EN.530.603 Applied Optimal Control
HW #1 Solutions

Gowtham Garimella

September 25, 2014

1. To find stationary points and determine whether they are maxima, minima or saddle points: For stationary points, the gradient has to be zero.

   (a)

$$L(x) = (1 - x_1)^2 + 200(x_2 - x_1^2)^2$$

$$\nabla L(x) = \begin{bmatrix} \frac{\partial L}{\partial x_1} \\ \frac{\partial L}{\partial x_2} \end{bmatrix} = \begin{bmatrix} 2(1 - x_1)(-1) + 400(x_2 - x_1^2)(-2x_1) \\ 2 \times 200(x_2 - x_1^2) \end{bmatrix} = \mathbf{0_{2\times 1}}$$

$$\Rightarrow x_2 = x_1^2 \quad \& \quad -1 + x_1 = 0$$

$$\Rightarrow x_1 = 1 \quad \& \quad x_2 = x_1^2 = 1$$

   Thus the stationary point is (1,1). Now to characterize the stationary point we look at the Hessian of the function:

$$\nabla^2 L(x) = \begin{bmatrix} 2\left(1 - 400x_2 + 1200x_1^2\right) & 2(-400x_1) \\ 400(-2x_1) & 400 \end{bmatrix}$$

$$\nabla^2 L(x)|_{(1,1)} = H = \begin{bmatrix} 1602 & -800 \\ -800 & 400 \end{bmatrix}$$

$$eig(H) = \{0.4, 2001.6\}$$

   Thus the Hessian is positive definite which implies the stationary point (1,1) is a **strict local minimum**

   (b)

$$L(u) = (u - 1)(u + 2)(u - 3)$$

$$\nabla L(u) = \frac{\partial L}{\partial u} = (u - 1)(u + 2) + (u - 1)(u - 3) + (u + 2)(u - 3) = 3u^2 - 4u - 5 = 0$$

$$\Rightarrow u^* = \{2.1196, -0.7863\}$$

Now to characterize the stationary points we look at the Hessian:

$$\nabla^2 L(u) = \frac{\partial^2 L}{\partial^2 u} = 6u - 4$$
$$\nabla^2 L(u)|_{u=3.36} == 8.718$$
$$\nabla^2 L(u)|_{u=-0.694} = -8.718$$

Clearly, point u = 2.1194 is **strict local minima** and point u = -0.7863 is **strict local maximum**

(c)

$$L(u) = (u_1^2 + 3u_1 - 4)(u_2^2 - u_2 + 3)$$
$$\nabla L(u) = \begin{bmatrix} \frac{\partial L}{\partial u_1} \\ \frac{\partial L}{\partial u_2} \end{bmatrix} = \begin{bmatrix} (2u_1 + 3)(u_2^2 - u_2 + 3) \\ (u_1^2 + 3u_1 - 4)(2u_2 - 1) \end{bmatrix} = \mathbf{0_{2\times 1}}$$
$$\forall u_2 \in \mathbb{R} \quad u_2^2 - u_2 + 3 > 0$$
$$\Rightarrow u_1 = -3/2$$
$$\Rightarrow u^* = \begin{bmatrix} -3/2 \\ 1/2 \end{bmatrix}$$

Now that we found the stationary point let us characterize it by looking at the Hessian of the above objective function:

$$\nabla^2 L(u) = \begin{bmatrix} 2(u_2^2 - u_2 + 3) & (2u_1 + 3)(2u_2 - 1) \\ (2u_1 + 3)(2u_2 - 1) & 2(u_1^2 + 3u_1 - 4) \end{bmatrix}$$
$$\nabla^2 L(u)|_{u=(-1.5,0.5)} = \begin{bmatrix} 5.5 & 0 \\ 0 & -12.5 \end{bmatrix}$$
$$\Rightarrow eig(\nabla^2 L(u)) = \{5.5, -12.5\}$$

Since the eigen values are both positive and negative at the stationary point u = (-3/2, 1/2), the point is a **saddle point**.

2. To find the stationary points and determine the maxima, minima or saddle points:

   (a) Define H using lagrangian multipliers as $H := L(x) + \lambda f(x)$ where $\lambda \in \mathbb{R}$ is a scalar. To find the stationary points, we have to equate the gradient to zero:

$$
\begin{aligned}
H(x) &= \frac{1}{2}(x_1^2 + x_2^2 + x_3^2) + \lambda(x_1 + x_2 + x_3) \\
\frac{\partial H}{\partial x_1} &= x_1 + \lambda = 0 \quad \text{(at stationary point)} \\
\frac{\partial H}{\partial x_2} &= x_2 + \lambda = 0 \\
\frac{\partial H}{\partial x_3} &= x_3 + \lambda = 0 \\
\text{From constraint f(x)} \quad & x_1 + x_2 + x_3 = 0 \\
\Rightarrow x_2^* = x_3^* &= x_1^* = -\lambda = 0
\end{aligned}
$$

   Thus the stationary point is (0,0,0). To characterize the stationary point we look at the Hessian of the H. Since $\lambda = 0$, $H = L$. Thus

$$
\nabla^2 H = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \Rightarrow eig(H) = \{1, 1, 1\}
$$

   Since the eigen values are all positive, the Hessian is positive definite and hence the stationary point (0,0,0) is a **strict local minimum**

   (b) Similar to above problem, define the augmented cost function and let's equate the gradient to zero:

$$
\begin{aligned}
& H(u) = L(u) + \lambda f(u) = (u_1^2 + 3u_1 - 4)(u_2^2 - u_2 + 3) + \lambda(u_1 - 2u_2) \\
& \frac{\partial H}{\partial u_1} = (2u_1 + 3)(u_2^2 - u_2 + 3) + \lambda = 0 \quad \text{(at stationary point)} \\
& \frac{\partial H}{\partial u_2} = (u_1^2 + 3u_1 - 4)(2u_2 - 1) - 2\lambda = 0 \\
& \frac{\partial H}{\partial \lambda} = u_1 - 2u_2 = 0 \\
& \Rightarrow u_1 = 2u_2 \quad \text{and} \\
& \Rightarrow (u_1^2 + 3u_1 - 4)(2u_2 - 1) + 2(2u_1 + 3)(u_2^2 - u_2 + 3) = 0 \\
& \Rightarrow (2u_2^2 + 3u_2 - 2)(2u_2 - 1) + (4u_2 + 3)(u_2^2 - u_2 + 3) = 0 \\
& \Rightarrow 8u_2^3 + 3u_2^2 + 2u_2 + 11 = 0
\end{aligned}
$$

   The above cubic equation yields only one real root i.e $u_2 = -1.1683$ i.e $u = (-2.3367, -1.1683)$. To characterize this stationary point we look at the second order variation of H only along the constraint directions:

$$\begin{aligned}
du^T\nabla^2 H du &= \begin{bmatrix} du_1 \\ du_2 \end{bmatrix}^T \begin{bmatrix} 2(u_2^2 - u_2 + 3) & (2u_1+3)(2u_2-1) \\ (2u_1+3)(2u_2-1) & 2(u_1^2+3u_1-4) \end{bmatrix} \begin{bmatrix} du_1 \\ du_2 \end{bmatrix} \\
df &= f_{u_1}du_1 + f_{u_2}du_2 = du_1 - 2du_2 = 0 \\
\Rightarrow du^T\nabla^2 H du|_{u=(-2.3367,-1.1683)} &= du_2 \begin{bmatrix} 2 & 1 \end{bmatrix} \begin{bmatrix} 11.0667 & 5.5834 \\ 5.5834 & -11.0999 \end{bmatrix} \begin{bmatrix} 2 \\ 1 \end{bmatrix} du_2 \\
&= 55.5du_2^2
\end{aligned}$$

Since the second order variation is always greater than zero for any non zero du_2, the stationary point (-2.3367, -1.1683) is a **strict local minimum**

3. (a) To optimize the quadratic cost function with constraints given as:

$$\begin{aligned}
L(x,u) &= \frac{1}{2}x^TQx + \frac{1}{2}u^TRu \\
f(x,u) &= Ax + Bu + c = 0 \quad x \in \mathbb{R}^n, c \in \mathbb{R}^m
\end{aligned}$$

Assume Q, R are symmetric. If not the skew symmetric part of the matrix does not alter the cost in any way and the matrices can be relabeled accordingly. Using the Lagrangian multipliers method the necessary conditions are:

$$\begin{aligned}
L'(x,u) &= \frac{1}{2}x^TQx + \frac{1}{2}u^TRu + \lambda^T(Ax + Bu + c) \\
\frac{\partial L'}{\partial x} &= x^TQ + \lambda^T(A) = \mathbf{0_{1\times n}} \\
\frac{\partial L'}{\partial u} &= u^TR + \lambda^T(B) = \mathbf{0_{1\times m}} \\
\frac{\partial L'}{\partial \lambda} &= Ax + Bu + c = 0
\end{aligned}$$

The sufficient conditions are:

$$\begin{aligned}
\begin{bmatrix} dx \\ du \end{bmatrix}^T \begin{bmatrix} Q & 0 \\ 0 & R \end{bmatrix} \begin{bmatrix} dx \\ du \end{bmatrix} > 0 \quad \text{s.t} \\
df = Adx + Bdu = 0
\end{aligned}$$

Assuming A is full rank, the optimal solution for x, u are found as follows:

$$\begin{aligned}
&u = -R^{-1}B^T\lambda \\
&Ax = (BR^{-1}B^T)\lambda - c \\
&\Rightarrow x = A^{-1}(BR^{-1}B^T)\lambda - A^{-1}c \\
&Qx + A^T\lambda = (QA^{-1}(BR^{-1}B^T) + A^T)\lambda - QA^{-1}c = 0 \\
&\Rightarrow A^{-1}((AQA^{-1})(BR^{-1}B^T) + AA^T)\lambda = QA^{-1}c
\end{aligned}$$

It can be shown that the matrix $((AQA^{-1})(BR^{-1}B^T) + AA^T)$ is positive definite under assumption that A is full rank and Q, R are positive semidefinite. Hence it is invertible and we can find λ as:

$$
\begin{aligned}
&\lambda = ((AQA^{-1})(BR^{-1}B^T) + AA^T)^{-1}AQA^{-1}c \\
&\text{OR} \\
&\lambda = (QA^{-1}(BR^{-1}B^T) + A^T)^{-1}QA^{-1}c \\
&\Rightarrow u = -R^{-1}B^T(QA^{-1}(BR^{-1}B^T) + A^T)^{-1}QA^{-1}c \\
&\Rightarrow x = A^{-1}(BR^{-1}B^T)(QA^{-1}(BR^{-1}B^T) + A^T)^{-1}QA^{-1}c
\end{aligned}
$$

Second order sufficient conditions can be simplified as:

$$
\begin{aligned}
&df = Adx + Bdu = 0 \\
&\Rightarrow dx = -A^{-1}Bdu \\
&\Rightarrow du^T \left((A^{-1}B)^T Q(A^{-1}B) + R\right) du > 0
\end{aligned}
$$

Since R is positive definite, Q is positive semidefinite, the combined matrix $((A^{-1}B)^T Q(A^{-1}B) + R)$ is positive definite. Hence the stationary point found is a strict local minimum.

(b) Given quadratic Cost function and linear constraints:

$$
\begin{aligned}
&L(y) = \frac{1}{2}y^T My \\
&\text{s.t} \\
&f(y) = Ay + c = 0
\end{aligned}
$$

Using the Lagrangian approach, the new cost function is written as:

$$H = \frac{1}{2}y^T My + \lambda^T(Ay + c)$$

The first order necessary conditions require:

$$
\begin{aligned}
H_y &= y^T M + \lambda^T A = 0 \\
&\Rightarrow y = M^{-1}A^T\lambda \\
H_\lambda = Ay + c &= (AM^{-1}A^T)\lambda + c = 0 \\
&\Rightarrow \lambda = -(AM^{-1}A^T)^{-1}c \\
&\Rightarrow y = -M^{-1}A^T(AM^{-1}A^T)^{-1}c
\end{aligned}
$$

It should be noted that given A is full rank and M is positive definite, implies that the matrix $AM^{-1}A^T$ is also positive definite and hence invertible. We next look at the second order conditions as follows:

$$
\begin{aligned}
&dy^T\nabla^2 H dy = dy^T M dy > 0 \\
&Ady + c = 0
\end{aligned}
$$

Since M is positive definite, the fist condition is satisfied irrespective of the constraint. Thus the stationary point found is already a strict local minimum. Now to say it is a global minimum, first we note that there are no other stationary points for the cost function except the one we found. Second the function is radially unbounded i.e

$$\lim_{\|y\|\to\infty} \frac{1}{2} y^T M y \to \infty$$

Hence there are no minimum possible at the boundaries either. Hence the only minimum and the global minimum is the stationary point found.

4. Matlab code for both the parts are given as follows:

```
function [] = Hw1(  )
%HW1 Summary of this function goes here
%   Detailed explanation goes here

%%%%%%%%%%%%%%%% Part A %%%%%%%%%%%%%%%
%Variables
syms x1 x2;
x = [x1;x2];

%Choose function
L = (1-x1)^2 + 100*(x2-x1^2)^2;
%Gradient and Hessian
gradL = jacobian(L,x)';
hessL = jacobian(gradL,x)';

%Matlab handles for values:
Lnum = matlabFunction(L,'vars',{x});
gradLnum = matlabFunction(gradL,'vars',{x});
hessLnum = matlabFunction(hessL,'vars',{x});

disp(['L = ' , char(L)]);
disp(['gradient L = ']); disp(gradL);
disp(['hessian L = ']); disp(hessL);



disp('Gradient descent');
%Choose step size
step = 0.002;
disp(['step = ',num2str(step)]);
%Initialpoint
itpoints = zeros(2,1000);
x0 = [0;0];
itpoints(:,1) = x0;
disp('Starting point = '); disp(x0);
%Finding the stationary points:
i = 1;
gradcurrent = gradLnum(x0);
%figure,clf,  hold on;
while ((i < 10000) && (norm(gradcurrent) > 1e-4))
    gradcurrent = gradLnum(x0);
```

```
    x0 = x0 - step*gradcurrent;
    disp(['iteration: ', num2str(i)]);
    %disp(['norm of gradient: ',num2str(norm(gradcurrent))]);
    %disp('gradient');
    %disp(gradcurrent);
    %plot
    %plot(i,norm(gradcurrent),'*');
    %pause(0.05);
    i = i+1;
    itpoints(:,i) = x0;
end
disp(['Final point found after',num2str(i),'is:']);
disp(x0);
disp('Value at final point:');
disp(Lnum(x0));
disp('final gradient');
disp(gradLnum(x0));
disp('hessian @ final point');
disp(hessLnum(x0));
disp(eig(hessLnum(x0)));
figure;
ezsurfc(L,[-2,2]);
hold on, plot3(itpoints(1,:),itpoints(2,:),Lnum(itpoints),'g*-');
figure;
ezcontour(L,[0,2],100);
hold on, plot(itpoints(1,:),itpoints(2,:),'mx-');
disp('Please press any key to continue');
pause;

%%%%%%%%%%%%%%%% Part B %%%%%%%%%%%%%%%
clear;
close all;
figure;

syms x u;
L = x^2 + 20*u^2;
f = x - 2*u + 3;

ezcontour(L,[-5,5],100);
hold on, ezplot(f,[-5,5]);


options = optimset([],'GradObj','on','GradConstr','on');
options.Display = 'iter';
%options.PlotFcns = @optimplotx;
y0 = [0; 2];
objfun(y0);
[yfinal] = fmincon(@objfun,y0,[],[],[],[],[],[], ...
                                    @noncon,options);
disp('yfinal'); disp(yfinal);
figure;
%plot the function:
syms x u;
y = [x;u];

```

```
L = x^2 + 20*u^2;

ezsurfc(L,[-2,2]);
hold on;
f = x - 2*u + 3;
ezplot(f,[-2,2]);
end


function [Lval,gradLval] = objfun(y1)

    syms x u;
    y = [x;u];

    L = x^2 + 20*u^2;
    gradL = jacobian(L,y)';

    Lnum = matlabFunction(L,'vars',{y});
    Lval = Lnum(y1);
    plot(y1(2),y1(1),'b*-');%plot on same figure the points
    if(nargout >1)
        gradLnum = matlabFunction(gradL,'vars',{y});
        gradLval = gradLnum(y1);
    end
end

function [c,ceq,gc,gceq] = noncon(y1)

    syms x u;
    y = [x;u];

    f = x - 2*u + 3;
    gradf = jacobian(f,y)';

    fnum = matlabFunction(f,'vars',{y});
    c = fnum(y1);

    if(nargout >1)
     ceq = [];
    end
    if(nargout > 2)
        gradfnum = matlabFunction(gradf,'vars',{y});
        gc = gradfnum(y1);
        gceq = [];
    end
end
```

5. Minimization problem stated as:

$$f^k(d) = f(x^k) + \nabla f(x^k)^T d + \frac{1}{2} d^T \nabla^2 f(x^k) d \quad d \in \mathbb{R}^n$$

$$s.t \quad \|d\| \leq \gamma^k$$

If the constraint is active it is treated as an equality and following Lagrange multipliers method:

$$
\begin{aligned}
f(d) &= \|d\| - \gamma^k = \sqrt{d^T d} - \gamma^k \\
L'(d) &= f(x^k) + \nabla f(x^k)^T d + \frac{1}{2} d^T \nabla^2 f(x^k) d + \lambda(\sqrt{d^T d} - \gamma^k) \quad \lambda \geq 0 \\
\frac{\partial L'}{\partial d} &= \nabla f(x^k)^T + d^T \nabla^2 f(x^k) + \lambda(\frac{1}{2\sqrt{d^T d}} d^T) = 0 \\
\frac{\partial L'}{\partial \lambda} &= \sqrt{d^T d} - \gamma^k = 0 \\
\Rightarrow -\nabla f(x^k)^T &= d^T \left(\nabla^2 f(x^k) + \frac{\lambda}{2\gamma^k} I \right) \\
\Rightarrow \left(\nabla^2 f(x^k) + \delta^k I \right) d &= -\nabla f(x^k) \quad \text{Since hessian and Identity are both symmetric}
\end{aligned}
$$

Thus we have shown that the constrained optimization problem is equivalent to solving the above form of matrix equations for δ^k and d. The value of δ^k is given as:

$$
\begin{aligned}
&\delta^k = 0 \quad \text{if } \|\nabla^2 f(x^k)^{-1} \nabla f(x^k)\| \leq \gamma^k \quad \text{else} \\
&\delta^k \text{ is given by } \| \left(\nabla^2 f(x^k) + \delta^k I \right)^{-1} \nabla f(x^k)\| = \|d\| = \gamma^k
\end{aligned}
$$

A reasonable choice for δ^k is found by assuming $\delta^k = max\{-eig(\nabla^2 f(x^k)), 0\} + \epsilon$ (eig(M) refers to list of eigen values of the M) and finding ϵ to ensure $\|d\| \leq \gamma^k$. With the form of δ chosen, the matrix $H' = \left(\nabla^2 f(x^k) + \delta^k I \right)^{-1}$ is symmetric positive definite with maximum eigen value being $1/\epsilon$. Then the matrix ${H'}^T H'$ is also positive definite with maximum eigen value being $1/\epsilon^2$.

$$
\begin{aligned}
&d^T d = \nabla f(x^k)^T {H'}^T H' \nabla f(x^k) \leq \frac{1}{\epsilon^2} \|\nabla f(x^k)\|^2 \\
&\text{Choose } \epsilon \text{ as} \\
&\epsilon = \frac{\|\nabla f(x^k)\|}{\gamma^k} \qquad \Rightarrow d^T d \leq {\gamma^k}^2 \\
&\delta^k = max\{-eig(\nabla^2 f(x^k)), 0\} + \frac{\|\nabla f(x^k)\|}{\gamma^k}
\end{aligned}
$$