

Convex and Non-Smooth Optimization
Spring, 2018
Project Report
Peter Will

Introduction

The Matrix Completion Problem consists of reconstructing the unknown or missing entries of a matrix from a relatively small set of its entries. This problem features prominently in the design of recommendation systems, where a company may wish to make personalized item recommendations to its users on the basis of historical user-item feedback, as in the well-publicized NetFlix Competition. In the general case, a full-rank matrix $\in \mathbb{R}^{m \times n}$ will have $m \times n$ degrees of freedom, and it is impossible to reconstruct it from only a sample of its entries. On the other hand, if a matrix is low rank, then sampled entries in a sense "encode" information about the full matrix, and accurate reconstruction becomes tractable. In 2008, Candes and Recht extended theoretical conditions under which perfect recovery of the original matrix was probabilistically bound. Among other things, the authors showed that a low rank matrix solution could be obtained via the convex relaxation:

$$\begin{aligned} \min \quad & \|X\|_* \\ \text{s.t.} \quad & X_{ij} = M_{ij}, \quad (i, j) \in \Omega \end{aligned} \tag{1}$$

Where $\|X\|_*$ is the nuclear norm of the low-rank reconstruction matrix X and Ω is the set of indices over which the elements of X are required to match those of the original matrix M . An interesting variation of this problem is the noisy matrix completion problem.

$$\begin{aligned} \min \quad & \|X\|_* \\ \text{s.t.} \quad & \|X_\Omega - M_\Omega\|_F \leq \delta \end{aligned} \tag{2}$$

In 2, the sampling constraint is relaxed so that entries of the original matrix are only required to be " δ -close" to those of the low-rank reconstruction. This formulation of the matrix completion problem is of real-world importance, as there is often a source of error in the observed values of the original matrix. Alternatively, if the original matrix is not itself low-rank, one may still be interested in finding an approximate low-rank representation for it.

The convex formulation of the matrix completion problem shown in 1 and 2 has spurred much investigation into performant algorithmic solutions. Candes and Recht demonstrated that the problems could be cast as semi-definite programs, and are therefore able to be solved by general interior point methods. However, these methods scale poorly and are therefore unsuitable for application to large datasets. Consequently, in recent years many algorithms have been proposed. Specifically, there has been a resurgence of interest in Alternating Direction Method algorithms for the matrix completion problem owing to its decomposable structure and analytically solvable sub problems.

ADM for MCC

Papers by Yang and Yuan (2009) and Tao and Yuan (2009) have studied the application of ADM methods to the noisy matrix completion problem and their results are summarized below. In the following, the noisy version of the problem given in 2 is first rewritten in ADM form. Expressions for the augmented Lagrangian and resulting sub-problems are then derived. Finally, the analytical solutions to these sub-problems along with explicit ADM iterate-updates are produced.

We begin by introducing an auxillary variable Y into 2 and writing the problem in ADM form as.

$$\begin{aligned} \min \quad & \|X\|_* \\ \text{s.t.} \quad & X - Y = 0 \\ & X \in \mathbb{R}^{m \times x}, Y \in \mathbf{B}_\delta \end{aligned} \quad (3)$$

where for $\delta \geq 0$

$$\mathbf{B}_\delta := \{U \in \mathbb{R}^{m \times n} : \|U_\Omega - M_\Omega\|_F \leq \delta\} \quad (4)$$

represents the convex subspace of $\mathbb{R}^{m \times n}$ over which the relaxed sampling constraints are satisfied. The augmented Lagrangian of 10 is given by

$$\mathcal{L}(X, Y, Z, \beta) := \|X\|_* - \langle Z, X - Y \rangle + \frac{\beta}{2} \|X - Y\|_F^2 \quad (5)$$

Where Z is the Lagrangian multiplier and $\beta \geq 0$ is the penalty paramater. At a high level, ADM works by alternating between minimizing 5 with respect to the variables X and Y . Because the objective function of 10 is decomposable and is constrained over a convex set, each of the resulting sub-problems can be solved analytically. The actual iterative scheme for ADM of 10 is as follows:

$$X^{k+1} = \arg \min_X \mathcal{L}(X^k, Y^k, Z^k, \beta) \quad (6a)$$

$$Y^{k+1} = \arg \min_Y \mathcal{L}(X^{k+1}, Y^k, Z^k, \beta) \quad (6b)$$

$$Z^{k+1} = Z^k - \beta (X^{k+1} - Y^{k+1}) \quad (6c)$$

Optimality conditions require that for 6a

$$0 \in \partial(\|X^k\|_*) - [Z^k - \beta(X^k - Y^k)] \quad (7)$$

Cai, Candes, Shen (2008) showed 7 can be solved via singular value shrinkage computation. More specifically, let

$$A^k = \frac{1}{\beta} Z^k + Y^k$$

And denote its singular value decomposition as

$$A^k = U^k \Sigma^k (V^k)^T$$

7 can be rewritten as

$$0 \in \frac{1}{\beta} \|X^k\|_* + X^k - A^k \quad (8)$$

Which is satisfied when

$$X^k = U^k \hat{\Sigma}^k (V^k)^T \quad (9)$$

Where $\hat{\Sigma}^k$ is the matrix of singular values given by $\text{diag}(\{(\sigma_i^k - \frac{1}{\beta})_+\})$. Thus, an analytical solution to update given in 6a exists. It requires a singular value decomposition of the matrix X which generally requires a $\mathcal{O}(n^3)$ complexity cost. This ultimately ends up dominating the ADM's algorithm costs.

Turning now to the sub-problem of 6b it is clear the minimization problem reduces to

$$\begin{aligned} \arg \min_Y \quad & \|X - Y\|_F^2 \\ \text{s.t.} \quad & Y \in \mathbf{B}_\delta \end{aligned} \quad (10)$$

That is, the minimizer Y is the euclidean projection of X onto \mathbf{B} .