

# Report on Visualization Project

Daiqi Linghu, Dunzhu Li, Yiran Ma

## 1 Introduction

## 2 Algorithm

### 2.1 Optimization

In the code (BSGD.m), we solve the following optimization problem:

$$\operatorname{argmin}_{U,V,a,b} \frac{\lambda}{2} (\|U\|^2 + \|V\|^2 + \|a\|^2 + \|b\|^2) + \sum_{(i,j) \in S} (Q_{i,j})^2$$

where  $Q_{i,j} = (u_i^T v_j + a_i + b_j + \mu - Y_{i,j})$ . Then,

$$\frac{\partial J}{\partial U_l} = \lambda U_l + \sum_{(i,j) \in S} 2Q_{i,j} 1_{i=l} V_j$$

### 2.2 Visualization

## 3 Implementation

### 3.1 Find U & V

To solve the optimization problem with Bias-SGD, we simply divide the data by 5 fold, use 4 folds as training data and 1 fold as test data to find the optimal damping parameter  $\lambda$  and number of iteration. The step size  $\eta$  is chosen as 1e-2 and is decreased by 10% for each iteration. From the results shown below (Fig. 1), we choose  $\lambda = 25$ . For  $\lambda = 25$ , the out-of-sample error is still decreasing at the 100th iteration (Fig. 1), however, the improvement is minimal. Therefore, we choose the number of iteration as 100.

### 3.2 Projection

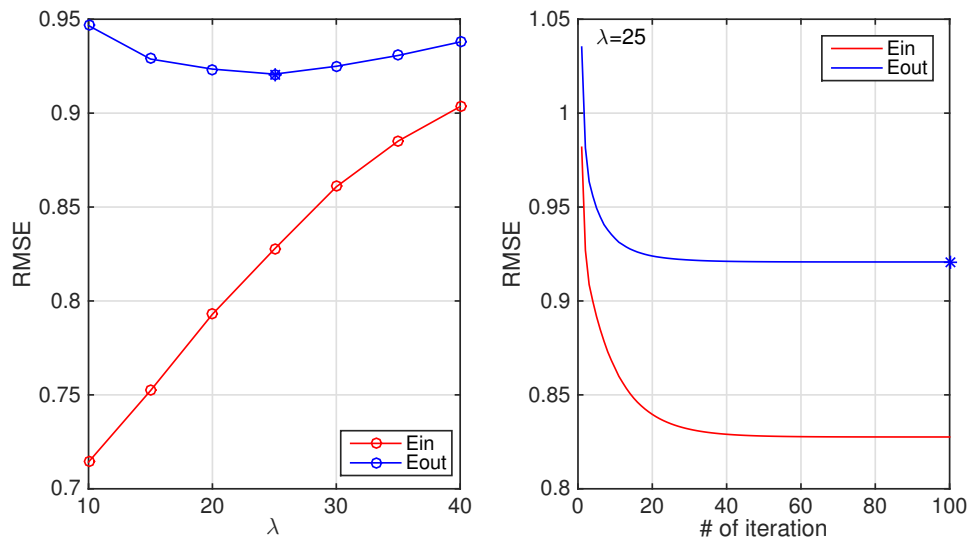


Figure 1: RMS error vs. Parameter