MATH 164 - Optimization - Midterm

Darren Tsang, 405433124

Produced on Sunday, Jul. 19 2020 @ 10:57:29 AM

Note: I referenced the course textbook as I was completing this midterm.

Question 1

Part A

Using the Golden Ratio

Iteration 1:

$$a_1 = a_0 + \rho(b_0 - a_0) = -1 + .38196(1 + 2) = -.2362, f(a_1) = .8454$$
$$b_1 = a_0 + (1 - \rho)(b_0 - a_0) = -1 + (1 - .38196)(1 + 2) = .2362, f(b_1) = 1.3222$$

Since $f(a_1) < f(b_1)$, the new search interval is now:

$$[a_0, b_1] = [-1, .2362]$$

Iteration 2:

$$a_2 = -1 + .38196(.2362 + 1) = -.5278, f(a_2) = .8685$$
$$b_2 = a_1 = -.2362, f(b_2) = .8454$$

Since $f(b_2) < f(a_2)$, the new search interval becomes:

$$[a_2, b_1] = [-.5278, .2362]$$

Iteration 3:

$$a_3 = b_2 = -.2362, f(a_3) = .8454$$
$$b_3 = -.5278 + (1 - .38196)(.2362 + .5278) = -.0556, f(b_3) = .9490$$

Since $f(a_3) < f(b_3)$, the new search interval becomes:

$$[a_2, b_3] = [-.5278, -.0556]$$

Iteration 4:

$$a_4 = -.5278 + .38196(-.0556 + .5278) = -.3471, f(a_4) = .8272$$
$$b_4 = a_3 = -.2362, f(b_4) = .8454$$

Since $f(a_4) < f(b_4)$, the new search interval becomes:

$$[a_2, b_4] = [-.5278, -.2362]$$

Iteration 5:

$$a_5 = -.5278 + .38196(-.2362 + .5278) = -.4164, f(a_5) = .8328$$

$$b_5 = a_4 = -.3471, f(b_5) = .8272$$

Since $f(b_5) < f(a_5)$, the new search interval becomes:

$$[-.4164, -.2362]$$

We can stop here because the distance between the interval is $|-.2362 + .4164| = .1802 < .2$.

Using the Bisection Method

We will be finding the root of the following function:

$$g(x) = f'(x) = e^x + 2x$$

Iteration 1:

$$a_0 = -1, g(-1) = -1.6321$$
$$b_0 = 1, g(1) = 4.7183$$
$$x_1 = 0, g(0) = 1$$

Iteration 2:

$$a_1 = -1, g(a_1) = -1.6321$$
$$b_1 = 0, g(b_1) = 1$$
$$x_2 = -.5, g(x_2) = -.3935$$

Iteration 3:

$$a_2 = -.5, g(a_2) = -.3935$$
$$b_2 = 0, g(b_2) = 1$$
$$x_3 = -.25, g(x_3) = .2788$$

Iteration 4:

$$a_3 = -.5, g(a_3) = -.3935$$
$$b_3 = -.25, g(b_3) = .2788$$
$$x_4 = -.375, g(x_4) = -.0627$$

The new interval will be $[-.375, -.25]$, which means we can stop because the distance between the two points is less than .2.

Part B

We take the limit:

$$\lim_{n\to\infty} \frac{\left(\frac{b-a}{2^{n+1}}\right)}{\left(\frac{b-a}{2^n}\right)^1} = \lim_{n\to\infty} \frac{1}{2} = \frac{1}{2} = \lambda$$

Since $0 < \lambda < \infty$, the convergence for the bisection method is linear.

Question 2

$$f(x) = \frac{1}{2}||Ax||^2 + \frac{1}{2}||x-b||^2 \tag{1}$$

Part A

We can expand (1):

$$f(x) = \frac{1}{2}[(Ax)^T(Ax)] + \frac{1}{2}[(x-b)^T(x-b)] = \frac{1}{2}[x^T(A^TA)x] + \frac{1}{2}[||x||^2 - 2b^Tx + ||b||^2]$$

Doing some more simplifying:

$$f(x) = \frac{1}{2}[x^T(A^TA)x + ||x||^2] + \frac{1}{2}[-2b^Tx + ||b||^2] = \frac{1}{2}[x^T(A^TA)x + x^TIx] - b^Tx + \frac{1}{2}||b||^2$$

We finally get:

$$f(x) = \frac{1}{2}x^T(A^TA+I)x - b^Tx + \frac{1}{2}||b||^2$$

Now, we can find the gradient and set it equal to 0:

$$\nabla f(x) = (A^TA+I)x - b = 0$$

Solving for x, we get:

$$x^* = (A^TA+I)^{-1}b,$$

as the minimizer of f.

Then, the Hessian is:

$$F(x) = A^TA + I$$

Let d be a nonzero vector. We want to show that the Hessian is positive definite:

$$d^TF(x)d = d^TA^TAd + d^TId = (Ad)^T(Ad) + d^Td = ||Ad||^2 + ||d||^2 > 0$$

We see that $x^* = (A^TA+I)^{-1}b$ satisfies the conditions of SOSC (as shown in above), thus x^* is a strict local minimizer for f. Furthermore, we know that x^* is the only critical point (as shown above), which means that x^* must be a global minimizer.

Part B

We plug in what we are given into f:

$$f(x) = \frac{1}{2}x^T\left(\begin{bmatrix}0 & 1\\ 1 & 1\\ 1 & 0\end{bmatrix}^T\begin{bmatrix}0 & 1\\ 1 & 1\\ 1 & 0\end{bmatrix} + \begin{bmatrix}1 & 0\\ 0 & 1\end{bmatrix}\right)x - \begin{bmatrix}1\\ 2\end{bmatrix}^T x + \frac{1}{2}\left|\left|\begin{bmatrix}1\\ 2\end{bmatrix}\right|\right|^2 = \frac{1}{2}x^T\begin{bmatrix}3 & 1\\ 1 & 3\end{bmatrix}x - \begin{bmatrix}1\\ 2\end{bmatrix}^T x + \frac{5}{2}$$

Then, we find the eigenvalues of $Q = \begin{bmatrix}3 & 1\\ 1 & 3\end{bmatrix}$:

$$det(Q - \lambda I) = \begin{bmatrix}3-\lambda & 1\\ 1 & 3-\lambda\end{bmatrix} = (3-\lambda)^2 - 1 = 0 \longrightarrow \lambda_1 = 2, \lambda_2 = 4$$

Thus, for any step size α such that $0 < \alpha < 2/4 = 1/2$ the gradient descent algorithm will converge.

Part C

The gradient of f with the A and b given in Part B:

$$\nabla f(x) = \begin{bmatrix} 3 & 1 \\ 1 & 3 \end{bmatrix} x - \begin{bmatrix} 1 \\ 2 \end{bmatrix}$$

Here is the algorithm for steepest descent:

$$x^{(k+1)} = x^{(k)} - \alpha \nabla f(x^{(k)}) = x^{(k)} - \alpha \begin{bmatrix} 3 & 1 \\ 1 & 3 \end{bmatrix} x^{(k)} + \alpha \begin{bmatrix} 1 \\ 2 \end{bmatrix},$$

If α is such that $0 < \alpha < 1/2$, the algorithm will converge.

Part D

The conditions for Theorem 9.1 hold, thus for all $x^{(1)}$ sufficiently close to x^*, Newton's method is well-defined for all k and converges to x^*. Furthermore, the conditions for Theorem 9.2 also hold, which means any initial $x^{(0)}$ will produce an algorithm that will always be descending towards the minimum. This means that Newton's method globally converge for this function.

Question 3

$$f(x) = \frac{1}{2}x_1^2 + \frac{1}{3}x_2^2 - \frac{2}{3}x_1x_2 - 2x_1 - x_2 \tag{1}$$

Part A

We can rewrite the equation (1) as:

$$f(x) = \frac{1}{2}x^TQx + xb = \frac{1}{2}\begin{bmatrix}x_1 & x_2\end{bmatrix}\begin{bmatrix}1 & -\frac{2}{3} \\ -\frac{2}{3} & \frac{2}{3}\end{bmatrix}\begin{bmatrix}x_1 \\ x_2\end{bmatrix} - \begin{bmatrix}x_1 & x_2\end{bmatrix}\begin{bmatrix}2 \\ 1\end{bmatrix}$$

Part B

We have:

$$g(x) = \nabla f(x) = \begin{bmatrix}x_1 - \frac{2}{3}x_2 - 2, & -\frac{2}{3}x_1 + \frac{2}{3}x_2 - 1\end{bmatrix}^T$$

Thus,

$$g^{(0)} = \nabla f(x^{(0)}) = \begin{bmatrix}-2 & -1\end{bmatrix}^T,$$
$$d^{(0)} = -g^{(0)} = \begin{bmatrix}2 & 1\end{bmatrix}^T,$$
$$\alpha_0 = -\frac{g^{(0)T}d^{(0)}}{d^{(0)T}Qd^{(0)}} = -\frac{-5}{2} = \frac{5}{2},$$
$$x^{(1)} = x^{(0)} + \alpha_0 d^{(0)} = \begin{bmatrix}5 & \frac{5}{2}\end{bmatrix}^T$$

The next step:

$$g^{(1)} = \nabla f(x^{(1)}) = \begin{bmatrix}\frac{4}{3} & \frac{-8}{3}\end{bmatrix}^T,$$
$$\beta_0 = \frac{g^{(1)T}Qd^{(0)}}{d^{(0)T}Qd^{(0)}} = \frac{32/9}{2} = \frac{16}{9}$$
$$d^{(1)} = -g^{(1)} + \beta_0 d^{(0)} = \begin{bmatrix}\frac{20}{9} & \frac{40}{9}\end{bmatrix}^T$$
$$\alpha_1 = -\frac{g^{(1)T}d^{(1)}}{d^{(1)T}Qd^{(1)}} = -\frac{-80/9}{400/81} = 1.8$$
$$x^{(2)} = x^{(1)} + \alpha_1 d^{(1)} = \begin{bmatrix}9 & 10.5\end{bmatrix}^T$$

Then:

$$g^{(2)} = \nabla f(x^{(2)}) = \begin{bmatrix}0 & 0\end{bmatrix}^T$$

Thus, the minimizer of f is $\begin{bmatrix}9 & 10.5\end{bmatrix}^T$.

Part C

Setting the gradient of $f(x)$ to 0:

$$\nabla f(x) = \begin{bmatrix}x_1 - \frac{2}{3}x_2 - 2, & -\frac{2}{3}x_1 + \frac{2}{3}x_2 - 1\end{bmatrix}^T = \begin{bmatrix}0 & 0\end{bmatrix}^T$$

We obtain the following:

$$x_1 - \frac{2}{3}x_2 - 2 = 0 \longrightarrow \frac{2}{3}x_2 = x_1 - 2$$
$$-\frac{2}{3}x_1 + \frac{2}{3}x_2 - 1 = -\frac{2}{3}x_1 + x_1 - 2 - 1 = 0 \longrightarrow x_1 = 9 \longrightarrow x_2 = 10.5$$

Thus, the minimizer of f is $\begin{bmatrix}9 & 10.5\end{bmatrix}^T$, which agrees with our result from Part B.