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).

\[
\begin{array}{c|c|c|c|c}
 & \text{movie 1} & \text{movie 2} & \text{movie 3} & \text{movie 4} \\
\hline
\text{user 1} & 3 & ? & 5 & ? \\
\hline
\text{user 2} & ? & 1 & ? & 2 \\
\hline
\text{user 3} & 2 & ? & 2 & ? \\
\hline
\text{user 4} & ? & 4 & ? & ? \\
\hline
\text{user 5} & 5 & ? & ? & 4 \\
\hline
\end{array}
\]

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 $r < \min(m,n)$, so that

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

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

$$Y_{ij} = \begin{cases} X_{ij} & (i,j) \in \mathcal{S} \\ \text{?} & \text{otherwise} \end{cases}, \quad \mathcal{S} \subset \{1, \ldots, m\} \times \{1, \ldots, n\}^2$$

where $\mathcal{S}$ 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?
Supply 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 $M \times r + N \times r = (M+N)r$, 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 $M \times r + (N-r)r = (M+N)r - r^2$ degrees of freedom.

If $M = N$, then the DoF $= 2Nr$, 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 do we need?
LRMC Problem Formulation

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

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

Let $P_2$ be the orthogonal projection onto the space of matrices supported on $\mathcal{I}$. Then we write

$$P_2 X = \begin{cases} X_{ij} & (i,j) \in \mathcal{I} \\ 0 & \text{otherwise} \end{cases}$$

An equality-constrained $\ell_1$-norm pursuit-type optimization problem is then

$$\min_X \|X\|_{1,1} \quad \text{s.t.} \quad P_2(X) = P_2(Y)$$

(1)
Alternatively, if we believe our observations are corrupted by noise, we may choose to solve the Lasso-type/Tikhonov regularized problem

$$\min_{\mathbf{x}} \frac{1}{2} \| \mathbf{P}_2(\mathbf{x}) - \mathbf{P}_2(\mathbf{y}) \|_F^2 + \lambda \| \mathbf{x} \|_1 . \quad (2)$$

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 on homework. Define the (split) augmented Lagrangian to be

$$L^s(\mathbf{x}, \mathbf{z}, \mathbf{L}) = \frac{1}{2} \| \mathbf{z} \|_2^2 \quad + \frac{1}{2} \| \mathbf{P}_2(\mathbf{x}) - \mathbf{P}_2(\mathbf{y}) \|_F^2 \quad + \langle \mathbf{L}, \mathbf{x} - \mathbf{z} \rangle \quad + \frac{\lambda}{2} \| \mathbf{x} - \mathbf{z} \|_2^2$$

where $\mathbf{L}$ 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

\[ X^{(k+1)}_{i,j} = \begin{cases} \frac{1}{\rho} Y_{i,j} & \text{if } (i,j) \in S \\ Z - L & \text{if } (i,j) \notin S \end{cases} \]

\[ Z^{(k+1)} = \text{prox}_{\lambda p} (X + L) \]

\[ L^{(k+1)} = L + X - Z \]