Robust Principal Component Analysis (PCA)

(or Rank-Sparsity Structure)

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

$$\min \|X - L\|_F \qquad \text{"2" Schatten-}p\text{ norm}$$

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

View $X = L + E$

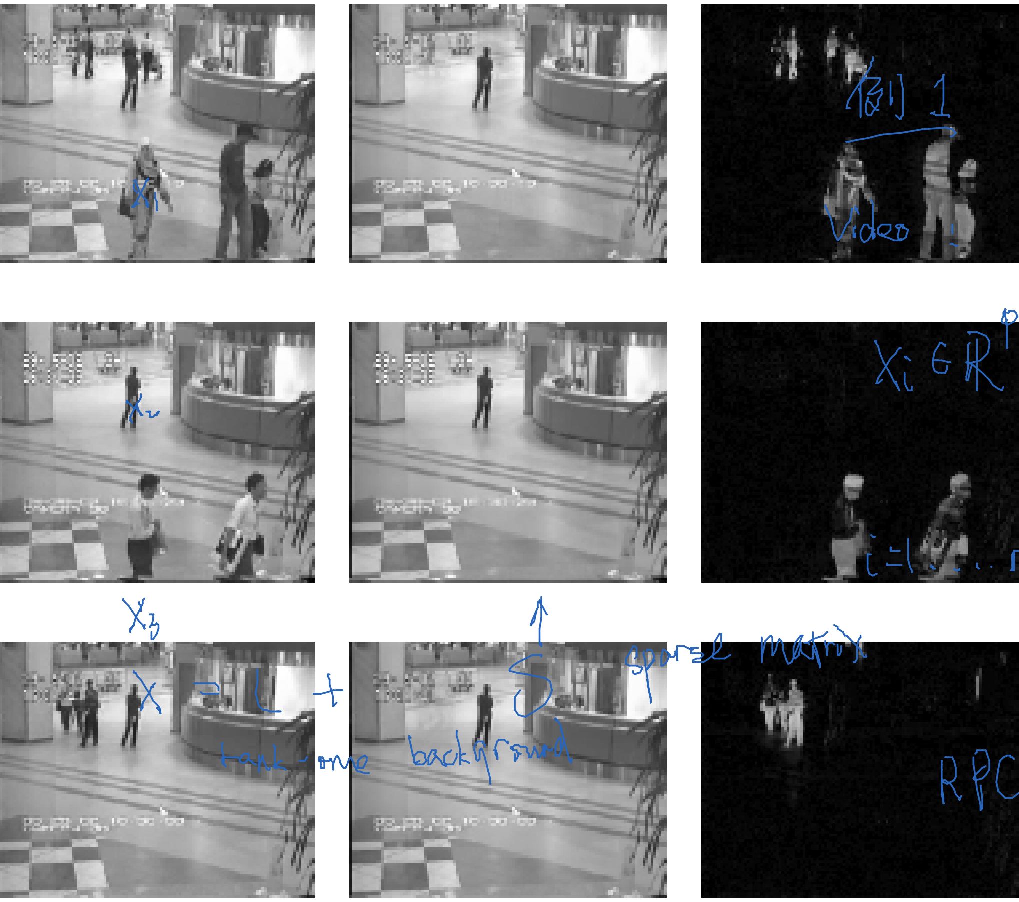
L: Low rank, E: noise/perturbation Gaussian $\|E\|_F$ sm.

" SVD approach for PCA "

新 $X = L + S$

L: lowrank, S: sparse

$$\#\{S_{ij} \neq 0\} \ll pn$$



例 1

Surveillance Video

$X^{p \times n}$

Video

$X_i \in \mathbb{R}^p$ image

p # pixels

$i = 1, \ldots, n$ frame time

X_1, X_2, X_3

$X = L + S$

rank-one background

sparse matrix

RPCA.

Ma. Candes

Wright.

例2 rank-1 model

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

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

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

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

例3 Gaussian Graphical Model

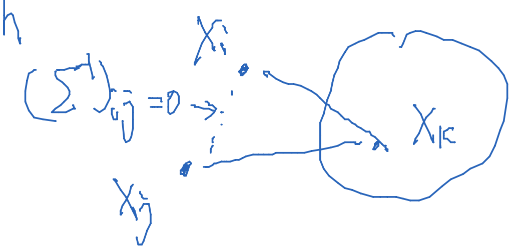
$$X_1, \dots, X_p \sim N(0, \Sigma)$$

$X_i \perp X_j$ conditionally independent given other variables

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

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

graph



$$\{X_i : i = 1, \cdots p\} = \underbrace{O}_{\{X_1 \cdots X_O\}} \cup \underbrace{H}_{\{X_1 \cdots X_H\}}$$

$$\Sigma = \left[\begin{array}{c|c} \Sigma_{OO} & \Sigma_{OH} \\ \hline \Sigma_{HO} & \Sigma_{HH} \end{array}\right] \Rightarrow \Omega = \Sigma^{-1} = \left[\begin{array}{c|c} \Omega_{OO} & \Omega_{OH} \\ \hline \Omega_{HO} & \Omega_{HH} \end{array}\right]$$

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

invisible hand

sparse

conditional indep.

较少 H 解释 O

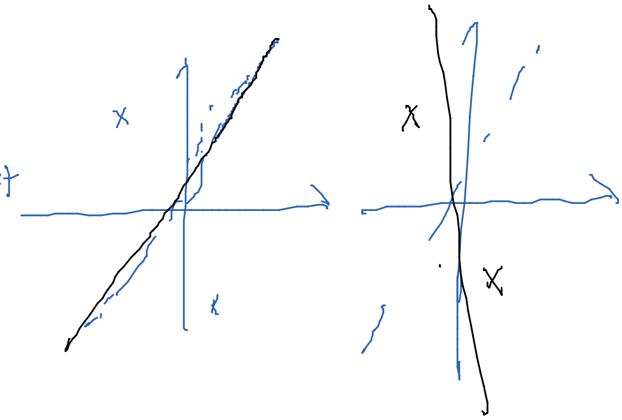
low rank

$X = L + S$?

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

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

robust



sensitive to outlier

RPCA

$$\min \|X-L\|_0 = \#\{(i,j) : S_{ij} \neq 0, \; S+L=X\}$$
$$\text{s.t. } \text{rank}(L) \le k$$

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) $\quad \min \|L\|_* + \lambda \|S\|_1$

$\quad$ s.t. $X = L + S$

Convex Prog.

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

SDP

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

Linear Matrix Inequality with S.D. constraints

LP

"CVX"

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

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

数据 X

"$X = \underline{\underline{L}} + \underline{\underline{S}}$" (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) : $x \in \mathbb{R}^n$, $c \in \mathbb{R}^n$

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

s.t. $Ax = b$ standard form

$x \geq 0$

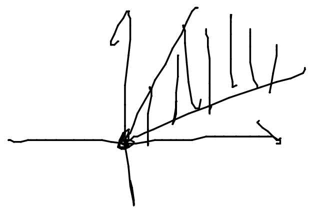
(SDP) $X \in \mathbb{R}^{n \times n}$, $C \in \mathbb{R}^{n \times n}$

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

Frobenius

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

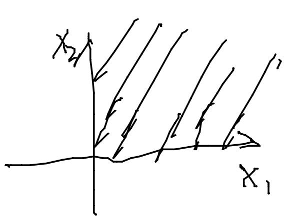
$X \succeq 0$ p.s.d.



"Conic Programming" $x \in C$

C is cone $\Leftrightarrow$ $x \in C \rightarrow \alpha x \in C$ $\alpha > 0$

$x \geq 0$



$\Leftarrow X^{n \times n} \succeq 0$ p.s.d.

Dual.

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

$\mu \geq 0$

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

"LD": $\max_{y} b^T y$

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.

$X = L_0 + S_0$

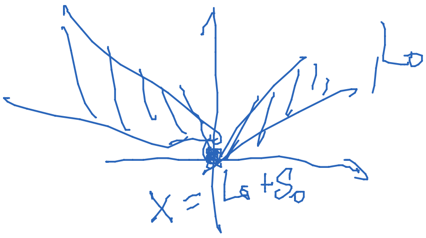
① L_0 low rank 不能 sparse

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

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

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

Thm (1) L_0 $n\times n$. 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. $1 - O(n^{-10})$

Partial observation.

$\min \ \|L\|_* + \lambda\|S\|_1$

s.t. $P_\Omega(X) = P_\Omega(L) + P_\Omega(S)$ $\quad\Omega$ 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 $\quad |\Omega| \le 0.1 n^2 \leftarrow$ observation sample $O(n^2)$

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

SDP $(\lambda = 1/\sqrt{0.1 n})$ 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.

$\alpha \le 6$.

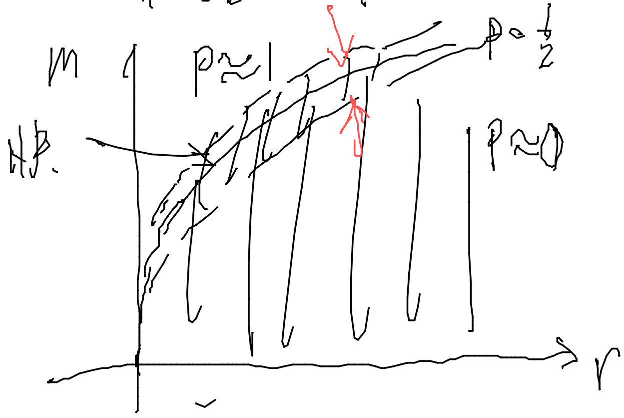
Gross '2011 $\quad |\Omega| \ge \mu n r \log^2 n$. $\quad$ nearly optimal

---

"Open": Phase-transition

matrix completion $\quad L_0$ low-rank

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



n = fixed

prob(exact recovery | Ω random)

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

Tropp '2013

Conic Geometry

Sparse PCA : SDP

Recall PCA

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

Σ : covariance matrix

nonconvex problem

* eigen-decomp.

SDP approach

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

$X = x x^T$ rank-one

SDP:

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

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

$$X \succeq 0$$

relaxation

PCA sparse.

$$\Sigma = \sigma_x^2 u u^T + \sigma_\Sigma^2 I_p$$

u; sparse

"eigenvector" $\longrightarrow$ sparse

SPCA

"SDP"

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

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

$$X \succeq 0$$

"linear"

MDS with Uncertainty : Sensor Network Localization (SNL)

Recall Classical MDS

Given d_{ij} pair distance $\forall (i,j)$ Complete

? $y_i \in \mathbb{R}^k$ s.t.

$$\min_{Y \in \mathbb{R}^{n\times k}} \sum_{i,j} \left(\|y_i - y_j\|^2 - d_{ij}^2 \right)^2$$

$$\Leftrightarrow \min_{Y} \| Y Y^T - B \|^2 \qquad B = -\frac{1}{2} H D H^T$$

$$D = [d_{ij}^2]$$

$$\Leftrightarrow \text{eigen-decomp}(B)$$

新的特点：SNL

$G = (V, E)$

V: sensor

① $(i,j) \in E$ iff d_{ij}

incomplete

② noise: $\tilde{d}_{ij} = d_{ij} + \epsilon_{ij}$ noise

③ anchor point $x_i = a_i$

partial

目标：

$$\|y_i - y_j\|^2 = d_{ij}^2 \qquad (i,j) \in E$$

$$\|a_i - y_j\|^2 = d_{ij}^2$$

? $\min_{Y} \sum_{(i,j)\in E} \left(\|y_i - y_j\|^2 - d_{ij}^2 \right)^2$ { gradient method / nonlinear opt.

不能 Eigen-decomp

练习: $\|y_i - y_j\|^2 = d_{ij}^2$ Quadratic equality

SD Relax ⇒ "Linear"

$y_i \in \mathbb{R}^k$

$$\|y_i - y_j\|^2 = (y_i - y_j)^T (y_i - y_j)$$

$Y = [y_1 \cdots y_n]^{k \times n}$

$$= (e_i - e_j)^T [Y^T Y]^{n \times n} (e_i - e_j)$$

$e_i = (0 \cdots 0 \overset{i\text{-th}}{1} 0 \cdots 0)$

$y_i = Y \cdot e_i$

$y_i - y_j = Y(e_i - e_j)$

$$= (e_i - e_j)(e_i - e_j)^T \bullet [Y^T Y]$$

$X = Y^T Y \succeq 0$

$$= (e_i - e_j)(e_i - e_j)^T \bullet X$$

LMI. SD

$X \succeq 0$

Relaxation

$$X = Y^T Y \Longrightarrow X \succeq Y^T Y \quad (X - Y^T Y \succeq 0)$$

$$\Longleftrightarrow \begin{bmatrix} I_k & Y \\ Y^T & X \end{bmatrix} \succeq 0, \quad X \succeq 0$$

Linear Matrix Inequality

Lemma $\|y_i - y_j\|^2 = d_{ij}^2, \quad (i,j) \in E$

SDR

LMI

$$\begin{cases} Z = \begin{bmatrix} I & Y \\ Y^T & X \end{bmatrix} \succeq 0 \\ Z_{1:k,1:k} = I \\ (0; e_i - e_j)(0; e_i - e_j)^T \bullet Z = d_{ij}^2 \quad (i,j) \in E \end{cases}$$

Note:

① $d_{ij}^2(1-\epsilon_{ij}) \le \|y_i - y_j\|^2 \le d_{ij}^2(1+\epsilon_{ij}) \Longrightarrow$ LMI. noise

② anchor point. $y_i = a_i$, $\|a_i - y_j\|^2 = d_{ij}^2 \Rightarrow (a_i; e_j)(a_i; e_j)^T \bullet Z = \tilde{d}_{ij}^2$

SDP approach → MDS

Matlab

Protein 3-D structure Reconstruction

CMDS Schoenberg

SDP-MDS Exact Recovery ? Yinyu Ye group.

P. Biswas.

A. So.

Recall SDP

(SDP) $\min\ C \bullet X$ $\quad C, X \in R^{n\times n}$

s.t. $A_i \bullet X = b_i \quad i=1,\dots,m \quad b = \begin{bmatrix} b_1 \\ \vdots \\ b_m \end{bmatrix}$

$X \succeq 0$

(SDD) $\max\ -b^T y$ $\quad y \in R^m,\ S \in R^{n\times n}$

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

$$F_p = \{ X \succeq 0 : A_i X = b_i \}$$

$$F_D = \{ (y, S) : S = C - \sum A_i y_i \succeq 0 \}$$

primal obj: $C \bullet X$ dual obj $b^T y$

Weak Duality $F_p \neq \emptyset,\ F_D \neq \emptyset$

$$\Rightarrow \underbrace{C \bullet X - b^T y}_{\text{duality gap}} \geq 0 \quad \forall X \in F_p,\ \forall (y, S) \in F_d$$

Strong Duality of SDP

(1) $F_p \neq \phi$, $F_d \neq \phi$

(2) F_p or F_d has an interior solution

$\Rightarrow$ $X^* \in F_p$. $(y^*, S^*) \in F_D$ is optimal soln.

iff $C \bullet X^* = b^T y^*$ duality gap is zero

$X^* S^* = 0$. Complementary

存在内点 Soln

$\underbrace{X^*, (y^*, S^*)}_{\text{"witness"}}$ $X^* \cdot S^* = 0$ opt.

primal-dual pair

(*) $\mathrm{rank}(X^*) + \mathrm{rank}(S^*) \leq n$

SDP-MDS.

$$Z = \begin{bmatrix} I_k & Y \\ Y^T & X \end{bmatrix} \succeq 0$$

假设 $d_{ij} = \|x_i - x_j\|$ $x_i \in R^k$

$$\exists\ Z^* = \begin{bmatrix} I_k & Y^* \\ Y^{*T} & X^* \end{bmatrix} \succeq 0$$ interior point soln

$\left.\begin{array}{l} \mathrm{rank}(Z^*) \geq k \\ \mathrm{rank}(X^*) + \mathrm{rank}(S^*) \leq n \end{array}\right\} \Rightarrow \mathrm{rank}(Z^*) + \mathrm{rank}(S^*) \leq k+n$

$\mathrm{rank}(S^*) \leq n$ $\Big\} \Rightarrow \mathrm{rank}(Z^*) \geq k$

$X^* = Y^T Y \quad Y \in R^{k\times n}$

$\Rightarrow \quad \mathrm{rank}(Z^*) = k \qquad (\geq k) \qquad$ minimal rank

代数 $\quad \mathrm{rank}(S^*) = n \qquad (\leq n) \qquad$ maximal rank.

几何："Universal Rigidity (UR)"

there is a unique embedding $y_i \in R^k \hookrightarrow R^l \quad l \geq k$

$(\underbrace{y_i}_{k}, \underbrace{0, \ldots, 0}_{l-k}) \quad$ s.t. $\quad \|y_i - y_j\| = d_{ij}, \ (i,j) \in E$

minimal dim embedding $k : R^k$

Schoenberg '1938 : G is complete $\xRightarrow{\text{spectrum}}$ UR.

G is incomplete

[So - Ye '2007] $\quad$ G general.

UR $\Leftrightarrow$ SDP maximal rank sol'n

$\mathrm{rank}(S^*) = n. \ \mathrm{rank}(Z^*) = k$

<u>Theorem</u> Equivalent statements

- (几何) G is UR or has a unique embedding in R^k
- (代数) SDP has a max-rank feasible sol'n $\mathrm{rank}(Z^*) = k$
  $(\mathrm{rank}(S^*) = n)$
- $X^* = Y^T Y \qquad$ or $\mathrm{trace}(X - Y^T Y) = 0$
  eigen decomp of X^*.

<u>UR</u>'s. polynomial $(n, k, \log(\frac{1}{\varepsilon}))$

Ye, ICCM '2010 ; Fields '2011

Maximum Variance Unfolding (Manifold Learning)

$$X = Y^T Y \Rightarrow X \succeq Y^T Y$$

MVU

$\underline{\underline{K}} = Y^T Y$ 度量 $d_{ij}^2 = K_{ii} + K_{jj} - 2K_{ij}$

$K_{ij} = \langle Y_i, Y_j \rangle$

"SDP"

$$\max \ \mathrm{trace}(K)$$
$$\text{s.t.} \quad K_{ii} + K_{jj} - 2K_{ij} = d_{ij}^2$$
$$\sum_j K_{ij} = 0$$
$$K \succeq 0$$

$K \Rightarrow Y^T Y$

Why max tr(K) ? 不能 work.

[So-Ye 2007]

unfolding

SDP.

k-manifold

(k+1)-lateration graph

"Unfold manifold"