Validating the LASSO Algorithm by Unmixing Spectral Signatures in Multicolor Phantoms

Daniel V. Samarov[a], Matthew Clarke[b], Ji Yoon Lee[b], David Allen[b], Maritoni Litorja[b] and Jeeseong Hwang[b]

[a]National Institute of Standards and Technology, Information Technology Laboratory, Statistical Engineering Division, 100 Bureau Drive, Gaithersburg, MD 20899, USA;
[b]National Institute of Standards and Technology, Optical Technology Division, Biophysics Group, 100 Bureau Drive, Gaithersburg, MD 20899, USA

ABSTRACT

As hyperspectral imaging (HSI) sees increased implementation into the biological and medical fields it becomes increasingly important that the algorithms being used to analyze the corresponding output be validated. While certainly important under any circumstance, as this technology begins to see a transition from benchtop to bedside ensuring that the measurements being given to medical professionals are accurate and reproducible is critical. In order to address these issues work has been done in generating a collection of datasets which could act as a test bed for algorithms validation. Using a microarray spot printer a collection of three food color dyes, acid red 1 (AR), brilliant blue R (BBR) and erioglaucine (EG) are mixed together at different concentrations in varying proportions at different locations on a microarray chip. With the concentration and mixture proportions known at each location, using HSI an algorithm should in principle, based on estimates of abundances, be able to determine the concentrations and proportions of each dye at each location on the chip. These types of data are particularly important in the context of medical measurements as the resulting estimated abundances will be used to make critical decisions which can have a serious impact on an individual's health.

In this paper we present a novel algorithm for processing and analyzing HSI data based on the LASSO algorithm (similar to "basis pursuit"). The LASSO is a statistical method for simultaneously performing model estimation and variable selection. In the context of estimating abundances in an HSI scene these so called "sparse" representations provided by the LASSO are appropriate as not every pixel will be expected to contain every endmember. The algorithm we present takes the general framework of the LASSO algorithm a step further and incorporates the rich spatial information which is available in HSI to further improve the estimates of abundance. We show our algorithm's improvement over the standard LASSO using the dye mixture data as the test bed.

Keywords: Sparse regression, LASSO, SPLASSO, hyperspectral image analysis

1. INTRODUCTION

Hyperspectral imaging until recently has been a technology more commonly associated with remote sensing type applications to identify ground materials with unique optical reflectance and/or absorbance properties. The use of hyperspectral imaging in biomedical settings, while not new, has seen slower adoption into more widespread practice. The reasons for this include open instrument and algorithmic challenges associated with imaging biomedical scenes to identify chemical and molecular substances. Some common problems encountered include controlling/accounting for reflective glare from surfaces (e.g. from various regions on an organ), large background signals (e.g. from cell media), shading from three-dimensional contours of objects and accounting for unknown elements in an image. Each of these can, have a significant impact on the observed signal, making it difficult to make sense of or reliably use the collected measurements for any type of quantitative analysis.

Send correspondence to:
Daniel V. Samarov: E-mail: daniel.samarov@nist.gov, Telephone: 301-975-3534

Inspite of the various challenges the potential benefit of hyperspectral imaging technology in the clinic and laboratory is undeniable and has already has begun to see increased use in both settings.[1,2] For example, while still largely qualitative, surgeons at the University of Texas Southwestern have been implementing hyperspectral imaging technology in kidney related surgical procedures to monitor changes in hemoglobin oxygenation (referred to as oximetry). The system allows the surgeons to gauge, in real time the relative oxygen content of the kidney before the clamping of the renal artery, which is done in order to avoid excessive bleeding and related complications, to after clamping of the renal artery. Here the pre-clamp oxygen levels are used as a representative baseline for 100% oxygenation. Current standard of care dictates that the surgeon has about 30 minutes to complete his procedure. With real time oxymetric measurements they have a better gauge of exactly what that window is, as well as having insight into other potentially relevant physiological changes in the patient.

As hyperspectral imaging begins to see increased use in clinics and laboratories, it is important to understand the accuracy of the algorithms being used to analyze them. One possible way to measure algorithm performance is through the use of well characterized benchmark data sets, preferably where some type of ground truth exists. The characteristics of a data set that would be well suited for the types of applications we are interested in here are those that have similar scale in terms of object sizes, ranges of spectral features and spatial complexity.

Recently a novel microarray printing methodology was introduced[3] for testing of absorption and/or reflectance of microscopy measurements. The instrument (SpotBot2, ArrayIt, Sunnyvale, CA) uses printed dyes to enable multiplexed testing of the spectral capability of hyperspectral instruments. The unique optical signatures associated with chemical properties and their associated spectra are commonly referred to as *endmembers*. The dye printing platform is appealing for our applications as it mimics many of the properties and challenges of biomedical measurements, i.e. there is a spatial component, it contains shading and contours (doming effect associated with the printing of the dyes), mixtures of different components at different concentrations and a highly variable background. In this way the microarray printing methodology can act as a standard data set for validating the performance of hyperspectral image analysis algorithms. In the work presented here we focus on endmember *abundance fraction* estimation (i.e. dye concentration).

Figure 1 illustrates the layout of one of the microarray printing platforms involving the mixture of three different food coloring dyes. The three dyes are acid red 1 (AR), brilliant blue R (BBR) and erioglaucine (EG).[4] The image on the left is what the actual array looks like and the image on the right shows the layout of the locations, relative proportions and concentrations of each of the dyes. Dye samples are initially prepared in water. These stock solutions are then further diluted and mixed with 75% poly(ethylene glycol, MW = 600 kD) (PEG) for a final PEG concentration of 50% (v/v). Here, the three right-most columns with "dilutions" written above, correspond to individual dyes at 100% down to 5% concentration. The columns to the left of this show mixtures of these dyes at varying concentrations.

These dyes were selected because of their spectral distributions, shown in Figure 2. Here relative absorbance measurements (y-axis) were taken at 61 wavelengths between 400 nm and 700 nm (x-axis) with 5nm, full width of the half maximum bandwidth at every wavelength. While the dyes are spectrally distinct there is still considerable overlap between them in certain spectral regions. This overlap presents a challenge when trying to determine the concentration and proportion of a particular dye at a given spatial location where several dyes have been mixed.

Note, the algorithms used here for determing how many and which endmembers are present in the dye-mixture data are the hyperspectral signal identification by minimum error (HYSIME[6]) and the simplex identification via split augmented Lagrangian (SISAL[5]) algorithms, respectively. Here 18 endmembers were estimated by these algorithms. While both these aspects of hyperspectral image analysis are extremely important, they are outside the scope of the work presented here which is focused on estimating the abundances of the endmembers. From here on we take the endmembers to be known.

1.1 Results and Discussion

Next we present the main results using the LASSO (least absolute shrinkage and selection operator[7]) and the SPLASSO (Spatial LASSO[8]) to estimate the concentrations and proportions of the dyes from the three-dye mixture data set described above. The LASSO and SPLASSO are both known as "sparse" regression methods,
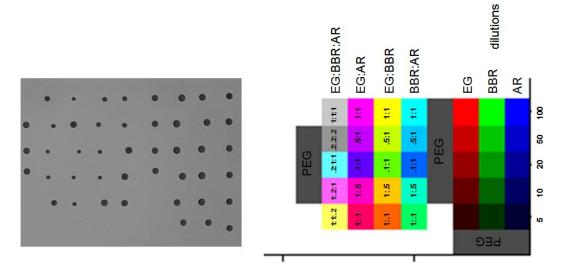
Figure 1. The design of the microarray printing platform for three dyes. The image on the left depicts the actual array and the image on the right shows the location, concentrations and proportion of each of the dyes.
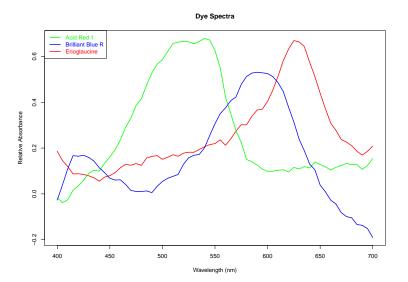

Figure 2. The spectral signatures of the three dyes used in this experiment. Signatures estimated using the SISAL algorithm.[5]

i.e. they have the property that some of the regression coefficients (i.e. abundances) are set to be exactly equal to 0, this property will be discussed in more detail in Section 2. Methods like these are particularly well suited to the task of abundance estimation in hyperspectral images as they exploit the fact that most pixels in a image are only composed of a subset of the total number of endmembers present (i.e. some will have 0 abundance).

Here algorithm performance is measured by looking at how well the ratio's from our estimates match up with the design in Figure 1. Figure 3 shows a heatmap of the estimated abundances using the LASSO and SPLASSO for each of the three dyes (we have zoomed in on the dye spots and removed most of the background regions to ease visualization of the results). The vertical white lines dilineate the different dye mixtures shown at the bottom and the color bar to the right of each figure shows the range of the estimated abundance fractions. Note that in each case the fractions are less than 1. There are a number of factors at play here which could cause this to happen. In particular the complexity of the background and other features, reflected in the 18 endmembers required to model the data, all have an effect on the measured absorbance.
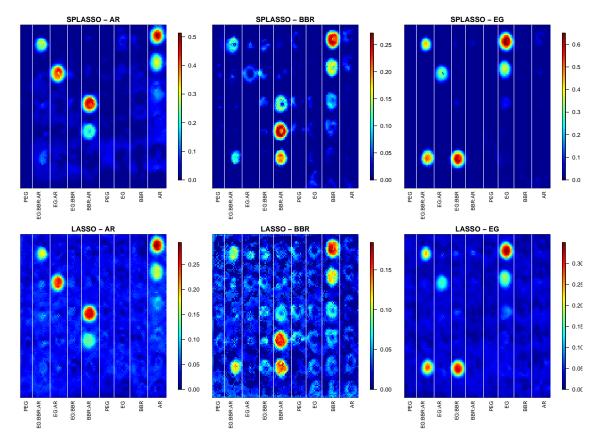

Figure 3. The abundance estimates of AR, BBR and EG. The color scale on the right indicates the estimated abundance fractions.

From a qualitative standpoint the SPLASSO algorithm produces a more visually appealing result as compared to the standard LASSO (less salt-and-pepper artifacts in the background). From a practical standpoint, this reduces the number of false-positive readings, saying a dye is present when it is not.

In order to estimate the concentration at each dye location we begin by estimating the 100% concentration values. This is done by calculating the average estimated abundance fractions from a 7×7 pixel region centered at the 100% concentration locations for AR, BBR and EG (see Figures 1 and 3); call these values AR_{100}, BBR_{100} and EG_{100}. Similar 7×7 regions are then selected and averages calculated at each spot; call these values AR_i, BBR_i and EG_i, where i denotes each spot location. The final estimated calibrated abundance fraction for each dye at each location is then calculated as
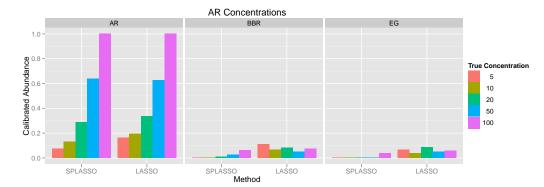
Figure 4. The concentrations in the AR dilution column. Each subplot displays the estimated calibrated abundance fractions within this column. Both methods are able to generally capture the dilution curve.
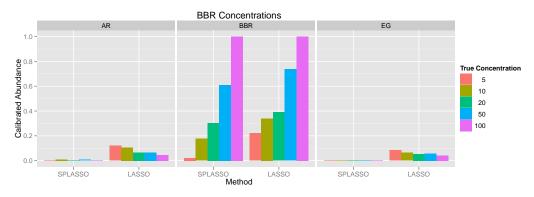

Figure 5. The concentrations in the BBR dilution column.

$$\frac{\text{AR}_i}{\text{AR}_{100}} \; , \; \frac{\text{BBR}_i}{\text{BBR}_{100}} \text{ and } \frac{\text{EG}_i}{\text{EG}_{100}}.$$

These results are shown in Figures 4, 5, 6 and 7. The first three figures display the results of the estimated calibrated abundance fractions for the serial dilutions of the pure dye mixtures (the three right most columns of Figure 1). Each of these are divided into three subplots corresponding to the three different dyes, with the x-axis in each seperated by the method used, LASSO or SPLASSO. For each method there is a barplot showing the estimated abundance fractions for each serial dilution, ranging from $5\%, \dots, 100\%$ (shown in the legend key to the right). In principle, only the subplot corresponding to the dye whose serial dilution we are currently estimating should have non-zero abundance fraction estimates.

So, for example Figure 4 shows the serial dilutions of AR, where we would expect that of the subplots only the one titled "AR" (on the left) would have non-zero abundance estimates. Of course in practice this is not the case, as is seen by the non-zero abundance fraction estimates in the BBR and EG subplots, however both methods are able to approximately capture the dilution curve. In general the SPLASSO produced fewer false positive estimates with respect to the other dyes.

EG appeared to be the most challenging dye to estimate for both methods (as can be seen by the difference between the estimated and true concentration values), in particular both the LASSO and SPLASSO were considerably less sensitive at the lower concentrations. Both approaches performed quite well on AR with BBR being a bit more challenging for the LASSO. As whole, for the pure dye mixtures both methods were able to do a good job of characterizing the serial dilutions, with the SPLASSO doing a much better job of producing fewer false positive readings.
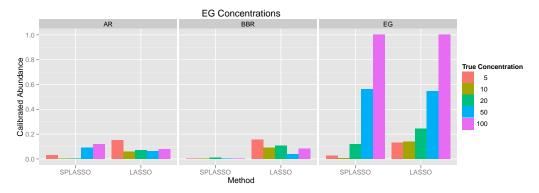
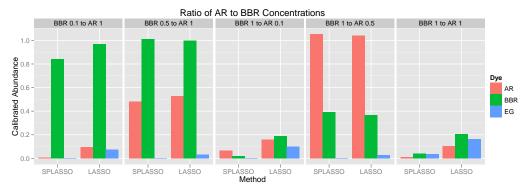Figure 6. The concentrations in the EG dilution column.

Figure 7. The concentrations in the BBR:AR dilution column.

Figure 7 shows the results from the mixture of AR and BBR at varying concentrations. In this plot each subfigure shows the estimated abundance fraction of each dye for each mixture (corresponding to the BBR:AR column in Figure 1). So, for example the left most plot displays the mixture of 10% AR with 100% BBR. Here the LASSO appears to do a better job over the SPLASSO. For both the 50% AR and 100% BBR and 100% AR and 50% BBR both methods do well in estimating the correct concentrations and proportions. However, for the cases where we have 100% AR and 10% BBR or 100% of both, both methods have difficulty estimating the correct concentrations and proportions.

There are several possible reasons for this, part of it may involve improvement in the image acquisition process, but also further development of pre-processing steps (e.g. background correction). Experimental errors may also play a role. For example, as the relative ratio of one dye increases, the unmixing and different surface wettability may introduce local heterogeneity within a single microarray spot and differential reflective scattering properties.

Within the scope of this work, we conclude that the SPLASSO algorithm is capable of extracting endmembers of two different chemical substances within the ratio range of 1:0.5 to 1:1. For the sake of space we have not included results for the remaining columns of mixtures, or replicate measurements (what we have shown is for one of three), however similar results hold.

2. METHODS

While not always the case in practice it is commonly assumed (as it is here) that at each pixel the spectral signature is a linear mixture of each of the endmembers present in the scene. Before providing a more formal description of linear mixing we begin by introducing some notation: define $\mathbf{y}_i = (y_{i1}, \ldots, y_{ip})^T$, $i = 1, \ldots, n$ to be the set of spectral response vectors, n corresonding to the total number of pixels in the image. Let $\mathbf{x}_j = (x_{1j}, \ldots, x_{pj})^T$, $j = 1, \ldots, m$ be the set endmembers (where each of the p entries maps to a specific

wavelength), which are collected in the matrix $\mathbf{X} = [\mathbf{x}_1, \ldots, \mathbf{x}_m]$. Finally let $\boldsymbol{\beta}_i = (\beta_{i1}, \ldots, \beta_{im})$, $i = 1, \ldots, n$ be the set of abundance vectors whose entries tell us the proportion and concentration of an endmember at a pixel. In order to ensure that these abundances have a physical meaning it is typically required that each element of $\boldsymbol{\beta}_i$ be nonnegative and that the sum of the elements of $\boldsymbol{\beta}_i$ are less than or equal to one. More generally

$$\mathbf{y}_i = \mathbf{X}\boldsymbol{\beta}_i, \text{ subject to }, \beta_{il} \geq 0 \text{ and } \sum_{l=1}^{m} \beta_{il} \leq 1. \tag{1}$$

As mentioned in the introduction "sparsity" in the abundance vectors $\boldsymbol{\beta}_i$ (i.e. possibly many β_{ij}'s being equal to 0) arises naturally in hyperspectral imaging as most pixels are typically composed of only a subset of the m endmembers. For example, in the the dye mixture data we know that many of the spots are made up of one or a mixture of two dyes. In some applications large dictionaries of endmembers specific to the types of objects being analyzed are available, with only a subset of the endmembers in the dictionary being present in the image at all. By explicitly taking into account the sparse nature of the endmember abundance vectors we are able to reduce the number of false positives (saying an endmember is present in a pixel when it is not) and therefore the accuracy of the estimation.

In Sections 2.1 and 2.2 we outline the LASSO and SPLASSO models respectively. Both of these approaches produce sparse estimates of the abundances and as illustrated in Section 1 produce very good results.

2.1 LASSO

Standard approaches to model building, such as ordinary least squares (OLS) do not produce sparse results and variable selection procedures traditionally used in conjunction with OLS, such as best subset selection, encounter difficulties when there are more than a few variables (as the number of possible combinations to consider quickly becomes intractable). Other shortcomings of subset selection methods are related to the discrete nature in which variables are added or removed from the model,[9].[10]

In order to effectively deal with these challenges regularization techniques which incorporate an l_1 penalty on the coefficient (abundance) vector, such as the LASSO[7] were developed. Through the introduction of the penalty term these methods are able to simultaneously perform prediction and variable selection. Sparse regression methods have been shown to be effective in practice across a wide range of applications. The form of the LASSO is quite similar to the linear mixing model described in (1) with the additional constraint that $|\boldsymbol{\beta}|_1 = \sum_{j=1}^{m} \beta_j \leq c$, for some constant c (where $|\cdot|_1$ is the l_1 norm). The loss function can then be expressed as

$$\hat{\boldsymbol{\beta}}_i(\text{LASSO}) = \arg\min_{\boldsymbol{\beta}_i} \left\| \mathbf{y}_i - \sum_{j=1}^{m} \mathbf{x}_j \beta_{ij} \right\|^2 + \lambda |\boldsymbol{\beta}_i|_1, \tag{2}$$

where λ is a nonnegative regularization parameter. The l_1 penalty term has the effect of continuously shrinking the coefficients toward 0 as λ increases and, for λ sufficiently large it can be shown that some coefficients are set exactly to 0. Extending this to the abundance estimation problem requires that the above estimation procedure be repeated for each i, $i = 1, \ldots, n$.

To gain some insight into how the LASSO is able to obtain estimates which are exactly 0, we consider the following special case. Suppose that the matrix of endmembers, $\mathbf{X}$ is orthonormal, i.e. $\mathbf{X}^T\mathbf{X} = \mathbf{I}$ and $\mathbf{I}$ is the identity matrix. Then it can be shown that the solution of the LASSO problem in (2) has the closed form solution

$$\hat{\beta}_{il}(\text{LASSO}) = \text{sgn}(\hat{\beta}_{il}(\text{OLS}))(|\hat{\beta}_{il}(\text{OLS})| - \lambda/2)_+, \ l = 1, \ldots, m \tag{3}$$

where $\hat{\beta}_{il}(\text{OLS}) = \mathbf{x}_l^T \mathbf{y}_i$ is the OLS estimate, and $(u)_+ = \max(0, u)$. Thus for $\lambda/2 \geq |\hat{\beta}_{il}(\text{OLS})|$, $\hat{\beta}_{il}(\text{LASSO}) = 0$.

For the more general case where we do not have orthonormality two fast and efficient algorithms have been proposed in the literature to solve (2): one based on least angle regression (LARS)[11] and a more recent adaptation based on the coordinate descent algorithm.[12] Both these methods provide a significant improvement in computational speed over standard linearly constrained, quadrattic programming approaches; in particular coordinate descent[12] has been shown to be very efficient for working with larger data sets. Straightforward modifications of either approach allow us to incorporate the positivity and sum to less than or equal to one constraints of linear mixing.

2.2 SPLASSO

One of the drawbacks of existing LASSO methods in the context of hyperspectral imaging is that they ignore the smoothly varying, spatial relationships between pixels and abundances. In order to effectively leverage this we introduce a spatial penalty term of the form $\sum_{j \in N(\mathbf{y}_i)} ||\boldsymbol{\beta}_i - \boldsymbol{\beta}_j||^2 w_{ij}$ into the LASSO objective (2) giving us the SPLASSO loss function

$$\hat{\boldsymbol{\beta}}_j(\text{SPLASSO}) = \arg\min_{\boldsymbol{\beta}_j} \sum_{i=1}^{n} ||\mathbf{y}_i - \mathbf{X}\boldsymbol{\beta}_i||^2 + \lambda_1 |\boldsymbol{\beta}_i|_1 + \lambda_2 \sum_{j \in N(\mathbf{y}_i)} ||\boldsymbol{\beta}_i - \boldsymbol{\beta}_j||^2 w_{ij}. \tag{4}$$

Here λ_1 and λ_2 are nonnegative regularization parameters, $N(\mathbf{y}_i)$ is the set of neighboring pixels about $\mathbf{y}_i$ and $w_{ij} \in [0, 1]$ is a spatial weight function capturing the similarity between observation i and its neighbors $j \in N(\mathbf{y}_i)$. The neighborhood defined by $N(\cdot)$ can take on a number of different forms; for our purposes we take $N = N_k$, the symmetric k-neighborhood on a regular 2D grid. To illustrate the form of N_k, suppose we are at grid point g_{rs} in a $M_1 \times M_2$ image, $1 \leq r \leq M_1$, $1 \leq s \leq M_2$. For $k = 1$ our neighborhood would be defined as the set of points $N_1 = \{g_{r-1,s}, g_{r+1,s}, g_{r,s-1}, g_{r,s+1}, g_{r-1,s+1}, g_{r+1,s+1}, g_{r-1,s-1}, g_{r+1,s-1}\}$.

The introduction of the penalty term $\sum_{j \in N_k(\mathbf{y}_i)} ||\boldsymbol{\beta}_i - \boldsymbol{\beta}_j||^2 w_{ij}$ in (4) has the effect of "encouraging" the $\boldsymbol{\beta}_i$'s to be similar to their k-neighbors, introducing a smoothness to the coefficient vectors. In hyperspectral unmixing this has several appealing aspects: in particular it allows our estimates to be more robust to instrument and sample variability. Intuitively this makes sense as the variability introduced from these different sources will tend to be smoothed out. Of course, as in any smoothing method, care needs to be taken to avoid removing actual features by oversmoothing.

For this reason appropriate selection of the weights w_{ij} and regularization parameters λ_1 and λ_2 is extremely important. In the application of the SPLASSO to hyperspectral imaging it is desirable to have a weight function which uses both spatial and spectral information. Let us suppose that the spectral signature, $\mathbf{y}_i$ whose abundances we are estimating corresponds to the rs^{th} pixel in the image (for illustrative purpose we refer to this point as $\mathbf{y}_{rs}$). The spatial component of the weight function can then be captured by

$$b_{rs}(lm) = \begin{cases} \frac{1}{(r-l)^2 + (s-m)^2}, l \in [r-k, r+k], m \in [s-k, s+k] & \text{if } (l, m) \notin (r, s), \\ 0 & \text{otherwise.} \end{cases} \tag{5}$$

Our decision to use (5) is that it provides a decrease in the effect a neighboring pixel has the further we move out from the current observation being estimated. However, the decrease is not so rapid as to make the contribution of the surrounding observations negligible. Next, to leverage spectral information we use the weights

$$c_{rs}(lm) = \frac{\mathbf{y}_{rs}^T \mathbf{y}_{lm}}{||\mathbf{y}_{rs}|| ||\mathbf{y}_{lm}||}, l \in [r-k, r+k], m \in [s-k, s+k], \tag{6}$$

which is the cosine of the angle between the spectra. This is a similarity measure commonly used in hyperspectral image analysis applications. For our purposes it is appealing because it allows our spatial weight function to be

adaptive to local features in the image, e.g. if we are at the edge of an object. We have also found it useful in practice to include a threshold on the angle between spectra, so that if $\mathrm{acos}(c_{rs}(lm)) > t$, $t \in [0, \pi]$ then $c_{rs}(lm) = 0$. Putting the spatial (5) and spectral (6) weights together the weight function is defined as

$$w_{rs}(lm) = b_{rs}(lm)c_{rs}(lm).$$

To gain insight into the properties of (4) and the role of the regularization parameter λ_2 we once again considering the case where $\mathbf{X}$ is taken to be orthonormal. Let $\gamma = 1/(1 + \lambda_2)$, $\sum_{j \in N_k(\mathbf{y}_i)} w_{ij} = 1$, (note, the latter does not need to hold in general, we do so here for illustrative purposes), $\alpha_{i,l} = \sum_{j \in N_k(\mathbf{y}_i)} \beta_{j,l} w_{ij}$ and

$$\hat{b}_{i,l} = \gamma \hat{\beta}_{i,l}(\mathrm{OLS}) + (1 - \gamma)\alpha_{i,l}$$

then it can be shown that

$$\hat{\beta}_{i,l}(\mathrm{SPLASSO}) = \mathrm{sgn}(\hat{b}_{i,l})\left(|\hat{b}_{i,l}| - \frac{\lambda_1}{2}\gamma\right)_+. \tag{7}$$

Looking at (7) we can see that it is quite similar to (3) except that now it the parameter γ controls the tradeoff between the OLS estimate and a smoothly weighted average of its neighboring pixels.

Similar approaches to solving the LASSO can also be applied to solving the SPLASSO; for details see.[8]

REFERENCES

[1] Zuzak, K., Francis, R., Wehner, E., Litorja, M., Cadeddu, J., and Livingston, E., "Active dlp hyperspectral illumination: A noninvasive, in vivo, system characterization visualizing tissue oxygenation at near video rates," *Analytical Chemistry* **83**(19), 7424–7430 (2011).

[2] Lee, J., Clarke, M., Tokumasu, F., Lesoine, J., Allen, D., Chang, R., Litorja, M., and Hwang, J., "Absorption-based hyperspectral imaging and analysis of single erythrocytes," *IEEE Journal of Selected Topics in Quantum Electronics* , 10 (2011).

[3] Clarke, M., Allen, D., Samarov, D., and Hwang, J., "Characterization of hyperspectral imaging and analysis via microarray printing of dyes," *Proceedings of SPIE* **7891** (2011).

[4] Green, F., [*The Sigma-Aldrich Handbook of Stains, Dyes and Indicators*], Aldrich Chem Co Library (1990).

[5] Bioucas-Dias, J., "A variable splitting augmented lagrangian approach to linear spectral unmixing," *First IEEE GRSS Workshop on Hyperspectral Image and Signal Processing-WHISPERS* (2009).

[6] Bioucas-Dias, J. and Nascimento, J., "Hyperspectral subspace identification," *IEEE Transactions on Geoscience and Remote Sensing* **46**(8), 2435–2445 (2008).

[7] Tibshirani, R., "Regression shrinkage and selection via the lasso," *Journal of the Royal Statistical Society (Series B)* **58**, 267–288 (1996).

[8] Samarov, D., Hwang, J., Lee, J., and Clarke, M., "The spatial lasso with applications to unmixing hyperspectral images," Working paper (2011).

[9] Breiman, L., "Better subset rergression using the nonnegative garotte," *Technometrics* **37**, 373–384 (1995).

[10] Fan, J. and Li, R., "Variable selection via nonconcave penalized likelihood and its oracle properties," *Journal of the American Statistical Association* **96**, 1348–1360 (2001).

[11] Efron, B., Hastie, T., Johnstone, I., and Tibshirani, R., "Least angle regression," *Annals of Statistics* **32**(2), 407–499 (2004).

[12] Friedman, J., Hastie, T., Hofling, H., and Tibshirani, R., "Pathwise coordinate optimization," *The Annals of Applied Statistics* **1**(2), 302–332 (2007).