Comments on "Automatic Target Detection for Sparse Hyperspectral Images" by Ahmad W. Bitar et al.

Ahmad W. Bitar, Ali Chehab, Jean-Philippe Ovarlez

Maroun Semaan Faculty of Engineering and Architecture

Comments on «Automatic Target Detection for Sparse Hyperspectral Images [1]» by Ahmad W. Bitar et al.

American University of Beirut

Technical Report

Ali Chehab
American University of Beirut
chehab@aub.edu.lb

Ahmad W. Bitar
American University of Beirut
ab76@aub.edu.lb

Jean-Philippe Ovarlez
ONERA & CentraleSupélec
jeanphilippe@centralesupelec.fr

June 3, 2020

Summary

In this technical report, we explain how our proposed sparse and low-rank matrix decomposition method for hyperspectral target detection, provided in our work «Automatic Target Detection for Sparse Hyperspectral Images [1]», can be extended to the l_q norm $(0 < q \leq 1)$. Since the use of the l_1 norm is still too far away from the ideal l_0 norm, many non-convex regularizers, interpolated between the l_0 norm and the l_1 norm, have been proposed to better approximate the l_0 norm.

Main Notations

Throughout this report, we depict vectors in lowercase boldface letters and matrices in uppercase boldface letters. The notation $(.)^T$ and $\mathrm{Tr}(.)$ stand for the transpose and trace of a matrix, respectively. In addition, $\mathrm{rank}(.)$ is for the rank of a matrix. A variety of norms on matrices will be used. For instance, $\mathbf{M}$ is a matrix, and $[\mathbf{M}]_{:,j}$ is the jth column. The matrix $l_{2,0}$ and $l_{2,q}$ $(0 < q \leq 1)$ norms are defined by $\|\mathbf{M}\|_{2,0} = \# \left\{ j : \left\| [\mathbf{M}]_{:,j} \right\|_2 \neq 0 \right\}$, and $\|\mathbf{M}\|_{2,q} = \left(\sum_j \left\| [\mathbf{M}]_{:,j} \right\|_2^q \right)^{(1/q)}$, respectively. The Frobenius norm and the nuclear norm (the sum of singular values of a matrix) are denoted by $\|\mathbf{M}\|_F$ and $\|\mathbf{M}\|_* = \mathrm{Tr} \left(\mathbf{M}^T \mathbf{M} \right)^{(1/2)}$, respectively.

1 Main contribution

Consider the following minimization problem:

$$\min_{\mathbf{L},\mathbf{C}} \left\{ \tau \operatorname{rank}(\mathbf{L}) + \lambda \left\|\mathbf{C}\right\|_{2,0} + \left\| \mathbf{D} - \mathbf{L} - (\mathbf{A}_t\mathbf{C})^T \right\|_F^2 \right\}, \tag{1}$$

where $\mathbf{D}$, $\mathbf{L}$, $(\mathbf{A}_t\mathbf{C})^T \in \mathbb{R}^{e\times p}$, $\mathbf{A}_t \in \mathbb{R}^{p\times N_t}$, $\mathbf{C} \in \mathbb{R}^{N_t\times e}$, τ controls the rank of $\mathbf{L}$, and λ the sparsity level in $\mathbf{C}$.

We relax the rank term and the $l_{2,0}$ norm to their convex proxies [1, 2, 3, 4, 5]. More precisely, we use the nuclear norm $||\mathbf{L}||_*$ as a surrogate for the rank($\mathbf{L}$) term, and the $l_{2,1}$ norm $\|\mathbf{C}\|_{2,1}$ as a surrogate for the $l_{2,0}$ norm $\|\mathbf{C}\|_{2,0}$.

$$\min_{\mathbf{L},\mathbf{C}} \left\{ \tau \left\|\mathbf{L}\right\|_* + \lambda \left\|\mathbf{C}\right\|_{2,1} + \left\| \mathbf{D} - \mathbf{L} - (\mathbf{A}_t\mathbf{C})^T \right\|_F^2 \right\}. \tag{2}$$

Problem (2) can be re-written as

$$\min_{\mathbf{L},\mathbf{C}} \left\{ \tau \sum_{i=1}^{\min(e,p)} \sigma_i(\mathbf{L}) + \lambda \sum_{j=1}^{e} \|[\mathbf{C}]_{:,j}\|_2 + \left\| \mathbf{D} - \mathbf{L} - (\mathbf{A}_t\mathbf{C})^T \right\|_F^2 \right\}, \tag{3}$$

where $\{\sigma_i\}_{i=1}^{\min(e,p)}$ are the singular values of $\mathbf{L}$.

Extension to the l_q norm $(0 < q \leq 1)$

We replace the nuclear norm and the $l_{2,1}$ norm by their q - norm proxies in Eq. (2), with $0 < q \leq 1$. More precisely [6, 7, 8]

$$\min_{\mathbf{L},\mathbf{C}} \left\{ \tau \sum_{i=1}^{\min(e,p)} (\sigma_i(\mathbf{L}) + \epsilon)^q + \lambda \sum_{j=1}^{e} \|[\mathbf{C}]_{:,j}\|_2^q + \left\| \mathbf{D} - \mathbf{L} - (\mathbf{A}_t\mathbf{C})^T \right\|_F^2 \right\}, \tag{4}$$

where $0 < \epsilon \ll 1$. Problem (4) is recasted into two sub-problems, and thus, at each iteration k we have

$$\min_{\mathbf{L}} \left\{ \left\| \mathbf{L} - \left(\mathbf{D} - \left(\mathbf{A}_t\, \mathbf{C}^{(k-1)} \right)^T \right) \right\|_F^2 + \tau \sum_{i=1}^{\min(e,\,p)} (\sigma_i(\mathbf{L}) + \epsilon)^q \right\}, \tag{5a}$$

$$\min_{\mathbf{C}} \left\{ \left\| \left(\mathbf{D} - \mathbf{L}^{(k)} \right)^T - \mathbf{A}_t\, \mathbf{C} \right\|_F^2 + \lambda \sum_{j=1}^{e} \|[\mathbf{C}]_{:,j}\|_2^q \right\}. \tag{5b}$$

1.1 Providing an optimal solution to sub-problem (5a)

For ease of notation, we consider the matrix $\mathbf{E}^{(k-1)} = \left(\mathbf{A}_t\, \mathbf{C}^{(k-1)}\right)^T$. Let us suppose $g_i(\mathbf{L}) = \sigma_i(\mathbf{L}) + \epsilon$, $f(g_i(\mathbf{L})) = (\sigma_i(\mathbf{L}) + \epsilon)^q$, and $h(l_{m,j}) = \left(l_{m,j} - \left(d_{m,j} - e_{m,j}^{(k-1)}\right)\right)^2$, with $i \in [1, \min(e, p)]$, $j \in [1, e]$, and $m \in [1, p]$.

The function $f\left(g_i\left(\mathbf{L}\right)\right)$ is concave, and thus, $-f\left(g_i\left(\mathbf{L}\right)\right)$ is convex. According to the definition of the subgradient of a convex function, we can write [6, 7, 8]

$$-f\left(g_i\left(\mathbf{L}\right)\right) \geq -f\left(g_i\left(\mathbf{L}^{(k-1)}\right)\right) + \left\langle -w_i^{(k-1)},\, g_i\left(\mathbf{L}\right) - g_i\left(\mathbf{L}^{(k-1)}\right)\right\rangle, \tag{6}$$

with $-w_i^{(k-1)} = \partial\left(-f\left(g_i\left(\mathbf{L}^{(k-1)}\right)\right)\right)$ or $w_i^{(k-1)} = -\partial\left(-f\left(g_i\left(\mathbf{L}^{(k-1)}\right)\right)\right)$.
We can re-write Eq. (6) as

$$f\left(g_i(\mathbf{L})\right) \leq f\left(g_i\left(\mathbf{L}^{(k-1)}\right)\right) + \left\langle w_i^{(k-1)},\, g_i\left(\mathbf{L}\right) - g_i\left(\mathbf{L}^{(k-1)}\right)\right\rangle. \tag{7}$$

The loss function $h\left(l_{m,j}\right)$ has a Lipshitz continous gradient, and thus, we can surrogate it as

$$h\left(l_{m,j}\right) \leq h\left(l_{m,j}^{(k-1)}\right) + \left\langle \nabla h\left(l_{m,j}^{(k-1)}\right),\, l_{m,j} - l_{m,j}^{(k-1)}\right\rangle + \frac{\mu}{2}\left(l_{m,j} - l_{m,j}^{(k-1)}\right)^2, \tag{8}$$

with $\mu > 0$. By combining Eqs. (7) and (8), the sub-problem (5a) is approximated as

$$\begin{aligned}
&\min_{\mathbf{L}}\left\{\tau \sum_{i=1}^{\min(e,p)}\left[\left(\sigma_i\left(\mathbf{L}^{(k-1)}\right)+\epsilon\right)^q + \left\langle w_i^{(k-1)},\, \sigma_i\left(\mathbf{L}\right) - \sigma_i\left(\mathbf{L}^{(k-1)}\right)\right\rangle\right]\right.\\
&\qquad\left. + \sum_{m=1}^{p}\sum_{j=1}^{e}\left[h\left(l_{m,j}^{(k-1)}\right) + \left\langle \nabla h\left(l_{m,j}^{(k-1)}\right),\, l_{m,j} - l_{m,j}^{(k-1)}\right\rangle + \frac{\mu}{2}\left(l_{m,j} - l_{m,j}^{(k-1)}\right)^2\right]\right\},\\
&\implies \min_{\mathbf{L}}\left\{\tau \sum_{i=1}^{\min(e,p)}\left[\left\langle w_i^{(k-1)},\, \sigma_i(\mathbf{L})\right\rangle\right]\right.\\
&\qquad\left. + \frac{\mu}{2}\sum_{m=1}^{p}\sum_{j=1}^{e}\left[l_{m,j} - \left(l_{m,j}^{(k-1)} - \frac{1}{\mu}\nabla h\left(l_{m,j}^{(k-1)}\right)\right)\right]^2\right\},\\
&\implies \min_{\mathbf{L}}\left\{\tau \sum_{i=1}^{\min(e,p)}\left[\left\langle w_i^{(k-1)},\, \sigma_i(\mathbf{L})\right\rangle\right]\right.\\
&\qquad\left. + \frac{\mu}{2}\sum_{m=1}^{p}\sum_{j=1}^{e}\left[l_{m,j} - \left(l_{m,j}^{(k-1)} - \frac{2}{\mu}\left(l_{m,j}^{(k-1)} - \left(d_{m,j} - e_{m,j}^{(k-1)}\right)\right)\right)\right]^2\right\},\\
&\implies \min_{\mathbf{L}}\left\{\tau \sum_{i=1}^{\min(e,p)}\left[\left\langle w_i^{(k-1)},\, \sigma_i\left(\mathbf{L}\right)\right\rangle\right]\right.\\
&\qquad\left. + \frac{\mu}{2}\left\|\mathbf{L} - \left(\mathbf{L}^{(k-1)} - \frac{2}{\mu}\left(\mathbf{L}^{(k-1)} - \left(\mathbf{D} - \mathbf{E}^{(k-1)}\right)\right)\right)\right\|_F^2\right\},
\end{aligned} \tag{9}$$

with $w_i^{(k-1)} = -\partial\left(-\left(\sigma_i\left(\mathbf{L}^{(k-1)}\right)+\epsilon\right)^q\right) = q\left(\sigma_i\left(\mathbf{L}^{(k-1)}\right)+\epsilon\right)^{q-1} = \frac{q}{\left(\sigma_i\left(\mathbf{L}^{(k-1)}\right)+\epsilon\right)^{1-q}}$.

Let us consider that $\mathbf{X}^{(k-1)} = \mathbf{L}^{(k-1)} - \frac{2}{\mu}\left(\mathbf{L}^{(k-1)} - \left(\mathbf{D} - \mathbf{E}^{(k-1)}\right)\right)$. Given $\mathbf{X}^{(k-1)} \in \mathbb{R}^{e\times p}$, $0 \leq w_1^{(k-1)} \leq \cdots \leq w_{\min(e,p)}^{(k-1)}$, and according to Theorem 2.3 in [9], the global optimal "unique" solution (if $\mathbf{X}^{(k-1)}$ has a unique singular value decomposition (SVD)) to the above optimization problem (9) is given by the adaptive SVD soft-thresholding operator

$$\boxed{\mathbf{L}^{(k)} = \mathcal{S}_{\frac{\tau\,\mathbf{w}^{(k-1)}}{\mu}}\left(\mathbf{X}^{(k-1)}\right) = \mathbf{U}^{(k-1)}\,\mathcal{S}_{\frac{\tau\,\mathbf{w}^{(k-1)}}{\mu}}\left(\mathbf{\Sigma}^{(k-1)}\right)\,\mathbf{V}^{(k-1)T}}$$

with $\mathbf{X}^{(k-1)} = \mathbf{U}^{(k-1)}\,\mathbf{\Sigma}^{(k-1)}\,\mathbf{V}^{(k-1)T}$, and
$\mathcal{S}_{\frac{\tau\,\mathbf{w}^{(k-1)}}{\mu}}\left(\mathbf{\Sigma}^{(k-1)}\right) = \mathrm{Diag}\left\{\left(\sigma_i\left(\mathbf{X}^{(k-1)}\right) - \frac{\tau\,w_i^{(k-1)}}{\mu}\right)_+ ,\ i \in [1,\,\min(e,\,p)]\right\}$.

Proof. Let $\mathbf{g} = \{g_i\}_{i=1}^{\min(e,p)} = \boldsymbol{\sigma}(\mathbf{L})$. According to Theorem 2.3 in [9], the optimization problem (9) can be equivalently written as

$$\min_{\mathbf{g}:\,g_1\geq\cdots\geq g_{\min(e,p)}\geq 0}\left\{\min_{\substack{\mathbf{L}\in\mathbb{R}^{e\times p}\\ \boldsymbol{\sigma}(\mathbf{L})=\mathbf{g}}}\left\{\frac{\mu}{2}\left\|\mathbf{L}-\mathbf{X}^{(k-1)}\right\|_F^2\right\} + \tau\sum_{i=1}^{\min(e,p)} w_i^{(k-1)}\,g_i\right\}. \tag{10}$$

For the inner minimization, we have the inequality

$$\begin{aligned}
\frac{\mu}{2}\left\|\mathbf{L}-\mathbf{X}^{(k-1)}\right\|_F^2 &= \frac{\mu}{2}\,\mathrm{Tr}\left[\left(\mathbf{L}-\mathbf{X}^{(k-1)}\right)\left(\mathbf{L}-\mathbf{X}^{(k-1)}\right)^T\right]\\
&= \frac{\mu}{2}\,\mathrm{Tr}\left[\mathbf{L}\mathbf{L}^T - \mathbf{L}\mathbf{X}^{(k-1)T} - \mathbf{X}^{(k-1)}\mathbf{L}^T + \mathbf{X}^{(k-1)}\mathbf{X}^{(k-1)T}\right]\\
&= \frac{\mu}{2}\,\mathrm{Tr}\left(\mathbf{L}\mathbf{L}^T\right) + \frac{\mu}{2}\,\mathrm{Tr}\left(\mathbf{X}^{(k-1)}\mathbf{X}^{(k-1)T}\right) - \mu\,\mathrm{Tr}\left(\mathbf{X}^{(k-1)}\mathbf{L}^T\right)\\
&= \frac{\mu}{2}\sum_{i=1}^{\min(e,p)} g_i^2 + \frac{\mu}{2}\sum_{i=1}^{\min(e,p)}\sigma_i^2\left(\mathbf{X}^{(k-1)}\right) - \mu\,\mathrm{Tr}\left(\mathbf{X}^{(k-1)}\,\mathbf{L}^T\right)\\
&\geq \frac{\mu}{2}\sum_{i=1}^{\min(e,p)} g_i^2 + \frac{\mu}{2}\sum_{i=1}^{\min(e,p)}\sigma_i^2\left(\mathbf{X}^{(k-1)}\right) - \mu\sum_{i=1}^{\min(e,p)} g_i\,\sigma_i\left(\mathbf{X}^{(k-1)}\right).
\end{aligned}$$

The optimization problem (10) is re-written as

$$\begin{aligned}
&\min_{\mathbf{g}:\,g_1\geq\cdots\geq g_{\min(e,p)}\geq 0}\sum_{i=1}^{\min(e,p)}\left(\frac{\mu}{2}\,g_i^2 + \frac{\mu}{2}\,\sigma_i^2\left(\mathbf{X}^{(k-1)}\right) - \mu\,g_i\,\sigma_i\left(\mathbf{X}^{(k-1)}\right) + \tau\,w_i^{(k-1)}\,g_i\right),\\
&\min_{\mathbf{g}:\,g_1\geq\cdots\geq g_{\min(e,p)}\geq 0}\sum_{i=1}^{\min(e,p)}\left(\frac{\mu}{2}\,g_i^2 + \left[-\mu\,\sigma_i\left(\mathbf{X}^{(k-1)}\right) + \tau\,w_i^{(k-1)}\right]g_i + \frac{\mu}{2}\,\sigma_i^2\left(\mathbf{X}^{(k-1)}\right)\right).
\end{aligned} \tag{11}$$

By computing the derivative w.r.t. g_i and setting it to zero, we have

$$\mu\,g_i - \mu\,\sigma_i\left(\mathbf{X}^{(k-1)}\right) + \tau\,w_i^{(k-1)} = 0\,,$$

and thus, the optimal solution to Eq. (11) is given by

$$g_i = \left(\sigma_i\left(\mathbf{X}^{(k-1)}\right) - \frac{\tau\,w_i^{(k-1)}}{\mu}\right)_+ .$$

Hence, the global optimal unique solution to the optimization problem (9) is given by $\mathbf{L}^{(k)} = \mathbf{U}^{(k-1)}\ \mathrm{Diag}\left(\left\{\left(\boldsymbol{\sigma}\left(\mathbf{X}^{(k-1)}\right) - \frac{\tau\,\mathbf{w}^{(k-1)}}{\mu}\right)_+\right\}\right)\ \mathbf{V}^{(k-1)T}$, and which concludes the proof.

■

1.2 Providing an optimal solution to sub-problem (5b)

Eq. (5b) can be solved by various methods, among which we adopt the alternating direction method of multipliers (ADMM) [10]. More precisely, we introduce an auxiliary variable $\mathbf{F}$ into sub-problem (5b) and recast it into the following form

$$\left(\mathbf{C}^{(k)}, \mathbf{F}^{(k)}\right) = \underset{s.t.\quad \mathbf{C}=\mathbf{F}}{\mathrm{argmin}} \left\{\left\|\left(\mathbf{D} - \mathbf{L}^{(k)}\right)^T - \mathbf{A}_t\,\mathbf{C}\right\|_F^2 + \lambda \sum_{j=1}^{e} \|[\mathbf{F}]_{:,j}\|_2^q\right\}. \tag{12}$$

Problem (12) is then solved as

$$\mathbf{C}^{(k)} = \underset{\mathbf{C}}{\mathrm{argmin}} \left\{\left\|\left(\mathbf{D} - \mathbf{L}^{(k)}\right)^T - \mathbf{A}_t\,\mathbf{C}\right\|_F^2 + \frac{\rho^{(k-1)}}{2}\left\|\mathbf{C} - \mathbf{F}^{(k-1)} + \frac{1}{\rho^{(k-1)}}\,\mathbf{Z}^{(k-1)}\right\|_F^2\right\}, \tag{13a}$$

$$\mathbf{F}^{(k)} = \underset{\mathbf{F}}{\mathrm{argmin}} \left\{\lambda \sum_{j=1}^{e} \|[\mathbf{F}]_{:,j}\|_2^q + \frac{\rho^{(k-1)}}{2}\left\|\mathbf{C}^{(k)} - \mathbf{F} + \frac{1}{\rho^{(k-1)}}\,\mathbf{Z}^{(k-1)}\right\|_F^2\right\}, \tag{13b}$$

$$\mathbf{Z}^{(k)} = \mathbf{Z}^{(k-1)} + \rho^{(k-1)}\left(\mathbf{C}^{(k)} - \mathbf{F}^{(k)}\right). \tag{13c}$$

where $\mathbf{Z} \in \mathbb{R}^{N_t \times e}$ is the Lagrangian multiplier matrix, and ρ is a positive scalar.

1.2.1 Solving sub-problem (13a)

$$-2\,\mathbf{A}_t^T\left(\left(\mathbf{D} - \mathbf{L}^{(k)}\right)^T - \mathbf{A}_t\,\mathbf{C}\right) + \rho^{(k-1)}\left(\mathbf{C} - \mathbf{F}^{(k-1)} + \frac{1}{\rho^{(k-1)}}\,\mathbf{Z}^{(k-1)}\right) = \mathbf{0},$$

$$\Rightarrow \left(2\,\mathbf{A}_t^T\,\mathbf{A}_t + \rho^{(k-1)}\,\mathbf{I}\right)\,\mathbf{C} = \rho^{(k-1)}\,\mathbf{F}^{(k-1)} - \mathbf{Z}^{(k-1)} + 2\,\mathbf{A}_t^T\left(\mathbf{D} - \mathbf{L}^{(k)}\right)^T.$$

This implies:

$$\mathbf{C}^{(k)} = \left(2\,\mathbf{A}_t^T\,\mathbf{A}_t + \rho^{(k-1)}\,\mathbf{I}\right)^{-1}\left(\rho^{(k-1)}\,\mathbf{F}^{(k-1)} - \mathbf{Z}^{(k-1)} + 2\,\mathbf{A}_t^T\left(\mathbf{D} - \mathbf{L}^{(k)}\right)^T\right)$$

1.2.2 Solving sub-problem (13b)

According to Lemma 3.3 in [11] and Lemma 4.1 in [12], problem (13b) admits the following closed-form solution:

$$[\mathbf{F}]_{:,j}^{(k)} = \max\left(\left\|[\mathbf{C}]_{:,j}^{(k)} + \tfrac{1}{\rho^{(k-1)}}\,[\mathbf{Z}]_{:,j}^{(k-1)}\right\|_2^{2-q} - \tfrac{\lambda}{q\,\rho^{(k-1)}},\, 0\right)\left(\frac{[\mathbf{C}]_{:,j}^{(k)} + \tfrac{1}{\rho^{(k-1)}}\,[\mathbf{Z}]_{:,j}^{(k-1)}}{\left\|[\mathbf{C}]_{:,j}^{(k)} + \tfrac{1}{\rho^{(k-1)}}\,[\mathbf{Z}]_{:,j}^{(k-1)}\right\|_2^{2-q}}\right)$$

Proof. At the jth column, sub-problem (13b) refers to

$$[\mathbf{F}]_{:,j}^{(k)} = \underset{[\mathbf{F}]_{:,j}}{\operatorname{argmin}} \left\{ \lambda \left\| [\mathbf{F}]_{:,j} \right\|_2^q + \frac{\rho^{(k-1)}}{2} \left\| [\mathbf{C}]_{:,j}^{(k)} - [\mathbf{F}]_{:,j} + \frac{1}{\rho^{(k-1)}} \; [\mathbf{Z}]_{:,j}^{(k-1)} \right\|_2^2 \right\} .$$

By finding the derivative w.r.t $[\mathbf{F}]_{:,j}$ and setting it to zero, we obtain

$$-\rho^{(k-1)} \left([\mathbf{C}]_{:,j}^{(k)} - [\mathbf{F}]_{:,j} + \frac{1}{\rho^{(k-1)}} \; [\mathbf{Z}]_{:,j}^{(k-1)} \right) + \frac{\lambda \, \partial \left(\left\| [\mathbf{F}]_{:,j} \right\|_2^q \right)}{\partial \, [\mathbf{F}]_{:,j}} = \mathbf{0}$$

$$\Rightarrow [\mathbf{C}]_{:,j}^{(k)} + \frac{1}{\rho^{(k-1)}} \; [\mathbf{Z}]_{:,j}^{(k-1)} = [\mathbf{F}]_{:,j} + \frac{\lambda \, \partial \left(\left\| [\mathbf{F}]_{:,j} \right\|_2^q \right)}{\rho^{(k-1)} \, \partial \, [\mathbf{F}]_{:,j}} . \tag{14}$$

Let $[\mathbf{F}]_{:,j} = [f_{1,j}, \cdots, f_{N_t,j}]^T \in \mathbb{R}^{N_t}$. We have

$$\begin{aligned}
\frac{\partial}{\partial f_{t,j}} \; \left\| [\mathbf{F}]_{:,j} \right\|_2^q &= \frac{\partial}{\partial f_{t,j}} \left(\left(\sum_{s=1}^{N_t} |f_{s,j}|^2 \right)^{1/2} \right)^q \\
&= \frac{\partial}{\partial f_{t,j}} \left(\sum_{s=1}^{N_t} |f_{s,j}|^2 \right)^{q/2} \\
&= \frac{q}{2} \left(\sum_{s=1}^{N_t} |f_{s,j}|^2 \right)^{\frac{q-2}{2}} \times \frac{\partial}{\partial f_{t,j}} \left(\sum_{s=1}^{N_t} |f_{s,j}|^2 \right) \\
&= \frac{q}{2} \left(\left(\sum_{s=1}^{N_t} |f_{s,j}|^2 \right)^{1/2} \right)^{q-2} \times \sum_{s=1}^{N_t} \left(2|f_{s,j}| \times \frac{\partial}{\partial f_{t,j}} |f_{s,j}| \right) \\
&= q \; \left\| [\mathbf{F}]_{:,j} \right\|_2^{q-2} \times \sum_{s=1}^{N_t} |f_{s,j}| \, \delta_{s,t} \, \frac{f_{s,j}}{|f_{s,j}|} = q \; \left\| [\mathbf{F}]_{:,j} \right\|_2^{q-2} \times f_{t,j} \\
&= \frac{f_{t,j}}{q \; \left\| [\mathbf{F}]_{:,j} \right\|_2^{2-q}}, \quad t \in [1, \; N_t] \; .
\end{aligned}$$

This implies $\frac{\partial}{\partial [\mathbf{F}]_{:,j}} \; \left\| [\mathbf{F}]_{:,j} \right\|_2^q = \frac{[\mathbf{F}]_{:,j}}{q \left\| [\mathbf{F}]_{:,j} \right\|_2^{2-q}}$. Hence, Eq. (14) is re-written as

$$[\mathbf{C}]_{:,j}^{(k)} + \frac{1}{\rho^{(k-1)}} \; [\mathbf{Z}]_{:,j}^{(k-1)} = [\mathbf{F}]_{:,j} + \frac{\lambda \; [\mathbf{F}]_{:,j}}{q \, \rho^{(k-1)} \; \left\| [\mathbf{F}]_{:,j} \right\|_2^{2-q}} . \tag{15}$$

By computing the $\|.\|_2^{2-q}$ norm of (15), we obtain

$$\left\| [\mathbf{C}]_{:,j}^{(k)} + \frac{1}{\rho^{(k-1)}} \; [\mathbf{Z}]_{:,j}^{(k-1)} \right\|_2^{2-q} = \left\| [\mathbf{F}]_{:,j} \right\|_2^{2-q} + \frac{\lambda}{q \, \rho^{(k-1)}} . \tag{16}$$

From Eqs. (15) and (16), we have

$$\frac{[\mathbf{C}]_{:,j}^{(k)} + \frac{1}{\rho^{(k-1)}} \; [\mathbf{Z}]_{:,j}^{(k-1)}}{\left\| [\mathbf{C}]_{:,j}^{(k)} + \frac{1}{\rho^{(k-1)}} \; [\mathbf{Z}]_{:,j}^{(k-1)} \right\|_2^{2-q}} = \frac{[\mathbf{F}]_{:,j}}{\left\| [\mathbf{F}]_{:,j} \right\|_2^{2-q}} . \tag{17}$$

Consider that

$$[\mathbf{F}]_{:,j} = \|[\mathbf{F}]_{:,j}\|_2^{2-q} \times \frac{[\mathbf{F}]_{:,j}}{\|[\mathbf{F}]_{:,j}\|_2^{2-q}} \,. \tag{18}$$

By replacing $\|[\mathbf{F}]_{:,j}\|_2^{2-q}$ from Eq. (16) into Eq. (18), and $\frac{[\mathbf{F}]_{:,j}}{\|[\mathbf{F}]_{:,j}\|_2^{2-q}}$ from Eq. (17) into Eq. (18), we conclude the proof. ■

1.3 Some Initializations and Convergence Criterion

We initialize $\mathbf{L}^{(0)} = \mathbf{0}$, $\mathbf{F}^{(0)} = \mathbf{C}^{(0)} = \mathbf{Z}^{(0)} = \mathbf{0}$, $\rho^{(0)} = 10^{-4}$ and update $\rho^{(k)} = 1.1\,\rho^{(k-1)}$. The criteria for convergence of sub-problem (5b) is $\left\|\mathbf{C}^{(k)} - \mathbf{F}^{(k)}\right\|_F^2 \leq 10^{-6}$.

For Problem (4), we stop the iteration when the following convergence criterion is satisfied:

$$\frac{\left\|\mathbf{L}^{(k)} - \mathbf{L}^{(k-1)}\right\|_F}{\|\mathbf{D}\|_F} \leq \epsilon \quad \text{and} \quad \frac{\left\|\left(\mathbf{A}_t\,\mathbf{C}^{(k)}\right)^T - \left(\mathbf{A}_t\,\mathbf{C}^{(k-1)}\right)^T\right\|_F}{\|\mathbf{D}\|_F} \leq \epsilon\,,$$

where $\epsilon > 0$ is a precision tolerance parameter.

References

[1] Bitar AW, Ovarlez J-P, Cheong L-F, and Chehab A, "Automatic Target Detection for Sparse Hyperspectral Images", in: Prasad S., Chanussot J. (eds) Hyperspectral Image Analysis. Advances in Computer Vision and Pattern Recognition. Springer, Cham, Apr 2020. Available on (Arxiv) and (HAL-CentraleSupélec)

[2] Bitar AW, Cheong L-F, and Ovarlez J-P, "Target and Background Separation in Hyperspectral Imagery for Automatic Target Detection", 2018 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Calgary, AB, pp. 1598-1602, Sep 2018. (Available Here)

[3] Bitar AW, Cheong L-F, and Ovarlez J-P, "Sparse and Low-Rank Matrix Decomposition for Automatic Target Detection in Hyperspectral Imagery", in IEEE Transactions on Geoscience and Remote Sensing, vol. 57, no. 8, pp. 5239-5251, Aug 2019. (Available Here)

[4] Bitar AW, Cheong L-F, and Ovarlez J-P, "Simultaneous sparsity-based binary hypothesis model for real hyperspectral target detection", 2017 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), New Orleans, LA, pp. 4616-4620, Mar 2017. (Available Here)

[5] Bitar AW, Ovarlez J-P, and Cheong L-F, "Sparsity-Based cholesky factorization and its application to hyperspectral anomaly detection", 2017 IEEE 7th International Workshop on Computational Advances in Multi-Sensor Adaptive Processing (CAMSAP), Curacao, pp. 1-5, Dec 2017. (Available Here)

[6] Wang J, Wang M, Hu X, and Yan S, "Visual data denoising with a unified Schatten-p norm and l_q norm regularized principal component pursuit", Pattern Recognition, vol. 48, no. 10, pp 3135-3144, Oct 2015.

[7] Canyi L, Yunchao W, Zhouchen L, and Shuicheng Y, "Proximal Iteratively Reweighted Algorithm with Multiple Splitting for Nonconvex Sparsity Optimization", Proceedings of the Twenty-Eighth AAAI Conference on Artificial Intelligence, pp 1251-1257, Dec 2013.

[8] Canyi L, Jinhui T, Shuicheng Y, and Zhouchen L, "Nonconvex Nonsmooth Low-Rank Minimization via Iteratively Reweighted Nuclear Norm", in IEEE Transactions on Image Processing, vol. 25, no. 2, pp. 829-839, Feb 2016.

[9] Chen K, Dong H, and Chan K, "Reduced rank regression via adaptive nuclear norm penalization", Biometrika, vol. 100, no. 4, pp 901-920, Dec 2013.

[10] Boyd S, Parikh N, Chu E, Peleato B, and Eckstein J, "Distributed Optimization and Statistical Learning via the Alternating Direction Method of Multipliers",Found. Trends Mach. Learn., vol. 3, no. 1, pp. 1-122, Jan 2011.

[11] Yang J, Yin W, Zhang Y, and Wang Y, "A Fast Algorithm for Edge-Preserving Variational Multichannel Image Restoration", SIAM Journal on Imaging Sciences, vol. 2, no. 2, pp. 569-592, Apr 2009.

[12] Liu G, Lin Z, Yan S, Sun J, Yu Y, and Ma Y, "Robust Recovery of Subspace Structures by Low-Rank Representation", IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 35, no. 1, pp. 171-184, Jan 2013.