Non-negative matrix factorization and Robust PCA

Lingyu Zhang

Mar 28, 2018

Theoretical part

0.1 Projected Gradient Descent for non-negative matrix factorization

We firstly define the convex set C for projection as below:

$$C = \{W \in R_+^{m*r} : ||W_j||_2 \leq 1, j = 1, 2, ...r\}$$

Therefore, problem (1) is equivalent to

$$\min_{W,H} \frac{1}{2}||WH^T - X||_F^2 + I_c(W)$$

The projection operator onto C can be written as

$$P_c(x) = \arg\min_{z \in C} ||x - z||^2$$

We use Frobenius norm to measure the approximation

$$\frac{1}{2}||WH^T - X||_F^2 = \frac{1}{2}\sum_{i,j}(WH^T - X)_{ij}^2$$

Therefore, at k-th iteration, we update the parameter as below:

$$W^{k+1} = P_c\Bigg(W^k - \eta\nabla\Big(\frac{1}{2}||W^kH^T - X||_F^2\Big)\Bigg) = P_c\Bigg(W^k - \eta(W^kH^T - X)H\Bigg)$$

0.2 Alternating minimization for non-negative matrix factorization

The objective function here is

$$\min_{W,H} \frac{1}{2}||WH^T - X||_F^2 + I_c(W)$$

We firstly initialize W and compute H, at k-th iteration

$$H^{k+1} = \arg\min_{H} ||W^k H^T - X||_F^2 \text{ , subject to } H \geq 0$$

$$W^{k+1} = \arg\min_{W} ||W(H^{k+1})^T - X||_F^2 \text{ , subject to } W \geq 0$$

We then use accelerated proximal gradient for the sub-problem.
For sub-problem 1, the objective function is

$$f_1(H) = ||W^k H^T - X||_F^2, \text{ subject to } H \geq 0$$

We split the objective function according to the colomns of H

$$f_1(H) = ||W^k H^T - X||_F^2 = \sum_{c=1}^{n} ||W^k H(:,c)^T - X(:,c)||_2^2$$

Then we apply accelerate proximal gradient for each column of H
For sub-problem 2, the objective function is

$$f_2(W) = ||W^k H^T - X||_F^2, \text{ subject to } W \geq 0$$

We split the objective function according to the rows of W

$$f_2(W) = ||W^k H^T - X||_F^2 = \sum_{i=1}^{m} ||W(i,:)^k H^T - X(i,:)||_2^2$$

Then we apply accelerate proximal gradient for each row of W

0.3 Alternating proximal gradient for non-negative matrix factorization

We firstly initialized W and H, at each iteration, we update the parameters as below:

$$W^{k+1} = \arg\min_{W} ||W(H^k)^T - X||_F^2 \text{ , subject to } W \geq 0$$

$$(H^T)^{k+1} = \arg\min_{H} ||W^{k+1} H^T - X||_F^2 \text{ , subject to } H \geq 0$$

Therefore,

$$W^{k+1} = \max\Big(0, W^k - \frac{\nabla_W F(W^k, (H^t)^k)}{L_{W^k}}\Big)$$

$$(H^T)^{k+1} = \max\Big(0, (H^T)^k - \frac{\nabla_{H^T} F(W^{k+1}, (H^t)^k)}{L_{(H^T)^k}}\Big)$$

For Lipschitz constants, we use spectral norm of $W^T W$ and $H^T * H$ to calculate it.

$$L_{W^k} = \sqrt{max\Bigg(eigen\Big((W^k)^T * W^k\Big)\Bigg)}$$

$$L_{(H^T)^k} = \sqrt{max\Bigg(eigen\Big((H^k)^T * H^k\Big)\Bigg)}$$

0.4 Proximal gradient for robust PCA

At each iteration, we update the parameters as below:

$$\tilde{L}^k = L^k + \eta(L^k - L^{k-1})$$

$$\tilde{S}^k = S^k + \eta(S^k - S^{k-1})$$

$$Y_L^k = \tilde{L}^k - \frac{\beta}{2}(\tilde{L}^k + \tilde{S}^k - X)$$

$$(U, S, V) = svd(Y_L^k)$$

$$L^{k+1} = U(S - \frac{1}{2}I)_+V^*$$

$$Y_S^k = \tilde{S}^k - \frac{\beta}{2}(\tilde{L}^k + \tilde{S}^k - X)$$

$$S^{k+1} = sign(Y_S^k)(|Y_S^k| - \frac{\lambda_1}{2}11^*)$$

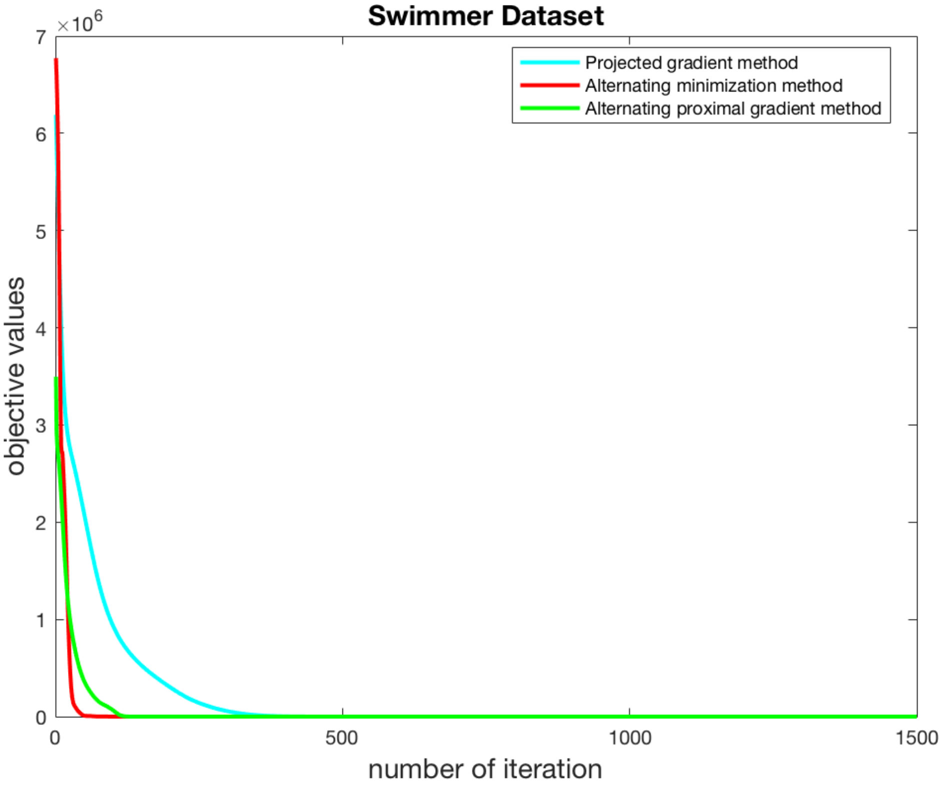
Experimental part

Compare Projected gradient, Alternating minimization and Alternating proximal gradient for NMF

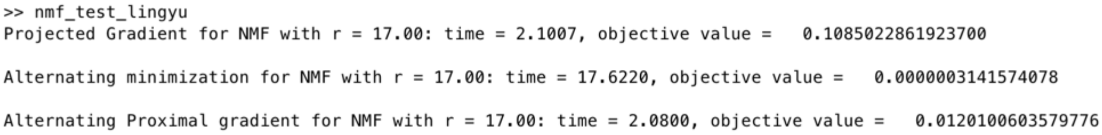
Firstly for data preparation, we reshape the Swimmer dataset to $[1024, 256]$ We initialize the parameters, the initial value of matrix H and W are randomized generated by normal distributed numbers. We set $r = 17$ which means that the dimension of H is $17 * 256$ and the dimension of W is $1024 * 17$
For projected gradient, we set the hyperparameters as below:

$$\text{learning rate } \eta = 0.0001$$

For alternating minimization, we applied Accelerated proximal gradient for the subrountine.
We run the 1000 iterations to do the compairson.
Below is the objective values plot for Projected gradient, Alternating minimization and Alternating proximal gradient for NMF.



We can find that the objective value is going down with the training iteration increases and finally converged. Below is the time and objective value compairson for 1000 iterations.



```
>> nmf_test_lingyu
Projected Gradient for NMF with r = 17.00: time = 2.1007, objective value =   0.1085022861923700

Alternating minimization for NMF with r = 17.00: time = 17.6220, objective value =   0.0000003141574078

Alternating Proximal gradient for NMF with r = 17.00: time = 2.0800, objective value =   0.0120100603579776
```

We also visualized each column of W as image, in order to do it, we reshape each column of W as $32 * 32$. We can find that each column of W represent a different limbs which is in different directions.
Below is the visualized columns of W learnt by alternating minimization algorithm.
Below is the visualized columns of W learnt by alternating proximal gradient algorithm.

Compare the performance of Alternating minimization and proximal gradient for Robust PCA

For data preparation, we firstly reshape the data to $20800 * 200$ dimension.We initialize the parameters, the initial value of matrix L and S are randomized generated by normal

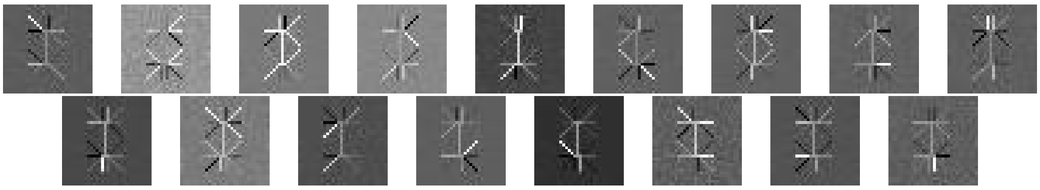


Figure 1: Visualization of all the columns of W by alternating minimization

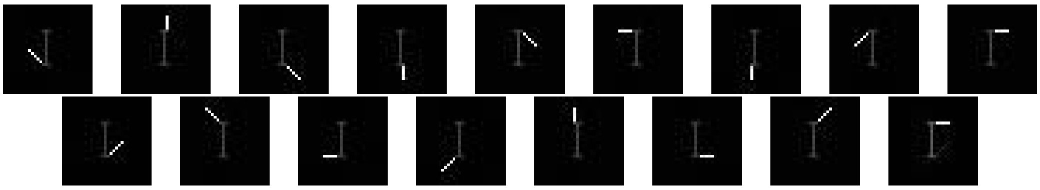


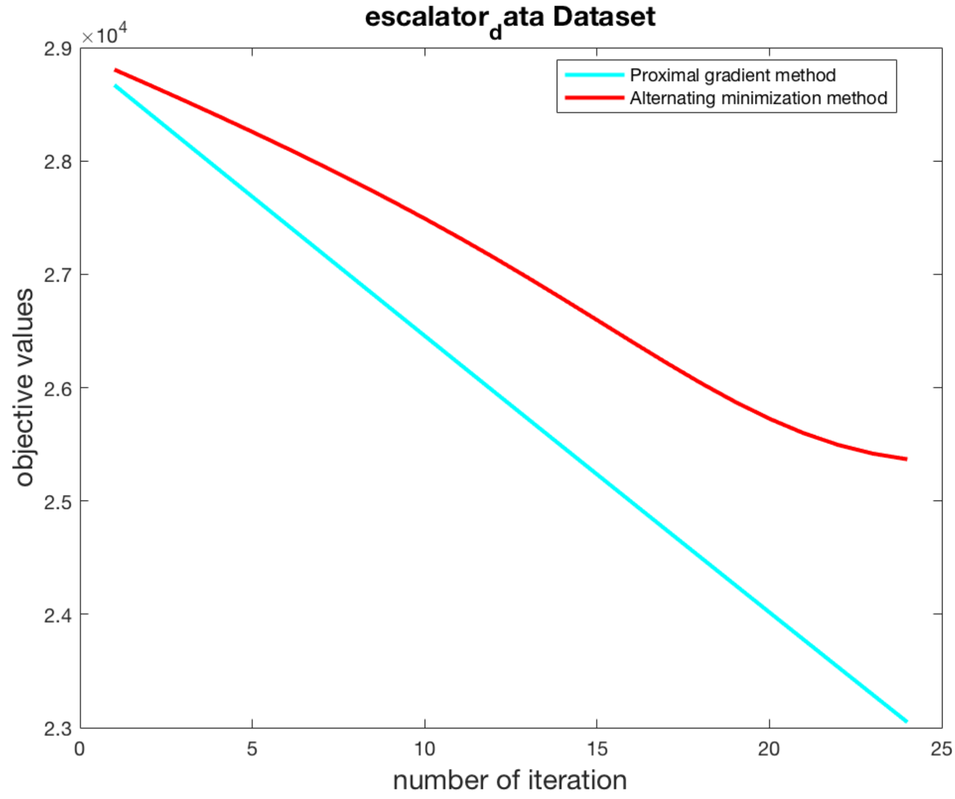
Figure 2: Visualization of all the columns of W by alternating proximal gradient

distributed numbers. We set the hyperparameters as below:

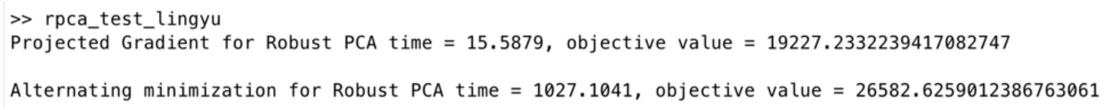
$$\beta = 1e - 15;$$

$$\lambda = 5e - 3;$$

Below is the objective values plot for stochastic gradient over iterations on both objective functions.



We can find that the objective value is going down with the training iteration increases and finally converged. Below is the time and objective value compairson for 20 iterations.



```
>> rpca_test_lingyu
Projected Gradient for Robust PCA time = 15.5879, objective value = 19227.2332239417082747

Alternating minimization for Robust PCA time = 1027.1041, objective value = 26582.6259012386763061
```