HYPERSPECTRAL UNMIXING BY BREGMAN SPLITTING CONSTRAINT ENFORCEMENT

RUSSELL WARREN[1] AND STANLEY OSHER[2]

Abstract. We have developed a method for hyperspectral (HS) image data unmixing that requires neither pure pixels nor any prior information about the image other than an estimate of the number of different materials present. Based on well-established techniques from nonlinear optimization theory, the algorithm estimates the spectral and spatial structure in the image through a single noise-tolerant algorithm, removing the need for separate endmember and spatial abundance estimation steps. The algorithm is illustrated on data collected by the SpecTIR HS imaging sensor.

1. Introduction. Hyperspectral imaging is an important technology for many commercial and military applications such as natural resource discovery and intelligence gathering. Images are collected having typically many thousands of discrete pixels, each with a wavelength spectrum of up to hundreds of spectral bands. Usually, the images contain many different topographical features such as soil, vegetation, and water in addition to man-made features such as roads, buildings, etc. Those features all have distinctive spectral properties that must be separated by processing algorithms, ideally using as little prior knowledge as possible. The term *unmixing* refers to the need for performing this separation.

Because many pixels will have wavelength spectra that are a composite of two or more basic components in the image, algorithms are needed for automatically performing the unmixing. The usual method for doing this unmixing is to locate the *endmember* pixels that have a pure spectrum from a single material. If those pure pixels exist and can be identified, least-squares methods can be used to find the image spatial regions associated with each material. Standard simplex-based algorithms such as N-findr [1] and VCA [2] work reasonably well if pure pixels exist under high SNR conditions.

The problem with standard endmember selection methods is that pure pixels may not exist due to inadequate spatial resolution of the imaging sensor, or noise that can lead to poor spectral estimates at a single pixel. Although approaches exist that do not require the pure pixel assumption, they typically either require training data for supervised learning of the material spectra, or the use of large *dictionaries* of possible spectral signatures. The latter algorithms use compressive sensing methods for selecting a sparse subset of spectra from the dictionary that adequately captures the spectral content of the image. Both methods require the use of additional information about the image that may not exist in practice. Other methods for relaxing the pure pixel assumption perform

[1] EO-Stat Inc. of North Carolina, 10010 Vail Dr., Chapel Hill, NC. 27517 (eostatinc@aol.com).
[2] Department of Mathematics, University of California, Los Angeles, CA. 90095 (sjo@math.ucla.edu).

endmember estimation by data transformations [3] or nonnegative matrix factorization [4] applied to simplex models for the data mixing model. Those models constrain the so-called *abundances* to be positive values summing to 1 over the number of endmembers at each pixel. Because of this constraint, simplex-based methods require an additional, nontrivial processing step to estimate the actual spatial concentrations for each endmember component once the endmember spectra are found. Our approach does both spectral and spatial estimations in one step.

The theoretical basis for the unmixing algorithm is an iterative approach that alternates between estimating the spatial distribution and spectral components of each assumed material while holding the other component fixed. Because this is an ill-posed problem without a unique solution in general, special care is needed in enforcing constraints on the spatial and spectral variables at each iteration of the algorithm to produce useful results. The constraints are positivity on the spatial and spectral estimates, and a unit-vector constraint on the spectrum of each material. For enforcing these constraints we use results from the recent paper by Lai and Osher [5] that introduces auxiliary variables to split the constraint enforcement from the estimation of concentration and spectra by Bregman iteration. The combination of variable splitting--to analytically impose positivity and unit-norm (in the case of the spectra) constraints--with Bregman iteration--to solve the unconstrained concentration and spectral estimation problems--is termed SOC (splitting orthogonality constraints) by the authors.

It is slightly more convenient to derive the spectral and spatial components of the unmixing algorithm using the augmented Lagrangian (AL) method [6, 7] rather than Bregman iteration. The AL algorithm is entirely equivalent to Bregman iteration; the latter is derived by replacing the Lagrangian constraint terms with Bregman distance functions.

One of the complications in processing hyperspectral imaging data is the sheer size of the data cubes. One test data set used to illustrate our approach has 192,000 pixels each having 360 spectral bands. Processing these data by an iterative algorithm is a challenge for a PC. We have found that substantial computation-time saving results from using uniform spatial subsampling to do the unmixing with little or no degradation to the spectral estimates. The spatial structure at full resolution can then be found by processing the full data cube with the estimated spectra.

The remainder of the paper is organized as follows. In section 2 we derive the HS unmixing algorithm using an augmented Lagrangian with the SOC method used for constraint enforcement. Section 3 discusses the spatial subsampling method. In section 4 the unmixing with and without subsampling is illustrated on HS test data collected by the SpecTIR sensor over Beltsville, MD. We summarize and discuss possible generalizations of the approach in section 5.

2. Hyperspectral Unmixing by SOC Constraint Enforcement. Given the HS image data $G \in \mathbf{R}^{M \times N}$ where M denotes the number of spectral bands, and $N \equiv N_1 N_2$ represents the number of pixels after stacking the N_1 rows and N_2 columns into a single

vector, we model G as $G = \rho C + n$ with $\rho \in \mathbf{R}_+^{M \times L}$ and $C \in \mathbf{R}_+^{L \times N}$ with L the number of assumed materials in the image, and n is an additive noise with bounded variance. We then have the constrained problem

$$\min_{\rho, C} \frac{1}{2} \left\| G - \rho C \right\|_F^2 \text{ s. t. } \rho \geq 0, \ \left\| \rho_l \right\| = 1, \ 1 \leq l \leq L, \ C \geq 0 . \tag{2.1}$$

Because of the coupling between ρ and C we have a nonconvex problem having in general many solutions, so the existence and variability of solutions to this problem starting with the initial estimate ρ^0 become central concerns. In particular, the method used to enforce the constraints is crucial to finding consistent solutions to (2.1) using different spectral initializations.

Analogous to the use of splitting for basis pursuit or other constrained convex optimizations, the SOC method introduces additional parameters into the objective function that allow a difficult constrained optimization to be broken into two simpler problems that often have either analytical or easily implemented iterative solutions; it is in effect a divide-and-conquer strategy.

For the HS unmixing problem we introduce two sets of parameters, one set $\rho = r$ for the spectra, and another $C = e$ for the spatial concentration. Letting ρ^k, r^k, C^k, e^k denote the parameter estimates at iteration k, we have the problem

$$\min_{\rho, r, C, e} J(C, \rho) \text{ s. t. } \rho = r, \ r \geq 0, \ \left\| r_l \right\| = 1, \ C = e, \ e \geq 0, \tag{2.2}$$

with $J(C, \rho) = \frac{1}{2} \left\| G - \rho C \right\|_F^2$, where the subscript F denotes the Frobenius matrix norm defined as $\left\| A \right\|_F^2 \equiv \text{tr}\left(AA^T \right)$.

The augmented Lagrangian for this problem is then

$$L(C, e, p, \rho, r, q) = J(C, \rho) + \langle p, e - C \rangle + \frac{\lambda_C}{2} \left\| e - C \right\|_F^2 + \langle q, r - \rho \rangle + \frac{\lambda_\rho}{2} \left\| r - \rho \right\|_F^2 , \tag{2.3}$$

with dual Lagrange multiplier parameters p and q that enforce the equality constraints. The third and fifth terms on the right represent the quadratic terms that augment the classical Lagrangian. Their addition makes the problem locally a saddle point problem, and provides a systematic method for locally optimizing both the multipliers and $\{\rho, C\}$ in parallel. By virtue of the additional parameters e and r we have unconstrained optimizations over C and ρ with the constraints easily enforced on e and r analytically.

From the structure of (2.3) we see that the total problem at a given iteration k breaks into two subproblems: finding $\left(C^k, e^k, p^k\right)$ given ρ^{k-1}, and finding $\left(\rho^k, r^k, q^k\right)$ given C^k. For the first subproblem we have

$$\left(C^k, e^k, p^k\right) = \arg\max_{p} \min_{C, e \geq 0} L\left(C, e, p, \rho^{k-1}, r^{k-1}, q^{k-1}\right). \tag{2.4}$$

From

$$\nabla_C L = -\rho^T\left(G - \rho C\right) - p + \lambda_C\left(C - e\right)\Big|_{\rho=\rho^{k-1}, p=p^{k-1}, e=e^{k-1}} = 0, \tag{2.5}$$

we find

$$\boxed{C^k = \left(\rho^{(k-1)T}\rho^{(k-1)} + \lambda_C I\right)^{-1}\left(\rho^{(k-1)T}G + p^{k-1} + \lambda_C e^{k-1}\right)}. \tag{2.6}$$

Similarly, $\nabla_e L = p^{k-1} - \lambda_C\left(C^k - e\right) = 0$ gives the projection onto the positive halfspace

$$\boxed{e^k = \left[C^k - p^{k-1}/\lambda_C\right]^+ \equiv \max\left(C^k - p^{k-1}/\lambda_C, 0\right)}. \tag{2.7}$$

From $p^k = \nabla_C J\left(C, \rho\right)\Big|_{C=C^k, \rho=\rho^{k-1}}$ and using (2.6) we get

$$\boxed{p^k = p^{k-1} - \lambda_C\left(C^k - e^k\right)}. \tag{2.8}$$

The second subproblem assumes the form

$$\left(\rho^k, r^k, q^k\right) = \arg\max_{q} \min_{\substack{\rho, r \geq 0 \\ \|r_n\|=1, 1 \leq n \leq L}} L\left(C^k, e^k, p^k, \rho, r, q\right). \tag{2.9}$$

Differentiating L with respect to ρ gives

$$\nabla_\rho L = -\left(G - \rho C\right)C^T - q + \lambda_\rho\left(\rho - r\right)\Big|_{C=C^k, q=q^{k-1}, r=r^{k-1}} = 0, \tag{2.10}$$

and the result

$$\boxed{\rho^k = \left(GC^{kT} + q^{k-1} + \lambda_\rho r^{k-1}\right)\left(C^k C^{kT} + \lambda_\rho I\right)^{-1}}. \tag{2.11}$$

From $\nabla_r L = q^{k-1} - \lambda_\rho\left(\rho^k - r\right) = 0$ and the positive-orthant sphere constraint we get

$$\boxed{r^k = \left[\rho^k - q^{k-1}/\lambda_\rho\right]^+ \Big/ \left\|\left[\rho^k - q^{k-1}/\lambda_\rho\right]^+\right\|_F} \tag{2.12}$$

Finally, from $q^k = \nabla_\rho J\left(C,\rho\right)\Big|_{\rho=\rho^k, C=C^k}$ and (2.11) we find

$$\boxed{q^k = q^{k-1} - \lambda_\rho\left(\rho^k - r^k\right)}. \tag{2.13}$$

Both subproblem recursions for C by (2.6)-(2.8) and ρ by (2.11)-(2.13) are iterated to convergence (with r, e, p, and q initialized at 0) within an outer loop that iterates between the subproblems. The spectral estimates ρ^0 are initialized by positive, unit-norm random vectors. Also, the spatial subsampling discussed in section 3 is used for estimating the low-resolution C and spectra followed by a final iteration of (2.6)-(2.8) for estimating C at full resolution. A summary of the overall algorithm in given in Box 1.

We make the following remarks about the algorithm in Box 1. The HS data cube is denoted by the three-dimensional array $R\left(j,k_1,k_2\right)$ where $1 \le j \le M$,$1 \le k_1 \le N_1$, and $1 \le k_2 \le N_2$. For the unmixing, R is subsampled by every N_{samp} pixels to get G after stacking the spatial columns into a single row vector. This operation is denoted by $\langle\ \rangle$. The last part of the algorithm estimates C at full resolution using the spectral estimates from the unmixing taking G to be the full data cube R after spatial stacking.

Formally, the recursions by iterated Bregman and augmented Lagrangians are related [8] by $b_C^k = -p^k/\lambda_C$ and $b_r^k = -q^k/\lambda_\rho$, where b_C^k and b_r^k are the gradients (subdifferentials, in general) appearing in the Bregman distance functions.

3. Spatial Subsampling. The hyperspectral unmixing algorithm derived above is necessarily iterative. As noted above, hyperspectral data cubes such as those produced by the SpecTIR sensor have typically about 200,000 pixels with hundreds (360 for SpecTIR) of spectral bands. Processing data cubes this large over one hundred or so iterations is not realistic for routine application. Because of the enormous amount of information present in these data cubes, one might hope that using a relatively small subset of the data could yield good spectral estimates. Those spectral estimates could then be used to generate the spatial components at full resolution much more efficiently than using all the data at once. Obviously, if one is looking for the spectral structure and location of small objects, this method will fail. But in that case a better approach would be to first estimate and remove the overall background structure by the method proposed here, and then look for the small objects in the fitting residuals without subsampling.

This section and the processing results in section 4 suggest that it is indeed possible to use only 1% of the spatial data to do the unmixing without compromising the quality of the spectral estimates. This section provides some theoretical justification for using spatial subsampling on huge data sets. We first look at the relation between the spectra and concentration from the AL algorithm derived above near convergence, and

Input: N_1, N_2, M, N_{samp}, R, λ_C, λ_ρ, ε, $\rho^{(0)}$, L, tol

Initialize: $C^{(0)} = e^{(0)} = p^{(0)} = 0,\ r^{(0)} = q^{(0)} = 0,\ \bar{\rho}^{(0)} = 0$

$G = \left\langle R\left(1\!:\!M, 1\!:\!N_{samp}\!:\!N_1, 1\!:\!N_{samp}\!:\!N_2\right)\right\rangle$

$D\rho = 1,\ k = 1$

while $D\rho > \varepsilon$

$\delta = 1$

while $\delta >$ tol

$$C^{(k)} = \left(\rho^{(k-1)T}\rho^{(k-1)} + \lambda_C I_L\right)^{-1}\left(\rho^{(k-1)T}G + p^{(k-1)} + \lambda_C e^{(k-1)}\right)$$

$$e^{(k)} = \left[C^{(k)} - p^{(k-1)}/\lambda_C\right]^+$$

$$p^{(k)} = p^{(k-1)} - \lambda_C\left(C^{(k)} - e^{(k)}\right)$$

$$\delta = \left\|C^{(k)} - C^{(k-1)}\right\|_F$$

end

$\delta = 1$

while $\delta >$ tol

$$\rho^{(k)} = \left(GC^{(k)T} + q^{(k-1)} + \lambda_\rho r^{(k-1)}\right)\left(C^{(k)}C^{(k)T} + \lambda_\rho I_L\right)^{-1}$$

$$r^{(k)} = \left[\rho^{(k)} - q^{(k-1)}/\lambda_\rho\right]^+ \Big/ \left\|\left[\rho^{(k)} - q^{(k-1)}/\lambda_\rho\right]^+\right\|_F$$

$$q^{(k)} = q^{(k-1)} - \lambda_\rho\left(\rho^{(k)} - r^{(k)}\right)$$

$$\delta = \left\|\rho^{(k)} - \rho^{(k-1)}\right\|_F$$

end

$\bar{\rho}^{(k)} = \rho^{(k)},\ D\rho = \left\|\bar{\rho}^{(k)} - \bar{\rho}^{(k-1)}\right\|_F$

$k \to k+1$

end

$G = \langle R \rangle$

$\delta = 1,\ n = 1,\ C^{(0)} = \bar{e}^{(0)} = \bar{p}^{(0)} = 0$

while $\delta >$ tol

$$C^{(n)} = \left(\bar{\rho}^{(k)T}\bar{\rho}^{(k)} + \lambda_C I_L\right)^{-1}\left(\bar{\rho}^{(k)T}G + \bar{p}^{(n-1)} + \lambda_C \bar{e}^{(n-1)}\right)$$

$$\bar{e}^{(n)} = \left[C^{(n)} - \bar{p}^{(n-1)}/\lambda_C\right]^+$$

$$\bar{p}^{(n)} = \bar{p}^{(n-1)} - \lambda_C\left(C^{(n)} - \bar{e}^{(n)}\right)$$

$$\delta = \left\|C^{(n)} - C^{(n-1)}\right\|_F,\quad n \to n+1$$

end

Box 1 SOC algorithm for hyperspectral unmixing.

then consider a statistical argument based on the Cramer-Rao lower bound on the covariance of the spectral estimates.

We first note that the estimate of the $M \times L$ spectral array ρ^k at the k-th iteration is given by (2.11)

$$\rho^k = \left(GC^{kT} + q^{k-1} + \lambda_\rho r^{k-1}\right)\left(C^k C^{kT} + \lambda_\rho I\right)^{-1} \tag{3.1}$$

in terms of the HS data *G* (after stacking), $L \times N$ spatial concentration array C^k, Lagrange dual variables q^{k-1}, SOC splitting variables r^{k-1}, and tuning parameter λ_ρ. Near convergence to a local minimum q^{k-1} is approximately given by the *r*-update equation

$$\nabla_r L = q^{k-1} - \lambda_\rho \left(\rho^k - r\right)\Big|_{r=r^{k-1}} \; ; \; 0 \tag{3.2}$$

as $q^{k-1} = \lambda_\rho \left(\rho^k - r^{k-1}\right)$. Substitution into

$$\rho^k \left(C^k C^{kT} + \lambda_\rho I\right) = GC^{kT} + q^{k-1} + \lambda_\rho r^{k-1} \tag{3.3}$$

gives the simple result

$$\rho^k C^k C^{kT} = GC^{kT} \quad \text{(at convergence)} . \tag{3.4}$$

We see that near convergence the spectral estimates are dependent on the data *G* and the concentration through the array $C^k C^{kT}$.

We next note that $C^k C^{kT}$ is, up to a constant factor, the Fisher information matrix for ρ at the *k*-th iteration. This can be seen from the hyperspectral unmixing model

$$G = \rho C + n \tag{3.5}$$

where the additive noise array *n* is assumed to be approximately multivariate normal, independent, zero-mean randomly distributed with variance σ^2, leading to the density for ρ conditioned on C^k:

$$f\left(\rho \middle| C^k\right) = \frac{1}{\left(2\pi\sigma^2\right)^{MN/2}} \exp\left[-\frac{1}{2\sigma^2}\left\|G - \rho C^k\right\|_F^2\right] \tag{3.6}$$

Following the treatment in Keener [9], for example, the Fisher information matrix for ρ conditioned on C^k is

$$I\left(\rho \middle| C^k\right) = -E\left[\nabla_\rho^2 \ln f\left(\rho \middle| C^k\right)\right] \tag{3.7}$$

where E denotes expectation and ∇_ρ^2 is the Hessian matrix of second order derivatives with respect to ρ. Substituting (3.6) then gives

$$\boxed{I\left(\rho\middle|C^k\right)=\frac{1}{\sigma^2}C^kC^{kT}} \tag{3.8}$$

and the conditional covariance given by the Cramer-Rao lower bound for unbiased estimates of ρ becomes

$$Cov\left(\rho\middle|C^k\right)\geq\left[I\left(\rho\middle|C^k\right)\right]^{-1}=\sigma^2\left(C^kC^{kT}\right)^{-1}. \tag{3.9}$$

We see that C^kC^{kT} or rather its normalized matrix norm $\left\|C^k/\sqrt{N}\right\|_F^2=\mathrm{tr}\left(C^kC^{kT}/N\right)$ is the relevant scalar quantity for comparing the dependence of estimates of ρ on spatial subsampling.

We compared $\left\|C^k/\sqrt{N}\right\|_F$ at convergence from running the unmixing algorithm for 50 trials on the data cube Scene 8 from the online site [10]. These data have $N=700\times608$ pixels and M = 31 spectral bands within the visible spectral regime. Figure 1 shows an RGB image of the data from the Manchester site.

Figure 1 RGB image of Scene 8.

Figure 2 shows the processing of the data for 6 materials comparing subsampling at every 10th pixel in both row and column with the estimates at full resolution. Random unit vector initializations were used for each trial. The sample mean of the ratios $\left\|C_{sub}^k/\sqrt{N_{sub}}\right\|_F\Big/\left\|C_{full}^k/\sqrt{N_{full}}\right\|_F$ was 0.9980. Here N_{sub} and N_{full} are the total number of image pixels for the subsampled C_{sub} and full C_{full} images, respectively. Evidently, for

subsampling that uses as little as 1% of the spatial information we expect quite similar spatial norms, and therefore similar spectral estimates.

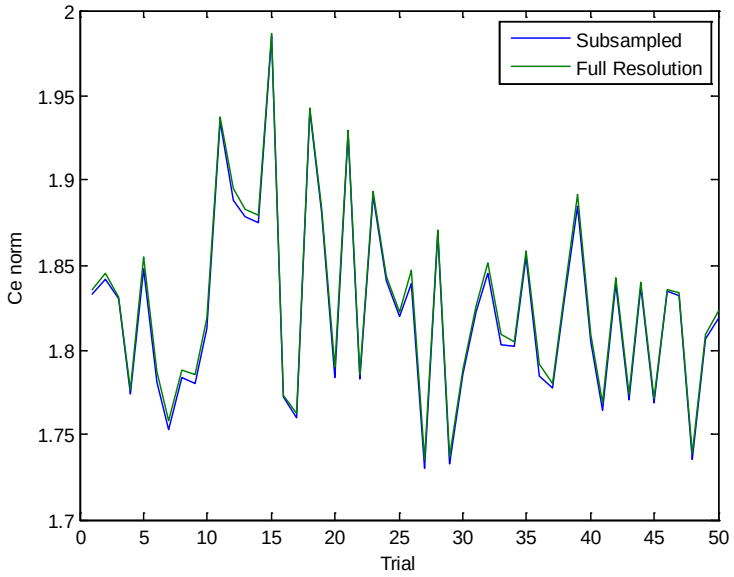

Figure 2 Comparison of the estimated concentration norms for subsampled and full resolution data.

4. Application to SpecTIR Data. Among the HS data analyzed using this method, the SpecTIR HS data available through the site www.spectir.com have been particularly useful. These data are well-suited for remote sensing algorithm development since they have typically 360 spectral bands over the range of 400-2500 nm, and have been collected from a high altitude platform. We show the analysis of one of these data sets: the agricultural and vegetation radiance data collected over Beltsville, MD. Figure 3 shows an image of a subset of the Beltsville data cube having 600 by 320 pixels with 360 bands with approximately 5 nm spectral resolution. The figure was taken from the SpecTIR Website. The data were reduced in amplitude by a factor of 5000 to match the dynamic range of our unmixing algorithm using the regularization parameters $\lambda_C = 0.1$ and $\lambda_\rho = 300$.

Implementation of the SOC HS unmixing algorithm in Matlab gave the spectral estimates, assuming 6 materials, shown in Figure 4. We used random unit vector initialization and spatial subsampling with every 10th pixel in both row and column. The spectra are labeled by color in the usual Matlab order: blue, green, red, cyan, magenta, and yellow, corresponding to spectra 1-6. The algorithm converged in 77 iterations, and required 4.83 s using the subsampled data. We note the nulls in the spectra around 1500 nm and 1900 nm that are due primarily to water vapor absorption in the atmosphere. Figure 5 shows the subsampled spatial distribution of the six materials. The spatial concentrations at full resolution are given in Figure 6. The full-resolution spatial calculation from the spectra shown in Figure 4 took an additional 13.86 s.

For comparison, Figures 7 and 8 show the spectral and spatial estimates from the SOC unmixing algorithm with no subsampling. There was convergence in 69 iterations.

As expected, there was not a significant difference in the results compared to Figures 4 and 6 except for the computation time of 901 s (15 m).

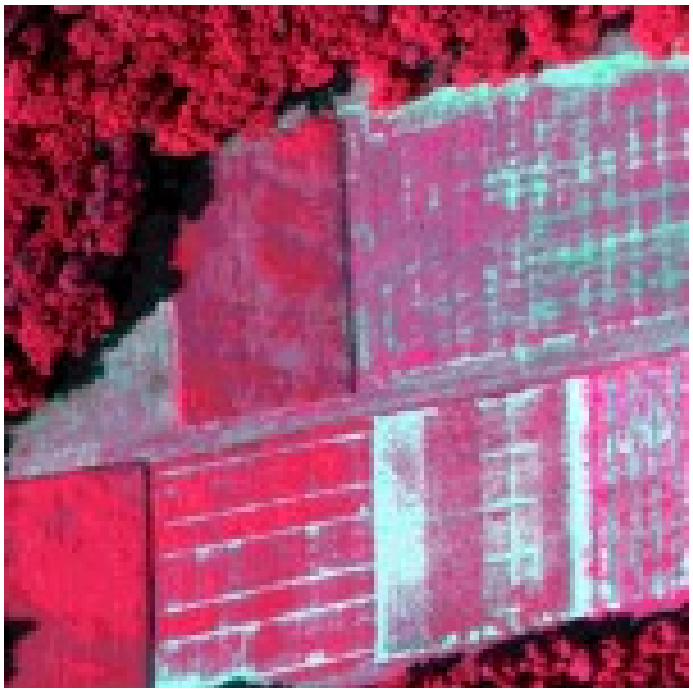

Figure 3 Beltsville, MD radiance data from the SpecTIR Website.

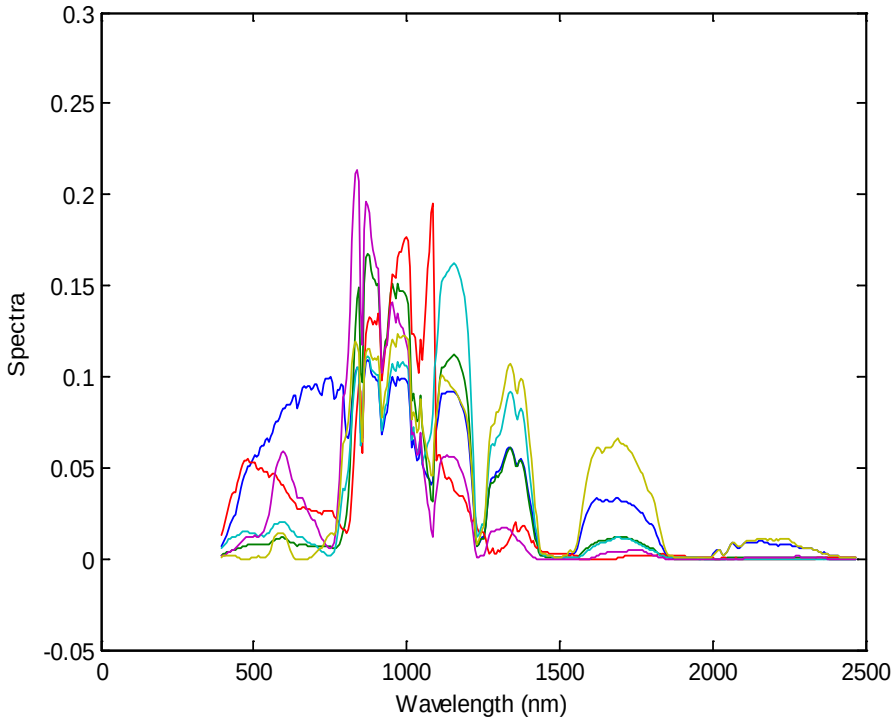

Figure 4 Spectral estimates from the Beltsville SpecTIR data using spatial subsampling.

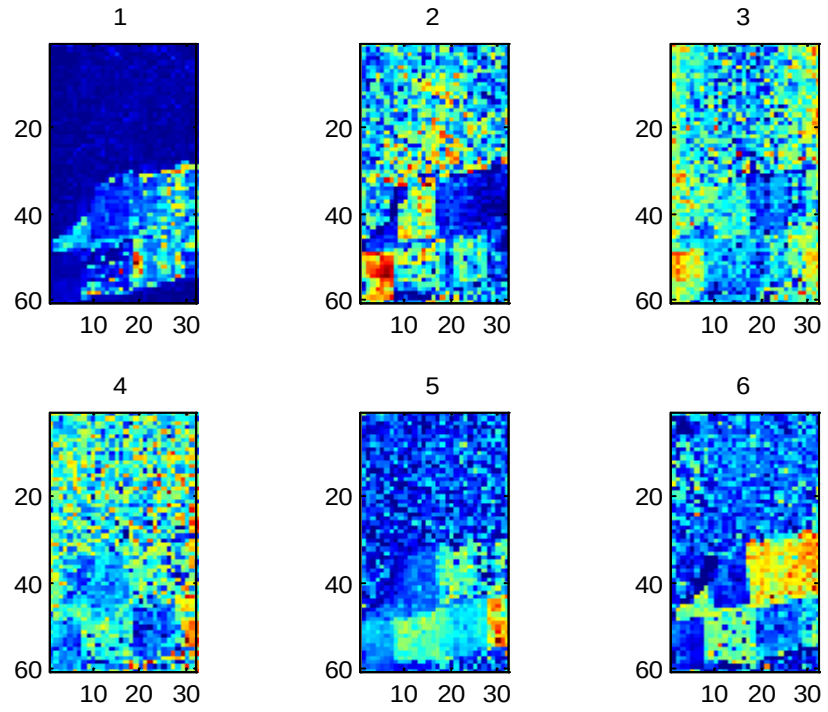

Figure 5 Spatial concentrations from the subsampled Beltsville SpecTIR data.

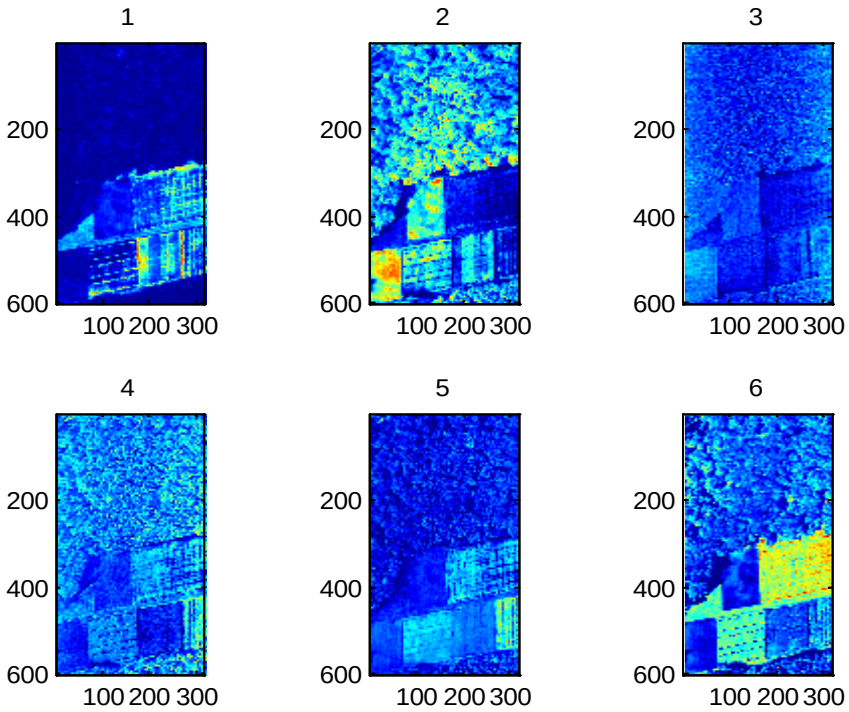

Figure 6 Spatial concentrations from Beltsville SpecTIR data at full resolution.

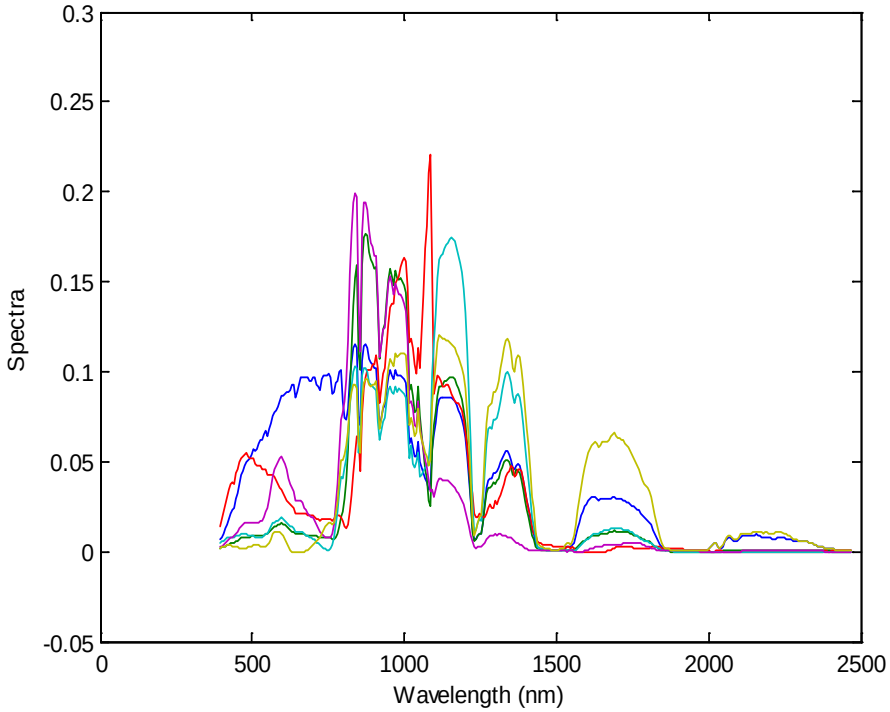

Figure 7 Spectral estimates from the Beltsville SpecTIR data without spatial subsampling.

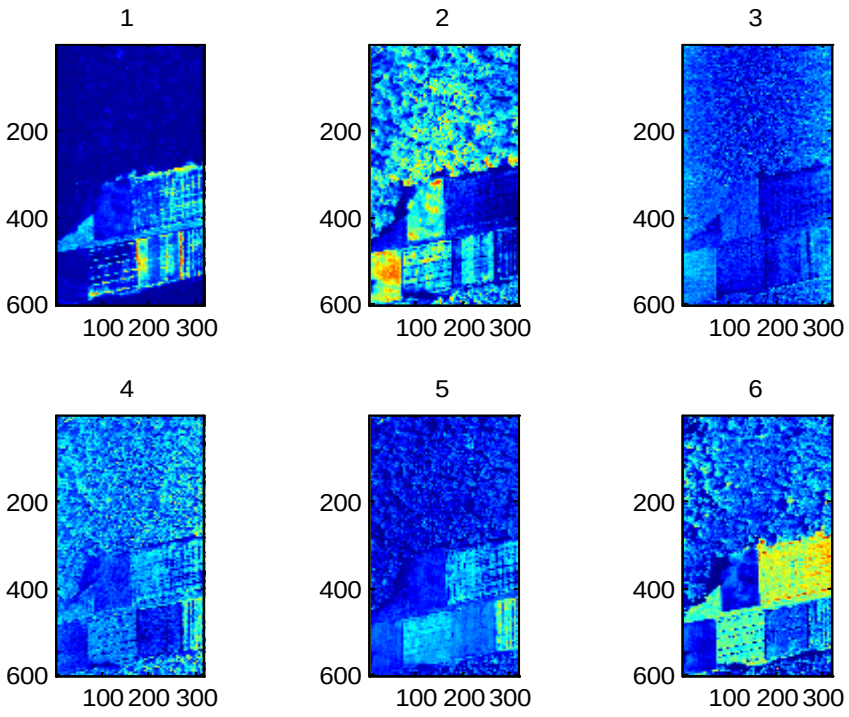

Figure 8 Spatial estimates from the Beltsville SpecTIR data without spatial subsampling.

5. Summary and Generalizations. The hyperspectral unmixing algorithm based on Bregman iteration (or, equivalently, augmented Lagrangians) using positivity and unit-vector constraint enforcement through the splitting orthogonality constraint (SOC) method of Lai and Osher has produced an efficient and reliable unmixing method that requires no prior information about the image such as spectral databases or the assumption of pure pixels. We have found that the use of spatial subsampling for the unmixing can give over two orders of magnitude improvement in the computation time with little degradation in the quality of the spectral estimates. Spatial estimates at full-image resolution are easily computed from the spectra by running the concentration-estimation half of the algorithm using the spectral estimates on the full data cube. We have illustrated the SOC-based unmixing algorithm on SpecTIR data assuming 6 materials. The processing of this and other data cubes with up to 12 materials has been done with no difficulty.

The basic method presented here for fixed data cubes can be extended in several ways. First, the size of the cube can be generalized to data collections over arbitrarily long tracks from pushbroom sensors by using overlapped blocks of row-data as they become available for fixed numbers of columns (detectors) and spectral bands. The approach has been applied successfully to the SHARE data set [11] having 2805 rows with 320 columns and 360 bands. Second, the spectral estimates from the unmixing can be used as feature vectors in a support vector machine classifier. Because different random spectral initializations give similar but distinct output spectra, we have developed an approach that uses the average of a few spectral initializations. We have found that using only 3-5 such initializations gives excellent classifier results. The efficiency of the spatial subsampling method makes this computationally feasible. Third, the approach suggested in section 3 for finding small or faint targets by first applying the basic algorithm to estimate and remove the overall background, and performing the unmixing on the residuals has been successful in detecting faint chemical plumes injected into hyperspectral background data. The same method could be used for small targets such as landmines. Finally, the methods developed here for hyperspectral data can and have been used for processing of lidar data containing mixtures of aerosol and vapor components.

Acknowledgment. Russell Warren's work is supported by UCLA under subcontract to NSF grant DMS-1118971. Stanley Osher's work is supported by NSF grants DMS-1118971 and DMS-0914561.

REFERENCES

[1] M. E. Winter, “N-findr: an algorithm for fast autonomous spectral end-member determination in hyperspectral data,” in *Imaging Spectrometry V.* 1999, vol. 3753, pp. 266-275, SPIE.

[2] J. M. P. Nascimento and J. M. Bioucas Dias, “Vertex component analysis: a fast algorithm to unmix hyperspectral data,” *IEEE Trans. Geosci. Remote Sens.,* vol. 43, no. 4, pp. 898-910, Apr. 2005.

[3] M. D. Craig, “Minimum-volume transforms for remotely sensed data,” *IEEE Trans. Geosci. Remote Sens.*, vol. 32, no. 3, pp. 542-552, May 1994.
[4] L. Miao and H. Qi, “Endmember extraction from highly mixed data using minimum volume constrained nonnegative matrix factorization,” *IEEE Trans. Geosci. Remote Sens.*, vol. 45, no. 3, pp. 765-777, 2007.
[5] R. Lai, and S. Osher, "A splitting method for orthogonally constrained problems," UCLA CAM Report 12-39, 2012.
[6] M. R. Hestenes, "Multiplier and gradient methods," *Journal of Optimization Theory and Applications*, Vol. 4, pp. 303-320, 1969.
[7] M. J. D. Powell, "A method for nonlinear constraints in minimization problems," *Optimization*, Edited by R. Fletcher, Academic Press, New York, New York, 1972.
[8] As shown in general on p. 154 of Y. Wotau, S. Osher, D. Goldfarb, and J. Darbon, "Bregman iterative algorithms for l_1-minimization with applications to compressive sensing," *SIAM J. Imaging Sci.*, 1,1, pp. 143-168, 2008.
[9] R. W. Keener, *Theoretical Statistics: Topics for a Core Course*, Springer, 2010.
[10] http://personalpages.manchester.ac.uk/staff/david.foster/Hyperspectral_images_of_natural_scenes_02.html.
[11] J. A. Herweg, J. P. Kerekes, O. Weatherbee, D. Messinger, J. van Aardt, E. J. Ientilucci, Z. Ninkov, J. Fauling, N. Raqueno,and J. Meda, SpecTIR Hyperspectral Airborne Rochester Experiment (SHARE) Data Collection Campaign, *Proc. SPIE* 8390 (2012).