Hyperspectral image unmixing via bilinear generalized approximate message passing

Jeremy Vila, Philip Schniter, and Joseph Meola

Supported in part by NSF-I/UCRC grant IIP-0968910 and NSF grant CCF-1218754

SPIE Defense, Security and Sensing @ Baltimore - 4/30/13

Hyperspectral Image Unmixing

- Goal: Estimate the N material spectra, i.e., endmembers, and the corresponding fractional abundances from a hyperspectral image (HSI) dataset $\boldsymbol{Y}$ of M spectral measurements taken across $T = T_1 \times T_2$ pixels.
- We write the received radiance data as the bilinear model

$$\boldsymbol{Y} = \boldsymbol{S}\boldsymbol{A} + \boldsymbol{W} \in \mathbb{R}^{M \times T},$$

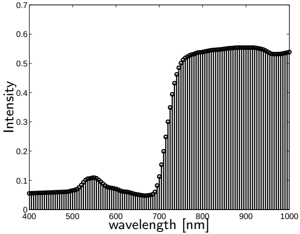
where the columns of $\boldsymbol{S} \in \mathbb{R}_+^{M \times N}$ are the non-negative (NN) endmembers, the rows of $\boldsymbol{A} \in \mathbb{R}_+^{N \times T}$ are the NN abundance maps, and $\boldsymbol{W}$ is noise.
- To satisfy sum-to-one constraints on the abundances, i.e., $\sum_n a_{nt} = 1 \; \forall t$, we augment the system model as

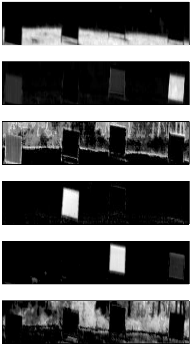
$$\underbrace{\begin{bmatrix} \boldsymbol{Y} \\ \mathbf{1}^{\mathsf{T}} \end{bmatrix}}_{\bar{\boldsymbol{Y}}} = \underbrace{\begin{bmatrix} \boldsymbol{S} \\ \mathbf{1}^{\mathsf{T}} \end{bmatrix}}_{\bar{\boldsymbol{S}}} \boldsymbol{A} + \begin{bmatrix} \boldsymbol{W} \\ \mathbf{0}^{\mathsf{T}} \end{bmatrix},$$

where $\mathbf{1}^{\mathsf{T}}$ and $\mathbf{0}^{\mathsf{T}}$ are rows of ones and zeros, respectively.

Spectral and Spatial Coherence

- In practice, there exists additional structure beyond NN constraints on $\boldsymbol{S}$ and NN & sum-to-one (i.e., simplex) constraints on $\boldsymbol{A}$.
- The amplitudes of each endmember are usually correlated, an aspect we call spectral coherence.





- Also, since each material typically inhabits a fraction of the T pixels in the scene, the abundances are sparse.
- When a material inhabits a given pixel, it is more likely to inhabit a nehigboring pixel, a property we call spatial coherence, i.e., structured sparsity.
- If we can account for these structures in our model, we can improve estimation of the endmembers and abundances.

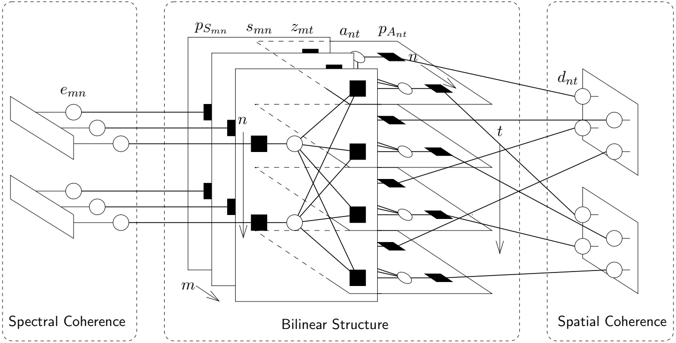
Example HSI unmixing approaches

- Endmember extraction algorithms such as vertex component analysis (VCA) [Nascimento '05] and fast separable NN matrix factorization (FSNMF) [Gillis '12] rely on the pure-pixel assumption, which may not hold in real-world data.
  - Abundances are then estimated separately, typically via fully constrained least squares (FCLS) [Heinz '01] to enforce simplex constraints.
  - This approach does not leverage spectral or spatial coherence.

- The Bayesian Linear Unmixing (BLU) [Dobigeon '09] algorithm jointly estimate the endmembers and abundances via Gibbs sampling techniques.
  - Spatially Constrained Unmixing (SCU) [Mittelman '12] expands upon BLU by employing a sticky hierarchical Dirichlet process prior to exploit spatial coherence.
  - Both BLU and SCU exhibit runtimes orders-of-magnitude larger than the "pure-pixel" approaches (with FCLS).

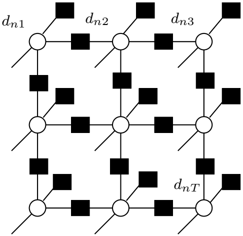
- We propose a Bayesian approach to HSI, called HSI-AMP, that jointly estimates the endmembers and abundances using the framework of loopy belief propagation.
  - We model each material's spectral amplitudes as a Markov chain, and abundances as structured-sparse with support governed by a Markov random field (MRF).
  - HSI-AMP exhibits complexities on par with "pure-pixel" approaches, with performance that exceeds them.

Proposed Approach: HSI-AMP



- The factor graph for the model assumed by HSI-AMP can be separated into three sub-graphs: spectral coherence, spatial coherence, and bilinear structure.
- Inference on the bilinear structure sub-graph is tackled using the Bilinear Generalized Approximate Message Passing (BiG-AMP) algorithm.
- We merge the three separate inference tasks using the "turbo-AMP" approach, [Schniter '12] where beliefs are exchanged between sub-graphs until they agree.

Signal Model



- We define binary support variables $d_{nt} \in \{0, 1\}$, indicating whether material n is present in pixel t.
- We assume that the abundances $\{a_{nt}\}$ are i.i.d conditional on $\{d_{nt}\}$ according to the sparse pdf

$$p_{A|D}(a_{nt}|d_{nt}) = (1 - d_{nt})\delta(a_{nt}) + d_{nt}h_A(a_{nt}),$$

where $h_A(\cdot)$ is the pdf on a_{nt} when active.

- For each n, we model the support pattern $\{d_{nt}\}_{t=1}^{T}$ as a MRF.
- To model correlation in spectral amplitudes $\{s_{mn}\}_{m=1}^{M}$, we introduce auxiliary variables $\{e_{mn}\}_{m=1}^{M}$ for each n, and model each using a Gauss-Markov chain, i.e.,

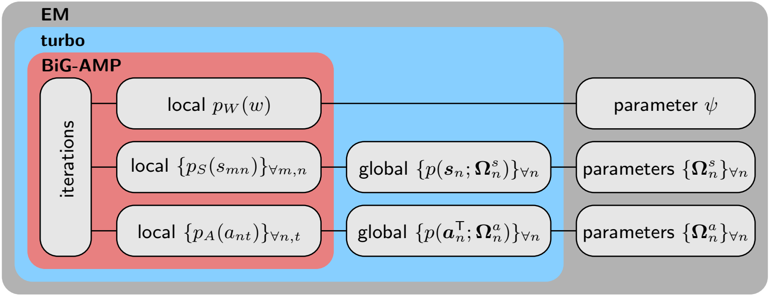
$$p(e_{mn}|e_{(m-1)n}) = \mathcal{N}\big(e_{mn}; (1 - \eta_n)e_{(m-1)n} + \eta_n\kappa_n, \eta_n^2\sigma_n^2 \tfrac{2-\eta_n}{\eta_n}\big),$$

where $\kappa_n \in \mathbb{R}$ is mean of the process, σ_n^2 is the variance, and $\eta_n \in [0, 1]$ is the correlation.

- Inference on the MRF and Gauss-Markov sub-graphs can be efficiently implemented using loopy BP [Li '09] and the backward-forward algorithm, respectively.

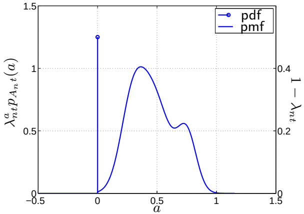
Turbo BiG-AMP and Expectation Maximization (EM)

- On its own, BiG-AMP is limited by two major assumptions:
  1. The priors are separable, e.g., $p(\boldsymbol{S}) = \prod_{m,n} p_S(s_{mn})$, $p(\boldsymbol{A}) = \prod_{n,t} p_A(a_{nt})$.
  2. The priors are perfectly-matched to the data.
- The "turbo" extension allows us to use BiG-AMP with non-separable priors
- The EM extension allowed us to tune the distributional parameters on the local priors and the Gauss-Markov and MRF priors.



BiG-AMP local priors

- We want the (EM-tuned) local priors to closely-match the true marginal distributions while yielding tractible BiG-AMP computations.
- We assign the local prior on s_{mn} as
$$p_{S_{mn}}(s) = \mathcal{N}_+(s; \theta^s_{mn}, \phi^s_{mn}),$$
where $\mathcal{N}_+(s;\theta,\phi)$ is a $\mathcal{N}(s;\theta,\phi)$ distribution truncated on $[0,\infty)$ and scaled appropriately.
- θ^s_{mn} and ϕ^s_{mn} are set to mean and variance of the most recent message from the e_{mn} node.



- We assign the local prior on a_{nt} as a Bernoulli non-negative Gaussian mixture pdf
$$p_{A_{nt}}(a) = (1-\lambda^a_{nt})\delta(a) + \lambda^a_{nt} \sum_{\ell=1}^{L} \omega^a_{n\ell} \mathcal{N}_+(a; \theta^a_{n\ell}, \phi^a_{n\ell}),$$
where λ^a_{nt} is set from the most recent message from the d_{nt} node.
- We assume that the coefficients of the noise $\boldsymbol{W}$ are i.i.d. Gaussian with variance ψ.
- The parameters $\{\{\omega^a_{n\ell}, \theta^a_{n\ell}, \phi^a_{n\ell}\}\forall n,\ell, \psi\}$ are all tuned via EM.

Initializations

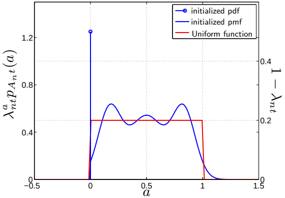
- Since the EM algorithm may converge to a local, rather than global, maximum of the likelihood, proper initialization is critical.

- We initialize the endmembers (i.e., $(\hat{\boldsymbol{S}})^0$) at the solutions provided by VCA.
- Using $(\hat{\boldsymbol{S}})^0$, we apply UCLS to initialize the abundance maps (i.e., $(\hat{\boldsymbol{A}})^0$)

- For the endmembers' NNG distributions, we set $(\boldsymbol{\theta}^a)^0 = (\hat{\boldsymbol{S}})^0$ and $(\boldsymbol{\phi}^a)^0 = \mathbf{1}$.

- For the abundances' BNNGM parameters, we set $(\lambda^a)^0 = \frac{1}{2}$ and $L = 3$. $\{\omega^a_{n\ell}\}$, $\{\theta^a_{n\ell}\}$, and $\{\phi^a_{n\ell}\}$ were set to best fit the uniform pdf on $[0, 1]$.

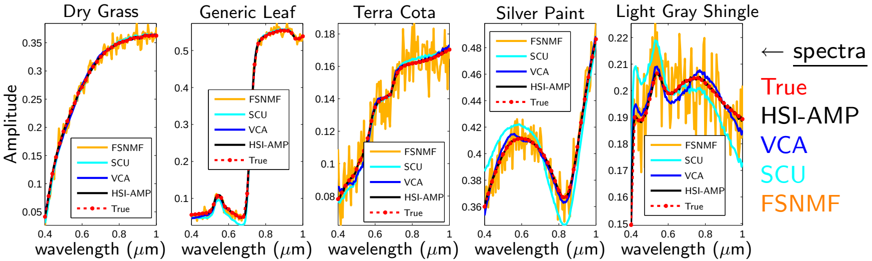
- Set noise variance as $\psi^0 = \|\boldsymbol{Y}\|_F^2/(MT(\text{SNR}^0 + 1))$, where without user input, we assume $\text{SNR}^0 = 100$.



- Automatic selection of the model order N is an important topic for future research.

Results: Pure-Pixel Synthetic Data

- Pure pixel abundance maps of size $T = 50 \times 50$ were generated with $N = 5$ materials residing in equal sized strips and the SNR was set to 30 dB.
- The endmember spectra were taken from a reflectance library.
- In one realization, shown below, HSI-AMP's estimates match the true endmembers.
- FSNMF estimates appear noisy, and all competing algorithms fail to recover silver paint and light gray shingle endmembers.

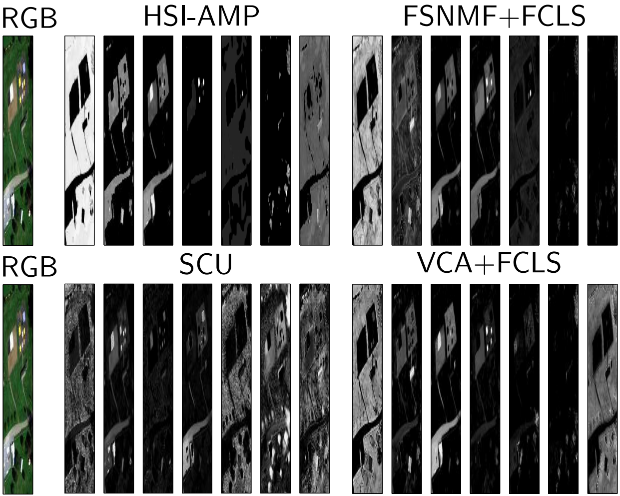


Results: Pure-Pixel Synthetic Data

- Averaging over $R = 10$ iterations, we reported the runtime and NMSE recovery.
- HSI-AMP outperformed the best competing technique (VCA+FCLS) by more than 16 dB in NMSE_S and about 7 dB in NMSE_A.
- HSI-AMP's runtime is comparable VCA+FCLS and FSNMF+FCLS, and 2-3 orders of magnitude faster than SCU.
- HSI-AMP is 2-3 orders slower than VCA+UCLS and FSNMF+UCLS, but their use of UCLS comes at the cost of 25 dB less accuracy in $\boldsymbol{A}$.

	$\boldsymbol{S}$ Runtime	$\boldsymbol{A}$ Runtime	Total Runtime	$\text{NMSE}_{\boldsymbol{S}}$	$\text{NMSE}_{\boldsymbol{A}}$
HSI-AMP	-	-	5.35 sec	**-57.1** dB	**-37.3 dB**
SCU	-	-	2808 sec	-30.6 dB	-20.5 dB
VCA + FCLS	0.05 sec	4.08	4.13 sec	-39.6 dB	-30.5 dB
VCA + UCLS	0.05 sec	0.0007 sec	0.05 sec	-39.6 dB	-12.0 dB
FSNMF + FCLS	0.002 sec	3.97 sec	3.97 sec	-25.3 dB	-12.5 dB
FSNMF + UCLS	0.002 sec	0.0008 sec	0.002 sec	-23.4 dB	-6.8 dB

Results: SHARE 2012 dataset



- Data consisted of $M = 360$ spectral bands, ranging from $400 - 2450$ nm, taken over scene of $T = 150 \times 100$ pixels.
- HSI-AMP appears to do a better job distinguishing among materials than these state-of-the-art unmixing algorithms.
- We're currently waiting on ground-truth data to enable a quantifiable comparison.

Conclusions

- VCA and FSNMF assume the presence of pure pixels, which may not exist in real data, and do not exploit the spatial and spectral coherence that usually do exist.
  - Abundance estimation is usually done separately via FCLS.
- SCU exploits spectral and spatial coherence and jointly estimates $\boldsymbol{S}$ and $\boldsymbol{A}$, but runtimes are orders-of-magnitude slower than competing approaches, and its Gibbs sampling appears to be finicky.
- HSI-AMP showed state-of-the-art joint estimation of $\boldsymbol{S}$ and $\boldsymbol{A}$ in two experiments, while exhibiting complexities on par with VCA+FCLS and FSNMF+FCLS.
- We attribute HSI-AMP's success to its ability to leverage known spectral and spatial coherence properties, while learning the prior parameters via EM.
- Automatic selection of the model order N is an important topic for future research.
- Detection of known materials is another potential area for future research.

Thanks to:

1. Nina Raqueno et al. at the Rochester Institute of Technology for providing the SHARE 2012 dataset
2. Jason Parker for his assistance with BiG-AMP

References

1. Nascimento, J. and Bioucas Dias, J., Vertex component analysis: A fast algorithm to unmix hyperspectral data, *Geoscience and Remote Sensing, IEEE Transactions on* 43(4), 898910 (2005).
2. Gillis, N. and Vavasis, S. A., Fast and robust recursive algorithms for separable nonnegative matrix factorization, *arXiv:1208.1237v2* (2012).
3. Dobigeon, N., Moussaoui, S., Coulon, M., Tourneret, J. Y., and Hero, A., Joint bayesian endmember extraction and linear unmixing for hyperspectral imagery, *Signal Processing, IEEE Transactions on* 57(11), 43554368 (2009).
4. Mittelman, R., Dobigeon, N., and Hero, A., Hyperspectral image unmixing using a multiresolution sticky HDP, *Signal Processing, IEEE Transactions on* 60(4), 16561671 (2012).
5. Heinz, D. and Chang, C.-I., Fully constrained least squares linear spectral mixture analysis method for material quantification in hyperspectral imagery, *Geoscience and Remote Sensing, IEEE Transactions on* 39(3), 529545 (2001).
6. Frey, B. J. and MacKay, D. J. C., A revolution: Belief propagation in graphs with cycles, *Proc. Neural Inform. Process. Syst. Conf.*, 479485 (1997).
7. Li, S. Z., [*Markov Random Field Modeling in Image Analysis*], Springer, London, 3rd ed. (2009).
8. Donoho, D. L., Maleki, A., and Montanari, A., Message passing algorithms for compressed sensing: I. Motivation and construction, *Proc. Inform. Theory Workshop*, 15 (Jan. 2010).
9. Rangan, S., Generalized approximate message passing for estimation with random linear mixing, *Proc. IEEE Int. Symp. Inform. Thy.*, (Aug. 2011). (See also *arXiv:1010.5141*).
10. Schniter, P., Turbo reconstruction of structured sparse signals, *Proc. Conf. Inform. Science & Syst.*, 16 (Mar. 2010).
11. Dempster, A., Laird, N. M., and Rubin, D. B., Maximum-likelihood from incomplete data via the EM algorithm, *J. Roy. Statist. Soc.* 39, 117 (1977).
12. Giannandrea, A., Raqueno, N., Messinger, D. W., Faulring, J., Kerekes, J. P., van Aardt, J., Canham, K., Hagstrom, S., Ontiveros, E., Gerace, A., Kaufman, J., Vongsy, K. M., Griffith, H., and Bartlett, B. D., The SHARE 2012 data collection campaign, (April 2013).