Primal-dual interior-point optimization for penalized least squares estimation of abundance maps in hyperspectral imaging

Saïd Moussaoui [1], Emilie Chouzenoux [2], Jérôme Idier [1]

[1] Ecole Centrale Nantes, IRCCyN (CNRS UMR 6597), France
[2] Univ. Paris-Est Marne-la-Vallée, LIGM (CNRS UMR 8049), France

4th Workshop on Hyperspectral Image and Signal Processing: Evolution in Remote Sensing

Shanghai, June 6th, 2012

1. Introduction

Let a data cube with L spectral bands and images of N pixels stored in a matrix $\boldsymbol{Y} \in \mathbb{R}^{L \times N}$.

1.1 Linear spectral unmixing

Explain each pixel spectrum as a linear combination of P endmember spectra $\boldsymbol{s}_p$ (columns of $\boldsymbol{S} \in \mathbb{R}^{L \times P}$)

$$\boldsymbol{y}_n = \sum_{p=1}^{P} a_{(n,p)} \boldsymbol{s}_p + \boldsymbol{e}_n, \quad \forall n \in \{1, \ldots, n\} \Longrightarrow \boldsymbol{Y} = \boldsymbol{S}\,\boldsymbol{A} + \boldsymbol{E}$$

$\boldsymbol{A}$ is the abundance matrix of size $P \times N$ (contains the abundance maps in its rows).
$\boldsymbol{E} \in \mathbb{R}^{L \times N}$ is a matrix representing model/measurement errors.

1.2 Constrained estimation of $\boldsymbol{A}$

$(C1)$ Non-negativity

$$a_{(n,p)} \geqslant 0, \quad \forall n \in \{1, \ldots, n\}, \forall p \in \{1, \ldots, P\}$$

$(C2)$ Sum-to-one

$$\sum_{p=1}^{P} a_{(n,p)} = 1, \quad \forall n \in \{1, \ldots, n\}$$

1.3 Motivations

① Fast processing of large data sets (high image sizes)

② Add prior information on the abundance maps (roughness, sparsity, structural spatial model)

③ Need to jointly unmix all the image pixels (rather than a sequential processing)

1.4 Our proposal ...

✓ Use regularization methods to add the spatial information

✓ Propose special-purpose inversion algorithms having a fast convergence rate

✓ Constrained optimization tools (Interior-point methods)

➥ Primal-dual interior-point methods for constrained convex optimization

Outline

❶ Introduction

❷ Proposed method of abundance maps estimation

- § Problem statement
- § Equality constraint integration
- § Primal-dual interior-point algorithm

❸ Illustrative example and discussion

- § Least squares case
- § Penalized least squares case

❹ Conclusions

2. Abundance maps estimation

2.1 Problem statement

$$\min_{\boldsymbol{A}\in\mathbb{R}^{P\times N}} \left(F(\boldsymbol{A}) = \|\boldsymbol{Y} - \boldsymbol{S}\boldsymbol{A}\|_F^2 \right) \quad s.t. \quad C1 \text{ and/or } C2$$

where $\|\cdot\|_F$ is the Frobenius norm.

$C1$	non-negative least squares (NNLS) [Lawson and Hanson ; 1974]
$C2$	sum-to-one constrained least squares (SCLS) [Settle and Drake ; 1993]
$C1\,\&\,C2$	fully constrained least squares (FCLS) [Heinz and Chaing ; 2002]
$C1\,\&\,C2$	Bayesian and RJ-MCMC method [Dobigeon et al. ; 2008]

⊛ Proposed approach

Minimize any strongly convex and differentiable criterion $F(\boldsymbol{A})$ subject to linear equality and inequality constraints, including C1 and C2.

- Add a variable substitution to integrate the equality constraint C2
- Use a fast interior-point method for solving iteratively the inequality constrained optimization

2.2 Sum-to-one constraint integration

⊛ **Proposition** [Armand ; 2000].

Let $\boldsymbol{A}^{(1)} \in \mathbb{R}^{P\times N}$ and $\boldsymbol{Z} \in \mathbb{R}^{P\times(P-1)}$ such that

- $\mathbf{1}_P^\top \boldsymbol{A} = \mathbf{1}_P^\top \quad \longrightarrow \boldsymbol{A}^{(1)}$ satisfies C2.
- $\boldsymbol{Z}^\top \mathbf{1}_P \ = \mathbf{0}_{P-1} \quad \longrightarrow \boldsymbol{Z}$ formed from the null-space of $\mathbf{1}_P^\top$.

Thus, for any $\boldsymbol{C} \in \mathbb{R}^{P-1\times N}$, $\boldsymbol{A} = \boldsymbol{A}^{(1)} + \boldsymbol{Z}\boldsymbol{C}$ will also satisfy the constraint C2

⊛ **Consequence**

The optimization problem is rewritten as

$$\min_{\boldsymbol{C}\in\mathbb{R}^{P-1\times N}} F(\boldsymbol{A}^{(1)} + \boldsymbol{Z}\boldsymbol{C}) \ \text{ s.t } \left(\boldsymbol{Z}\boldsymbol{C} + \boldsymbol{A}^{(1)}\right) \geqslant \mathbf{0} \iff \min_{\boldsymbol{c}\in\mathbb{R}^{NP-N}} \Phi(\boldsymbol{c}) \quad \text{s.t.} \quad \boldsymbol{T}\boldsymbol{c} + \boldsymbol{t} \geqslant \mathbf{0}$$

with $\boldsymbol{c} = \mathrm{vect}(\boldsymbol{C})$, $t = \mathrm{vect}(\boldsymbol{A}^{(1)})$ and $\boldsymbol{T} = \boldsymbol{I}_N \otimes \boldsymbol{Z}$.

The equality constraint will be implicitly satisfied during the optimization

2.3 Primal-dual interior-point algorithm

⊛ Optimality conditions

The optimal values of $\boldsymbol{c}$ and the Lagrange multipliers $\boldsymbol{\lambda}$ associated to the constraints should satisfy the Karush-Kuhn-Tucker (KKT) conditions

$$① \ \nabla\Phi(\boldsymbol{c}) - \boldsymbol{T}^{\intercal}\boldsymbol{\lambda} = \boldsymbol{0}, \qquad ② \ \boldsymbol{\Lambda}(\boldsymbol{Tc}+\boldsymbol{t}) = \boldsymbol{0}, \qquad ③ \ \boldsymbol{Tc}+\boldsymbol{t} \geqslant \boldsymbol{0}, \qquad ④ \ \boldsymbol{\lambda} \geqslant \boldsymbol{0}$$

where $\boldsymbol{\Lambda} = \mathrm{Diag}(\boldsymbol{\lambda})$.

⊛ Interior-point methods

- Keep the solution inside the feasible domain by adding a logarithmic barrier function
- Iteratively estimate $(\boldsymbol{c}_k, \boldsymbol{\lambda}_k)$ from perturbed versions of the KKT conditions

$$① \ \nabla\Phi(\boldsymbol{c}) - \boldsymbol{T}^{\intercal}\boldsymbol{\lambda} = \boldsymbol{0}, \qquad ② \ \boldsymbol{\Lambda}(\boldsymbol{Tc}+\boldsymbol{t}) = \boldsymbol{\mu}_k, \qquad ③ \ \boldsymbol{Tc}+\boldsymbol{t} \geqslant \boldsymbol{0}, \qquad ④ \ \boldsymbol{\lambda} \geqslant \boldsymbol{0}$$

- The perturbation parameter $\boldsymbol{\mu}_k = \mu_k \boldsymbol{1}_{N(P-1)}$ is chosen such that $\lim_{k\to+\infty} \mu_k = 0$.

⊛ In practice ... two steps

① Calculation of $(\boldsymbol{c}_{k+1}, \boldsymbol{\lambda}_{k+1})$ using $(\boldsymbol{c}_k, \boldsymbol{\lambda}_k)$ by solving approximately the perturbed KKT system using a descent direction method,

② Calculation of μ_{k+1} using an update rule guaranteeing the convergence.

① Approximate resolution of the perturbed KKT system using a Newton step

$$(\boldsymbol{c}_{k+1}, \boldsymbol{\lambda}_{k+1}) = (\boldsymbol{c}_k + \alpha_k \boldsymbol{d}_k^c, \boldsymbol{\lambda}_k + \alpha_k \boldsymbol{d}_k^\lambda)$$

• The directions $\boldsymbol{d}_k^c$ and $\boldsymbol{d}_k^\lambda$ are obtained after a first order development of the perturbed KKT system equalities

$$\begin{bmatrix} \nabla^2 \Phi(\boldsymbol{c}_k) & -\boldsymbol{T}^\intercal \\ \boldsymbol{\Lambda}_k \boldsymbol{T} & \mathrm{Diag}(\boldsymbol{T}\boldsymbol{c}_k + \boldsymbol{t}) \end{bmatrix} \begin{bmatrix} \boldsymbol{d}_k^c \\ \boldsymbol{d}_k^\lambda \end{bmatrix} = \begin{bmatrix} \boldsymbol{T}^\intercal \boldsymbol{\lambda}_k - \nabla \Phi(\boldsymbol{c}_k) \\ \boldsymbol{\mu}_k - \boldsymbol{\Lambda}_k(\boldsymbol{T}\boldsymbol{c}_k + \boldsymbol{t}) \end{bmatrix}$$

• The stepsize α_k should be calculated to ensure the inequalities fulfillment and make a sufficient decrease of a primal-dual merit function

$$\Psi_\mu(\boldsymbol{c}, \boldsymbol{\lambda}) = \Phi(\boldsymbol{c}) - \mu \sum_{i=1}^{NP} \ln([\boldsymbol{T}\boldsymbol{c} + \boldsymbol{t}]_i) + \boldsymbol{\lambda}^\intercal(\boldsymbol{T}\boldsymbol{c} + \boldsymbol{t}) - \mu \sum_{i=1}^{NP} \ln(\lambda_i [\boldsymbol{T}\boldsymbol{c} + \boldsymbol{t}]_i)$$

The search is performed by backtracking and the sufficient decrease is checked using Armijo condition applied to $\Psi_{\mu_k}(\boldsymbol{c}_k + \alpha \boldsymbol{d}_{k+1}^c, \boldsymbol{\lambda}_k + \alpha \boldsymbol{d}_k^\lambda)$.

② Perturbation parameter update

The parameter μ_k is updated according to $\mu_{k+1} = \theta \frac{\delta_{k+1}}{NP}$, where $\delta_{k+1} = (\boldsymbol{T}\boldsymbol{c}_{k+1} + \boldsymbol{t})^\intercal \boldsymbol{\lambda}_{k+1}$ is the duality gap and $\theta \in (0, 1)$ [El Bakary; 1996].

3. Illustrative example and discussion

3.1 Mixture synthesis

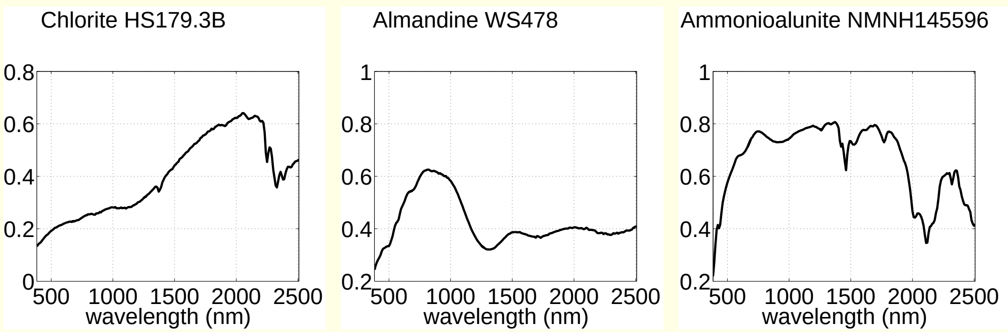
① Randomly select endmember spectra from the USGS library, $L = 224$ and $P \in \{3, 5, 10\}$

② Simulate abundance maps as a superposition of several Gaussian patterns, $N \in \{64^2, \ldots, 256^2\}$

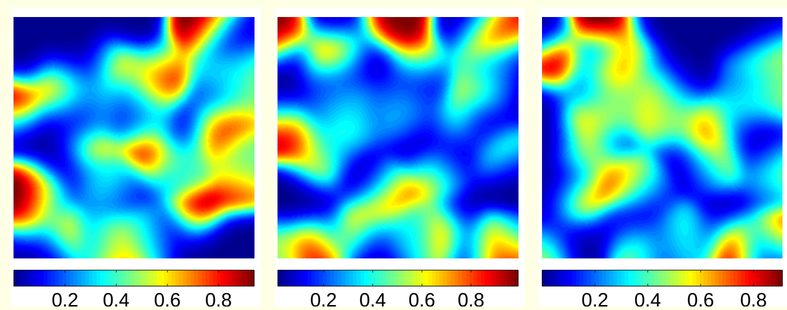
③ Add a Gaussian noise to get some signal-to-noise ratio, SNR $\in \{20, 15, 10, 5\}$ dB

④ Monte Carlo simulation with 30 independent realizations.

⊛ Example



pure spectra



abundance maps

⊛ Unmixing

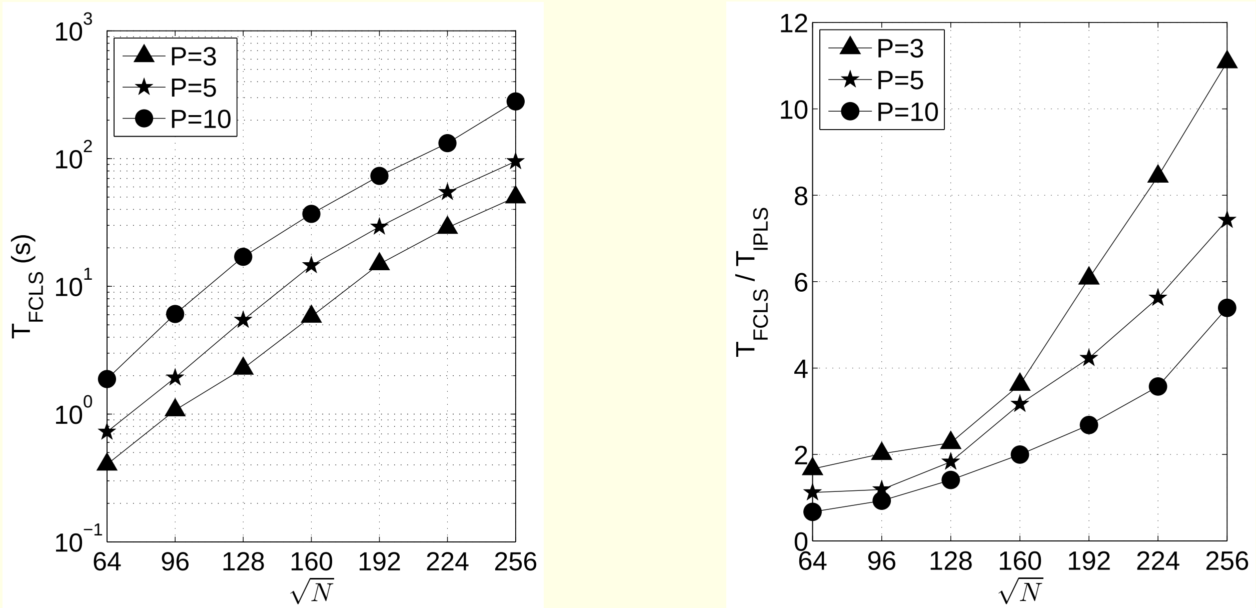
❶ Use an endmember extraction algorithm (VCA [Nascimento and Bioucas Dias ; 2005]),

❷ Estimate the abundance maps using the FCLS and the proposed method (IPLS).

3.2 Constrained least squares estimation

$$\text{Solve} \quad \min_{\boldsymbol{A} \in \mathrm{R}^{(P \times N)}} \left(F_{\mathrm{LS}}(\boldsymbol{A}) = \|\boldsymbol{Y} - \boldsymbol{S}\boldsymbol{A}\|_F^2 \right) \quad \text{s.t} \quad \boldsymbol{A} \geqslant \boldsymbol{0} \text{ and } \boldsymbol{1}_P^{\intercal} \boldsymbol{A} = \boldsymbol{1}_N^{\intercal}$$

⊛ **Computation time:** [MacBookPro - Intel Core 2 Duo 2.4 GHz processor, 4 GB RAM (667 MHz)].



(**Left**) Computation time of FCLS. (**Right**) Obtained speedup with the primal-dual approach.

⊛ **Comments**

- Both algorithms are suitable for parallel implementation,
- IPLS can also take into account the sum less or equal to one constraint $\boldsymbol{1}_P^{\intercal} \boldsymbol{A} \leqslant \boldsymbol{1}_N^{\intercal}$.

3.3 Constrained penalized least squares estimation

⊛ Problem formulation

Solve $\min_{\boldsymbol{A} \in \mathrm{R}^{(P \times N)}} \left(F_{\mathrm{PLS}}(\boldsymbol{A}) = \|\boldsymbol{Y} - \boldsymbol{S}\boldsymbol{A}\|_F^2 + \beta R(\boldsymbol{A}) \right)$ s.t $\boldsymbol{A} \geqslant \mathbf{0}$ and $\mathbf{1}_P^\intercal \boldsymbol{A} = \mathbf{1}_N^\intercal$

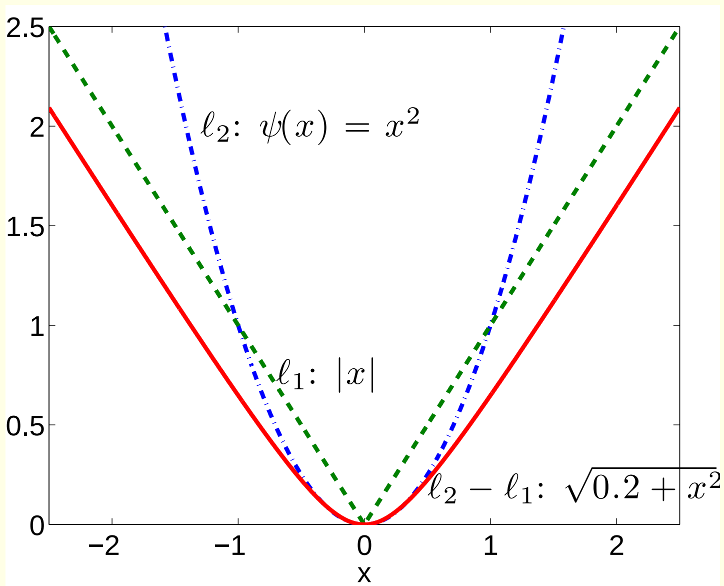
where $R(\boldsymbol{A})$ is a regularization criterion

Roughness penalty: $$R(\boldsymbol{A}) = \sum_{p=1}^{P} \psi \left(\boldsymbol{\Delta} \boldsymbol{a}_p \right)$$

with $\boldsymbol{\Delta}$ a spatial derivative operator and $\psi(\cdot)$ a convex, symmetric and differentiable function.

In image restoration, $(\ell_2 - \ell_1)$ functions are preferred for edge-preserving regularization

Regularization parameter β allows a tradeoff between data fidelity and solution roughness



⊛ Estimation accuracy: $P = 5, L = 244, N = 256^2, \psi(x) = x^2, \beta = 0.1$

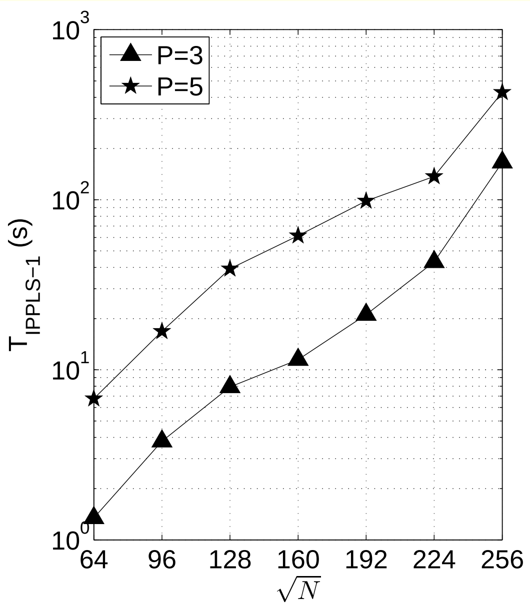
$$\mathrm{NMSE}(\%) = \frac{100}{P} \sum_{p=1}^{P} \left(\frac{\|\boldsymbol{a}_p - \hat{\boldsymbol{a}}_p\|^2}{\|\boldsymbol{a}_p\|^2} \right)$$

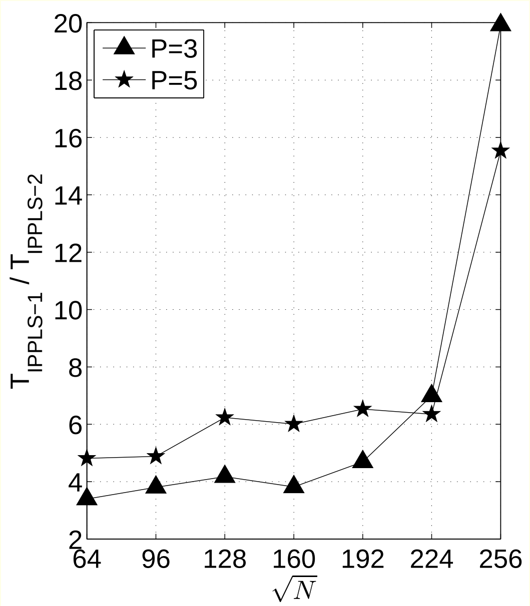
	SNR (dB)			
Method	20	15	10	5
FCLS	0.18	0.46	1.34	3.64
IPLS	0.18	0.46	1.33	3.63
IPPLS	**0.08**	**0.23**	**0.68**	**2.01**

⊛ Computation time:

IPPLS-1 : Exact Newton step

IPPLS-2 : Inexact Newton step (approximate resolution of primal-dual system)





⊛ Computation of the primal and dual directions

$$\begin{bmatrix} \nabla^2\Phi(\boldsymbol{c}_k) & -\boldsymbol{T}^\intercal \\ \boldsymbol{\Lambda}_k\boldsymbol{T} & \mathrm{Diag}(\boldsymbol{T}\boldsymbol{c}_k+\boldsymbol{t}) \end{bmatrix} \begin{bmatrix} \boldsymbol{d}_k^c \\ \boldsymbol{d}_k^\lambda \end{bmatrix} = \begin{bmatrix} \boldsymbol{T}^\intercal\boldsymbol{\lambda}_k - \nabla\Phi(\boldsymbol{c}_k) \\ \boldsymbol{\mu}_k - \boldsymbol{\Lambda}_k(\boldsymbol{T}\boldsymbol{c}_k+\boldsymbol{t}) \end{bmatrix}$$

Variable substitution leads to

$$\boldsymbol{d}_k^\lambda = \mathrm{Diag}(\boldsymbol{T}\boldsymbol{c}_k+\boldsymbol{t})^{-1}\left[\boldsymbol{\mu}_k - \boldsymbol{\Lambda}_k(\boldsymbol{T}\boldsymbol{c}_k+\boldsymbol{t}) - \boldsymbol{\Lambda}_k\boldsymbol{T}\boldsymbol{d}_k^c\right]$$

and

$$\left[\nabla^2\Phi(\boldsymbol{c}_k) + \boldsymbol{T}^\intercal\mathrm{Diag}(\boldsymbol{T}\boldsymbol{c}_k+\boldsymbol{t})^{-1}\boldsymbol{\Lambda}_k\boldsymbol{T}\right]\boldsymbol{d}_k^c = \boldsymbol{T}^\intercal\mathrm{Diag}(\boldsymbol{T}\boldsymbol{c}_k+\boldsymbol{t})^{-1}\boldsymbol{\mu}_k - \nabla\Phi(\boldsymbol{c}_k)$$

- Use a preconditioned conjugate gradient to make an approximate resolution of this system,

- The convergence proof for such an inexact Newton scheme is established whatever the number of gradient method iterations.

3.4 Application to Cuprite data set (AVIRIS'97)

- Cube size $[250 \times 191$ pixels; 188 bands$]$
- The endmembers are determined using VCA.

⊛ Computation time (in seconds)

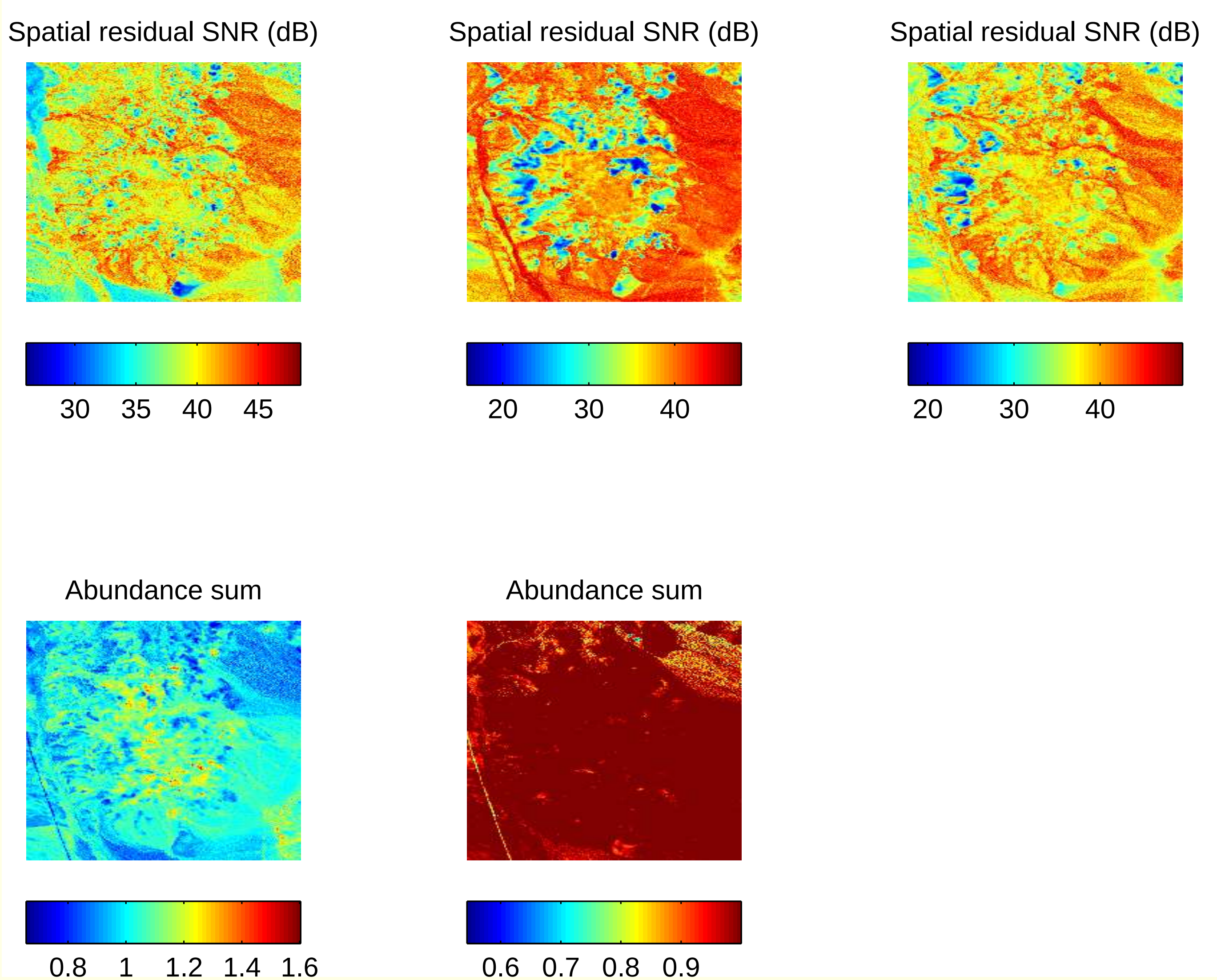
P	Least squares		Penalized least squares	
	FCLS	IPLS	IPPLS-1	IPPLS-2
3	29	4	337	7
4	39	7	645	14
5	51	11	623	21
6	62	15	1520	30
7	76	20	2260	39
8	93	25	–	47
10	116	39	–	73
12	161	61	–	107

⊛ Discussion

- Computation time reduction in the constrained least squares case,
- The approximate resolution of the primal-dual system reduces the computation cost,
- Preconditioned conjugate gradient is used but alternative methods will be preferred to reduce the memory storage.

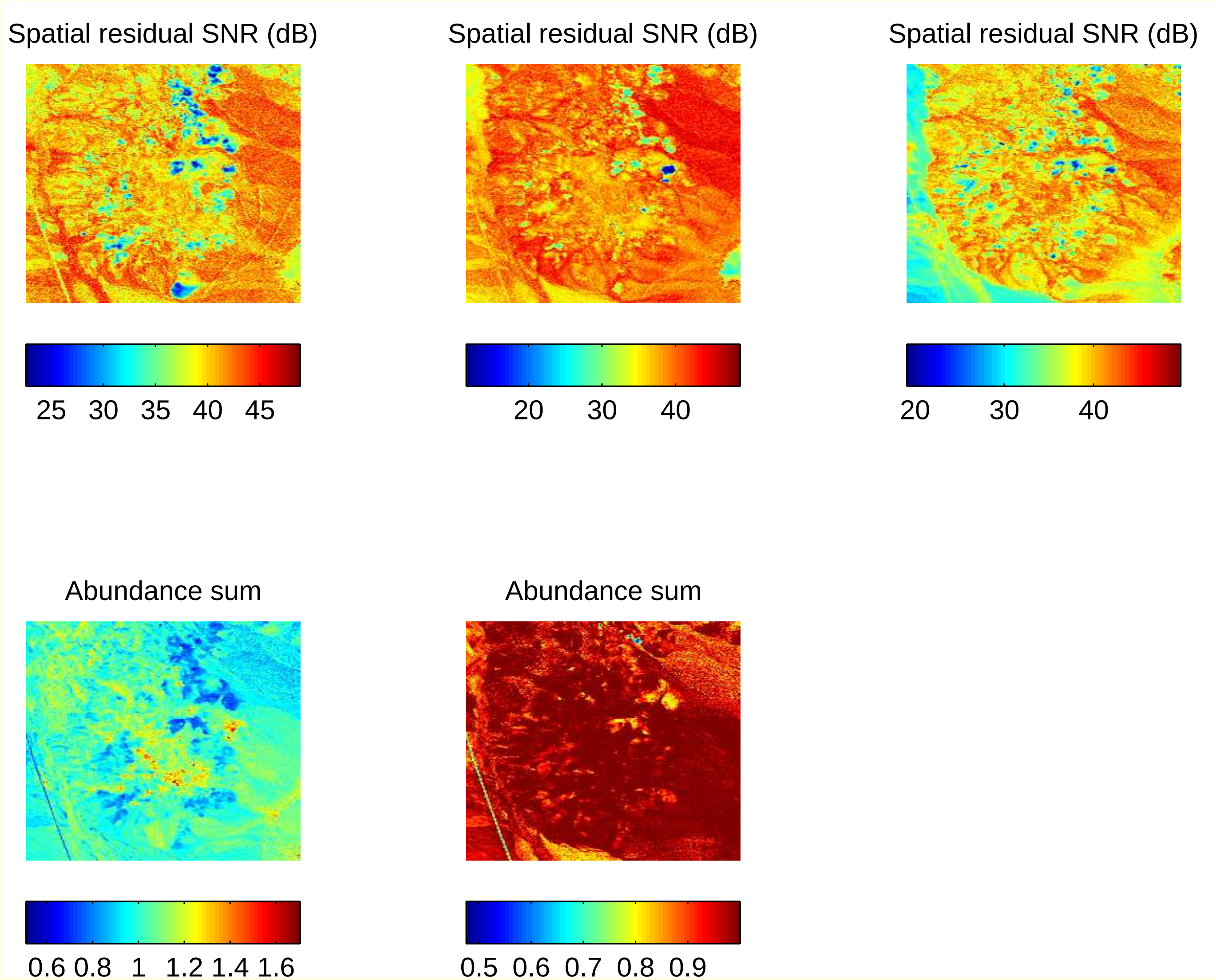
⊛ Spatial SNR and abundance sum (15 endmembers)

Constraints	non-negativity	sum less to one	sum-to-one
Time (s)	53	137	138



⊛ Spatial SNR and abundance sum (20 endmembers)

Constraints	non-negativity	sum less to one	sum-to-one
Time (s)	70	262	237



4. Conclusions

⊛ **Fast interior-point algorithm for the estimation of abundance maps**

✓ Theoretically convergent and faster than the FCLS algorithm,

✓ Allows to account for any linear equality or inequality constraint,

✓ Can be applied for non-Gaussian (but strongly convex) neg-log likelihood criterion,

✓ Adapted to non-quadratic regularization functions.

⊛ **Perspectives**

- GPU implementation of the constrained least squares estimation algorithm,

- Find a more efficient approximate resolution of the primal-dual system,

- Application to non-linear mixing models (criterion convexity ? convergence issues ?)