Supplementary Material for Integrated Low Rank Based Discriminative Feature Learning for Recognition

Pan Zhou, Zhouchen Lin, *Senior Member, IEEE,* and Chao Zhang, *Member, IEEE*

I. Proof of Theorem 3.2

Before we prove Theorem 3.2, we introduce a lemma, which is described as follows.

Lemma 1.1 ([1], [2]): Suppose that $f : R^m \to R$ is a continuously differentiable function with Lipschitz continuous gradient whose Lipschitz constant is L. Then for any $x, y \in R^m$ and $\gamma \geq L$,

$$f(x) \leq f(y) + \langle x - y, \nabla f(y)\rangle + \frac{\gamma}{2}\|x - y\|_2^2. \tag{1}$$

Now, we prove Theorem 3.2.

proof Let $f(Q_i) = \| \tilde{H}_i - Q_i \|_2^2$ $(i = 1, \cdots, r)$. It is easy to check that $f(Q_i)$ satisfies Lemma 1.1, $\nabla f(Q_i) = -2(\tilde{H}_i - Q_i)$ and $L = 2$. We first prove the following inequality:

$$\langle Q_i^{k+1} - Q_i^k, \nabla f(Q_i^k)\rangle + \frac{\gamma}{2}\|Q_i^{k+1} - Q_i^k\|_2^2 + \frac{\alpha}{(g_i^k)^2}\|Q_i^{k+1}\|_2^2 \leq \frac{\alpha}{(g_i^k)^2}\|Q_i^k\|_2^2, \quad (\forall k, i), \tag{2}$$

where k denotes the number of iterations. Define $\beta_k = \frac{(g_i^{k-1})^2}{(g_i^{k-1})^2+\alpha}$, then $Q_i^k = \frac{(g_i^{k-1})^2}{(g_i^{k-1})^2+\alpha}\tilde{H}_i = \beta_k \tilde{H}_i$. We consider inequality (2).

$$\begin{aligned}
c =& \left\langle Q_i^{k+1} - Q_i^k, \nabla f(Q_i^k)\right\rangle + \frac{\gamma}{2}\|Q_i^{k+1} - Q_i^k\|_2^2 + \frac{\alpha}{(g_i^k)^2}\|Q_i^{k+1}\|_2^2 - \frac{\alpha}{(g_i^k)^2}\|Q_i^k\|_2^2 \\
=& \left\langle Q_i^{k+1} - Q_i^k, \nabla f(Q_i^k) + \frac{\gamma}{2}(Q_i^{k+1} - Q_i^k) + \frac{\alpha}{(g_i^k)^2}(Q_i^{k+1} + Q_i^k)\right\rangle \\
=& \left\langle Q_i^{k+1} - Q_i^k, \left(2 - \frac{\gamma}{2} + \frac{\alpha}{(g_i^k)^2}\right) Q_i^k + \left(\frac{\gamma}{2} + \frac{\alpha}{(g_i^k)^2}\right) Q_i^{k+1} - 2\tilde{H}_i\right\rangle \\
=&(\beta_{k+1} - \beta_k)\left[\left(2 - \frac{\gamma}{2} + \frac{\alpha}{(g_i^k)^2}\right)\beta_k + \left(\frac{\gamma}{2} + \frac{\alpha}{(g_i^k)^2}\right)\beta_{k+1} - 2\right]\|\tilde{H}_i\|_2^2 \\
=&(\beta_{k+1} - \beta_k)\left[\left(2 - \frac{\gamma}{2} + \frac{\alpha}{(g_i^k)^2}\right)(\beta_k - \beta_{k+1}) + \left(\frac{\gamma}{2} + \frac{\alpha}{(g_i^k)^2}\right)\beta_{k+1} + \left(2 - \frac{\gamma}{2} + \frac{\alpha}{(g_i^k)^2}\right)\beta_{k+1} - 2\right]\|\tilde{H}_i\|_2^2 \\
=&(\beta_{k+1} - \beta_k)\left[\left(2 - \frac{\gamma}{2} + \frac{\alpha}{(g_i^k)^2}\right)(\beta_k - \beta_{k+1}) + 2\frac{\alpha + (g_i^k)^2}{(g_i^k)^2}\beta_{k+1} - 2\right]\|\tilde{H}_i\|_2^2 \\
=&(\beta_{k+1} - \beta_k)\left[\left(2 - \frac{\gamma}{2} + \frac{\alpha}{(g_i^k)^2}\right)(\beta_k - \beta_{k+1}) + 2\frac{\alpha + (g_i^k)^2}{(g_i^k)^2}\frac{(g_i^k)^2}{\alpha + (g_i^k)^2} - 2\right]\|\tilde{H}_i\|_2^2 \\
=& -(\beta_{k+1} - \beta_k)^2\left(2 - \frac{\gamma}{2} + \frac{\alpha}{(g_i^k)^2}\right)\|\tilde{H}_i\|_2^2.
\end{aligned} \tag{3}$$

So if γ satisfies $L = 2 < \gamma \leq 4$, we have that $c \leq 0$, i.e., inequality (2) holds.

Note that g^{k+1} is the optimal solution to problem (4):

$$g^{k+1} = \operatorname*{argmin}_{\sum_{i=1}^r g_i = t, g_i \geq 0} \sum_{i=1}^r \frac{\alpha}{g_i^2} \| Q_i^{k+1} \|_2^2 . \tag{4}$$

So the following inequality holds.

$$\sum_{i=1}^r \frac{\alpha}{(g_i^k)^2}\|Q_i^{k+1}\|_2^2 \geq \sum_{i=1}^r \frac{\alpha}{(g_i^{k+1})^2}\|Q_i^{k+1}\|_2^2. \tag{5}$$

P. Zhou, Z. Lin, and C. Zhang are with Key Lab. of Machine Perception (MoE), School of EECS, Peking University, P. R. China. Z. Lin is also with Cooperative Medianet Innovation Center, Shanghai, China. Emails: pzhou@pku.edu.cn, zlin@pku.edu.cn, and chzhang@cis.pku.edu.cn. Z. Lin is the corresponding author.

Then, combining inequalities (2) and (5), we can further obtain the following inequality:

$$\begin{aligned} -\sum_{i=1}^{r}\langle Q_i^{k+1}-Q_i^k,\nabla f(Q_i^k)\rangle &\geq \sum_{i=1}^{r}\left(\frac{\gamma}{2}\|Q_i^{k+1}-Q_i^k\|_2^2+\frac{\alpha}{(g_i^k)^2}\|Q_i^{k+1}\|_2^2-\frac{\alpha}{(g_i^k)^2}\|Q_i^k\|_2^2\right) \\ &\geq \sum_{i=1}^{r}\left(\frac{\gamma}{2}\|Q_i^{k+1}-Q_i^k\|_2^2+\frac{\alpha}{(g_i^{k+1})^2}\|Q_i^{k+1}\|_2^2-\frac{\alpha}{(g_i^k)^2}\|Q_i^k\|_2^2\right). \end{aligned} \tag{6}$$

Since $f(Q_i)$ $(i=1,\cdots,r)$ satisfies Lemma 1.1, the following inequality holds:

$$\begin{aligned} \sum_{i=1}^{r}\left(f(Q_i^k)-f(Q_i^{k+1})\right) &\geq \sum_{i=1}^{r}\left(-\langle Q_i^{k+1}-Q_i^k,\nabla f(Q_i^k)\rangle-\frac{L}{2}\|Q_i^{k+1}-Q_i^k\|_2^2\right) \\ &\geq \sum_{i=1}^{r}\left(\frac{\alpha}{(g_i^{k+1})^2}\|Q_i^{k+1}\|_2^2-\frac{\alpha}{(g_i^k)^2}\|Q_i^k\|_2^2+\frac{\gamma-L}{2}\|Q_i^{k+1}-Q_i^k\|_2^2\right). \end{aligned} \tag{7}$$

Thus, we can obtain the following inequality:

$$\begin{aligned} F(Q^k,g^k)-F(Q^{k+1},g^{k+1}) &= \sum_{i=1}^{r}\left(f(Q_i^k)-f(Q_i^{k+1})+\frac{\alpha}{(g_i^k)^2}\|Q_i^k\|_2^2-\frac{\alpha}{(g_i^{k+1})^2}\|Q_i^{k+1}\|_2^2\right) \\ &\geq \frac{\gamma-L}{2}\sum_{i=1}^{r}\|Q_i^{k+1}-Q_i^k\|_2^2=\frac{\gamma-L}{2}\|Q^{k+1}-Q^k\|_F^2\geq 0. \end{aligned} \tag{8}$$

Therefore, $F(Q^k,g^k)$ is monotonically decreasing. So $F(Q^k,g^k)=\sum_{i=1}^{r}\left(\| \tilde{H}_i-Q_i^k \|_2^2+\frac{\alpha}{(g_i^k)^2}\| Q_i^k \|_2^2\right)\leq F(Q^1,g^1)$. Thus $\{Q^k\}$ is bounded. Summing all the inequality (8) for all $k\geq 1$, we obtain

$$F(Q^1,g^1)-F(Q^{k+1},g^{k+1})\geq\frac{\gamma-L}{2}\sum_{j=1}^{k}\|Q^{j+1}-Q^j\|_F^2. \tag{9}$$

As $\gamma>L$, the above implies that $\lim_{k\to\infty}\|Q^{k+1}-Q^k\|_F^2=0$. As $g_i^k=\frac{t\|Q_i^k\|_2^{\frac{2}{3}}}{\sum_{i=1}^{r}\|Q_i^k\|_2^{\frac{2}{3}}}$ $(\forall i=1,\cdots,r)$, $\lim_{k\to\infty}\|g^{k+1}-g^k\|_2^2=0$. As $\sum_{i=1}^{r}g_i=t>0$ and $g_i\geq 0$, the sequence $\{g_k\}$ is bounded.

II. Proof of Theorem 3.3

Proof In Theorem 3.2, we have proved that the sequence $\{Q^k,g^k\}$ is bounded. For any accumulation point (Q^*,g^*) of $\{Q^k,g^k\}$, suppose a subsequence (Q^{k_j},g^{k_j}) fulfills $\lim_{j\to\infty}Q^{k_j}=Q^*$ and $\lim_{j\to\infty}g^{k_j}=g^*$. In each iteration, we denote

$$\tau^k=\frac{2\alpha}{t^3}\left(\sum_{i=1}^{r}\| Q_i^k \|_2^{\frac{2}{3}}\right)^3. \tag{10}$$

Then there exists τ^* such that $\lim_{j\to\infty}\tau^{k_j}=\tau^*$. In our iteration process, $\nabla_Q F(Q^{k_j+1},g^{k_j})=0$, $\nabla_g F(Q^{k_j+1},g^{k_j+1})+\tau^{k_j+1}=0$, and $\sum_{i=1}^{r}g_i^{k_j}=t$. Letting $j\to\infty$, we have $\nabla_Q F(Q^*,g^*)=0$, $\nabla_g F(Q^*,g^*)+\tau^*=0$, and $\sum_{i=1}^{r}g_i^*=t$. So (Q^*,g^*) is a KKT point.

III. Fast Algorithm for Robust PCA [3]

In this section, we introduce the ℓ_1-filtering for solving Robust PCA problem. We sketch it below.

ℓ_1-filtering first randomly samples a $s_r r\times s_c r$ submatrix X^s from X, where $s_r>1$ and $s_c>1$ are the row and column oversampling rates, respectively. For simplicity, we assume that X^s is at the top left corner of matrix X. Then accordingly X, A, and E is partitioned into:

$$X=\begin{bmatrix}X^s & X^c\\ X^r & \hat{X}^s\end{bmatrix},\ A=\begin{bmatrix}A^s & A^c\\ A^r & \hat{A}^s\end{bmatrix},\ E=\begin{bmatrix}E^s & E^c\\ E^r & \hat{E}^s\end{bmatrix}, \tag{11}$$

where A is the low rank matrix we need to recover and E is the sparse error matrix.

Then ℓ_1-filtering recovers A^s, called the seed matrix, from X^s by solving a small-sized Robust PCA problem. Since $s_r r$ and $s_c r$ are both small as compared with m and n, the computation of recovering A^s is much cheaper than recovering the whole A.

Next, as the rank of A and A^s are both r, there must exist matrix Q and P satisfying the following equations (12)

$$A^c = A^s Q, \quad A^r = P^T A^s. \tag{12}$$

Since the matrix E is sparse, the matrices E^c and E^r are also sparse. So we can find matrix Q and P by minimizing the following problems:

$$\min_{E^c, Q} \parallel E^c \parallel_1, \quad s.t.\ X^c = A^s Q + E^c, \tag{13}$$

and

$$\min_{E^r, P} \parallel E^r \parallel_1, \quad s.t.\ X^r = P^T A^s + E^r, \tag{14}$$

respectively. For these two problems, using the alternating direction method (ADM) [4] to solve them is efficient. So we can get P and Q, thus A^c and A^r can be also obtained. Finally, only $\hat{A}^s$ needs to be computed. By the low-rankness of A, we can obtain

$$\hat{A}^s = P^T A^s Q. \tag{15}$$

In summary, the matrix A can be recovered with a complexity of $O(r^2(d+m))$ at each iteration [3], which is much lower than $O(rmd)$ when d and m are large.

References

[1] D. P. Bertsekas, *Nonlinear programming*. Athena Scientific, 1999.

[2] A. Beck and M. Teboulle, "A fast iterative shrinkage-thresholding algorithm for linear inverse problems," *SIAM Journal on Imaging Sciences*, vol. 2, no. 1, pp. 183–202, 2009.

[3] R. Liu, Z. Lin, Z. Su, and J. Gao, "Linear time principal component pursuit and its extensions using l_1 filtering," *Neurocomputing*, vol. 142, pp. 529–541, 2014.

[4] Z. Lin, R. Liu, and Z. Su, "Linearized alternating direction method with adaptive penalty for low-rank representation," in *Proceedings of the 25th Annual Conference on Neural Information Processing Systems*, Granada, Spain, Dec. 2011, pp. 612–620.