HYPERSPECTRAL UNMIXING BY THE ALTERNATING DIRECTION METHOD OF MULTIPLIERS

RUSSELL E. WARREN

EO-Stat Inc. of North Carolina 10010 Vail Dr., Chapel Hill, NC, 27517

STANLEY J. OSHER

Department of Mathematics University of California, Los Angeles, CA, 90095

ABSTRACT. We have developed a method for hyperspectral image data unmixing that requires neither pure pixels nor any prior knowledge about the data. Based on the well-established Alternating Direction Method of Multipliers, the problem is formulated as a biconvex constrained optimization with the constraints enforced by Bregman splitting. The resulting algorithm estimates the spectral and spatial structure in the image through a numerically stable iterative approach that removes the need for separate endmember and spatial abundance estimation steps. The method is illustrated on data collected by the SpecTIR imaging sensor.

1. Introduction. Hyperspectral (HS) imaging is an important technology for many commercial and military applications such as natural resource discovery and intelligence gathering. Images are collected having typically hundreds of thousands of discrete pixels, each having a wavelength spectrum of 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 contain wavelength spectra that are a composite of two or more basic components in the image, algorithms are needed to automatically perform 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. This was already recognized by the likely originator of the term "endmember," Schowengerdt [3]: "... endmembers only exist as a conceptual convenience and as idealizations in real images." Also, the concept of a pure pixel may not have meaning in the LWIR spectral region.

²⁰¹⁰ Mathematics Subject Classification. xxxx.

Key words and phrases. xxxx.

Nevertheless, an enormous literature has been produced with endmembers as one of the central concepts underlying contemporary hyperspectral processing. The recent treatise on hyperspectral processing methods by Cheng [4] with over 500 references, discusses endmember-based unmixing methods extensively and emphasizes their importance.

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 attempt to use sophisticated learning 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 endmember estimation by data transformations [5] or nonnegative matrix factorization [6] applied to simplex models for the data mixing model. Those models constrain the *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 performs both spectral and spatial estimations in one step.

There are several challenges to developing an efficient and reliable algorithm for HS unmixing without prior knowledge about the spectral or spatial structure of the data. The most significant challenge is that the unknown spectral and spatial components enter the signal model as a product, thereby producing a biconvex optimization problem: convex in either component given the other, but not jointly convex. Fortunately, there is a well-established optimization framework, the Alternating Direction Method of Multipliers [7] (ADMM), that can accommodate the biconvex unmixing problem. Its implementation here leads to a very natural alternating structure with constraints on the parameter arrays enforced through an augmented Lagrangian with Bregman splitting.

Constraints are needed to produce physically meaningful and reliable results. Those constraints are chosen here to be positivity on the spatial and spectral estimates, and a unit-vector constraint on the spectrum of each material. The latter replaces the spatially convex constraints used for abundances, and are felt to be more natural physically. We have used equality constraints for the unit-norms of the spectra, and a Bregman splitting method SOC (Splitting Orthogonal Constraints) of Lai and Osher [8] that introduces auxiliary variables to split the positivity constraints from the estimation of concentration and the spectra. We have also learned that improved results are obtained by introducing an l_1 regularization of the spectral estimates through total variation (TV) also implemented through Bregman splitting and soft thresholding following Goldstein and Osher [9].

Another processing challenge is the sheer size of the typical data cube. The 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 indeed 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 ADMM with an augmented Lagrangian constructed

to implement the constraints and TV regularization of the spectral array. Section 3 assesses the performance of the ADMM unmixing algorithm with respect to several issues using HS test data collected by the SpecTIR sensor over Beltsville, MD. We summarize and discuss generalizations in section 4. Section 5 is an appendix deriving a spectral expansion method for numerically solving the Sylvester equation required by the unmixing algorithm.

2. ADMM unmixing algorithm derivation. As noted above, our optimization problem is biconvex, and as such we cannot expect to achieve globally unique estimates of the spectral and spatial arrays ρ and C; we must settle for finding local minimizers that will be dependent on the initialization of the spectral array ρ . Because of the need to estimate both ρ and C alternately from the same data cube, ADMM is the natural framework to use for the optimization.

Given the HS image data $G \in \mathbb{R}^{M \times N}$ where M denotes the number of spectral bands, and $N = N_1 N_2$ represents the number of pixels after stacking the N_1 rows and N_2 columns into a single vector, we model G as $G = \rho C + \nu$ with $\rho \in \mathbb{R}^{M \times L}_+$ and $C = \mathbb{R}^{L \times N}_+$. Here L is the number of assumed materials in the image, and ν is an additive noise with zero-mean and bounded variance. We then have the constrained problem

(1)
$$\min_{\rho,C} \frac{1}{2} \|G - \rho C\|_F^2 + \alpha \|\nabla \rho\|_1 + \langle m, c(\rho) \rangle + \frac{\lambda_m}{2} \|c(\rho)\|^2 \quad \text{s.t. } \rho \ge 0, C \ge 0,$$

with m a dual Lagrange multiplier, and $c_l(\rho) = \|\rho_l\| - 1, 1 \le l \le L$, is a unit norm equality constraint on the spectrum of each material. The first term is the data fidelity component, the second is an l_1 -regularization term on the derivative (total variation) of the spectral wavelength dependence, and the last two terms are the augmented Lagrange terms that enforce the unit normalization. The parameter α controls the balance between data fidelity and spectral smoothness.

An important feature of our method is the use of Bregman splitting to enforce the non-negativity constraints on ρ and C as well as the TV regularization of ρ . Analogous to the use of splitting for basis pursuit or other convex optimizations, the SOC (Splitting Orthogonality Constraint) method introduces additional parameters into the objective function that allow a difficult constrained optimization to be broken into simpler problems that often have either analytical or easily implemented iterative solutions.

For the HS unmixing problem we introduce three new sets of parameters: one set $r = \rho$ for the spectra, another set $s = \nabla \rho$ for their derivatives, and a set e = Cfor the spatial concentration array. Letting $\rho^k, r^k, C^k, e^k, s^k$ denote the parameter estimates at iteration k, we have the problem

(2)
$$\min_{\rho,r,C,e,s} J(C,\rho) + \alpha \|s\|_1 + \langle m, c(\rho) \rangle + \frac{\lambda_m}{2} \|c(\rho)\|^2$$

such that $\rho = r, r \ge 0, \|r_l\| = 1, C = e, e \ge 0, s = \nabla \rho,$

with $J(C, \rho) = \frac{1}{2} ||G - \rho C||_F^2$, where the subscript *F* denotes the Frobenius (Hilbert-Schmidt) matrix norm.

The augmented Lagrangian for this problem is then

$$L(C, e, p, \rho, r, q, s, n, m) = J(C, \rho) + \alpha \|s\|_{1} + \langle m, c(\rho) \rangle + \frac{\lambda_{m}}{2} \|c(\rho)\|^{2} +$$
(3)

$$\operatorname{tr}\langle p, e - C \rangle + \frac{\lambda_{C}}{2} \|e - C\|_{F}^{2} + I_{+}(e) + I_{+}(r) +$$

$$\operatorname{tr}\langle q, r - \rho \rangle + \frac{\lambda_{\rho}}{2} \|r - \rho\|_{F}^{2} + tr \langle n, s - \nabla \rho \rangle + \frac{\lambda_{s}}{2} \|s - \nabla \rho\|_{F}^{2}$$

with dual Lagrange multiplier parameters m, p, q, and n that enforce the equality constraints. The Lagrange multipliers are updated within the iterative ADMM framework by the method of multipliers developed independently by Hestenes [10] and Powell [11] for equality constraints, and later generalized to inequality constraints by Rockafellar [12]. The fourth, sixth, tenth and twelfth terms on the right represent the quadratic penalty terms that augment the classical Lagrangian. Their addition promotes the constraint enforcements in (2), adds numerical stability to the solution, and provides a systematic method for locally optimizing both the multipliers and $\{\rho, C\}$ in parallel. By virtue of the additional parameters e, r, and s, we have an unconstrained optimization over C and equality-constrained optimization over ρ with the remaining constraints easily enforced on e, r, and s analytically. The parameters $\lambda_m, \lambda_C, \lambda_\rho$, and λ_s are termed penalty parameters, and within the context of augmented Lagrangian theory, are updated at each iteration k by a factor $\gamma > 1$ such that $\lambda_i^k = \gamma \lambda_i^{k-1}$. Reference [13] discusses methods for iteratively updating the penalty parameters. Increasing the penalty functions at each iteration has the effect of significantly improving the convergence of the algorithm. The seventh and eighth terms represent indicator functions defined as $I_{+}(e) = 0$, for $e \geq 0$, $=\infty$, for e < 0.

From the structure of (3) we see that the total problem at a given iteration k breaks into two saddle point subproblems: finding (C^k, e^k, p^k) given ρ^{k-1} , et al., and finding $(\rho^k, r^k, q^k, s^k, n^k, m^k)$ given C^k , et al. For the first subproblem we have

(4)
$$(C^k, e^k, p^k) = \arg \max_{p} \min_{C, e \ge 0} L(C, e, p, \rho^{k-1}, r^{k-1}, q^{k-1}, s^{k-1}, n^{k-1}, m^{k-1}).$$

To solve (4) we use an iterative approach, computing updates to C, e, and p in that order with $||C^k - C^{k-1}||$ as the convergence criterion. In more detail, from

(5)
$$\nabla_C L = -\rho^T \left(G - \rho C \right) - p + \lambda_C \left(C - e \right) \Big|_{\rho = \rho^{k-1}, p = p^{k-1}, e = e^{k-1}} = 0,$$

we find

(6)
$$C^{k} = \left(\rho^{(k-1)T}\rho^{k-1} + \lambda_{C}^{k-1}I\right)^{-1} \left(\rho^{(k-1)T}G + p^{k-1} + \lambda_{C}^{k-1}e^{k-1}\right)$$

Similarly, $\nabla_e L = p^{k-1} - \lambda_C^{k-1} (C^k - e) = 0$ and the positivity constraint give the projection onto the positive halfspace

(7)
$$e^{k} = \left[C^{k} - p^{k-1}/\lambda_{C}^{k-1}\right]^{+} \equiv \max\left(C^{k} - p^{k-1}/\lambda_{C}^{k-1}, 0\right)$$

From the general theory of augmented Lagrangians [13] we get

(8)
$$p^{k} = p^{k-1} - \lambda_{C}^{k-1} \left(C^{k} - e^{k} \right), \quad \lambda_{C}^{k} = \gamma \lambda_{C}^{k-1}$$

The second subproblem assumes the form

(9)
$$(\rho^k, r^k, q^k, s^k, n^k, m^k) = \arg \max_{q,n,m} \min_{\substack{\rho, s, r \ge 0 \\ \|r_i\| = 1, 1 \le i \le L}} L(C^k, e^k, p^k, \rho, r, q, s, n, m).$$

Analogous to the first subproblem, we solve (9) by iteratively updating ρ, m, r, q, s , and n in that order with $\|\rho^k - \rho^{k-1}\|$ as the convergence criterion. Specifically, differentiating L with respect to ρ gives

$$\nabla_{\rho}L = -(G - \rho C)C^{T} - q + \lambda_{\rho} (\rho - r) - \nabla^{T}n - \lambda_{s}\nabla^{T} (s - \nabla\rho) + m\rho/\|\rho\| + (10) \qquad \lambda_{m}\rho(\|\rho\| - 1)/\|\rho\||_{C = C^{k}, q = q^{k-1}, r = r^{k-1}, \rho = \rho^{k-1}, s = s^{k-1}, m = m^{k-1}, n = n^{k-1} = 0,$$

and, after collecting terms with the $\|\rho\|$ terms evaluated at $\rho^{k-1},$ we find the equation for ρ^k

(11)
$$\rho^{k} \left(C^{k} C^{kT} + \lambda_{\rho}^{k-1} I_{L} + m^{k-1} / \| \rho^{k-1} \| + \lambda_{m}^{k-1} (\| \rho^{k-1} \| - 1) / \| \rho^{k-1} \| \right) + \lambda_{s}^{k-1} \nabla^{T} \nabla \rho^{k} = G C^{kT} + q^{k-1} + \lambda_{\rho}^{k-1} r^{k-1} + \nabla^{T} \left(n^{k-1} + \lambda_{s}^{k-1} s^{k-1} \right).$$

Equation (11) for ρ^k is called Sylvester's equation, and can be solved numerically in Matlab using the Hessenberg-Schur algorithm [14] or the simpler but possibly less numerically stable spectral method given in the appendix.

The associated Lagrange parameter update equations for m and λ_m are

(12)
$$m^{k} = m^{k-1} + \lambda_{m}^{k-1} \left(\left\| \rho^{k} \right\| - 1 \right), \quad \lambda_{m}^{k} = \gamma \lambda_{m}^{k-1}$$

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

(13)
$$u \equiv \rho^{k} - q^{k-1} / \lambda_{\rho}^{k-1}, \quad r^{k} = [u]^{+} / \left\| [u]^{+} \right\|_{F}$$

The Lagrange multiplier updates for q and λ_{ρ} are then

(14)
$$q^{k} = q^{k-1} - \lambda_{\rho}^{k-1} \left(\rho^{k} - r^{k}\right), \quad \lambda_{\rho}^{k} = \gamma \lambda_{\rho}^{k-1}$$

For updating s we use soft thresholding through the shrinkage mapping S(x, a) to find

(15)
$$s^{k} = S\left(\nabla\rho^{k} - \frac{n^{k-1}}{\lambda_{s}^{k-1}}, \frac{\alpha}{\lambda_{s}^{k-1}}\right)$$

with

$$S(x,a) \equiv \left\{ \begin{array}{c} x-a, x > a\\ 0, -a \le x \le a\\ x+a, x < -a \end{array} \right\}$$

Finally, the Lagrange dual variable n and λ_s are updated as

(16)
$$n^{k} = n^{k-1} + \lambda_{s}^{k-1} \left(s^{k} - \nabla \rho^{k} \right), \quad \lambda_{s}^{k} = \gamma \lambda_{s}^{k-1}.$$

Both subproblem recursions for C by (6)-(8) and ρ by (11)-(16) are iterated to convergence (with r, e, s, p, q, m and n initialized at 0) within an outer loop that iterates between the subproblems. The spectral estimates ρ^0 are initialized by positive, unit-norm vectors computed from the VCA algorithm. An alternative to VCA initialization would be random positive unit-norm vectors. VCA estimates tend to give faster convergence. The differentiation operator ∇ is implemented as a two-point difference approximation. Also, spatial subsampling is used with typically every tenth pixel for estimating the spectra from (11)-(16) followed by a final iteration of (6)-(8) for estimating C at full resolution. A summary of the overall algorithm is 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 $[R(j, k_1, k_2)]$ where $1 \leq j \leq M$, $1 \leq k_1 \leq N_1$, and $1 \leq k_2 \leq N_2$. 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 $\langle \cdot \rangle$. 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.

$$\begin{split} & \text{Input} : N_1, N_2, M, N_{samp}, R, \lambda_C^{(0)}, \lambda_p^{(0)}, \lambda_s^{(0)}, \alpha, \varepsilon, \rho^{(0)}, L, \text{tol}, \gamma \\ & \text{Initialize} : C^{(0)} = e^{(0)} = p^{(0)} = 0, r^{(0)} = q^{(0)} = s^{(0)} = n^{(0)} = m^{(0)} = 0, \bar{\rho}^{(0)} = 0 \\ & G = \langle R(1:M, 1:N_{samp}:N_1, 1:N_{samp}:N_2) \rangle \\ & D\rho = 1, k = 1 \\ & \text{while } D\rho > \varepsilon \\ & \delta = 1 \\ & \text{while } \delta > \text{tol} \\ & C^{(k)} = \left(\rho^{(k-1)T} \rho^{(k-1)} + \lambda_C^{(k-1)} I_L \right)^{-1} \left(\rho^{(k-1)T}G + p^{(k-1)} + \lambda_C^{(k-1)}e^{(k-1)} \right) \\ & e^{(k)} = \left[C^{(k)} - p^{(k-1)} / \lambda_C^{(k-1)} \right]^+ \\ & p^{(k)} = p^{(k-1)} - \lambda_C^{(k-1)} \left(C^{(k)} - e^{(k)} \right) \\ & \delta = \| C^{(k)} - C^{(k-1)} \|_F \\ & \text{end} \\ & \delta = 1 \\ & \text{while } \delta > \text{tol} \\ & p^{(k)} \left(C^{(k)C(k)T} + \lambda_p^{(k-1)} I_L + m^{(k-1)} / \| \rho^{(k-1)} \| + \\ & \lambda_m^{(k-1)} \left(\| \rho^{(k-1)} \| - 1 \right) / \| \rho^{(k-1)} \| \right) + \lambda_s^{(k-1)} \nabla^T \nabla \rho^{(k)} \\ & = GC^{(k)T} + q^{(k-1)} + \lambda_p^{(k-1)} r^{(k-1)} + \nabla^T \left(n^{(k-1)} + \lambda_s^{(k-1)} s^{(k-1)} \right) \\ & m^{(k)} = m^{(k-1)} + \lambda_m^{(k-1)} \left(\| \rho^{(k)} \| - 1 \right) \\ & u \equiv \rho^{(k)} - q^{(k-1)} / \lambda_p^{(k-1)}, \quad r^{(k)} = [u]^+ / \| [u]^+ \|_F \\ & q^{(k)} = q^{(k-1)} - \lambda_p^{(k-1)} \left(\rho^{(k)} - \nabla \rho^{(k)} \right) \\ & s^{(k)} = S \left(\nabla \rho^{(k)} - n^{(k-1)} \right) / \lambda_s^{(k)} - \nabla \rho^{(k)} \right) \\ & \delta = \| \rho^{(k)} - \rho^{(k-1)} \|_F \\ & \text{end} \\ & \bar{\rho}^{(k)} = \rho^{(k)}, \quad D\rho = \| \bar{\rho}^{(k)} - \bar{\rho}^{(k-1)} \|_F \\ & \lambda_m^{(k)} = \gamma \lambda_m^{(k-1)}, \quad \lambda_C^{(k)} = \gamma \lambda_C^{(k-1)}, \quad \lambda_s^{(k)} = \gamma \lambda_s^{(k-1)} \\ & k \to k + 1 \\ & \text{end} \\ & G = \langle R \rangle \\ & \delta = 1, m = 1, \quad C^{(0)} = \bar{e}^{(0)} = \bar{p}^{(0)} = 0 \\ & \text{while } \delta > \text{tol} \\ & C^{(m)} = \left(\bar{\rho}^{(k)T} \bar{\rho}^{(k)} + \lambda_C^{(m-1)} I_L \right)^{-1} \left(\bar{\rho}^{(k)T} G + \bar{p}^{(m-1)} + \lambda_C^{(m-1)} \bar{p}^{(m-1)} \right) \\ \end{split}$$

$$\begin{split} \bar{e}^{(m)} &= \left[C^{(m)} - \bar{p}^{(m-1)} \middle/ \lambda_C^{(m-1)} \right]^+ \\ \bar{p}^{(m)} &= \bar{p}^{(m-1)} - \lambda_C^{(m-1)} \left(C^{(m)} - \bar{e}^{(m)} \right), \quad \lambda_C^{(m)} = \gamma \lambda_C^{(m-1)} \\ \delta &= \left\| C^{(m)} - C^{(m-1)} \right\|_F, \quad m \to m+1 \\ \text{end} \end{split}$$

Box 1. TV-regularized ADMM unmixing algorithm using splitting constraint enforcement.

3. **ADMM Unmixing assessment using HS data.** Given the biconvexity of the ADMM unmixing algorithm above, there are several potential issues with the method that need to be addressed. They include:

- The dependence of the spectral estimates on the use of spatial subsampling versus the full data cube,
- The variability of the spectral estimates to different initializations,
- The stability of the algorithm to increasing the number of materials sought,
- The physical significance of the spectral estimates,
- The value of adding a unit-vector constraint for ρ and total variation regularization to the spectral estimation.

Because of the difficulty in providing analytic answers to these questions, we use Matlab calculations on a data cube collected by the state-of-the-art SpecTIR sensor using 360 spectral bands with about 5 nm resolution between about 393 nm and 2.5 μ m. We first address the spatial subsampling question. Figure 1 shows an image of radiance data from the Beltsville, MD cube having 600 by 320 pixels. The picture and data were taken from the SpecTIR Website, www.spectir.com.

We compared the subsampled (by $N_{samp} = 10$ pixels) and full resolution spectral estimates on these data using 20 trials with VCA initialization and initial penalty parameter choices $\lambda_{\rho} = 300$, $\lambda_{C} = 0.01$, $\alpha = 0.3$, $\lambda_{s} = 0.3$, $\lambda_{m} = 0.04$, and $\gamma = 1.1$ for 6 materials. Figures 2 and 3 compare the full resolution mean and standard error estimates (top) with the spatially subsampled estimates (bottom). Although there was some difference in the standard error estimates, the mean values are seen to be very close. The mean unmixing times were 782.4 s (full) and 1.026 s (subsampled). The results indicate that using only 1% of the spatial information gave satisfactory results on these data in much shorter times.

A related concern is the variability in the spectral estimates to the initialization spectra. We address that issue next using the Beltsville data cube with the ADMM algorithm (with spatial subsampling) run 20 times for different VCA initializations using the parameter choices above for 6 materials. Figure 4 plots a superposition of the spectral estimates after permuting the color order to match that of the first trial. We see that the spectral variability from initialization is similar to that seen in Figure 3. It is also similar to the variability observed from running the VCA endmember algorithm repeatedly on the same data. As a side point, we note that the spectral variability from different initializations can be reduced greatly by averaging the estimates from a few (3–5) trials.

Figures 5 and 6 compare the spatial estimates at the subsampled and full resolution from the spectra at the first trial. The discrete pixel estimates are clearly evident in Figure 5 but not Figure 6. We next examine the unmixing stability as the number of material components is increased from 2 to 20. Figure 7 plots the fitting error J at convergence with $J \equiv \left\|G - \bar{\rho}\bar{C}\right\|_F^2 / MN$ and \bar{C} computed at full spatial resolution. The error is seen to decrease smoothly with material number. This indicates that the algorithm remains stable as the number of materials increases. In effect, the unmixing algorithm is extracting more information from the data set with increasing material numbers until the noise floor in the data is reached. Evidently, about 10–12 materials are enough to capture most of the structure in the image, but the algorithm can still produce stable results for at least 20 materials.

Figures 8 and 9 show the spectral and spatial estimates for 20 materials. Three plots are used for the spectra of materials 1:7, 8:14, and 15:20, respectively. The unmixing algorithm converged in 34 iterations to a fitting error of 2.1e-5 in 8.96 s for the unmixing time. Processing was done on a Dell XPS8500 running at 3.4 GHz.

The fitting error J provides a means for selecting the number of materials, and the results here suggest that the exact number chosen is not significant due to the numerical stability of the algorithm. Figure 10 plots the monotonic decrease in spectral norm difference $\Delta \rho^k \equiv \|\rho^k - \rho^{k-1}\|_F$ between iterations for 20 materials. This monotonic decrease is another indication of the stability of the algorithm.

The question of the physical significance of the unmixing results would best be addressed using data having known spectral signatures. In the absence of such data, we applied the unmixing algorithm twice, once to estimate ρ and C, and again to unmix data made from their product ρC . Applying this idea to the Beltsville data for 4 materials gave the results in Figures 11 and 12. Figure 11 compares the initially inferred spectra (top) with their estimates from the second unmixing (bottom). Figure 12 compares the spatial estimates for the first component as input (left) and inferred (right). We see close agreement in both cases, suggesting that the estimates have an intrinsic meaning.

Lastly, we look at the value of adding the explicit equality constraint¹ for the unit norm of ρ and the total variation l_1 -regularization of ρ to the unmixing algorithm. In Figure 13 the fitting error at convergence for 20 trials of the present algorithm (HS-8) with both unit-norm and TV regularization of ρ is compared with an earlier algorithm [15] (HS-4) having neither, but otherwise the same. The TV norm of ρ is defined as $\|\nabla\rho\|_{1,l} \equiv \sum_{j=1}^{M} |\nabla\rho_{j,l}|$. The number of materials was 6 here. The substantial improvement in the fitting error from adding the TV and unit-norm terms for ρ to the augmented Lagrangian is evident.

4. Summary and generalizations. The hyperspectral unmixing algorithm described here is based on ADMM using a biconvex augmented Lagrangian with positivity and unit-norm constraint enforcement through the splitting orthogonality constraint (SOC) method of Lai and Osher [8] with total variation l_1 regularization of the spectral estimates. The resulting algorithm is an efficient and reliable unmixing method that requires no prior information about the data such as its spectral content or the assumption of pure pixels. We have found that the use of spatial subsampling that uses as little as 1% of the spatial information for the unmixing can

¹Our earlier unmixing algorithms (such as HS-4) enforced the spectral unit-norm constraint only through the auxiliary variable r. The present algorithm (HS-8) supplements this with an explicit constraint on ρ . The question then is what is gained by this extra feature and the TV regularization.

give almost three orders of magnitude improvement in the computation time with little or no degradation in the quality of the spectral estimates. Spatial estimates at full-image resolution are easily computed from the spectra by running the spatial estimation half of the ADMM algorithm on the full data cube. We have achieved numerically stable results from the ADMM unmixing algorithm on SpecTIR data with up to 20 materials.

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 [16] 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 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 classification results. The efficiency of the spatial subsampling method makes this computationally feasible. Third, the unmixing approach can be easily adapted for finding faint or small targets by first applying the basic algorithm to estimate and remove the clutter background, and performing the unmixing again on the residuals. The method has been successful in detecting faint chemical plumes in background clutter [17]. The same method could be used for small targets such as landmines. Finally, the methods developed here for hyperspectral data have been used for processing multispectral lidar data containing mixtures of aerosol and vapor components.

5. Appendix: Spectral solution of Sylvester's equation. The ADMM implementation of the HS unmixing algorithm in section 2 requires the solution of Sylvester's equation

$$\rho A + B\rho = C,$$

for $\rho \in \mathbb{R}^{M \times L}_+$, where $A \in \mathbb{R}^{L \times L}_+$, $B \in \mathbb{R}^{M \times M}_+$, and $C \in \mathbb{R}^{M \times L}$ are input matrices. One method of solving this equation is to use a Kronecker matrix product construction that involves the inversion of an $ML \times ML$ matrix. For small matrices this works well, but for implementation with M = 360 and L = 16, say, the processing times are too large for practical use with an iterative algorithm such as our ADMM unmixer.

We have developed a very efficient method for solving Sylvester's equation for the special case of A and B symmetric matrices, as they are in our unmixing problem. A more general algorithm could be based on the SVD expansion. In our case we represent A and B by their spectral representations

$$A = \sum_{k} \alpha_k \phi_k \phi_k^T, \quad B = \sum_{j} \beta_j \psi_j \psi_j^T,$$

where α and β are the eigenvalue spectra and ϕ and ψ are the orthonormal eigenvectors of dimension L and M, respectively. We expand ρ and C in these eigenvectors as

$$\rho = \sum_{j,k} a_{jk} \psi_j \phi_k^T, \quad C = \sum_{j,k} c_{jk} \psi_j \phi_k^T$$

where the expansion coefficients of C are given by orthonormality as

$$c_{jk} = \psi_j^T C \phi_k$$

Substituting the ρ expansion into those of A and B, we have

$$\rho A = \sum_{j,k} a_{jk} \psi_j \phi_k^T \sum_{k'} \alpha_{k'} \phi_{k'} \phi_{k'}^T = \sum_{j,k} a_{jk} \alpha_k \psi_j \phi_k^T,$$

$$B \rho = \sum_{j'} \beta_{j'} \psi_{j'} \psi_{j'}^T \sum_{j,k} a_{jk} \psi_j \phi_k^T = \sum_{j,k} a_{jk} \beta_j \psi_j \phi_k^T,$$

where $\phi_k^T \phi_{k'} = \delta_{kk'} \equiv 1$, for $k = k', \equiv 0$, for $k \neq k'$ and similarly for ψ were used. From Sylvester's equation we get

$$\rho A + B\rho = \sum_{j,k} a_{jk} \left(\alpha_k + \beta_j \right) \psi_j \phi_k^T = C = \sum_{j,k} c_{jk} \psi_j \phi_k^T$$

Therefore, comparing coefficients we find

$$a_{jk} = \frac{c_{jk}}{\alpha_k + \beta_j}, \quad \text{for } \alpha_k \neq -\beta_j.$$

However, since $\alpha_k \ge 0$, and $\beta_j \ge 0$, $\alpha_k = -\beta_j$ would require both to vanish. This can be avoided by retaining only eigenvalues greater than the machine precision.

We further note that for our iterative application, only A changes between iterations, so, since M > L, the larger matrix expansion for B needs to be done only once. The resulting algorithm applied to random (symmetric) matrices for M = 360 and L = 16 required 0.1 s compared to the 2.7 s needed by the Kronecker method. The results differed by less than 2e-10. When applied to a test data set of hyperspectral data with just L = 4 materials and 360 spectral bands, the Kronecker method required 1297 s for the unmixing compared with 13 s using the eigenfunction algorithm.

Acknowledgments. Russell Warren's work was supported by UCLA under subcontract to NSF grant DMS-1118971. Stanley Osher's work was supported by NSF grant DMS-1118971.

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.*, 43 (2005), 898–910.
- [3] R. A. Schowengerdt, Remote Sensing: Models and Methods for Image Processing, 2nd Ed., Academic Press, NY, 1997.
- [4] C.-I Cheng, Hyperspectral Data Processing: Algorithm Design and Analysis, John Wiley & Sons, Hoboken, NJ, 2013.
- [5] M. D. Craig, Minimum-volume transforms for remotely sensed data, *IEEE Trans. Geosci. Remote Sens.*, 32 (1994), 542–552.
- [6] L. Miao and H. Qi, Endmember extraction from highly mixed data using minimum volume constrained nonnegative matrix factorization, *IEEE Trans. Geosci. Remote Sens.*, 45 (2007), 765–777.
- [7] S. Boyd, N. Parkh, E. Chu, B. Peleato and J. Eckstein, Distributed optimization and statistical learning via the alternating direction method of multipliers, *Foundations and Trends in Machine Learning*, 3 (2010).
- [8] R. Lai, and S. Osher, A splitting method for orthogonally constrained problems, UCLA CAM Report 12-39, 2012.
- T. Goldstein and S. Osher, The split Bregman method for l-1 regularized problems, SIAM J. on Imaging Sciences, 2 (2009).
- [10] M. R. Hestenes, Multiplier and gradient methods, Journal of Optimization Theory and Applications, 4 (1969).
- [11] M. J. D. Powell, A method for nonlinear constraints in minimization problems, in *Optimiza*tion (ed. R. Fletcher), Academic Press, New York, 1972.

- [12] R. T. Rockafellar, The multiplier method of Hestenes and Powell applied to convex programming, Journal of Optimization Theory and Applications, 12 (1973).
- [13] J. Nocedal and S. J. Wright, Numerical Optimization, Section 17.4, Springer, 1999.
- [14] G. H. Golub, S. Nash and C. van Loan, A Hessenberg-Schur method for the problem AX + XB = C, *IEEE Trans. Automatic Control*, **24** (1979).
- [15] The earlier algorithm was developed under Purchase Order 1010 P PA379 to EO-Stat Inc. from UCLA under NSF grant DMS-1118971.
- [16] 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.
- [17] Wide Area Arial Reconnaissance (WAAR) Project sponsored by the Edgewood Chemical Biological Center with data made available through the Advanced Threat Detection Program run by DTRA and NSF. Data collected by the Applied Physics Laboratory with a Telops Hyper-Cam sensor in 2009. Information supplied by Alison Carr of APL.

Received xxx xxxx; revised xxx xxxx.



FIGURE 1. Beltsville, MD data cube from the SpecTIR Website.



FIGURE 2. Mean spectral estimates from full (top) and subsampled (bottom) data.



FIGURE 3. Estimated standard error from full (top) and subsampled (bottom) data.



FIGURE 4. Spectral estimates from 20 trials of Beltsville data unmixing.



FIGURE 5. Subsampled spatial estimates from the 1st trial.



FIGURE 6. Spatial estimates at full resolution from the 1st trial.



FIGURE 7. Fitting error at convergence versus number of assumed materials.



FIGURE 8. Spectral estimates for the Beltsville data run for 20 materials.



FIGURE 9. Spatial estimates at full resolution for the Beltsville data for 20 materials.



FIGURE 10. Norm difference in spectral estimates for unmixing with 20 materials.



FIGURE 11. Input and inferred spectral estimates from the Beltsville data.



FIGURE 12. Input and inferred spatial estimates for the 1st material of the Beltsville data.



FIGURE 13. Fitting errors of the current unmixer (HS-8) and an earlier version (HS-4).