620-361 Operations Research Techniques and Algorithms

Assignment 2 Solutions

1. (8 marks) **Consider the problem**

$$\textbf{min } x_1^2x_2^2 - 2x_1x_2^2 + x_1^2 + 3x_2^2 - 2x_1 - 4x_2.$$

Apply two iterations of the

(a) steepest descent method;

(b) Newton's method;

(c) BFGS method

to this problem, starting from the point $x^0 = (0, 0)$**. Which method produces the most accurate estimate?**

We call the objective function $f(x)$. The gradient of f is

$$\nabla f(x) = (2x_1x_2^2 - 2x_2^2 + 2x_1 - 2, 2x_1^2x_2 - 4x_1x_2 + 6x_2 - 4).$$

When this is $(0, 0)$, the first component gives us

$$2(x_1 - 1)(x_2^2 + 1) = 0$$

which implies $x_1 = 1$. The second component gives

$$2(x_1 - 1)^2 + 4(x_2 - 1) = 0$$

which implies $x_2 = 1$.

The Hessian of f is

$$\nabla^2 f(x) = \begin{bmatrix} 2x_2^2 + 2 & 4x_1x_2 - 4x_2 \\ 4x_1x_2 - 4x_2 & 2x_1^2 - 4x_1 + 6 \end{bmatrix}.$$

At the point $(1, 1)$ this is

$$\nabla^2 f(1, 1) = \begin{bmatrix} 4 & 0 \\ 0 & 4 \end{bmatrix}$$

which is clearly positive definite; therefore $(1, 1)$ is the global minimum.

(a) $\nabla f(0,0) = (-2,-4)$, so $d^0 = (2,4)$. Minimising $f(2t, 4t)$ gives the step size $t_0 = 0.286$ and so

$$x^1 = (0,0) + 0.286(2,4) = (0.571, 1.142).$$

Then $\nabla f(0.571, 1.142) = (-1.977, 0.988)$, so $d^1 = (1.977, -0.988)$. Minimising $f(0.571 + 1.977t, 1.142 - 0.988t)$ gives the step size $t_1 = 0.202$, so our final iterate is

$$x^2 = (0.571, 1.142) + 0.202(1.977, -0.988) = (0.970, 0.943).$$

(b)

$$\nabla^2 f(0,0) = \begin{bmatrix} 2 & 0 \\ 0 & 6 \end{bmatrix},$$

which is positive definite. The Newton direction is

$$d^0 = -\nabla^2 f(0,0)^{-1} \nabla f(0,0) = -\begin{bmatrix} \frac{1}{2} & 0 \\ 0 & \frac{1}{6} \end{bmatrix} \begin{bmatrix} -2 \\ -4 \end{bmatrix} = \begin{bmatrix} 1 \\ \frac{2}{3} \end{bmatrix}.$$

Minimising $f(t, \frac{2t}{3})$ gives the step size $t_0 = 1.172$ and so

$$x^1 = (0,0) + 1.172(1, \frac{2}{3}) = (1.172, 0.781).$$

Then

$$d^1 = -\nabla^2 f(1.172, 0.781)^{-1} \nabla f(1.172, 0.781) = -\begin{bmatrix} 3.22 & 0.536 \\ 0.536 & 4.059 \end{bmatrix}^{-1} \begin{bmatrix} 0.553 \\ -0.829 \end{bmatrix} = \begin{bmatrix} -0.210 \\ 0.232 \end{bmatrix}.$$

Minimising $f(1.172 - 0.210t, 0.781 + 0.232t)$ gives the step size $t_1 = 0.886$. So our final iterate is

$$x^2 = (1.172, 0.781) + 0.886(-0.210, 0.232) = (0.985, 0.987).$$

(c) The first iterate of the BFGS method is the same as the steepest descent method, as we are taking $H_0 = I$. So our first iterate is

$$x^1 = (0.571, 1.142).$$

Updating H, we find that $s^0 = (0.571, 1.142)$, $g^0 = (0.023, 4.988)$, and $r^0 = (0.004, 0.874)$. This gives us

$$H_1 = \begin{bmatrix} 1.301 & 0.108 \\ 0.108 & 0.228 \end{bmatrix}.$$

Then our descent direction is

$$d^1 = -H_1 \nabla f(0.571, 1.142) = -\begin{bmatrix} 1.301 & 0.108 \\ 0.108 & 0.228 \end{bmatrix} \begin{bmatrix} -1.977 \\ 0.988 \end{bmatrix} = \begin{bmatrix} 2.466 \\ -0.011 \end{bmatrix}.$$

Minimising $f(0.571 + 2.466t, 1.142 - 0.011t)$ gives the step size $t_1 = 0.174$, so our final iterate is

$$x^2 = (1.001, 1.140).$$

To compare the estimates, we look at the normed error.

Method	Normed error
Steepest descent	0.064
Newton's	0.020
BFGS	0.140

Therefore Newton's method is the most accurate.

2. (6 marks) **The BFGS method is constructed so that all the approximate inverse Hessians, H_k, are symmetric and positive definite.**

   (a) Show by direct calculation that the H_k's you obtained in question 1(c) are positive definite.

   (b) Explain why we want H_k to be positive definite.

   (c) Prove from the BFGS formula that H_k is symmetric.

   (a) Obviously $H_0 = I$ is positive definite. We apply the leading principal minor test to show that H_1 is positive definite. Obviously $1.301 > 0$, and $det(H_1) = 0.286 > 0$. Therefore H_1 is positive definite.

   (b) We want H_k to be positive definite so that the direction we search in, $-H_k \nabla f(x^k)$, is a descent direction.

   (c) We use induction. Obviously H_0 is symmetric. Assume that H_k is symmetric. Then

$$\begin{aligned} H_{k+1}^T &= H_k^T + \frac{1 + \langle r^k, g^k \rangle}{\langle s^k, g^k \rangle} [s^k (s^k)^T]^T - [s^k (r^k)^T]^T - [r^k (s^k)^T]^T \\ &= H_k + \frac{1 + \langle r^k, g^k \rangle}{\langle s^k, g^k \rangle} s^k (s^k)^T - r^k (s^k)^T - s^k (r^k)^T \\ &= H_{k+1}. \end{aligned}$$

   Thus H_{k+1} is symmetric and H_k is symmetric for all k by induction.

3. (6 marks) **Implement the steepest descent method in MATLAB code. For each iteration, find the step size using the algorithm which finds a step size satisfying the Armijo-Goldstein and Wolff conditions.**

   See the file steepestDescent.m on the website.