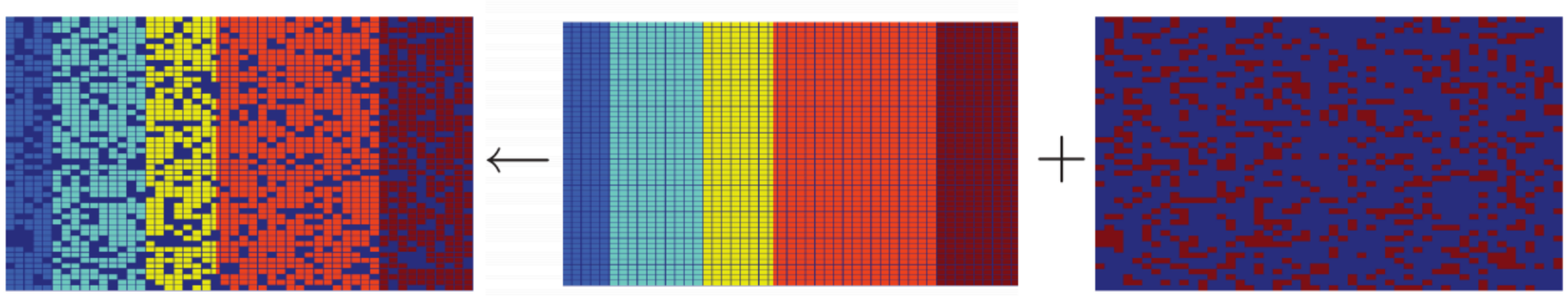
Robust PCA

Dalin Guo
March 22nd

Problem definition



From Emmanuel Candès: http://www.ihes.fr/~comdev/liens/Chaire_Schlumberger/candes

$$M = L_0 + S_0$$

- L0 has low-rank and S0 is sparse
- Both components are of arbitrary magnitude
- Unknown dimension of L0, number of and locations of non-zero entries of S0
- Can we recover L0 and S0 both accurately?

Candès, E. J., Li, X., Ma, Y., & Wright, J. (2011). Robust principal component analysis?. *Journal of the ACM (JACM)*, *58*(3), 11.

Application

- Video Surveillance: identify activities that stand out from the background

- Face Recognition: identify the low-dimensional subspace; remove defects in face images (e.g. specularities)

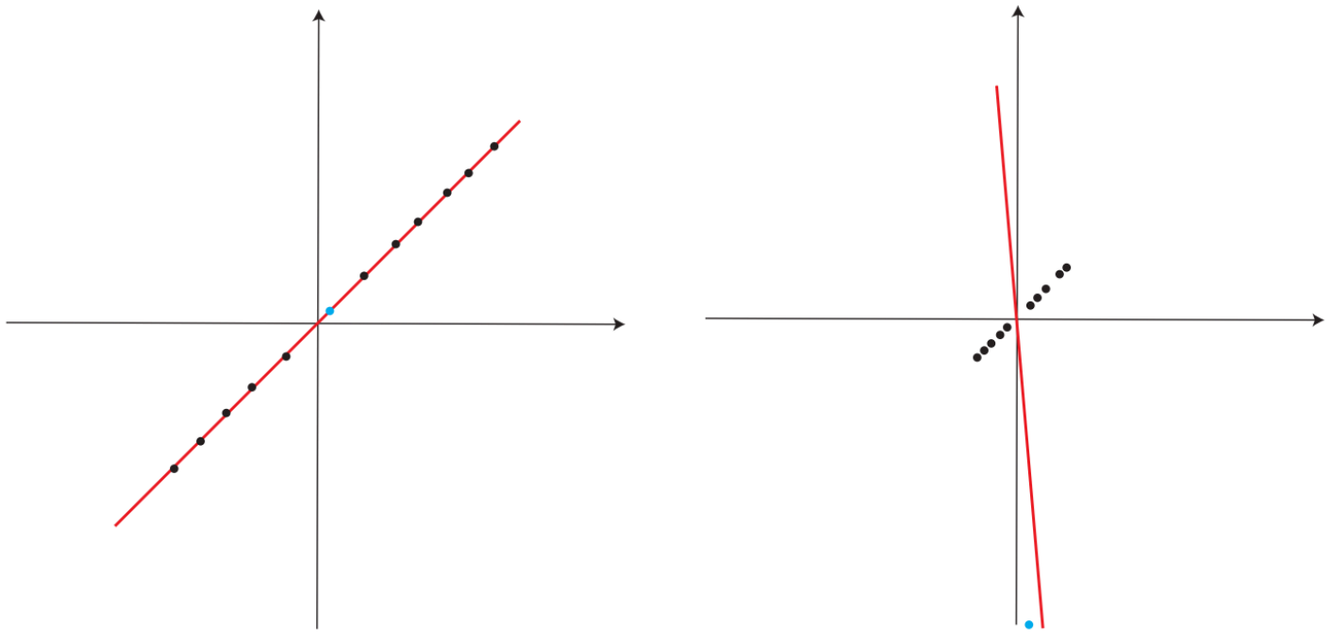
- Latent Semantic Indexing: L0 captures common words used in all documents; S0 captures the the few keywords that best distinguish each document

- Ranking and Collaborative Filtering: use incomplete rankings to predict the preference

Candès, E. J., Li, X., Ma, Y., & Wright, J. (2011). Robust principal component analysis?. *Journal of the ACM (JACM)*, *58*(3), 11.

Classical PCA

$$M = L_0 + N_0$$



From Emmanuel Candès: http://www.ihes.fr/~comdev/liens/Chaire_Schlumberger/candes

- N0: small and i.i.d. Gaussian
- Solution given by truncated SVD
- Very sensitive to outliers

Candès, E. J., Li, X., Ma, Y., & Wright, J. (2011). Robust principal component analysis?. *Journal of the ACM (JACM)*, *58*(3), 11.

Principal Component Pursuit (PCP)

Recovery via (convex) PCP

$$\begin{array}{ll} \text{minimize} & \|L\|_* + \lambda \|S\|_1 \\ \text{subject to} & L + S = M \end{array}$$

See also Chandrasekaran, Sanghavi, Parrilo, Willsky ('09)

- nuclear norm: $\|L\|_* = \sum_i \sigma_i(L)$ (sum of sing. values)
- ℓ_1 norm: $\|S\|_1 = \sum_{ij} |S_{ij}|$ (sum of abs. values)

From Emmanuel Candès: http://www.ihes.fr/~comdev/liens/Chaire_Schlumberger/candes

Candès, E. J., Li, X., Ma, Y., & Wright, J. (2011). Robust principal component analysis?. *Journal of the ACM (JACM)*, *58*(3), 11.

Guarantee

Theorem

- L_0 *is* $n \times n$ *of rank*$(L_0) \leq \rho_r n \, \mu^{-1} (\log n)^{-2}$
- S_0 *is* $n \times n$, *random sparsity pattern of cardinality* $m \leq \rho_s n^2$

Then with probability $1 - O(n^{-10})$, *PCP with* $\lambda = 1/\sqrt{n}$ *is exact:*

$$\hat{L} = L_0, \quad \hat{S} = S_0$$

Same conclusion for rectangular matrices with $\lambda = 1/\sqrt{\text{max dim}}$

- Exact
  - whatever the magnitudes of L_0!
  - whatever the magnitudes of S_0!
- No tuning parameter!

Can achieve stronger probabilities of success, e. g. $1 - O(n^{-\beta})$, $\beta > 0$

From Emmanuel Candès: http://www.ihes.fr/~comdev/liens/Chaire_Schlumberger/candes

Candès, E. J., Li, X., Ma, Y., & Wright, J. (2011). Robust principal component analysis?. *Journal of the ACM (JACM)*, *58*(3), 11.

Algorithm

- Augmented Lagrangian

$$\mathcal{L}(L, S; Y) = \|L\|_* + \lambda\|S\|_1 + \frac{1}{\tau}\langle Y, M - L - S\rangle + \frac{1}{2\tau}\|M - L - S\|_F^2$$

Easy to minimize over L and S separately

$$\arg\min_L \mathcal{L}(L, S, Y) = \mathcal{D}_\tau(M - S + Y)$$

$$\arg\min_S \mathcal{L}(L, S, Y) = \mathcal{S}_{\lambda\tau}(M - L + Y)$$

Scalar shrinkage: $\mathcal{S}_\tau[x] = \mathrm{sgn}(x)\max(|x| - \tau, 0)$

- Componentwise thresholding $\mathcal{S}_\tau(X)$
- Singular value thresholding $\mathcal{D}_\tau(X)$

$$\mathcal{D}_\tau(X) = U\mathcal{S}_\tau(\Sigma)V^* \qquad X = U\Sigma V^*$$

From Emmanuel Candès: http://www.ihes.fr/~comdev/liens/Chaire_Schlumberger/candes

Candès, E. J., Li, X., Ma, Y., & Wright, J. (2011). Robust principal component analysis?. *Journal of the ACM (JACM)*, *58*(3), 11.

Algorithm

- Alternating Directions

initialize: S_0, Y_0 and $\tau > 0$
while not converged

1. $L_k = \mathcal{D}_\tau(M - S_{k-1} + Y_{k-1})$ (shrink singular values)
2. $S_k = \mathcal{S}_{\lambda\tau}(M - L_k + Y_{k-1})$ (shrink scalar entries)
3. $Y_k = Y_{k-1} + (M - L_k - S_k)$

end while
output: L, S

From Emmanuel Candès: http://www.ihes.fr/~comdev/liens/Chaire_Schlumberger/candes

Candès, E. J., Li, X., Ma, Y., & Wright, J. (2011). Robust principal component analysis?. *Journal of the ACM (JACM)*, *58*(3), 11.

Simulation Result

Dimension n	$\text{rank}(L_0)$	$\|S_0\|_0$	$\text{rank}(\hat{L})$	$\|\hat{S}\|_0$	$\frac{\|\hat{L}-L_0\|_F}{\|L_0\|_F}$	# SVD	Time(s)
500	25	12,500	25	12,500	1.1×10^{-6}	16	2.9
1,000	50	50,000	50	50,000	1.2×10^{-6}	16	12.4
2,000	100	200,000	100	200,000	1.2×10^{-6}	16	61.8
3,000	250	450,000	250	450,000	2.3×10^{-6}	15	185.2

$\text{rank}(L_0) = 0.05 \times n,\ \|S_0\|_0 = 0.05 \times n^2.$

Dimension n	$\text{rank}(L_0)$	$\|S_0\|_0$	$\text{rank}(\hat{L})$	$\|\hat{S}\|_0$	$\frac{\|\hat{L}-L_0\|_F}{\|L_0\|_F}$	# SVD	Time(s)
500	25	25,000	25	25,000	1.2×10^{-6}	17	4.0
1,000	50	100,000	50	100,000	2.4×10^{-6}	16	13.7
2,000	100	400,000	100	400,000	2.4×10^{-6}	16	64.5
3,000	150	900,000	150	900,000	2.5×10^{-6}	16	191.0

$\text{rank}(L_0) = 0.05 \times n,\ \|S_0\|_0 = 0.10 \times n^2.$

Table 1: Correct recovery for random problems of varying size. Here, $L_0 = XY^* \in \mathbb{R}^{n \times n}$ with $X, Y \in \mathbb{R}^{n \times r}$; X, Y have entries i.i.d. $\mathcal{N}(0, 1/n)$. $S_0 \in \{-1, 0, 1\}^{n \times n}$ has support chosen uniformly at random and independent random signs; $\|S_0\|_0$ is the number of nonzero entries in S_0. Top: recovering matrices of rank $0.05 \times n$ from 5% gross errors. Bottom: recovering matrices of rank $0.05 \times n$ from 10% gross errors. In all cases, the rank of L_0 and ℓ_0-norm of S_0 are correctly estimated. Moreover, the number of partial singular value decompositions (# SVD) required to solve PCP is almost constant.

Candès, E. J., Li, X., Ma, Y., & Wright, J. (2011). Robust principal component analysis?. *Journal of the ACM (JACM)*, *58*(3), 11.

Face Image De-nosing

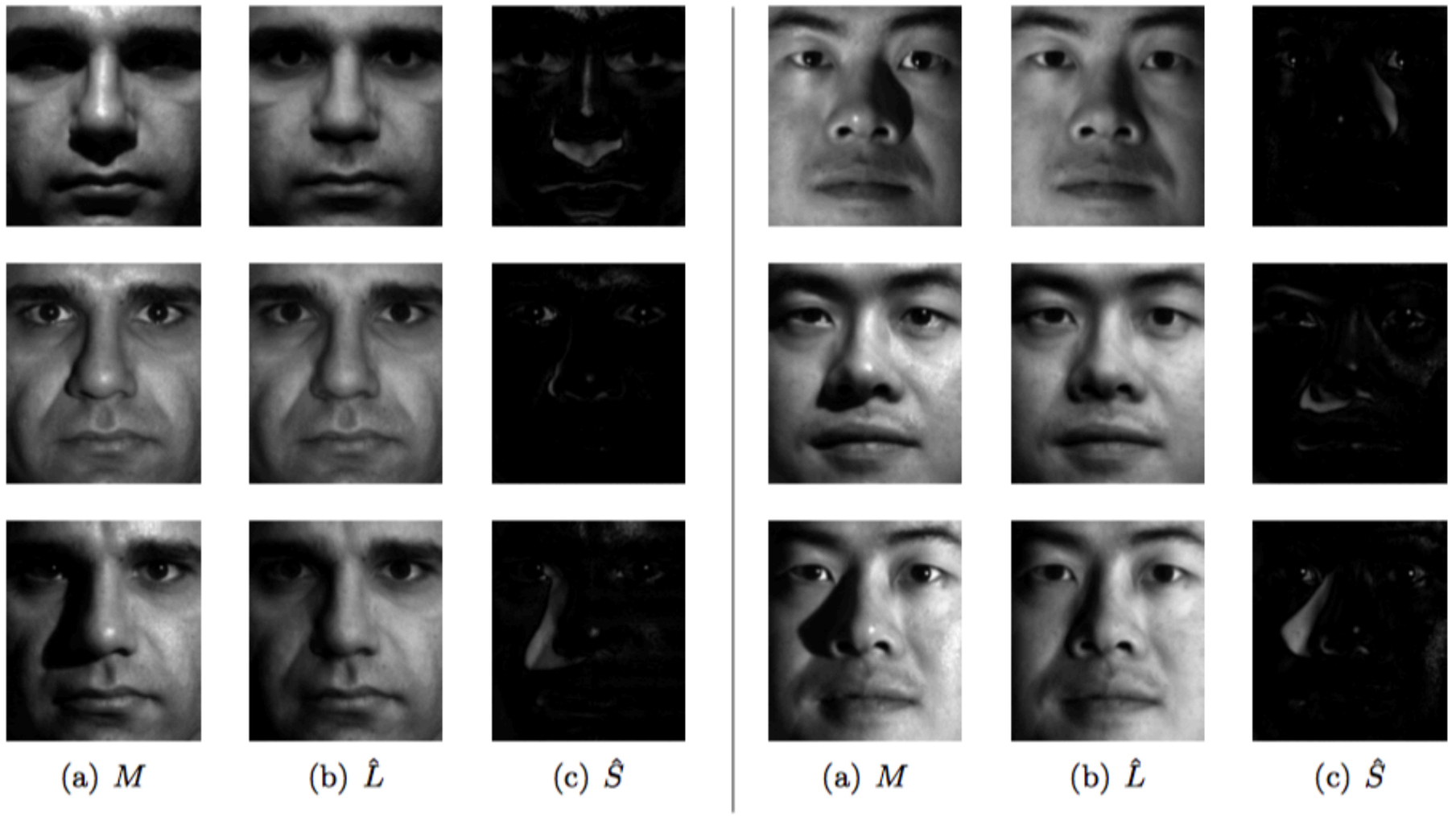


Figure 4: Removing shadows, specularities, and saturations from face images. (a) Cropped and aligned images of a person's face under different illuminations from the Extended Yale B database. The size of each image is 192×168 pixels, a total of 58 different illuminations were used for each person. (b) Low-rank approximation $\hat{L}$ recovered by convex programming. (c) Sparse error $\hat{S}$ corresponding to specularities in the eyes, shadows around the nose region, or brightness saturations on the face. Notice in the bottom left that the sparse term also compensates for errors in image acquisition.

Candès, E. J., Li, X., Ma, Y., & Wright, J. (2011). Robust principal component analysis?. *Journal of the ACM (JACM)*, *58*(3), 11.

Another Problem

$$Y = L + S + E$$

- In practice, measurement noise exists everywhere within the matrix

- low-rank, sparse and dense noise components: L0, S0, N0

- Solved by a Bayesian framework
  - can infer full posterior
  - can account for correlation (e.g. video, Markov property)

- Posterior Inference: MCMC/VB

Ding, X., He, L., & Carin, L. (2011). Bayesian robust principal component analysis. *IEEE Transactions on Image Processing*, *20*(12), 3419-3430.

Another Problem

$$Y = L + S + E$$

- Solved by a Bayesian framework
  - noise statistics may be inferred, no tunings of hyper parameters
  - allow non-stationary noise
- Posterior Inference:
  - MCMC converge very fast (10 iterations in the example)
  - ~~VB very sensitive to initialization~~

Ding, X., He, L., & Carin, L. (2011). Bayesian robust principal component analysis. *IEEE Transactions on Image Processing*, *20*(12), 3419-3430.