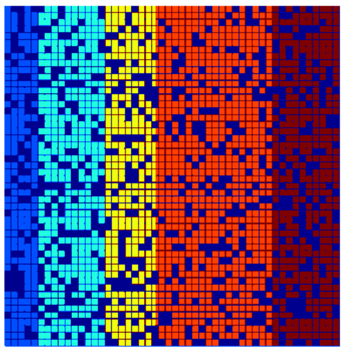
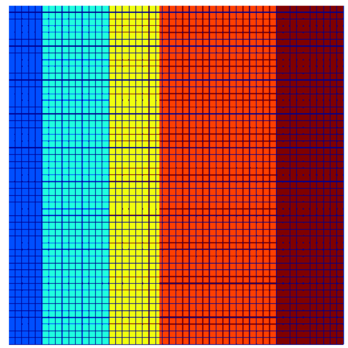
ROBUST PCA

D - observation



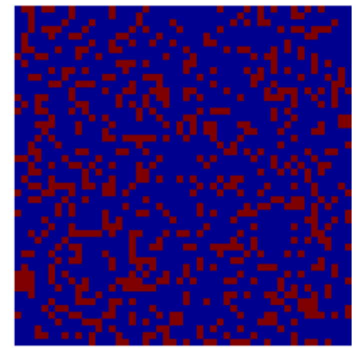
$=$

A_0 – low-rank



$+$

E_0 – sparse

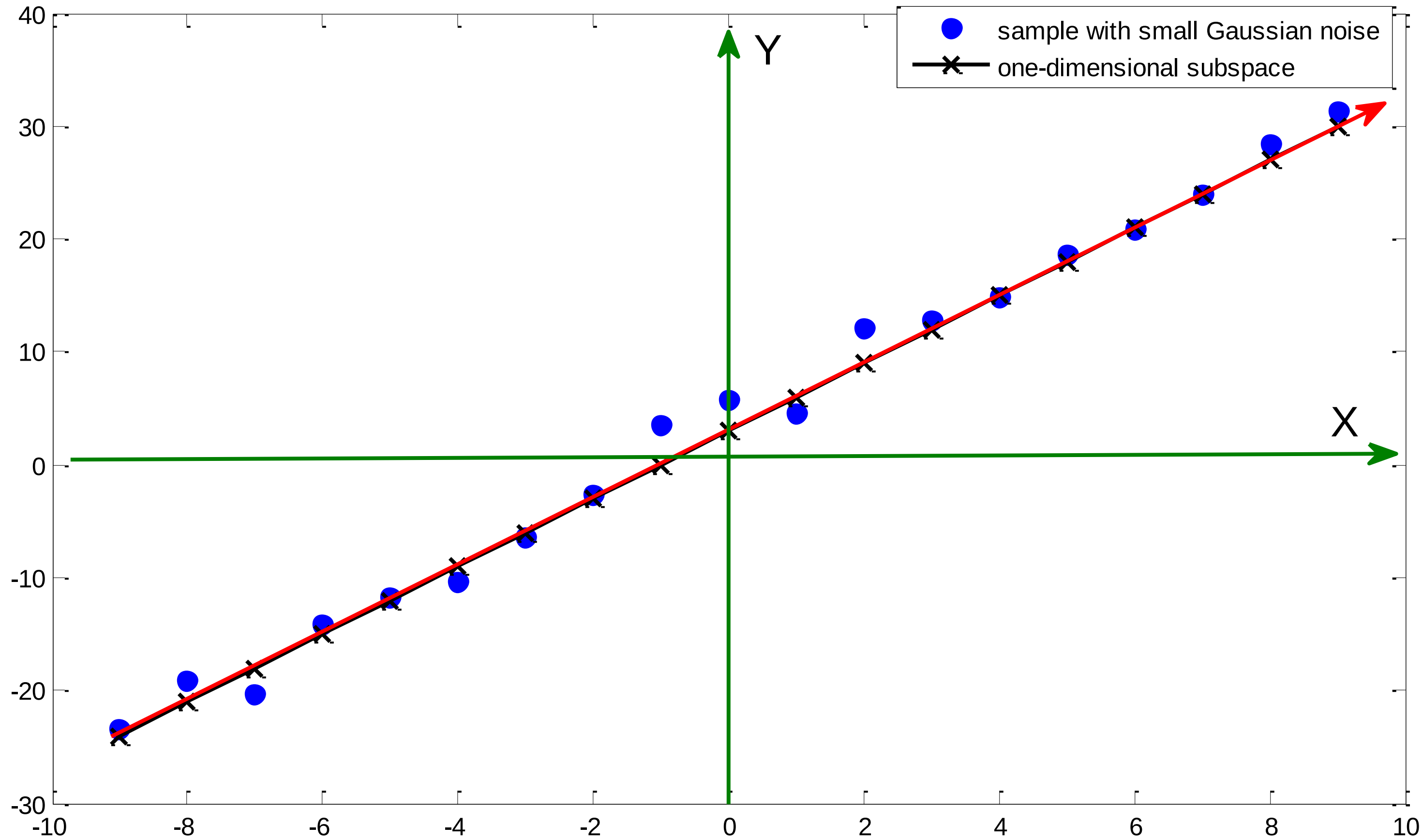


Problem: Given $D = A_0 + E_0$, recover A_0 and E_0

Low-rank component   **Sparse component (gross errors)**

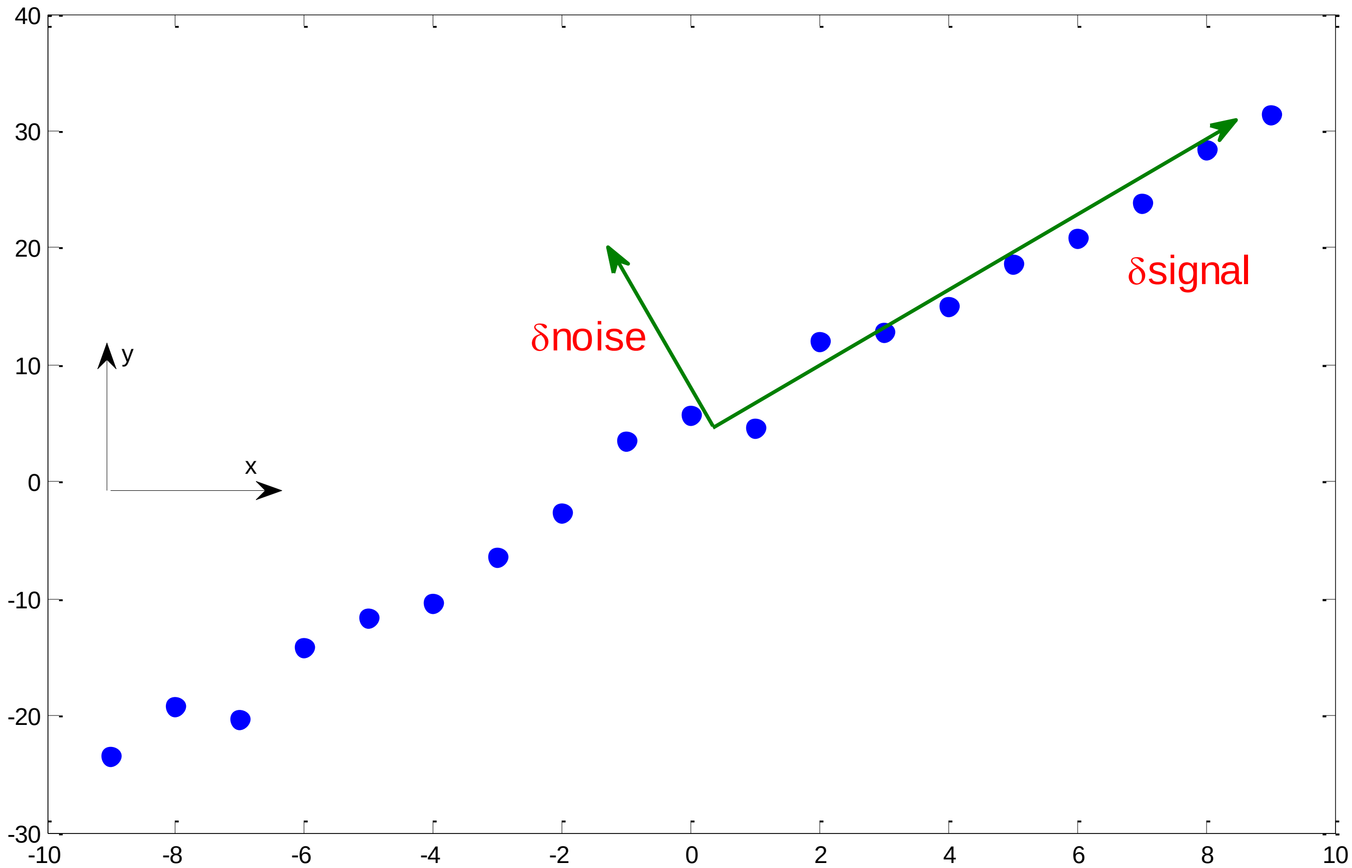
PCA

PCA直观理解


sample with small Gaussian noise
one-dimensional subspace
Y
X
40
30
20
10
0
-10
-20
-30
-10
-8
-6
-4
-2
0
2
4
6
8
10

PCA

PCA直观理解

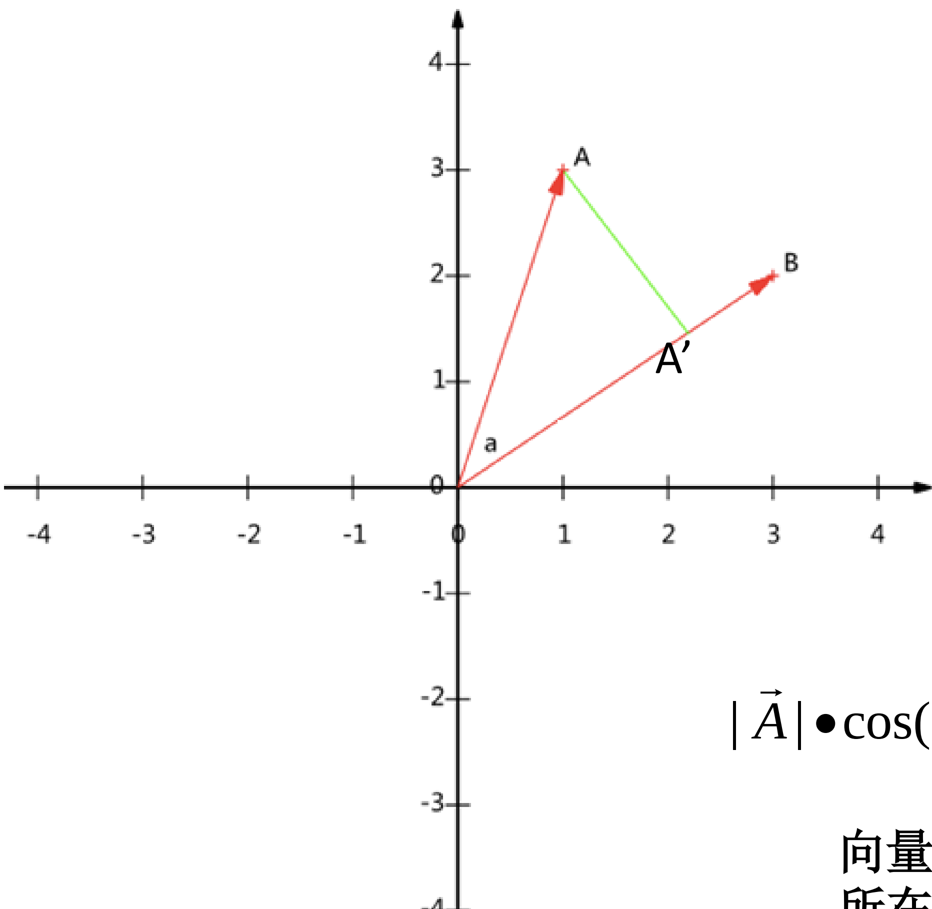


$$\begin{pmatrix} \mathrm{var}(x) & \mathrm{cov}(x,y) \\ \mathrm{cov}(x,y) & \mathrm{var}(y) \end{pmatrix}$$

$$\begin{pmatrix} \mathrm{var}(a) & 0 \\ 0 & \mathrm{var}(b) \end{pmatrix}$$

PCA

矩阵相乘与投影



$$|\vec{A}| \bullet |\vec{B}| \bullet \cos(\alpha)$$
$$= \vec{A} \bullet \vec{B}$$
$$= A_x \bullet B_x + A_y \bullet B_y$$
$$= \begin{pmatrix} A_x & A_y \end{pmatrix} \begin{pmatrix} B_x \\ B_y \end{pmatrix}$$

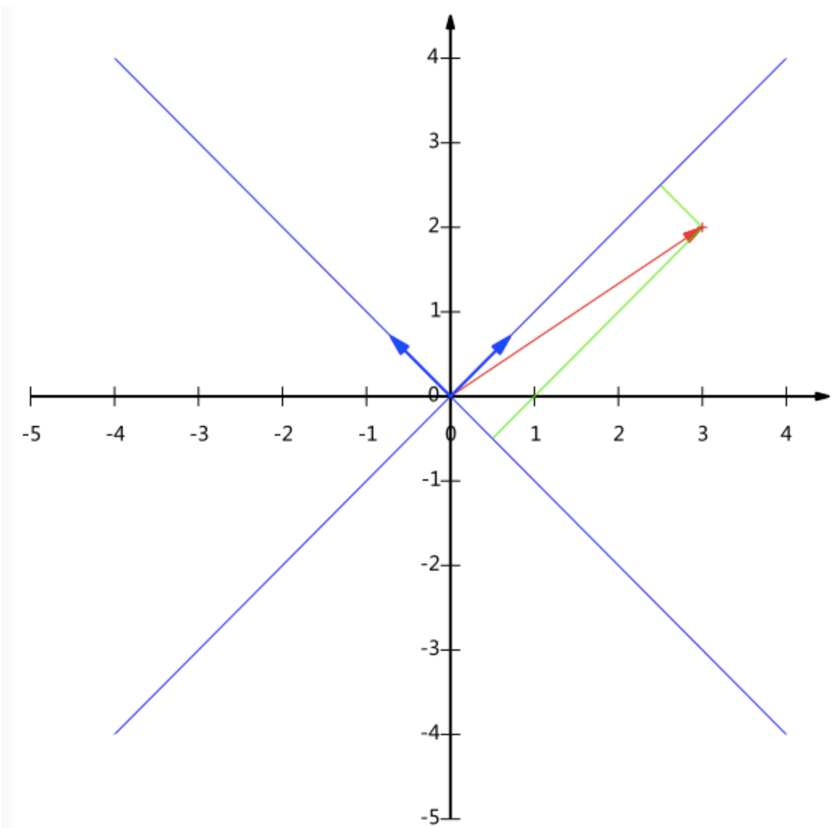
$$if \quad \sqrt{B_x{}^2 + B_y{}^2} = 1$$

$$|\vec{A}| \bullet \cos(\alpha) = \quad |\vec{A}| \bullet |\vec{B}| \bullet \cos(\alpha) = \begin{pmatrix} A_x & A_y \end{pmatrix} \begin{pmatrix} B_x \\ B_y \end{pmatrix}$$

向量B的模为1，则A与B的内积值等于A向B所在直线投影的矢量长度

PCA

矩阵相乘与坐标空间映射



$$\begin{pmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \\ -\frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \end{pmatrix} \begin{pmatrix} 3 \\ 2 \end{pmatrix} = \begin{pmatrix} \frac{5}{\sqrt{2}} \\ -\frac{1}{\sqrt{2}} \end{pmatrix}$$

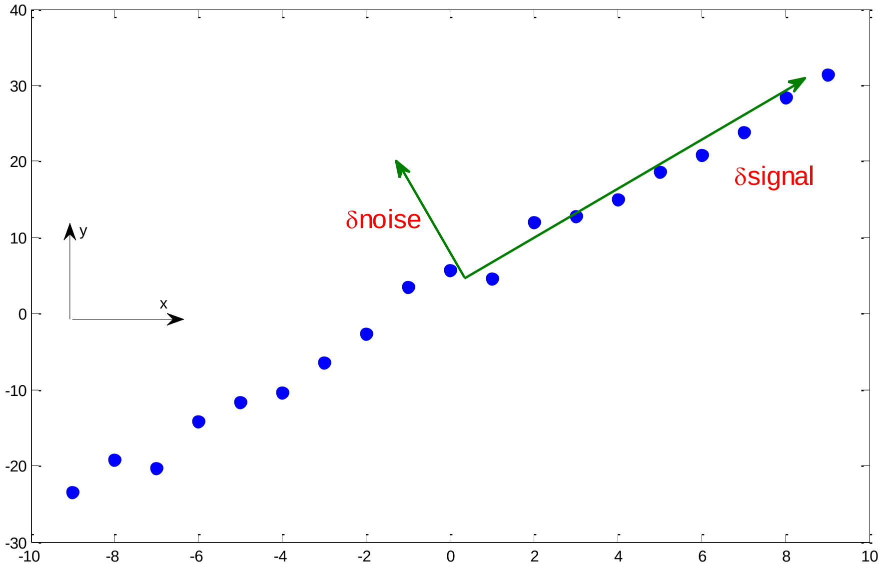
$$= \begin{pmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \\ -\frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \end{pmatrix} \begin{pmatrix} 5 & 0 \\ 0 & -1 \end{pmatrix}$$

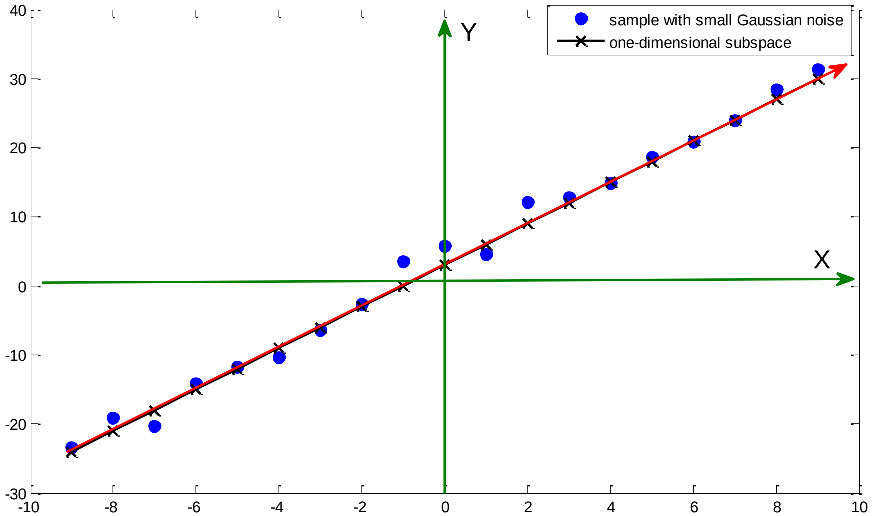
$$\alpha = \left(\frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}} \right), \beta = \left(-\frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}} \right)$$

α β为单位向量，相互正交

$$\begin{pmatrix} \alpha \\ \beta \end{pmatrix} \begin{pmatrix} 3 \\ 2 \end{pmatrix} = \begin{pmatrix} \alpha \\ \beta \end{pmatrix} \begin{pmatrix} 5 & 0 \\ 0 & -1 \end{pmatrix}$$

PCA





$$\begin{pmatrix}\alpha\\ \beta\end{pmatrix}\begin{pmatrix}\mathrm{var(x)} & \mathrm{cov(x,y)}\\ \mathrm{cov(x,y)} & \mathrm{var(y)}\end{pmatrix}$$

$$=\begin{pmatrix}\alpha\\ \beta\end{pmatrix}\begin{pmatrix}\mathrm{var(a)} & 0\\ 0 & \mathrm{var(b)}\end{pmatrix}$$

$$\begin{pmatrix}\mathrm{var(x)} & \mathrm{cov(x,y)}\\ \mathrm{cov(x,y)} & \mathrm{var(y)}\end{pmatrix}\alpha = \mathrm{var(a)}\alpha$$

PCA

SVD

$A_{m\times n}$ $\qquad (A^T A)V_i = \lambda V_i$

$$(AV_i)^T (AV_i) = V_i^T (A^T AV_i) = V_i^T \lambda V_i = \lambda E$$

$$\Rightarrow | AV_i |= \lambda$$

$$U_i = \frac{AV_i}{\sqrt{AV_i}} = \frac{AV_i}{\sqrt{\lambda_i}} \qquad \sigma_i = \sqrt{\lambda_i} \qquad U_i = \frac{AV_i}{\sigma_i}$$

$$U_i^T U_i = (\frac{AV_i}{\sigma_i})^T \frac{AV_i}{\sigma_i} = \frac{V_i^T A^T AV_i}{\sigma_i^2} = \frac{V_i^T \lambda V_i}{\lambda} = E$$

为单位向量，

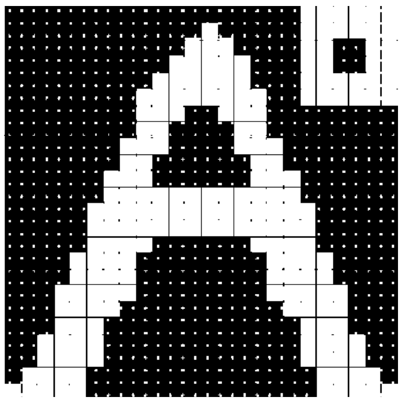
$$U_i^T U_j = (\frac{AV_i}{\sigma_i})^T \frac{AV_j}{\sigma_j} = \frac{V_i^T A^T AV_i}{\sigma_i \sigma_j} = \frac{V_i^T \lambda V_j}{\lambda} = 0$$

相互正交

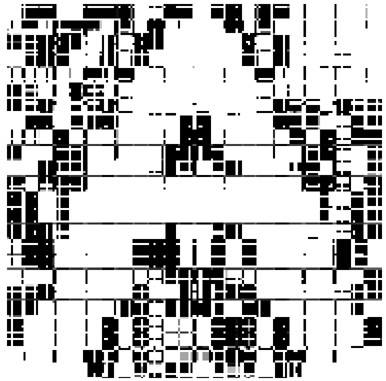
$$U_i = \frac{AV_i}{\sigma_i} \Rightarrow A = U\Sigma V^{-1} = U\Sigma V^T$$

PCA

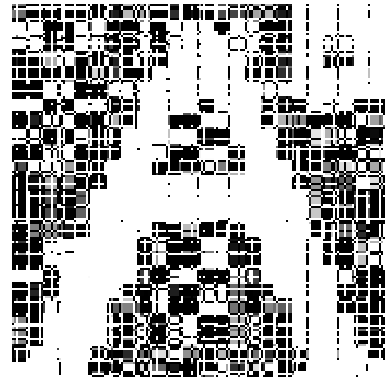
原图



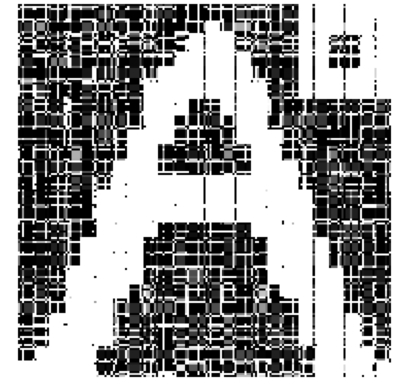
前 10特 征



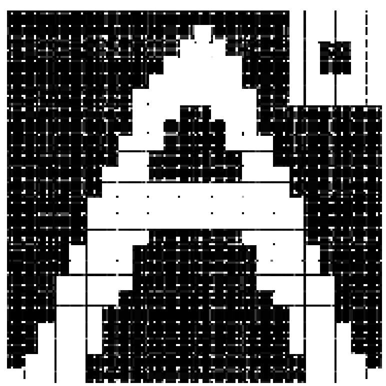
前 20特 征



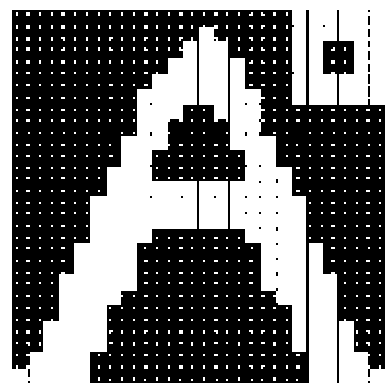
前 50特 征



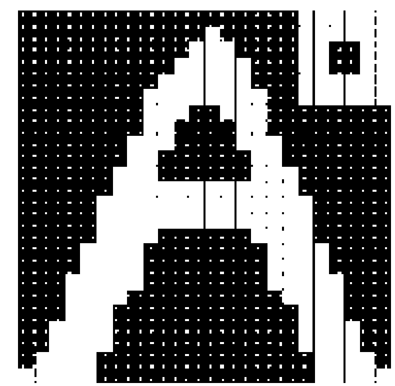
前 100特 征



前 150特 征

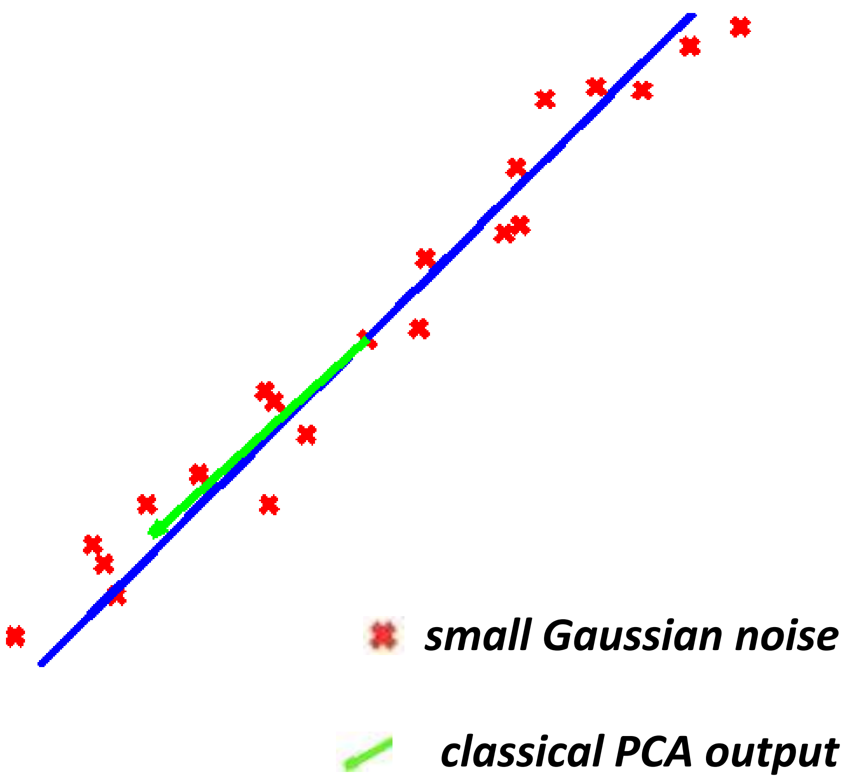


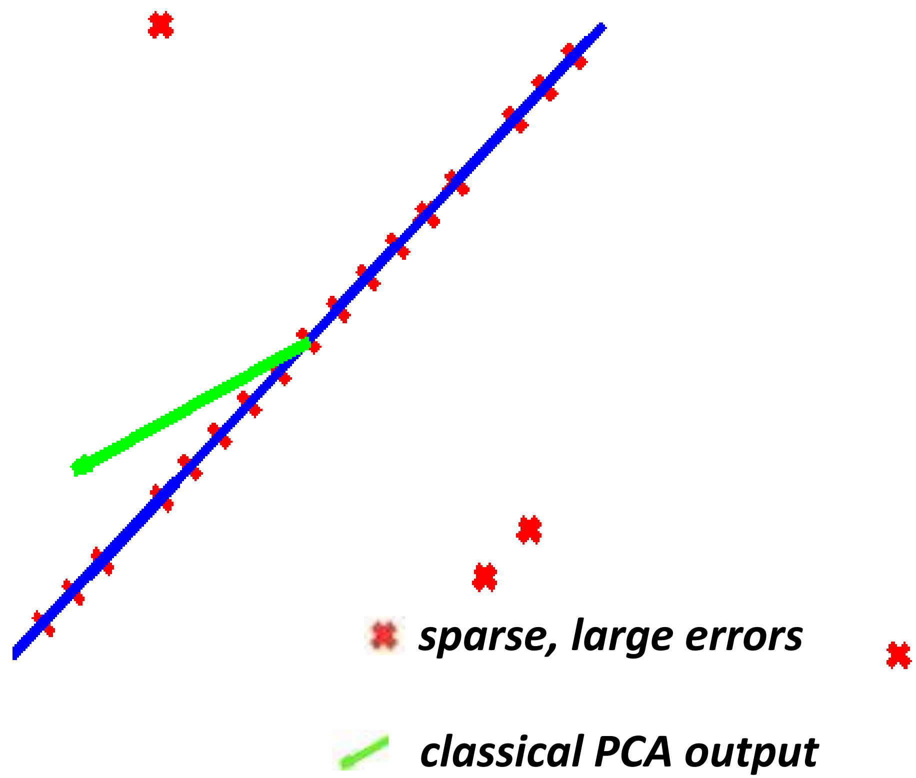
前 200特 征



ROBUST PCA

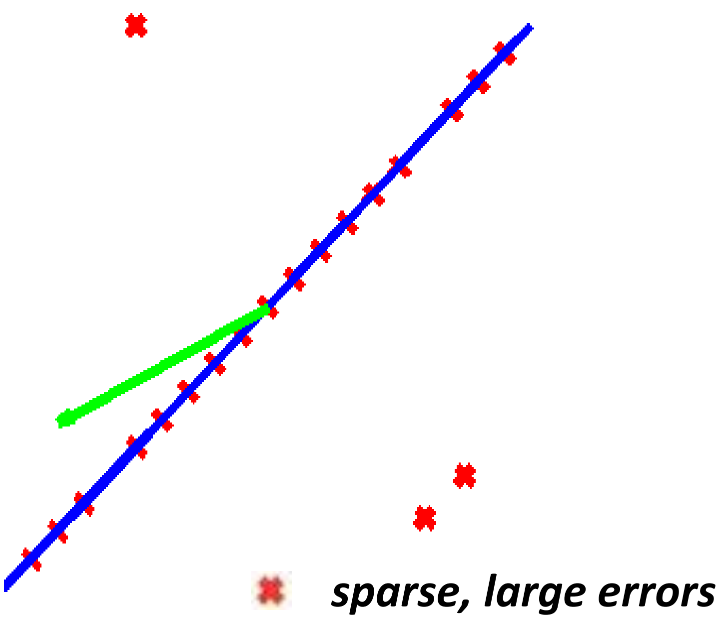
经典PCA局限性





ROBUST PCA

提出 Robust PCA



给出：$D = \underbrace{A_0}_{\text{Low-rank component}} + \underbrace{E_0}_{\text{Sparse component (gross errors)}}$，恢复 A_0 和 E_0

$$\min \ \operatorname{rank}(A) + \gamma\|E\|_0 \quad \text{subj} \quad A + E = D.$$

低秩：$\operatorname{rank}(A) = \#\{\sigma_i(A) \neq 0\}.$

稀疏：$\|E\|_0 = \#\{E_{ij} \neq 0\}.$

！ **Not always** – original problem is NP-hard

ROBUST PCA

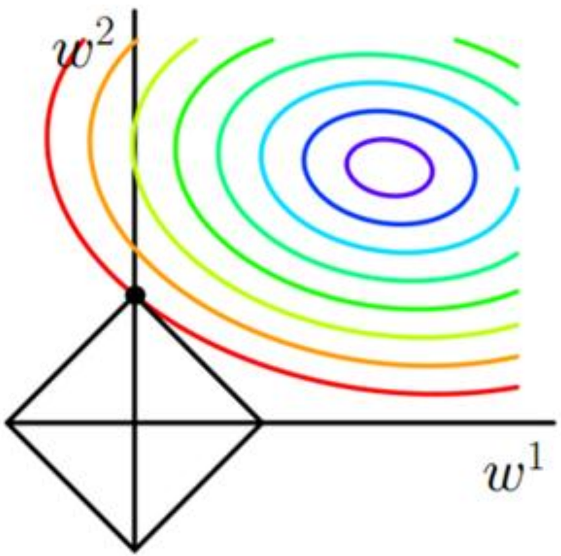
Convex relaxation

0范数(L0)： $\|E\|_0 = \#\{E_{ij} \neq 0\}.$ → **1范数(L1)：** $\|E\|_1 = \sum_{ij} |E_{ij}|.$

最优凸近似

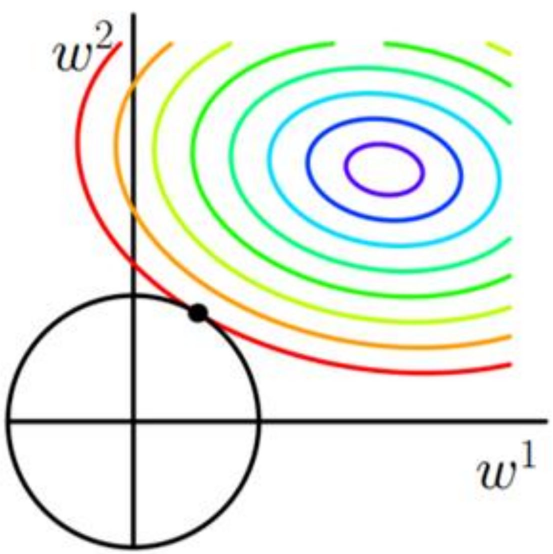
L1范数约束依然稀疏

$$\min_w \frac{1}{n} \| y - Xw \|^2, s.t \| w \|_1 \leq C$$



L2范数约束无稀疏特性

$$\min_w \frac{1}{n} \| y - Xw \|^2, s.t \| w \|_2 \leq C$$



ROBUST PCA

Convex relaxation

$$\min\ \operatorname{rank}(A) + \gamma\|E\|_0 \quad \text{subj} \quad A + E = D.$$

Convex ⇓ relaxation

$$\operatorname{rank}(A) = \#\{\sigma_i(A) \neq 0\}.$$

最优 ↓↓ 凸近似

$$\|A\|_* = \sum_i \sigma_i(A).$$

$$\|E\|_0 = \#\{E_{ij} \neq 0\}.$$

最优 ↓↓ 凸近似

$$\|E\|_1 = \sum_{ij} |E_{ij}|.$$

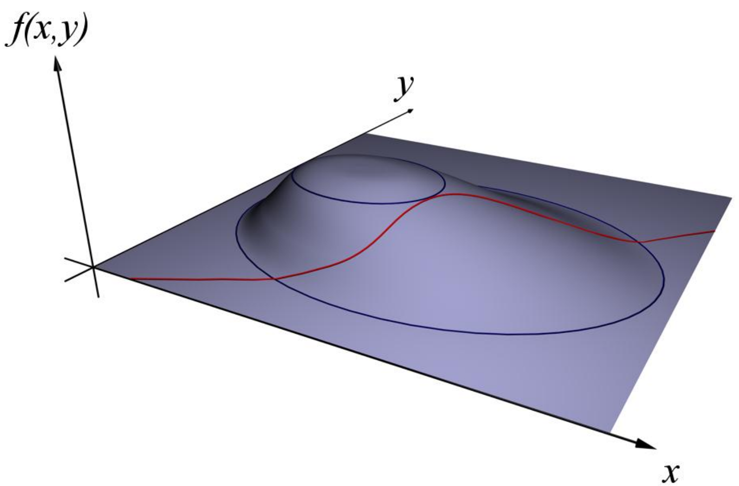
$$\min \quad \|A\|_* + \gamma \|E\|_1 \quad subj \quad A + E = D$$

NP问题转化为凸函数约束问题

拉格朗日乘子法原理

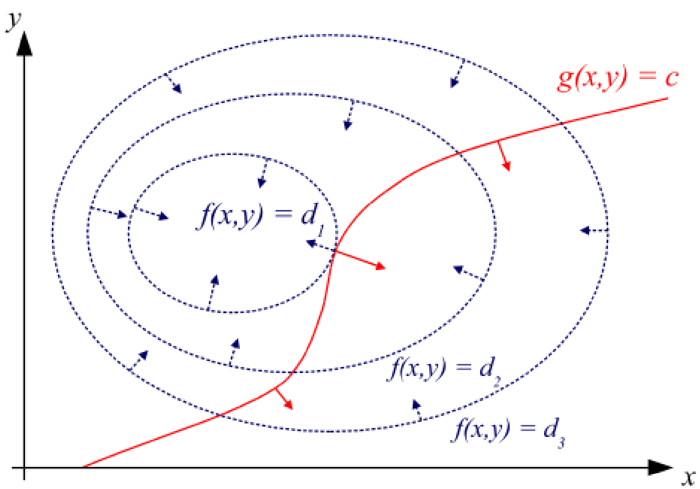
约束问题

$$\max \quad f(x, y)$$
$$s.t. \quad g(x, y) = c$$



极值点位置：

$$\nabla[f(x, y) + \lambda(g(x, y) - c)] = 0$$



约束问题

$$\max \quad f(x)$$
$$s.t. \quad g_i(x) = 0 \quad i = 1, 2 \ldots, n$$

无约束的目标函数替代原约束问题

极值问题

$$\max \quad L(x, \lambda_i) = f(x) + \sum_{i=1}^{n} \lambda_i g_i(x)$$

ROBUST PCA

增广拉格朗日乘子法

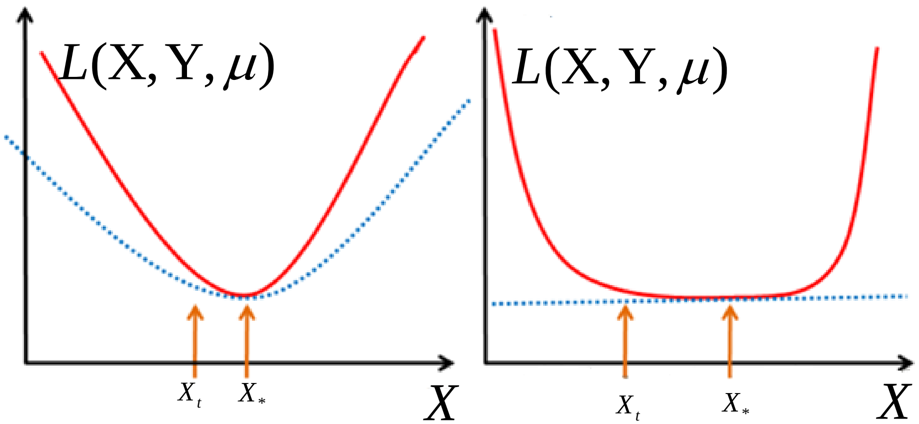
迭代解法

1: $\rho \geq 1$.
2: **while** not converged **do**
3: Solve $X_{k+1} = \arg\min_X L(X, Y_k, \mu_k)$.
4: $Y_{k+1} = Y_k + \mu_k h(X_{k+1})$;
5: Update μ_k to μ_{k+1}.
6: **end while**
Output: X_k.

$$\min \quad f(x), \quad s.t \quad h(x) = 0$$

$$L(X, Y, \mu) = f(X) + \langle Y, h(X) \rangle + \frac{\mu}{2} \| h(X) \|_2^2$$

强凸



ROBUST PCA

RPCA via the Inexact ALM Method

$$\min\ \operatorname{rank}(A) + \gamma\|E\|_0 \quad \text{subj} \quad A + E = D.$$

$$f(\mathrm{x}) = \|\mathrm{A}\|_* + \lambda\|\mathrm{E}\|_1, \textit{and} \qquad h(\mathrm{x}) = \mathrm{D} - \mathrm{A} - \mathrm{E}$$

$$L(\mathrm{A},\mathrm{E},\mathrm{Y},\mu) = \|\mathrm{A}\|_* + \lambda\|\mathrm{E}\|_1 + \langle Y, D - A - E \rangle + \frac{\mu}{2}\|\mathrm{D}-\mathrm{A}-\mathrm{E}\|_2^2$$

Input: Observation matrix $D \in \mathbb{R}^{m\times n}$, λ.
1: $Y_0 = D/J(D)$; $E_0 = 0$; $\mu_0 > 0$; $\rho > 1$; $k = 0$.
2: **while** not converged **do**
3: // Lines 4-5 solve $A_{k+1} = \arg\min_A L(A, E_k, Y_k, \mu_k)$.
4: $(U, S, V) = \operatorname{svd}(D - E_k + \mu_k^{-1} Y_k)$;
5: $A_{k+1} = U\mathcal{S}_{\mu_k^{-1}}[S]V^T$.
6: // Line 7 solves $E_{k+1} = \arg\min_E L(A_{k+1}, E, Y_k, \mu_k)$.
7: $E_{k+1} = \mathcal{S}_{\lambda\mu_k^{-1}}[D - A_{k+1} + \mu_k^{-1} Y_k]$.
8: $Y_{k+1} = Y_k + \mu_k(D - A_{k+1} - E_{k+1})$.
9: Update μ_k to μ_{k+1}.
10: $k \leftarrow k + 1$.
11: **end while**
Output: (A_k, E_k).

ROBUST PCA

RPCA求解

For a 1000x1000 matrix of rank 50, with 10% (100,000) entries randomly corrupted: $\min \|A\|_* + \lambda\|E\|_1 \quad \text{subj} \quad A + E = D.$

Algorithms	Accuracy	Rank	||E||_0	# iterations	time (sec)
IT	5.99e-006	50	101,268	8,550	119,370.3
DUAL	8.65e-006	50	100,024	822	1,855.4
APG	5.85e-006	50	100,347	134	1,468.9
APG_P	5.91e-006	50	100,347	134	82.7
$EALM_P$	2.07e-007	50	100,014	34	37.5
$IALM_P$	3.83e-007	50	99,996	23	11.8

10,000 times speedup!

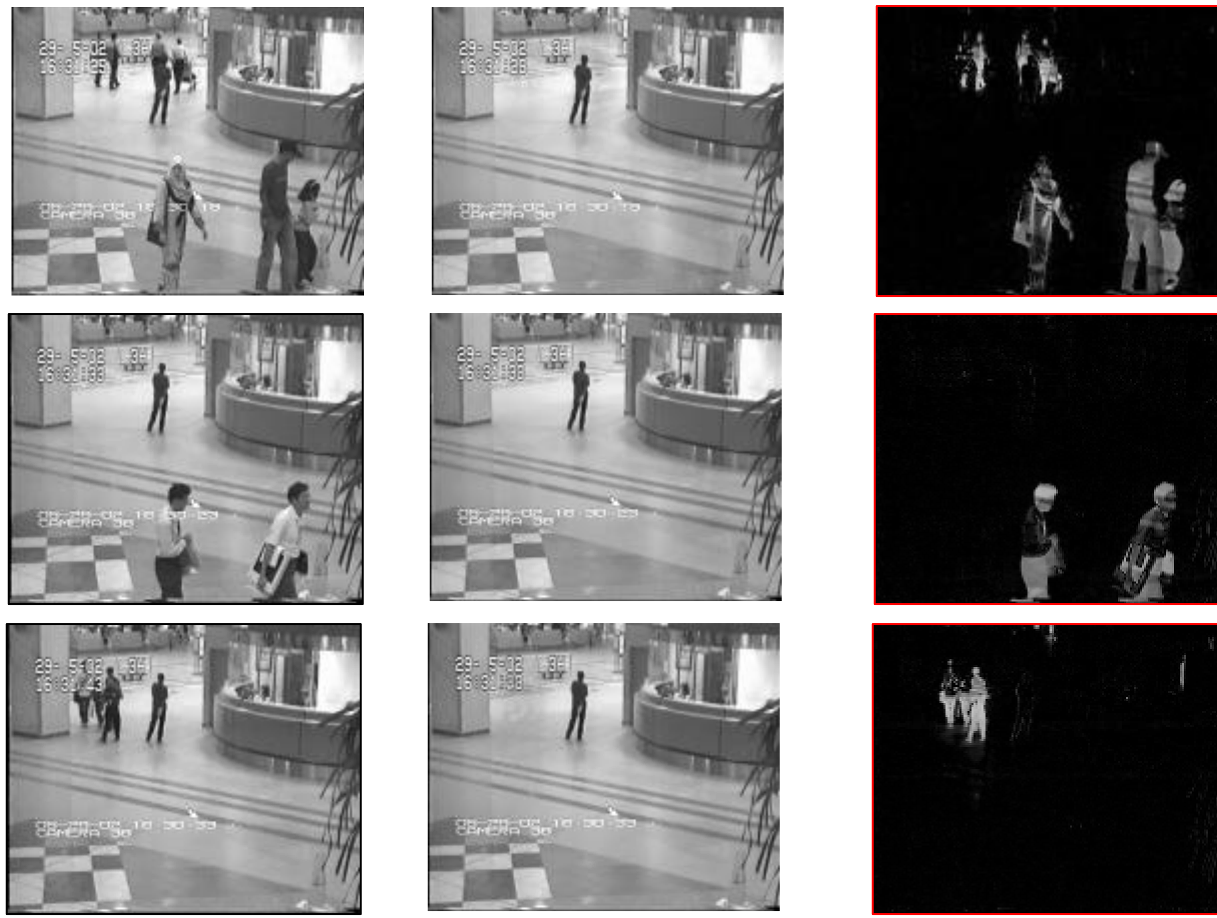
ROBUST PCA

应用

Background Modeling

$$D = A + E$$

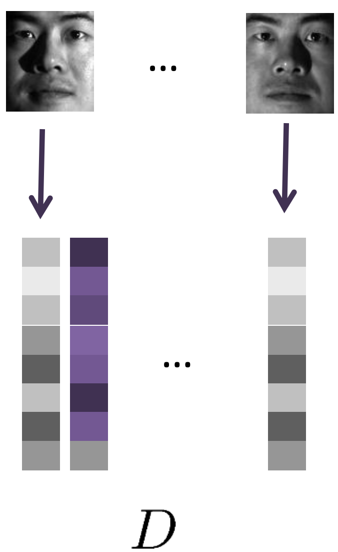
Video D | Low-rank appx. A | Sparse error E

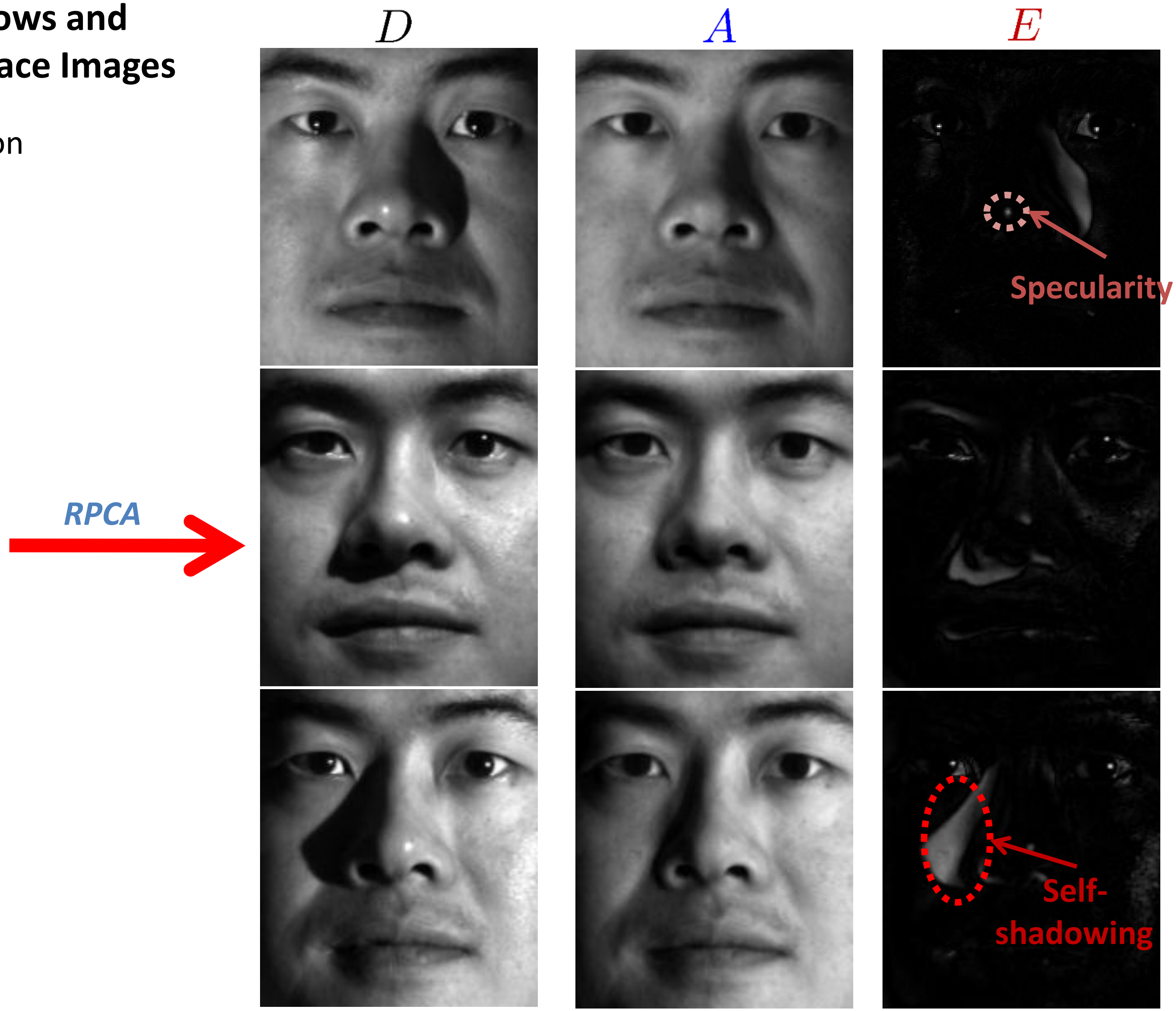


ROBUST PCA

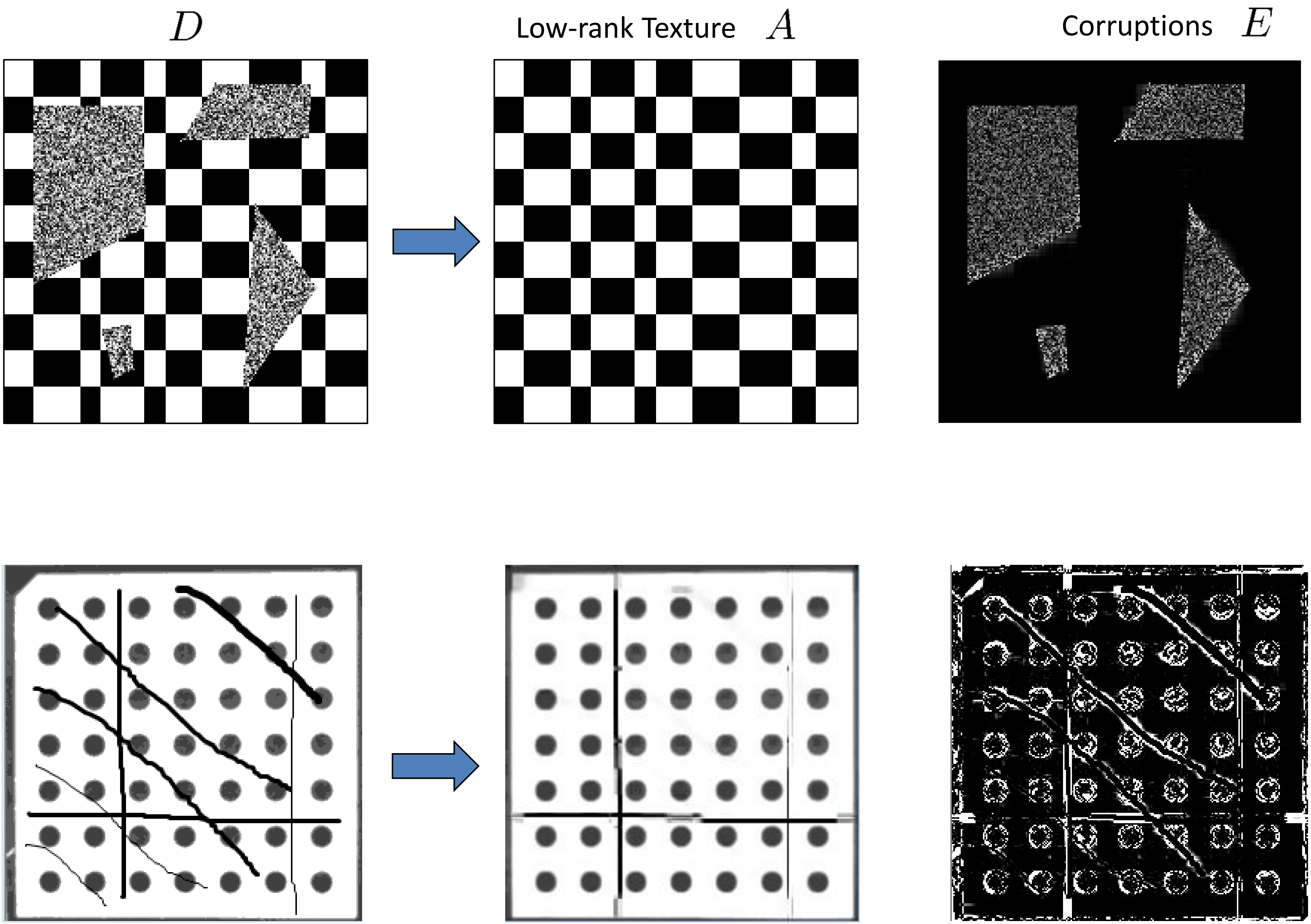
Removing Shadows and Specularities from Face Images

58 images of one person under varying lighting:



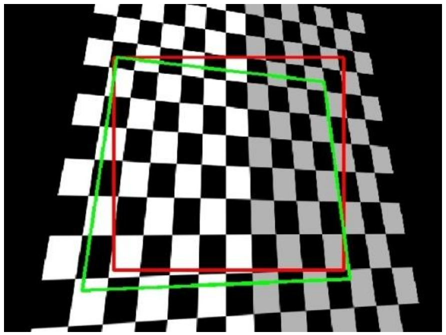


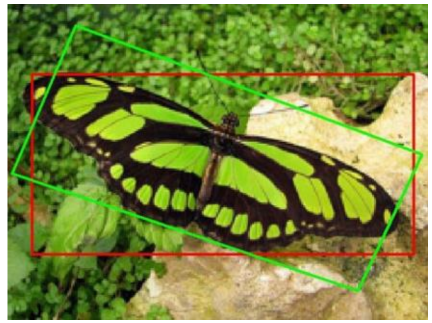
ROBUST PCA



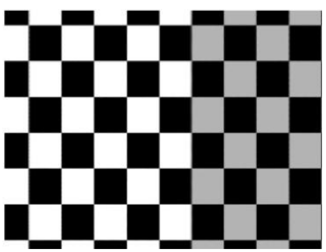
ROBUST PCA

Input (red window D)





Output (rectified green window A)



ROBUST PCA

参考：

Prof. Yi Ma	http://perception.csl.illinois.edu/matrix-rank/sample_code.html
github	https://github.com/andrewssobral/lrslibrary
Optimization	http://www.stat.purdue.edu/~vishy/introml/notes/Optimization.pdf
Lagrange multiplier	https://en.wikipedia.org/wiki/Lagrange_multiplier
RPCA	Robust Principal Component Analysis?
IALM	The Augmented Lagrange Multiplier Method for Exact Recovery of Corrupted Low-Rank Matrices