Convex Assignment 1

Subhankar Chakraborty

May 15 2020

Introduction

A few points to note for this report.

- Codes for some questions have been included in the report. Not everything has been included for the sake of verbosity.
- The screenshots of the MATLAB command line outputs are attached with the zipped file only. They have been omitted from the report for the sake of neatness.
- Each question along with its plots, screenshots and the files it depends on is present in its own folder.

Question 1

It can be easily seen that the following problem can be reduced to the form

$$\underset{z\in\mathbb{R}^3}{\text{minimize}} \qquad \|\mathbf{Az}-\mathbf{b}\|^2$$

with the matrices A, b and z given by

$$\mathbf{A} = \begin{bmatrix} 2\mathrm{x}_1 & 2\mathrm{y}_1 & 1 \\ 2\mathrm{x}_2 & 2\mathrm{y}_2 & 1 \\ . & . & . \\ . & . & . \\ 2\mathrm{x}_m & 2\mathrm{y}_m & 1 \end{bmatrix}$$

$$\mathbf{b} = \begin{bmatrix} \mathrm{x}_1^2 + y_1^2 \\ \mathrm{x}_2^2 + y_2^2 \\ . \\ . \\ \mathrm{x}_m^2 + y_m^2 \end{bmatrix}$$

$$\mathbf{z} = \begin{bmatrix} \mathrm{x}_c \\ \mathrm{y}_c \\ \mathrm{r} \text{ - } \mathrm{x}_c^2 + y_c^2 \end{bmatrix}$$

Note that we do not need to specifically impose that r> 0 as it can be shown mathematically that for any choice of the coordinates of centre, the best fitting radius is always positive. The code for this question is mentioned below.

```
circle_fit;
number_of_variables=3;
A = ones(length(x), number_of_variables);
A(:,1) = 2*x;
A(:,2) = 2*y;
b = x.^2+y.^2;
cvx_begin
    variable p %xc
    variable q %yc
    variable r %c**2-xc**2-yc**2
    minimize (norm(A*[p;q;r]- b,2))
cvx_end
```

The best centre coordinates are **[-2.5671, 6.4680]** and the best value of the radius is **1.3214**. The plot of the best fit circle is shown in figure 1.

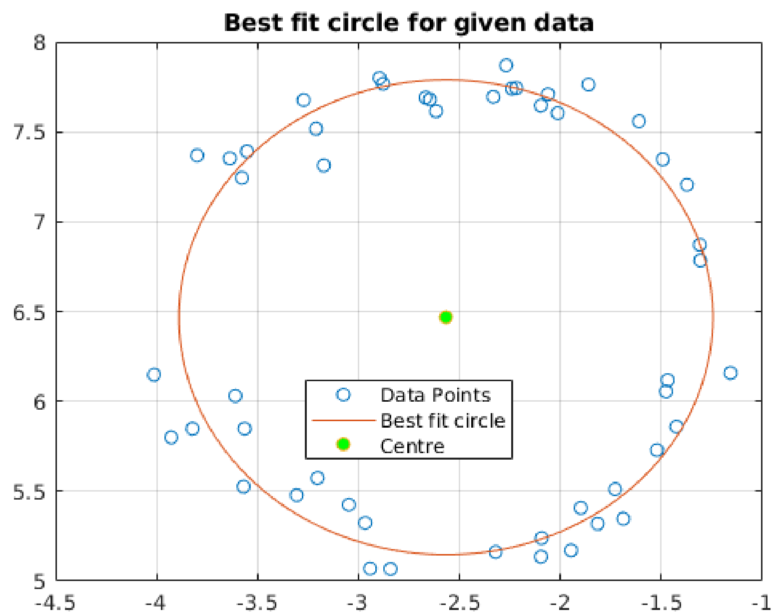

Figure 1: Best fit circle

Question 2

Gradient Descent with fixed step size of 0.01

The code is mentioned below.

```
x0=[-1; 0.7];
x = x0;
all_x=x0;
step_size=0.1;
niters=0;
while norm([exp(x(1)+3*x(2)-0.1)+exp(x(1)-3*x(2)-0.1)\
-exp(-x(1)-0.1); 3*exp(x(1)+3*x(2)-0.1)-
\3*exp(x(1)-3*x(2)-0.1)],2)>0.01
    del_x = -[exp(x(1)+3*x(2)-0.1)+\
    exp(x(1)-3*x(2)-0.1)-exp(-x(1)-0.1);
    3*exp(x(1)+3*x(2)-0.1)-3*exp(x(1)-3*x(2)-0.1)];
    x = x+step_size*del_x;
    all_x = [all_x, x];
    niters=niters+1;
end
```

Some observations are mentioned below.

- The optimal value of the objection function is **2.5593**
- The optimal value is attained at [x1, x2] = [-0.3497, 0]
- The number of iterations required for convergence is 19

The convergence plot is shown in figure 2.

Gradient descent with backtracking line search

Some observations are mentioned below.

- The optimal value of the objection function is **2.5593**
- The optimal value is attained at [x1, x2] = [-0.3440, -0.0006]
- The number of iterations required for convergence is 11

The convergence plot is shown in figure 3.

Newton's method with backtracking line search

Some observations are mentioned below.

- The optimal value of the objection function is **2.5593**
- The optimal value is attained at [x1, x2] = [-0.3456, -0.0005]

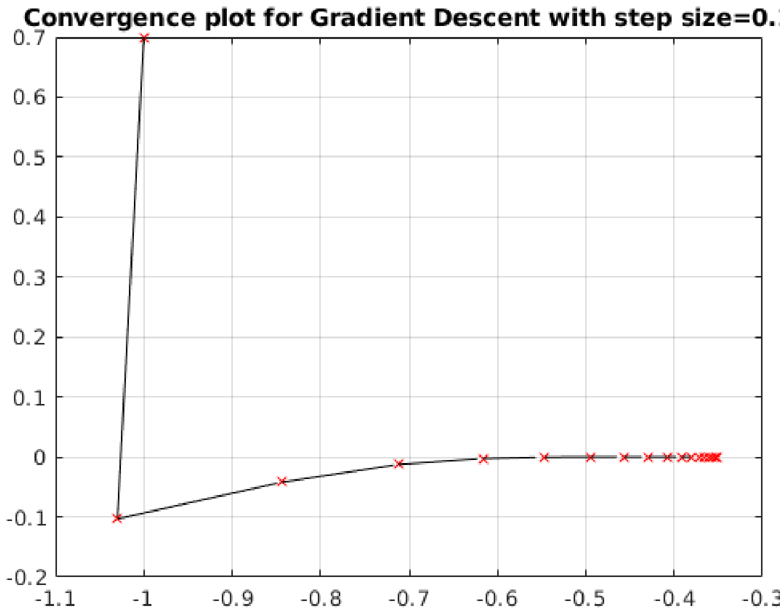

Figure 2: Convergence Plot

- The number of iterations required for convergence is 3
- Note that Newton's method converges in much fewer iteration than the other two methods. One issue is, however, the calculation of the inverse of the hessian matrix which has a complexity of $\mathcal{O}(n^3)$ which can quickly become the limiting step when we have larger number of variables.

The convergence plot is shown in figure 4.

Question 3

(a)

Consider the two sets

1. $\{x_i | x_i \in 0, 1\}$;   i=1,2...,n
2. $\{x_i | x_i(1 - x_i) = 0\}$;   i=1,2...,n

Consider $x_i \in$ set 1 $\implies x_i = 0 \vee x_i = 1 \implies x_i(1 - x_i) = 0 \implies x_i \in$ set 2
Consider $x_i \in$ set 2 $\implies x_i(1 - x_i) = 0 \implies x_i = 0 \vee x_i = 1 \implies x_i \in 0, 1$ $\implies x_i \in$ set 1
Quite clearly, the sets 1 and 2 are equal. So the problems as mentioned in (5) and (6) in the questions PDF are equivalent.

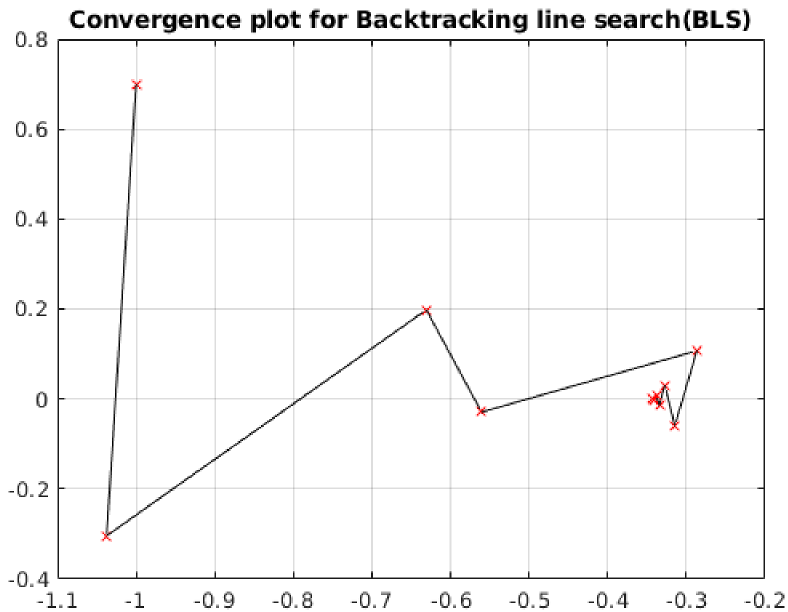

Figure 3: Convergence Plot

(b)

Quite clearly, problem (6) is not convex as the equality constraints are not affine.

(c)

The primal problem is given by

$$\begin{aligned} \underset{x \in \mathbb{R}^n}{\text{minimize}} \quad & c^T x \\ \text{subject to} \quad & Ax \leq b \\ & x_i(1 - x_i) = 0,\ i = 1, \ldots, n \end{aligned}$$

The above problem can be reformulated as

$$\begin{aligned} \underset{x \in \mathbb{R}^n}{\text{minimize}} \quad & c^T x \\ \text{subject to} \quad & Ax \leq b \\ & (x_i - 1)x_i = 0,\ i = 1, \ldots, n \end{aligned}$$

The Lagrangian is defined as

$$L(x, \lambda, \mu) = c^T x + \lambda^T (Ax - b) + \sum_{i=1}^{n} \mu_i (x_i - 1) x_i$$

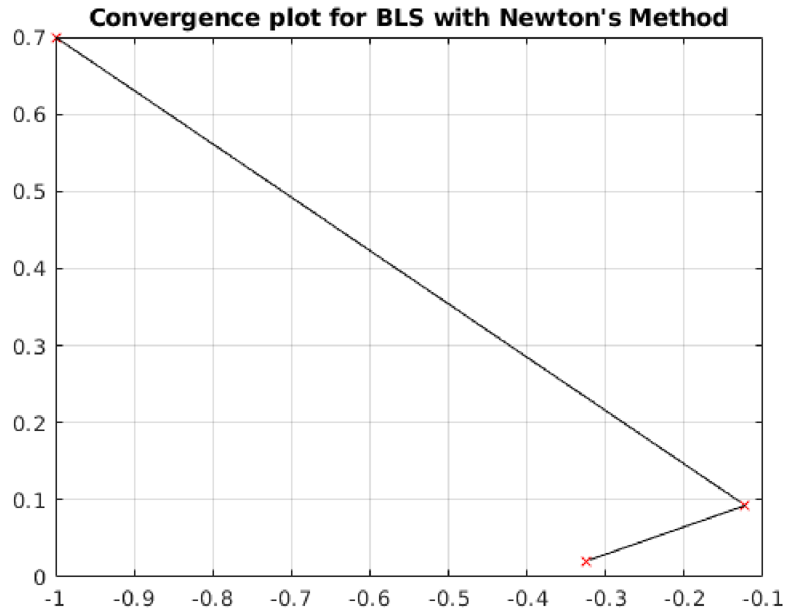

Figure 4: Convergence Plot

The dual function is given by

$$g(\lambda, \mu) = \inf_{x \in \mathbb{R}^n} \quad L(x, \lambda, \mu)$$
$$\text{subject to} \quad \lambda \geq 0$$

We can clearly see that $L(x, \lambda, \mu)$ is a quadratic in each x_i and hence, a non trivial infimum exists only if the coefficients of x_i's are positive. Also, if $\mu_i = 0$ for some i, the infimum is -∞ as we can set $x_j = 0$, $j \neq i$ and we have an affine function in x_i, whose infimum is -∞ except when $\lambda^T A_i + c_i = 0$, which need not always be the case.

$$\implies \mu > 0$$

We also notice that there are no cross terms in L, ie, each minimizer x_i can be found independent of the others. Quite clearly,

$$x_i^* = \frac{\mu_i - c_i - a_i^T \lambda}{2\mu_i}$$
$$i = 1, \ldots, n$$
$$a_i \in \mathbb{R}^m \text{ is the } i\text{th column of A.}$$

Substituting the corresponding values of x_i^*, we get

$$\mathrm{g}(\lambda, \mu) = -b^T \lambda \text{ -} \frac{1}{4} \sum_{i=1}^{i=n} \frac{(\mu_i - c_i - a_i^T \lambda)^2}{\mu_i}$$

$$\lambda \geq 0$$
$$\mu > 0$$

The dual objective can hence be defined as

$$\begin{aligned} \underset{\lambda \in \mathbb{R}^m, \mu \in \mathbb{R}^n}{\text{maximize}} \quad & g(\lambda, \mu) \\ \text{subject to} \quad & \lambda \geq 0 \\ & \mu > 0 \end{aligned}$$

Note that the -$\frac{1}{4}\sum_{i=1}^{i=n} \frac{(\mu_i - c_i - a_i^T\lambda)^2}{\mu_i}$ term in $g(\lambda, \mu)$ is always negative. Hence, the maximum for a given λ occurs when $\sum_{i=1}^{i=n} \frac{(\mu_i - c_i - a_i^T\lambda)^2}{\mu_i}$ is minimized with respect to μ as the other term is independent of μ.
Consider the following cases

1. $c_i + a_i^T\lambda > 0$
   Quite clearly, in this case we can make the second term zero. Hence g is maximized for $\mu_i = c_i + a_i^T\lambda$.

2. $c_i + a_i^T\lambda \leq 0$
   Clearly in this case we cannot set $\mu_i = c_i + a_i^T\lambda$ and make the second term zero as that would violate the $\mu > 0$ restriction. Consider the substitution $y_i = c_i + a_i^T$. Consider for a particular i

   $$\frac{(\mu_i - y_i)^2}{\mu_i}$$
   $$= \frac{y_i^2}{\mu_i} + \mu_i + \text{-}2y_i$$

   It is minimized at $\mu_i = -y_i$ and the value is given by $-4y_i = -4(c_i + a_i^T)$

Hence the problem becomes

$$\begin{aligned} \underset{\lambda \in \mathbb{R}^m}{\text{maximize}} \quad & -b^T\lambda + \sum_{i=1}^{i=n} min(c_i + a_i^T\lambda, 0) \\ \text{subject to} \quad & \lambda \geq 0 \end{aligned}$$

(d)

The dual problem is given by

$$\begin{aligned} \underset{\lambda \in \mathbb{R}^m}{\text{maximize}} \quad & -b^T\lambda + \sum_{i=1}^{i=n} min(c_i + a_i^T\lambda, 0) \\ \text{subject to} \quad & \lambda \geq 0 \end{aligned}$$

Equivalently,

$$\begin{aligned} \underset{\lambda \in \mathbb{R}^m}{\text{minimize}} \quad & b^T\lambda - \sum_{i=1}^{i=n} min(c_i + a_i^T\lambda, 0) \\ \text{subject to} \quad & -\lambda \leq 0 \end{aligned}$$

Note the following points.

- $b^T\lambda$ is convex in λ.
- $min(c_i + a_i^T\lambda, 0)$ is the pointwise infimum of two concave functions and is hence, concave.
- $\sum_{i=1}^{i=n} min(c_i + a_i^T\lambda, 0)$ is concave as is it the sum of concave functions.
- Hence, -$\sum_{i=1}^{i=n} min(c_i + a_i^T\lambda, 0)$ is convex in λ as it is the negation of a concave function.
- The sum of convex functions is convex. Hence, $b^T\lambda - \sum_{i=1}^{i=n} min(c_i + a_i^T\lambda, 0)$ is clearly a convex function in λ.
- The inequality constraints are affine and hence, they are convex.
- Hence, the problem is a convex optimization problem.

(e)

The matlab code for this section is mentioned below.

```
boolean;
m=size(A,1);
n=size(A,2);
summer = ones(n,1);
cvx_begin
    variables l(m)
    maximize (-b'*l + summer'*min(0, c+A'*l))
    subject to
        l>=0
cvx_end
```

The optimal value of the dual function is -3 and the optimal value occurs at λ=0 (vector containing m zeroes).

(f)

The optimal values of the dual, ie, d^* and the primal, ie, p^* are equal.

Question 4

Single Objective Optimization

The final optimization problem is

$$\underset{A\in\mathbb{R}^{nxn}, B\in\mathbb{R}^{nxm}}{\text{minimize}} \quad \sum_{\mathbf{t=1}}^{\mathbf{T-1}} \|\mathbf{x(t+1) - Ax(t) - Bu(t)}\|_\mathbf{2}^\mathbf{2}$$

The code for the given section is mentioned below.

```
y = xs(:,2:100);
x = xs(:,1:99);
n = size(x(:,1),1);
m = size(us(:,1),1);
cvx_begin quiet
    variables A(n,n)
    variables B(n,m)
    minimize (sum(sum_square_abs(y-A*x-B*us)))
cvx_end
```

The optimal value of the objective function is **844.8594**. The values of A and B are omitted here due to the sake of verbosity.

Multi Objective Optimization

This is a multi objective optimization problem. For different values of the paramter $\lambda > 0$, we get different Pareto Optimal Points. The Optimization Problem is formulated as mentioned below.

$$\underset{A\in\mathbb{R}^{nxn}, B\in\mathbb{R}^{nxm}}{\text{minimize}} \qquad \sum_{\mathbf{t=1}}^{\mathbf{T-1}} \|\mathbf{x(t+1) - Ax(t) - Bu(t)}\|_\mathbf{2}^\mathbf{2} + \lambda(\mathbf{card(A)} + \mathbf{card(B)})$$

This formalization, however, is not convex as card(X), ie, the number of non zero entries in matrix X is not a convex function in X. We can relax the optimization problem to a convex problem by using the L1 norm instead of card(X). The reformulated problem is

$$\underset{A\in\mathbb{R}^{nxn}, B\in\mathbb{R}^{nxm}}{\text{minimize}} \qquad \sum_{\mathbf{t=1}}^{\mathbf{T-1}} \|\mathbf{x(t+1) - Ax(t) - Bu(t)}\|_\mathbf{2}^\mathbf{2} + \lambda(\|\mathbf{A}\|_\mathbf{1} + \|\mathbf{B}\|_\mathbf{1})$$

The Pareto Optimal frontier for the optimization problem is mentioned in figure 5.

The chosen point on the frontier is marked in green. It was chosen as upto that value of Squared Error, we see significant drops in the value of non-zero entries. However, at that point, we see no change in the number of non-zero entries for an increase in squared error.

Question 5

The final optimization problem is

$$\underset{z\in\mathbb{R}^{K+1}}{\text{minimize}} \qquad \|\mathbf{A(x)_K^T z - sin(x)}\|_1$$

where

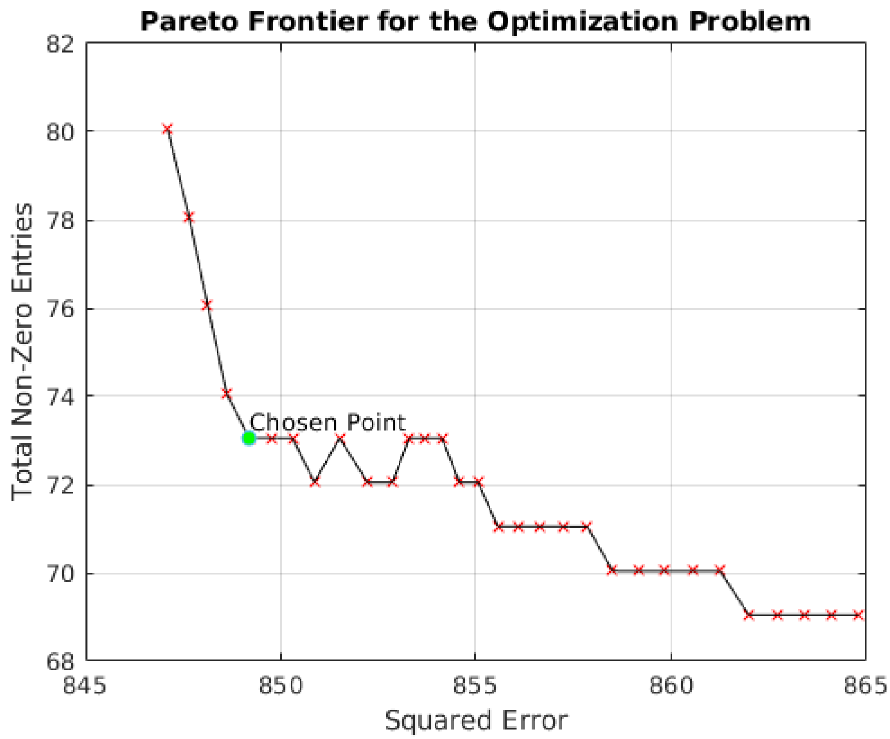

Figure 5: Pareto Optimal Frontier

- $$\mathbf{x} = \begin{bmatrix} \mathrm{x}_1 \\ \mathrm{x}_2 \\ . \\ . \\ \mathrm{x}_N \end{bmatrix}$$

  is a vector of the points over which the value of sin(x) is calculated.
- $A(x)_K$ is the Vandermonde Matrix of degree K for the given data points.
- $$\mathbf{sin(x)} = \begin{bmatrix} \sin(\mathrm{x}_1) \\ \sin(\mathrm{x}_2) \\ . \\ . \\ \sin(\mathrm{x}_N) \end{bmatrix}$$

The code for this question is mentioned below.

```
load('Q5_data.mat');
act_vals=sin(a);
A = flip(A,2);
cvx_begin quiet
    variables coefficients(K+1)
    minimize (norm(A*coefficients-act_vals', 1))
cvx_end
```

The optimal value of the objective function is **1.3121**. The plot is shown in figure 6.

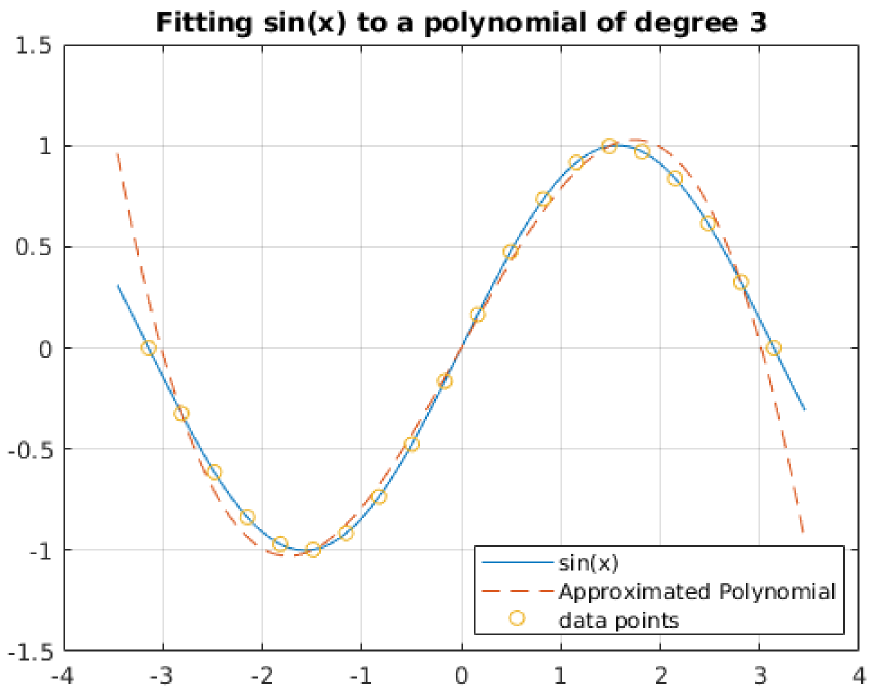

Figure 6: 3rd degree polynomial approximation to sin(x)

Question 6

The final formalization for the optimization problem is given by

$$\begin{aligned} \underset{C \in \mathbb{R}_+^{nxn}}{\text{minimize}} \quad & \|\mathbf{C} - \hat{\mathbf{C}}\|_{fro} \\ \text{subject to} \quad & C_{ii} = \hat{C}_{ii} = 1, \; i = 1, \dots, n \\ & C \succeq 0. \end{aligned}$$

This question is a typical case of an SDP. The code is mentioned below.

```
C_hat = [1 -0.76 0.07 -0.96; -0.76 1 0.18 0.07; 0.07
0.18 1 0.41; -0.96 0.07 0.41 1];
m = size(C_hat,1);
cvx_begin sdp
    variable C(m,m) semidefinite
    minimize (norm(C-C_hat, 'fro'))
    subject to
        C(1,1)==1
        C(2,2)==1
        C(3,3)==1
        C(4,4)==1
cvx_end
```

The optimal value of the objective function is **0.358138** and is attained at

$$\mathbf{C} = \begin{bmatrix} 1.0000 & -0.6427 & -0.0053 & -0.7871 \\ -0.6427 & 1.0000 & 0.1381 & 0.1661 \\ -0.0053 & 0.1381 & 1.0000 & 0.3483 \\ -0.7871 & 0.1661 & 0.3483 & 1.0000 \end{bmatrix}$$