Robust Matrix Completion via Robust Gradient Descent

Andrew Singh
Carnegie Mellon University, Pittsburgh, PA

Abstract
We consider the problem of robust matrix completion, where the objective is to recover a partially observed matrix of low rank that has been corrupted by an adversary. We present a matrix completion method that is robust to corruption from an arbitrary outlier distribution. Our method performs gradient descent in an alternating minimization scheme, robustly estimating the mean of the sample gradients in each gradient update. Our experiments show that our method is substantially more robust than conventional techniques for matrix completion.

Introduction
Matrix completion is the problem of reconstructing a matrix for which only a fraction of the entries are observed. With no prior assumptions on the matrix, this problem is unsolvable since the missing entries could be arbitrary values. However if we assume that the matrix has some underlying structure, that is, the matrix is of low rank, it becomes possible to recover the matrix exactly even when only a small fraction of its entries are observed.

The problem can be formulated as follows. Let \( L \in \mathbb{R}^{m \times n} \) be a matrix of rank \( r \). Suppose we only observe a \( p \) fraction of entries in \( L \). The objective of the problem is, given these sparse observations, to find factors \( U \in \mathbb{R}^{m \times r} \) and \( V \in \mathbb{R}^{n \times r} \) such that the reconstruction error \( \| L - UV \|_F \) is minimized.

While there are several established methods for solving this problem when the data is clean, these methods can easily break down when a small fraction of malicious noise has been added to the data. We consider the problem of robust matrix completion, where a fraction of our observations have been corrupted by an adversary. Instead of observing a fraction of the entries of \( L \), we now observe a fraction of \( M = L + \delta \), where \( \delta \in \mathbb{R}^{m \times n} \) is a matrix of arbitrary sparse corruptions.

Matrix completion has several real-world applications, a popular one of which is recommender systems [1]. Consider \( n \) users, \( m \) movies, and the \( n \times m \) ratings matrix \( R \) where \( R_{ij} \) is user \( i \)'s rating for movie \( j \). Since each user has seen only a small fraction of movies, this matrix is very sparse. The goal of the recommender system is to predict users’ ratings for unseen movies; that is, predict the values of the missing entries in the ratings matrix.

Methods
One of the most successful approaches to matrix completion is alternating minimization [2][3]. Rather than a specific algorithm, AltMin is a general framework for matrix completion, and has been applied to other optimization problems as well. Recall that the objective of matrix completion is to find factors \( U, V \) such that their product \( UV \) best approximates the full matrix. The framework consists of two stages: fixing \( V \) and optimizing for \( U \), and fixing \( U \) and optimizing for \( V \). The algorithm alternates between these two stages until convergence.

The problem with conventional AltMin in the noisy setting is that it uses the traditional L2 norm as its loss function, which is very sensitive to outliers. One solution is to modify the loss function to a more robust alternative, such as the L1 norm, but even this breaks down under noise with bias, as shown in our experiments.

Our approach employs the technique of robust gradient estimation proposed by Prasad et al. [4]. The key idea of this method is that when performing gradient descent on noisy data, instead of modifying the loss function, we can robustly estimate the sample gradients. In batch gradient descent, a gradient update is simply the mean of the sample gradients. By robustly estimating the gradient mean, we can build a model that better handles arbitrary corruptions to the data.

Our algorithm is based on alternating minimization, but employs robust gradient descent for the optimization. Let \( Y_0 \) be the sampling operator on observed indices \( \Omega \) such that \( Y_0(M) \) is the matrix of observed entries of \( M \) (the unobserved entries are set to \( 0 \)).

Algorithm 1: AltMin with Robust Gradient Descent

Input : Observed matrix \( Y_0(M) \in \mathbb{R}^{m \times n} \), rank \( r \), corruption fraction \( \eta \)
Output : Optimized factors \( U \in \mathbb{R}^{m \times r} \), \( V \in \mathbb{R}^{n \times r} \)
while \( \| Y_0(M) - Y_0(UV) \|_F \leq \delta \) do
while \( U \) has not converged do
for row \( u \in U \) do
\[ \text{grads} = \frac{1}{2} \left( Y_0(M) - Y_0(UV) \right)_\Omega \]
\( u \leftarrow u - \text{RobustMean(grads, } \eta \) \)
while \( V \) has not converged do
for row \( v \in V \) do
\[ \text{grads} = \frac{1}{2} \left( Y_0(M) - Y_0(UV) \right)_\Omega \]
\( v \leftarrow v - \text{RobustMean(grads, } \eta \) \)
return \( U, V \)

Experiments
We compare the performance of our algorithm to two conventional methods for matrix completion. All three methods are under the general alternating minimization framework, but differ in their loss function, or in the case of our robust algorithm, how the sample gradients are averaged. The standard AltMin algorithm uses the L2 norm as its loss function. Another popular alternative to the L2 norm is the L1 norm, which minimizes the sum of absolute values of errors rather than the sum of squared errors. We compare both of these methods to our robust estimator.

We generated random factors \( U, V \) from a standard normal distribution to obtain low-rank matrix \( L = UV \). We then sampled a random \( p \) fraction of entries from each row and column of \( L \) yielding \( Y_0(L) \). Lastly, we added random corruptions from a uniform \([0, 100]\) distribution to a random \( \eta \) fraction of observed entries in each row and column of \( Y_0 \) to obtain matrix \( Y_0(M) \), the input to our algorithms. We varied rank \( r \), sampling probability \( p \), and corruption fraction \( \eta \) at a time while keeping all other parameters fixed, and evaluated the change in reconstruction error, measured as root mean squared error (RMSE), for each of the three algorithms.

Conclusions
We presented a matrix completion method that is robust to arbitrary corruptions in the data. At the core of our method is robust gradient descent, where we robustly estimate the mean of the sample gradients at each gradient update. We empirically showed that our method recovers the true matrix better than the L2 and L1 norms in the alternating minimization scheme.

Acknowledgements
I would like to thank Adarsh Prasad and Professor Pradeep Ravikumar for giving me the opportunity to work on this project and for advising me through the research process. Their support and guidance was an essential part of making this project a reality.

References