Robust Computation of Linear Modeling

Teng Zhang

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

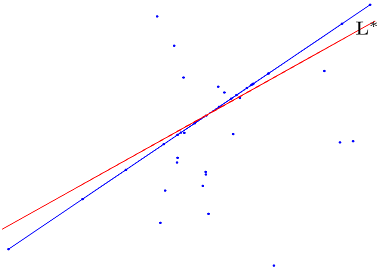
May 21, 2011

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 $\mathcal{X} = \{\mathbf{x}_i\}_{i=1}^N \subset \mathbb{R}^D$, which contains some points sampled from L^* (we call them inliers) and outliers sampled from $\mathbb{R}^D \setminus \mathrm{L}^*$.
- 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)

History

- 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)).

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

  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 $\mathbf{L}$.
- 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:

$$\hat{\mathrm{L}} = \arg\min \sum_{i=1}^{N} \mathrm{dist}(\mathbf{x}_i, \mathrm{L})^2.$$

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

$$\hat{\mathrm{L}} = \arg\min \sum_{i=1}^{N} \mathrm{dist}(\mathbf{x}_i, \mathrm{L}).$$

- Nonconvex optimization–no tractable algorithm.

Formulation

- Rewrite the optimization problem as:

$$\hat{\mathrm{L}} = \underset{d\text{-dimensional subspaces } \mathrm{L}}{\arg\min} \sum_{i=1}^{N} \operatorname{dist}(\mathbf{x}_i, \mathrm{L}) = \underset{\mathrm{L}}{\arg\min} \sum_{i=1}^{N} \|\mathbf{P}_{\mathrm{L}^\perp} \mathbf{x}_i\|.$$

- (Z & Lerman, 11) Use $\mathbf{Q}$ as the convex relaxation of $\mathbf{P}_{\mathrm{L}^\perp}$, we have

$$\hat{\mathbf{Q}} = \underset{\mathbf{Q} \in \mathbb{H}}{\arg\min}\, F(\mathbf{Q}), \text{ where } F(\mathbf{Q}) := \sum_{i=1}^{N} \|\mathbf{Q}\mathbf{x}_i\|, \quad (1)$$

where

$$\mathbb{H} := \{\mathbf{Q} \in \mathbb{R}^{D \times D} : \mathbf{Q} = \mathbf{Q}^T, \operatorname{tr}(\mathbf{Q}) = 1\}. \quad (2)$$

- The condition $\operatorname{tr}(\mathbf{Q}) = 1$ 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 $\dim(\mathrm{L}^*)$ by d. If the following two conditions are satisfied, then we have $\mathrm{L}^* \subseteq \ker(\hat{\mathbf{Q}})$.

- $$\min_{\mathbf{Q}\in\mathbb{H},\mathbf{Q}\mathbf{P}_{\mathrm{L}^{*\perp}}=\mathbf{0}} \sum_{\mathbf{x}\in\mathcal{X}_1} \|\mathbf{Q}\mathbf{x}\| > \sqrt{2} \min_{\mathbf{v}\in\mathrm{L}^{*\perp},\|\mathbf{v}\|=1} \sum_{\mathbf{x}\in\mathcal{X}_0} |\mathbf{v}^T\mathbf{x}|, \quad (3)$$

- $$\min_{\mathbf{Q}\in\mathbb{H},\mathbf{Q}\mathbf{P}_{\mathrm{L}^{*\perp}}=\mathbf{0}} \sum_{\mathbf{x}\in\mathcal{X}_1} \|\mathbf{Q}\mathbf{x}\| > \sqrt{2} \max_{\mathbf{v}\in\mathrm{L}^{*},\|\mathbf{v}\|=1} \sum_{\mathbf{x}\in\mathcal{X}_0} |\mathbf{v}^T\mathbf{x}|. \quad (4)$$

Theoretical justification using deterministic conditions

If the following condition is also satisfied, then we recover L^* exactly: $\mathrm{L}^* = \ker(\hat{\mathbf{Q}})$

- Any minimizer of the following oracle problem

$$\hat{\mathbf{Q}}_0 := \underset{\mathbf{Q} \in \mathbb{H}, \mathbf{Q}\mathbf{P}_{\mathrm{L}^*} = \mathbf{0}}{\arg\min} F(\mathbf{Q}) \tag{5}$$

  satisfies

$$\operatorname{rank}(\hat{\mathbf{Q}}_0) = D - d. \tag{6}$$

Some remarks

- This method can obtain the dimension of the subspace by the number of zero eigenvalues of $\hat{\mathbf{Q}}$.
- 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

$$\hat{\mathbf{Q}} = \arg\min_{\mathbf{Q}\in\mathbb{H}} F(\mathbf{Q}), \text{ where } F(\mathbf{Q}) := \sum_{i=1}^{N} \|\mathbf{Q}\mathbf{x}_i\|, \tag{7}$$

where

$$\mathbb{H} := \{\mathbf{Q} \in \mathbb{R}^{D\times D} : \mathbf{Q} = \mathbf{Q}^T, \operatorname{tr}(\mathbf{Q}) = 1\}. \tag{8}$$

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

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

- We propose Reaper algorithm, which minimizers

$$\hat{\mathbf{Q}} = \arg\min_{\mathbf{Q}\in\mathbb{H}_1} F(\mathbf{Q}). \tag{9}$$

A probabilistic model

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

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

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

then $\hat{\mathbf{Q}} = \mathbf{P}_{\mathrm{L}^{*\perp}} = \mathbf{I} - \mathbf{P}_{\mathrm{L}^*}$ 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

$$\rho_{in} > \tilde{C}_1 + \tilde{C}_2\beta + \tilde{C}_3\rho_{\mathrm{out}}$$

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

Algorithm

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

$$\mathbf{Q}_{\text{new}} = \arg\min_{\mathbf{Q}} \sum_{i=1}^{N} \frac{\|\mathbf{Q}\mathbf{x}_i\|^2}{\|\mathbf{Q}_{\text{old}}\mathbf{x}_i\|}.$$

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

$$\mathbf{Q}_{\text{new}} = \left(\sum_{i=1}^{N} \frac{\mathbf{x}_i\mathbf{x}_i^T}{\|\mathbf{Q}_{\text{old}}\mathbf{x}_i\|}\right)^{-1} / \operatorname{tr}\left(\left(\sum_{i=1}^{N} \frac{\mathbf{x}_i\mathbf{x}_i^T}{\|\mathbf{Q}_{\text{old}}\mathbf{x}_i\|}\right)^{-1}\right)$$

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

$$\mathbf{Q}_{\text{new}} = c_2 \min\left(c_1\mathbf{I}, \left(\sum_{i=1}^{N} \frac{\mathbf{x}_i\mathbf{x}_i^T}{\|\mathbf{Q}_{\text{old}}\mathbf{x}_i\|}\right)^{-1}\right);$$

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

Regularization in algorithm

- The update formula fails when $\|\mathbf{Q}_{old}\mathbf{x}_i\| = 0$
- Use IRLS with regularized weight:

$$\mathbf{Q}_{\text{new}} = \arg\min_{\mathbf{Q}} \sum_{i=1}^{N} \frac{\|\mathbf{Q}\mathbf{x}_i\|^2}{\max(\|\mathbf{Q}_{\text{old}}\mathbf{x}_i\|, \delta)}.$$

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

$$\mathbf{Q}_{\text{new}} = \left(\sum_{i=1}^{N} \frac{\mathbf{x}_i\mathbf{x}_i^T}{\max(\|\mathbf{Q}_{\text{old}}\mathbf{x}_i\|, \delta)}\right)^{-1} \Big/ \operatorname{tr}\left(\left(\sum_{i=1}^{N} \frac{\mathbf{x}_i\mathbf{x}_i^T}{\max(\|\mathbf{Q}_{\text{old}}\mathbf{x}_i\|, \delta)}\right)^{-1}\right)$$

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

Convergence of algorithm

- The algorithm converges to the minimizer of the objective function: $\|\mathbf{Q}_k - \mathbf{Q}^*\| \to 0$.
- The proof of the convergence depends on the assumption that

$$\{\mathcal{X} \cap \mathrm{L}_1\} \cup \{\mathcal{X} \cap \mathrm{L}_2\} \neq \mathcal{X}, \text{ for all } (D-1)\text{-dimensional subspaces } \mathrm{L}_1, \mathrm{L}_2$$

- This condition holds when $N \geq 2D - 1$ and $\{\mathbf{x}_i\}_{i=1}^N$ 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 downsampled to 20×20
- 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