Bayesian Fusion of Multi-band Images Beyond Pansharpening

Jean-Yves Tourneret

Joint work with Qi Wei, Jose M. Bioucas-Dias and Nicolas Dobigeon

University of Toulouse, IRIT/INP-ENSEEIHT & TéSA

23 Juin 2016, Strasbourg, France

Panchromatic Image (50cm)

Thanks to Mathias Ortner from Airbus Defence and Space

Multispectral Image (2m)

Thanks to Mathias Ortner from Airbus Defence and Space

Pansharpened Image (50cm)

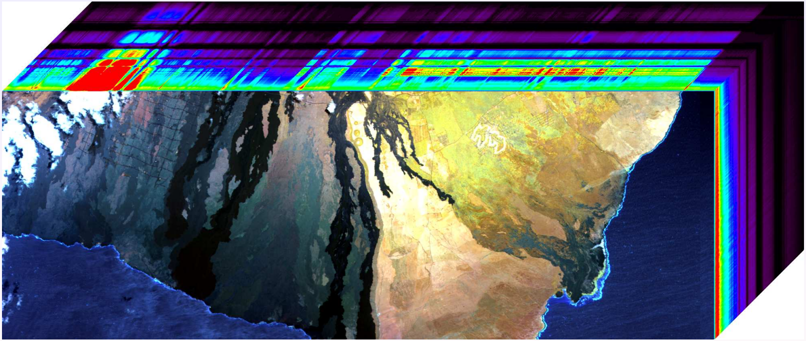
Thanks to Mathias Ortner from Airbus Defence and Space

Hyperspectral Imagery

Hyperspectral Images

- Spectral: same scene observed at different wavelengths
- Spatial: pixel represented by a vector of hundreds of measurements.

Hyperspectral Cube



Problem Statement

Figure: (a) Hyperspectral Image (size: 99 $\times$ 46 $\times$ 224, res.: 80m $\times$ 80m) (b) Panchromatic Image (size: 396 $\times$ 184 $\times$ 1 res.: 20m $\times$ 20m) (c) Target (size: 396 $\times$ 184 $\times$ 224 res.: 20m $\times$ 20m)

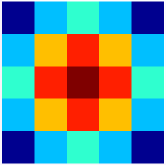
Name	AVIRIS (HS)[1]	SPOT-5 (MS)	Pleiades (MS)	WorldView-3 (MS)
Res. (m)	20	10	2	1.24
# bands	224	4	4	8

[1] R. O. Green *et al.*, "Imaging spectroscopy and the airborne visible/infrared imaging spectrometer (AVIRIS)," *Remote Sens. of Environment*, 1998.

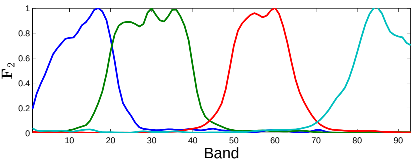
Forward model

$$\mathbf{Y}_{\mathrm{H}} = \mathbf{XBS} + \mathbf{N}_{\mathrm{H}}, \quad \mathbf{Y}_{\mathrm{M}} = \mathbf{RX} + \mathbf{N}_{\mathrm{M}}$$

- **$\mathbf{X} \in \mathbb{R}^{m_\lambda \times n}$: full resolution unknown image**
- $\mathbf{Y}_{\mathrm{H}} \in \mathbb{R}^{m_\lambda \times m}$ and $\mathbf{Y}_{\mathrm{M}} \in \mathbb{R}^{n_\lambda \times n}$: observed HS and MS images
- $\mathbf{B} \in \mathbb{R}^{n \times n}$: cyclic convolution operator acting on the bands
- $\mathbf{S} \in \mathbb{R}^{n \times m}$: downsampling matrix
- $\mathbf{R} \in \mathbb{R}^{n_\lambda \times m_\lambda}$: spectral response of the MS sensor
- $\mathbf{N}_{\mathrm{H}} \in \mathbb{R}^{m_\lambda \times m}$ and $\mathbf{N}_{\mathrm{M}} \in \mathbb{R}^{n_\lambda \times n}$: HS and MS noises, assumed to be a band-dependent Gaussian sequence



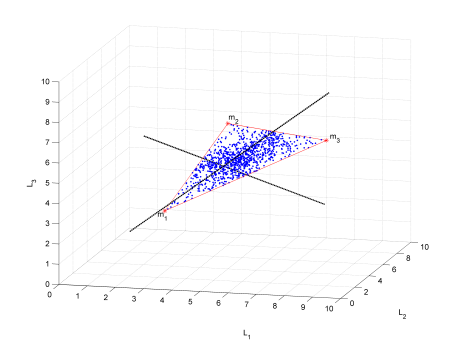
(a) Kernel of **B** (Spatial blurring)



(b) **R** (Spectral blurring)

Reparameterization

Dimensionality reduction



Projection of the data **X** in a lower-dimensional subspace ($\mathbb{R}^{\widetilde{m}_\lambda}$): $\mathbf{X} = \mathbf{H}\mathbf{U}$, where **H** is an $\widetilde{m}_\lambda \times m_\lambda$ projection matrix[2].

[2] J. M. Bioucas-Dias *et al.*, "Hyperspectral subspace identification," IEEE Trans. Geosci. and Remote Sens., vol. 46, no. 8, pp. 2435-2445, 2008.

Likelihoods

- Likelihood of the observations[3]

$$\mathbf{Y}_{\mathrm{H}}|\mathbf{U}, s_{\mathrm{H}}^2 \sim \mathcal{MN}_{m_\lambda, m}(\mathbf{HUBS}, \mathrm{diag}\left(s_{\mathrm{H}}^2\right), \mathbf{I}_m)$$
$$\mathbf{Y}_{\mathrm{M}}|\mathbf{U}, s_{\mathrm{M}}^2 \sim \mathcal{MN}_{n_\lambda, n}(\mathbf{RHU}, \mathrm{diag}\left(s_{\mathrm{M}}^2\right), \mathbf{I}_n)$$

where $s_{\mathrm{H}}^2 = \left[s_{\mathrm{H},1}^2, \ldots, s_{\mathrm{H},m_\lambda}^2\right]^T$ and $s_{\mathrm{M}}^2 = \left[s_{\mathrm{M},1}^2, \ldots, s_{\mathrm{M},n_\lambda}^2\right]^T$.

- Joint likelihood

$$f\left(\mathbf{Y}_{\mathrm{H}}, \mathbf{Y}_{\mathrm{M}}|\mathbf{U}, \boldsymbol{s}^2\right) = f\left(\mathbf{Y}_{\mathrm{H}}|\mathbf{U}, s_{\mathrm{H}}^2\right) f\left(\mathbf{Y}_{\mathrm{M}}|\mathbf{U}, s_{\mathrm{M}}^2\right)$$

with $\boldsymbol{s}^2 = \left\{s_{\mathrm{H}}^2, s_{\mathrm{M}}^2\right\}$

[3] The probability density function of a matrix normal distribution is defined by

$$p(\mathbf{X}|\mathbf{M}, \boldsymbol{\Sigma}_r, \boldsymbol{\Sigma}_c) = \frac{\exp\left(-\frac{1}{2}\mathrm{tr}\left[\boldsymbol{\Sigma}_c^{-1}(\mathbf{X}-\mathbf{M})^T\boldsymbol{\Sigma}_r^{-1}(\mathbf{X}-\mathbf{M})\right]\right)}{(2\pi)^{np/2}|\boldsymbol{\Sigma}_c|^{n/2}|\boldsymbol{\Sigma}_r|^{p/2}}$$

Outline

Outline

Parameter Priors

- Pixel vectors in the lower dimensional subspace: independent conjugate Gaussian priors

$$\mathbf{U}|\bar{\mathbf{U}}, \boldsymbol{\Sigma} \sim \mathcal{MN}\left(\bar{\mathbf{U}}, \boldsymbol{\Sigma}, \mathbf{I}_n\right)$$

- Noise variances: independent conjugate inverse-gamma priors

$$s^2_{\mathrm{H},\ell} \ \& \ s^2_{\mathrm{M},\ell}|\nu, \gamma \sim \mathcal{IG}\left(\frac{\nu}{2}, \frac{\gamma}{2}\right)$$

Flexible distribution whose shape can be adjusted from (ν, γ)

Assumptions

- $\bar{\mathbf{U}}$: fixed using an interpolated hyperspectral image (obtained using splines) projected onto the subspace
- ν: fixed (disappears later)

Hyperparameter Prior

Hyperparameter vector: $\boldsymbol{\Phi} = \{\boldsymbol{\Sigma}, \gamma\}$

- Hyperparameter $\boldsymbol{\Sigma}$: Inverse-Wishart ($\mathcal{IW}$) distribution

$$\boldsymbol{\Sigma} \sim \mathcal{W}^{-1}(\boldsymbol{\Psi}, \eta)$$

where $\boldsymbol{\Psi}$ and η are fixed to provide a non-informative prior

- Hyperparameter γ: Jeffreys' non-informative prior

$$f(\gamma) \propto \frac{1}{\gamma} \mathbf{1}_{\mathbb{R}^+}(\gamma)$$

Joint Posterior

Using Bayes theorem, the joint posterior distribution is

$$f(\boldsymbol{\theta}, \boldsymbol{\Phi} | \mathbf{Y}_{\mathrm{H}}, \mathbf{Y}_{\mathrm{M}}) \propto f(\mathbf{Y}_{\mathrm{H}}, \mathbf{Y}_{\mathrm{M}} | \boldsymbol{\theta}) f(\boldsymbol{\theta} | \boldsymbol{\Phi}) f(\boldsymbol{\Phi})$$

where

- unknown parameters: $\boldsymbol{\theta} = \left\{\mathbf{U}, s_{\mathrm{H}}^2, s_{\mathrm{M}}^2\right\}$
- unknown hyperparameters: $\boldsymbol{\Phi} = \{\boldsymbol{\Sigma}, \gamma\}$

How can we estimate $\boldsymbol{\theta}$ and $\boldsymbol{\Phi}$?

- Marginalize the hyperparameter γ
- Sample according to the joint posterior $f\left(\mathbf{U}, \boldsymbol{s}^2, \Sigma | \mathbf{Y}_{\mathrm{H}}, \mathbf{Y}_{\mathrm{M}}\right)$ by using a block Gibbs sampler, which can be easily implemented since all the conditional distributions associated with $f\left(\mathbf{U}, \boldsymbol{s}^2, \Sigma | \mathbf{Y}_{\mathrm{H}}, \mathbf{Y}_{\mathrm{M}}\right)$ are simple.

Outline

Block Gibbs sampler[4]

for $t = 1$ to N_{MC} **do**

% Sampling the image covariance matrix

Sample $\Sigma^{(t)}$ from $f(\Sigma|\mathbf{U}^{(t-1)}, \boldsymbol{s}^{2^{(t-1)}}, \mathbf{Y}_{\mathrm{H}}, \mathbf{Y}_{\mathrm{M}})$

% Sampling the multispectral noise variances

for $\ell = 1$ to n_{λ} **do**

Sample $s_{\mathrm{M},\ell}^{2(t)}$ from $f(s_{\mathrm{M},\ell}^{2}|\mathbf{U}, \mathbf{Y}_{\mathrm{M}})$,

end for

% Sampling the hyperspectral noise variances

for $\ell = 1$ to m_{λ} **do**

Sample $s_{\mathrm{H},\ell}^{2(t)}$ from $f(s_{\mathrm{H},\ell}^{2}|\mathbf{U}, \mathbf{Y}_{\mathrm{H}})$,

end for

% Sampling the high-resolved image

Sample $\mathbf{U}^{(t)}$ using a Hamiltonian Monte Carlo algorithm

end for

[4] Q. Wei *et al.*, "Bayesian fusion of multi-band images," *IEEE J. Sel. Topics Signal Process.*, vol. 9, no. 6, pp. 1117-1127, Sept. 2015.

Conditional Distributions

- Covariance matrix of the image $\boldsymbol{\Sigma}$

$$\boldsymbol{\Sigma}|\mathbf{u}, \boldsymbol{s}^2, \mathbf{Y}_{\mathrm{H}}, \mathbf{Y}_{\mathrm{M}} \sim \mathcal{W}^{-1}\left(\boldsymbol{\Psi} + \sum_{i=1}^{m_{\mathrm{x}} m_{\mathrm{y}}} (\boldsymbol{u}_i - \boldsymbol{\mu}_{\boldsymbol{u}}^{(i)})^T (\boldsymbol{u}_i - \boldsymbol{\mu}_{\boldsymbol{u}}^{(i)}), n + \eta\right)$$

- Noise variance vector $\boldsymbol{s}^2$

$$s_{\mathrm{H},\ell}^2|\mathbf{U}, \mathbf{Y}_{\mathrm{H}} \sim \mathcal{IG}\left(\frac{m}{2}, \frac{\left[\|\mathbf{Y}_{\mathrm{H}} - \mathbf{HUBS}\|_F^2\right]_\ell}{2}\right)$$

$$s_{\mathrm{M},\ell}^2|\mathbf{U}, \mathbf{Y}_{\mathrm{H}} \sim \mathcal{IG}\left(\frac{n}{2}, \frac{\left[\|\mathbf{Y}_{\mathrm{M}} - \mathbf{RHU}\|_F^2\right]_\ell}{2}\right)$$

Conditional Distributions (Cont.)

- Highly-resolved image **U**

$$-\log f(\mathbf{U}|\Sigma, \boldsymbol{s}^2, \mathbf{Y}_{\mathrm{H}}, \mathbf{Y}_{\mathrm{M}}) = \frac{1}{2}\|\Lambda_{\mathrm{H}}^{-\frac{1}{2}}(\mathbf{Y}_{\mathrm{H}} - \mathbf{HUBS})\|_F^2 + \frac{1}{2}\|\Lambda_{\mathrm{M}}^{-\frac{1}{2}}(\mathbf{Y}_{\mathrm{M}} - \mathbf{RHU})\|_F^2 + \frac{1}{2}\|\Sigma^{-\frac{1}{2}}(\mathbf{U} - \mu_{\mathbf{U}})\|_F^2 + \mathrm{C}$$

  - Not a matrix normal distribution but a normal distribution in vector form: huge covariance matrix
  - Very difficult to draw samples directly from the conditional distribution w.r.t. **U**
  - A Hamiltonian Monte Carlo method[5] is used to sample this high dimensional Gaussian distribution.
    - Other techniques, such as PO, is also possible.

[5] Neal, Radford M. "MCMC using Hamiltonian dynamics." *Handbook of Markov Chain Monte Carlo 2*, 2011.

Hamiltonian Monte Carlo Methods

Classical Metropolis-Hastings moves

- Classical proposal: random walk
- Accept/reject procedure

Can be inefficient for sampling large vectors (low acceptance rate and mixing properties)

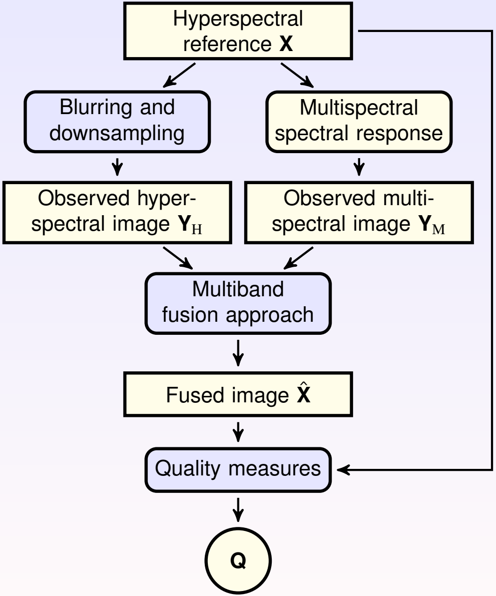
Deterministic gradient based methods

- Well adapted to update vector/matrix elements simultaneously
- Local behavior of a cost function

Hamiltonian Monte Carlo methods

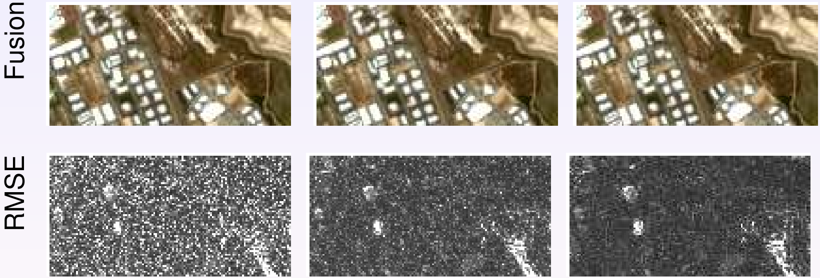
- Consideration of the local curvature of the target density to build an accurate proposal distribution for sampling vector/matrix elements simultaneously

Wald's protocol



Qualitative Results (AVIRIS dataset)

(a) HS (b) MS (c) Reference



(d) MAP[Hardie2004] (e) Wavelet MAP[Zhang2012] (f) MCMC

Quantitative Performance Measures

- RMSE/RSNR (Root Mean Square Error): a similarity measure between the target image $\mathbf{X}$ and the fused image $\hat{\mathbf{X}}$

$$\mathrm{RMSE}(\mathbf{X}, \hat{\mathbf{X}}) = \frac{1}{nm_\lambda} \|\mathbf{X} - \hat{\mathbf{X}}\|_F^2$$

$$\mathrm{RSNR}(\mathbf{X}, \hat{\mathbf{X}}) = \log \frac{1}{nm_\lambda} \frac{\|\mathbf{X}\|_F^2}{\mathrm{RMSE}}$$

  The smaller RMSE/larger RSNR, the better the fusion quality.

- SAM (Spectral Angle Mapper): spectral distortion between the actual and estimated images

$$\mathrm{SAM}(\boldsymbol{x}_n, \hat{\boldsymbol{x}}_n) = \arccos\left(\frac{\langle \boldsymbol{x}_n, \hat{\boldsymbol{x}}_n \rangle}{\|\boldsymbol{x}_n\|_2 \|\hat{\boldsymbol{x}}_n\|_2}\right)$$

  The overall SAM is obtained by averaging the SAMs computed from all image pixels. The smaller the absolute value of SAM, the less important the spectral distortion.

Quantitative Performance Measures

- UIQI (Universal Image Quality Index): related to the correlation, luminance distortion and contrast distortion of the estimated image w.r.t. the reference image. The UIQI between two images $\mathbf{a}$ and $\hat{\mathbf{a}}$ is

$$\mathrm{UIQI}(\mathbf{a}, \hat{\mathbf{a}}) = \frac{4\sigma^2_{a\hat{a}}\mu_a\mu_{\hat{a}}}{(\sigma^2_a + \sigma^2_{\hat{a}})(\mu^2_a + \mu^2_{\hat{a}})}$$

  where $(\mu_a, \mu_{\hat{a}}, \sigma^2_a, \sigma^2_{\hat{a}})$ are the sample means and variances of a and $\hat{a}$, and $\sigma^2_{a\hat{a}}$ is the sample covariance of $(a, \hat{a})$. The range of UIQI is $[-1, 1]$. The larger the UIQI, the better the fusion result.

- DD (degree of distortion): DD between two images $\mathbf{X}$ and $\hat{\mathbf{X}}$ is defined as

$$\mathrm{DD}(\mathbf{X}, \hat{\mathbf{X}}) = \frac{1}{nm_\lambda}\|\mathrm{vec}(\mathbf{X}) - \mathrm{vec}(\hat{\mathbf{X}})\|_1.$$

  The smaller DD, the better the fusion.

Quantitative Performance Measures

- ERGAS The relative dimensionless global error in synthesis (ERGAS) calculates the amount of spectral distortion in the image. This measure of fusion quality is defined as

$$\mathrm{ERGAS} = 100 \times \frac{1}{d^2}\sqrt{\frac{1}{m_\lambda}\sum_{i=1}^{m_\lambda}\left(\frac{\mathrm{RMSE}(i)}{\mu_i}\right)}$$

  where $1/d^2$ is the ratio between the pixel sizes of the MS and HS images, μ_i is the mean of the ith band of the HS image, and m_λ is the number of HS bands. The smaller ERGAS, the smaller the spectral distortion.

Quantitative Results (AVIRIS dataset)

Table: Performance of HS+MS fusion methods in terms of: RSNR (db), UIQI, SAM (deg), ERGAS and DD($\times 10^{-2}$) (AVIRIS dataset).

Methods	RSNR	UIQI	SAM	ERGAS	DD	Time(s)
MAP [6]	23.33	0.9913	5.05	4.21	4.87	**1.6**
Wavelet [7]	25.53	0.9956	3.98	3.95	3.89	31
Proposed	**26.74**	**0.9966**	**3.40**	**3.77**	**3.33**	530

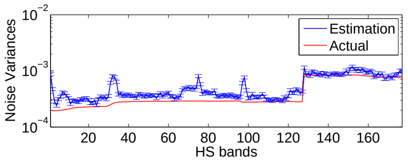
Advantages

- Samples generated by the proposed method can be used to compute uncertainties about the estimates (confidence interval)
- Generalization to more complex problems (non-Gaussianities, nonlinearity, etc)
- Noise variance estimation

[6] Hardie *et al.*, "Application of the Stochastic Mixing Model to Hyperspectral Resolution Enhancement," *IEEE Trans. Image Process.*, vol. 13, no. 9, Sept. 2004.

[7] Zhang *et al.*, "Noise-Resistant Wavelet-Based Bayesian Fusion of Multispectral and Hyperspectral Images," *IEEE Trans. Geosci. and Remote Sens.*, vol. 47, no. 11, Nov. 2009.

Noise Variance Estimation



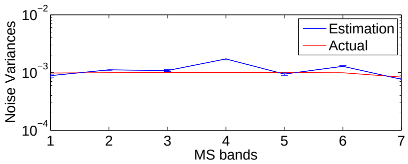


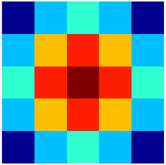
Figure: Noise variances and their MMSE estimates. (Top) HS image. (Bottom) MS image.

- Good estimation performance
- Track the variations of the noise variances within tolerable discrepancy

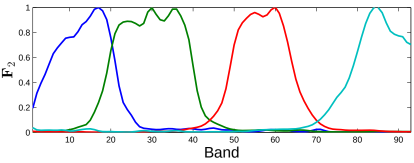
Forward model

$$\mathbf{Y}_{\mathrm{H}} = \mathbf{XBS} + \mathbf{N}_{\mathrm{H}}, \quad \mathbf{Y}_{\mathrm{M}} = \mathbf{RX} + \mathbf{N}_{\mathrm{M}}$$

- **$\mathbf{X} \in \mathbb{R}^{m_\lambda \times n}$: full resolution unknown image**
- $\mathbf{Y}_{\mathrm{H}} \in \mathbb{R}^{m_\lambda \times m}$ and $\mathbf{Y}_{\mathrm{M}} \in \mathbb{R}^{n_\lambda \times n}$: observed HS and MS images
- $\mathbf{B} \in \mathbb{R}^{n \times n}$: cyclic convolution operator acting on the bands
- $\mathbf{S} \in \mathbb{R}^{n \times m}$: downsampling matrix
- $\mathbf{R} \in \mathbb{R}^{n_\lambda \times m_\lambda}$: spectral response of the MS sensor
- $\mathbf{N}_{\mathrm{H}} \in \mathbb{R}^{m_\lambda \times m}$ and $\mathbf{N}_{\mathrm{M}} \in \mathbb{R}^{n_\lambda \times n}$: HS and MS noises, assumed to be a band-dependent Gaussian sequence



(a) Kernel of **B** (Hyperspectral)



(b) **R** (Multispectral)

Outline

Block Gibbs sampler with unknown R

for $t = 1$ to N_{MC} **do**

% Sampling the image covariance matrix

Sample $\Sigma_{\mathbf{u}}^{(t)}$ from $f(\Sigma|\mathbf{U}^{(t-1)}, \boldsymbol{s}^{2\,(t-1)}, \mathbf{Y}_{\mathrm{H}}, \mathbf{Y}_{\mathrm{M}})$

% Sampling the multispectral noise variances

for $\ell = 1$ to n_λ **do**

Sample $s_{\mathrm{M},\ell}^{2(t)}$ from $f(s_{\mathrm{M},\ell}^{2}|\mathbf{U}^{(t-1)}, \mathbf{Y}_{\mathrm{M}})$,

end for

% Sampling the hyperspectral noise variances

for $\ell = 1$ to m_λ **do**

Sample $s_{\mathrm{H},\ell}^{2(t)}$ from $f(s_{\mathrm{H},\ell}^{2}|\mathbf{U}^{(t-1)}, \mathbf{Y}_{\mathrm{H}})$,

end for

% Sampling the pseudo spectral response

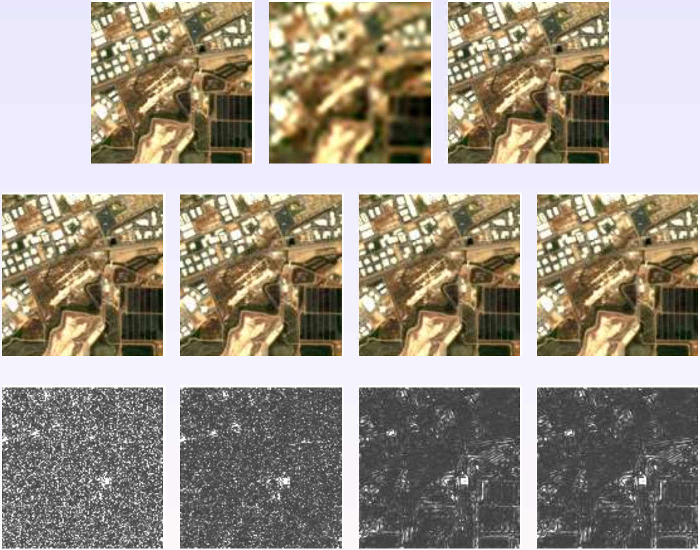
Sample $\underline{\mathbf{R}}$ from $f(\underline{\mathbf{R}}|\mathbf{U}^{(t-1)}, s_{\mathrm{M}}^{2\,(t)}, \mathbf{Y}_{\mathrm{M}})$[8]

% Sampling the high-resolved image

Sample $\mathbf{U}^{(t)}$ using a Hamiltonian Monte Carlo algorithm

end for

[8] Q. Wei *et al.*, "Bayesian fusion of multispectral and hyperspectral images with unknown sensor spectral response", in *ICIP*, Paris, France, Oct. 2014.



(a) MAP (b) Wavelet MAP (c) MCMC with known **R** (d) MCMC with unknown **R**

Quantitative fusion results

Table: Performance of the compared fusion methods: RSNR (in dB), UIQI, SAM (in degree), ERGAS, DD (in 10^{-2}) and Time (in second)(AVIRIS dataset).

Methods	RSNR	UIQI	SAM	ERGAS	DD	Time
MAP	16.655	0.9336	5.739	3.930	2.354	3
Wavelet MAP	19.501	0.9626	4.186	2.897	1.698	73
MCMC with known **R**	21.913	0.9771	3.094	2.231	1.238	8811
MCMC with mismatched **R**[9]	21.804	0.9764	3.130	2.260	1.257	8388
MCMC with unknown **R**	21.897	0.9769	3.101	2.234	1.244	10471

[9] FSNR= 10dB.

Outline

The negative logarithm of the joint posterior distribution $p\left(\boldsymbol{\theta}, \boldsymbol{\Sigma}|\boldsymbol{\mathcal{Y}}\right)$ is given as

$$\begin{aligned}
&L(\mathbf{U}, \boldsymbol{s}^2, \boldsymbol{\Sigma}) \\
&= -\log p\left(\boldsymbol{\theta}, \boldsymbol{\Sigma}|\boldsymbol{\mathcal{Y}}\right) \\
&= -\log p\left(\mathbf{Y}_{\mathrm{H}}|\boldsymbol{\theta}\right) - \log p\left(\mathbf{Y}_{\mathrm{M}}|\boldsymbol{\theta}\right) - \sum_{l=1}^{n} \log p\left(\boldsymbol{u}_l|\boldsymbol{\Sigma}\right) \\
&- \sum_{i=1}^{m_\lambda} \log p\left(s^2_{\mathrm{H},i}\right) - \sum_{j=1}^{n_\lambda} \log p\left(s^2_{\mathrm{M},j}\right) - \log p\left(\boldsymbol{\Sigma}\right) - C
\end{aligned}$$

- MAP estimator: minimizing the function $L(\mathbf{U}, \boldsymbol{s}^2, \boldsymbol{\Sigma})$ with respect to $\mathbf{U}$, $\boldsymbol{s}^2$ and $\boldsymbol{\Sigma}$ iteratively
- use a Block coordinated descent (BCD) algorithm [10]

[10] Q. Wei *et al.*, “Bayesian fusion of multispectral and hyperspectral images using a block coordinate descent Method”, in *IEEE GRSS Workshop on Hyperspectral Image and SIgnal Processing: Evolution in Remote Sensing (WHISPERS)*, Tokyo, Japan, Jun. 2015.

Block Coordinated Descent for HS and MS image fusion

Input: $\mathbf{Y}_{\mathrm{H}}$, $\mathbf{Y}_{\mathrm{M}}$, $\widetilde{m}_\lambda$, $\mathbf{B}$, $\mathbf{S}$, $\mathbf{R}$, $\boldsymbol{s}_0^2$, $\boldsymbol{\Sigma}_0$

- $\hat{\mathbf{H}} \leftarrow \mathrm{PCA}(\mathbf{Y}_{\mathrm{H}}, \widetilde{m}_\lambda)$; `/* Subspace transform matrix */`

for $t = 1, 2, \ldots$ **to** *stopping rule* **do**

$\mathbf{U}_t = \arg\min_{\mathbf{U}} L(\mathbf{U}, \boldsymbol{s}_{t-1}^2, \boldsymbol{\Sigma}_{t-1})$; `/* Optimize w.r.t.` $\mathbf{U}$ `*/`

$\boldsymbol{s}_t^2 = \arg\min_{\boldsymbol{s}^2} L(\mathbf{U}_t, \boldsymbol{s}^2, \boldsymbol{\Sigma}_{t-1})$; `/* Optimize w.r.t.` $\boldsymbol{s}^2$ `*/`

$\boldsymbol{\Sigma}_t = \arg\min_{\boldsymbol{\Sigma}} L(\mathbf{U}_t, \boldsymbol{s}_t^2, \boldsymbol{\Sigma})$; `/* Optimize w.r.t.` $\boldsymbol{\Sigma}$ `*/`

end

Output: $\hat{\mathbf{X}} = \hat{\mathbf{H}}\hat{\mathbf{U}}$ (High resolution HS image)

Remarks

The convergence is guaranteed[11].

[11] D. P. Bertsekas. Nonlinear Programming. Athena Scientific Belmont, 1999.

Minimization w.r.t. U

Using the linear model, dimensionality reduction, fusing the HS and MS images can be formulated as finding **U** minimizing the cost function

$$L_{\mathbf{U}}(\mathbf{U}) = \frac{1}{2}\|\boldsymbol{\Lambda}_{\mathrm{H}}^{-\frac{1}{2}}(\mathbf{Y}_{\mathrm{H}} - \mathbf{HUBS})\|_F^2 + \frac{1}{2}\|\boldsymbol{\Lambda}_{\mathrm{M}}^{-\frac{1}{2}}(\mathbf{Y}_{\mathrm{M}} - \mathbf{RHU})\|_F^2 + \frac{1}{2}\|\boldsymbol{\Sigma}^{-\frac{1}{2}}(\mathbf{U} - \boldsymbol{\mu}_{\mathbf{U}})\|_F^2.$$

- First two terms: data fidelity terms for the HS+MS images (likelihoods)
- Last term: penalty ensuring appropriate regularization (prior)

Difficulties

- Large dimensionality of **U**
- Diagonalization of the linear operators $\mathbf{H}(\cdot)\mathbf{BS}$ not possible

Alternating Direction Method of Multipliers (ADMM)

Idea: transform the unconstrained optimization with respect to **U** into a constrained one via a variable splitting "trick", and then attack this constrained problem using an augmented Lagrangian (AL) method[12]

- Splittings: $\mathbf{H}_1 = \mathbf{UB}$, $\mathbf{H}_2 = \mathbf{U}$ and $\mathbf{H}_3 = \mathbf{U}$
- Respective scaled dual variable: $\mathbf{G}_1, \mathbf{G}_2, \mathbf{G}_3$

$$
\begin{aligned}
& L(\mathbf{U}, \mathbf{H}_1, \mathbf{H}_2, \mathbf{H}_3, \mathbf{G}_1, \mathbf{G}_2, \mathbf{G}_3) \qquad \boxed{\textit{deconvolution}} \\
= \quad & \frac{1}{2} \left\| \Lambda_{\mathrm{H}}^{-\frac{1}{2}} (\mathbf{Y}_{\mathrm{H}} - \mathbf{H}\mathbf{H}_1 \mathbf{S}) \right\|_F^2 + \boxed{\frac{\mu}{2} \left\| \mathbf{UB} - \mathbf{H}_1 - \mathbf{G}_1 \right\|_F^2} + \\
& \frac{1}{2} \left\| \Lambda_{\mathrm{M}}^{-\frac{1}{2}} (\mathbf{Y}_{\mathrm{M}} - \mathbf{R}\mathbf{H}\mathbf{H}_2) \right\|_F^2 + \boxed{\frac{\mu}{2} \left\| \mathbf{U} - \mathbf{H}_2 - \mathbf{G}_2 \right\|_F^2} + \\
& \frac{1}{2} \left\| \Sigma^{-\frac{1}{2}} (\mu_{\mathbf{U}} - \mathbf{H}_3) \right\|_F^2 + \boxed{\frac{\mu}{2} \left\| \mathbf{U} - \mathbf{H}_3 - \mathbf{G}_3 \right\|_F^2}
\end{aligned}
$$

[12] M. Afonso *et al.*, "An augmented Lagrangian approach to the constrained optimization formulation of imaging inverse problems," *IEEE Trans. Image Process.*, vol. 20, no. 3,pp. 681-695, 2011.

Alternating Direction Method of Multipliers (ADMM)

Idea: transform the unconstrained optimization with respect to $\mathbf{U}$ into a constrained one via a variable splitting "trick", and then attack this constrained problem using an augmented Lagrangian (AL) method

- Splittings: $\mathbf{H}_1 = \mathbf{UB}$, $\mathbf{H}_2 = \mathbf{U}$ and $\mathbf{H}_3 = \mathbf{U}$
- Respective scaled dual variable: $\mathbf{G}_1, \mathbf{G}_2, \mathbf{G}_3$

upsampling

$$
\begin{aligned}
&L(\mathbf{U}, \mathbf{H}_1, \mathbf{H}_2, \mathbf{H}_3, \mathbf{G}_1, \mathbf{G}_2, \mathbf{G}_3) \\
= \; & \boxed{\frac{1}{2}\left\|\Lambda_{\mathrm{H}}^{-\frac{1}{2}}(\mathbf{Y}_{\mathrm{H}} - \mathbf{H}\mathbf{H}_1\mathbf{S})\right\|_F^2 + \frac{\mu}{2}\left\|\mathbf{UB} - \mathbf{H}_1 - \mathbf{G}_1\right\|_F^2} + \\
& \frac{1}{2}\left\|\Lambda_{\mathrm{M}}^{-\frac{1}{2}}(\mathbf{Y}_{\mathrm{M}} - \mathbf{R}\mathbf{H}\mathbf{H}_2)\right\|_F^2 + \frac{\mu}{2}\left\|\mathbf{U} - \mathbf{H}_2 - \mathbf{G}_2\right\|_F^2 + \\
& \frac{1}{2}\left\|\Sigma^{-\frac{1}{2}}\left(\mu_{\mathbf{U}} - \mathbf{H}_3\right)\right\|_F^2 + \frac{\mu}{2}\left\|\mathbf{U} - \mathbf{H}_3 - \mathbf{G}_3\right\|_F^2
\end{aligned}
$$

Alternating Direction Method of Multipliers (ADMM)
Idea: transform the unconstrained optimization with respect to $\mathbf{U}$ into a constrained one via a variable splitting "trick", and then attack this constrained problem using an augmented Lagrangian (AL) method

- Splittings: $\mathbf{H}_1 = \mathbf{U}\mathbf{B}$, $\mathbf{H}_2 = \mathbf{U}$ and $\mathbf{H}_3 = \mathbf{U}$
- Respective scaled dual variable: $\mathbf{G}_1, \mathbf{G}_2, \mathbf{G}_3$

spectral unmixing

$$
\begin{aligned}
&L(\mathbf{U}, \mathbf{H}_1, \mathbf{H}_2, \mathbf{H}_3, \mathbf{G}_1, \mathbf{G}_2, \mathbf{G}_3) \\
= \quad & \frac{1}{2}\left\|\Lambda_{\mathrm{H}}^{-\frac{1}{2}}(\mathbf{Y}_{\mathrm{H}} - \mathbf{H}\mathbf{H}_1\mathbf{S})\right\|_F^2 + \frac{\mu}{2}\left\|\mathbf{U}\mathbf{B} - \mathbf{H}_1 - \mathbf{G}_1\right\|_F^2 + \\
& \boxed{\frac{1}{2}\left\|\Lambda_{\mathrm{M}}^{-\frac{1}{2}}(\mathbf{Y}_{\mathrm{M}} - \mathbf{R}\mathbf{H}\mathbf{H}_2)\right\|_F^2 + \frac{\mu}{2}\left\|\mathbf{U} - \mathbf{H}_2 - \mathbf{G}_2\right\|_F^2} + \\
& \frac{1}{2}\left\|\boldsymbol{\Sigma}^{-\frac{1}{2}}\left(\boldsymbol{\mu}_{\mathbf{U}} - \mathbf{H}_3\right)\right\|_F^2 + \frac{\mu}{2}\left\|\mathbf{U} - \mathbf{H}_3 - \mathbf{G}_3\right\|_F^2
\end{aligned}
$$

Alternating Direction Method of Multipliers (ADMM)
Idea: transform the unconstrained optimization with respect to $\mathbf{U}$ into a constrained one via a variable splitting "trick", and then attack this constrained problem using an augmented Lagrangian (AL) method

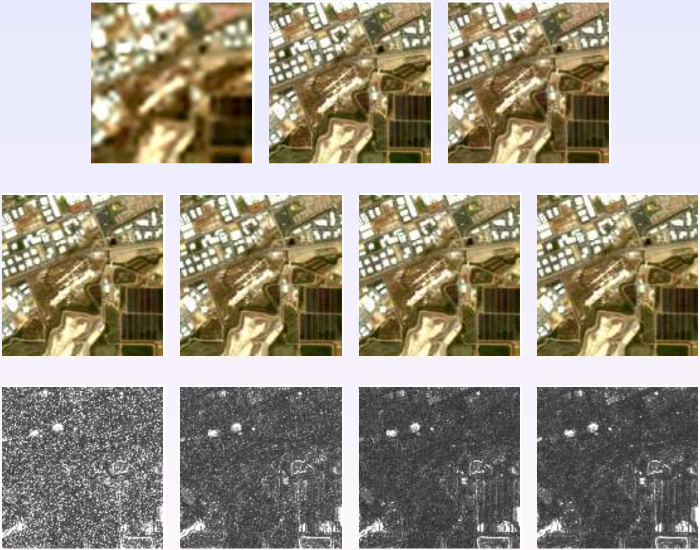
- Splittings: $\mathbf{H}_1 = \mathbf{U}\mathbf{B}$, $\mathbf{H}_2 = \mathbf{U}$ and $\mathbf{H}_3 = \mathbf{U}$
- Respective scaled dual variable: $\mathbf{G}_1, \mathbf{G}_2, \mathbf{G}_3$

$$L(\mathbf{U}, \mathbf{H}_1, \mathbf{H}_2, \mathbf{H}_3, \mathbf{G}_1, \mathbf{G}_2, \mathbf{G}_3) \qquad \boxed{\textit{denoising}}$$

$$= \frac{1}{2}\left\|\Lambda_{\mathrm{H}}^{-\frac{1}{2}}(\mathbf{Y}_{\mathrm{H}} - \mathbf{H}\mathbf{H}_1\mathbf{S})\right\|_F^2 + \frac{\mu}{2}\left\|\mathbf{U}\mathbf{B} - \mathbf{H}_1 - \mathbf{G}_1\right\|_F^2 +$$

$$\frac{1}{2}\left\|\Lambda_{\mathrm{M}}^{-\frac{1}{2}}(\mathbf{Y}_{\mathrm{M}} - \mathbf{R}\mathbf{H}\mathbf{H}_2)\right\|_F^2 + \frac{\mu}{2}\left\|\mathbf{U} - \mathbf{H}_2 - \mathbf{G}_2\right\|_F^2 +$$

$$\boxed{\frac{1}{2}\left\|\boldsymbol{\Sigma}^{-\frac{1}{2}}\left(\mu_{\mathbf{U}} - \mathbf{H}_3\right)\right\|_F^2 + \frac{\mu}{2}\left\|\mathbf{U} - \mathbf{H}_3 - \mathbf{G}_3\right\|_F^2}$$



(a) MAP (b) Wavelet MAP (c) MCMC (d) BCD

Table: Performance of the compared fusion methods: RSNR (in dB), UIQI, SAM (in degree), ERGAS, DD (in 10^{-2}) and time (in second) (AVIRIS dataset).

Methods	RSNR	UIQI	SAM	ERGAS	DD	Time
MAP	23.14	0.9932	5.147	3.524	4.958	3
Wavelet MAP	24.91	0.9956	4.225	3.282	4.120	72
MCMC	25.92	0.9971	3.733	2.926	3.596	6228
Proposed	25.85	0.9970	3.738	2.946	3.600	96

- Promising results for the considered quality measures
- Significant reduction in computation time: Save a lot of time!

Outline

Sparse Regularization

Motivation

Self-similarity property of natural image patches

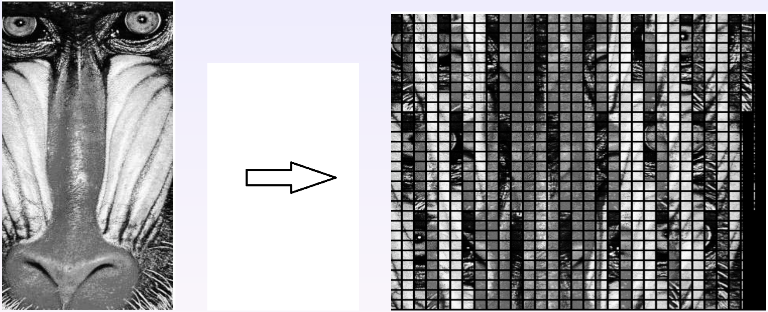


image patches

Remote Sensing Image

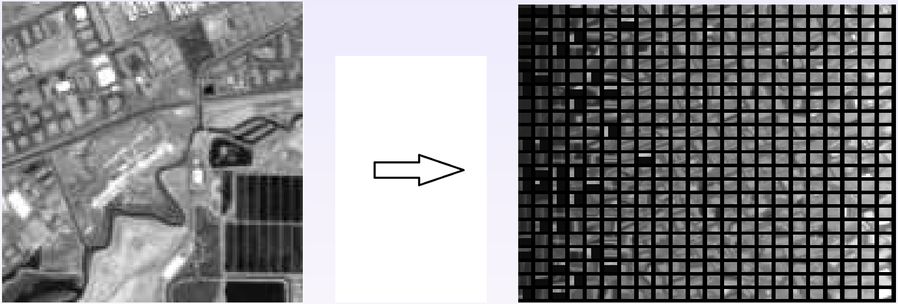
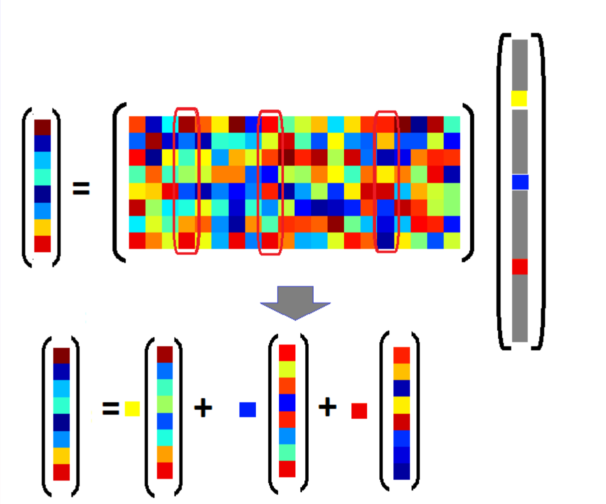


image patches

Sparse Regularization

The patches of the target image **U** can be sparsely approximated on an over-complete dictionary (with columns referred to as atoms).



Penalized inverse problem

Based on the linear model and dimensionality reduction, fusing the HS and MS images can be formulated as the following inverse problem:

$$\min_{\mathbf{U}} \underbrace{\frac{1}{2}\left\|\Lambda_{\mathrm{H}}^{-\frac{1}{2}}(\mathbf{Y}_{\mathrm{H}} - \mathbf{HUBS})\right\|_F^2}_{\substack{\text{HS data term} \\ \propto \ln p(\mathbf{Y}_{\mathrm{H}}|\mathbf{U})}} + \underbrace{\frac{1}{2}\left\|\Lambda_{\mathrm{H}}^{-\frac{1}{2}}(\mathbf{Y}_{\mathrm{M}} - \mathbf{RHU})\right\|_F^2}_{\substack{\text{MS data term} \\ \propto \ln p(\mathbf{Y}_{\mathrm{M}}|\mathbf{U})}} + \underbrace{\lambda \phi(\mathbf{U})}_{\substack{\text{regularizer} \\ \propto \ln p(\mathbf{U})}},$$

Sparse Regularization

Regularizer

$$\phi(\mathbf{U}) = \frac{1}{2}\left\|\mathbf{U} - \bar{\mathbf{U}}\left(\mathbf{D}, \mathbf{A}\right)\right\|_F^2$$

Separating each band of the target image leads to

$$\phi(\mathbf{U}) = \frac{1}{2}\sum_{i=1}^{\widetilde{m}_\lambda}\left\|\mathbf{U}_i - \mathcal{P}\left(\mathbf{D}_i\mathbf{A}_i\right)\right\|_F^2$$

- $\mathbf{U}_i \in \mathbb{R}^n$ is the ith band (or row) of $\mathbf{U} \in \mathbb{R}^{\widetilde{m}_\lambda \times n}$
- $\mathbf{D}_i \in \mathbb{R}^{n_\mathrm{p} \times n_\mathrm{at}}$ is the dictionary dedicated to the ith band of $\mathbf{U}$ (n_p is the patch size and n_at is the number of atoms) and $\mathbf{D} = \left[\mathbf{D}_1, \cdots, \mathbf{D}_{\widetilde{m}_\lambda}\right]$
- $\mathbf{A}_i \in \mathbb{R}^{n_\mathrm{at} \times n_\mathrm{pat}}$ is the ith band's code (n_pat is the number of patches associated with the ith band) and $\mathbf{A} = \left[\mathbf{A}_1, \cdots, \mathbf{A}_{\widetilde{m}_\lambda}\right]$
- $\mathcal{P}(\cdot)$ is a linear operator that averages the overlapping patches of each band to restore the target image

How can we obtain the dictionary **D** and the code **A**?

Dictionary Learning and Sparse Coding

Dictionary Learning

Learn[13] the set of over-complete dictionaries $\mathbf{D} = \left[\mathbf{D}_1, \cdots, \mathbf{D}_{\widetilde{m}_\lambda}\right]$: applying a DL algorithm on the rough estimation of $\mathbf{U}$ (constructed from the MS and HS images)

- K-SVD method
- Online Dictionary Learning (ODL) method

Sparse Coding

- Orthogonal Matching Pursuit (OMP): to estimate the sparse code $\mathbf{A}_i$ (with $n_{\max}$ coefficients) for each band $\mathbf{U}_i$
- Support ($\mathbf{\Omega}_i \subset \mathbb{N}^2, i = 1, \cdots, \widetilde{m}_\lambda$): The positions of the non-zero elements of the code $\mathbf{A}_i$ are also identified

[13] M. Elad *et al.*, "Image denoising via sparse and redundant representations over learned dictionaries," *IEEE Trans. Image Process.*, vol. 15, no. 12, pp. 3736–3745, 2006.

Re-estimation of the sparse code

Inspired by hierarchical models frequently encountered in Bayesian inference, we propose to include the code **A** within the estimation process.

$$\phi(\mathbf{U}, \mathbf{A}) = \frac{1}{2} \sum_{i=1}^{\widetilde{m}_\lambda} \left\| \mathbf{U}_i - \mathcal{P}\left(\mathbf{D}_i \mathbf{A}_i\right) \right\|_F^2 + \mu_a \left\| \mathbf{A}_i \right\|_0$$

NP hard!

where $\|.\|_0$ is the ℓ_0 counting function (or ℓ_0 norm) and μ_a is a regularization parameter.

By fixing the supports $\mathbf{\Omega}_i$, the ℓ_0 norm reduces to a constant. Hence,

$$\phi(\mathbf{U}, \mathbf{A}) = \frac{1}{2} \sum_{i=1}^{\widetilde{m}_\lambda} \left\| \mathbf{U}_i - \mathcal{P}\left(\mathbf{D}_i \mathbf{A}_i\right) \right\|_F^2 \text{ s.t. } \mathbf{A}_{i,\setminus \mathbf{\Omega}_i} = 0$$

where $\mathbf{A}_{i,\setminus \mathbf{\Omega}_i} = \{\mathbf{A}_i(l,k) \mid (l,k) \notin \mathbf{\Omega}_i\}$.

Final Optimization Problem

Joint optimization with respect to **U** and **A**

$$\min_{\mathbf{U},\mathbf{A}} L(\mathbf{U},\mathbf{A}) = \frac{1}{2}\left\|\Lambda_{\mathrm{H}}^{-\frac{1}{2}}(\mathbf{Y}_{\mathrm{H}} - \mathbf{HUBS})\right\|_F^2 + \frac{1}{2}\left\|\Lambda_{\mathrm{M}}^{-\frac{1}{2}}\mathbf{Y}_{\mathrm{M}} - \mathbf{RHU}\right\|_F^2 + \frac{\lambda}{2}\sum_{i=1}^{\widetilde{m}_\lambda}\left(\left\|\mathbf{U}_i - \mathcal{P}\left(\mathbf{D}_i\mathbf{A}_i\right)\right\|_F^2\right), \text{ s.t. } \mathbf{A}_{i,\setminus\boldsymbol{\Omega}_i} = 0$$

Solution

- Solved by minimizing w.r.t. **U** and **A** alternatively
- Each sub-problem is strictly convex

Outline

Optimization with respect to U

$$\min_{\mathbf{U}} L(\mathbf{U}) = \frac{1}{2}\left\|\mathbf{\Lambda}_{\mathrm{H}}^{-\frac{1}{2}}\left(\mathbf{Y}_{\mathrm{H}} - \mathbf{HUBS}\right)\right\|_F^2 + \frac{1}{2}\left\|\mathbf{\Lambda}_{\mathrm{M}}^{-\frac{1}{2}}\left(\mathbf{Y}_{\mathrm{M}} - \mathbf{RHU}\right)\right\|_F^2 + \frac{\lambda}{2}\sum_{i=1}^{\widetilde{m}_\lambda}\left(\left\|\mathbf{U}_i - \mathcal{P}\left(\mathbf{D}_i\mathbf{A}_i\right)\right\|_F^2\right),$$

Difficulties

- Large dimensionality of **U**
- Diagonalization of the linear operators $\mathbf{H}(\cdot)\mathbf{BS}$ and $\mathcal{P}(\cdot)$ not possible

Solution

Alternating Direction Method of Multipliers (ADMM)

Optimization with respect to A

Optimization with respect to $\mathbf{A}_i$ $(i = 1, \cdots, \widetilde{m}_\lambda)$ conditional on $\mathbf{U}_i$

$$\min_{\mathbf{A}_i} \left\| \mathbf{U}_i - \mathcal{P}(\mathbf{D}_i \mathbf{A}_i) \right\|_F^2 \text{ s.t. } \mathbf{A}_{i,\setminus \boldsymbol{\Omega}_i} = 0$$

Remarks

- The optimization with respect to $\mathbf{A}_i$ considers only the non-zero elements of $\mathbf{A}_i$, denoted as $\mathbf{A}_{i,\boldsymbol{\Omega}_i} = \{\mathbf{A}_i(l,k) \mid (l,k) \in \boldsymbol{\Omega}_i\}$
- Standard least square (LS) problem which can be solved analytically

Alternate Optimization Scheme[14]

Input: $\mathbf{Y}_{\mathrm{H}}$, $\mathbf{Y}_{\mathrm{M}}$, $\mathbf{B}$, $\mathbf{S}$, $\mathbf{R}$, $\mathrm{SNR}_{\mathrm{H}}$, $\mathrm{SNR}_{\mathrm{M}}$, $\widetilde{m}_{\lambda}$, $n_{\max}$

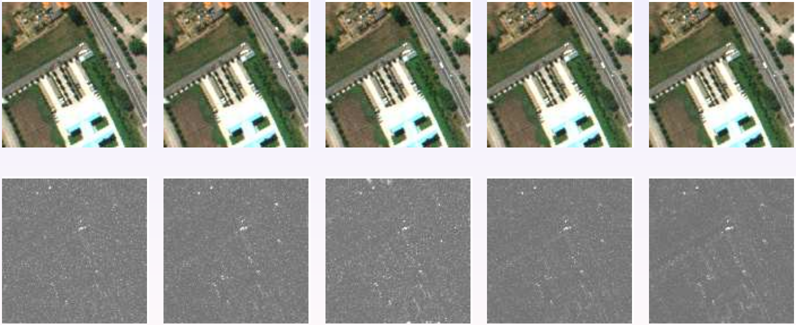
- Approximate $\bar{\mathbf{U}}$ using $\mathbf{Y}_{\mathrm{M}}$ and $\mathbf{Y}_{\mathrm{H}}$ /* *Rough estimation of* $\mathbf{U}$ */
- $\hat{\mathbf{D}} \leftarrow \mathrm{ODL}(\bar{\mathbf{U}})$ /* *Online dictionary learning* */
- $\hat{\mathbf{A}} \leftarrow \mathrm{OMP}(\hat{\mathbf{D}}, \bar{\mathbf{U}}, n_{\max})$ /* *Sparse coding* */
- $\hat{\Omega} \leftarrow \hat{\mathbf{A}} \neq 0$ /* *Computing support* */
- $\hat{\mathbf{H}} \leftarrow \mathrm{PCA}(\mathbf{Y}_{\mathrm{H}}, \widetilde{m}_{\lambda})$ /* *Computing subspace transform matrix* */

/* *Start alternate optimization* */
for $t = 1, 2, \ldots$ **to** *stopping rule* **do**
  $\hat{\mathbf{U}}_t \in \{\mathbf{U} : L(\mathbf{U}, \hat{\mathbf{A}}_{t-1}) \leq L(\hat{\mathbf{U}}_{t-1}, \hat{\mathbf{A}}_{t-1})\}$ /* *solved with ADMM* */
  $\hat{\mathbf{A}}_t \in \{\mathbf{A} : L(\hat{\mathbf{U}}_t, \mathbf{A}) \leq L(\hat{\mathbf{U}}_t, \hat{\mathbf{A}}_{t-1})\}$ /* *solved with LS* */
end
$\hat{\mathbf{X}} = \hat{\mathbf{H}}\hat{\mathbf{U}}$
Output: $\hat{\mathbf{X}}$ (high resolution HS image)

[14] Q. Wei *et al.*, "Hyperspectral and multispectral image fusion based on a sparse representation", *IEEE Trans. Geosci. and Remote Sens.*, vol. 53, no. 7, pp. 3658-3668, July 2015.

Qualitative results (Pavia dataset)

(a) Ref (b) HS (c) MS



(d) MAP (e) Wavelet (f) CNMF (g) Gaussian (h) Sparse

Quantitative results (Pavia dataset)

Table: Performance of different MS + HS fusion methods (Pavia dataset): RMSE (in 10^{-2}), UIQI, SAM (in degree), ERGAS, DD (in 10^{-3}) and Time (in second).

Methods	RMSE	UIQI	SAM	ERGAS	DD	Time
MAP	1.148	0.9875	1.962	1.029	8.666	3
Wavelet MAP	1.099	0.9885	1.849	0.994	8.349	75
CNMF	1.119	0.9857	2.039	1.089	9.007	14
Gaussian	1.011	0.9903	1.653	0.911	7.598	6003
Sparse	0.947	0.9913	1.492	0.850	7.010	282

The proposed method provides promising results for the considered quality measures.

Outline

Transforming optimization to solving a Sylvester Equation

Forward model

$$\mathbf{Y}_{\mathrm{H}} = \mathbf{XBS} + \mathbf{N}_{\mathrm{H}}, \quad \mathbf{Y}_{\mathrm{M}} = \mathbf{RX} + \mathbf{N}_{\mathrm{M}}$$

$$\text{s.t. } \mathbf{X} = \mathbf{HU}$$

Negative log-likelihood (in subspace)

$$\begin{aligned} -\log p(\mathcal{Y}|\mathbf{U}) &= -\log p(\mathbf{Y}_{\mathrm{H}}|\mathbf{U}) - \log p(\mathbf{Y}_{\mathrm{M}}|\mathbf{U}) + C \\ &= \tfrac{1}{2}\|\Lambda_{\mathrm{H}}^{-\frac{1}{2}}(\mathbf{Y}_{\mathrm{H}} - \mathbf{HUBS})\|^2 + \tfrac{1}{2}\|\Lambda_{\mathrm{M}}^{-\frac{1}{2}}(\mathbf{Y}_{\mathrm{M}} - \mathbf{RHU})\|^2 + C \end{aligned}$$

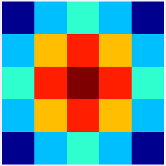
Minimizing the likelihood w.r.t. $\mathbf{U}$ $\Leftrightarrow$ solve a generalized Sylvester matrix equation

$$\left[\mathbf{H}^H \Lambda_{\mathrm{H}}^{-1} \mathbf{H}\right] \mathbf{U} \left[\mathbf{BS}(\mathbf{BS})^H\right] + \left[(\mathbf{RH})^H \Lambda_{\mathrm{M}}^{-1} (\mathbf{RH})\right] \mathbf{U} = \text{term ind. on } \mathbf{U}$$

- $\mathbf{BS}(\mathbf{BS})^H$ is not diagonalizable!

Assumption 1

The blurring matrix $\mathbf{B}$ *is a block circulant matrix with circulant blocks (BCCB).*



Assumption 2

The decimation matrix $\mathbf{S}$ *corresponds to downsampling the original signal and its conjugate transpose* $\mathbf{S}^H$ *interpolates the decimated signal with zeros, e.g.,*

$$\mathbf{S} = \begin{bmatrix} 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 1 \end{bmatrix}$$

These two assumptions are used to compute an explicit solution of the Sylvester equation.

$$\left[\mathbf{H}^H \Lambda_{\mathrm{H}}^{-1} \mathbf{H}\right] \mathbf{U} \left[\mathbf{BS}\left(\mathbf{BS}\right)^H\right] + \left[\left(\mathbf{RH}\right)^H \Lambda_{\mathrm{M}}^{-1} \left(\mathbf{RH}\right)\right] \mathbf{U} = \text{term ind. on } \mathbf{U}$$

3 Main Steps

- Left multiply by $\left(\mathbf{H}^H \Lambda_{\mathrm{H}}^{-1} \mathbf{H}\right)^{-1}$: $\mathbf{UC_2} + \mathbf{C_1 U} = \mathbf{C_3}$, where $\mathbf{C_2} = \mathbf{BS}\left(\mathbf{BS}\right)^{\mathbf{H}}$.

> *Lemma 1*
> *The equality* $\mathbf{F}^H \underline{\mathbf{S}} \mathbf{F} = \frac{1}{d} \mathbf{J}_d \otimes \boldsymbol{I}_m$ *holds, where* $\underline{\mathbf{S}} = \mathbf{S}\mathbf{S}^H$, $\mathbf{J}_d$ *is the* $d \times d$ *matrix of ones and* $\boldsymbol{I}_m$ *is the* $m \times m$ *identity matrix.*

- Diagonalize $\mathbf{C_1}$ and use Lemma 1 to simplify $\mathbf{C}_2$:

$$\bar{\mathbf{U}}\mathbf{M} + \Lambda_C \bar{\mathbf{U}} = \bar{\mathbf{C}}_3$$

with a diagonal matrix Λ_C and $\mathbf{M} = \frac{1}{d} \begin{bmatrix} \sum_{i=1}^{d} \underline{\mathbf{D}}_i & \underline{\mathbf{D}}_2 & \cdots & \underline{\mathbf{D}}_d \\ \mathbf{0} & \mathbf{0} & \cdots & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \cdots & \mathbf{0} \end{bmatrix}$

$\underline{\mathbf{D}}_i$: $m \times m$ diagonal matrix, d: downsampling ratio, m: number of image pixels

Theorem 2

[a] *Let* $(\bar{\mathbf{C}}_3)_{l,j}$ *denotes the* j*th block of the* l*th band of* $\bar{\mathbf{C}}_3$ *for any* $l = 1, \cdots, \widetilde{m}_\lambda$*. Then, the solution* $\bar{\mathbf{U}}$ *of the SE can be decomposed as*

$$\bar{\mathbf{U}} = \begin{bmatrix} \bar{\mathbf{u}}_{1,1} & \bar{\mathbf{u}}_{1,2} & \cdots & \bar{\mathbf{u}}_{1,d} \\ \bar{\mathbf{u}}_{2,1} & \bar{\mathbf{u}}_{2,2} & \cdots & \bar{\mathbf{u}}_{2,d} \\ \vdots & \vdots & \ddots & \vdots \\ \bar{\mathbf{u}}_{\widetilde{m}_\lambda,1} & \bar{\mathbf{u}}_{\widetilde{m}_\lambda,2} & \cdots & \bar{\mathbf{u}}_{\widetilde{m}_\lambda,d} \end{bmatrix}$$

with

$$\bar{\mathbf{u}}_{l,j} = \begin{cases} (\bar{\mathbf{C}}_3)_{l,j} \left(\frac{1}{d} \sum_{i=1}^{d} \underline{\mathbf{D}}_i + \lambda_C^l \boldsymbol{I}_m \right)^{-1}, & j = 1, \\ \frac{1}{\lambda_C^l} \left[(\bar{\mathbf{C}}_3)_{l,j} - \frac{1}{d} \bar{\mathbf{u}}_{l,1} \underline{\mathbf{D}}_j \right], & j = 2, \cdots, d. \end{cases}$$

[a] Q. Wei *et al.*, “Fast multi-band image fusion based on solving a Sylvester equation”, *IEEE Trans. Image Process.*, vol. 24, no. 11, pp. 4109-4121, Nov. 2015.

Fast fUsion based on a Sylvester Equation (FUSE)

Input: $\mathbf{Y}_{\mathrm{M}}, \mathbf{Y}_{\mathrm{H}}, \boldsymbol{\Lambda}_{\mathrm{M}}, \boldsymbol{\Lambda}_{\mathrm{H}}, \mathbf{R}, \mathbf{B}, \mathbf{S}, \mathbf{H}$

- $\mathbf{D} \leftarrow \mathrm{Dec}\,(\mathbf{B})$ and $\underline{\mathbf{D}} = \mathbf{D}^{*}\mathbf{D}$ /*Circulant matrix: $\mathbf{B} = \mathbf{F}\mathbf{D}\mathbf{F}^{H}$ */
- $\mathbf{C}_1 \leftarrow \left(\mathbf{H}^{H}\boldsymbol{\Lambda}_{\mathrm{H}}^{-1}\mathbf{H}\right)^{-1}\left((\mathbf{R}\mathbf{H})^{H}\boldsymbol{\Lambda}_{\mathrm{L}}^{-1}\mathbf{R}\mathbf{H}\right);$
- $(\mathbf{Q}, \boldsymbol{\Lambda}_C) \leftarrow \mathrm{EigDec}\,(\mathbf{C}_1)$ /* Eigen-dec of $\mathbf{C}_1$: $\mathbf{C}_1 = \mathbf{Q}\boldsymbol{\Lambda}_C\mathbf{Q}^{-1}$ */
- $\bar{\mathbf{C}}_3 \leftarrow$
  $\mathbf{Q}^{-1}\left(\mathbf{H}^{H}\boldsymbol{\Lambda}_{\mathrm{H}}^{-1}\mathbf{H}\right)^{-1}(\mathbf{H}^{H}\boldsymbol{\Lambda}_{\mathrm{H}}^{-1}\mathbf{Y}_{\mathrm{H}}\,(\mathbf{B}\mathbf{S})^{H} + (\mathbf{R}\mathbf{H})^{H}\boldsymbol{\Lambda}_{\mathrm{L}}^{-1}\mathbf{Y}_{\mathrm{M}})\mathbf{B}\mathbf{F}\mathbf{P}^{-1};$

for $l = 1$ **to** $\widetilde{m}_{\lambda}$ **do**

$$\bar{\mathbf{u}}_{l,1} = (\bar{\mathbf{C}}_3)_{l,1}\left(\frac{1}{d}\sum_{i=1}^{d}\underline{\mathbf{D}}_i + \lambda_C^{l}\mathbf{I}_n\right)^{-1};$$

for $j = 2$ **to** d **do**

$$\bar{\mathbf{u}}_{l,j} = \frac{1}{\lambda_C^{l}}\left((\bar{\mathbf{C}}_3)_{l,j} - \frac{1}{d}\bar{\mathbf{u}}_{l,1}\underline{\mathbf{D}}_j\right);$$

end

end

Output: $\mathbf{X} = \mathbf{H}\mathbf{Q}\bar{\mathbf{U}}\mathbf{P}\mathbf{D}^{-1}\mathbf{F}^{H}$

Outline

Context

Gaussian prior modeling

Dictionary-based sparse prior modeling

Fast fUsion based on solving a Sylvester Equation (FUSE)
- From maximum likelihood estimator...
- ... to maximum a posteriori estimator

Conclusion

From ML to MAP estimators

Generalized to Bayesian estimators[15]

- $\phi(\mathbf{X})$: Gaussian prior based on interpolation[16]
- $\phi(\mathbf{X})$: Sparse representation based on dictionary learning[17]
- $\phi(\mathbf{X})$: Total variation (TV)[18]

---

[15] Q. Wei *et al.*, “Fast multi-band image fusion based on solving a Sylvester equation”, *IEEE Trans. Image Process.*, vol. 24, no. 11, pp. 4109-4121, Nov. 2015.

[16] Q. Wei *et al.*, “Bayesian fusion of multi-band images,” *IEEE J. Sel. Topics Signal Process.*, vol. 9, no. 6, pp. 1117-1127, Sept. 2015.

[17] Q. Wei *et al.*, “Hyperspectral and multispectral image fusion based on a sparse representation”, *IEEE Trans. Geosci. and Remote Sens.*, vol. 53, no. 7, pp. 3658-3668, July 2015.

[18] M. Simões *et al.*, “A convex formulation for hyperspectral image superresolution via subspace-based regularization”, *IEEE Trans. Geosci. and Remote Sens.*, vol. 53, no. 6, pp. 3373-3388, June 2015.

Gaussian prior

Gaussian prior: Sylvester equation embedded in BCD (FUSE-BCD)

Input: $\mathbf{Y}_{\mathrm{H}}$, $\mathbf{Y}_{\mathrm{M}}$, $\widetilde{m}_\lambda$, $\mathbf{B}$, $\mathbf{S}$, $\mathbf{R}$, s_0^2, Σ_0

- $\hat{\mathbf{H}} \leftarrow \mathrm{PCA}(\mathbf{Y}_{\mathrm{H}}, \widetilde{m}_\lambda)$; /* Subspace transform matrix */

for $t = 1, 2, \ldots$ **to** *stopping rule* **do**

$\mathbf{U}_t = \arg\min_{\mathbf{U}} L(\mathbf{U}, s_{t-1}^2, \Sigma_{t-1})$; /* Sylvester equation */

$s_t^2 = \arg\min_{s^2} L(\mathbf{U}_t, s^2, \Sigma_{t-1})$; /* Optimize w.r.t. s^2 */

$\Sigma_t = \arg\min_{\Sigma} L(\mathbf{U}_t, s_t^2, \Sigma)$; /* Optimize w.r.t. Σ */

end

Output: $\hat{\mathbf{X}} = \hat{\mathbf{H}}\hat{\mathbf{U}}$ (High resolution HS image)

Sparse representation

Sparse prior: Sylvester equation embedded in BCD (FUSE-BCD)

Input: $\mathbf{Y}_{\mathrm{H}}$, $\mathbf{Y}_{\mathrm{M}}$, $\mathbf{B}$, $\mathbf{S}$, $\mathbf{R}$, $\mathrm{SNR}_{\mathrm{H}}$, $\mathrm{SNR}_{\mathrm{M}}$, $\widetilde{m}_{\lambda}$, $n_{\max}$
Output: $\hat{\mathbf{X}}$ (high resolution HS image)

- Approximate $\bar{\mathbf{U}}$ using $\mathbf{Y}_{\mathrm{M}}$ and $\mathbf{Y}_{\mathrm{H}}$ /* *Rough estimation of* $\mathbf{U}$ */
- $\hat{\mathbf{D}} \leftarrow \mathrm{ODL}(\bar{\mathbf{U}})$ /* *Online dictionary learning* */
- $\hat{\mathbf{A}} \leftarrow \mathrm{OMP}(\hat{\mathbf{D}}, \bar{\mathbf{U}}, n_{\max})$ /* *Sparse coding* */
- $\hat{\Omega} \leftarrow \hat{\mathbf{A}} \neq 0$ /* *Computing support* */
- $\hat{\mathbf{H}} \leftarrow \mathrm{PCA}(\mathbf{Y}_{\mathrm{H}}, \widetilde{m}_{\lambda})$ /* *Computing subspace transform matrix* */

/* *Start alternate optimization* */
for $t = 1, 2, \ldots$ **to** *stopping rule* **do**
  $\hat{\mathbf{U}}_t \in \{\mathbf{U} : L(\mathbf{U}, \hat{\mathbf{A}}_{t-1}) \leq L(\hat{\mathbf{U}}_{t-1}, \hat{\mathbf{A}}_{t-1})\}$ /* *solved with SE* */
  $\hat{\mathbf{A}}_t \in \{\mathbf{A} : L(\hat{\mathbf{U}}_t, \mathbf{A}) \leq L(\hat{\mathbf{U}}_t, \hat{\mathbf{A}}_{t-1})\}$ /* *solved with LS* */
end
$\hat{\mathbf{X}} = \hat{\mathbf{H}}\hat{\mathbf{U}}$

Non-Gaussian prior

Non-Gaussian prior, such as (TV)[19]

$$\arg\min_{\mathbf{U}} \underbrace{\frac{1}{2}\|\Lambda_{\mathrm{H}}^{-\frac{1}{2}}\left(\mathbf{Y}_{\mathrm{H}}-\mathbf{HUBS}\right)\|_F^2}_{\text{HS data term}} + \underbrace{\frac{1}{2}\|\Lambda_{\mathrm{M}}^{-\frac{1}{2}}\left(\mathbf{Y}_{\mathrm{M}}-\mathbf{RHU}\right)\|_F^2}_{\text{MS data term}} + \underbrace{\lambda \mathrm{TV}\left(\mathbf{U}\right)}_{\text{regularizer}}.$$

can be equivalently solved as:

$$\arg\min_{\mathbf{U},\mathbf{V}} \frac{1}{2}\|\Lambda_{\mathrm{H}}^{-\frac{1}{2}}\left(\mathbf{Y}_{\mathrm{H}}-\mathbf{HUBS}\right)\|_F^2 + \frac{1}{2}\|\Lambda_{\mathrm{M}}^{-\frac{1}{2}}\left(\mathbf{Y}_{\mathrm{M}}-\mathbf{RHU}\right)\|_F^2 + \lambda \mathrm{TV}\left(\mathbf{V}\right)$$

s.t. $\mathbf{U} = \mathbf{V}$

- ADMM algorithm: Sylvester equation + proximity operator
- Sylvester equation embedded in ADMM (FUSE-ADMM)

[19] M. Simões *et al.*, “A convex formulation for hyperspectral image superresolution via subspace-based regularization”, *IEEE Trans. Geosci. and Remote Sens.*, vol. 53, no. 6, pp. 3373-3388, June 2015.

Performance and Computational Times

Table: Performance of HS+MS fusion methods: RSNR (in dB), UIQI, SAM (in degree), ERGAS, DD (in 10^{-3}) and time (in second).

Regularization	Methods	RSNR	UIQI	SAM	ERGAS	DD	Time
supervised naive Gaussian	ADMM	29.321	0.9906	1.555	0.888	7.115	126.83
	FUSE	29.372	0.9908	1.551	0.879	7.092	0.38
unsupervised naive Gaussian	ADMM-BCD	29.084	0.9902	1.615	0.913	7.341	99.55
	FUSE-BCD	29.077	0.9902	1.623	0.913	7.368	1.09
sparse representation	ADMM-BCD	29.582	0.9911	1.423	0.872	6.678	162.88
	FUSE-BCD	29.688	0.9913	1.431	0.856	6.672	73.66
TV	ADMM	29.473	0.9912	1.503	0.861	6.922	134.21
	FUSE-ADMM	29.631	0.9915	1.477	0.845	6.788	90.99

- The computational time is decreased significantly!

Table: Characteristics of the three datasets[20]

dataset	dimensions	spatial res	N	instrument
Moffett	PAN 185 × 395 HS 37 × 79	20m 100m	224	AVIRIS
Camargue	PAN 500 × 500 HS 100 × 100	4m 20m	125	HyMap
Garons	PAN 400 × 400 HS 80 × 80	4m 20m	125	HyMap

[20] L. Loncan, L. B. Almeida, J. M. Bioucas-Dias, X. Briottet, J. Chanussot, N. Dobigeon, S. Fabre, W. Liao, G. Licciardi, M. Simoes, J-Y. Tourneret, M. Veganzones, G. Vivone, Q. Wei and N. Yokoya, “Hyperspectral pansharpening: a review”, *IEEE Geosci. and Remote Sens. Mag.*, vol. 3, no. 3, pp. 27-46, Sept. 2015.

Figure: Camargue. (a) Ref, (b) interpolation, (c) SFIM, (d) MTF GLP HPM, (e) GSA, (f) PCA, (g) GFPCA, (h) CNMF, (i) Bayesian Sparse, (j) HySure.

Table: Quality measures for the Moffett field dataset[21]

method	CC	SAM	RMSE	ERGAS	Time(sec)
SFIM	0.92955	9.5271	365.2577	6.5429	1.26
MTF-GLP	0.93919	9.4599	352.1290	6.0491	1.86
MTF-GLP-HPM	0.93817	9.3567	354.8167	6.1992	1.71
GS	0.90521	14.1636	443.4351	7.5952	4.77
GSA	0.93857	11.2758	363.7090	6.2359	5.52
PCA	0.89580	14.6132	463.2204	7.9283	3.46
GFPCA	0.91614	11.3363	404.2979	7.0619	2.58
CNMF	0.95496	9.4177	314.4632	5.4200	10.98
Supervised Gaussian	0.97785	7.1308	220.0310	3.7807	1.31
Sparse represent.	0.98168	6.6392	200.3365	3.4281	133.61
HySure	0.97059	7.6351	254.2005	4.3582	140.05

[21] red: best green: second best blue: third best

Table: Quality measures for the Camargue dataset[22]

method	CC	SAM	RMSE	ERGAS	Time(sec)
SFIM	0.91886	4.2895	637.1451	3.4159	3.47
MTF-GLP	0.92397	4.3378	622.4711	3.2666	4.26
MTF-GLP-HPM	0.92599	4.2821	611.9161	3.2497	4.25
GS	0.91262	4.4982	665.0173	3.5490	8.29
GSA	0.92826	4.1950	587.1322	3.1940	8.73
PCA	0.90350	5.1637	710.3275	3.8680	8.92
GFPCA	0.89042	4.8472	745.6006	4.0001	8.51
CNMF	0.93000	4.4187	591.3190	3.1762	47.54
Supervised Gaussian	0.95195	3.6428	489.5634	2.6286	7.35
Sparse represent.	0.95882	3.3345	448.1721	2.4712	485.13
HySure	0.94650	3.8767	511.8525	2.8181	296.27

[22] red: best green: second best blue: third best

Table: Quality measures for the Garons dataset[23]

method	**CC**	**SAM**	**RMSE**	**ERGAS**	**Time**(sec)
SFIM	0.77052	6.7356	1036.4695	5.1702	2.74
MTF-GLP	0.80124	6.6155	956.3047	4.8245	4.00
MTF-GLP-HPM	0.79989	6.6905	962.1076	4.8280	2.98
GS	0.80347	6.6627	1037.6446	5.1373	5.56
GSA	0.80717	6.7719	928.6229	4.7076	5.99
PCA	0.81452	6.6343	1021.8547	5.0166	6.09
GFPCA	0.63390	7.4415	1312.0373	6.3416	4.36
CNMF	0.82993	6.9522	893.9194	4.4927	23.98
Supervised Gaussian	0.86857	5.8749	784.1298	3.9147	3.07
Sparse represent.	0.87834	5.6377	750.3510	3.7629	259.44
HySure	0.86080	6.0224	778.1051	4.0454	177.60

[23] red: best green: second best blue: third best

Conclusions

- Fusion of multi band images formulated as a linear inverse problem, that exploits explicitly the forward model
- Constrain the estimated image in a lower-dimensional space
- Definition of multiple priors within a (hierarchical) Bayesian framework
  - Gaussian prior
  - Sparse prior from dictionary learning
- Estimation of noise variances is possible with the proposed algorithm
- The spectral response of the MS image can be included in the estimation at the price of a higher computational complexity
- Toward fast fusion by using a closed-form solution of Sylvester equation: can be generalized to various priors

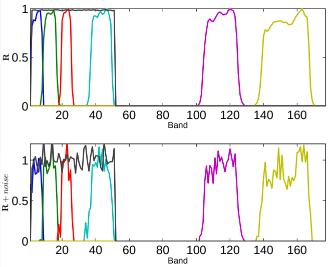
Ongoing work

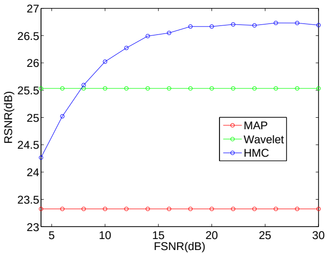
- Forward model
  - joint estimation of the HS and MS degradation operators: $\mathbf{B}$ and $\mathbf{R}$
  - incorporating other physical models: unmixing, MRF, etc.
- Real data
  - misregistration: different sensors, platforms
  - nonlinear degradation: translation, rotation, stretching
  - HISUI, EnMAP: satellites to be launched
  - regularization parameters: included within the estimation scheme
- Sequential inference
  - 4-D (spatial, spectral, temporal) datacube: compressive sensing
  - exploiting 4-D data: super-resolution, change detection, etc.

Robustness with respect to R

FSNR: defined to adjust the knowledge of **R**

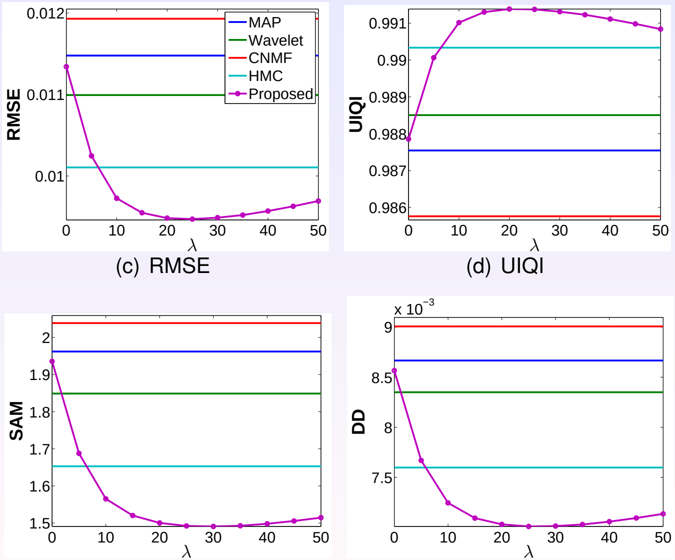
$$\mathrm{FSNR} = 10\log_{10}\left(\frac{\|\mathbf{R}\|_F^2}{m_\lambda n_{\lambda,2} s_2^2}\right)$$





When FSNR is above 8dB, the proposed method outperforms the MAP and wavelet-based MAP methods.

Performance versus λ



(c) RMSE (d) UIQI

(e) SAM (f) DD