Matrix completion

Gursimran singh

Thapar university

June 5,2012

Singular value decmoposition

$$M = U\sum V^T$$

Where

$U = \{u_1, u_2, u_3 \ldots . u_m\}$ *are the left singular vectors of MM*

$V = \{v_1, v_2, v_3 \ldots . v_n\}$ *are the right singular vectors of M*

$\begin{bmatrix} s_1 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & s_r \end{bmatrix}$ *are the r singular values of M*

Such that

$$Mu_i = s_i v_i$$

Satisfying these constraints

$U^T U = I$ *U is orthonormal*

$V^T V = I$ *V is orthonormal*

$s_{ij} = 0\,;\ if\ i \neq j$ *i.e S is diagnol matrix*

Singular value decmoposition ... continued

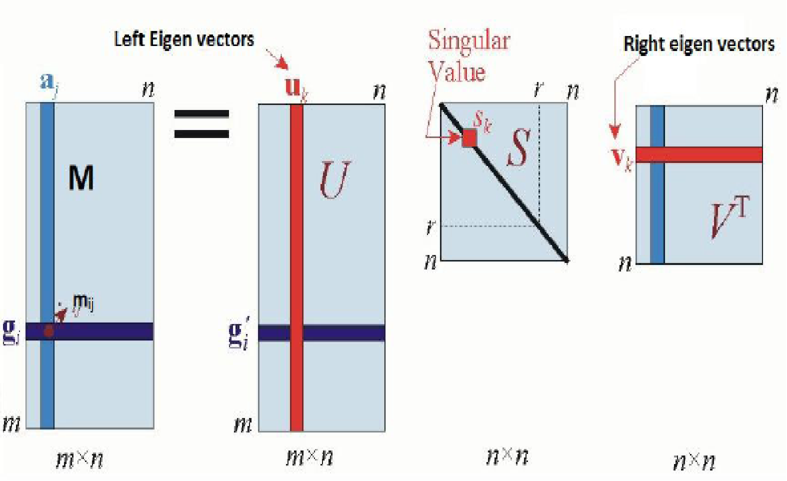


Figure 2.1 - Partial singular value decomposition

Principal Component Analysis

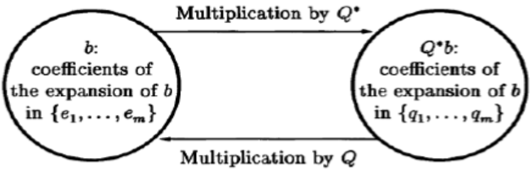


Figure 2.2 Basis change equation

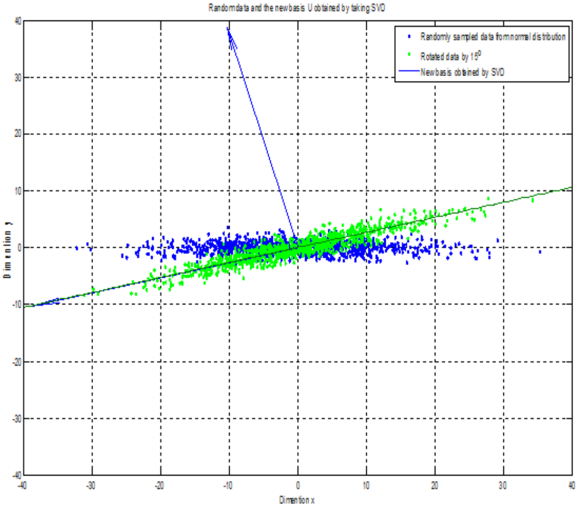
$$b' = Q^* b$$

b' is the new dataset in new orthonormal basis Q

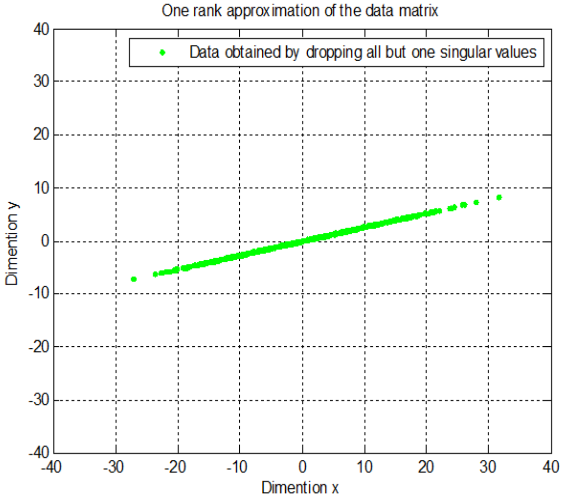
b is the original dataset in I

Columns of Q contains new basis vectors

Principal Component Analysis



Principal Component Analysis



Matrix completion problem

$\boxed{R^{nxn}}$

Figure 3.1 - Matrix completion

Degrees of freedom in a matrix(n×n) of rank r

r(2n-r)

Proof of $r(2n-r)$

$\Rightarrow nr + (n-r)r$
$\Rightarrow nr + nr - r^2$
$\Rightarrow 2nr - r^2$

- This will be of Θ(nr) for a n×n matrix of rank r.
- which is much smaller than $\Theta(n^2)$.

Example

For a matrix 1000x1000 of rank 20 we need just 40000 numbers instead of 1000000 numbers.

Which set of $r(2n-r)$ numbers?

Ofcourse not all sets of such numbers will work.

1. If a complete row or column is missing, then you can't recover. There is no hope.
2. If the matrix is something like this.

$$M = e_1 e_n^* = \begin{bmatrix} 0 & \cdots & 1 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & 0 \end{bmatrix}$$

3. or

$$M = e_1 x^* = \begin{bmatrix} x_1 & \cdots & x_n \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}$$

4. So there are bunch of cases in which matrix recovery isn't going to work, whatsoever the method is..

Sample the matrix at random?

- Only some combinitions of $r(2n-r)$ numbers are going to be the right choice. (Good cases/ Bad cases)
  - Matrix.
  - No of enteries. $(> r(2n-r))$
  - Which combination of these enteries.

- What is **Probability** of exact matrix completion. Randomly sample
  1. Matrix - Number of combinations of numbers that are GOOD cases.
  2. Number of samples $(> r(2n-r)$, otherwise its 0)

Central Idea!!

If I want the matrix to get completed for **most of cases (fixed probability)**, we will need **different amount of enteries** $\mathbf{(}> r(2n-r)\mathbf{)}$ for different matrices.

$$Samlpes \propto \Psi_X(M) \text{ given some P}$$

Coherence

Matrix U should be uniformly distributed across all canonical basis

$$L \in R^{n \times n} = U \Sigma V^*$$

Coherence parameter $\mu \geq 1$

$\| P_U e_i \|^2 \leq \frac{\mu r}{n}$ $\qquad$ $\| P_V e_i \|^2 \leq \frac{\mu r}{n}$ where

- P_i is projection onto corrosponding space.
- e_i is i_{th} unit vector.

Alternatively

$$\mu(U) = \frac{n}{r} \max(\| P_U e_i \|)$$

- Weak coherence property, 2008.
- Strong coherence property, 2009. Result improved.

Mathematical results

Lemma: Information theoretic limit (C and Tao, 2009)

Coupon collector's effect, No method whatsoever can work if no of sampled enteries in the matrix is

$$m \lesssim \mu \times nr \times \log(n)$$

Candes and Recht, 2008

Recovering M exactly is possible with probability at least $1 - cn^-3$ when sampled enteries

$$m \gtrsim \mu \times n^{6/5} \times r \log(n) \qquad \text{sometimes } (\tfrac{6}{5} \textit{ instead of } \tfrac{5}{4})$$

Improving the result, Candes and Tao, 2009

Recovering M exactly is possible with probability at least $1 - n^-10$ when sampled enteries

$$m \gtrsim \mu \times n \times r \log^a(n) \qquad \text{a} \leq 6 \text{ (sometimes 2)}$$

If we have right $r(2n-r)$ numbers

If we chose the right numbers, we can hope to find only one matrix that will be the solution to the problem below.
Otherwise there could be multiple matrices that are consistant with those missing values and we can never know which one out of them all.

Rank minimisation problem

$$\text{minimize } \operatorname{rank}(X)$$
$$\text{subject to } X_{ij} = M_{ij} \quad \forall (i,j) \in \Omega_{known}$$

Conclusions

- If there are more than one matrix that satisfies the above convex optimisation problem then it means you can't recover.
- It means the choice of number is not right.

How to minimise rank, NP hard

We were minimising the L_2 norm using gradient descent, which can be visualized as rank minimisation due to constraints. **(not sure though)**

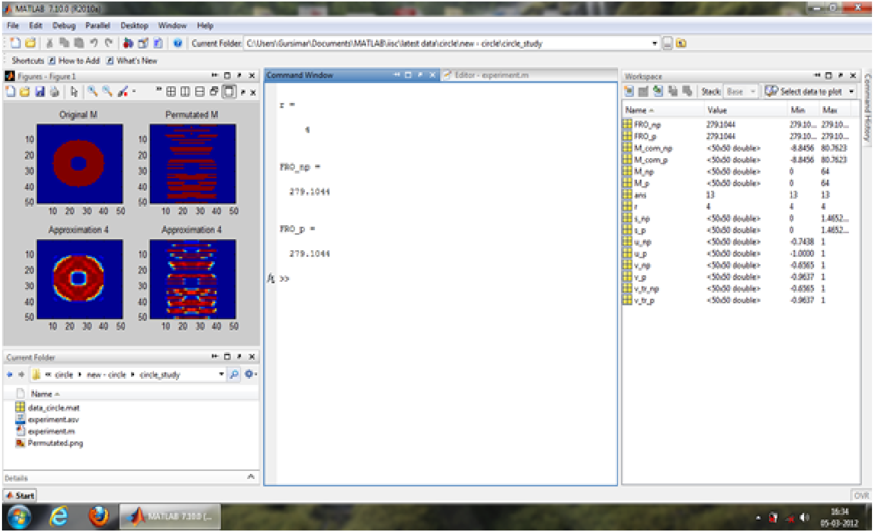
$$min \parallel M' - X \parallel \quad \forall (i,j) \in \Omega_{known}$$

Nuclear norm

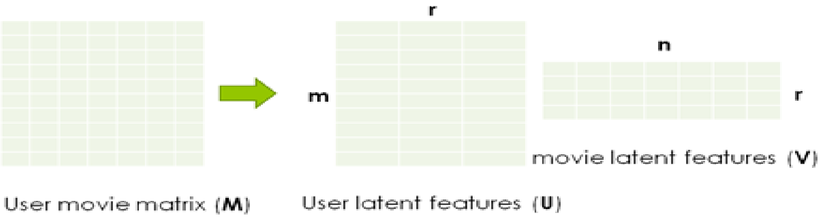
$$minimize \parallel X \parallel_*$$
$$subject\ to\ X_{ij} = M_{ij}\ \forall (i,j) \in \Omega_{known}$$
$$where, \parallel X \parallel_* = \sum_{i=1}^{r} \sigma_i^2$$

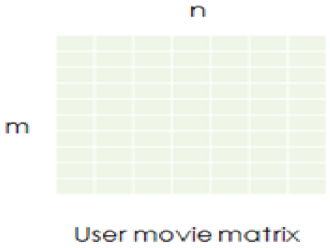
- This will be proxy for l_0 norm of singular values matrix.
- l_0 norm is number of non zero values, which is essentially rank of matrix.

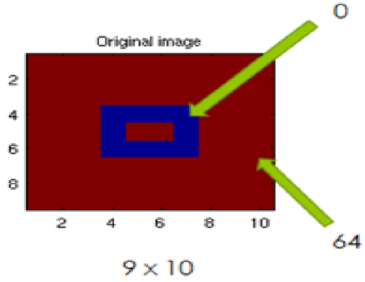
Experiments



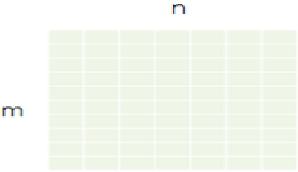
Rectangular matrix



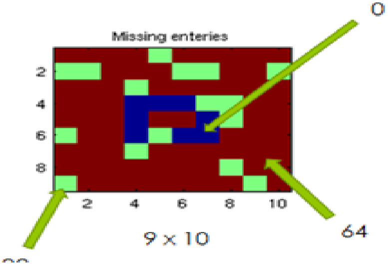


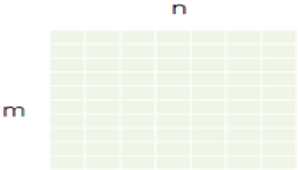


Rectangular matrix - Initial matrix

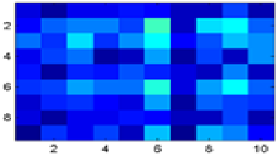


User movie matrix



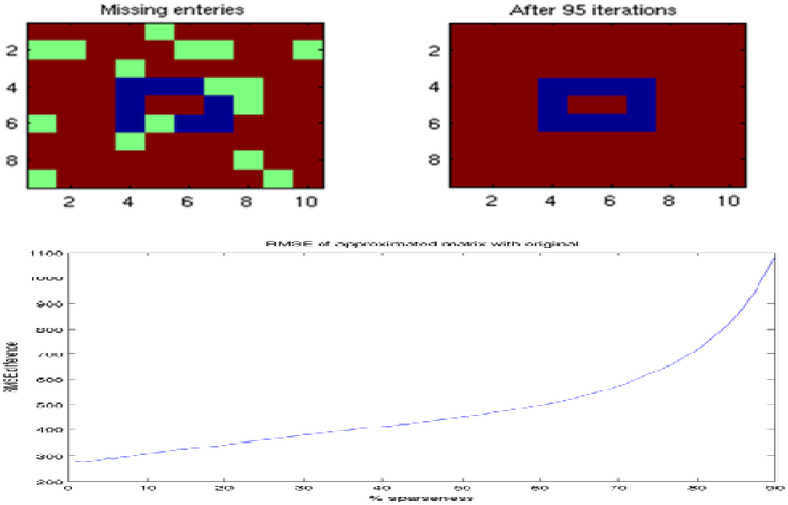


User data matrix

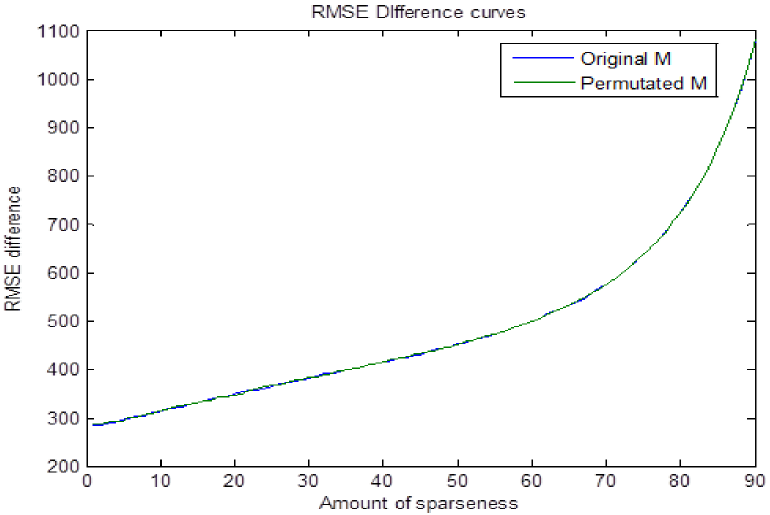


9 x 10

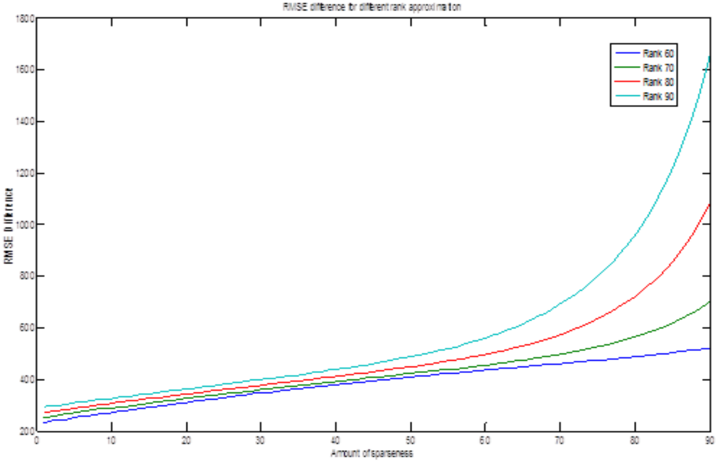
Rectangular matrix - Recovery



Rectangular matrix - Permutation

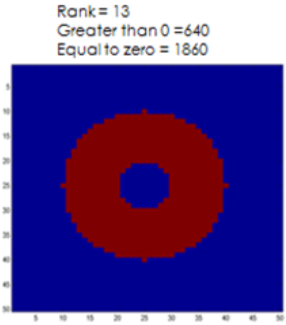


Rectangular matrix - Comaprison with r¿R

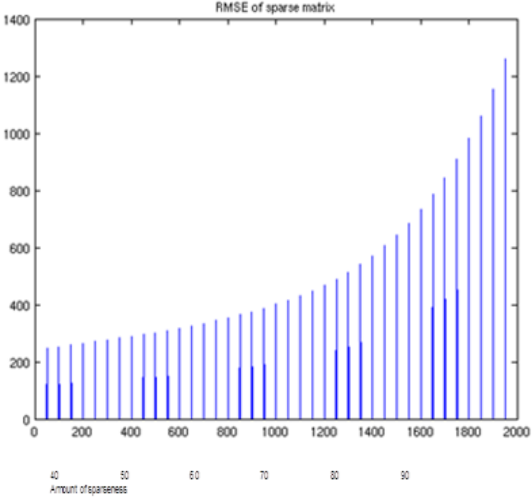


Amount of sparseness

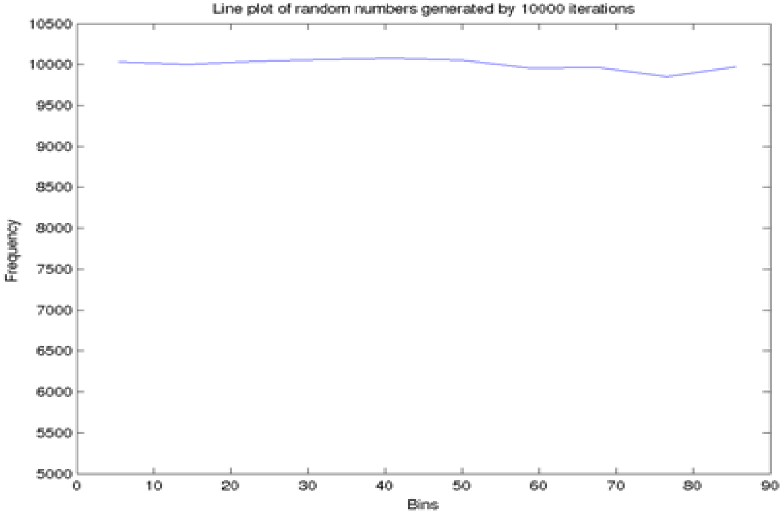
Rectangular matrix - Circle



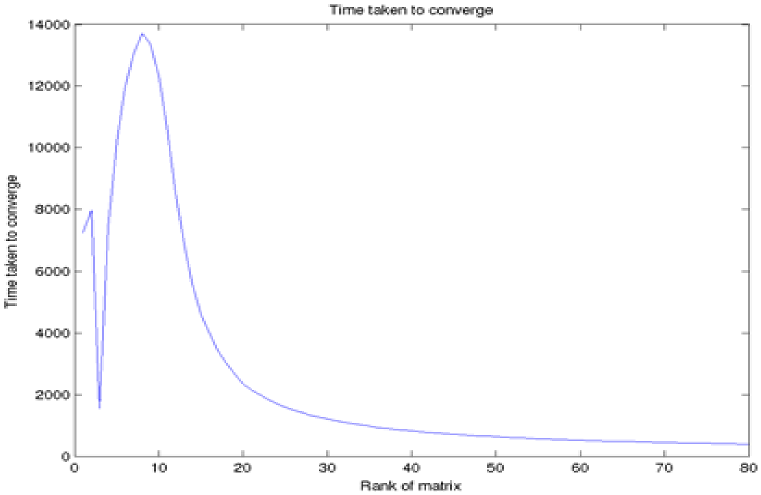
- R=260
- mc_itr =13000
- l_lim= 2000/50
- Time ~ 43hrs



Testing - distribution testing



Testing - Time vs rank



Applications

- High dimentionality, but low rank structure.
  1. Netflix matrix
  2. Triangulation with sensor net.
  3. Quantum state tomography
  4. Machine learning
  5. Corrupted videos
- Robost PCA
  1. Recovering foreground from background
  2. Removing shadows from face images.
  3. Batch face alignment
- Transform invariant low rank textures.
- 3D reconstruction
- Latent semantic analysis