Homework 4

Noé Hollande, Yujun Kim

June 10, 2022

Question 1:

First, the feasible set of P is not empty, for instance (x, y) with $x = (1, 0, ..., 0)$ and $y = (0, 1, 0, ..., 0)$ is feasible, now, suppose that (x_1, y_1),(x_2, y_2) are two feasible points, we'll that that P is convex, i.e. that $\lambda(x_1, y_1) + (1-\lambda)(x_2, y_2)$ is a feasible point for any $\lambda \in [0, 1]$. Indeed :, by linearity of the norm

$$(\lambda x_1 + (1-\lambda)x_2)^t(\lambda x_1 + (1-\lambda)x_2) = \lambda(x_1^t x_1) + (1-\lambda)(x_2^t x_2) = 1.$$

The same argument applies for the second condition, finally, by linearity once again :

$$(\lambda x_1 + (1-\lambda)x_2)^t(\lambda y_1 + (1-\lambda)y_2) = \lambda x_1^t y_1 + (1-\lambda)x_2^t y_2 = 0.$$

Therefore the convexity is proved.

Now for the LICQ, we derivate the three conditions to have $(-2x, 0), (0, -2y), (y, x)$, on a feasible point, x and y are of norm 1, therefore different than 0 and therefore orthogonal, hence they're linearly indepedent, therefore $(-2x, 0), (0, -2y), (y, x)$ are (there's no solution to $-2x = \lambda y$).

Question 2 :

The optimal value of the relaxed problem is $1/2(a_1 + b_1)$ where a_1, b_1 are respectively the minimal eigenvalues of A and B. This follows from Linear Algebra II, we know it's a finite value because the matrices are SPD, to attain it, we need $(x, y) = (v_{a_1}, v_{b_1})$ where those are the eigenvectors of a_1 and b_1 respectively of norm 1. The optimal value of the relaxed problem is better or equal to the one of the target problem (which is logical since it has less conditions), they're equal if and only if $v_{a_1}; v_{b_1}$ are orthogonal (note that there can be multiple candidates for eigenvectors of norm 1), since that will make them respect the third condition and therefore be feasible in the target problem.

Question 3 :

From now on, let $z = (x, y)$ for simpler notation.

The following expression works :

$$L(z, \mu) = \frac{1}{2} z^t \begin{bmatrix} A - 2\mu_1 & \mu_3 \\ \mu_3 & B - 2\mu_2 \end{bmatrix} z + \mu_1 + \mu_2$$

where μ_i designates the i-th coordinate of μ, the symmetry of A and B guarantees the symmetry of the matrix, that we'll designate by H from now on.

Question 4:

We write the dual function $L_D(\mu) = \inf L(z,\mu) = \inf\ (1/2)z^tHz + \mu^t \begin{bmatrix}1\\1\\0\end{bmatrix}$. For $L_D(\mu)$ to be finite, we need H to be SPD (which we can check with a computer if we have $A, B$and μ since they're the only parameters), then we'd have that $L_D(\mu) \geq \mu^t \begin{bmatrix}1\\1\\0\end{bmatrix}$ which proves it's finite.Then, the optimal value for $maxL_D(\mu)$ can be rephrased :

$$max\mu_1 + \mu_2 \text{ such that } H \text{ is SPD.}$$

If it's not SPD , then for any z such that $z^tHz = a < 0$, $\lambda z^tH\lambda z = \lambda^2 a < 0$, since λ can be arbitrarly large, the solution is not finite. Hence, this is a necessary and sufficient condition.

Question 5:

If H is not SPD, then there's no optimal solution, if it is, then we can just take $z = 0$ and get the solution $\mu_1 + \mu_2$, however, we know that this must be lesser or equal than $1/2(a_1 + b_1)$ due to weak duality and Q2 (here a_1 and b_1 represent the smallest eigeinvalues of A and B respectively); this makes sense since a necessary condition for H to be SPD is that $a_1 \geq 2\mu_1$ and $b_1 \geq \mu_2$ (otherwise, if $a_1 < 2\mu_1$ for instance, then $v_{a_1}^t H v_{a_1} = a_1 - 2\mu_1 < 0$ would be a contradiction).

Hence, $\mu_1 + \mu_2 \leq (1/2)a_1 + (1/2)b_1$, therefore weak duality is respected. In addition, the upper bound is reached if $\mu_3 = 0$, making the optimal solution : $\begin{bmatrix}\frac{a_1}{2}\\ \frac{b_1}{2}\\ 0\end{bmatrix}$

Question 6:

The existence of a KKT point condition is satisfied, indeed, we know from Q1 that the LICQ is verified, plus we're working on a compact set (set the norm of x and y must be equal to 1), which guarantees the existence of a local minimum where a constraint qualification holds, and therefore by Theorem 8.26, the existence of a KKT point.

However, we'll need the computer to find valid Lagrange multipliers, and we need to solve the problem to compute them, hence we can't guarantee this condition before computing the problem (and neither can we confirm condition number 2 since it need Lagrange's multipliers as arguments).

Question 7:

We get

$$\nabla_x L_\beta(x, y, \mu) = Ax - 2\mu_1 x + \mu_3 y - 2\beta x(1 - x^t x) + \beta y x^t y.$$

$$\nabla_y L_\beta(x, y, \mu) = By - 2\mu_2 y + \mu_3 x - 2\beta y(1 - y^t y) + \beta x x^t y$$

Question 8

Refer the attached code(also in the appendix of this file). alm.m generates the Lagrangian and its gradient.

Question 9

Refer the attached code(also in the appendix of this file). With given initial settings, we have the following results.

```
z_min =
   -1.3695
   -0.2535
   -0.0107
    0.0825
    1.4721
   -1.2693
    0.9636
   -1.0786
    0.7410
    0.5569

L_min = -26.4171
```

Here, `z = [x;y]`, so

$$x = \begin{bmatrix} -1.3695 \\ -0.2535 \\ -0.0107 \\ 0.0825 \\ 1.4721 \end{bmatrix}, y = \begin{bmatrix} -1.2693 \\ 0.9636 \\ -1.0786 \\ 0.7410 \\ 0.5569 \end{bmatrix}, L_{min} = -26.4171$$

Question 10

Each row of following is $z_k = [x_k \; y_k]$.

```
R1 =
    1.3349    0.4913   -0.4391    0.1087   -1.4509    0.9716   -0.7564    ...
       1.2123   -0.9557   -0.3235
    1.0365    0.4027   -0.3792    0.1000   -1.1276    0.7603   -0.5929    ...
       0.9649   -0.7630   -0.2499
    0.8481    0.3517   -0.3503    0.0978   -0.9235    0.6284   -0.4911    ...
       0.8132   -0.6450   -0.2035
    0.7350    0.3248   -0.3393    0.0989   -0.8012    0.5507   -0.4314    ...
       0.7253   -0.5764   -0.1761
    0.6708    0.3114   -0.3364    0.1007   -0.7318    0.5075   -0.3984    ...
       0.6767   -0.5383   -0.1610
    0.6360    0.3050   -0.3362    0.1022   -0.6941    0.4845   -0.3809    ...
       0.6508   -0.5180   -0.1530
    0.6177    0.3019   -0.3365    0.1031   -0.6743    0.4727   -0.3718    ...
       0.6374   -0.5074   -0.1489
    0.6083    0.3003   -0.3368    0.1036   -0.6642    0.4666   -0.3673    ...
       0.6306   -0.5020   -0.1468
    0.6036    0.2996   -0.3370    0.1038   -0.6591    0.4636   -0.3649    ...
       0.6272   -0.4992   -0.1458
```

At each iteration, f is calculated as following, converging around -6.4100(the last value).

```
R2 =
  -27.0433
  -16.6912
  -11.5137
   -8.9232
   -7.6269
   -6.9781
   -6.6536
   -6.4912
   -6.4100
```

Also, each row of the following is h calculated at each iteration. Notice that the row values converges to [0 0 0]. The norm of h converges to 0.

```
R3 =
   -3.3330   -3.0037    0.7584
   -1.6619   -1.5051    0.3890
   -0.8281   -0.7548    0.2002
   -0.4126   -0.3786    0.1027
   -0.2057   -0.1899    0.0523
   -0.1027   -0.0951    0.0265
   -0.0513   -0.0476    0.0133
   -0.0256   -0.0238    0.0067
   -0.0128   -0.0119    0.0033
```

Lastly, βh calculated at each iteration is shown in each row of the following. Here, theses values does not converges to 0, but rather kept stable.

```
R4 =
   -3.3330   -3.0037    0.7584
   -3.3239   -3.0102    0.7779
   -3.3124   -3.0192    0.8006
   -3.3009   -3.0291    0.8216
   -3.2916   -3.0376    0.8371
   -3.2853   -3.0436    0.8468
   -3.2816   -3.0473    0.8523
   -3.2795   -3.0493    0.8552
   -3.2785   -3.0504    0.8566
```

Question 11

Each row of following is $z_k = [x_k \; y_k]$.

```
S1 =
    1.3349    0.4911   -0.4391    0.1089   -1.4509    0.9717   -0.7564    ...
       1.2122   -0.9557   -0.3235
    0.5926    0.2931   -0.3230    0.0962   -0.6474    0.4849   -0.3805    ...
       0.6199   -0.4859   -0.1617
    0.6017    0.2962   -0.3318    0.1020   -0.6570    0.4603   -0.3623    ...
       0.6228   -0.4960   -0.1446
    0.5987    0.2988   -0.3371    0.1041   -0.6538    0.4609   -0.3629    ...
       0.6237   -0.4963   -0.1449
```

```
    0.5988    0.2988   -0.3372    0.1041   -0.6539    0.4605   -0.3626    ...
       0.6237   -0.4965   -0.1447
    0.5988    0.2988   -0.3372    0.1041   -0.6539    0.4605   -0.3626    ...
       0.6237   -0.4965   -0.1447
    0.5988    0.2988   -0.3372    0.1041   -0.6539    0.4605   -0.3626    ...
       0.6237   -0.4965   -0.1447
    0.5988    0.2988   -0.3372    0.1041   -0.6539    0.4605   -0.3626    ...
       0.6237   -0.4965   -0.1447
    0.5988    0.2988   -0.3372    0.1041   -0.6539    0.4605   -0.3626    ...
       0.6237   -0.4965   -0.1447
```

At each iteration, f is calculated as following, converging around -6.3288(the last value).

```
S2 =
  -27.0430
   -6.3392
   -6.3353
   -6.3290
   -6.3289
   -6.3288
   -6.3288
   -6.3288
   -6.3288
```

Also, each row of the following is h calculated at each iteration. Notice that the row values converges to [0 0 0]. The norm of h converges to 0 and the convergence is faster that the quadratic penalty method.

```
S3 =
   -3.3330   -3.0036    0.7586
    0.0302   -0.0265    0.0335
   -0.0019    0.0020    0.0074
    0.0004   -0.0004    0.0003
   -0.0000    0.0000    0.0000
    0.0000   -0.0000    0.0000
   -0.0000    0.0000    0.0000
    0.0000   -0.0000   -0.0000
   -0.0000    0.0000   -0.0000
```

Lastly, μ_k is recorded in each row of the following.

```
S4 =
         0         0         0
   -3.3330   -3.0036    0.7586
   -3.2727   -3.0566    0.8256
   -3.2803   -3.0485    0.8552
   -3.2773   -3.0516    0.8578
   -3.2774   -3.0515    0.8582
   -3.2774   -3.0515    0.8582
   -3.2774   -3.0515    0.8582
   -3.2774   -3.0515    0.8582
```

Question 12

μ we figured out is [-3.2774 -3.0515 0.8582]' Recall that

$$L(z,\mu) = \frac{1}{2} z^t \begin{bmatrix} A - 2\mu_1 & \mu_3 \\ \mu_3 & B - 2\mu_2 \end{bmatrix} \quad z + \mu_1 + \mu_2$$

$\begin{bmatrix} A - 2\mu_1 & \mu_3 \\ \mu_3 & B - 2\mu_2 \end{bmatrix}$ has following eigenvalues.

```
ans =
   -1.1210
   -0.0000
    0.8468
    1.5289
    4.7816
    7.8188
    9.7074
   10.4000
   32.7670
   34.5590
```

As it has one negative eigenvalue, it is not positive definite. Thus, L is not convex.

From question 1, $LICQ$ holds. Thus, the strong duality follows.

Appendix: Code

0.1 main.mlx

```
global A B n
A = [ 2 4 4 5 9;
4 4 8 7 5;
4 8 2 8 5;
5 7 8 6 5
9 5 5 5 2 ];
B = [ 4 8 3 6 8;
8 4 4 5 2;
3 4 2 8 5;
6 5 8 4 7;
8 2 5 7 8 ];
n = 5;

eig(A)
eig(B)

%Problem 8
[L, gradL] = alm(randn(10, 1), randn(3, 1), 1);

%Problem 9
beta = 1.42;
mu = [1; 2; -3];
x0 = [1; 0; -1; 2; 1];
y0 = [1; 2; 0;  1; 2];
z0 = [x0; y0];
[z_min, L_min] = fminunc(@(z) alm(z, mu, beta), z0)

f = @f;
f(z_min);
h = @h;
```

```
h(z_min);

%Problem 10
%
z0 = randi(5, 10, 1);
mu = [0; 0; 0];
R = cell(9, 4);

for i = 1:9
    beta = 2^(i-1);
    z_min = fminunc(@(z) alm(z, mu, beta), z0);

    R(i, 1) = {z_min};
    R(i, 2) = {f(z_min)};
    R(i, 3) = {h(z_min)};
    R(i, 4) = {beta*h(z_min)};
    z0 = z_min;
end

R1 = cell2mat(R(:, 1)')'
R2 = cell2mat(R(:, 2)')'
R3 = cell2mat(R(:, 3)')'
R4 = cell2mat(R(:, 4)')'
f(z_min)
h(z_min)

%Problem 11
% ALM Method
z0 = randi(5, 10, 1);
mu = [0; 0; 0];
S = cell(9, 4);
for i = 1:9
    beta = 2^(i-1);
    z_min = fminunc(@(z) alm(z, mu, beta), z0);
    %R(i, :) = [z_min, f(z_min), h(z_min), beta*h(z_min)];
    S(i, 1) = {z_min};
    S(i, 2) = {f(z_min)};
    S(i, 3) = {h(z_min)};
    S(i, 4) = {mu};

    mu = mu + beta*h(z_min);
    z0 = z_min;
end

S1 = cell2mat(S(:, 1)')'
S2 = cell2mat(S(:, 2)')'
S3 = cell2mat(S(:, 3)')'
S4 = cell2mat(S(:, 4)')'
f(z_min)
h(z_min)

%Problem 12
mu
test = [0 0 0]
X = zeros(10);
X(1:n, 1:n) = A-2*mu(1)*eye(n)
X(n+1:2*n, n+1:2*n) = B-2*mu(2)*eye(n)
X(1:n, n+1:2*n) = mu(3)*eye(n)
X(n+1:2*n, 1:n) = mu(3)*eye(n)
```

```
eig(X)
```

0.2 alm.m

```
function [L,gradL] = alm(z, mu, beta)
%Returns L_\beta(z, mu), grad L_\beta(z, mu)
    global A B n
    f = @f;
    h = @h;

    x = z(1:n, 1);
    y = z(n+1:end, 1);

    L = f(z) + mu'*h(z) + beta/2*norm(h(z))^2;
    gradLx = A*x - 2*mu(1)*x + mu(3)*y - 2*beta*x*(1-x'*x) + beta*y*x'*y;
    gradLy = B*y - 2*mu(2)*y + mu(3)*x - 2*beta*x*(1-y'*y) + beta*x*x'*y;
    gradL = [gradLx; gradLy];
end
```

0.3 f.m

```
function f = f(z)
    global A B n
    x = z(1:n, 1);
    y = z(n+1:end, 1);
    f = 1/2*(x'*A*x+y'*B*y);
end
```

0.4 h.m

```
function h = h(z)
    global n
    x = z(1:n, 1);
    y = z(n+1:end, 1);
    h = [1-x'*x; 1-y'*y; x'*y];
end
```