

# Constrained regularization for noninvasive glucose sensing using Raman spectroscopy

Wei-Chuan Shih

*Department of Electrical and Computer Engineering  
Department of Biomedical Engineering, University of Houston  
4800 Calhoun Rd., Houston, TX 77204, USA  
wshih@uh.edu*

Received 13 September 2014

Accepted 24 November 2014

Published 30 December 2014

Multivariate calibration is an important tool for spectroscopic measurement of analyte concentrations. We present a detailed study of a hybrid multivariate calibration technique, constrained regularization (CR), and demonstrate its utility in noninvasive glucose sensing using Raman spectroscopy. Similar to partial least squares (PLS) and principal component regression (PCR), CR builds an implicit model and requires knowledge only of the concentrations of the analyte of interest. Calibration is treated as an inverse problem in which an optimal balance between model complexity and noise rejection is achieved. Prior information is included in the form of a spectroscopic constraint that can be obtained conveniently. When used with an appropriate constraint, CR provides a better calibration model compared to PLS in both numerical and experimental studies.

*Keywords:* Glucose; noninvasive; multivariate calibration; partial least squares; principal component regression; Raman spectroscopy; constrained regularization.

## 1. Introduction

Multivariate calibration is a powerful analytical technique for extracting analyte concentrations in complex chemical systems with many hidden factors.<sup>1</sup> Numerous applications of multivariate calibration can be found in analyzing spectroscopic data, where information about all the analytes can be collected simultaneously at many wavelengths. Explicit calibration methods such as ordinary least

squares (OLS) and classical least squares (CLS) are often used when all of the constituent spectra can be individually measured or pre-calibrated using pure constituents. In contrast, implicit multivariate calibration techniques such as principal component regression (PCR) and partial least squares (PLS) allow the construction of a numerical model entirely based on spectroscopic and concentration data.<sup>2</sup> They are powerful because virtually no prior information

is needed. Explicit or implicit, the goal of multivariate calibration is to obtain a vector consisting of regression coefficients,  $\mathbf{b}$ , such that an analyte's concentration,  $c$ , can be accurately predicted by taking the scalar product of  $\mathbf{b}$  with an experimental spectrum,  $\mathbf{s}$ :

$$c = \mathbf{s}^T \cdot \mathbf{b}, \quad (1)$$

where lowercase boldface type denotes a column vector, uppercase boldface type a matrix; and the superscript<sup>T</sup> denotes transpose. The regression vector,  $\mathbf{b}$ , has the same length as the measured spectra, and thus can provide spectroscopic information regarding the calibration.  $\mathbf{b}$  is unique in an ideal noise-free linear system without constituent correlations, and the goal of both implicit and explicit schemes is to find an accurate approximation to  $\mathbf{b}$  for the experimental system of interest.

Recently, we proposed a novel hybrid technique, called constrained regularization (CR), where implicit multivariate calibration is viewed as an inverse problem.<sup>3</sup> It is hybrid owing to salient features of both explicit and implicit techniques. CR requires only the spectroscopic and concentration data, yet allows the incorporation of the target analyte spectrum as prior information. Starting with the inverse mixture model as the forward problem, we define the inverse problem with solution  $\mathbf{b}$ . Instabilities associated with the inversion process are removed by regularization, and prior information is included by means of a spectroscopic constraint.

Raman scattering provides molecular “fingerprinting” capability due to the inelastic interaction between incident photon and molecular vibration.<sup>2,4–7</sup> Raman spectroscopy offers several advantages such as no need for reagents and separation, nondestructiveness to the specimen, high sensitivity with plasmonic enhancement, capability of qualitative and quantitative measurements, providing molecular structure information with high spatial resolution, etc.<sup>8–12</sup> We demonstrated that CR provides improved performance using numerical models as well as experimental Raman spectra acquired from human subjects. We show that with CR both the standard error obtained using leave-one-out cross validation (SEV) and the standard error of prediction (SEP) improve compared to results obtained using PLS. To the best of our knowledge, this is the first time that CR has been applied to a

Raman skin model and *in vivo* data for noninvasive glucose sensing.

## 2. Theory

The forward problem for our calibration model is defined by the linear inverse mixture model for a single analyte:

$$\mathbf{c} = \mathbf{S}^T \mathbf{b}, \quad (2)$$

where  $\mathbf{S}$  is a set of calibration spectra, with each spectrum occupying a column of  $\mathbf{S}$ , associated with several known concentrations of the analyte of interest that are expressed as a column vector,  $\mathbf{c}$ , the  $j$ th element of which corresponds to the  $j$ th column of  $\mathbf{S}$ . The goal of the calibration procedure is to use the set of data  $[\mathbf{S}, \mathbf{c}]$  to obtain an accurate  $\mathbf{b}$  by inverting Eq. (2). The resulting  $\mathbf{b}$  can then be used in Eq. (1) to predict the analyte concentration,  $c$ , of an independent sample by measuring its spectrum,  $\mathbf{s}$ . The “accuracy” of  $\mathbf{b}$  is usually judged by its ability to correctly predict concentrations prospectively via Eq. (1).

CR enabled a convenient way to accomplish inversion with the flexibility of incorporation of a spectral constraint,  $\mathbf{b}_0$ , as described as the minimization of a quadratic cost function,  $\Phi$ <sup>13</sup>:

$$\Phi(\lambda, \mathbf{b}_0) = \|\mathbf{S}^T \mathbf{b} - \mathbf{c}\|^2 + \lambda \|\mathbf{b} - \mathbf{b}_0\|^2, \quad (3)$$

with  $\|\mathbf{a}\|$  the Euclidean norm (i.e., magnitude) of  $\mathbf{a}$ , and a spectral constraint that introduces prior information about  $\mathbf{b}$ . The first term of  $\Phi$  is the model approximation error, and the second term the norm of the difference between the solution and the constraint, which controls the smoothness of the solution and its deviation from the constraint. If  $\mathbf{b}_0$  is zero, the inversion is identical to standard Tikhonov regularization. As mentioned above, for  $\lambda = 0$  only least-norm (LM) solution is obtained. In the other limit, in which  $\lambda$  goes to infinity, the solution is simply  $\mathbf{b} = \mathbf{b}_0$ . In a previous volunteer study using PLS calibration, we retrospectively exploited the similarity between  $\mathbf{b}$  and  $\mathbf{g}$ , the spectrum of the analyte of interest, by calculating the correlation coefficient of  $\mathbf{b}$  with  $\mathbf{g}$ .<sup>14</sup> The result indicated that significant correlation between  $\mathbf{b}$  and  $\mathbf{g}$  offers confidence that the calibration is based on the analyte of interest rather than spurious effects such as co-variations among constituents. Therefore, the

spectrum of the analyte of interest is a good choice for the spectral constraint.

### 3. Experimental Methods

We have employed multivariate calibration techniques such as PLS to noninvasive glucose sensing using Raman spectroscopy.<sup>2,7,14</sup> Raman spectroscopy provides detailed quantitative information about sample composition, and is thus a promising technique for study of biological systems.<sup>15–21</sup> Glucose is a convenient analyte to study because its concentration can be conveniently altered and monitored in human volunteers. As shown below, numerical simulations and *in vivo* data collected noninvasively from the forearm of volunteers demonstrate that CR, when used with appropriate constraints, provides a calibration with significantly improved prediction accuracy compared to PLS.

**Numerical Study I.** The numerical data were chosen to closely simulate the human data. The predominant Raman spectral features sampled from the forearm are indicative of skin [Fig. 1(a)]. To simulate the forearm spectrum, we employed a model composed of nine representative constituents of the skin–blood–tissue matrix and the spectrum of glucose dissolved in water (Fig. 2). The choice of these constituents was based on the known composition of skin, and the relative amplitudes were

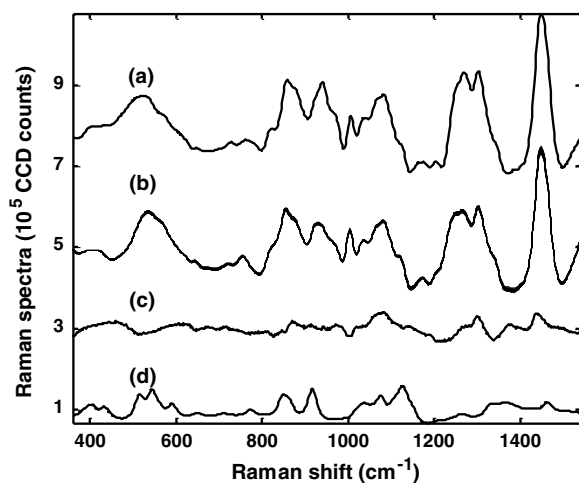


Fig. 1. (a) Typical Raman spectrum of skin with background removed; (b) typical simulated Raman spectra, 25 sample spectra are overlaid; (c) difference between the first two spectra in (b), magnified 10 $\times$ ; (d) glucose Raman spectrum, 90 mg/dL, magnified 100 $\times$ . The spectra are displaced vertically for better visualization.

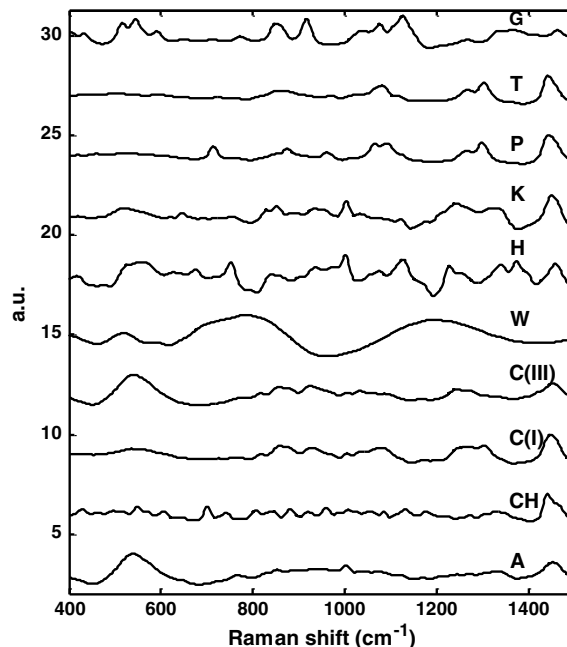


Fig. 2. Raman spectra of the 10 components used in the simulation: (A): actin; (CH): cholesterol; (CI): collagen I; (CIII): collagen III; (W): water; (H): hemoglobin; (K): keratin; (P): phosphatidylcholine; (T): triolein; (G): glucose.

chosen to approximate those of the skin–blood–tissue matrix. Glucose was included at physiological concentrations of 70 mg/dL to 210 mg/dL. The resulting simulations exhibited Raman signal ratios of glucose to the total matrix varying from 0.2% to 0.6%, which is the typical range measured in skin.<sup>22,23</sup> Such relative amounts of glucose were confirmed by studies in our laboratory employing minced samples of porcine skin, a good spectral model of human skin, with elevated levels of glucose. In simulating sample-to-sample variations, we varied all of the background constituent concentrations in a random fashion (SD 5% of the design spectral weights in Table 1), ensuring that there is no significant correlation between pairwise model constituents ( $r^2 \sim 0.02$ ). An appropriate amount of Gaussian random noise (SD 130 counts), estimated from the volunteer data, was added to each noiseless sample. We define the signal as the norm of the spectrum of interest. Total Raman signal-to-noise-ratio (SNR) (12,000) and glucose Raman SNR (24–72) can then be calculated by dividing the norm of the total Raman signal ( $1.5 \times 10^6$  counts) and the glucose signal (3120–9360 counts) by the noise magnitude (130 counts), respectively. Finally, to simulate reference concentration measurement error, Gaussian random error (SD 6 mg/dL) was added to the glucose concentrations, as well. The

Table 1. Model parameters.

Model constituents and spectral weights (%)				
Actin	Cholesterol	Collagen I	Collagen III	Water
1	2	49	7	3
Hb	Keratin	Phosphatidylcholine	Trolein	Glucose
6	15	4	13	0.2–0.6
Glucose range		70–210 mg/dL		
Spectral noise		130 counts		
Reference error		6 mg/dL		
Number of samples per calibration		25		
Total Raman SNR		~12,000		
Glucose Raman SNR		24–72		

model constituents and parameters are summarized in Table 1. Since these parameters are similar to their experimentally observed counterparts, we expect the numerical data to closely simulate the *in vivo* Raman spectra.

A typical Raman spectrum of human skin is shown in Fig. 1(a). The broad slowly varying background was fit to a fifth-order polynomial and subtracted off from the spectrum. A total of 25 simulated Raman spectra are shown in Fig. 1(b), with glucose and other constituents varied within the above design constraints. As can be seen, they approximate the observed features of skin Raman spectra very well. Figure 1(c) shows the difference between the first two simulated spectra, magnified 10 $\times$ , and Fig. 1(d) shows the model glucose spectrum at 90 mg/dL, magnified 100 $\times$ . The Raman signature of glucose is not apparent in either the sample spectra nor their difference, thus necessitating the use of multivariate calibration techniques.

**Numerical Study II.** In Numerical Study I, all model constituents were varied in an approximately random fashion. To study the effectiveness of CR with spurious correlations present, a second numerical study was performed, with significant constituent co-variations present in the calibration sample set: Strong correlation ( $r^2 \sim 0.72$ ) between hemoglobin and glucose concentrations, and exponential decays in the total Raman signal level from the first sample to the last. (The volunteer spectra manifested behavior of this type.<sup>14</sup>) All other model parameters were identical to those of the first numerical study.

**Data Analysis.** In both numerical studies, PLS, OLS and CR calibration methods were applied to these simulated spectra using code written in MATLAB (The MathWorks, Natick, MA). In each

case, a set of 25 calibration sample spectra were generated and a specific method was applied, with leave-one-out cross validation, to calculate the SEV and  $\mathbf{b}$  vector. This  $\mathbf{b}$  vector was then used to predict the concentrations of a set of 25 independent field samples, generated using the same model parameters, and the SEP calculated. SEV/P were calculated using the formula:

$$\text{SEV/P} = \sqrt{\frac{\sum_{j=1}^n \|\mathbf{a}_j - \mathbf{c}_j\|^2}{n}}, \quad (4)$$

with  $\mathbf{a}_j$  and  $\mathbf{c}_j$  the calculated and reference concentrations, respectively, and  $n$  the number of samples. This procedure was repeated 50 times and the results were averaged to obtain the estimates of mean error and variance. The selection of the regularization parameter was implemented as an automated search within the cross-validation framework, and the exact value has been selected when the SEV for a particular dataset is minimized. Therefore, it varied across different datasets.

**Volunteer Study.** We applied PLS and CR calibrations to *in vivo* transcutaneous Raman spectra collected from 17 *nondiabetic* human volunteers.<sup>14</sup> The full details of the protocol are given in Ref. 14. Briefly, volunteers fasted for 12 h prior to the experiment. Shortly after the start of the experiment, the subjects drank a highly concentrated glucose solution (Sun-dex, 75 g glucose). Raman spectra were collected from a  $\sim 1 \text{ mm}^2$  region of the forearm of a subject, excited by 830 nm light. The sampling volume was  $\sim 1 \text{ mm}^3$ . Each spectrum was the average of 90 consecutive two-second acquisitions. Spectra were acquired every 5 min over a period of 2 to 3 h. During this period, the blood glucose concentration typically doubled and then returned to its initial level. Reference capillary blood samples

were collected every 10 min and analyzed by means of a Hemocue glucose analyzer with a precision of SD 6 mg/dL. Spline interpolation was used to provide reference values at the 5-min intervals. Raw spectra were smoothed and then the wavelength was calibrated using an indene reference.

#### 4. Experimental Results

**Numerical Study I.** Table 2 summarizes the results from the application of each calibration method to the numerical data, using the results of PLS with 9 factors as a baseline. We observed the following: First, without regularization, the SEV is reduced only slightly, with solution led to simply the LN solution. The SEV is reduced only slightly when using CR with collagen I, the largest constituent of the simulation model, as the constraint, indicating that collagen I is a poor constraint. Finally, using CR with either the glucose spectrum or  $\mathbf{b}_{OLS}$  as the constraint, there is significant improvement in accuracy. Results for OLS are further discussed in Sec. 5.

**Numerical Study II.** Table 3 summarizes the results for the second numerical study in which the simulated data exhibit spurious correlations among model constituents. CR with appropriate constraints works much better than PLS or CR with poor constraints. This demonstrates that proper prior spectral information indeed improves the calibration results even when confounding effects such

Table 2. Comparison of different calibration methods with simulated data. Column 1 indicates the constraint used, columns 2 and 3 list the SEV and the SEP values obtained, with corresponding error reduction compared to PLS in parentheses, using the OLS (noise-added) and SEP (7.2) as the best achievable values. OLS (noiseless) indicates OLS in the absence of spectral noise; OLS (noise-added) denotes OLS with the same amount of spectral noise as for the PLS/CR data. Both OLS results include the same amount of concentration error as for PLS/CR data.

	Constraint ( $\mathbf{b}_0$ )	SEV (mg/dL)	SEP (mg/dL)
PLS	N/A	12.7	12.3
LN	N/A	12.4 (5%)	11.8 (10%)
CR	Collagen I	12.4 (5%)	11.8 (10%)
CR	Glucose	9.2 (64%)	8.9 (67%)
CR	$\mathbf{b}_{OLS}$	8.2 (82%)	8.2 (80%)
OLS (noise-added)	N/A	N/A	7.2
OLS (noiseless)	N/A	N/A	6.7

Table 3. Comparison of different calibration methods using simulated data with substantial spurious correlations. See text for details. Column designations are the same as for Table 2.

	Constraint ( $\mathbf{b}_0$ )	SEV(mg/dL)	SEP(mg/dL)
PLS	N/A	13.9	26.0
LN	N/A	14.0 (−1%)	25.7 (2%)
CR	Collagen I	14.0 (−1%)	25.8 (1%)
CR	Glucose	11.9 (30%)	16.3 (52%)
CR	$\mathbf{b}_{OLS}$	11.2 (40%)	13.0 (69%)

as spurious correlations exist. Since the  $\mathbf{b}$  vector of OLS is not derived from the calibration sample set, spurious correlations have no effect on it. Thus, the OLS results are identical to those of Table 2.

**Volunteer Study.** Results for the application of PLS and CR without constraint or with glucose as the constraint to 17 volunteers of the transcutaneous study are summarized in Table 4 which shows the SEV's of data from two representative volunteers and the average over all 17 volunteers. However, due to nonlinear effects such as differences in tissue composition, autofluorescence background, sampling volume, and physiological conditions, different regularization parameters were selected to minimize the SEV for each individual. This also reflects that better results can be obtained if the prior information is weighed more in some volunteers, while less in others. A potential future direction is to assess whether CR can better handle individuals with different conditions such as pre-diabetic or with skin issues. Averaging reduces variations among different individuals due to physiological and skin diversities, unmodeled in the numerical studies. CR with glucose as the constraint shows the most significant error reduction compared to the result of PLS. Figure 3 shows the glucose spectrum, the PLS  $\mathbf{b}$  vector, and the CR  $\mathbf{b}$  vector with the glucose spectrum as the constraint. Both PLS and CR  $\mathbf{b}$

Table 4. Comparison of different calibration methods with the volunteer data. Column 1 indicates the constraint used, columns 2 to 4 list the SEV values obtained from 2 representative volunteers and the average over all volunteers, with corresponding error reduction compared to PLS in parentheses. See text for details.

	Constraint ( $\mathbf{b}_0$ )	Volunteer A	Volunteer B	Average
PLS	N/A	14.7	13.3	13.2
LN	N/A	14.2 (7%)	10.3 (49%)	12.2 (17%)
CR	Glucose	13.1 (21%)	9.8 (57%)	11.8 (23%)

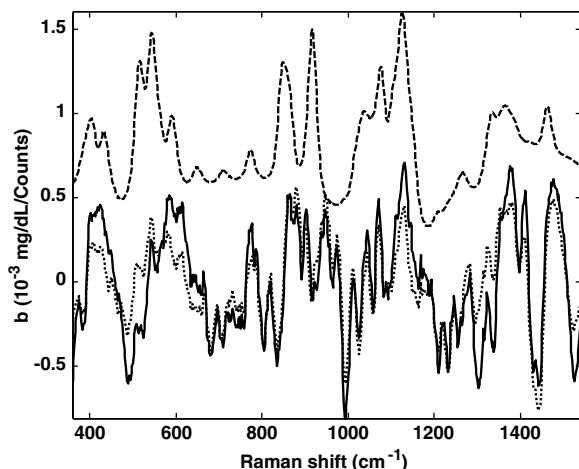


Fig. 3. Comparison of the glucose spectrum (dashed line) with the smoothed PLS  $\mathbf{b}$  vector (dotted line) and CR  $\mathbf{b}$  vector (solid line) for data from one volunteer.

vectors have been smoothed using a Savitzky–Golay 13-point filter to remove high-frequency fluctuations that are likely caused by noise.

## 5. Discussion

PLS and PCR are valuable tools in multivariate calibration. For spectral analysis of linear chemical systems, they allow flexible experimental design and are powerful in extracting correlations between spectra and analyte concentrations. To utilize prior information, hybrid methods such as CR allows the flexibility of incorporating spectroscopic information about the analyte of interest. In the broader context, regularization methods tend to perform slightly better than either PLS or PCR.<sup>24</sup> A heuristic explanation is that regularization provides a continuous “knob”, and therefore can be used to find a better balance between model complexity and noise rejection. Our results show that in addition to this intrinsic advantage, an even more significant improvement can be obtained by incorporating a meaningful solution constraint.

In Numerical Study I, we showed that CR with appropriate constraints outperforms PLS under “normal” conditions, i.e., no spurious correlations among constituents. In this case, spectroscopic noise and concentration error are the major factors determining prediction accuracy. In Numerical Study II, CR in the presence of spurious correlations was investigated. The results showed that CR with appropriate constraints gives rise to  $\mathbf{b}$  vectors that are

significantly less affected by such spurious correlations, and thus provide much better predictions than PLS. Further, compared to the scenario simulated in Numerical Study I in which no significant constituent co-variation was present, these results suggest that CR effectively reduces the detrimental influence due to spurious correlations.

Table 2 also lists the OLS results for constituents of different spectral quality. For OLS (noiseless), the  $\mathbf{b}$  vector was calculated from noiseless constituent spectra, whereas for OLS (noise-added),  $\mathbf{b}$  was derived from constituent spectra with SNR similar to that of the calibration sample set used by the implicit methods. Therefore, the results of Table 2 show that the performance of CR with a well-chosen constraint approaches that of OLS (noise-added) with equivalent SNR. The fact that CR cannot predict as well as OLS (noiseless) is not surprising because, unlike OLS (noiseless), the CR  $\mathbf{b}$  vectors are derived from noise-added data. The OLS (noiseless) result is considered the fundamental limit, given the SNR of the field spectra and the accuracy of reference concentration measurements. When the only error source is noise in the field spectra, the ultimate prediction error for OLS is inversely proportional to both the SNR and an overlap factor which quantifies spectral overlap among constituents. When other sources of error such as concentration error exist, the total prediction error can be estimated by the error terms in quadrature.<sup>25,26</sup> To closely simulate the volunteer data in the numerical studies, 6 mg/dL was used as the reference concentration error, and apparently it dominates the OLS (noiseless) prediction error, 6.7 mg/dL. However, such a limit can be further improved with more accurately measured reference concentrations.

In the numerical studies, the size of the calibration sample set (25) was chosen to match the average number of samples obtained from each individual in the volunteer study. This makes the sample-to-PLS-factor ratio  $\sim 3$ , smaller than the standard recommended value between 5 and 10.<sup>27</sup> To confirm that 9 is an appropriate number of factors for PLS, a numerical study with larger calibration sample size (50) was conducted. The SEP’s for PLS, CR with glucose as the constraint, and CR with  $\mathbf{b}_{\text{OLS}}$  as the constraint, are 8.9, 7.9 and 7.8 mg/dL, respectively. Since more samples in the calibration set effectively increases the SNR for the calibration and reduces the degree of ill-posedness, the larger sample set gives lower prediction errors for all methods. However, the

results suggest that CR with proper constraints can achieve the same quality calibration with fewer calibration samples as compared to PLS.

Finally, it appears that CR results in a much greater improvement over PLS when applied to the numerical studies than when applied to the volunteer study. One possible explanation is that the spectral constraints used for the volunteer study are obtained from clear samples, e.g., glucose in water or other individual pure constituents, whereas, transcutaneous Raman spectra are distorted by tissue turbidity. Therefore, the constraints employed in the volunteer study may not be the best choices. Incorporating tissue turbidity into constraints is currently under investigation in our laboratory. Another possible explanation is that the intense slowly varying background, not present in the spectral constraints, becomes a nonlinear confounding factor in the volunteer study. We are presently seeking a better way to model this background.

## 6. Conclusion

CR enjoys all advantages of implicit calibration while effectively avoids potential pitfalls such as spurious correlations among constituent concentrations. As has been demonstrated, there is flexibility in the choice of constraints, a convenient one being the spectrum of the analyte of interest itself. This flexibility is crucial because it is difficult, if not impossible, to quantify high-fidelity spectra and the relative “importance” of individual constituents in living tissues. We have shown that CR significantly outperforms PLS when analyzing Raman spectra from human subjects, as well as simulated data with or without spurious correlations among constituents. In the volunteer study, CR reduced the SEV by  $\sim 20\%$  on average compared to PLS. In the numerical study, the error reduction is  $\sim 73.5\%$  without spurious correlations, and  $\sim 61\%$  in the presence of spurious correlations.

Further, there is no particular condition that must be satisfied in order for CR to function (apart from system linearity), which suggests that it can serve as a general method suitable for any scenario in which implicit calibration is performed. Its application to transcutaneous measurement of glucose concentration using near infrared absorption reflectance/transmission spectroscopy is presently under investigation in our laboratory.

## Acknowledgments

WCS acknowledges funding from the National Science Foundation (NSF) CAREER Award (CBET-1151154), the National Aeronautics and Space Administration (NASA) Early Career Faculty Grant (NNX12AQ44G), Gulf of Mexico Research Initiative (GoMRI-030), and Cullen College of Engineering at the University of Houston. This work was partially performed at the MIT Laser Biomedical Research Center supported by the NIH National Center for Research Resources, Grant No. P41-RR02594. WCS thanks Prof. Michael Feld and Dr. Kate Bechtel for helpful discussions.

## References

1. H. Martens, T. Naes, *Multivariate Calibration*, Wiley (1992).
2. W.-C. Shih, K. Bechtel, M. Feld, M. Arnold, G. Small, *In Vivo Glucose Measurements*, John Wiley & Sons, Inc., pp. 331–356, (2009).
3. W. C. Shih, K. L. Bechtel, M. S. Feld, “Constrained regularization: Hybrid method for multivariate calibration,” *Anal. Chem.* **79**, 234–239 (2007).
4. O. R. Scepanovic, K. L. Bechtel, A. S. Haka, W. C. Shih, T. W. Koo, A. J. Berger, M. S. Feld, “Determination of uncertainty in parameters extracted from single spectroscopic measurements,” *J. Biomed. Optics* **12**, 064012 (2007).
5. K. L. Bechtel, W. C. Shih, M. S. Feld, “Intrinsic Raman spectroscopy for quantitative biological spectroscopy Part II: Experimental applications,” *Opt. Express* **16**, 12737–12745 (2008).
6. W. C. Shih, K. L. Bechtel, M. S. Feld, “Intrinsic Raman spectroscopy for quantitative biological spectroscopy Part I: Theory and simulations,” *Opt. Express* **16**, 12726–12736 (2008).
7. W. C. Shih, K. L. Bechtel, M. S. Feld, *Handbook of Optical Sensing of Glucose in Biological Fluids and Tissues*, CRC Press, pp. 353–386 (2008).
8. J. Qi, W.-C. Shih, “Parallel Raman microspectroscopy using programmable multipoint illumination,” *Opt. Lett.* **37**, 1289–1291 (2012).
9. J. Qi, P. Motwani, M. Gheewala, C. Brennan, J. C. Wolfe, W.-C. Shih, “Surface-enhanced Raman spectroscopy with monolithic nanoporous gold disk substrates,” *Nanoscale* **5**, 4105–4109 (2013).
10. M. Li, F. Zhao, J. Zeng, J. Qi, J. Lu, W.-C. Shih, “Microfluidic surface-enhanced Raman scattering sensor with monolithically integrated nanoporous gold disk arrays for rapid and label-free biomolecular detection,” *J. Biomed. Optics* **19**, 111611 (2014).

11. J. Qi, W. C. Shih, "Performance of line-scan Raman microscopy (LSRM) for high-throughput chemical imaging of cell population," *Appl. Optics* **53**, 2881–2885 (2014).
12. J. Zeng, J. Qi, F. Bai, J. C. Chung Yu, W.-C. Shih, "Analysis of ethyl and methyl centralite vibrational spectra for mapping organic gunshot residues," *Analyst* **139**, 4270–4278 (2014).
13. M. Bertero, P. Boccacci, *Introduction to Inverse Problems in Imaging*, Institute of Physics Pub., Bristol, UK; Philadelphia, PA (1998).
14. A. M. K. Enejder, T. G. Scecina, J. Oh, M. Hunter, W.-C. Shih, S. Sasic, G. Horowitz, M. S. Feld, "Raman spectroscopy for non-invasive glucose measurements," *J. Biomed. Optics* **10**, 031114 (2005).
15. C. V. Raman, K. S. Krishnan, *Nature* **121**, 501–502 (1928).
16. G. J. Puppels, F. F. M. Demul, C. Otto, J. Greve, M. Robertnicoud, D. J. Arndtjovin, T. M. Jovin, "Studying single living cells and chromosomes by confocal Raman microspectroscopy," *Nature* **347**, 301–303 (1990).
17. E. B. Hanlon, R. Manoharan, T. W. Koo, K. E. Shafer, J. T. Motz, M. Fitzmaurice, J. R. Kramer, I. Itzkan, R. R. Dasari, M. S. Feld, "Prospects for *in vivo* Raman spectroscopy," *Phys. Med. Biol.* **45**, R1–R59 (2000).
18. J. Qi, J. Li, W.-C. Shih, "High-speed hyperspectral Raman imaging for label-free compositional microanalysis," *Biomed. Opt. Express* **4**, 2376–2382 (2013).
19. M. Li, J. Lu, J. Qi, F. Zhao, J. Zeng, J. C.-C. Yu, W.-C. Shih, "Stamping surface-enhanced Raman spectroscopy for label-free, multiplexed, molecular sensing and imaging," *J. Biomed. Optics* **19**, 050501 (2014).
20. C.-H. Liu, J. Qi, J. Lu, S. Wang, C. Wu, W.-C. Shih, K. Larin, "Improvement of tissue analysis and classification using optical coherence tomography combined with Raman spectroscopy," *J. Innov. Opt. Health Sci.* 1550006 (2014).
21. N. Sudheendran, J. Qi, E. Young, A. Lazar, D. Lev, R. Pollock, K. Larin, W.-C. Shih, "Line-scan Raman microscopy complements optical coherence tomography for tumor boundary detection," *Laser Phys. Lett.* **11**, 105602 (2014).
22. A. C. Guyton, J. E. Hall, *Human Physiology and Mechanisms of Disease*, Saunders, Philadelphia (1997).
23. J. N. Roe, B. R. Smoller, "Bloodless glucose measurements," *Crit. Rev. Ther. Drug Carrier Syst.* **15**, 199–241 (1998).
24. I. E. Frank, J. H. Friedman, "A statistical view of some chemometrics regression tools," *Technometrics* **35**, 109–135 (1993).
25. A. Lorber, B. R. Kowalski, "Estimation of prediction error for multivariate calibration," *J. Chemom.* **2**, 93–109 (1988).
26. D. Qi, A. J. Berger, "Quantitative concentration measurements of creatine dissolved in water and urine using Raman spectroscopy and a liquid core optical fiber," *J. Biomed. Optics* **10**, 031115 (2005).
27. T. Hastie, R. Tibshirani, J. H. Friedman, *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*, Springer, New York (2001).