Robust Computation of Linear Modeling

Teng Zhang

University of Minnesota, Institute of Mathematics and its Applications
SIAM Conference on Imaging Science

May 21, 2011
Outline

- Background: Robust Principal Components Analysis (PCA)
- A new formulation for robust PCA
- Theory for exact recovery of the subspace
- Algorithm development
- Experiments
Problem Formulation

- Given: a linear subspace $L^*$ and a data set $\mathcal{X} = \{x_i\}_{i=1}^N \subset \mathbb{R}^D$, which contains some points sampled from $L^*$ (we call them inliers) and outliers sampled from $\mathbb{R}^D \setminus L^*$.
- Goal: recover $L^*$ using $\mathcal{X}$.
- Fact: PCA is sensitive to outliers:
History

- Covariance estimators in statistics community: $M$-estimator, $S$-estimator, MVD (minimum volume ellipsoid) estimator, MCD (minimum covariance determinant) estimator, Stahel-Donoho estimator. See review by Maronna et al. (06)
- Projection Pursuit: Li & Chen (85), Ammann (93), McCoy & Tropp (10)
- Outlier detection and removal: Torre & Black (01), Xu et al. (10)
Convex optimization based on nuclear norm: Xu et al. (10), McCoy & Tropp (10) (inspired by related works by Chandrasekaran et al. (09), Candès et al. (09)).

\[
\text{minimize } \|L\|_* + \lambda \|O\|_{(2,1)}, \text{ s.t. } X_{N \times D} = L + O,
\]

where \(\| \cdot \|_*\) and \(\| \cdot \|_{(2,1)}\) are the nuclear norm and sum of \(l_2\) norms of rows respectively.

- Recover \(L^*\) by the span of the rows of \(L\).
- Theoretical guarantee on exact recovery of \(L^*\), and tractable algorithm.
Motivation of the new formulation

- The classical method minimizes the sum of the squared residual:

\[ \hat{L} = \arg \min_{L} \sum_{i=1}^{N} \text{dist}(x_i, L)^2. \]

- For robustness to outliers, people use the sum of unsquared residual (Spath and Watson, 87; Nyquist, 88):

\[ \hat{L} = \arg \min_{L} \sum_{i=1}^{N} \text{dist}(x_i, L). \]

- Nonconvex optimization—no tractable algorithm.
Formulation

- Rewrite the optimization problem as:

\[
\hat{L} = \arg\min_{d\text{-dimensional subspaces } L} \sum_{i=1}^{N} \text{dist}(x_i, L) = \arg\min_{L} \sum_{i=1}^{N} \|P_{L^\perp}x_i\|.
\]

- \((Z \& Lerman, 11)\) Use \(Q\) as the convex relaxation of \(P_{L^\perp}\), we have

\[
\hat{Q} = \arg\min_{Q \in \mathbb{H}} F(Q), \text{ where } F(Q) := \sum_{i=1}^{N} \|Qx_i\|, \quad (1)
\]

where

\[
\mathbb{H} := \{Q \in \mathbb{R}^{D \times D} : Q = Q^T, \text{tr}(Q) = 1\}. \quad (2)
\]

- The condition \(\text{tr}(Q) = 1\) guarantees that the solution is not a zero matrix.
Property of formulation

- Convex
- No parameter required
- Can not handle arbitrarily large outliers
Theoretical justification using deterministic conditions

(Z & Lerman, 11) Denote the set of inliers and outliers by $\mathcal{X}_1$ and $\mathcal{X}_0$ respectively, and denote $\dim(L^*)$ by $d$. If the following two conditions are satisfied, then we have $L^* \subseteq \ker(\hat{Q})$.

$$\min_{Q \in H, Q p_{L^* \perp} = 0} \sum_{x \in \mathcal{X}_1} \|Qx\| > \sqrt{2} \min_{v \in L^* \perp, \|v\| = 1} \sum_{x \in \mathcal{X}_0} |v^T x|,$$ \hspace{1cm} (3)

$$\min_{Q \in H, Q p_{L^* \perp} = 0} \sum_{x \in \mathcal{X}_1} \|Qx\| > \sqrt{2} \max_{v \in L^*, \|v\| = 1} \sum_{x \in \mathcal{X}_0} |v^T x|.$$ \hspace{1cm} (4)
If the following condition is also satisfied, then we recover $L^*$ exactly: $L^* = \ker(\hat{Q})$

- Any minimizer of the following oracle problem

$$\hat{Q}_0 \coloneqq \arg\min_{Q \in \mathcal{H}, Q P_{L^*} = 0} F(Q)$$

satisfies

$$\text{rank}(\hat{Q}_0) = D - d.$$
Some remarks

- This method can obtain the dimension of the subspace by the number of zero eigenvalues of $\hat{Q}$.
- Is there a stronger method if we know the dimension of the subspace in advance?
When we know the dimension $d$

- Recall that we minimize

$$
\hat{Q} = \arg \min_{Q \in \mathcal{H}} F(Q), \text{ where } F(Q) := \sum_{i=1}^{N} \|Qx_i\|,
$$

where

$$
\mathcal{H} := \{Q \in \mathbb{R}^{D \times D} : Q = Q^T, \text{tr}(Q) = 1\}.
$$

- $Q$ is a convex relaxation of $P_{L^\perp}$.
- The convex hull of $P_{L^\perp}$ for all $d$-dimensional subspace $L$ is:

$$
\mathcal{H}_1 = \{Q \in \mathbb{R}^{D \times D} : Q = Q^T, 0 \leq Q \leq I, \text{tr}(Q) = D - d\}.
$$

- We propose Reaper algorithm, which minimizers

$$
\hat{Q} = \arg \min_{Q \in \mathcal{H}_1} F(Q).
$$
A probabilistic model

- Gaussian inliers and Gaussian outliers
- $N_{\text{in}}$: number of inliers, $N_{\text{out}}$: number of outliers
- $\sigma_{\text{in}}^2$: variance of inliers, $\sigma_{\text{out}}^2$: variance of outliers
- $\rho_{\text{in}} = N_{\text{in}}/d$, $\rho_{\text{out}} = N_{\text{out}}/D$

(Lerman, Z, McCoy & Tropp, 12) When

$$\rho_{\text{in}} > C_1 + C_2 \beta + C_3 \frac{\sigma_{\text{out}}}{\sigma_{\text{in}}} (\rho_{\text{out}} + 1 + 4 \beta)$$

then $\hat{Q} = P_{L^* \perp} = I - P_{L^*}$ w.p. at least $1 - 4e^{-\beta d}$.

We can estimate that $C_1 < 13$, $C_2 < 7$ and $C_3 < 16$. 
S-reaper algorithm: a variant

- We can obtain extra robustness by first projecting all points to the sphere.
- This variant can handle outliers with arbitrarily large magnitude.
- We call this variant S-reaper algorithm.

(Lerman, Z, McCoy & Tropp, 12) When

\[ \rho_{in} > \tilde{C}_1 + \tilde{C}_2 \beta + \tilde{C}_3 \rho_{out} \]

then \( \hat{Q} = P_{L^* \perp} = I - P_{L^*} \) w.p. at least \( 1 - 4e^{-\beta d} \).
Algorithm

- Recall the objective function $F(Q) = \sum_{i=1}^{N} \|Qx_i\|$
- Heuristics proposal for IRLS (iteratively reweighted least square) algorithm:

$$Q_{\text{new}} = \arg\min_Q \sum_{i=1}^{N} \frac{\|Qx_i\|^2}{\|Q_{\text{old}}x_i\|}.$$

- When $Q \in \mathbb{H}$, the update formula is

$$Q_{\text{new}} = \left(\sum_{i=1}^{N} \frac{x_i x_i^T}{\|Q_{\text{old}}x_i\|}\right)^{-1} \frac{\text{tr}\left(\left(\sum_{i=1}^{N} \frac{x_i x_i^T}{\|Q_{\text{old}}x_i\|}\right)^{-1}\right)}{\text{tr}\left(\left(\sum_{i=1}^{N} \frac{x_i x_i^T}{\|Q_{\text{old}}x_i\|}\right)^{-1}\right)}.$$ 

- When $Q \in \mathbb{H}_1$ (i.e., Reaper algorithm),

$$Q_{\text{new}} = c_2 \min\left(c_1 I, \left(\sum_{i=1}^{N} \frac{x_i x_i^T}{\|Q_{\text{old}}x_i\|}\right)^{-1}\right);$$

$c_1$ and $c_2$ are chosen such that $Q_{\text{new}} \in \mathbb{H}_1$, i.e., $\|Q_{\text{new}}\|_2 = 1$, $\text{tr}(Q_{\text{new}}) = D - d$. 

15 / 21
Regularization in algorithm

- The update formula fails when $\|Q_{old}x_i\| = 0$
- Use IRLS with regularized weight:

$$Q_{new} = \arg\min_Q \sum_{i=1}^N \frac{\|Qx_i\|^2}{\max(\|Q_{old}x_i\|, \delta)}.$$  

- Then the update formula for the case $Q \in \mathbb{H}$ is

$$Q_{new} = \left(\sum_{i=1}^N \frac{x_ix_i^T}{\max(\|Q_{old}x_i\|, \delta)}\right)^{-1} / \text{tr} \left(\left(\sum_{i=1}^N \frac{x_ix_i^T}{\max(\|Q_{old}x_i\|, \delta)}\right)^{-1}\right)$$

Reaper algorithm (i.e., $Q \in \mathbb{H}_1$) can be regularized similarly.
Convergence of algorithm

- The algorithm converges to the minimizer of the objective function: \( \| Q_k - Q^* \| \to 0 \).
- The proof of the convergence depends on the assumption that
  \[ \mathcal{X} \cap L_1 \cup \mathcal{X} \cap L_2 \neq \mathcal{X}, \text{ for all } (D - 1)\text{-dimensional subspaces } L_1, L_2 \]
- This condition holds when \( N \geq 2D - 1 \) and \( \{x_i\}_{i=1}^N \) lie in general positions.
- Empirically Reaper and S-reaper algorithms converge linearly.
Experiment

- 64 images of a single face under different illuminations from the Extended Yale Face database (used as inliers)
- 400 additional random images from the BACKGROUND/Google folder of the Caltech101 database (used as outliers)
- Resolution downsampling to $20 \times 20$
- The face images lie on a nine-dimensional subspace (Basri & Jacobs, 03)
- Learn the subspace from a data set that contain 32 face images and 400 other random images.
Experiment

We compare our Reaper and S-reaper algorithms with PCA, SPCA (PCA with spherical projection), LLD (the convex method based on nuclear norm):
Conclusions

- We proposed a new formulation for robust PCA.
- We gave theoretical guarantee on exact recovery of the subspace.
- We have fast implementations.
Collaborators:

- Gilad Lerman (University of Minnesota)
- Michael Mccoy (California Institute of Technology)
- Joel Tropp (California Institute of Technology)

Email address: zhang620@umn.edu