Supplementary Materials for the Paper "A Convergent Incoherent Dictionary Learning Algorithm for Sparse Coding, ECCV, Zurich, 2014"

Chenglong Bao, Yuhui Quan, Hui Ji

Department of Mathematics
National University of Singapore

Abstract. This document accompanies the paper referenced as "A convergent incoherent dictionary learning algorithm for sparse coding, ECCV, Zurich, 2014". We provide a complete proof of Theorem 1 with all the details.

1 Statement of Problem

Minimization model. The paper is about solving the following minimization problem:

$$\min_{\boldsymbol{D},\boldsymbol{C}} \quad \frac{1}{2}\|\boldsymbol{Y}-\boldsymbol{D}\boldsymbol{C}\|_F^2+\lambda\|\boldsymbol{C}\|_0+\frac{\alpha}{2}\|\boldsymbol{D}^\top\boldsymbol{D}-\boldsymbol{I}\|_F^2, \text{ s.t. } \boldsymbol{D}\in\mathcal{D},\ \boldsymbol{C}\in\mathcal{C}, \tag{1}$$

where $\mathcal{D}=\{\boldsymbol{D}=(\boldsymbol{d}_1,\dots,\boldsymbol{d}_m)\in\mathbb{R}^{n\times m}:\|\boldsymbol{d}_j\|_2=1,\ 1\le j\le m\}$ and $\mathcal{C}=\{\boldsymbol{C}=(\boldsymbol{c}_1^\top,\dots,\boldsymbol{c}_m^\top)^\top\in\mathbb{R}^{m\times p},\|\boldsymbol{c}_i\|_\infty\le M,\ 1\le i\le m\}$. Let $\delta_{\mathcal{X}}$ denotes the indicate function of $\mathcal{X}$ such that $\delta_{\mathcal{X}}(x)=0$ if $x\in\mathcal{X}$ and $+\infty$ otherwise. Then, the problem (1) can be re-written as

$$\min_{\boldsymbol{Z}:=(\boldsymbol{C},\boldsymbol{D})} H(\boldsymbol{Z})=F(\boldsymbol{C})+Q(\boldsymbol{Z})+G(\boldsymbol{D}). \tag{2}$$

where

$$\begin{cases} F(\boldsymbol{C})=\lambda\|\boldsymbol{C}\|_0+\delta_{\mathcal{C}}(\boldsymbol{C}),\\ Q(\boldsymbol{C},\boldsymbol{D})=\frac{1}{2}\|\boldsymbol{Y}-\boldsymbol{D}\boldsymbol{C}\|_F^2+\frac{\alpha}{2}\|\boldsymbol{D}^\top\boldsymbol{D}-\boldsymbol{I}\|_F^2,\\ G(\boldsymbol{D})=\delta_{\mathcal{D}}(\boldsymbol{D}). \end{cases} \tag{3}$$

Algorithm 1. Based on the so-called *proximal operator* [5] defined by

$$\mathrm{P}rox_t^F(x):=\arg\min_u F(u)+\frac{t}{2}\|u-x\|_F^2,$$

the proposed hybrid alternating proximal algorithm for solving (2) is summarized as follows,

$$\begin{cases} \boldsymbol{c}_j^{(k+1)}\in\mathrm{P}rox_{\mu_j^{k+1}}^{F(\boldsymbol{U}_j^{k+1})+Q(\boldsymbol{U}_j^{k+1},\boldsymbol{D}^{(k)})}(\boldsymbol{c}_j^{(k)}), \quad 1\le j\le m,\\ \boldsymbol{d}_j^{(k+1)}\in\mathrm{P}rox_{\lambda_j^{k+1}}^{G(\boldsymbol{S}_j^{k+1})}(\boldsymbol{d}_j^{(k)}-\frac{1}{\lambda_j^{k+1}}\nabla_{\boldsymbol{d}_j}Q(\boldsymbol{C}^{(k)},\boldsymbol{V}_j^{k+1})), \end{cases}, \quad 1\le j\le m, \tag{4}$$

where $\mu_j^{k+1}, \lambda_j^{k+1} \in (a,b)$, a, b are some positive constants and

$$\begin{cases} \boldsymbol{U}_j^k = (\boldsymbol{c}_1^{(k)\top}, \dots, \boldsymbol{c}_{j-1}^{(k)\top}, \boldsymbol{c}_j^\top, \boldsymbol{c}_{j+1}^{(k-1)\top}, \dots, \boldsymbol{c}_m^{(k-1)\top})^\top, \\ \boldsymbol{S}_j^k = (\boldsymbol{d}_1^{(k)}, \dots, \boldsymbol{d}_{j-1}^{(k)}, \boldsymbol{d}_j, \boldsymbol{d}_{j+1}^{(k-1)}, \dots, \boldsymbol{d}_m^{(k-1)}), \\ \boldsymbol{V}_j^k = (\boldsymbol{d}_1^{(k)}, \dots, \boldsymbol{d}_{j-1}^{(k)}, \boldsymbol{d}_j^{(k)}, \boldsymbol{d}_{j+1}^{(k-1)}, \dots, \boldsymbol{d}_m^{(k-1)}). \end{cases} \tag{5}$$

The parameter sequence λ_j^k is chosen so as to $\lambda_j^k > L(\boldsymbol{d}_j^{(k)})$ where $L(\boldsymbol{d}_j^{(k)})$ is defined by

$$\|\nabla_{\boldsymbol{d}_j} Q(\boldsymbol{C}^{(k)}, \bar{\boldsymbol{D}}_j^1) - \nabla_{\boldsymbol{d}_j} Q(\boldsymbol{C}^{(k)}, \bar{\boldsymbol{D}}_j^2)\|_F \le L(\boldsymbol{d}_j^k)\|\boldsymbol{d}_j^1 - \boldsymbol{d}_j^2\|_F, \tag{6}$$

for all $\boldsymbol{d}_j^1, \boldsymbol{d}_j^2 \in \mathbb{R}^n$ where $\bar{\boldsymbol{D}}_j^i = (\boldsymbol{d}_1^{(k)}, \dots, \boldsymbol{d}_{j-1}^{(k)}, \boldsymbol{d}_j^i, \boldsymbol{d}_{j+1}^{(k-1)}, \dots, \boldsymbol{d}_m^{(k-1)}), i = 1, 2$. Let $\boldsymbol{Z}^{(k)} := (\boldsymbol{C}^{(k)}, \boldsymbol{D}^{(k)})$ be the sequence generated by (4), in the next, we will first define the critical point for a non-convex function, then show that $\{\boldsymbol{Z}^{(k)}\}_{k\in\mathbb{N}}$ is a Cauchy sequence and converges to the critical point of (2).

Theorem 1. *The sequence $\{(\boldsymbol{C}^{(k)}, \boldsymbol{D}^{(k)})\}_{k\in\mathbb{N}}$ generated by the algorithm 1 is a Cauchy sequence and converges to the critical point of* (2).

2 Preliminaries

Definition 1. *([4]) Given the non-convex function $f : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ is a proper and lower semi-continuous function and* $\mathrm{dom} f = \{x \in \mathbb{R}^n : f(x) < +\infty\}$.

– *For $x \in \mathrm{dom} f$, its* Frechét subdifferential *of f is defined as*

$$\hat{\partial} f(x) = \{u : \liminf_{y\to x, y\neq x} (f(y) - f(x) - \langle u, y - x\rangle)/(\|y - x\|) \ge 0\},$$

and $\hat{\partial} f(x) = \emptyset$ if $x \notin \mathrm{dom} f$.

– *The* Limiting Subdifferential *of f at x is defined as*

$$\partial f(x) = \{u \in \mathbb{R}^n : \exists x^k \to x, f(x^k) \to f(x) \ \textit{and} \ \ u^k \in \hat{\partial} f(x^k) \to u\}.$$

– *The point x is a* critical point *of f if* $0 \in \partial f(x)$.

Remark 1.
– If x is a local minimizer of f then $0 \in \partial f(x)$.
– If f is the convex function, then $\partial f(x) = \hat{\partial} f(x) = \{u | f(y) \ge f(x) + \langle u, y - x\rangle, \forall y \in \mathrm{dom} f\}$. In that case, $0 \in \partial f(x)$ is the first order optimal condition.
– The Limiting subgradient of H defined in (2) is given by

$$\partial H(\boldsymbol{Z}) = (\partial F(\boldsymbol{C}) + \nabla_{\boldsymbol{C}} H(\boldsymbol{Z}), \partial G(\boldsymbol{D}) + \nabla_{\boldsymbol{D}} H(\boldsymbol{Z})). \tag{7}$$

The proof of Theorem 1 is built upon Theorem 2.9 in [3].

Theorem 2. *([3]) Assume $H(z)$ is a proper and lower semi-continuous function with* $\inf H > -\infty$*, the sequence $\{z^{(k)}\}_{k\in\mathbb{N}}$ is a Cauchy sequence and converges to the critical point of $H(z)$, if the following four conditions hold:*

(P1) **Sufficient decrease condition.** *There exists some positive constant* ρ_1*, such that*

$$H(z^{(k)}) - H(z^{(k+1)}) \geq \rho_1 \|z^{(k+1)} - z^{(k)}\|_F^2, \ \forall k = 1, 2, \ldots.$$

(P2) **Relative error condition.** *There exists some positive constant* $\rho_2 > 0$*, such that*

$$\|w^{(k)}\|_F \leq \rho_2 \|z^{(k)} - z^{(k-1)}\|_F, \ w^{(k)} \in \partial H(z^{(k)}), \ \forall k = 1, 2, \ldots.$$

(P3) **Continuity condition.** *There exists a subsequence* $\{z^{(k_j)}\}_{j \in \mathbb{N}}$ *and* $\bar{z}$ *such that*

$$z^{(k_j)} \to \bar{z}, \ H(z^{(k_j)}) \to H(\bar{z}), \quad \textit{as } j \to +\infty.$$

(P4) $H(z)$ **is a KL function**. $H(z)$ *satisfies the* Kurdyka-Lojasiewicz *property in its effective domain.*

3 Proof of Theorem 1

The proof of Theorem 1 is built upon Theorem 2.9 in [3], i.e. Theorem 2. Let $\boldsymbol{Z}^{(k)} := (\boldsymbol{C}^{(k)}, \boldsymbol{D}^{(k)})$ denote the sequence generated by the iteration (4). First of all, it can be seen that the objective function $H(\boldsymbol{Z}) = F(\boldsymbol{C}) + Q(\boldsymbol{Z}) + G(\boldsymbol{D})$ is the proper, lower semi-continuous function and bounded below by 0 by the definition (3). Secondly, the sequence $\{\boldsymbol{Z}^{(k)}\}_{k \in \mathbb{N}}$ generated by iteration (4) is bounded since $\boldsymbol{D}^{(k)} \in \mathcal{D}$ and $\boldsymbol{C}^{(k)} \in \mathcal{C}$ for all $k = 1, 2, \ldots$. In the next, we show one by one that the sequence $\{\boldsymbol{Z}^{(k)}\}_{k \in \mathbb{N}}$ satisfies the condition (P1)-(P4). Then, Theorem 1 is proved by directly calling Theorem 2.

3.1 Proof of condition (P1)

Lemma 1. *The sequence* $\{\boldsymbol{Z}^{(k)}\}_{k \in \mathbb{N}}$ *generated by* (4) *satisfies the following property, for* $1 \leq j \leq m$*,*

$$\begin{cases} H(\boldsymbol{T}_j^k, \boldsymbol{D}^{(k-1)}) \leq H(\boldsymbol{T}_{j-1}^k, \boldsymbol{D}^{(k-1)}) - \frac{\mu_j^k}{2} \|\boldsymbol{c}_j^{(k)} - \boldsymbol{c}_j^{(k-1)}\|_F^2, \\ H(\boldsymbol{C}^{(k)}, \boldsymbol{V}_j^k) \leq H(\boldsymbol{C}^{(k)}, \boldsymbol{V}_{j-1}^k) - \frac{\lambda_j^k - L(\boldsymbol{d}_j^{(k)})}{2} \|\boldsymbol{d}_j^{(k)} - \boldsymbol{d}_j^{(k-1)}\|_F^2, \end{cases} \tag{8}$$

where

$$\begin{cases} \boldsymbol{T}_j^k = (\boldsymbol{c}_1^{(k)\top}, \ldots, \boldsymbol{c}_j^{(k)\top}, \boldsymbol{c}_{j+1}^{(k-1)\top}, \ldots, \boldsymbol{c}_m^{(k-1)\top})^\top, \ \boldsymbol{T}_0^{(k)} = \boldsymbol{C}^{(k-1)}, \\ \boldsymbol{V}_j^k = (\boldsymbol{d}_1^{(k)}, \ldots, \boldsymbol{d}_j^{(k)}, \boldsymbol{d}_{j+1}^{(k-1)}, \ldots, \boldsymbol{d}_m^{(k-1)}), \ \boldsymbol{V}_0^{(k)} = \boldsymbol{D}^{(k-1)}. \end{cases} \tag{9}$$

Proof. From the fist step in (4), we know

$$\boldsymbol{c}_j^{(k)} \in \arg\min_{\boldsymbol{c}_j \in \mathcal{C}} F(\bar{\boldsymbol{c}}_j^k) + Q(\boldsymbol{U}_j^k, \boldsymbol{D}^{(k-1)}) + \frac{\mu_j^k}{2} \|\boldsymbol{c}_j - \boldsymbol{c}_j^{(k-1)}\|_F^2, \tag{10}$$

By the optimality of $\boldsymbol{c}_j^{(k)}$ in (10), we have

$$F(\boldsymbol{c}_j^k) + Q(\boldsymbol{T}_j^{(k)}, \boldsymbol{D}^{(k-1)}) + \frac{\mu_j^k}{2} \|\boldsymbol{c}_j^{(k)} - \boldsymbol{c}_j^{(k-1)}\| \leq F(\boldsymbol{c}_{j-1}^k) + Q(\boldsymbol{T}_{j-1}^{(k)}, \boldsymbol{D}^{(k-1)}).$$

Sum $G(\boldsymbol{D}^{(k-1)})$ on both sides of the above inequality, we have the first inequality in (8). From the second step in (4), we know

$$\boldsymbol{d}_j^{(k)} \in \arg\min_{\boldsymbol{d}_j \in \mathcal{D}} G(\boldsymbol{S}_j^k) + \langle \nabla_{\boldsymbol{d}_j} Q(\boldsymbol{C}^{(k)}, \boldsymbol{V}_{j-1}^{(k)}), \boldsymbol{d}_j - \boldsymbol{d}_j^{(k-1)} \rangle + \frac{\lambda_j^k}{2} \|\boldsymbol{d}_j - \boldsymbol{d}_j^{(k-1)}\|_F^2.$$

The above inequality implies

$$G(\boldsymbol{d}_j^k) + \langle \nabla_{\boldsymbol{d}_j} Q(\boldsymbol{C}^{(k)}, \boldsymbol{V}_{j-1}^k), \boldsymbol{d}_j^{(k)} - \boldsymbol{d}_j^{(k-1)} \rangle + \frac{L(\boldsymbol{d}_j^k)}{2} \|\boldsymbol{d}_j^{(k)} - \boldsymbol{d}_j^{(k-1)}\|_F^2 \leq G(\boldsymbol{V}_{j-1}^k).$$

From (6), we have

$$\begin{aligned} Q(\boldsymbol{C}^{(k)}, \boldsymbol{V}_j^k) \leq & Q(\boldsymbol{C}^{(k)}, \boldsymbol{V}_{j-1}^k) + \langle \nabla_{\boldsymbol{d}_j} Q(\boldsymbol{C}^{(k)}, \boldsymbol{V}_{j-1}^k), \boldsymbol{d}_j^{(k)} - \boldsymbol{d}_j^{(k-1)} \rangle \\ & + \frac{L(\boldsymbol{d}_j^k)}{2} \|\boldsymbol{d}_j^{(k)} - \boldsymbol{d}_j^{(k-1)}\|_F^2. \end{aligned} \tag{11}$$

Together with (3.1), the second inequality in (8) is satisfied.

Sum up the above inequalities, we can obtain

$$\begin{aligned} & H(\boldsymbol{C}^{(k-1)}, \boldsymbol{D}^{(k-1)}) - H(\boldsymbol{C}^{(k)}, \boldsymbol{D}^{(k)}) \\ \geq & \sum_{j=1}^{m} \left(\frac{\mu_j^k}{2} \|\boldsymbol{c}_j^{(k)} - \boldsymbol{c}_j^{(k-1)}\|_F^2 + \frac{\lambda_j^k - L(\boldsymbol{d}_j^{(k)})}{2} \|\boldsymbol{d}_j^{(k)} - \boldsymbol{d}_j^{(k-1)}\|_F^2 \right). \end{aligned} \tag{12}$$

Using the fact that there exist $a, b > 0$ such that $a < \mu_j^k, \lambda_j^k < b$ and $\lambda_j^k > L(\boldsymbol{d}_j^{(k)})$, we can establish the sufficient decreasing property (P1) for $\{\boldsymbol{Z}^{(k)}\}_{k \in \mathbb{N}}$ from (12).

3.2 Proof of condition (P2)

Lemma 2. *Let* $\boldsymbol{\omega}_{\boldsymbol{C}}^{(k)} = (\boldsymbol{\omega}_{\boldsymbol{C}}^{1\top}, \ldots, \boldsymbol{\omega}_{\boldsymbol{C}}^{m\top})^\top$ *and* $\boldsymbol{\omega}_{\boldsymbol{D}}^{(k)} = (\boldsymbol{\omega}_{\boldsymbol{D}}^1, \ldots, \boldsymbol{\omega}_{\boldsymbol{D}}^m)$ *where*

$$\begin{cases} \boldsymbol{\omega}_{\boldsymbol{C}}^j = \nabla_{\boldsymbol{c}_j} Q(\boldsymbol{Z}^{(k)}) - \nabla_{\boldsymbol{c}_j} Q(\boldsymbol{T}_j^k, \boldsymbol{D}^{(k-1)}) - \mu_j^k (\boldsymbol{c}_j^{(k)} - \boldsymbol{c}_j^{(k-1)}), \\ \boldsymbol{\omega}_{\boldsymbol{D}}^j = \nabla_{\boldsymbol{d}_j} Q(\boldsymbol{Z}^{(k)}) - \nabla_{\boldsymbol{d}_j} Q(\boldsymbol{C}^{(k)}, \boldsymbol{V}_j^k) - \lambda_j^k (\boldsymbol{d}_j^{(k)} - \boldsymbol{d}_j^{(k-1)}), \end{cases} \tag{13}$$

and $\boldsymbol{T}_j^k, \boldsymbol{V}_j^k$ *is defined in* (9)*. Then,* $\boldsymbol{\omega}^k := (\boldsymbol{\omega}_{\boldsymbol{C}}^{(k)}, \boldsymbol{\omega}_{\boldsymbol{D}}^{(k)}) \in \partial H(\boldsymbol{Z}^{(k)})$ *and there exists a constant* $\rho > 0$*, such that*

$$\|\boldsymbol{\omega}^k\|_F \leq \rho \|\boldsymbol{Z}^{(k)} - \boldsymbol{Z}^{(k-1)}\|_F.$$

Proof. The optimality condition of the first minimization problem in (4) is

$$\nabla_{\boldsymbol{c}_j} Q(\boldsymbol{T}_j^k, \boldsymbol{D}^{(k-1)}) + \mu_j^k (\boldsymbol{c}_j^{(k)} - \boldsymbol{C}^{(k-1)}) + u_j^k = 0, \tag{14}$$

where $u_j^k \in \partial_{\boldsymbol{c}_j} F(\boldsymbol{T}_j^k)$. Therefore, the following holds

$$\boldsymbol{u}_j^k = -(\nabla_{\boldsymbol{c}_j} Q(\boldsymbol{T}_j^k, \boldsymbol{D}^{(k-1)}) + \mu_j^k (\boldsymbol{c}_j^{(k)} - \boldsymbol{C}^{(k-1)})) \tag{15}$$

Since $F(\boldsymbol{C}) = \|\boldsymbol{C}\|_0 = \sum_{j=1}^{m} \|\boldsymbol{c}_j\|_0$, we have $\boldsymbol{u}_j^k \in \partial_{\boldsymbol{c}_j} F(\boldsymbol{C}^{(k)})$. From (7), it is easy to know $\boldsymbol{u}_j^k + \nabla_{\boldsymbol{c}_j} Q(\boldsymbol{Z}^{(k)}) \in \partial_{\boldsymbol{c}_j} H(\boldsymbol{Z}^{(k)})$. Therefore, we have

$$\nabla_{\boldsymbol{c}_j} Q(\boldsymbol{Z}^{(k)}) - \nabla_{\boldsymbol{c}_j} Q(\boldsymbol{T}_j^k, \boldsymbol{D}^{(k-1)}) - \mu_j^k(\boldsymbol{c}_j^{(k)} - \boldsymbol{c}_j^{(k-1)}) \in \partial_{\boldsymbol{c}_j} H(\boldsymbol{Z}^{(k)}).$$

Similarly, by optimality condtion of the second minimization problem in (4), we have

$$\nabla_{\boldsymbol{d}_j} Q(\boldsymbol{C}^{(k)}, \boldsymbol{V}_j^k) + \lambda_j^k(\boldsymbol{d}_j^{(k)} - \boldsymbol{d}_j^{(k-1)}) + \boldsymbol{v}_j^k = 0, \tag{16}$$

where $\boldsymbol{v}_j^k \in \partial_{\boldsymbol{d}_j} G(\boldsymbol{V}_j^k)$. Since $\mathcal{D} = \bigcap_{j=1}^{m} \{\boldsymbol{D} : \|\boldsymbol{d}_j\|_2 = 1\}$, we have $\boldsymbol{v}_j^k \in \partial_{\boldsymbol{d}_j} G(\boldsymbol{D}^{(k)})$. From (7), we know $\boldsymbol{v}_j^k + \nabla_{\boldsymbol{d}_j} Q(\boldsymbol{Z}^{(k)}) \in \partial_{\boldsymbol{d}_j} H(\boldsymbol{Z}^{(k)})$. Consequently, we have

$$\nabla_{\boldsymbol{d}_j} Q(\boldsymbol{Z}^{(k)}) - \nabla_{\boldsymbol{d}_j} Q(\boldsymbol{C}^{(k)}, \boldsymbol{V}_j^k) - \lambda_j^k(\boldsymbol{d}_j^{(k)} - \boldsymbol{d}_j^{(k-1)}) \in \partial_{\boldsymbol{d}_j} H(\boldsymbol{Z}^{(k)}).$$

Since $\boldsymbol{C}^{(k)} \in \mathcal{C}$ and $\boldsymbol{D}^{(k)} \in \mathcal{D}$ for all $k \in \mathbb{N}$, the sequence $\{\boldsymbol{Z}^{(k)}\}_{k\in\mathbb{N}}$ is a bounded sequence. Let $\{\boldsymbol{Z}^{(k)}\} \subseteq \mathcal{Z}$, the following inequality holds: there exists $L > 0$, such that

$$\|\nabla_{\boldsymbol{Z}} Q(\boldsymbol{Z}_1) - \nabla_{\boldsymbol{Z}} Q(\boldsymbol{Z}_2)\|_F \leq L\|\boldsymbol{Z}_1 - \boldsymbol{Z}_2\|_F, \ \forall \boldsymbol{Z}_1, \boldsymbol{Z}_2 \in \mathcal{Z}, \tag{17}$$

since Q has lipschitz continuous gradient. Therefore, we have

$$\begin{aligned}
\|\boldsymbol{\omega}_{\boldsymbol{C}}^j\| \leq& \mu_j^k \|\boldsymbol{c}_j^{(k)} - \boldsymbol{c}_j^{(k-1)}\|_F + \|\nabla_{\boldsymbol{c}_j} Q(\boldsymbol{Z}^{(k)}) - \nabla_{\boldsymbol{c}_j} Q(\boldsymbol{T}_j^k, \boldsymbol{d}^{(k-1)})\|_F \\
\leq& b\|\boldsymbol{c}_j^{(k)} - \boldsymbol{c}_j^{(k-1)}\|_F + L(\sum_{i=j}^{m} \|\boldsymbol{c}_i^{(k)} - \boldsymbol{c}_i^{(k-1)}\| + \|\boldsymbol{d}^{(k)} - \boldsymbol{d}^{(k-1)}\|_F) \\
=& (b + (m-j)L)\|\boldsymbol{c}_j^{(k)} - \boldsymbol{c}_j^{(k-1)}\|_F + L\|\boldsymbol{d}^{(k)} - \boldsymbol{d}^{(k-1)}\|_F \\
\leq& ((m+1)L + b)\|\boldsymbol{Z}^{(k)} - \boldsymbol{Z}^{(k-1)}\|_F
\end{aligned} \tag{18}$$

Similarly, we also have

$$\begin{aligned}
\|\boldsymbol{\omega}_{\boldsymbol{D}}^j\| \leq& \lambda_j^k \|\boldsymbol{d}_j^{(k)} - \boldsymbol{d}_j^{(k-1)}\|_F + \|\nabla_{\boldsymbol{d}_j} Q(\boldsymbol{Z}^{(k)}) - \nabla_{\boldsymbol{d}_j} Q(\boldsymbol{C}^{(k)}, \boldsymbol{V}_j^k)\|_F \\
\leq& b\|\boldsymbol{d}_j^{(k)} - \boldsymbol{d}_j^{(k-1)}\|_F + L(\sum_{i=j}^{m} \|\boldsymbol{d}_i^{(k)} - \boldsymbol{d}_i^{(k-1)}\|_F) \\
\leq& (mL + b)\|\boldsymbol{Z}^{(k)} - \boldsymbol{Z}^{(k-1)}\|_F
\end{aligned} \tag{19}$$

Therefore, by $\boldsymbol{\omega}^k = (\boldsymbol{\omega}_{\boldsymbol{C}}^{(k)}, \boldsymbol{\omega}_{\boldsymbol{D}}^{(k)})$, we have

$$\|\boldsymbol{\omega}^k\|_F = \sum_{j=1}^{m} \|\boldsymbol{\omega}_{\boldsymbol{C}}^j\|_F + \|\boldsymbol{\omega}_{\boldsymbol{D}}^j\|_F \leq \rho\|\boldsymbol{Z}^{(k)} - \boldsymbol{Z}^{(k-1)}\|_F, \tag{20}$$

where $\rho = m((2m+1)L + 2b)$.

3.3 Proof of condition (P3)

Lemma 3. *The sequence* $\{\boldsymbol{Z}^{(k)}\}_{k\in\mathbb{N}}$ *satisfies the Continuity condition: there exists* $\bar{\boldsymbol{Z}}$ *such that*

$$\boldsymbol{Z}^{(k_j)} \to \bar{\boldsymbol{Z}},\ H(\boldsymbol{Z}^{(k_j)}) \to H(\bar{\boldsymbol{Z}}), \quad as\ j \to +\infty.$$

.

Proof. Since $\boldsymbol{C}^{(k)} \in \mathcal{C}$ and $\boldsymbol{D}^{(k)} \in \mathcal{D}$ for all $k \in \mathbb{N}$, the sequence $\{\boldsymbol{Z}^{(k)}\}_{k\in\mathbb{N}}$ is a bounded sequence and there exists a sub-sequence $\{\boldsymbol{Z}^{(k_j)}\}_{j\in\mathbb{N}}$ such that $\boldsymbol{Z}^{(k_j)} \to \bar{\boldsymbol{Z}} = (\boldsymbol{U}, \bar{\boldsymbol{D}})$. Since $\boldsymbol{Z}^{(k_j-1)}$ is also a bounded sequence, without loss of generality, assume $\boldsymbol{Z}^{(k_j-1)} \to \bar{\boldsymbol{Z}}_1$. In the next, we first show that $\bar{\boldsymbol{Z}} = \bar{\boldsymbol{Z}}_1$. By the lemma 1, we have

$$H(\boldsymbol{Z}^{(k-1)}) - H(\boldsymbol{Z}^{(k)}) \geq \rho_1 \|\boldsymbol{Z}^{(k)} - \boldsymbol{Z}^{(k-1)}\|_F^2,$$

where $\rho_1 > b$. So, $H(\boldsymbol{Z}^{(k)})$ is a decreasing sequence and from the fact that $H(\boldsymbol{Z}^{(0)}) < +\infty, H(\boldsymbol{Z}) \geq 0$, we have $\lim\limits_{k\to+\infty} H(\boldsymbol{Z}^{(k)}) = \bar{H}$,where $\bar{H}$ is some constant. Summing from $k = 0$ to N, we have

$$H(\boldsymbol{Z}^{(0)}) - H(\boldsymbol{Z}^{(N)}) \leq \rho_1 \sum_{k=1}^{N} \|\boldsymbol{Z}^{(k)} - \boldsymbol{Z}^{(k-1)}\|_F^2,$$

let $N \to +\infty$ in the above, we have

$$\sum_{k=1}^{+\infty} \|\boldsymbol{Z}^{(k)} - \boldsymbol{Z}^{(k-1)}\|_F^2 \leq \frac{H(\boldsymbol{Z}^{(0)}) - \bar{H}}{\rho_1} < +\infty,$$

which implies $\lim\limits_{k\to+\infty} \|\boldsymbol{Z}^{(k)} - \boldsymbol{Z}^{(k-1)}\|_F = 0$. So, for any $\epsilon > 0$, there exists $J \in \mathbb{N}$, such that for all $j > N$, $\|\boldsymbol{Z}^{(k_j)} - \boldsymbol{Z}^{(k_j-1)}\|_F < \epsilon/2$ and $\|\boldsymbol{Z}^{(k_j)} - \bar{\boldsymbol{Z}}\|_F < \epsilon/2$. It implies

$$\|\boldsymbol{Z}^{(k_j-1)} - \bar{\boldsymbol{Z}}\|_F \leq \|\boldsymbol{Z}^{(k_j)} - \boldsymbol{Z}^{(k_j-1)}\|_F + \|\boldsymbol{Z}^{(k_j)} - \bar{\boldsymbol{Z}}\|_F < \epsilon.$$

Consequently, $\boldsymbol{Z}^{(k_j-1)} \to \bar{\boldsymbol{Z}}$ as $j \to +\infty$.

Let $F(\boldsymbol{C}) = \sum_{j=1}^{m} f_j(\boldsymbol{c}_j)$, where $f_j(\boldsymbol{c}_j) = \|\boldsymbol{c}_j\|_0$. From the iterative step (4), we have for all k,

$$\boldsymbol{c}_j^{(k)} \in \mathop{\arg\min}_{\boldsymbol{c}_j\in\mathcal{C}} f_j(\boldsymbol{c}_j) + Q(\boldsymbol{U}_j^k, \boldsymbol{D}^{(k-1)}) + \frac{\mu_j^k}{2}\|\boldsymbol{c}_j - \boldsymbol{c}_j^{(k-1)}\|_F^2,$$

Let $\boldsymbol{c}_j = \boldsymbol{u}_j$ in the above inequality, we have

$$\begin{aligned} &f_j(\boldsymbol{c}_j^{(k)}) + Q(\boldsymbol{T}_j^k, \boldsymbol{D}^{(k-1)}) + \frac{\mu_j^k}{2}\|\boldsymbol{c}_j^{(k)} - \boldsymbol{c}_j^{(k-1)}\|_F^2 \\ \leq &f_j(\boldsymbol{u}_j) + Q(\boldsymbol{U}_j^k, \boldsymbol{D}^{(k-1)}) + \frac{\mu_j^k}{2}\|\boldsymbol{u}_j - \boldsymbol{c}_j^{(k-1)}\|_F^2, \end{aligned} \tag{21}$$

where $\boldsymbol{U}_j^k = (\boldsymbol{c}_1^{(k)\top}, \ldots, \boldsymbol{c}_{j-1}^{(k)\top} \boldsymbol{u}_j^\top, \boldsymbol{c}_{j+1}^{(k-1)\top}, \ldots, \boldsymbol{c}_m^{(k-1)\top})^\top$. Choose $k = k_j$ and let $j \to +\infty$ in (21), using the fact that $\boldsymbol{Z}^{(k_j-1)} \to \bar{\boldsymbol{Z}}$, we have

$$\limsup_{j \to +\infty} f_j(\boldsymbol{c}_j^{(k_j)}) \leq f_j(\boldsymbol{u}_j).$$

Since f_j is a lower semicontinuous function, we have $\lim_{j \to +\infty} f_j(\boldsymbol{c}_j^{(k_j)}) = f_j(\boldsymbol{u}_j)$. By the same argument, we have for all $j = 1, \ldots, m$, $\lim_{j \to +\infty} f_j(\boldsymbol{c}_j^{(k_j)}) = f_j(\boldsymbol{u}_j)$. Since Q is a smooth function and $G(\boldsymbol{D}^{(k)}) = 0,\ \forall k \in \mathbb{N}$, we have

$$\lim_{j \to +\infty} Q(\boldsymbol{Z}^{(k_j)}) = Q(\bar{\boldsymbol{Z}}), \quad \lim_{j \to +\infty} G(\boldsymbol{D}^{(k_j)}) = G(\bar{\boldsymbol{D}}).$$

This implies

$$\lim_{j \to +\infty} H(\boldsymbol{Z}^{(k_j)}) = \lim_{j \to +\infty} F(\boldsymbol{C}^{(k_j)}) + Q(\boldsymbol{Z}^{(k_j)}) + G(\boldsymbol{D}^{(k_j)}) = H(\bar{\boldsymbol{Z}}).$$

3.4 Proof of condition (P4)

For the property (P4), see [4] for the definition. An important class of functions that satisfies the Kurdyka-Lojasiewicz property is the so-called semi-algebraic functions [4].

Definition 2. *(Semi-algebraic sets and functions [4, 2]) A subset S of $\mathbb{R}^n$ is called the semi-algebraic set if there exists a finite number of real polynomial functions g_{ij}, h_{ij} such that*

$$S = \bigcup_j \bigcap_i \{x \in \mathbb{R}^n : g_{ij}(x) = 0, h_{ij}(x) < 0\}.$$

A function f is called the semi-algebraic function if its graph $\{(x,t) \in \mathbb{R}^n \times \mathbb{R}, t = f(x)\}$ is a semi-algebraic set.

Theorem 3. ([4]) *Let f is a proper and lower semicontinuous function. If f is semi-algebraic then it satisfies the K-L property at any point of* dom$f$.

Lemma 4. *All the function $F(\boldsymbol{C})$,$Q(\boldsymbol{Z})$ and $G(\boldsymbol{D})$ defined in* (3) *are semi-algebraic functions. Moreover, $H(\boldsymbol{Z}) = F(\boldsymbol{C}) + Q(\boldsymbol{Z}) + G(\boldsymbol{D})$ is the semi-algebraic function.*

Proof. For $Q(\boldsymbol{C}, \boldsymbol{D}) = \frac{1}{2}\|\boldsymbol{Y} - \boldsymbol{D}\boldsymbol{C}\|_F^2 + \frac{\alpha}{2}\|\boldsymbol{D}^\top\boldsymbol{D} - \boldsymbol{I}\|_F^2$ is a real polynomial function, $Q(\boldsymbol{C}, \boldsymbol{D})$ is a semi-algebraic function [4].

It is easy to notice that the set $\mathcal{D} = \{\boldsymbol{Y} \in \mathbb{R}^{n \times m} : \|\boldsymbol{d}_k\|_2 = 1, 1 \leq k \leq m\} = \bigcap_{k=1}^m \{\boldsymbol{Y} : \sum_{j=1}^n y_{kj}^2 = 1\}$ is a semi-algebraic set. And the set $\mathcal{C} = \{\boldsymbol{C} \in \mathbb{R}^{m \times p} | \|\boldsymbol{c}_k\|_\infty \leq M\} = \bigcup_{j=1}^M \bigcup_{k=1}^p \{\boldsymbol{C} : \|\boldsymbol{c}_k\|_\infty = j\}$ is a semi-algebraic set. Therefore, the indicator functions $\delta_{\mathcal{D}}(\boldsymbol{C})$ and $\delta_{\mathcal{D}}(\boldsymbol{D})$ are semi-algebraic functions from the fact that the indicator function for semi-algebraic sets are semi-algebraic functions [1].

For the function $F(\boldsymbol{C}) = \|\boldsymbol{C}\|_0$. The graph of F is $S = \bigcup_{k=0}^{mp} L_k \triangleq \{(\boldsymbol{C}, k) : \|\boldsymbol{C}\|_0 = k\}$. For each $k = 0, \cdots, mp$, let $\mathcal{S}_k = \{J : J \subseteq \{1, \cdots, mp\}, |J| = k\}$, then

$L_k = \bigcup_{J \in \mathcal{S}_k} \{(\boldsymbol{C}, k) : \boldsymbol{C}_{J^c} = 0, \boldsymbol{C}_J \neq 0\}$. It is easy to know the set $\{(\boldsymbol{C}, k) : \boldsymbol{C}_{J^c} = 0, \boldsymbol{C}_J \neq 0\}$ is a semi-algebraic set in $\mathbb{R}^{m \times p} \times \mathbb{R}$. Thus, $F(\boldsymbol{C}) = \|\boldsymbol{C}\|_0$ is a semi-algebraic function since the finite union of the semi-algebraic set is still semi-algebraic.

Consequently, $H(\boldsymbol{Z})$ is a semi-algebraic function since the finite summation of semi-algebraic functions are still semi-algebraic [4].

References

1. Attouch, H., Bolte, J.: On the convergence of the proximal algorithm for nonsmooth functions involving analytic features. Mathematical Programming (2009)
2. Attouch, H., Bolte, J., Redont, P., Soubeyran, A.: Proximal alternating minimization and projection methods for nonconvex problems: An approach based on the kurdyka-lojasiewicz inequality. Math. Oper. Res. 35(2), 438–457 (2010)
3. Attouch, H., Bolte, J., Svaiter, B.F.: Convergence of descent methods for semi-algebraic and tame problems: proximal algorithms, forward–backward splitting, and regularized gauss–seidel methods. Math. Program. Ser. A. 137(1-2), 91–129 (2013)
4. Bolte, J., Sabach, S., Teboulle, M.: Proximal alternating linearized minimization for nonconvex and nonsmooth problems. Math. Program. Ser. A. pp. 1–36 (2013)
5. Rockafellar, R.T., Wets, R.J.B.: Variational analysis: grundlehren der mathematischen wissenschaften, vol. 317. Springer (1998)