

1. Do **matrix completion** in two ways:

(a) Use the low-rank factorization model  $A \approx XY^\top$  and the objective function of the form

$$F(X, Y) = \frac{1}{2} \|P_\Omega(A - XY^\top)\|_F^2 + \frac{\lambda}{2} (\|X\|_F^2 + \|Y\|_F^2).$$

Try values of  $\lambda$  0.1, 1, and 10, and  $\text{rank}(X) = \text{rank}(Y) = k$ ,  $k = 1, 2, \dots, 7$ . Find  $X$  and  $Y$  using alternating iteration

$$X^{m+1} = \arg \min_X F(X, Y^m), \quad (1)$$

$$Y^{m+1} = \arg \min_Y F(X^{m+1}, Y). \quad (2)$$

Each of these steps can be further decomposed into a collection of small linear least squares problems. For example, at each substep of (1), we solve the linear least squares problem to compute the row  $i$  of  $X$ :

$$\mathbf{x}_i^\top = \arg \min_{\mathbf{x}} \frac{1}{2} \left\| \mathbf{x}^\top Y_{\Omega_i}^\top - a_{\Omega_i} \right\|_2^2 + \frac{\lambda}{2} \|\mathbf{x}\|^2, \quad (3)$$

where  $\Omega_i := \{j \mid (i, j) \in \Omega\}$ ,  $Y_{\Omega_i}^\top$  is the set of columns of  $Y^\top$  with indices in  $\Omega_i$ , and  $a_{\Omega_i}$  is the set of known entries of  $A$  in its row  $i$ . A similar problem can be set up for each column of  $Y$ . Work out solutions to these problems in a manner similar to the one in Sections 5.3.1 and 5.3.2 of `LinearAlgebra.pdf` except that there will be no constraint requiring the entries to be positive. Implement the resulting algorithm. Comment on how the value of  $\lambda$  and the choice of rank affects the result. Which values of the rank and  $\lambda$  seem the most reasonable to you? You can judge by your own row.

**Solution:** We will update the row  $i$  of  $X$  by solving the linear system

$$(Y_{\Omega_i}^\top Y_{\Omega_i} + \lambda I) \mathbf{x}_i = Y_{\Omega_i}^\top a_{\Omega_i}$$

and likewise for row  $j$  of  $Y$ ,

$$(X_{\Omega_j}^\top X_{\Omega_j} + \lambda I) \mathbf{y}_j = X_{\Omega_j}^\top a_{\Omega_j}$$

suggesting the code [here](#). We find that smaller values of  $\lambda$  achieve smaller values of  $\|P_\Omega(A - M)\|_2$ ; on other hand, the larger the value of  $k$ , the smaller the value of  $\|P_\Omega(A - M)\|_2$ . Of the values of  $\lambda$  and  $k$  suggested,  $\lambda = 0.1$  and  $k = 7$  produces the most reasonable result with respect to the norm over the known values of  $A$ .

(b) Use the approach of penalizing the nuclear norm in Section 6.3 of `LinearAlgebra.pdf` and the iteration

$$M^{j+1} = S_\lambda(M^j + P_\Omega(A - M^j))$$

Experiment with different values of  $\lambda$ .

**Solution:** We implement the method of penalizing the nuclear norm via the code [here](#). We find that the smaller the value of  $\lambda$ , the smaller the norm  $\|P_\Omega(A - M)\|_2$ .

Compare these two approaches for matrix completion. Which one gives more sensible results? Which one is easier to use? Which one do you find more efficient?

**Solution:** The method of penalizing the nuclear yields both better results and was much easier to implement and reason about.

2. Adding the nuclear norm into the objective function for the matrix completion problem promotes low rank. The reason is that the level set of  $\|A\|_* = a$  tend to have corners, and the minimizers of the objective function including the nuclear norm penalty term tend to lie at these corners. Check experimentally that nuclear norm level set for a  $2 \times 2$  matrix

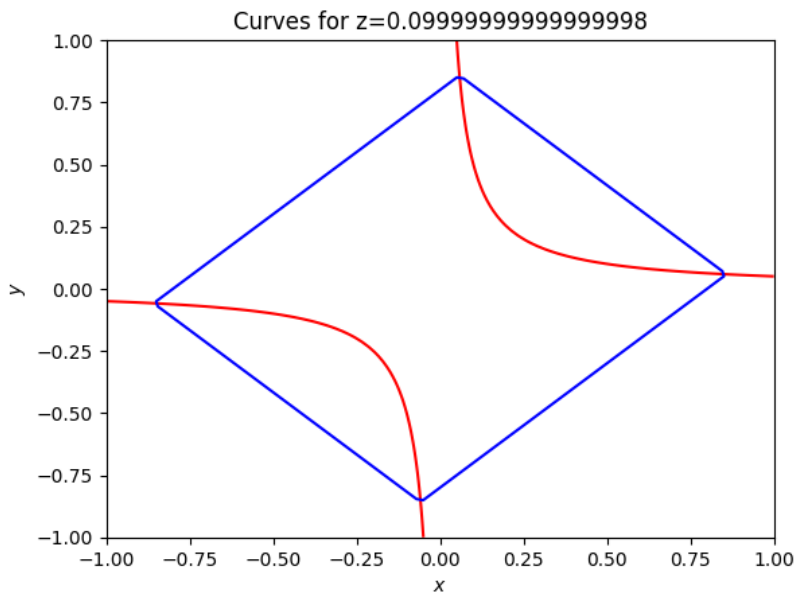
$$A = \begin{bmatrix} 0.5 & x \\ y & z \end{bmatrix}$$

have corners at points where  $\det A = 0$ . Proceed as follows. Let  $a = 1$ . Set a grid for the values of  $z$ :

$$z \in \{-0.4, -0.3, -0.2, -0.1, 0, 0.1, 0.2, 0.3, 0.4\}$$

For each grid value of  $z$  plot two curves on the same figure. Curve 1: the zero level set of the determinant of  $A$ . Curve 2: the level set  $\|A\|_* = a$ . Note at which points these curves intersect.

**Solution:** We plot the curves for all specified values of  $z$  via [this code](#). We plot the case of  $z = 0.1$  below.



Note that the points of intersection are at the corners of the level set  $\|A\|_* = 1$ . This observation holds for the other specified values of  $z$  as well.

3. Let  $A$  be a symmetric positive definite matrix and let  $\{p_0, \dots, p_{n-1}\}$  be a set of vectors such that  $p_j^\top A p_k = 0$  for all  $j \neq k$ ,  $0 \leq j, k \leq n-1$ . We say that this set of vectors is *conjugate* with respect to the matrix  $A$ . Prove that this set of vectors must be linearly independent.

**Solution:** Let  $\{p_0, \dots, p_{n-1}\}$  be conjugate, so that  $p_j^\top A p_k = 0$  for all  $j \neq k$ , and toward a contradiction, suppose  $\{p_0, \dots, p_{n-1}\}$  is linearly dependent.

Then there exist coefficients  $c_0, \dots, c_{n-1}$  not all zero such that

$$c_0 p_0 + \dots + c_{n-1} p_{n-1} = \vec{0}. \quad (4)$$

In particular, we can choose  $k$  such that  $c_k \neq 0$ . Then left multiplying (4) by  $p_k^\top A$ , on the right-hand side we see that  $p_k^\top A \vec{0} = 0$ , but on the left-hand side, by conjugacy we have that

$$p_k^\top A (c_0 p_0 + \dots + c_{n-1} p_{n-1}) = c_k p_k^\top A p_k$$

and hence  $c_k p_k^\top A p_k = 0$ . Finally, since  $A$  is symmetric positive definite,  $p_k^\top A p_k > 0$ , and hence  $c_k = 0$ , contradicting  $c_k \neq 0$ ; since our assumption led to a contradiction, it must be that  $\{p_0, \dots, p_{n-1}\}$  is linearly independent.