Consider the Netflix problem, where we are given a collection of users, each of whom has rated some subset of the available movies/shows on Netflix. If we arrange this information in a matrix, it may look like this (obviously Netflix has more users and movies than this).

	movie 1	movie 2	movie 3	movie 4
user 1	3	?	5	?
user 2	?	1	?	2
user 3	2	?	2	?
user 4	?	4	?	?
user 5	5	?	?	4

Obviously, Netflix wishes to fill in these missing entries so it knows what movies to recommend to a given user. One method of filling in these missing entries leverages low-rank structure in the matrix. This is called low-rank matrix completion (LRMC).

Observation Model

Assume $X \in \mathbb{R}^{M \times N}$ has rank $r \ll \min(M, N)$, so that

$$X = \sum_{i=1}^{r} \sigma_i u_i v_i^T .$$

Suppose we observe a subset of the entries in X, so we are given a matrix $Y \in \mathbb{R}^{M \times N}$ such that

$$Y_{ij} = \begin{cases} X_{ij} & (i,j) \in \Omega \\ ? & \text{otherwise} \end{cases}, \quad \Omega \subset \{1, \dots, M\} \times \{1, \dots, N\}$$

where Ω is a set of known sampling locations. We may then ask

1) Can we recover X from Y?

2) How do we do it?

3) How well does it work?

Sampling Conditions

To answer the first question, assume we receive exact (noiseless) measurements. If X has rank r, we can write

$$X = \tilde{U}_r \tilde{V}_r^T,$$

where $\tilde{U}_r$ has size $M \times r$ and $\tilde{V}_r$ has size $N \times r$. In this light, the degrees of freedom in X are $Mr + Nr = (M+N)r$, which is typically much smaller than MN. Alternatively, suppose the first r columns are linearly independent, and the next $N-r$ columns are dependent entirely on the first r. This gives $Mr + (N-r)r = (M+N)r - r^2$ degrees of freedom.

If $M \approx N$, then the DoF $\approx 2Nr$, so we need at least $O(Nr)$ samples to have any hope of recovery.

Q: If X is $N \times N$, approximately how many samples per column do we need?

LRMC Problem Formulation

How should we formulate the LRMC problem as an optimization problem? We have two goals for an estimate $\hat{X}$:

1) $\hat{X}_{ij} = X_{ij}$ for $(i,j) \in \Omega$ (observed entries match)
2) $\operatorname{rank}(\hat{X}) = r$ (estimate is low-rank)

Let P_Ω be the orthogonal projection onto the space of matrices supported on Ω. Then we write

$$P_\Omega X = \begin{cases} X_{ij} & (i,j) \in \Omega \\ 0 & \text{otherwise} \end{cases}$$

An equality-constrained / basis pursuit-type optimization problem is then

$$\min_X \ \|X\|_* \qquad (1)$$
$$\text{ST} \quad P_\Omega(X) = P_\Omega(Y)$$

Alternatively, if we believe our observations are corrupted by noise, we may choose to solve the Lasso-type / Tikhonov regularized problem

$$\min_X \frac{1}{2}\|P_\Omega(X) - P_\Omega(Y)\|_F^2 + \lambda\|X\|_* . \quad (2)$$

[Note the similarity between (1-2) and our optimization problems for sparse regression.]

Algorithms for LRMC

We can solve the LRMC problem using both ADMM and IRLS. The latter will appear in our last demo. Define the (split) augmented Lagrangian to be

$$\mathcal{L}_\rho(X, Z, L) = \lambda\|Z\|_* + \frac{1}{2}\|P_\Omega(X) - P_\Omega(Y)\|_F^2 + \langle L, X - Z\rangle + \frac{\rho}{2}\|X - Z\|_F^2$$

where L is the matrix of Lagrange multipliers. The ADMM updates are given below. Their derivation is likely to be a homework problem. Note the similarity to the Lasso updates!

ADMM for LRMC

$$X_{ij}^{(k+1)} = \begin{cases} \frac{1}{\rho+1} Y_{ij} & \text{if } (i,j) \in \Omega \\ Z - L & \text{if } (i,j) \notin \Omega \end{cases}$$

$$Z^{(k+1)} = D_{\lambda/\rho}(X + L)$$

$$L^{(k+1)} = L + X - Z$$