October 5, 2019

Summary of Results

Two relative errors are evaluated as the metric for comparing various algorithms. They are defined as

$$Error_1 = \frac{\|\mathcal{P}_\Omega(\hat{M} - M^*)\|_F}{\|\mathcal{P}_\Omega(M^*)\|_F},$$
$$Error_2 = \frac{\|\hat{M} - M^*\|_F}{\|M^*\|_F}.$$

Let M^* be a 10-by-10 matrix of rank 5 with entries generated independently from the standard Normal distribution, M_{obs} has uniformly selected 20% missing entries. Table 1 summarizes the gap between the two errors as well as the time cost for each algorithm to reach such accuracy.
The hyperparameters of each algorithm are listed below.

- Vannila GD: learning rate $= 0.04$, # iteration $= 10000$;
- Regularized GD: learning rate $= 0.04$, # iteration $= 10000$, $C_t = 5, C_d = 6500, C_1 = 0.4, C_2 = 1$;
- Projected GD: learning rate $= 0.04$, # iteration $= 10000$, $c = 1$, $\mu = 1$;
- GD on Manifold: # iteration $= 10000$,
  - Optimization over S: learning rate $= 1$, # iteration $= 50$;
- AltMin: learning rate $= 0.06$, # iteration $= 100$,
  - Optimization over L and R: # iteration $= 100$.

From the table, it is showed that (i) Regularized GD and Projected GD need well-selected tuning parameters in order to perform well as stated in the theorems, (ii) GD on manifold is computationally expensive due to the SVD step and the optimization over S, (iii) there is a gap of accuracy between the observed matrix and the true matrix.

Table 1: Comparison of Algorithms

	Vanilla GD	Regularized GD	Projected GD	GD on Manifold	AltMin
Error1	$4.24 \cdot 10^{-5}$	$4.24 \cdot 10^{-5}$	$4.24 \cdot 10^{-5}$	$1.71 \cdot 10^{-5}$	$1.27 \cdot 10^{-5}$
Error2	0.0010	0.0010	0.0010	0.0004	0.0002
Time(s)	0.27	4.79	7.81	68.72	0.27

However, when the missing fraction increases to 30%, none of the algorithms is able to recover the true matrix (Table 2). Note that

$$\frac{\|M_{obs} - M^*\|_F}{\|M^*\|_F} = 0.57.$$

Table 2: Results of Algorithms with missing fraction 30%

	Vanilla GD	Regularized GD	Projected GD	GD on Manifold	AltMin
Error1	0.00041	0.00041	0.00041	0.00840	0.00236
Error2	0.77	0.77	0.77	0.85	0.74
Time(s)	0.27	4.85	8.24	68.72	0.25

Moreover, Figure 1 shows that as iteration number grows, the predictor is closer and closer to M_{obs} but further and further away from M^*.

Figure 1: Change of error1 using Vanilla GD