Advances in Pure Mathematics, 2019, 9, 523-533
http://www.scirp.org/journal/apm
ISSN Online: 2160-0384
ISSN Print: 2160-0368

Continuous Iteratively Reweighted Least Squares Algorithm for Solving Linear Models by Convex Relaxation

Xian Luo, Wanzhou Ye

Department of Mathematics, College of Science, Shanghai University, Shanghai, China
Email: 18817615683@163.com, wzhy@shu.edu.cn

How to cite this paper: Luo, X. and Ye, W.Z. (2019) Continuous Iteratively Reweighted Least Squares Algorithm for Solving Linear Models by Convex Relaxation. *Advances in Pure Mathematics*, **9**, 523-533.
https://doi.org/10.4236/apm.2019.96024

Received: May 13, 2019
Accepted: June 24, 2019
Published: June 27, 2019

Abstract

In this paper, we present continuous iteratively reweighted least squares algorithm (CIRLS) for solving the linear models problem by convex relaxation, and prove the convergence of this algorithm. Under some conditions, we give an error bound for the algorithm. In addition, the numerical result shows the efficiency of the algorithm.

Keywords

Linear Models, Continuous Iteratively Reweighted Least Squares, Convex Relaxation, Principal Component Analysis

1. Introduction

Low-dimensional linear models have broad applications in data analysis problems such as computer vision, pattern recognition, machine learning, and so on [1] [2] [3]. In most of these applications, the data set is noisy and contains numbers of outliers so that they are distributed in higher dimensional. Meanwhile, principal component analysis is the standard method [4] for finding a low-dimensional linear model. Mathematically, the problem can be present as

$$\min \sum_{x\in\chi} \|x-\Pi x\|^2 \quad \text{subject to} \quad \Pi \text{ is an orthoprojector and } tr\Pi = d \tag{1}$$

where χ is the data set consisting of N points in $\mathbb{R}^D$, $x \in \mathbb{R}^D$, a target dimension $d \in \{1,2,\cdots,D-1\}$, $\|\cdot\|^2$ denotes the l_2 norm on vectors, and tr refers to the trace.

Unfortunately, if the data set contains a large number of noise in the inliers and a substantial number of outliers, these nonidealities points can interfere with the linear models. To guard the subspace estimation procedure against outliers, statistics have proposed to replace the l_2 norm [5] [6] [7] with l_1 norm that is less sensitive to outliers. This idea leads the following optimization problem

$$\min \sum_{x\in\chi} \|x - \Pi x\| \quad \text{subject to} \quad \Pi \text{ is an orthoprojector and } tr\Pi = d \tag{2}$$

The optimization problem (2) is not convex, and we have no right to expect that the problem is tractable [8] [9]. Wright [10] proved that most matrices can be efficiently and exactly recovered from most error sign-and-support patterns by solving a simple convex program, for which they give a fast and provably convergent algorithm. Later, Candes [11] presented that under some suitable assumptions, it is possible to recover the subspace by solving a very convenient convex program called Principal Component Pursuit. Base on convex optimization, Lerman [12] proposed to use a relaxation of the set of orthogonal projectors to reach the convex formulation, and give the linear model as follows

$$\min \sum_{x\in\chi} \|x - Px\| \quad \text{subject to} \quad 0 \preceq P \preceq I \text{ and } trP = d \tag{3}$$

where the matrix P is the relaxation of orthoprojector Π, whose eigenvalues lie in the interval $[0,1]$ form a convex set. The curly inequality $\preceq$ denotes the semidefinite order: for symmetric matrices A and B, we write $A \preceq B$ if and only if $B - A$ is positive semidefinite, the problem (3) is called REAPER.

To obtain a d-dimensional linear model from a minimizer of REAPER, we need to consider the auxiliary problem

$$\min \sum_{x\in\chi} \|P_\star - \Pi\|_{S_1} \quad \text{subject to} \quad \Pi \text{ is an orthoprojector and } tr\Pi = d \tag{4}$$

where $P_\star$ is an optimal point of REAPER, $\|\cdot\|_{S_1}$ denotes Schatten 1-norm, in other words, the orthoprojector $\Pi_\star$ is closest to $P_\star$ in the Schatten 1-norm, and the range of $\Pi_\star$ is the linear model we want. Fortunately, Lerman has given the error bound between the d-dimensional subspace with the d-dimensional orthoprojector $\Pi_\star$ in Theorem 2.1 [12].

In this paper, we improve that the algorithm calls continuous iteratively reweighted least squares algorithm for solving REAPER (3), and under a weaker assumption on the data set, we can prove the algorithm is convergent. In the experiment part, we compare the algorithm with the IRLS algorithm [12].

The rest of the paper is organized as follows. In Section 2, we develop the CIRLS algorithm for solving problem (3). We present a detail convergence analysis for the CIRLS algorithm in Section 3. An efficient numerical is reported in Section 4. Finally, we conclude this paper in Section 5.

2. The CIRLS Algorithm

We give the CIRLS algorithm for solving optimization problem (3), and the algorithm is summarized as **Algorithm 1**.

Algorithm 1 Continuous iteratively reweighted least squares for solving problem(3)

Require: data set χ, $x \in \mathbb{R}^D$, choose parameters $\delta > 0$, $0 < \eta < 1$, stopping tolerance $\varepsilon > 0$
Ensure: $P^{(k)}$

Initialize $k = 0$, $\alpha^{(0)} = +\infty$, $\omega_{\delta_k,x} = 1, \forall x \in \chi$
For $k = 0, 1, 2, \cdots$ do
  Updata $P^{(k)} \leftarrow \underset{\substack{0 \preceq P \preceq I \\ trP = d}}{\arg\min} \sum_{x\in\chi} \omega_{\delta_k,x} ||x - P^{(k-1)}x||^2$;
  Updata $\omega_{\delta_k,x} \leftarrow \frac{1}{\max\{\delta_k, ||x - P^{(k)}x||\}}$, *where* $\delta_k = \eta^k + \delta$;
  Updata $\alpha^{(k)} \leftarrow \sum_{x\in\chi} \omega_{\delta_k,x} ||x - P^{(k)}x||^2$;
  Until $\alpha^{(k)} \geq \alpha^{(k-1)} - \varepsilon$
End for

In the next section, we prove the sequence $\{P^{(k)}\}_{k\geq 1}$ is convergent to P_δ. Furthermore, we also present the P_δ satisfied the bound with the optimal point $P_\star$ of REAPER.

3. Convergence of CIRLS Algorithm

In this section, we prove that the sequence $\{P^{(k)}\}_{k\geq 1}$ generated by Algorithm 1 is convergent to P_δ and we provide the P_δ satisfied the bound with the optimal point $P_\star$ of REAPER. Firstly, we start from the Lemma prepare for the proof of the following theorem.

Lemma 1 ([12], *Theorem* 4.1) *Assume that the set* χ *of observations does not lie in the union of two strict subspaces of* $\mathbb{R}^D$. *Then the iterates of IRLS Algorithm with* $\varepsilon = 0$ *converge to a point* P_δ *that satisfies the constraints of the REAPER problem. Moreover, the objective value at* P_δ *satisfies the bound*

$$\sum_{x\in\chi} \|x - P_\delta x\| - \sum_{x\in\chi} \|x - P_\star x\| \leq \frac{1}{2}\delta|\chi|, \tag{5}$$

where $P_\star$ *is an optimal point of REAPER.*

In lemma 1, under the assumption that the set χ of observations does not lie in the union of two strict subspaces of $\mathbb{R}^D$, they consider the convergence of the IRLS algorithm. However, verify whether a data set satisfies this assumption requires amounts of computation in theory, so we give a assumption that is easier to verify in following theorem.

Theorem 1 *Assume that the set* χ *of observations satisfies* $span\{\chi\} = \mathbb{R}^D$, *if* P_δ *is the limit point of the sequence* $\{P^{(k)}\}_{k\geq 1}$ *generated by* $CIRLS_\delta$ *algorithm, then* P_δ *is an optimal point of the following optimization problem*

$$\min\left(\sum_{\{x:\|x-Px\|\geq\delta\}} \|x - Px\| + \frac{1}{2}\sum_{\{x:\|x-Px\|<\delta\}} \left\{\frac{\|x - Px\|^2}{\delta} + \delta\right\}\right) \tag{6}$$
$$\text{subject to } 0 \preceq P \preceq I \text{ and } trP = d,$$

and satisfies

$$\sum_{x\in\chi} \|x - P_\delta x\| - \sum_{x\in\chi} \|x - P_\star x\| \leq \frac{1}{2}\delta|\chi|, \tag{7}$$

where $P_\star$ *is an optimal point of REAPER*(3).

Proof. Firstly, we consider the optimization model with **Algoritnm 1**,

$$\min\left(\sum_{\{x:\|x-Px\|\geq\delta\}}\|x-Px\|+\frac{1}{2}\sum_{\{x:\|x-Px\|<\delta\}}\left\{\frac{\|x-Px\|^2}{\delta}+\delta\right\}\right) \tag{8}$$
$$\text{subject to } 0\preceq P\preceq I \text{ and } trP=d,$$

then, we define iterative point $P^{(k+1)}$ of the optimization problem

$$P^{(k+1)}:=\arg\min\sum_{x\in\chi}\omega_{\delta_k,x}\|x-Px\|^2 \quad \text{subject to } 0\preceq P\preceq I \text{ and } trP=d. \tag{9}$$

where $\omega_{\delta_k,x}:=\dfrac{1}{\max\left\{\delta_k,\left\|x-P^{(k)}x\right\|\right\}}$.

For convenience, let $Q=I-P$, and the optimization model(9) convert into

$$Q^{(k+1)}:=\underset{\substack{0\preceq Q\preceq I\\ trQ=D-d}}{\arg\min}\sum_{x\in\chi}\omega_{\delta_k,x}\|Qx\|^2. \tag{10}$$

where $\omega_{\delta_k,x}:=\dfrac{1}{\max\left\{\delta_k,\left\|Q^{(k)}x\right\|\right\}}$.

Similarly, we convert the optimization model (8) into

$$\min F_\delta(Q)=\min\left(\sum_{\{x:\|Qx\|\geq\delta\}}\|Qx\|+\frac{1}{2}\sum_{\{x:\|Qx\|<\delta\}}\left(\frac{\|Qx\|^2}{\delta}+\delta\right)\right) \tag{11}$$
$$\text{subject to } 0\preceq Q\preceq I \text{ and } trQ=D-d.$$

Next we prove the convergence of $P^{(k)}$, consider the Huber-like function

$$H_{\delta_k}(x,y)=\begin{cases}\frac{1}{2}\left(\frac{x^2}{\delta_k}+\delta_k\right) & 0\leq y<\delta_k\\ \frac{1}{2}\left(\frac{x^2}{y}+y\right) & y\geq\delta_k\end{cases} \tag{12}$$

since $\frac{1}{2}\left(\frac{x^2}{y}+y\right)\geq|x|$, for any $y>0$, then

$$H_{\delta_k}(x,y)\geq H_{\delta_k}(x,|x|). \tag{13}$$

holds for any $y>0$, $x\in R$. We introduce the convex function

$$\begin{aligned}F_{\delta_k}(Q)&:=\sum_{x\in\chi}H_{\delta_k}(\|Qx\|,\|Qx\|)\\&=\sum_{\{x\in\chi:\|Qx\|\geq\delta_k\}}\|Qx\|+\frac{1}{2}\sum_{\{x\in\chi:\|Qx\|<\delta_k\}}\left(\frac{\|Qx\|^2}{\delta_k}+\delta_k\right),\end{aligned} \tag{14}$$

note that F is continuously differentiable at each matrix Q, and the gradient is

$$\begin{aligned}\nabla F_{\delta_k}(Q)&=\sum_{\{x\in\chi:\|Qx\|\geq\delta_k\}}\frac{Qxx^{\mathrm{T}}}{\|Qx\|}+\sum_{\{x\in\chi:\|Qx\|<\delta_k\}}\frac{Qxx^{\mathrm{T}}}{\delta_k}\\&=\sum_{x\in\chi}\frac{Qxx^{\mathrm{T}}}{\max\{\delta_k,\|Qx\|\}}.\end{aligned} \tag{15}$$

in addition, we introduce the function

$$
\begin{aligned}
&G_{\delta_k}\left(Q,Q^{(k)}\right) \\
&:= \sum_{x\in\chi} H_{\delta_k}\left(\|Qx\|,\left\|Q^{(k)}x\right\|\right) \\
&= \sum_{\left\{x\in\chi:\left\|Q^{(k)}x\right\|\ge\delta_k\right\}} \frac{1}{2}\left(\frac{\|Qx\|^2}{\left\|Q^{(k)}x\right\|}+\left\|Q^{(k)}x\right\|\right)+\sum_{\left\{x\in\chi:\left\|Q^{(k)}x\right\|<\delta_k\right\}} \frac{1}{2}\left(\frac{\|Qx\|^2}{\delta_k}+\delta_k\right),
\end{aligned}
\tag{16}
$$

then

$$
\nabla G_{\delta_k}\left(Q,Q^{(k)}\right)=\sum_{x\in\chi}\frac{Qxx^{\mathrm{T}}}{\max\left\{\delta_k,\left\|Q^{(k)}x\right\|\right\}}. \tag{17}
$$

By the definition of $G_{\delta_k}\left(Q,Q^{(k)}\right)$ we know that

$$
G_{\delta_k}\left(Q^{(k)},Q^{(k)}\right)=\sum_{x\in\chi} H_{\delta_k}\left(\left\|Q^{(k)}x\right\|,\left\|Q^{(k)}x\right\|\right)=F_{\delta_k}\left(Q^{(k)}\right), \tag{18}
$$

and

$$
\nabla G_{\delta_k}\left(Q^{(k)},Q^{(k)}\right)=\sum_{x\in\chi}\frac{Q^{(k)}xx^{\mathrm{T}}}{\max\left\{\delta_k,\left\|Q^{(k)}x\right\|\right\}}=\nabla F_{\delta_k}\left(Q^{(k)}\right), \tag{19}
$$

it is obvious that $G_{\delta_k}\left(\cdot,Q^{(k)}\right)$ is a smooth quadratic function, we may relate $G_{\delta_k}\left(\cdot,Q^{(k)}\right)$ through the expansion in $Q^{(k)}$ as follows

$$
\begin{aligned}
G_{\delta_k}\left(Q,Q^{(k)}\right)=&F_{\delta_k}\left(Q^{(k)}\right)+\left\langle Q-Q^{(k)},\nabla F_{\delta_k}\left(Q^{(k)}\right)\right\rangle \\
&+\frac{1}{2}\left\langle Q-Q^{(k)},C_k\left(Q^{(k)}\right)\left(Q-Q^{(k)}\right)\right\rangle,
\end{aligned}
\tag{20}
$$

where $C_k\left(Q^{(k)}\right)=\sum_{x\in\chi}\frac{xx^{\mathrm{T}}}{\max\left\{\delta_k,\left\|Q^{(k)}x\right\|\right\}}$ is the Hessian matrix of $G_{\delta_k}\left(Q,Q^{(k)}\right)$.

By the definition of $G_{\delta_k}\left(Q,Q^{(k)}\right)$ we know that $F_{\delta_k}(Q)=G_{\delta_k}(Q,Q)$, combines with (13) we have

$$
\begin{aligned}
F_{\delta_k}(Q)&=\sum_{x\in\chi}H_{\delta_k}\left(\|Qx\|,\|Qx\|\right) \\
&\le\sum_{x\in\chi}H_{\delta_k}\left(\|Qx\|,\left\|Q^{(k)}x\right\|\right) \\
&=G_{\delta_k}\left(Q,Q^{(k)}\right),
\end{aligned}
\tag{21}
$$

and since the optimization model(10) is equivalent to the model

$$
Q^{(k+1)}:=\underset{\substack{0\preceq Q\preceq I\\ trP=D-d}}{\arg\min}\, G_{\delta_k}\left(Q,Q^{(k)}\right), \tag{22}
$$

then we have the monotonicity property

$$
F_{\delta_k}\left(Q^{(k+1)}\right)\le G_{\delta_k}\left(Q^{(k+1)},Q^{(k)}\right)\le G_{\delta_k}\left(Q^{(k)},Q^{(k)}\right)=F_{\delta_k}\left(Q^{(k)}\right). \tag{23}
$$

We note that when $x\ge 0$, we have

$$
H_{\delta_k}(x,x)=\begin{cases} x & x\ge\delta_k \\ \frac{1}{2}\left(\frac{x^2}{\delta_k}+\delta_k\right) & 0\le x<\delta_k \end{cases} \tag{24}
$$

$$H_{\delta_{k+1}}(x,x)=\begin{cases} x & x\geq \delta_{k+1} \\ \frac{1}{2}\left(\frac{x^2}{\delta_{k+1}}+\delta_{k+1}\right) & 0\leq x<\delta_{k+1} \end{cases} \tag{25}$$

since $\delta_{k+1}<\delta_k$, on the one hand, when $x\geq \delta_{k+1}$, $H_{\delta_{k+1}}(x,x)=x$ holds, and $H_{\delta_k}(x,x)\geq x$ holds for any $x\in R$, therefore the inequality $H_{\delta_{k+1}}(x,x)\leq H_{\delta_k}(x,x)$ holds for $x\geq \delta_{k+1}$. On the other hand, when $x<\delta_{k+1}$, $H_{\delta_{k+1}}(x,x)=\frac{1}{2}\left(\frac{x^2}{\delta_{k+1}}+\delta_{k+1}\right)$, and $H_{\delta_k}(x,x)=\frac{1}{2}\left(\frac{x^2}{\delta_k}+\delta_k\right)$ holds, then we have

$$\begin{aligned} H_{\delta_k}(x,x)-H_{\delta_{k+1}}(x,x) &= \frac{1}{2}\left(\frac{x^2}{\delta_k}-\frac{x^2}{\delta_{k+1}}+\delta_k-\delta_{k+1}\right) \\ &= \frac{1}{2}(\delta_k-\delta_{k+1})\left(1-\frac{x^2}{\delta_k\delta_{k+1}}\right). \end{aligned} \tag{26}$$

since $\delta_k>\delta_{k+1}$, and $x<\delta_{k+1}$, $H_{\delta_k}(x,x)-H_{\delta_{k+1}}(x,x)>0$ holds for all $x\geq 0$, therefore

$$\begin{aligned} F_{\delta_{k+1}}\left(Q^{(k+1)}\right) &= \sum_{x\in\chi} H_{\delta_{k+1}}\left(\left\|Q^{(k+1)}x\right\|,\left\|Q^{(k+1)}x\right\|\right) \\ &\leq \sum_{x\in\chi} H_{\delta_k}\left(\left\|Q^{(k+1)}x\right\|,\left\|Q^{(k+1)}x\right\|\right) \\ &= F_{\delta_k}\left(Q^{(k+1)}\right), \end{aligned} \tag{27}$$

and according to (23), we have the following result

$$F_{\delta_{k+1}}\left(Q^{(k+1)}\right)\leq F_{\delta_k}\left(Q^{(k)}\right). \tag{28}$$

Since

$$Q^{(k+1)}=\underset{\substack{0\preceq Q\preceq I \\ trQ=D-d}}{\arg\min}\, G_{\delta_k}\left(Q,Q^{(k)}\right), \tag{29}$$

combined with the convex optimization variational inequalities with some constraints, we have

$$0\leq\left\langle Q-Q^{(k+1)},\nabla G_{\delta_k}\left(Q^{(k+1)},Q^{(k)}\right)\right\rangle \ \forall 0\preceq Q\preceq I, trQ=D-d, \tag{30}$$

base on (20), we have the equation

$$\nabla G_{\delta_k}\left(Q,Q^{(k)}\right)=\nabla F_{\delta_k}\left(Q^{(k)}\right)+C_k\left(Q^{(k)}\right)\left(Q-Q^{(k)}\right), \tag{31}$$

then we have

$$0\leq\left\langle Q^{(k)}-Q^{(k+1)},\nabla F_{\delta_k}\left(Q^{(k)}\right)+C_k\left(Q^{(k)}\right)\left(Q^{(k+1)}-Q^{(k)}\right)\right\rangle, \tag{32}$$

therefore

$$\begin{aligned} & G_{\delta_k}\left(Q^{(k+1)},Q^{(k)}\right) \\ &= F_{\delta_k}\left(Q^{(k)}\right)+\left\langle Q^{(k+1)}-Q^{(k)},\nabla F_{\delta_k}\left(Q^{(k)}\right)\right\rangle \\ &\quad +\frac{1}{2}\left\langle Q^{(k+1)}-Q^{(k)},C_k\left(Q^{(k)}\right)\left(Q^{(k+1)}-Q^{(k)}\right)\right\rangle \\ &\leq F_{\delta_k}\left(Q^{(k)}\right)-\left\langle Q^{(k+1)}-Q^{(k)},C_k\left(Q^{(k)}\right)\left(Q^{(k+1)}-Q^{(k)}\right)\right\rangle \end{aligned}$$

$$
\begin{aligned}
&+\frac{1}{2}\left\langle Q^{(k+1)}-Q^{(k)},C_{\delta_k}\left(Q^{(k)}\right)\left(Q^{(k+1)}-Q^{(k)}\right)\right\rangle \\
&\le F_{\delta_k}\left(Q^{(k)}\right)-\frac{1}{2}\left\langle Q^{(k+1)}-Q^{(k)},C_k\left(Q^{(k)}\right)\left(Q^{(k+1)}-Q^{(k)}\right)\right\rangle,
\end{aligned} \tag{33}
$$

and according to(23), then we get the inequality

$$
F_{\delta_k}\left(Q^{(k+1)}\right)\le F_{\delta_k}\left(Q^{(k)}\right)-\frac{1}{2}\left\langle Q^{(k+1)}-Q^{(k)},C_k\left(Q^{(k)}\right)\left(Q^{(k+1)}-Q^{(k)}\right)\right\rangle, \tag{34}
$$

add to $F_{\delta_{k+1}}\left(Q^{(k+1)}\right)\le F_{\delta_k}\left(Q^{(k+1)}\right)$, then we have

$$
\frac{1}{2}\left\langle Q^{(k+1)}-Q^{(k)},C_{\delta_k}\left(Q^{(k)}\right)\left(Q^{(k+1)}-Q^{(k)}\right)\right\rangle \le F_{\delta_k}\left(Q^{(k)}\right)-F_{\delta_{k+1}}\left(Q^{(k+1)}\right). \tag{35}
$$

Let $m:=\max_{k\ge 1}\left\{\delta_k,\max_{x\in\chi}\{\|x\|\}\right\}$, since $Q^{(k+1)}$, $Q^{(k)}$ and $C_k\left(Q^{(k)}\right)$ are all symmetric matrix, then the inequality

$$
\begin{aligned}
&\left\langle Q^{(k+1)}-Q^{(k)},C_k\left(Q^{(k)}\right)\left(Q^{(k+1)}-Q^{(k)}\right)\right\rangle \\
&\ge tr\left(\left(Q^{(k+1)}-Q^{(k)}\right)^2 C_k\left(Q^{(k)}\right)\right),
\end{aligned} \tag{36}
$$

holds, in addition, $0\preceq Q^{(k+1)}\preceq I$, $0\preceq Q^{(k)}\preceq I$, thus $\left(Q^{(k+1)}-Q^{(k)}\right)^2$ is positive semidefinite matrix, and $\left\|Q^{(k)}x\right\|\le\left\|Q^{(k)}\right\|\cdot\|x\|\le\|x\|$, then we have $\max\left\{\delta_k,\left\|Q^{(k)}x\right\|\right\}\le\max\left\{\delta_k,\|x\|\right\}\le m$, therefore

$$
C_k\left(Q^{(k)}\right)=\sum_{x\in\chi}\frac{xx^{\mathrm{T}}}{\max\left\{\delta_k,\left\|Q^{(k)}x\right\|\right\}}\succeq\frac{1}{m}\sum_{x\in\chi}xx^{\mathrm{T}}\succeq\frac{1}{m}\lambda_{\min}\left(\sum_{x\in\chi}xx^{\mathrm{T}}\right)I, \tag{37}
$$

so we have

$$
\left(Q^{(k+1)}-Q^{(k)}\right)^2 C_k\left(Q^{(k)}\right)\succeq\frac{1}{m}\lambda_{\min}\left(\sum_{x\in\chi}xx^{\mathrm{T}}\right)\left(Q^{(k+1)}-Q^{(k)}\right)^2, \tag{38}
$$

thus we can get the inequality as follows

$$
tr\left(Q^{(k+1)}-Q^{(k)}\right)^2 C_k\left(Q^{(k)}\right)\ge\frac{1}{m}\lambda_{\min}\left(\sum_{x\in\chi}xx^{\mathrm{T}}\right)\left\|Q^{(k+1)}-Q^{(k)}\right\|_F^2. \tag{39}
$$

Let $\mu=\frac{1}{2m}\lambda_{\min}\left(\sum_{x\in\chi}xx^{\mathrm{T}}\right)$, then $\mu>0$, and we have

$$
\left\|Q^{(k+1)}-Q^{(k)}\right\|_F^2\le\frac{1}{\mu}\left(F_{\delta_k}\left(Q^{(k)}\right)-F_{\delta_{k+1}}\left(Q^{(k+1)}\right)\right), \tag{40}
$$

since $F_{\delta_k}(Q)\ge 0$, so we have

$$
\sum_{k=1}^{\infty}\left\|Q^{(k+1)}-Q^{(k)}\right\|_F^2\le\frac{1}{\mu}F_{\delta_1}\left(Q^{(1)}\right)<+\infty, \tag{41}
$$

therefore, we get the following limits

$$
\lim_{k\to\infty}\left\|Q^{(k+1)}-Q^{(k)}\right\|_F=0. \tag{42}
$$

Since the set $\{Q:0\preceq Q\preceq I,trQ=D-d\}$ is bounded and closed convex set, and $Q^{(k+1)}\in\{Q:0\preceq Q\preceq I,trQ=D-d,k=1,2,\cdots\}$, then the sequence $\left\{Q^{(k)}\right\}_{k\ge 1}$ exist convergent subsequences, we suppose $\left\{Q^{(k_i)}\right\}_{i\ge 1}$ is one of the subsequences and

$$\tilde{Q} = \lim_{i \to +\infty} \left\{ Q^{(k_i)} \right\}. \tag{43}$$

Since for any $i \geq 1$, we have

$$\begin{aligned} \left\| Q^{(k_i+1)} - \tilde{Q} \right\|_F &= \left\| Q^{(k_i+1)} - Q^{(k_i)} + Q^{(k_i)} - \tilde{Q} \right\|_F \\ &\leq \left\| Q^{(k_i+1)} - Q^{(k_i)} \right\|_F + \left\| Q^{(k_i)} - \tilde{Q} \right\|_F, \end{aligned} \tag{44}$$

and

$$\lim_{k \to \infty} \left\| Q^{(k+1)} - Q^{(k)} \right\|_F = 0, \tag{45}$$

so we have

$$\lim_{i \to +\infty} \left\| Q^{(k_i+1)} - \tilde{Q} \right\|_F = 0, \tag{46}$$

in other words

$$\lim_{i \to +\infty} Q^{(k_i+1)} = \tilde{Q}. \tag{47}$$

According to the definition of $Q^{(k_i+1)}$ and (10), for any $Q: 0 \preceq Q \preceq I, trQ = D - d$, we have

$$0 \leq \left\langle Q - Q^{(k_i+1)}, \nabla F_{\delta_{k_i}}\left(Q^{(k_i)}\right) + C_{k_i}\left(Q^{(k_i)}\right)\left(Q^{(k_i+1)} - Q^{(k_i)}\right) \right\rangle. \tag{48}$$

Since

$$\begin{aligned} \lim_{i \to +\infty} \nabla F_{\delta_{k_i}}\left(Q^{(k_i)}\right) &= \lim_{i \to +\infty} \sum_{x \in \chi} \frac{Q^{(k_i)} x x^{\mathrm{T}}}{\max\left\{\delta_{k_i}, \left\| Q^{(k_i)} x \right\|\right\}} \\ &= \sum_{x \in \chi} \lim_{i \to +\infty} \frac{Q^{(k_i)} x x^{\mathrm{T}}}{\max\left\{\delta_{k_i}, \left\| Q^{(k_i)} x \right\|\right\}} \\ &= \sum_{x \in \chi} \frac{\tilde{Q} x x^{\mathrm{T}}}{\max\left\{\delta, \left\| \tilde{Q} x \right\|\right\}} \\ &= \nabla F_{\delta}\left(\tilde{Q}\right) \end{aligned} \tag{49}$$

and

$$\begin{aligned} &\left\| C_{k_i}\left(Q^{(k_i)}\right)\left(Q^{(k_i+1)} - Q^{(k_i)}\right) \right\|_F \\ &= \left\| \sum_{x \in \chi} \frac{x x^{\mathrm{T}}}{\max\left\{\delta_{k_i}, \left\| Q^{(k_i)} x \right\|\right\}} \left(Q^{(k_i+1)} - Q^{(k_i)}\right) \right\|_F \\ &\leq \sum_{x \in \chi} \frac{\left\| x x^{\mathrm{T}} \right\|_F}{\max\left\{\delta_{k_i}, \left\| Q^{(k_i)} x \right\|\right\}} \cdot \left\| Q^{(k_i+1)} - Q^{(k_i)} \right\|_F \\ &\leq \left(\sum_{x \in \chi} \frac{\left\| x x^{\mathrm{T}} \right\|_F}{\delta_{k_i}} \right) \cdot \left\| Q^{(k_i+1)} - Q^{(k_i)} \right\|_F \\ &\leq \frac{1}{\delta} \left(\sum_{x \in \chi} \left\| x x^{\mathrm{T}} \right\|_F \right) \cdot \left\| Q^{(k_i+1)} - Q^{(k_i)} \right\|_F \end{aligned} \tag{50}$$

as well as $\left\|Q^{(k_i+1)}-Q^{(k_i)}\right\|_F \to 0(i\to+\infty)$, so we have

$$\lim_{i\to+\infty}\left\|C_{k_i}\left(Q^{(k_i)}\right)\left(Q^{(k_i+1)}-Q^{(k_i)}\right)\right\|_F=0, \tag{51}$$

thus

$$\lim_{i\to+\infty}C_{k_i}\left(Q^{(k_i)}\right)\left(Q^{(k_i+1)}-Q^{(k_i)}\right)=0. \tag{52}$$

Taking the limit at both ends of the inequality (48) and using the continuity of the inner product, for any $Q:0\preceq Q\preceq I, trQ=D-d$, we have

$$0\le\left\langle Q-\tilde{Q},\nabla F_\delta\left(\tilde{Q}\right)\right\rangle, \tag{53}$$

the variational inequalities demonstrate that

$$\tilde{Q}=\arg\min F_\delta(Q)\quad\text{subject to } 0\le Q\le I,\ trQ=D-d, \tag{54}$$

it means that the limit points Q_δ of any convergent subsequence generated by $\left\{Q^{(k)}\right\}_{k\ge1}$ is an optimal point of the following optimization problem

$$Q_\delta=\arg\min F_\delta(Q)\quad\text{subject to } 0\le Q\le I,\ trQ=D-d. \tag{55}$$

Now we set $P_\delta=I-Q_\delta$ and define $F_0(P):=\sum_{x\in\chi}\left\|(I-P)x\right\|$ with $0\le P\le I$ and $trP=d$. And we define $F_\delta(P):=\sum_{x\in\chi}H_\delta\left(\left\|(I-P)x\right\|,\left\|(I-P)x\right\|\right)$, and $P_\star:=\arg\min F_0(P)$ with respect to the feasible set $0\le P\le I$, and $trP=d$, then we have

$$P_\delta=\arg\min F_\delta(P),\quad\text{subject to } 0\le P\le I\ ,trQ=d. \tag{56}$$

According to the lemma 3.1, we have

$$0\le F_0(P_\delta)-F_0(P_\star)\le\frac{1}{2}\delta|\chi|, \tag{57}$$

where $|\chi|$ denotes the number of elements of χ.

4. Numerical Experiments

In this section, we present a numerical experiment to show the efficiency of the CIRLS algorithm for solving problem (3). We compare the performance of our algorithm with IRLS on the data generated from the following model. In the test, we randomly choose *Nin* inliers sampled from the d-dimensional Multivariate Normal distribution $N(0,\Pi_L)$ on subspace L and add *Nout* outliers sampled from a uniform distribution on $[0,1]^D$, we also add a Gaussian noise $N(0,\varepsilon^2 I)$. The experiment is performed in R and all the experimental results were averaged over 10 independent trials.

The parameters were set the same as [12] that $\varepsilon=10^{-15}$, $\varepsilon=0.01$ and $\delta=10^{-10}$, we choose η from $[0,1]$ and we set $\eta=\frac{1}{2}$ in general. All the other parameters of the two algorithms were set to be the same, *Nout* means the number of outliers and *Nin* is the number of inliers, *D* is the ambient dimension and *d* is the subspace dimension we want. From the model above, we get the data sets with values in the table below, then we calculate the iterative number

Table 1. The iterative number of the CIRLS and IRLS for different dimension.

D	Methods	d	Nin	$Nout$	n (iterative)
10	IRLS	2	50	200	22
	CIRLS	2	50	200	13
50	IRLS	10	50	200	51
	CIRLS	10	50	200	30
100	IRLS	10	50	200	51
	CIRLS	10	50	200	31
150	IRLS	10	50	200	53
	CIRLS	10	50	200	33
200	IRLS	10	50	200	53
	CIRLS	10	50	200	35
250	IRLS	10	50	200	51
	CIRLS	10	50	200	30
300	IRLS	10	70	280	49
	CIRLS	10	70	280	30

$n(\text{iterative})$ with the two algorithms through the data sets and the given parameters. The results are shown in Table 1.

We focus on the convergence speed of the two algorithms. Table 1 reports the numerical results of the two algorithms for different space dimension. From the result, we can see that in different dimension, the CIRLS algorithm performs better than IRLS algorithm in convergent efficiency.

5. Conclusion

In this paper, we propose an efficient continuous iteratively reweighted least squares algorithm for solving REAPER problem, and we prove the convergence of the algorithm. In addition, we present a bound between the convergent limit and the optimal point of REAPER problem. Moreover, in the experiment part, we compare the algorithm with the IRLS algorithm to show that our algorithm is convergent and performs better than IRLS algorithm in the rate of convergence.

Conflicts of Interest

The authors declare no conflicts of interest regarding the publication of this paper.

References

[1] Torre, F.D. and Black, M.J. (2001) Robust Principal Component Analysis for Computer Vision. *Proceedings of Eighth IEEE International Conference on Computer Vision*, Vancouver, 7-14 July 2001, 362-369. https://doi.org/10.1109/ICCV.2001.937541

[2] Wanger, A., Wright, J. and Ganesh, A. (2009) Towards a Practical Face Recognition System: Robust Registration Tillumination by Sparse Represention. 2009 *IEEE Conference on Computer Vision and Pattern Recognition*, Miami, FL, 20-25 June 2009, 20-25. https://doi.org/10.1109/CVPR.2009.5206654

[3] Ho, J., Yang, M., Lim, J., Lee, K. and Kriegman, D. (2003) Clustering Appearances of Objects under Varying Illumination Conditions. 2003 *IEEE Computer Vision and Pattern Recognition*, Madison, WI, 18-20 June 2003, 11-18.

[4] Jolliffe, I.T. (2002) Principal Component Analysis. *Journal of Marketing Research*, **25**, 513.

[5] Watson, G.A. (2002) On the Gauss-Newton Method for l1 Orthogonal Distance Regression. *IMA Journal of Numerical Analysis*, **22**, 345-357. https://doi.org/10.1093/imanum/22.3.345

[6] Ding, C., Zhou, D., He, X. and Zha, H. (2006) R1-PCA: Rotational Invariant L1-Norm Principal Component Analysis for Robust Subspace Factorization. *The 23rd International Conference on Machine Learning*, **23**, 281-288. https://doi.org/10.1145/1143844.1143880

[7] Kwak, N. (2008) Principal Component Analysis Based on L1-Norm Maximization. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Pittsburgh, PA, 25-29 June 2006, 1672-1680. https://doi.org/10.1109/TPAMI.2008.114

[8] Maronna, R.A., Martin, D.R. and Yohai, V.J. (2006) Robust Statistics. Wiley Series in Probability and Statistics. Wiley, Chichester. https://doi.org/10.1002/0470010940

[9] Zhang, T., Szlam, A. and Lerman, G. (2009) Median K-Flats for Hybrid Linear Modeling with Many Outliers. *IEEE International Conference on Computer Vision Workshops*, **12**, 234-241.

[10] Wright, J., Peng, Y.G. and Ma, Y. (2009) Robust Principal Component Analysis: Exact Recovery of Corrupted Low-Rank Matrices by Convex Optimization. 2009 *23rd Annual Conference on Neural Information Processing Systems*, Vancouver, Canada, 7-10 December 2009, 2080-2088.

[11] Candes, E.J., Li, X., Ma, Y. and Wright, J. (2011) Robust Principal Component Analysis? *Journal of the ACM*, **58**, Article No. 11. https://doi.org/10.1145/1970392.1970395

[12] Lerman, G., McCoy, M.B., Tropp, J.T. and Zhang, T. (2015) Robust Computation of Linear Models by Convex Relaxation. *Foundations of Computational Mathematics*, **15**, 363-410. https://doi.org/10.1007/s10208-014-9221-0