Supplementary Material:
Partial Sum Minimization of Singular Values in Robust PCA: Algorithm and Applications

Tae-Hyun Oh Student Member, IEEE, Yu-Wing Tai Senior Member, IEEE,
Jean-Charles Bazin Member, IEEE, Hyeongwoo Kim Student Member, IEEE,
and In So Kweon Member, IEEE

SUPPLEMENTARY MATERIAL

In this supplementary material, we prove Proposition 1, and we provide the pseudo code of the algorithm of image recovery application. We also present an additional experimental result not included in the main paper due to space limitation. All the parameters are the same as in the main paper or the referred papers, except if stated otherwise.

1 PROOFS OF PROPOSITION 1.

Lemma 4 (Lipschitz continuous of PSSV). The function of the partial sum of singular values $h(X) = \|X\|_p = \sum_{i=p+1}^{\min(m,n)} \sigma_i(X)$ (where $p \in \mathbb{N}$ denotes the target rank) for $X \in \mathbb{R}^{m \times n}$ is Lipschitz continuous. Namely, there exists a constant scalar $K$ satisfying

$$\|h(X_1) - h(X_2)\| \leq K \cdot \|X_1 - X_2\|_F \text{ for all } X_1, X_2 \in \mathbb{R}^{m \times n}.$$ 

Proof. Let the nuclear norm be $f(X) = \|X\|_* = \sum_{i=1}^{\min(m,n)} \sigma_i(X)$, and the Ky-Fan $p$-norm be $g(X) = \|X\|_{Ky(p)} = \sum_{i=1}^{p} \sigma_i(X)$. By definition, $h(X) = \|X\|_p = f(X) - g(X)$, and we know that the nuclear norm $f(\cdot)$ and the Ky-Fan $p$-norm $g(\cdot)$ are Lipschitz continuous (derivation of the Lipschitz continuity for Ky-Fan matrix norm is straightforward). Therefore we have

$$|f(X_1) - f(X_2)| \leq K_f \cdot \|X_1 - X_2\|_F,$$

$$|g(X_1) - g(X_2)| \leq K_g \cdot \|X_1 - X_2\|_F,$$

where $K_f$ and $K_g$ are Lipschitz constants for $f$ and $g$ respectively.

We see that

$$|h(X_1) - h(X_2)| = |f(X_1) - g(X_1) - (f(X_2) - g(X_2))|$$

$$= |f(X_1) - f(X_2) + (g(X_2) - g(X_1))|$$

$$\leq |f(X_1) - f(X_2)| + |g(X_1) - g(X_2)| \quad \text{(by triangle inequality)}$$

$$\leq (K_f + K_g) \cdot \|X_1 - X_2\|_F.$$

Since the constant $K = K_f + K_g$ satisfies the inequality, $h(X) = \|X\|_p$ is Lipschitz continuous. \hfill \qed

Since PSSV $\|\cdot\|_p$ is a non-convex function, the typical subdifferential for convex functions (a.k.a. Fenchel-Moreau subdifferential for convex functions. Refer to Daniilidis et al. [3]) would be the empty set. Therefore, we introduce a generalized subdifferential (a.k.a. Clarke subdifferential, see Definition 3.2 in [1]) for non-convex locally Lipschitz continuous functions. This is useful for deriving stationary points in the convergence proof.

Definition 2 (Generalized subgradients). Let $f : \mathbb{R}^n \to \mathbb{R}$ be a locally Lipschitz continuous function at a point $x \in \mathbb{R}^n$. Then the subdifferential of $f$ at $x$ is the set $\partial_C f(x)$ of vectors $z \in \mathbb{R}^n$ such that

$$\partial_C f(x) = \{ z : f^*(x;d) \geq \langle z, d \rangle \text{ for all } d \in \mathbb{R}^n \},$$

where each vector $z \in \partial_C f(x)$ is called a subgradient of $f$ at $x$, and the directional subgradient of $f$ at $x$ in the direction vector $d \in \mathbb{R}^n$ is defined as $f^*(x;d) = \limsup_{y \to x} \frac{f(y+td) - f(y)}{t}$.
Remark D.2.1. Definition 2 can be generalized to matrix cases analogously.

Remark D.2.2. Regardless of non-convexity or non-smoothness, the generalized subdifferential (here, Clarke subdifferential) always exists for locally Lipschitz continuous functions. Also, $\partial_C \| \cdot \|_p$ is well defined in $\mathbb{R}^{m \times n}$, because $\| \cdot \|_p$ is a Lipschitz continuous function as shown in Lemma 4.

The following Lemma is also important for the convergence proof of Proposition 1.

**Lemma 5** (Convexity and compactness of subdifferential [1]) Let $f : \mathbb{R}^{m \times n} \to \mathbb{R}$ be a locally Lipschitz continuous function at $X$. Then the subdifferential $\partial_C f(X)$ is a non-empty, convex and compact set.

Remark L.5.1. Basic properties of the generalized subdifferential are identical to those in convex case, and most of subdifferential calculus rules hold. Note that Karush-Kuhn-Tucker (KKT) optimality conditions are also properly defined with the generalized subdifferential [1].

Since our problem in Eq. (3, Main) does not have any inequality constraint, but an equality constraint, the KKT conditions are reduced to stationary and primal feasibility conditions. Armed with the above lemmas and definitions, we can now propose and prove the convergence of Alg. 1.

**Proposition 1** (Convergence). Let $S_k = (A_k, E_k, Y_k, \hat{Y}_k)$, where $Y_{k+1} = Y_k + \mu_k(O - A_{k+1} - E_k)$ and $\{S_k\}_{k=1}^{\infty}$ is a set of intermediate solutions of Alg. 1. Suppose that $\{Y_k\}_{k=1}^{\infty}$ and $\{\hat{Y}_k\}_{k=1}^{\infty}$ are bounded, $\lim_{k \to \infty} (Y_{k+1} - Y_k) = 0$, and $\mu_k$ is non-decreasing, then any accumulation point of $\{S_k\}_{k=1}^{\infty}$ satisfies the following KKT conditions:

(C1) $Y^* \in \partial_C \| A^* \|_p$,

(C2) $Y^* \in \partial \| \lambda E^* \|_1$,

(C3) $O - A^* - E^* = 0$,

(C4) $\partial_C \| A^* \|_p \cap \partial \| \lambda E^* \|_1 \neq \emptyset$,

where $Y^*$, $A^*$ and $E^*$ represent each cluster points. In particular, whenever $\{S_k\}_{k=1}^{\infty}$ converges, it converges to a KKT point of Eq. (2, Main).

**Proof.** For $Y$, we have $\mu_k^{-1}(Y_{k+1} - Y_k) = O - A_{k+1} - E_{k+1}$. Since $\lim_{k \to \infty} (Y_{k+1} - Y_k) = 0$ and $\mu_k$ is non-decreasing, we have $O - A_{k+1} - E_{k+1} = \mu_k^{-1}(Y_{k+1} - Y_k) \to 0$, which satisfies (C3).

Since $A_{k+1}$ obtained by the soft-thresholding operator [2] minimizes $L_{\mu_k}(A_{k+1}, E, Y_k)$ (refer to Eq. (10, Main)) by definition, we have

$$0 \in \partial \| \lambda E_{k+1} \|_1 - Y_k - \mu_k(O - A_{k+1} - E_{k+1})$$

$$= \partial \| \lambda E_{k+1} \|_1 - Y_{k+1}$$

(by definition of $Y_{k+1}$)

$$\Rightarrow Y_{k+1} \in \partial \| \lambda E_{k+1} \|_1,$$

which satisfies (C2).

Since $A_{k+1}$ optimally obtained by the PSVT minimizes $L_{\mu_k}(A, E_k, Y_k)$ by Theorem 1 and Lemma 5, we have

$$0 \in \partial_C \| A_{k+1} \|_p - Y_k - \mu_k(O - A_{k+1} - E_k)$$

$$= \partial_C \| A_{k+1} \|_p - Y_k - \mu_k(O - A_{k+1} - E_{k+1}) - \mu_k(E_{k+1} - E_k)$$

$$= \partial_C \| A_{k+1} \|_p - Y_{k+1} - \mu_k(E_{k+1} - E_k)$$

(by definition of $Y_{k+1}$)

$$\Rightarrow Y_{k+1} + \mu_k(E_{k+1} - E_k) \in \partial_C \| A_{k+1} \|_p,$$

which satisfies (C2).

Since $\{Y_k\}_{k=1}^{\infty}$ and $\{\hat{Y}_k\}_{k=1}^{\infty}$ are bounded, there must exist a scalar $c > 0$ such that $\|Y_{k+1}\|_F \leq c$ and $\|\hat{Y}_{k+1}\|_F \leq c$. Then,

$$Y_{k+1} - \hat{Y}_{k+1} = \mu_k^{-1}(Y_{k+1} - Y_k) - (O - A_{k+1} - E_k)$$

$$= \mu_k((O - A_{k+1} - E_{k+1}) - (O - A_{k+1} - E_k)) = \mu_k(E_{k+1} - E_k)$$

$$\Rightarrow \|E_{k+1} - E_k\|_F = \mu_k^{-1}\|Y_{k+1} - \hat{Y}_{k+1}\|_F$$

$$\leq \mu_k^{-1}(\|Y_{k+1}\|_F + \|\hat{Y}_{k+1}\|_F)$$

(by triangle inequality)

$$\leq 2c \mu_k^{-1} \to 0$$

(since $\mu_k$ is non-decreasing).

Thus, $Y_{k+1} \to Y^* \in \partial_C \| A^* \|_p$ from Eq. (5) and $Y^* \in \partial_C \| A^* \|_p \cap \partial \| \lambda E^* \|_1 \neq \emptyset$ which satisfy (C1) and (C4). The sequence $\{S_k\}_{k=1}^{\infty}$ gradually satisfies the KKT conditions, which completes the proof.

**Remark P.1.1.** In Alg. 1 (and Proposition 1), the assumption for non-decreasing $\mu_k$ is always satisfied by the update rule $\mu_{k+1} = \rho \mu_k$ (with $\rho > 1$), and the boundness of the sequences $\{Y_k\}_{k=1}^{\infty}$ and $\{\hat{Y}_k\}_{k=1}^{\infty}$ is satisfied by the below Lemmas 6...
and 7 in conjunction with a similar manner of Lemma 1 in Lin et al. [4]. Therefore, we see that Alg. 1 converges as long as $Y_k$ converges.

**Lemma 6** (Boundness of $|f^o(\cdot; \cdot)|$ [1]) Let $f : \mathbb{R}^{m \times n} \to \mathbb{R}$ be a locally Lipschitz continuous function at $X$, and $d \in \mathbb{R}^{m \times n}$. Then, there exists a scalar $B$ such that

$$|f^o(X; d)| \leq B\|d\|_F. \quad (7)$$

**Proof.** Refer to the proof of Theorem 3.1 in [1].

**Lemma 7** (Boundness of Clarke subgradient) Let $f : \mathbb{R}^{m \times n} \to \mathbb{R}$ be a locally Lipschitz continuous function at $X$. Then, a subgradient $W \in \partial_C f(X)$ is bounded as

$$\|W\|_F \leq B, \quad (8)$$

where $B$ is the same constant as in Lemma 6.

**Proof.** By the definition of the Clarke subdifferential, $W$ satisfies $f^o(X; d) \geq \langle W, d \rangle$ for all $d \in \mathbb{R}^{m \times n}$. By setting $d = W$, we get $f^o(X; W) \geq \langle W, W \rangle = \|W\|_F^2$. Then, by Lemma 6, we have

$$\|W\|_F^2 \leq |f^o(X; W)| \leq B\|W\|_F \Rightarrow \|W\|_F \leq B \quad (9)$$

2 ALGORITHM FOR IMAGE RECOVERY

**Algorithm 2** ADMM for Image Recovery

| Input : $O \in \mathbb{R}^{m \times n}$, the index map $\Omega$, the constraint rank $N$. Initialize $A_0 = O$, $B_0 = Z_0 = 0$, $\mu_0 > 0$, $\rho > 1$ and $k = 0$. while not converged do $A_{k+1} = P_{\mathcal{N},\mu_k^{-1}}[B_k - \mu_k^{-1}Z_k]$. $B_{k+1} = \arg\min_{P_{\Omega}(B)=P_{\Omega}(O)} \|B - (A_{k+1} + \mu_k^{-1}Z_k)\|_F^2$. $Z_{k+1} = Z_k + \mu_k(A_{k+1} - B_{k+1})$. $\mu_{k+1} = \rho\mu_k$. $k = k + 1$. end while Output : $A_k$. |

REFERENCES


