# Note_linearizedADMM

Keisuke Ozawa[1]

[1]Research and Development, DENSO IT Laboratory

September 25, 2024

# Note of Theoretical Analysis on The Convergence of Linearized ADMM for Separable Reweighted Sparse Hyperspectral Unmixing

**Keisuke Ozawa**[1*]

[1]Research and Development, DENSO IT Laboratory, Shimbashi, Minato-ku, Tokyo 105-0004, Japan

## Abstract

Sparse hyperspectral unmixing is a popular technique used for earth observation. Nonconvex sparse prior is often considered and the linearized ADMM is practically used to solve the problems. This note provides a detailed analysis that supports a recent consequence and experimental results dedicated to this class of algorithms [1]. We prove the convergence of the linearized ADMM for separable reweighted sparse hyperspectral unmxing but the extended algorithm and the proposition also cover relaxed scenarios where the variables are non-separable.

## 1 Introduction

Hyperspectral unmixing is a core technology for earth observation [2, 3, 4]. The purpose is to estimate the abundance maps: the distributions of composite materials over the land. In the literature of hyperspectral image analysis, the composite materials are called endmembers, of which the reflectance spectra span the basis set for observed spectra. There are two categories of hyperspectral unmixing: blind unmixing and dictionary-based unmixing. Blind unmixing does not assume the full information of endmember spectra or some part of them and requires to estimate both the endmember spectra and the abundance maps. Researchers' effort over the years has provided the dictionaries of endmembers. A hyperspectral image is assumed to be unmimxed using a dictionary. However, the dictionaries are usually over-complete, and thus sparsity promoting regularaizations have been introduced to estimate the endmembers. Sparse unmixing basically belongs to this category of unmixing.

A number of studies have proposed to optimize elaborate cost functions using the alternating direction method of multipliers (ADMM) with sometimes modifying it by using inner-outer loop techniques. However, rigorous consideration for the covnergence of this class of algorithms is still lacked. This note provides a detail of the theoretical analysis in the consideration on the convergence of linearized ADMM for separable reweighted sparse hyperspectral unmixing [1].

## 2 Preliminary and related work

In this section, we review existing sparse unmixing methods. Please refer comprehensive articles like [2, 3, 4] for the basis of sparse hyperspectral unmixing and further details and literature. Given a measured hyperspectral image $\mathbf{M} \in \mathbb{R}^{B\times N}$ of $B$ bands and $N$ pixels, the linear sparse unmixing model assumes a spectral dictionary $\mathbf{S} \in \mathbb{R}^{B\times D}$ with total $D$ (typically of hundreds) spectral signatures to unmix $\mathbf{M}$ as

$$\mathbf{M} = \mathbf{S}\mathbf{X} + \mathbf{N} \ \text{ s.t. } \ \mathbf{X} \geq 0, \tag{1}$$

where $\mathbf{X} \in \mathbb{R}^{D\times N}$ is the abundance matrix and $\mathbf{N}$ is the observation noise. The nonnegative assumption is often replaced with the sum-to-one constraint, which holds when irradiance and material concentration are uniform and thus a pixel stores the averaged spectra within the sensor resolution. However, this is not always true and some limitations are known in practice [2, 5, 6]. We herein assume only the nonnegative constraint and give results under that assumption.

The fundamental assumption of sparse unmixing is that only a small number of endmembers should have nonzero abundance. Such a sparsity comes with a constraint of a bounded cardinarity of abundance. From the Lagrange duality, there exists a coefficient on the $L_0$ norm added to a cost function. It is well known in the long history of sparse modelling that $L_0$ norm is difficult to optimize and thus various relaxations have been developed. Both convex and nonconvex relaxations can be consiered in the context of sparse unmixing, where, in general, the sparse unmixing can be written as

$$\min_{\mathbf{X}} \frac{1}{2}\|\mathbf{M} - \mathbf{S}\mathbf{X}\|_F^2 + \mu\Psi(\mathbf{X}) \ \text{ s.t. } \ \mathbf{X} \geq 0. \tag{2}$$

SUnSAL [7, 8] is the pioneering work that endows the pixel-wise sparsity on abundance maps. The regularization is $\Psi(\mathbf{X}) = \|\mathbf{X}\|_{1,1}$, which enforces every pixel to have zero abundance. While SUnSAL is a standard and suggestive study with a effective ADMM algorithm, it is insufficient in terms of reducing the number of selected endmembers. To circumvent this, collaborative sparsity has been also used [9]. The regularization $\Psi(\mathbf{X}) = \|\mathbf{X}\|_{2,1}$ has a collaborative behaviour to put all the pixels of some endmembers to zero, which often better reflects the assumption of sparse unmixing. They are convex and can be solved efficiently by using the soft-shrinkage or vector-soft-shrinkage operators. On the other hand, some work introduced nonconvex relaxation of $L_0$ norm. $L_p$ penalty $\Psi(\mathbf{X}) = \|\mathbf{X}\|_p^p$ is a tighter nonconvex relaxation than $L_1$ norm and a balanced characteristic was deduced for $p = 1/2$. Reweighted sparse regularization is also used well in sparse unmixing. Reweighted scheme was originally introduced by Candès et al. [10] to practically solve nonconvex optimization problem, e.g. using log-sum penalty. However, the convergence of ADMM using the reweighted sparsity has not been shown specifically under the assumptions of sparse hyperspectral unmixing, and there are only empirical reports of converging behaviours of the algorithms. On the same line, double-reweighted sparse regularization was presented to incorporate the effect of collaborative sparsity on abundance [11]. Although this regularization shows the better regression accuracy and sparseness promoting tendency, the technique is rather empirical and no analytical justification has been provided. Another partly lacked investigation is that existing work has not provided the functions $\Psi(\mathbf{X})$ explicitly as well as the analytical properties of these regularizations, which has been partly discussed in [1] based on an established result [12]. Although applying ADMM with the reweighted techniques shows the empirical convergence, the convergence analysis still remains a common issue.

## 3 Convergence analysis of linearized ADMM for sparse unmixing with nonconvex nonsmooth penalties

In this note, we revisit the reweighted sparsity promoting regularizations for sparse unmixing. First, we see a unified view to design sparse unmixing algorithms with nonconvex regularizations. It includes the reweighted sparsity and some varieties of double reweighted sparsity promoting regularizations as the update rules to solve the nonconvex problems. We give the explicit form of the corresponding nonconvex functions. Our analysis could provide a better understanding for the rather empirical flavor of double reweighted schemes. Second, we prove that ADMM reaches a solution under conditions satisfied in sparse unmixing.

### 3.1 Nonconvex penalties and reweighted sparsity

We assume the following class of nonconvex functions that are smooth almost everywhere excluding the origin. Local linear approximation is a widely used technique to optimize such nonconvex functions.

Candès et al. [10] considered a problem as $\min_x f(|x|) \Leftrightarrow \min_{x,u} f(u)$ s.t. $|x| \leq u$. Following this work for solving such nonconvex, nonsmooth optimization problems, there has been developed a more general framework, based on which we solve sparse unmixing problems.

Let $\Psi = \varphi \circ g$ with $\varphi$ concave, smooth function and $g$ convex, possibly nonsmooth function that has a prox operator. The tangent is thus $\varphi'(g(x'))(g(x) - g(x')) + \varphi(g(x'))$. We assume that $g(x) = |x|$ in this study. We extend this technique to the specific form appeared in sparse hyperspectral unmixing: $\Psi(\mathbf{X}) = \varphi(\mathbf{g}(\mathbf{X}))$, where $\varphi : \mathrm{Im}(\mathbf{g}) \to \mathbb{R}$ is differentiable and concave w.r.t. every $g_{ij}$ and $\mathbf{g} : \mathbb{R}^{n\times m} \to \mathbb{R}^{n\times m}$ takes $g_{ij}(\mathbf{X}) = |X_{ij}|$, for which the concavity of $\varphi$ is equivalent to the concavity w.r.t. every $g_{ij}(\mathbf{X})$. There are some ways to show that the following functions are concave w.r.t. each variable, by saying that the tangent is increasing and positive, or that the composite of a concave function followed by a nondecreasing function is concave, etc. The tangent of $\Psi$ w.r.t. $\mathbf{X}_{ij}$ at $\mathbf{X}$ is then $\partial\varphi/\partial g_{ij}(\mathbf{X})$.

Here are some examples that appear in the literature of hyperspectral unmixing.

#### 3.1.1 Log-sum penalty for $L_1$-reweighted sparsity

$$\Psi(\mathbf{X}) = \sum_{i,j} \log(|X_{ij}| + \varepsilon), \tag{3}$$

where the constant $\varepsilon > 0$, leads a linearly approximated upper bound for the $t$-th update as

$$\sum_{ij} \frac{\partial\varphi}{\partial g_{ij}} g_{ij}(\mathbf{X}) = \sum_{i,j} \frac{1}{|X_{ij}^{(t)}| + \varepsilon} |X_{ij}|. \tag{4}$$

#### 3.1.2 Modified log-sum penalty for $L_1$-collaborative double reweighted sparsity

$$\Psi(\mathbf{X}) = \sum_{i,j} \frac{\log(|X_{ij}| + \varepsilon) - \log(\sum_{j'} |X_{ij'}| + \varepsilon')}{\sum_{j'\neq j} |X_{ij'}| + \varepsilon' - \varepsilon}, \tag{5}$$

where the constants $\varepsilon, \varepsilon' > 0$, leads for the $t$-th update as

$$\sum_{ij} \frac{\partial\varphi}{\partial g_{ij}} g_{ij}(\mathbf{X}) = \frac{1}{\left(|X_{ij}^{(t)}| + \varepsilon\right)\left(\sum_{j'} |X_{ij'}^{(t)}| + \varepsilon'\right)} |X_{ij}|. \tag{6}$$

#### 3.1.3 Modified log-sum penalty for $L2$-collaborative double reweighted sparsity

$$\Psi(\mathbf{X}) = \sum_{i,j} \frac{-\mathrm{asinh}\left(C_{ij}^{1/2}\left(C_{ij} - \varepsilon|X_{ij}|\right)\right) / \left(||X_{ij}| + \varepsilon|C_{ij}\right)}{\left(C_{ij} + \varepsilon^2\right)^{1/2}}, \tag{7}$$

where $C_{ij} := \sum_{j'\neq j} X_{ij'}^2 + \varepsilon'$ and $\varepsilon > 0$, leads

$$\sum_{ij} \frac{\partial\varphi}{\partial g_{ij}} g_{ij}(\mathbf{X}) = \frac{1}{\left(|X_{ij}^{(t)}| + \varepsilon\right)\sqrt{\sum_{j'} |X_{ij'}^{(t)}|^2 + \varepsilon'}} |X_{ij}|. \tag{8}$$

#### 3.1.4 Arctangent penalty for $L_2$-$L_1$-reweighted sparsity

$$\Psi(\mathbf{X}) = \sum_{i,j} \frac{\mathrm{atan}(|X_{ij}|/\sqrt{a})}{\sqrt{a}}, \tag{9}$$

where $\varphi(x) = \arctan(x/\sqrt{a})/\sqrt{a}$ with a constant $a > 0$.

### 3.2 Reweighted ADMM for sparse unmixing with nonconvex nonsmooth penalties

Based on the nonconvex, nonsmooth penalty $\Psi(\mathbf{X}) = \varphi(\mathbf{g}(\mathbf{X}))$, we solve the sparse unmixig problem using ADMM [13, 14].

$$\min_{\mathbf{X}} \frac{1}{2}\|\mathbf{M} - \mathbf{S}\mathbf{X}\|_F^2 + \mu\varphi(\mathbf{g}(\mathbf{X})) + \iota_{\mathbb{R}\geq 0}(\mathbf{X}), \tag{10}$$

where $\iota_{\mathbb{R}\geq 0}(\mathbf{X}) = \sum_{ij} \iota_{\mathbb{R}\geq 0}(X_{ij})$ is the nonnegative constraint. Since this cannot be solved directly, variable splitting technique is often used. ADMM allows a flexible splitting into variables to efficiently solve the subproblems. However, there are two difficulties to solve the nonconvex, nonsmooth sparse unmixing.

- $\Psi$ is nonconvex and nonsmooth, and thus minimization subproblems including this function is intractable in general.
- The functions $g$ and $h$ are not differentiable. There are gaps between convex problems and nonconvex ones in showing convergence when we split variables for the functions $g$ and $h$.

The following technique and assumption are introduced as described in [15] to efficiently solve the problem as well as to guarantee a convergence property for general scenarios but the algorithms used specifically in nonconvex sparse hyperspectral unmixing is not covered.

- Optimize the variable with $\Psi$ approximately via the linearized cost function tangent to $\Psi$ at the current coordinate.
- The prox operators for $g$ and $h$ are assumed to satisfy $\mathrm{prox}_{g+h} = \mathrm{prox}_h \circ \mathrm{prox}_g$. This assumption enables to split the problem into subproblems of smooth one and nonsmooth but proximable one, and so the dual variable in the augmented Lagrangian is shown to be upper bounded.

We introduce an auxiliary variable $\mathbf{Y}$ under the constraint of $\mathbf{X} = \mathbf{Y}$. Using the auxiliary variable $\mathbf{Y}$ and a dual variable $\mathbf{P}$, the augmented Lagrangian is

$$\mathcal{L}(\mathbf{X},\mathbf{Y},\mathbf{P}) = \frac{1}{2}\|\mathbf{M}-\mathbf{S}\mathbf{Y}\|_F^2 + \mu\varphi(\mathbf{g}(\mathbf{X})) + \sum_{ij}\iota_{\mathbb{R}\geq 0}(X_{ij}) + \langle \mathbf{P}, \mathbf{X}-\mathbf{Y}\rangle + \frac{\rho}{2}\|\mathbf{X}-\mathbf{Y}\|_F^2 \quad (11)$$

and the following subproblems are considered:

$$\mathbf{X}^{(t+1)} = \arg\min_{\mathbf{X}} \tilde{\mathcal{L}}(\mathbf{X},\mathbf{Y}^{(t)},\mathbf{P}^{(t)}), \quad (12)$$

$$\mathbf{Y}^{(t+1)} = \arg\min_{\mathbf{Y}} \mathcal{L}(\mathbf{X}^{(t+1)},\mathbf{Y},\mathbf{P}^{(t)}), \quad (13)$$

$$\mathbf{P}^{(t+1)} = \mathbf{P}^{(t)} + \rho(\mathbf{X}^{(t+1)}-\mathbf{Y}^{(t+1)}), \quad (14)$$

where $\tilde{\mathcal{L}}$ is a reweighted function w.r.t. $\Psi$,

$$\tilde{\mathcal{L}}(\mathbf{X},\mathbf{Y}^{(t)},\mathbf{P}^{(t)}) = \mu\sum_{ij}\frac{\partial\varphi}{\partial g_{ij}}(\mathbf{X}^{(t)})g_{ij}(\mathbf{X}) + \sum_{ij}\iota_{\mathbb{R}\geq 0}(X_{ij}) + \langle \mathbf{P}^{(t)}, \mathbf{X}-\mathbf{Y}^{(t)}\rangle + \frac{\rho}{2}\|\mathbf{X}-\mathbf{Y}^{(t)}\|_F^2. \quad (15)$$

From the properties of the prox operators of $g$ and $h$ and from Equation 12, updating $\mathbf{X}$ follows

$$X_{ij}^{(t+1)} = \max\left(\mathcal{S}\left(X_{ij}'^{(t)}, \frac{\mu}{\rho}\frac{\partial\varphi}{\partial g}(X_{ij}^{(t)})\right), 0\right), \quad (16)$$

where $\mathbf{X}'^{(t)} = \mathbf{Y}^{(t)} - \mathbf{P}^{(t)}/\rho$, $\mathcal{S}(x,\sigma)$ is the soft-thresholding operator [16], and $\max(x,y)$ returns the larger one.

From Equation 13, updating $\mathbf{Y}$ follows

$$\mathbf{Y}^{(t+1)} = (\mathbf{S}^\top\mathbf{S}+\rho\mathbf{I})^{-1}\left\{\mathbf{S}^\top\mathbf{M} + \rho\left(\mathbf{X}^{(t+1)} + \frac{\mathbf{P}^{(t)}}{\rho}\right)\right\}, \quad (17)$$

where $\mathbf{I}$ is the identity matrix.

The whole procedures are summarized in Algorithm 1.

**Algorithm 1** ADMM for reweighted SHU

**Input:** Hyperspectral image $\mathbf{M} \in \mathbb{R}^{B\times N}$, spectral dictionary $\mathbf{S} \in \mathbb{R}^{B\times D}$ (pruned if necessary)
*Initialization* : $\mathbf{X} = \mathbf{O}$;
*Parameters* : $\rho > 0$, tol, max iteration ($t_{\max}$)
**Output:** Abundance $\mathbf{X}^* \in \mathbb{R}^{D\times N}$
1: **while** $t < t_{\max}$ and the updates in variables are under a threshold **do**
2: Update $\mathbf{X}$ using Equation 16.
3: Update $\mathbf{Y}$ using Equation 17.
4: Update $\mathbf{P}$ using Equation 14.
5: Continue updates 2-4.
6: **end while**

### 3.3 Convergence analysis of the linearized ADMM

Our interest is the convergence of the algorithm. In order to see what properties of functions and constraints conclude the analytical consequence, we slightly generalize the problem as it includes the sparse unmixing problem as a special case. Note that our analysis follows the styles of established theoretical results [15, 17] but is considered for assumptions and update rules used in sparse hyperspectral unmixing. With simplifying the notation in vectors,

$$\mathcal{L}(\mathbf{x},\mathbf{y},\mathbf{p}) = f(\mathbf{y}) + \varphi(\mathbf{g}(\mathbf{x})) + h(\mathbf{x}) + \langle \mathbf{p}, \mathbf{A}\mathbf{x}+\mathbf{B}\mathbf{y}\rangle + \frac{\rho}{2}\|\mathbf{A}\mathbf{x}+\mathbf{B}\mathbf{y}\|^2 \quad (18)$$

is considered. The regularization weight $\mu$ is included in $\varphi$ for simplicity. We assume $\mathbf{x} \in \mathbb{R}^q$ and $\mathbf{g} : \mathbb{R}^q \to \mathbb{R}^q$ with $g_i(\mathbf{x}) = |x_i|$, and $h(x_i) = \iota_{\mathbb{R}\geq 0}(x_i)$. We denote $g(x_i) = |x_i|$. We assume $f(\mathbf{y}) = \|\mathbf{M}-\mathbf{S}\mathbf{Y}\|_F^2/2$ for hyperspectral unmixing problems and $\mathbf{y} = \mathrm{vec}(\mathbf{Y})$; $\mathbf{x} = \mathrm{vec}(\mathbf{X})$; $\mathbf{A} = \mathbf{I}$ and $\mathbf{B} = -\mathbf{I}$ with the identity matrix $\mathbf{I}$. This is actually Lipschitz smooth as the difference between its gradients are bounded above with the maximum eigenvalue of $\mathbf{S}^\top\mathbf{S}$. Here, we replaced the spectral dictionary $\mathbf{S}$ by $\mathrm{diag}(\mathbf{S},\cdots,\mathbf{S})$ in reformulating the original problem to that using vectors. This does not affect the following analysis because each block matrix operates on each columns of the matrix variable and thus what appears in the analaysis, the largest eigenvalue, does not change in this reformulation. We assume that $\mathbf{B}$ is full-rank, $\mathrm{Im}(\mathbf{A}) \subseteq \mathrm{Im}(\mathbf{B})$, and $\sum_i g_i(\mathbf{x}) + \frac{1}{2}\|\mathbf{A}\mathbf{x}-\bar{\mathbf{x}}\|_F^2$ has a unique solution for any $\bar{\mathbf{x}}$. Then our claims hold, and these assumptions hold for the sparse unmixing problems, too. Some coefficients in the following claims would be the same under the specific assumptions.

The updates of the proposed linearized ADMM is

$$x_i^{(t+1)} = \arg\min_{x_i} \tilde{\mathcal{L}}(\mathbf{x}_{<i}^{(t+1)}, x_i, \mathbf{x}_{>i}^{(t)}, \mathbf{y}^{(t)}, \mathbf{p}^{(t)}), \quad (19)$$

$$\mathbf{y}^{(t+1)} = \arg\min_{\mathbf{y}} \mathcal{L}(\mathbf{x}^{(t+1)}, \mathbf{y}, \mathbf{p}^{(t)}), \quad (20)$$

$$\mathbf{p}^{(t+1)} = \mathbf{p}^{(t)} + \rho(\mathbf{A}\mathbf{x}^{(t+1)} + \mathbf{B}\mathbf{y}^{(t+1)}), \quad (21)$$

where the linearized Lagrangian is defined as

$$\tilde{\mathcal{L}}(\mathbf{x}_{<i}^{(t+1)}, x_i, \mathbf{x}_{>i}^{(t)}, \mathbf{y}^{(t)}, \mathbf{p}^{(t)}) = \sum_i \frac{\partial\varphi(\mathbf{g}(\mathbf{x}_{<i}^{(t+1)}, \mathbf{x}_{\geq i}^{(t)})}{\partial g_i} g(x_i) + h(x_i) + \langle \mathbf{p}^{(t)}, \mathbf{A}(\mathbf{x}_{<i}^{(t+1)\top}, x_i, \mathbf{x}_{>i}^{(t)\top})^\top + \mathbf{B}\mathbf{y}^{(t)}\rangle + \frac{\rho}{2}\|\mathbf{A}(\mathbf{x}_{<i}^{(t+1)\top}, x_i, \mathbf{x}_{>i}^{(t)\top})^\top + \mathbf{B}\mathbf{y}^{(t)}\|^2. \quad (22)$$

Our analytical result will suggest that the linearized ADMM formulations appeared in sparse hyperspectral unmixing can be regarded as an approximation of (19) to (12) with a practical benefit of computational cost. For the most conventional reweighted sparsity coming from the Log-sum penalty, the practical algorithm and our analytically sufficient algorithm are the same, since the variables in $\mathbf{X}$ or $\mathbf{x}$ are independent and thus the subgradient vector is the same. When variables interact e.g. by introducing a group-sparsity-promoting term, the analytical problem is that the convexity does not hold, and thus we have a sufficient condition for convergence when each variable $x_i$ is updated incrementally.

**Assumption 1.** We assume the following class of functions $f$, $g$, and $h$.

1. $f$ is Lipschitz smooth with constant $L_f$.
2. $\sum_i g_i(\mathbf{x}) + \|\mathbf{A}\mathbf{x}-\bar{\mathbf{x}}\|^2/2$ for any $\bar{\mathbf{x}}$ has a unique solution. Let the solution is given $\mathrm{prox}_{\mathbf{g}}(\bar{\mathbf{x}})$. Then, $\sum_i g_i(\mathbf{x}) + \sum_i h_i(\mathbf{x}) + \|\mathbf{A}\mathbf{x}-\bar{\mathbf{x}}\|^2/2$ has a unique solution as $\mathrm{prox}_{g+h}(\bar{\mathbf{x}}) = \mathrm{prox}_h \circ \mathrm{prox}_{\mathbf{g}}(\bar{\mathbf{x}})$.

3. Finally, $\varphi$ is concave and differentiable w.r.t. every $g_i$ within Im($\mathbf{g}$).

We denote the sequence that the algorithm generates by $\left\{\boldsymbol{\vartheta}^{(t)}\right\}_{0\le t<\infty} := \left\{\mathbf{x}^{(t)}, \mathbf{y}^{(t)}, \mathbf{p}^{(t)}\right\}_{0\le t<\infty}$.

We have the following lemmas. Again, note that our analysis follows the style of [15, 17] and Lemma 1 and 3 follow directly from [15] and some other part was also inspired by [17].

**Lemma 1.** The change of the dual variable **p** from the update is upper bounded by the change of the variable **y** up to the scaling with a constant:

$$\|\mathbf{p}^{(t+1)} - \mathbf{p}^{(t)}\| \le C_{\mathbf{B},f}\|\mathbf{y}^{(t+1)} - \mathbf{y}^{(t)}\|, \tag{23}$$

where $C_{\mathbf{B},f} = \underline{\sigma}_+^{-1/2}(\mathbf{B}^\top\mathbf{B})L_f$, $\underline{\sigma}_+(\mathbf{B}^\top\mathbf{B})$ being the smallest strictly-positive eigenvalue of $\mathbf{B}^\top\mathbf{B}$.

**Lemma 2.** The Lagrangian satisfies a sufficient descent with the update of the variable **x** if $\mathbf{A}^\top\mathbf{A}$ is diagonal:

$$\begin{aligned}\mathcal{L}_\rho(\mathbf{x}^{(t)}, \mathbf{y}^{(t)}, \mathbf{p}^{(t)}) - \mathcal{L}_\rho(\mathbf{x}^{(t+1)}, \mathbf{y}^{(t)}, \mathbf{p}^{(t)})&\\ \ge \frac{\rho}{2}\|\mathbf{A}\mathbf{x}^{(t)} - \mathbf{A}\mathbf{x}^{(t+1)}\|^2&\end{aligned} \tag{24}$$

**Lemma 3.** The Lagrangian satisfies a sufficient descent with the updates of the variable **y** and the dual **p** if the Lagrange multiplier $\rho$ is sufficiently large so that,

$$\rho\sigma_{\min}(\mathbf{B}^\top\mathbf{B}) - L_f - 2C_{\mathbf{B},f} > 0, \text{ and then}$$

$$\begin{aligned}&\mathcal{L}_\rho(\mathbf{x}^{(t+1)}, \mathbf{y}^{(t)}, \mathbf{p}^{(t)}) - \mathcal{L}_\rho(\mathbf{x}^{(t+1)}, \mathbf{y}^{(t+1)}, \mathbf{p}^{(t+1)})\\ &\ge \left(\frac{\rho\sigma_{\min}(\mathbf{B}^\top\mathbf{B})}{2} - \frac{L_f}{2} - C_{B,f}\right)\|\mathbf{y}^{(t)} - \mathbf{y}^{(t+1)}\|^2,\end{aligned} \tag{25}$$

where $\sigma_{\min}(\mathbf{B}^\top\mathbf{B})$ is the smallest eigenvalue of $\mathbf{B}^\top\mathbf{B}$.

**Lemma 4.** The Lagrangian $\mathcal{L}_\rho$ is lower-bounded for sufficiently large $\rho$'s.

**Lemma 5.** The subgradient of the Lagrangian $\mathcal{L}_\rho$ w.r.t. the primal and dual variables is bounded above by the change of the variables:

$$\nabla_{\mathbf{y}}\mathcal{L}_\rho(\boldsymbol{\vartheta}^{(t+1)}) = \mathbf{B}^\top(\mathbf{p}^{(t+1)} - \mathbf{p}^{(t)}) \tag{26}$$

$$\nabla_{\mathbf{p}}\mathcal{L}_\rho(\boldsymbol{\vartheta}^{(t+1)}) = \frac{1}{\rho}(\mathbf{p}^{(t+1)} - \mathbf{p}^{(t)}) \tag{27}$$

$$\begin{aligned}&\partial_i \left\{\varphi(\mathbf{g}(\mathbf{x})) + h(\mathbf{x})\right\}_{\mathbf{x}^{(t+1)}}\\ &- \partial\left\{\frac{\varphi(\mathbf{g}(\mathbf{x}_{<i}^{(t+1)}, \mathbf{x}_{\ge i}^{(t)}))}{g_i} g(\mathbf{x}) + h(\mathbf{x})\right\}_{x_i^{(t+1)}}\\ &+ \mathbf{A}^\top(\mathbf{p}^{(t+1)} - \mathbf{p}^{(t)}) + \rho\mathbf{B}\left(\mathbf{y}^{(t+1)} - \mathbf{y}^{(t)}\right) \in \frac{\partial\mathcal{L}_\rho}{\partial x_i}(\boldsymbol{\vartheta}^{(t+1)})\end{aligned} \tag{28}$$

Then, $\|\nabla_{\mathbf{y}}\mathcal{L}_\rho(\boldsymbol{\vartheta}^{(t+1)})\| \to 0$, $\|\nabla_{\mathbf{p}}\mathcal{L}_\rho(\boldsymbol{\vartheta}^{(t+1)})\| \to 0$, and there exists $\mathbf{d}^{(t+1)} \in \partial_{\mathbf{x}}\mathcal{L}_\rho(\boldsymbol{\vartheta}^{(t+1)})$ that $\|\mathbf{d}^{(t+1)}\| \to 0$ as $t \to \infty$.

From these Lemmas, we conclude the main proposition.

**Proposition 1.** Under the Assumption 3, the sequence $\left\{\boldsymbol{\vartheta}^{(t)}\right\}_{0\le t<\infty}$ has a limit point and converges to a stationary point.

# 4 Proofs of Lemmas

## 4.1 Proof of Lemma 1

*Proof.* From the optimality of $\mathbf{y}^{(t)}$,

$$\nabla f(\mathbf{y}^{(t)}) + \mathbf{B}^\top\mathbf{p}^{(t-1)} + \rho\mathbf{B}^\top(\mathbf{A}\mathbf{x}^{(t)} + \mathbf{B}\mathbf{y}^{(t)}) = 0.$$

Using the update rule of **p**, we obtain

$$\mathbf{B}^\top\mathbf{p}^{(t)} = -\nabla f(\mathbf{y}^{(t)}).$$

The change of **p** in each update is bounded from above as follows:

$$\begin{aligned}\|\mathbf{p}^{(t+1)} - \mathbf{p}^{(t)}\| &\le \underline{\sigma}_+^{-1/2}(\mathbf{B}^\top\mathbf{B})\|\mathbf{B}^\top(\mathbf{p}^{(t+1)} - \mathbf{p}^{(t)})\|\\ &= \underline{\sigma}_+^{-1/2}(\mathbf{B}^\top\mathbf{B})\|\nabla f(\mathbf{y}^{(t+1)}) - \nabla f(\mathbf{y}^{(t)})\|\\ &\le \underline{\sigma}_+^{-1/2}(\mathbf{B}^\top\mathbf{B})L_f\|\mathbf{y}^{(t+1)} - \mathbf{y}^{(t)}\|.\end{aligned}$$

This concludes Lemma 1. □

## 4.2 Proof of Lemma 2

*Proof.* Since $h(\mathbf{x}^{(t)}) = 0$ is satisfied for all $t \ge 0$, the change of Lagrangian is

$$\begin{aligned}&\mathcal{L}_\rho(\mathbf{x}^{(t)}) - \mathcal{L}_\rho(\mathbf{x}^{(t+1)})\\ &= \varphi(\mathbf{g}(\mathbf{x}^{(t)})) - \varphi(\mathbf{g}(\mathbf{x}^{(t+1)})) + \langle\mathbf{p}^{(t)}, \mathbf{A}(\mathbf{x}^{(t)} - \mathbf{x}^{(t+1)})\rangle\\ &+ \frac{\rho}{2}\|\mathbf{A}\mathbf{x}^{(t)} + \mathbf{B}\mathbf{y}^{(t)}\|^2 - \frac{\rho}{2}\|\mathbf{A}\mathbf{x}^{(t+1)} + \mathbf{B}\mathbf{y}^{(t)}\|^2.\end{aligned}$$

We introduce a subderivative vector

$$\mathbf{d}^{(t)} := -\mathbf{A}^\top\mathbf{p}^{(t)} - \rho\mathbf{A}^\top(\mathbf{A}\mathbf{x}^{(t+1)} + \mathbf{B}\mathbf{y}^{(t)})$$

and then have

$$\begin{aligned}\mathcal{L}_\rho(\mathbf{x}^{(t)}) - \mathcal{L}_\rho(\mathbf{x}^{(t+1)}) &= \varphi(\mathbf{g}(\mathbf{x}^{(t)})) - \varphi(\mathbf{g}(\mathbf{x}^{(t+1)}))\\ &-\langle\mathbf{d}^{(t)}, \mathbf{x}^{(t)} - \mathbf{x}^{(t+1)}\rangle + \frac{\rho}{2}\|\mathbf{A}(\mathbf{x}^{(t)} - \mathbf{x}^{(t+1)})\|^2.\end{aligned}$$

If $\mathbf{A}^\top\mathbf{A}$ is diagonal, which holds when $\mathbf{A} = \mathbf{I}$ as assumed for hyperspectral sparse unmixing, each entry is

$$d_i^{(t)} = (-\mathbf{A}^\top\mathbf{p}^{(t)})_i - \rho(\mathbf{A}^\top\mathbf{A}\mathbf{x}^{(t+1)})_i - \rho(\mathbf{A}^\top\mathbf{B}\mathbf{y}^{(t)})_i,$$

depends only on $x_i^{(t+1)}$ among the entries from $\mathbf{x}^{(t+1)}$. Under this condition and using the stationary condition

$$0 \in \frac{\partial\tilde{\mathcal{L}}(\mathbf{x}_{\le i}^{(t+1)}, \mathbf{x}_{>i}^{(t)}, \mathbf{y}^{(t)}, \mathbf{p}^{(t)})}{\partial x_i},$$

we have

$$d_i^{(t)} \in \partial_i\left(\frac{\partial\varphi(\mathbf{g}(\mathbf{x}_{<i}^{(t+1)}, \mathbf{x}_{\ge i}^{(t)}))}{\partial g_i} g(x_i^{(t+1)}) + h(x_i^{(t+1)})\right)$$

and equivalently

$$d_i^{(t)} \in \frac{\partial\varphi(\mathbf{g}(\mathbf{x}_{<i}^{(t+1)}, \mathbf{x}_{\ge i}^{(t)}))}{\partial g_i}\partial_i g(x_i^{(t+1)}) + \partial_i h(x_i^{(t+1)})$$

as a set for the right hand of the inclusion. Note that, if $\mathbf{A}^\top\mathbf{A}$ is not diagonal and so mixes up different $x_i$'s, the above inclusion occasionally fails since there are some $x_i^{(t)}$'s instead of $x_i^{(t+1)}$'s that define $d_i^{(t)}$'s.

If the following inequality holds

$$\varphi(\mathbf{g}(\mathbf{x}^{(t)})) - \varphi(\mathbf{g}(\mathbf{x}^{(t+1)})) - \langle \mathbf{d}^{(t)}, \mathbf{x}^{(t)} - \mathbf{x}^{(t+1)} \rangle \geq 0, \qquad (\star)$$

we have $\mathcal{L}_\rho(\mathbf{x}^{(t)}) - \mathcal{L}_\rho(\mathbf{x}^{(t+1)}) \geq \frac{\rho}{2}\|\mathbf{A}(\mathbf{x}^{(t)} - \mathbf{x}^{(t+1)})\|^2$.

To show that $(\star)$ actually holds, it is sufficient to show that

$$\begin{aligned}&\varphi(\mathbf{g}(\mathbf{x}^{(t+1)}_{<i}, \mathbf{x}^{(t)}_{\geq i})) \\ &\geq \varphi(\mathbf{g}(\mathbf{x}^{(t+1)}_{\leq i}, \mathbf{x}^{(t)}_{>i})) - \langle d^{(t)}_i, x^{(t+1)}_i - x^{(t)}_i \rangle \qquad (\star')\end{aligned}$$

holds, since the iterative evaluation of this inequality starting from $i = 1$ to $i = \dim(\mathbf{x})$ finally leads $(\star)$. Here, we used the aforementioned fact that $d^{(t)}_i$ depends only on $x_i$ among $\mathbf{x}$ and so $\sum_i \langle d^{(t)}_i, x^{(t)}_i - x^{(t+1)}_i \rangle = \langle \mathbf{d}^{(t)}, \mathbf{x}^{(t)} - \mathbf{x}^{(t+1)} \rangle$. We show that the inequality $(\star')$ actually holds by two cases: $x^{(t+1)}_i > 0$ and $x^{(t)}_i > 0$.

When $x^{(t+1)}_i > 0$, since $h(x^{(t+1)}_i) = 0$ and $\partial_i g(x^{(t+1)}_i) = 1$, we have $d^{(t)}_i = \frac{\partial \varphi(\mathbf{g}(\mathbf{x}^{(t+1)}_{<i}, \mathbf{x}^{(t)}_{\geq i}))}{\partial g_i}$. The inequality $(\star')$ holds from the concavity of $\varphi$ w.r.t. $g(x_i)$ and that $g(x^{(t)}_i) = x^{(t)}_i$, $g(x^{(t+1)}_i) = x^{(t+1)}_i$.

When $x^{(t+1)}_i = 0$, we first note that $x^{(t+1)}_i - x^{(t)}_i \leq 0$. Since we consider functions that take finite values, by considering the epigraph sectioned by $x^{(t+1)}_i$, we have

$$\begin{aligned} d^{(t)}_i &= \partial_i \left( \frac{\partial \varphi(g(\mathbf{x}^{(t+1)}_{<i}, \mathbf{x}^{(t)}_{\geq i}))}{\partial g_i} g(x^{(t+1)}_i) + h(x^{(t+1)}_i) \right) \\ &= \left\{ \boldsymbol{\xi} + \frac{\partial \varphi(g(\mathbf{x}^{(t+1)}_{<i}, \mathbf{x}^{(t)}_{\geq i}))}{\partial g_i} \partial_i g(x^{(t+1)}_i); \boldsymbol{\xi} \leq 0 \right\}, \end{aligned}$$

where $\partial_i g(x^{(t+1)}_i) = [-1, 1]$. Using this fact, we have

$$\begin{aligned} &\varphi(\mathbf{g}(\mathbf{x}^{(t+1)}_{<i}, \mathbf{x}^{(t)}_{\geq i})) \\ &\geq \varphi(\mathbf{g}(\mathbf{x}^{(t+1)}_{\leq i}, \mathbf{x}^{(t)}_{>i})) - \langle d^{(t)}_i, g(x^{(t+1)}_i) - g(x^{(t)}_i) \rangle \\ &= \varphi(\mathbf{g}(\mathbf{x}^{(t+1)}_{\leq i}, \mathbf{x}^{(t)}_{>i})) - \langle d^{(t)}_i, x^{(t+1)}_i - x^{(t)}_i \rangle \end{aligned}$$

which leads $(\star')$. We used the fact that $\partial_i g(\mathbf{x}^{(t+1)}) \leq 1$, $x^{(t+1)}_i - x^{(t)}_i \leq 0$, and the concavity of $\varphi$ w.r.t. $g(x_i)$ to show the inequality. This concludes the proof. □

### 4.3 Proof of Lemma 3

*Proof.*

$$\begin{aligned} &\mathcal{L}_\rho(\mathbf{x}^{(t+1)}, \mathbf{y}^{(t)}, \mathbf{p}^{(t)}) - \mathcal{L}_\rho(\mathbf{x}^{(t+1)}, \mathbf{y}^{(t+1)}, \mathbf{p}^{(t)}) \\ &= f(\mathbf{y}^{(t)}) - f(\mathbf{y}^{(t+1)}) + \langle \mathbf{p}^{(t)}, \mathbf{B}(\mathbf{y}^{(t)} - \mathbf{y}^{(t+1)}) \rangle \\ &+ \frac{\rho}{2}\|\mathbf{A}\mathbf{x}^{(t+1)} + \mathbf{B}\mathbf{y}^{(t)}\|^2 - \frac{\rho}{2}\|\mathbf{A}\mathbf{x}^{(t+1)} + \mathbf{B}\mathbf{y}^{(t+1)}\|^2 \\ &= f(\mathbf{y}^{(t)}) - f(\mathbf{y}^{(t+1)}) - \langle \nabla f(\mathbf{y}^{(t+1)}), \mathbf{y}^{(t)} - \mathbf{y}^{(t+1)} \rangle \\ &+ \frac{\rho}{2}\|\mathbf{B}(\mathbf{y}^{(t)} - \mathbf{y}^{(t+1)})\|^2, \qquad \text{(P3-1)} \end{aligned}$$

where we used the fact that $\mathbf{B}^\top \mathbf{p}^{(t+1)} = -\nabla f(\mathbf{y}^{(t+1)})$ to show the final equality. Using the Lipschitz property of $f$:

$$\begin{aligned} &f(\mathbf{y}^{(t)}) - f(\mathbf{y}^{(t+1)}) - \langle \nabla f(\mathbf{y}^{(t+1)}), \mathbf{y}^{(t)} - \mathbf{y}^{(t+1)} \rangle \\ &\geq -\frac{L_f}{2}\|\mathbf{y}^{(t)} - \mathbf{y}^{(t+1)}\|^2 \qquad \text{(P3-2)} \end{aligned}$$

Next, we evaluate

$$\begin{aligned} &\mathcal{L}_\rho(\mathbf{x}^{(t+1)}, \mathbf{y}^{(t+1)}, \mathbf{p}^{(t)}) - \mathcal{L}_\rho(\mathbf{x}^{(t+1)}, \mathbf{y}^{(t+1)}, \mathbf{p}^{(t+1)}) \\ &= \langle \mathbf{p}^{(t)} - \mathbf{p}^{(t+1)}, \mathbf{A}\mathbf{x}^{(t+1)} + \mathbf{B}\mathbf{y}^{(t+1)} \rangle \\ &= -\frac{\rho}{2}\|\mathbf{p}^{(t)} - \mathbf{p}^{(t+1)}\| \\ &\geq -C_{\mathbf{B},f}\|\mathbf{y}^{(t+1)} - \mathbf{y}^{(t)}\| \qquad \text{(P3-3)} \end{aligned}$$

By combining (P3-1) to (P3-3), we conclude the claim. □

### 4.4 Proof of Lemma 4

*Proof.* Since $\mathrm{Im}(A) \subseteq \mathrm{Im}(B)$, there exists some $\mathbf{y}'$ that satisfies $\mathbf{A}\mathbf{x}^{(t)} = -\mathbf{B}\mathbf{y}'$ and

$$\begin{aligned} \mathcal{L}_\rho &= f(\mathbf{y}^{(t)}) + \varphi(\mathbf{g}(\mathbf{x})) + h(\mathbf{x}^{(t)}) + \langle \mathbf{p}^{(t)}, \mathbf{A}\mathbf{x}^{(t)} + \mathbf{B}\mathbf{y}^{(t)} \rangle \\ &+ \frac{\rho}{2}\|\mathbf{A}\mathbf{x}^{(t)} + \mathbf{B}\mathbf{y}^{(t)}\|^2 \\ &= f(\mathbf{y}') + \varphi(\mathbf{g}(\mathbf{x})) + h(\mathbf{x}^{(t)}) + \frac{\rho}{2}\|\mathbf{A}\mathbf{x}^{(t)} + \mathbf{B}\mathbf{y}^{(t)}\|^2 \\ &+ \left[ f(\mathbf{y}^{(t)}) - f(\mathbf{y}') + \langle \nabla f(\mathbf{y}^{(t)}), \mathbf{y}' - \mathbf{y}^{(t)} \rangle \right] \\ &\geq f(\mathbf{y}') + \varphi(\mathbf{g}(\mathbf{x})) + h(\mathbf{x}^{(t)}) + \frac{\rho}{2}\|\mathbf{A}\mathbf{x}^{(t)} + \mathbf{B}\mathbf{y}^{(t)}\|^2 \\ &- \frac{L_f}{2}\sigma_{\min}\|\mathbf{B}\mathbf{y}' - \mathbf{B}\mathbf{y}^{(t)}\|^2 \\ &= f(\mathbf{y}') + \varphi(\mathbf{g}(\mathbf{x})) + h(\mathbf{x}^{(t)}) + \frac{\rho}{2}\|\mathbf{A}\mathbf{x}^{(t)} + \mathbf{B}\mathbf{y}^{(t)}\|^2 \\ &- \frac{L_f}{2}\sigma_{\min}\|\mathbf{A}\mathbf{x}^{(t)} - \mathbf{B}\mathbf{y}^{(t)}\|^2 \\ &\geq -\infty \;\text{ if }\; \rho > L_f\sigma_{\min}. \end{aligned}$$

This concludes the proof. □

### 4.5 Proof of Lemma 5

*Proof.* Using the optimality and Equation 21,

$$\begin{aligned} \nabla_{\mathbf{y}}\mathcal{L}_\rho(\boldsymbol{\vartheta}^{(t+1)}) &= \nabla f(\mathbf{y}^{(t+1)}) + \mathbf{B}^\top \mathbf{p}^{(t+1)} + \rho \mathbf{B}^{(\top)}\left(\mathbf{A}\mathbf{x}^{(t+1)} + \mathbf{B}\mathbf{y}^{(t+1)}\right) \\ &= \mathbf{B}^\top(\mathbf{p}^{(t+1)} - \mathbf{p}^{(t)}) \end{aligned}$$

The second equation is a direct consequence of derivative. Both of them go to zeros as $t \to \infty$ from Lemma 4 and the combination of Equation 23-25.

Finally,

$$\begin{aligned} \frac{\partial \mathcal{L}_\rho}{\partial x_i}(\boldsymbol{\vartheta}^{(t+1)}) &= \partial_i \left\{ \varphi(\mathbf{g}(\mathbf{x})) + h(\mathbf{x}) \right\}_{\mathbf{x}^{(t+1)}} \\ &+ \left(\mathbf{A}^\top \mathbf{p}^{(t+1)}\right)_i + \rho\left(\mathbf{A}^\top(\mathbf{x}^{(t+1)} + \mathbf{B}\mathbf{y}^{(t+1)})\right)_i. \end{aligned}$$

From the optimality of $\tilde{\mathcal{L}}_\rho$ w.r.t. $x_i$, we have

$$\begin{aligned} 0 \in \partial &\left\{ \frac{\varphi(\mathbf{g}(\mathbf{x}^{(t+1)}_{<i}, \mathbf{x}^{(t)}_{\geq i}))}{g_i} g(\mathbf{x}) + h(\mathbf{x}) \right\}_{x^{(t+1)}_i} \\ &+ \left(\mathbf{A}^\top \mathbf{p}^{(t)}\right)_i + \rho\left(\mathbf{A}^\top(\mathbf{x}^{(t+1)} + \mathbf{B}\mathbf{y}^{(t)})\right)_i. \end{aligned}$$

Then, we have Equation 28

$$\begin{aligned} &\left[ \partial_i \left\{ \varphi(\mathbf{g}(\mathbf{x})) + h(\mathbf{x}) \right\}_{\mathbf{x}^{(t+1)}} - \partial \left\{ \frac{\varphi(\mathbf{g}(\mathbf{x}^{(t+1)}_{<i}, \mathbf{x}^{(t)}_{\geq i}))}{g_i} g(\mathbf{x}) + h(\mathbf{x}) \right\}_{x^{(t+1)}_i} \right] \\ &+ \mathbf{A}^\top(\mathbf{p}^{(t+1)} - \mathbf{p}^{(t)}) + \rho\mathbf{B}\left(\mathbf{y}^{(t+1)} - \mathbf{y}^{(t)}\right) \in \frac{\partial \mathcal{L}_\rho}{\partial x_i}(\boldsymbol{\vartheta}^{(t+1)}) \end{aligned}$$

holds, where the second and third terms go to zeros as $t \to \infty$ again from Lemma 4 and Equation 23-25, $0 \in \frac{\partial \tilde{\mathcal{L}}_\rho}{\partial x_i}(\boldsymbol{\vartheta}^{(t+1)})$, and finally for the first term, from the facts that $\|\mathbf{x}^{(\mathbf{t})} - \mathbf{x}^{)(t+1)}\| \to 0$ as $t \to \infty$ and their epigraphs are the same [18], from the continuity of $g$, the first term also goes to zero as $t \to \infty$.

This concludes the proof. □

## References


[1] Keisuke Ozawa. On the convergence of linearized admm for separable reweighted sparse hyperspectral unmixing. In *2022 12th Workshop on Hyperspectral Imaging and Signal Processing: Evolution in Remote Sensing (WHISPERS)*, pages 1–5. IEEE, 2022.

[2] Marian-Daniel Iordache, José M Bioucas-Dias, and Antonio Plaza. Sparse unmixing of hyperspectral data. *IEEE Transactions on Geoscience and Remote Sensing*, 49(6): 2014–2039, 2011.

[3] José M Bioucas-Dias, Antonio Plaza, Nicolas Dobigeon, Mario Parente, Qian Du, Paul Gader, and Jocelyn Chanussot. Hyperspectral unmixing overview: Geometrical, statistical, and sparse regression-based approaches. *IEEE journal of selected topics in applied earth observations and remote sensing*, 5(2):354–379, 2012.

[4] Danfeng Hong, Wei He, Naoto Yokoya, Jing Yao, Lianru Gao, Liangpei Zhang, Jocelyn Chanussot, and Xiao Zhu. Interpretable Hyperspectral Artificial Intelligence: When nonconvex modeling meets hyperspectral remote sensing. In *IEEE Geoscience and Remote Sensing Magazine*, volume 9-2, pages 52–87, 2021.

[5] Alina Zare and KC Ho. Endmember variability in hyperspectral analysis: Addressing spectral variability during spectral unmixing. *IEEE Signal Processing Magazine*, 31 (1):95–104, 2013.

[6] Marian-Daniel Iordache, José M Bioucas-Dias, and Antonio Plaza. Total variation spatial regularization for sparse hyperspectral unmixing. *IEEE Transactions on Geoscience and Remote Sensing*, 50(11):4484–4502, 2012.

[7] José M. Bioucas-Dias and Mario A. T. Figueiredo. Alternating direction algorithms for constrained sparse regression: Application to hyperspectral unmixing. In *2nd Workshop on Hyperspectral Image and Signal Processing: Evolution in Remote Sensing WHISPERS*, pages 21–24, 2010.

[8] Mario Parente and Marian-Daniel Iordache. Sparse Unmixing of Hyperspectral Data: The Legacy of SUnSAL. In *IEEE International Geoscience and Remote Sensing Symposium IGARSS*, pages 21–24, 2021.

[9] Marian-Daniel Iordache, José M Bioucas-Dias, and Antonio Plaza. Collaborative sparse regression for hyperspectral unmixing. *IEEE Transactions on geoscience and remote sensing*, 52(1):341–354, 2013.

[10] Emmanuel J. Candès, Michael B. Wakin, and Stephen P. Boyd. Enhancing Sparsity by Reweighted $l_1$ Minimization. In *Journal of Fourier Analysis and Applications*, volume 14, page 877–905, 2008.

[11] Rui Wang, Heng-Chao Li, Wenzhi Liao, and Aleksandra Pižurica. Double reweighted sparse regression for hyperspectral unmixing. In *2016 IEEE International Geoscience and Remote Sensing Symposium (IGARSS)*, pages 6986–6989. IEEE, 2016.

[12] Ivan W Selesnick and Ilker Bayram. Sparse signal estimation by maximally sparse convex optimization. *IEEE Transactions on Signal Processing*, 62(5):1078–1092, 2014.

[13] Jonathan Eckstein and Dimitri P. Bertsekas. On the Douglas-Rachford splitting method and the proximal point algorithm for maximal monotone operators. In *Mathematical Programming*, volume 5, page 293–318, 1992.

[14] Stephen Boyd, Neal Parikh, Eric Chu, Borja Peleato, and Jonathan Eckstein. Distributed optimization and statistical learning via the alternating direction method of multipliers. In *Foundations and Trends in Machine Learning*, volume 3, page 1–122, 2011.

[15] Yu Wang, Watao Yin, and Jinshan Zeng. Global Convergence of ADMM in Nonconvex Nonsmooth Optimization. In *Journal of Scientific Computing*, volume 78, page 29–63, 2019.

[16] Robert Tibshirani. Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society: Series B (Methodological)*, 58:267–288, 1996.

[17] Qinghua Liu, Xinyue Shen, and Yuantao Gu. Linearized admm for nonconvex nonsmooth optimization with convergence analysis. *IEEE access*, 7:76131–76144, 2019.

[18] R Tyrrell Rockafellar and Roger J-B Wets. *Variational analysis*, volume 317. Springer Science & Business Media, 2009.