables is always Gaussian [19], we can write hypotheses in the vector form:

$$H_1: \mathbf{s}_{out} = \mathbf{r}_1,$$
$$H_2: \mathbf{s}_{out} = \mathbf{r}_2, \quad (2)$$

where $\mathbf{r}_1, \mathbf{r}_2$ are K-variate Gaussian vectors. By the Gaussian assumption, a sample $\mathbf{s}_{out}$ has the following conditional probability density function under hypothesis H_i, i=1,2 [25]:

$$p(\mathbf{s}_{out}|H_i) = \frac{1}{(2\pi)^{K/2}|\mathbf{\Sigma}_i|^{1/2}} e^{-\frac{1}{2}(\mathbf{s}_{out}-\mathbf{m}_i)^T \mathbf{\Sigma}_i^{-1}(\mathbf{s}_{out}-\mathbf{m}_i)} \quad (3)$$

where the mean vectors $\mathbf{m}_i$ and K*K covariance matrices are defined as:

$$\mathbf{m}_i \triangleq E(\mathbf{r}_i),$$
$$\mathbf{\Sigma}_i \triangleq E\left((\mathbf{m}_i - \mathbf{r}_i)(\mathbf{m}_i - \mathbf{r}_i)^T\right), i = 1,2. \quad (4)$$

The likelihood ratio test [25] decides between hypotheses based on comparison of the likelihood ratio $\Lambda(\mathbf{s}_{out})$ with a threshold η:

$$\Lambda(\mathbf{s}_{out}) \triangleq \frac{p(\mathbf{s}_{out}|H_1)}{p(\mathbf{s}_{out}|H_0)} \underset{H_0}{\overset{H_1}{\lessgtr}} \eta. \quad (5)$$

The threshold η depends on the chosen performance criteria (e.g., minimization of total error as in maximum aposteriori probability test). By taking logarithm and arranging the terms that contain the observed output signal, from Eq. (5) we obtain the following log-likelihood test, which represents the optimal classifier under the Gaussian assumption:

$$l(\mathbf{s}_{out}) = \mathbf{s}_{out}^T \mathbf{A}\, \boldsymbol{s_{out}} + \mathbf{b}^T \boldsymbol{s_{out}} \underset{H_1}{\overset{H_2}{\lessgtr}} \gamma. \quad (6)$$

where:

$$\mathbf{A} \triangleq \frac{1}{2}(\mathbf{\Sigma}_1^{-1} - \mathbf{\Sigma}_2^{-1})$$
$$\mathbf{b} \triangleq \mathbf{\Sigma}_2^{-1}\mathbf{m}_2 - \mathbf{\Sigma}_1^{-1}\mathbf{m}_1 \quad (7)$$
$$\gamma \triangleq ln\, \eta + \frac{1}{2}(\ln|\mathbf{\Sigma}_2| - \ln|\mathbf{\Sigma}_1| + \mathbf{m}_2^T\mathbf{\Sigma}_2^{-1}\mathbf{m}_2 - \mathbf{m}_1^T\mathbf{\Sigma}_1^{-1}\mathbf{m}_1).$$

Note that when the statistical parameters of the output signal, Eq. (4), are known, the log-likelihood test, Eq. (6), results in decision boundary *quadratic* in terms of the observed output vector of the system.

Assuming the availability of a sufficiently large number (n>K+1) of experimental realizations, means and the invertible covariance matrices, Eq. (4), can be estimated from experimental data [25]. The estimates can be subsequently plugged into Eq. (6)-(7). Alternatively, an approximately optimal classifier can be obtained using support vector machines (SVM) [26] with the following polynomial kernel:

$$\kappa(\mathbf{x}, \mathbf{y}) = (\mathbf{x}^T\mathbf{y} + 1)^2 \quad (8)$$

where **x, y** are K dimensional feature vectors.

III. Experimental Results

A. Experimental Setup

We utilized Andor Mechelle ME5000 spectrograph with an ICCD camera (iStar, Andor Technology, DH734-18F 03), see Fig. 3. The spectral resolution was R=4000 corresponding to 4 pixels FWHM [27]. The total number of channels was 26,040. The wavelength range was 199.04—974.83nm. The spectrometer used orders m=21-100. The grating had 52.13 line/mm with grating constant d≈5-30μm, blazed at 32.35 degrees. The spectra were collected 50 ns after the laser pulse with integration time of 700 μs by an onboard digital delay generator (DDG) of the spectrograph. The CCD was kept at a stable temperature at -10 ^{o}C using a

Thermoelectric (TE) cooler of the spectrograph to reduce dark signal. To excite plasma in Laser-induced Breakdown Spectroscopy (LIBS) [28], a broadband CPA-Series Ti-Sapphire ultra-short laser (Clark-MXR, Inc, Model: 2210) generating 150 fs long pulses operating at 775nm was used. For experiments with dark signal, the laser beam was blocked.

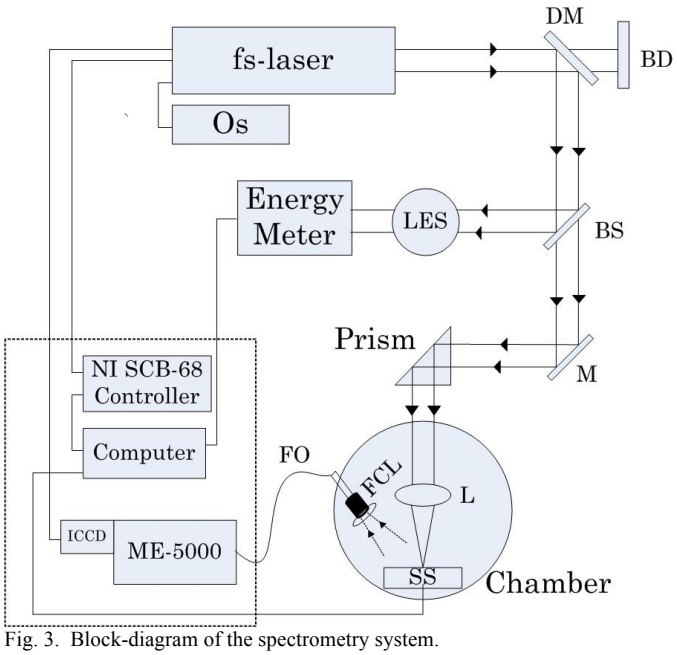


Fig. 3. Block-diagram of the spectrometry system.

B. Experiments of "Dark Signal"

To quantify characteristics of CCD sensor, 1000 dark spectra were acquired with no source of light incident to the sensor. The goal was to test the following hypotheses:

H_{01}: $s_{out}(\lambda_k)$ follows Gaussian distribution, $\lambda_k \in [200.33\text{nm}, 909.45\text{nm}]$. Note that outside this range the spectrometer provided signals equal to zero for all realizations. The total range of considered wavelengths included 24,650 discrete values.

H_{02}: $s_{out}(\lambda_i)$, $s_{out}(\lambda_j)$ are uncorrelated when $\lambda_i \neq \lambda_j$.

To test H_{01}, we used Kolmogorov-Smirnov [29] and Lilliefors test [30]. In addition, we computed skewness and kurtosis for observations $s_{out,i}(\lambda_k)$, i=1,…,1000 at each wavelength. The Kolmogorov-Smirnov test indicated that H_{01} can be rejected at 25 out of 24,650 wavelengths at the significance level α=0.05. The Lilliefors test indicated that H_{01} can be rejected at 1622, 366 and 212 wavelengths, with α=0.05, α=0.01 and α=0.005, respectively.

For the 25 wavelengths where H_{01} was rejected using the Kolmogorov-Smirnov test, we visually examined the histograms of 1,000 realizations. For wavelengths 211.9nm, 228.19nm, 303.82nm, the histograms indicated that the distribution of $s_{out}(\lambda_k)$ may be bimodal. For the other wavelengths, the histograms indicate presence of obvious outliers. These outliers (the maximal values) corresponded to eight realizations that were subsequently removed from the dataset.

The skewness and kurtosis [31] were calculated for each $s_{out}(\lambda_k)$ using the remaining 992 realizations. Fig. 4 and 5 show histograms of the obtained skewness and kurtosis.

To test H_{02}, we estimated normalized sample autocorrelation [32] of signals $s_{out}(\lambda_k)$ in the domain of discretized wavelengths λ_k, k=1,...,K. First, for each spectral order m, we determined discrete wavelengths $\lambda_{m,1} < \lambda_{m,2} < \cdots < \lambda_k < \cdots \lambda_{m,m_k}$ satisfying m=round($\frac{20139}{\lambda_k}$) (where λ_k is given in nanometers) [27]. Then, we computed sample autocorrelations $r_m(l)$ for signals $s(\lambda_{m,1}), \ldots, s(\lambda_{m,m_k})$ where the signals in each realizations were normalized to have the zero mean. Finally, we averaged normalized correlations $r_m(l)/r_m(0)$ for m=21,…100. The averaged normalized correlations for lags -20,…,20 are shown in Fig. 6.

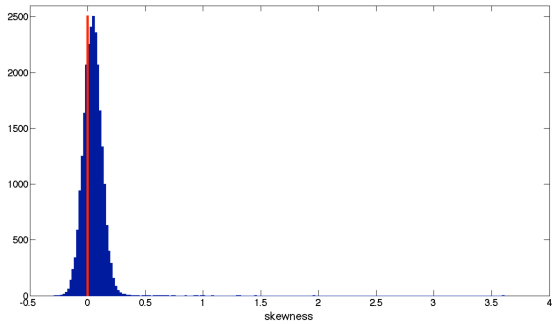


Fig. 4. Histogram of estimated skewness of "dark signal" at all observed wavelengths. The skewness of zero, characteristic for Gaussian distribution, is denoted by red line.

C. Experiments with Standardized Data

We measured laser-induced breakdown spectroscopy (LIBS) spectra [28] of NIST standardized glass. According to National Institute of Standards & Technology (NIST), the nominal composition of the standard reference wafer 612 used in this work is 72% SiO_2, 12% CaO, 14%

Na_2O, and 2% Al_2O_3. Total 61 trace elements are included in the glass support matrix. The reference wafer is specifically intended for evaluating analytical techniques used to determine trace elements in inorganic matrices [33]. For a sample of NIST standardized glass, we performed n =150 realizations of spectra. This was repeated seven times, for seven different samples.

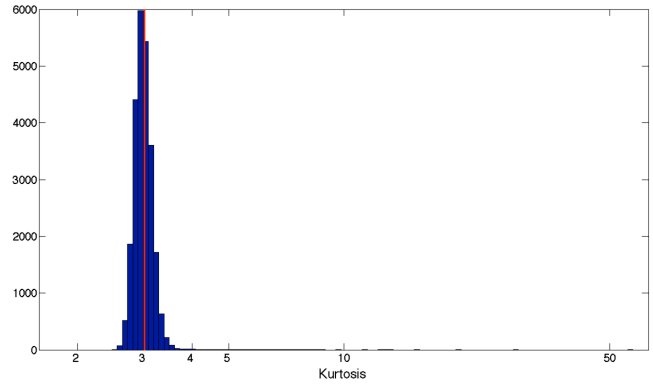


Fig. 5. Histogram of estimated kurtosis of "dark signal" at all observed wavelengths. The kurtosis of 3, characteristic for Gaussian distribution, is denoted by red line.

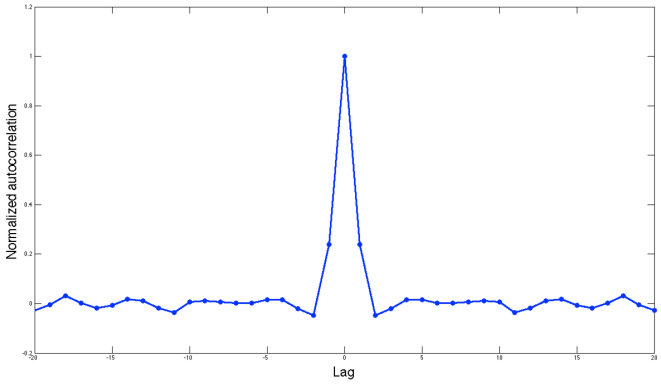


Fig. 6. Estimated normalized autocorrelation of "dark signal".

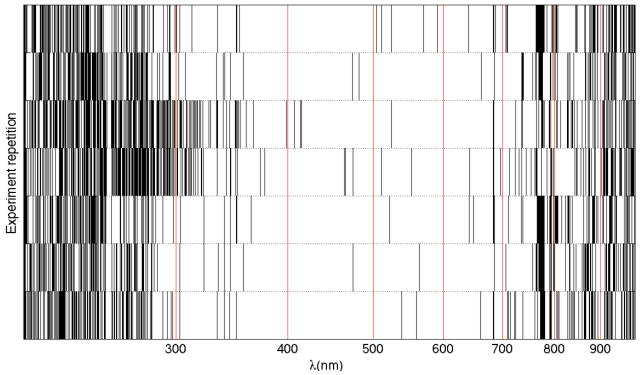
We repeated procedure indicated in Section III.B to test H_{01} (Gaussianity). Results of Kolmogorov-Smirnov test for all 7 repetitions and considered wavelengths are shown in Fig. 7.a for significance level α=0.05. Note that a "dark line" in the figure indicates that H_{01} can be rejected at the corresponding wavelength. Results of Lilliefors test for all 7 repetitions and considered wavelengths are shown in Fig. 7. b, c. for significance levels α=0.05, and α=0.005, respectively.

IV. Discussion

A. *Experiments of "Dark Signal"*

Kolmogorov-Smirnov (KS) test indicates that the hypothesis of Gaussian distribution of "dark signal" holds for almost all wavelengths (the H_{01} could not be rejected even with very large significance level α). Lilliefors test indicates that the number of wavelengths where the Gaussian distribution holds is smaller than the number indicated by the KS test. It is shown [34], that KS test tends to be inferior to Lilliefors test when the parameters of the Gaussian distribution are unknown. In such a case, the Lilliefors test has higher power (smaller probability of false acceptance of H_{01}). Hence, there is no wonder that the number of wavelengths where H_{01} is rejected (Gaussian distribution is not satisfied) is larger with the Lilliefors test than with the Kolmogorov-Smirnov test (for the same α=0.05). Histogram of estimated skewness, Fig. 4, indicates that the mode of the skewness is slightly larger than 0 (the skewness of normal distribution). From Fig. 5, the mode of kurtosis is around 3 (the kurtosis of the normal distribution). Based on these results, it can be concluded that the probability distribution of CCD noise is approximately normal for a large percentage of wavelengths.

The estimated normalized autocorrelation of "dark signal", Fig. 6., indicates that the dark noise samples are observably correlated only with the samples at adjacent wavelengths. Hence, H_{02} cannot be completely accepted. The assumption of whiteness (H_{02}), however, is *not* needed in our model.



a)

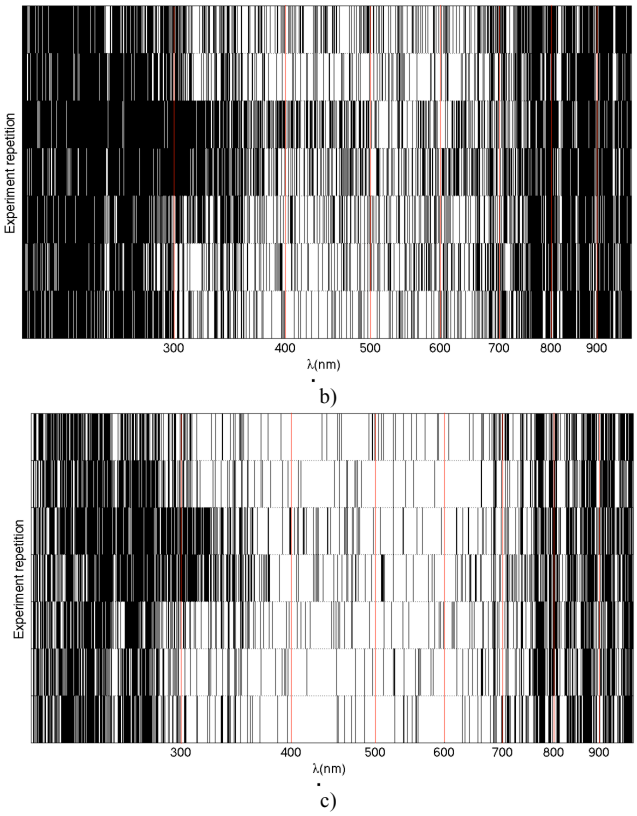


Fig. 7. Results of test of Gaussianity for 7 repetitions and all considered wavelengths. Dark lines indicate that H_0 (the distribution is Gaussian) can be rejected with specified significance. a)Kolmogorov-Smirnov test significance 0.05; b) Lilliefors test, significance 0.05; c) test, significance 0.005.

B. Experiments with NIST Glass

The results of Kolmogorov-Smirnov and Lilliefors tests on LIBS spectra of standardized NIST glass, Fig. 7, indicate that the distributions of signals $s_{out}(\lambda_k)$ can be considered approximately Gaussian for a large range of λ_k (notably, when $\lambda_k \in$[400nm, 700nm]). Due to observed Gaussian distribution of the dark signal, this leads to conclusion that $s_i(\lambda_k)$ in the considered case have approximate Gaussian distribution in this range of wavelengths.

C. Applicability of the Optimal Classifier

The optimal classifier presented in the paper is relatively simple (classification is performed by computing a quadratic function of observed discrete spectral components). This highly contrasts with sophisticated and complex classifiers previously attempted in literature [5, 7—8]. However, the applicability of the optimal classifiers depends on our ability to estimate statistical parameters of the output signal, Eq. (4). The estimation of covariance matrices $\mathbf{\Sigma}_i$ generally depends on the number of realizations available. Therefore, the direct application of the optimal classifier is possible if K is relatively small (i.e., considered wavelengths are chosen by an expert of by methods of feature selection [e.g., 35]). If K is large, the direct estimation of covariance matrices is not feasible, unless some additional assumptions about the structure of the matrices are imposed (e.g., only a few non-diagonal elements are different from zero; the matrices are Toeplitz, etc).

By employing Support Vector Machines (SVMs), we can estimate a hypothesis drawn from the function class of polynomials that both separates the data and achieves the maximum margin. SVMs carry the benefit of the descriptive power afforded by models with large degrees of freedom while incurring the complexity (VC-dimension) of a relatively small number of support vectors. In the SVM formulation, through "the kernel trick", a transformation of input space is implemented through the definition of its inner product over the set of in-sample data points. The kernel can be thought of as a transformation of the input space to a high dimensional representational space (or feature space). This also has the effect of further reducing the computational burden by avoiding computation of inner products in a high dimensional feature space. The resulting kernelized linear model in the feature space represents the equivalent non-linear model in the input space achieved by the inverse transformation on the final hypothesis. Note that classification of spectroscopy data using SVMs was successfully attempted in [36]. Note, however, that for large K, the actual estimation of model coefficients may require excessive computational power.

Eq. (6) represents the optimal classifier if the assumption of Gaussian distribution holds. Our experimental results indicate that the Gaussian distribution holds for noise and for *specific* spectroscopy signal in a range of wavelengths. In reality, signals $s_i(\lambda)$ have Poisson distribution. If distributions of the signals $s_i(\lambda)$ at two different wavelengths are *independent*, the signal components $s_f(\lambda)$ before sampling will also have Poisson distribution that can be approximated by Gaussian. However, if the distributions are *dependent*, $s_f(\lambda)$ as an integral of dependent

Poisson variables does not *have* to be Poisson random variable [37]. Further, $n^*(\lambda_k)$ may not be Gaussian random variables. If the assumptions of Gaussian distribution are not satisfied techniques of classification of non-Gaussian signals in generalized (non-Gaussian) noise need be considered [38—40].

Finally, the optimal classifier presented in this paper assumes that the signal flow in the spectrometer can be represented by *linear* systems $H_d(\lambda,\lambda'), H_c(\lambda,\lambda')$ and $H_s(\lambda)$. Further research is needed to develop the optimal classifier if the assumption of linearity does not hold.

V. Conclusions

We discussed the optimal classifier for a signal acquisition model in echelle spectrograph and validated model assumptions in a case of specific LIBS signal. We indicated that the optimal classifier has a quadratic decision boundary and can be approximated using SVMs with a quadratic kernel. Work in progress includes development of the optimal classifier when assumptions of Gaussianity and linearity are relaxed.

Acknowledgment

This work was supported in part by US Department of Defense Breast Cancer Research Program (HBCU Partnership Training Award #BC083639), the US National Science Foundation (CREST grant #HRD-1242067), CREST-8763, and NASA URC 7658 and the US Department of Defense/Department of Army (45395-MA-ISP, #54412-CI-ISP, W911NF-11-2-0046). Authors also want to thank Dr. Andrew Maidment and Dr. Predrag Bakic (Univ. Pennsylvania) and Dr. Vojislav Kecman (Virginia Commonwealth Univ.).

References

[1] S. Crouch, D. A. Skoog, *Principles of instrumental analysis*. Australia: Thomson Brooks/Cole, 2007.

[2] A.E. Nikulin, B. Dolenko, T. Bezabeh, R.L. Somorjai. "Near-Optimal Region Selection for Feature Space Reduction: Novel Preprocessing Methods for Classifying MR spectra," *NMR Biomed.* vol. *11*, no. *4-5*, pp. *209–216*, 1998.

[3] C. Beleites, G. Steiner, M.G. Sowa, R. Baumgartner, S. Sobottka, G. Schackert, R. Salzer. "Classification of Human Gliomas by Infrared Imaging Spectroscopy and Chemometric Image Processing," *Vibrational Spectroscopy*, vol. *38*, no. *1-2*, pp. *143–149*, 2005.

[4] B.K. Lavine, C.E. Davidson, A.J. Moores, P.R. Griffiths. "Raman Spectroscopy and Genetic Algorithms for the Classification of Wood Types," *Appl. Spectrosc*., vol. 55, no. 8, pp. *960-966*, 2001.

[5] E.G. Snyder, C.A. Munson, J.L. Gottfried, F.C. Jr. De Lucia, B. Gullett, A. Miziolek. "Laser-Induced Breakdown Spectroscopy for the Classification of Unknown Powders," *Appl. Opt.,* vol. *47*, no. *31*, pp. *G80-G87*, 2008.

[6] S. Sunku, E. N. Rao, G. M. Kumar, S.P. Tewari, S. V. Rao. "Discrimination methodologies using femtosecond LIBS and correlation techniques," Proc. SPIE 8726, Next-Generation Spectroscopic Technologies VI, pp. 87260H-87267H., 2013.

[7] T. Vance, D. Pokrajac, A. Marcano, Y. Markushin, S. McDaniel, N. Melikechi, A. Lazarevic. "Classification of LIBS Protein Spectra using Multi-Layer Perceptrons," *Transactions on Mass-Data Analysis of Images and Signals*, vol. *2*, no. *1*, pp. 96-111, 2010.

[8] D. Pokrajac, T. Vance, A. Lazarevic, A. Marcano, Y. Markushin, N. Melikechi, N. Reljin. "Performance of multilayer perceptrons for classification of LIBS protein spectra," Proc. 10th Symp. Neural Network Applications in Electrical Engineering (NEUREL), Belgrade, Serbia, pp. 171-174, 2010.

[9] T. Vance T, N. Reljin N, A. Lazarevic, D. Pokrajac, V. Kecman, N. Melikechi, A. Marcano, Y. Markushin, S. McDaniel. "Classification of LIBS protein spectra using support vector machines and adaptive local hyperplanes," Proc. 2010 IEEE World Congress on Computational Intelligence, Barcelona, Spain, pp. 1-7, July 2010.

[10] S. Dharmaraj, A.S. Jamaludin, H.M. Razak. R. Valliappan, N. A. Ahman, G. L. Harn, Z. Ismail. "The Classification of Phyllanthus Niruri Linn. According to Location by Infrared Spectroscopy," *Vibrational Spectroscopy*, vol. *41*, no. *1*, pp. *68–72*, 2006.

[11] M. Tripathi, "Echelle Spectrographs: a flexible tool for spectroscopy: Raman and LIBS Spectroscopy," Andor Technology, Belfast, UK, 2005, http://www.andor.com/pdfs/echelle_spectrograph.pdf.

[12] C. Palmer, E. Loewen, *Diffraction Grating Handbook,* 6th ed., Rochester, NY: Newport Corporation, 2005.

[13] E. Loewen and E. Popov, *Diffraction Gratings and Applications*, New York, NY: Marcel Dekker, Inc. 1997.

[14] M. Bottema, "Echelle efficiency and blaze characteristics, " in Proc. Periodic structures, gratings, moire patterns, and diffraction phenomena; San Diego, CA, July 29-August 1, 1980. (SPIE Proceedings. Volume 240), 1981, p. 171-176.

[15] K Faraji and W. James MacLean, "CCD Noise Removal in Digital Images", *IEEE Transaction of Image Processing*, vol. *5,* no. *9,*, pp *2676-2685*, 2006.

[16] -," CCD Image Sensor Noise Sources," Eastman Kodak Company, Rochester, NY, application note MTD/PS-0233, 2001.

[17] S. K. Mitra, *Digital Signal Processing: A Computer-based Approach,* New York NY: McGraw-Hill, 2006.

[18] L. Mandel, "Fluctuations of photon beams: the distribution of photo-electrons," *Proc. Phys. Soc.*, vol. *74*, no *3,* pp. *233-243,* 1959.

[19] G. Grimmett, D. Welsh: *Probability: An Introduction*, 1986, Oxford, UK: Oxford Science Publications, 1986.

[20] F. A. Haight, *Handbook of the Poisson Distribution,* Hoboken, NJ: Wiley, 1967.

[21] K A Jain, *Fundamental of Digital Image Processing*, Englewood Cliffs, NJ: Prentice-Hall, 1989.

[22] H. Tian, "Noise analysis in CMOS image sensors," Ph.D. dissertation, App. Physics Dept., Stanford Univ., Stanford, CA, 2000.

[23] J. G. Proakis, D. G. Manolakis, *Digital Signal Processing*, 3rd ed., Englewood Cliffs, NJ: Prentice-Hall, 1996.

[24] A. El Gamal, B. Fowler, H. Min, X. Liu, "Modeling and Estimation of FPN Components in CMOS Image Sensors," Proc. SPIE Solid State Sensor Arrays: Development and Applications II, Vol. 3301, pp. 168-177, 1998.

[25] H. L. Van Trees, K. L. Bell, *Detection Estimation and Modulation Theory, Part I*, 2nd ed., Hoboken, NJ: Wiley, 2013.

[26] N. Cristianini and J. Shawe-Taylor. *An Introduction to Support Vector Machines and other Kernel-based Learning Methods*, New York, NY: Cambridge University Press, 2000.

[27] -, "Andor™ Mechelle users guide," Andor Technology, Belfast, UK, version 3.0 Aug. 2008.

[28] A, W. Miziolek, V. Palleschi, I. Schechater, *Laser-induced Breakdown Spectroscopy (LIBS) fundamentals and Applications*, New York, NY: Cambridge University Press, 2006.

[29] F.J. Massey, "The Kolmogorov-Smirnov Test for Goodness of Fit," *Journal of the American Statistical Association*, vol. *46*, no. *253*, pp. *68–78*, 1951.

[30] H.W. Lilliefors, "On the Kolmogorov-Smirnov test for normality with mean and variance unknown," *Journal of the American Statistical Association*, Vol. 62, pp. *399–402*, , 1967.
[31] Y. Dodge, Y. *The Oxford Dictionary of Statistical Terms*, Oxford, UK: Oxford University Press, 2006.
[32] G. E. P. Box, G. M. Jenkins, G. C. Reinsel, *Time Series Analysis: Forecasting and Control*, 3rd ed, Englewood Cliffs, NJ: Prentice Hall, 1994.
[33] https://www-s.nist.gov/srmors/viewTableH.cfm?tableid=90 N.
[34] M. Razali, Y. N. Wah, "Power comparisons of Shapiro-Walk, Kolmogorov-Smirnov, Lilliefors and Anderson-Darling tests," *Journal of Statistical Modeling and Analytics*, vol. *2*, no. *1*, pp. *21-33*, 2011.
[35] H. Liu, H. Motoda, *Feature Selection for Knowledge Discovery and Data Mining*, Norwell, MA, USA: Kluwer, 1998.
[36] A. Bamgbade, R. Somorjai, B. Dolenko, E. Pranckeviciene, A. Nikulin, R. Baumgartner. "Evidence accumulation to identify discriminatory signatures in biomedical spectra," Proc. 10th Conf. on Artificial Intelligence in Medicine, AIME 2005, Aberdeen, UK pp. 463 – 467, July 23-27, 2005
[37] J. Jacod, "Two dependent Poisson processes whose sum is still a Poisson process*," J. Appl. Prob*. vol *12*, pp. *170-172*, 1975.
[38] D. Kletter, P.M. Schultheiss, H. Messer, "Optimal detection of non-Gaussian random signals in Gaussian noise,". Proc. ICASSP-91, Toronto, Ontario, Canada, vol. 2, 1305 – 1308, 14 May 1991-17 May 1991.
[39] A. H. Nuttall, "Optimum detection of random signal in non-Gaussian noise for low input signal to noise ratio," Naval Undersea Warfare Center Division, Newport, RI, NUWC-NPT Technical Report 11,512, 2/23/2004, http://www.dtic.mil/dtic/tr/fulltext/u2/a422595.pdf
[40] D. Middleton, *Non-Gaussian Statistical Communication Theory*, Hoboken, NJ: Wiley, 2012.