Robust Principal Component Analysis (PCA)

(or Rank-Sparsity Structure)

Recall PCA : $X \in \mathbb{R}^{p \times n}$ centered data matrix

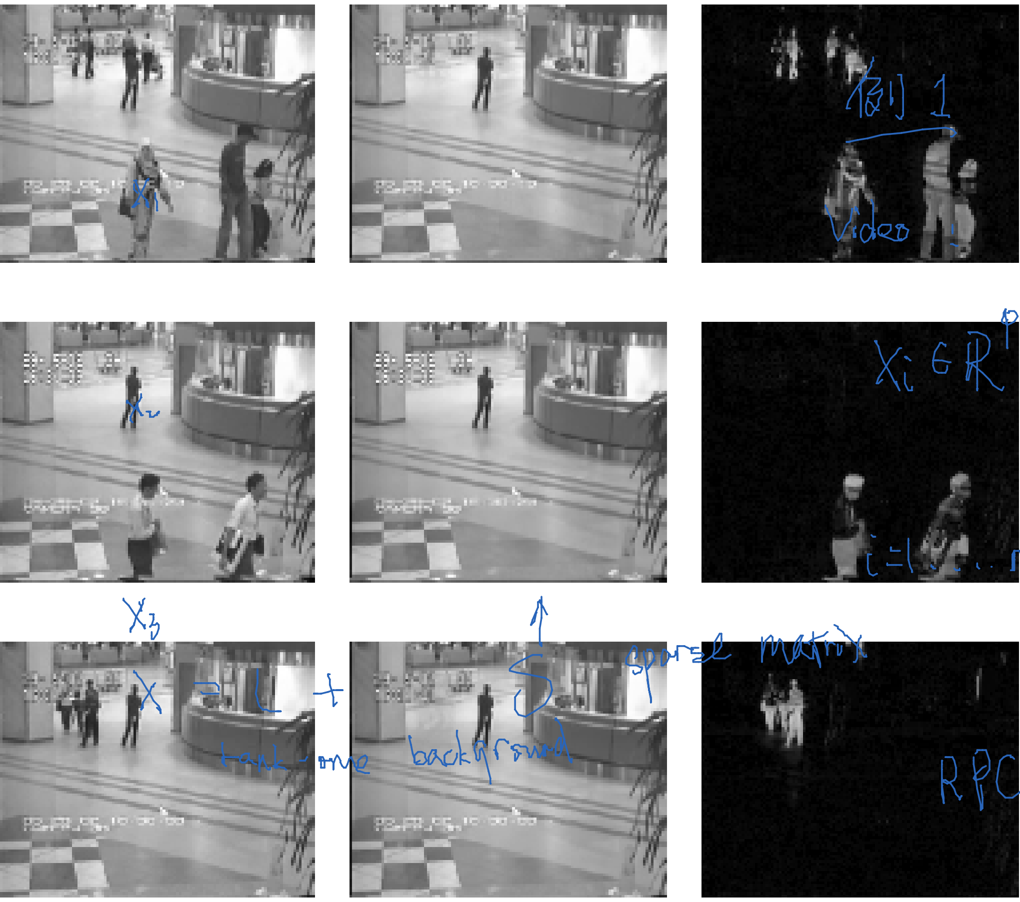
$$\min \|X - L\|_{*}$$ "2" Schatten-p norm

$$\text{s.t. } \operatorname{rank}(L) \le k$$ $X = U\Sigma V^T$

View $X = \underset{\text{Low rank}}{L} + \underset{\text{noise/perturbation}}{E}$ Gaussian $\|E\|_F$ a.

"SVD approach for PCA"

新 $X = \underset{\text{lowrank}}{L} + \underset{\text{sparse}}{S}$ $\#\{S_{ij} \neq 0\}$ $\ll pn$



Surveillance Video

$X^{p \times n}$

image

p # pixels

frame time

RPCA.

Ma. Candes

Wright.

例2 rank-1 model

$$Y_i = a_i u + \varepsilon_i \qquad u \in \mathbb{R}^p, \ \|u\|_2 = 1$$

$$Y = [Y_1 \cdots Y_n] \qquad Y_i \sim N(0, \Sigma) \qquad a_i \sim N(0, \sigma_x^2)$$

$$\varepsilon_i \sim N(0, \sigma_\varepsilon^2)$$

$$\Sigma = \underset{L}{\sigma_x^2 u u^T} + \underset{S}{\sigma_\varepsilon^2 I_p}$$

例3 Gaussian Graphical Model

$X_1, \cdots, X_p \sim N(0, \Sigma)$

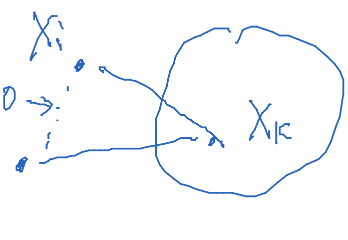
$X_i \perp X_j$ conditionally independent given other variables

$(X_i \perp X_j \mid X_{-\{i,j\}})$

$$\Longleftrightarrow (\Sigma^{-1})_{ij} = 0$$

graph

$(\Sigma^{-1})_{ij} = 0 \rightarrow$



$$\{X_i : i = 1, \cdots p\} = O \cup H$$

$$O = \{X_1 \cdots X_O\} \qquad H = \{X_1 \cdots X_H\}$$

$$\Sigma = \begin{bmatrix} \Sigma_{OO} & \Sigma_{OH} \\ \Sigma_{HO} & \Sigma_{HH} \end{bmatrix} \Rightarrow \Omega = \Sigma^{-1} = \begin{bmatrix} \Omega_{OO} & \Omega_{OH} \\ \Omega_{HO} & \Omega_{HH} \end{bmatrix}$$

$$\Sigma_{OO}^{-1} = \Omega_{OO} + \Omega_{OH} \Omega_{HH}^{-1} \Omega_{HO} = L$$

Ω_{OO}: sparse, conditional indep.

Ω_{HH}^{-1}: invisible hand

较少 H 解释 O

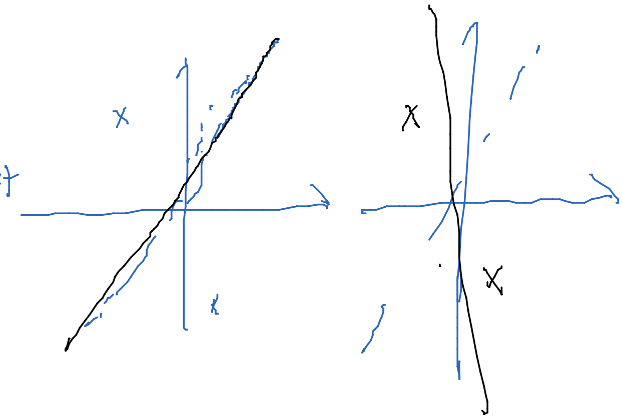
low rank

$X = L + S$?

PCA $\min \|X-L\|_F$

s.t. $\text{rank}(L) \le k$

robust



sensitive to outlier

RPCA $\min \|X-L\|_0 = \#\{(i,j) : S_{ij} \neq 0, \ S+L=X\}$

s.t. $\text{rank}(L) \le k$ NP-hard

Convex Relaxation

$X = L + S$

1° $\|S\|_0 = \#\{S_{ij} \neq 0\} \Longrightarrow \|S\|_1 = \sum_{i,j} |S_{ij}|$

2° $\text{rank}(L) = \#\{\sigma_i(L) \neq 0\} \Longrightarrow \|L\|_* = \sum_{i,j} \sigma_i(L)$

(P) $\min \|L\|_* + \lambda \|S\|_1$ Convex Prog.

s.t. $X = L + S$

$\|L\|_* = \min \frac{1}{2}(\text{trace}(W_1) + \text{trace}(W_2))$

SDP. s.t. $\begin{bmatrix} W_1 & L \\ L^T & W_2 \end{bmatrix} \succeq 0$ p.s.d.

Linear Matrix Inequality with S.D. constraints

LP

"CVX"

$$\min \ \frac{1}{2}\mathrm{tr}(\underline{\underline{W_1}}) + \frac{1}{2}\mathrm{tr}(\underline{\underline{W_2}}) + \lambda \|S\|_1$$

$$\text{s.t.} \begin{bmatrix} W_1 & L \\ L^T & W_2 \end{bmatrix} \succeq 0$$

数据 X

$$\text{"} X = \underline{\underline{L}} + \underline{\underline{S}} \text{"} \qquad (\text{float pt.})$$

SDP $\begin{cases} \text{linear inequality / equality} \\ \text{S.D. constraint} \end{cases}$

Basics of SDP vs. LP

S. Boyd. "Convex Optimization"

(LP) : $x \in \mathbb{R}^n$, $c \in \mathbb{R}^n$

$\min \ c^T x$ ⇐ $\langle c, x \rangle$ vector inner product

s.t. $Ax = b$ Standard form

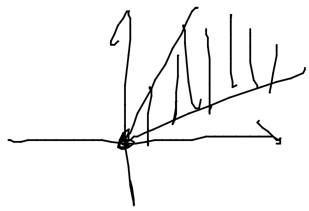
$x \geq 0$

(SDP) $X \in \mathbb{R}^{n\times n}$, $C \in \mathbb{R}^{n \times n}$

$\min \ C \bullet X = \sum_{i,j} C_{ij} X_{ij}$ Hadamard inn. prod. Frobenius

s.t. $A_i \bullet X = b_i$ $i = 1, \dots, m$

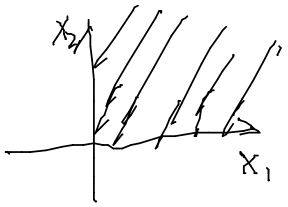
$X \succeq 0$ p.s.d.



"Conic Programming" $x \in C$

C is cone ⇔ $x \in C \to \alpha x \in C$ $\alpha > 0$

$x \geq 0$



⇐ $X^{n\times n} \succeq 0$ p.s.d.

Dual.

LD $\max\limits_{\mu \geq 0, y} \min\limits_{x} L(x; y, \mu) = c^T x + y^T(b - Ax) - \mu^T x$

$\mu \geq 0$

$\frac{\partial L}{\partial x} = c - A^T y - \mu = 0$ $\mu = c - A^T y \geq 0$

"LD" : $\max\limits_{y} \ b^T y$

s.t. $\mu = c - A^T y \geq 0$

"$n < 100$"

CVX

"SDD" $\max \ b^T y$

s.t. $S = C - \sum_{i=1}^{m} A_i y_i \succeq 0$ p.s.d.

Given $X = L_0 + S_0$

L_0 ? S_0 ?

SDP $\hat{L} = L_0$, $\hat{S} = S_0$ Exact Recovery Theory ?

RPCA Candes, Wright, Parrilo.

$X = L_0 + S_0$

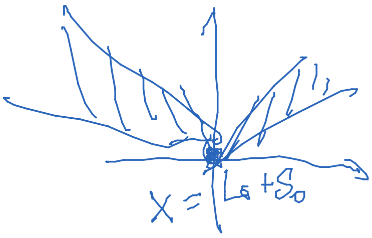
① L_0 low rank 不能 sparse

② S_0 sparse " low-rank

} identiability

Idea: Low-rank $L_0 = U\Sigma_k V^T$ cone

Sparse $S_0 = \#\{S_{ij} \neq 0\}$



" Geometry "

" Incoherence " (Candes-Recht '09)

$\exists \mu \geq 1$ s.t. $\forall e_i = (0\ 0\ 0\ \underset{i\text{th comp.}}{1}\ 0\ 0)^T \in \mathbb{R}^n$

① $\|U^T e_i\|^2 \leq \frac{\mu r}{n}$ $L_0 = U\Sigma V^T$ $\operatorname{rank}(L_0) = r$

② $\|V^T e_i\|^2 \leq \frac{\mu r}{n}$

③ $|UV^T|_{ij}^2 \leq \frac{\mu r}{n^2}$

Thm (1) L_0 $n\times n$. $\operatorname{rank}(L_0) \leq \rho_r n \mu^{-1} (\lg n)^{-2}$

(2) S_0 uniform sparse $\|S_0\|_0 \leq \rho_s n^2$

$SDP(\lambda = 1/\sqrt{n})$ exactly L_0, S_0 ($\hat{L} = L_0$, $\hat{S} = S_0$) with prob $\geq 1 - O(n^{-10})$

Partial observation.

$$\min \ \|L\|_* + \lambda \|S\|_1$$
$$\text{s.t.} \quad P_\Omega(X) = P_\Omega(L) + P_\Omega(S) \qquad \Omega \text{ subset of } \{(i,j)\}$$

<u>Thm</u> (1) $L_0^{n\times n}$ $\quad rk(L_0) \le \rho_r n \mu^{-1} (\log n)^{-2}$

(2) Ω uniform $\quad |\Omega| = 0.1 n^2$ ← observation sample $O(n^2)$

(3) corrupted observation with $\tau \le \tau_s$

SDP ($\lambda = 1/\sqrt{0.1n}$) w.h.p. recover L_0..

Rectangular matrix. $X^{n_1 \times n_2}$, $\lambda = 1/\sqrt{\max(n_1, n_2)}$

Candes-Tao '2010. $\quad |\Omega| \ge \mu n r \log^\alpha n \quad$ almost linear obs.

$\alpha \le 6$.

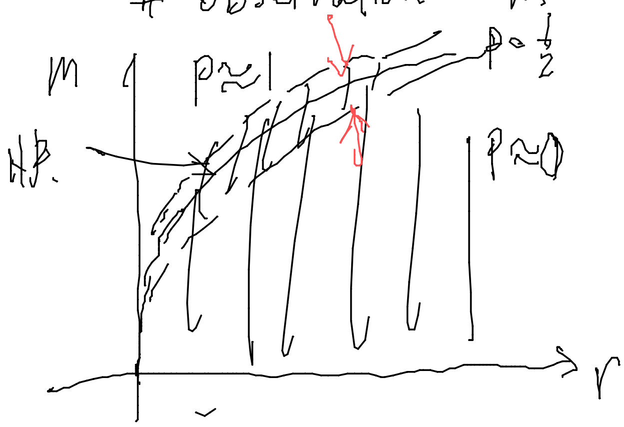
Gross '2011 $\quad |\Omega| \ge \mu n r \log^2 n. \quad$ nearly optimal

---

"Open": Phase-transition

matrix completion L_0 low-rank

\# observation $= m$ $\qquad$ rank$(L_0) = r$



n = fixed

prob(exact recovery) (random)

$\begin{cases} m = f(r) & \text{刻画?} \\ \text{band} & \text{宽?} \end{cases}$

Tropp '2013

Conic Geometry

Sparse PCA : SDP

Recall PCA

$$\max \cancel{x}^T \Sigma x \quad \text{s.t. } \|x\|_2 = 1$$

Σ : covariance matrix

nonconvex problem

* eigen-decomp.

SDP approach

$\text{trace } x^T \Sigma x = \text{trace } (\Sigma \underline{x x^T})$

$X = x x^T$ rank-one

SDP:

$$\max \operatorname{tr}(\Sigma X) \quad \leftarrow \text{线性}$$

$$\text{s.t. } \operatorname{trace}(X) = 1 \quad \leftarrow \text{线性}$$

$$X \succeq 0$$

relaxation

PCA sparse. $\Sigma = \sigma_x^2 \underline{u u^T} + \sigma_\Sigma^2 I_p$ u_i sparse

"eigenvector" $\longrightarrow$ sparse

SPCA

"SDP"

$$\max \operatorname{tr}(\Sigma X) - \lambda \|X\|_1$$

$$\text{s.t. } \operatorname{tr}(X) = 1$$

$$X \succeq 0$$

"linear"