Optimization HW 7

Michael Nameika

April 2023

Section 11.5 Problems

1. Consider the problem

$$\text{minimize} \quad f(x_1, x_2) = (x_1 - 2x_2)^2 + x_1^4$$

(i) Suppose a Newton's method with a line search is used to minimize the function, starting from the point $x = (2, 1)^T$. What is the Newton search direction at this point?

Recall that the Newton search direction is given by solving the linear system

$$\nabla^2 f(x_k) p_k = -\nabla f(x_k)$$

For our problem, we find the following for the gradient and Hessian of f:

$$\nabla f(x_1, x_2) = \begin{pmatrix} 2(x_1 - 2x_2) + 4x_1^3 \\ -4(x_1 - 2x_2) \end{pmatrix}$$

$$\nabla^2 f(x_1, x_2) = \begin{pmatrix} 2 + 12x_1^2 & -4 \\ -4 & 8 \end{pmatrix}$$

and with $x_0 = (2, 1)^T$, we have

$$\nabla f(x_0) = \begin{pmatrix} 32 \\ 0 \end{pmatrix}$$

$$\nabla^2 f(x_0) = \begin{pmatrix} 50 & -4 \\ -4 & 8 \end{pmatrix}$$

Then we find

$$p_0 = \begin{pmatrix} -2/3 \\ -1/3 \end{pmatrix}$$

(ii) Suppose a backtracking line search is used. Does the trial step $\alpha = 1$ satisfy the sufficient decrease condition for $\mu = 0.2$? For what values of μ does $\alpha = 1$ satisfy the sufficient decrease condition?

Recall the sufficient decrease condition:

$$f(x_k + \alpha_k p_k) \leq f(x_k) + \mu \alpha_k p_k^T \nabla f(x_k)$$

For $\mu = 0.2$, $\alpha = 1$, and p_k given by p_0 in part (i), we find the following:

$$\begin{aligned} f(x_k + \alpha_k p_k) &= f(4/3, 2/3) = \frac{256}{81} \\ f(x_k) + \mu \alpha_k p_k^T \nabla f(x_k) &= 16 + (0.2)(1)(-2/3, -1/3)(32, 0)^T \\ &= \frac{176}{15} \end{aligned}$$

Observe that $\frac{256}{81} \approx 3$ and $\frac{176}{15} \approx 11$, so the sufficient decrease condition is satisfied for $\mu = 0.2$. Now we wish to find all values of μ such that the sufficient decrease condition is satisfied. That is, we must solve

$$f(x_k + \alpha_k p_k) \leq f(x_k) + \mu\alpha p_k^T \nabla f(x_k)$$

for μ with the added condition that $\mu > 0$. Using $x_k = (2,1)^T$, $p_k = (-2/3, -1/3)^T$, $\alpha_k = 1$, we find

$$\begin{aligned} \mu &\leq \frac{256/81 - 16}{-64/3} \\ &= \frac{65}{108} \end{aligned}$$

and so The values of μ that the trial step $\alpha = 1$ satisfies the decrease condition are

$$0 < \mu \leq \frac{65}{108}$$

2. Let

$$f(x_1, x_2) = 2x_1^2 + x_2^2 - 2x_1x_2 + 2x_1^3 + x_1^4$$

(i) Suppose that the function is minimized starting from $x_0 = (0, -2)^T$. Verify that $p_0 = (0, 1)^T$ is a direction of descent.

Recall that p_0 is a direction of descent in case

$$p_0^T \nabla f(x_0) < 0$$

First, let us find $\nabla f(x_0)$. For the gradient we have

$$\nabla f(x_1, x_2) = \begin{pmatrix} 4x_1 - 2x_2 + 6x_1^2 + 4x_1^3 \\ 2x_2 - 2x_1 \end{pmatrix}$$

and so

$$\nabla f(x_0) = \begin{pmatrix} 4 \\ -4 \end{pmatrix}$$

and we have

$$\begin{aligned} p_0^T \nabla f(x_0) &= (0, 1) \begin{pmatrix} 4 \\ -4 \end{pmatrix} \\ &= -4 < 0 \end{aligned}$$

So $p_0 = (0, 1)^T$ is a descent direction of f at $x_0 = (0, -2)^T$.

(ii) Suppose that a line search is used to minimize the function $F(\alpha) = f(x_0 + \alpha p_0)$, and that a backtracking line search is used to find the optimal step length α. Does $\alpha = 1$ satisfy the sufficient decrease condition for $\mu = 0.5$? For what values of μ does $\alpha = 1$ satisfy the sufficient decrease condition?

Recall again the sufficient descent condition:

$$f(x_0 + \alpha p_0) \leq f(x_0) + \mu\alpha p_0^T \nabla f(x_0)$$

Evaluating the right-hand side, we find

$$\begin{aligned} f(x_0) + \mu\alpha p_0^T \nabla f(x_0) &= 4 + (0.5)(1)(-4) \\ &= 2 \end{aligned}$$

and the left hand side:

$$f(x_0 + \alpha p_0) = f(0, -1) = 1$$

Clearly, $1 \leq 2$, so the sufficient decrease condition is satisfied. Now we wish to find the values of μ that satisfy the sufficient decrease condition for $\alpha = 1$. Well, from above, we can see we wish to find μ that satisfy (with the added condition that $\mu > 0$

$$1 \leq 4 - 4\mu$$

Clearly, we require $\mu \leq 3/4$. Then the values of μ that satisfy the sufficient decrease condition are

$$0 < \mu \leq \frac{3}{4}$$

3. Consider the quadratic function

$$f(x) = \frac{1}{2}x^T Q x - c^T x,$$

where Q is a positive definite matrix. Let p be a direction of descent for f at the point x. Prove that the solution of the exact line search problem

$$\underset{\alpha > 0}{\text{minimize}} \quad f(x + \alpha p)$$

is

$$\alpha = -\frac{p^T \nabla f(x)}{p^T Q p}.$$

Proof: Let f and p be defined as above. Define

$$F(\alpha) \equiv f(x + \alpha p)$$

Notice that F is quadratic in α and since Q is positive definite, we have that a minimizer exists to F. Rewriting our minimization problem in terms of F, we wish to solve

$$\underset{\alpha > 0}{\text{minimize}}\, F(\alpha)$$

This equates to solving when the derivative of F is equal to zero. That is, solve

$$F'(\alpha) = 0$$

Notice

$$F'(\alpha) = p^T \nabla f(x + \alpha p)$$

and that

$$\nabla f(x) = Qx - c$$

Then $F'(\alpha) = p^T(Q(x + \alpha p) - c) = 0$. Simplifying, we find

$$\begin{aligned}
p^T Q x + \alpha p^T Q p - p^T c &= 0 \\
\alpha p^T Q p &= p^T c - p^T Q x \\
\alpha &= \frac{p^T c - p^T Q x}{p^T Q p}
\end{aligned}$$

but since $\nabla f(x) = Qx - c$, it is clear that $p^T c - p^T Q x = -p^T \nabla f(x)$ and our solution is

$$\alpha = -\frac{p^T \nabla f(x)}{p^T Q p}$$

Which is what we sought to show.

Section 12.2 Problems

1. Use the steepest-descent method to solve

$$\text{minimize} \quad f(x_1, x_2) = 4x_1^2 + 2x_2^2 + 4x_1x_2 - 3x_1,$$

starting from the point $(2, 2)^T$. Perform three iterations.

Writing a MATLAB script to implement the steepest descent method, we find after 17 iterations the minimum of $f(x)$ occurs at

$$x_* = \begin{pmatrix} -3/4 \\ -3/4 \end{pmatrix}$$

With the associated value of $f(x_*)$:

$$f(x_*) = -\frac{9}{8}$$

2. Apply the steepest-descent method, with an exact line search, to the three-dimensional quadratic function $f(x) = \frac{1}{2}x^T Q x - c^T x$ with

$$Q = \begin{pmatrix} 1 & 0 & 0 \\ 0 & \gamma & 0 \\ 0 & 0 & \gamma^2 \end{pmatrix} \quad \text{and} \quad c = \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}$$

Here γ is a parameter that can be varied. Try $\gamma = 1, 10, 100, 1000$. How do your results compare with the convergence theory developed above? (If you do this by hand, perform four iterations; if you are using a computer, then it is feasible to perform more iterations.)

Using $\gamma = 1$, we see that $Q = I_3$ and so the iteration will converge after one iteration to the minimizer $x_* = (1, 1, 1)^T$. Clearly, $\text{cond}(Q) = 1$ and so the rate constant is zero, which would correspond to superlinear convergence, or very fast convergence, which is what we see here. (since it converges in one iteration!)

For $\gamma = 10$, we can see that $\text{cond}(Q) = 100$ and so we would expect slow convergence, as the upper bound for the rate constant is given by $C = 0.960788$. This significant decrease in convergence is detailed in the steepest descent script, where it took 1050 iterations to reach the tolerance $\|\nabla f(x_k)\| < 10^{-10}$.

```
x0 =

    1.0000
    0.1000
    0.0100

>> f(x0)

ans =

   -0.5550

>> gradf(x0)

ans =

   1.0e-10 *

   -0.8012
   -0.0000
   -0.5788
```

```
count =

        1050
```

For $\gamma = 100$, we have $\text{cond}(Q) = 10000$ and the corresponding upper bound for the rate constant is $C = 0.999600$, so we expect even slower convergence. Using the script, we find it took 103,376 iterations to reach the tolerance $\|\nabla f(x_k)\| < 10^{-10}$.

```
x0 =

    1.0000
    0.0100
    0.0001

>> f(x0)

ans =

   -0.5050

>> gradf(x0)

ans =

   1.0e-10 *

   -0.8157
   -0.0000
   -0.5779
count =

      103376
```

Finally, for $\gamma = 1000$, we have $\text{cond}(Q) = 10^6$ and the associated upper bound for the rate constant is given by $C = 0.999996$. Using the script, we find it took approximately 10.3 million iterations to reach the tolerance $\|\nabla f(x_k)\| < 10^{-10}$.

```
x0 =

   0.999999999916986
   0.001000000000000
   0.000001000000000

>> f(x0)

ans =

  -0.500500500000000

>> gradf(x0)

ans =

   1.0e-10 *

  -0.830140400864821
  -0.000334177130412
  -0.557549562074655
```

```
count =

    10313424
```

Section 12.3 Problems

1. Apply the symmetric rank-one quasi-Newton method to solve

$$\text{minimize} \quad f(x) = \frac{1}{2}x^T Q x - c^T x$$

with

$$Q = \begin{pmatrix} 5 & 2 & 1 \\ 2 & 7 & 3 \\ 1 & 3 & 9 \end{pmatrix} \quad \text{and} \quad c = \begin{pmatrix} -9 \\ 0 \\ -8 \end{pmatrix}$$

Initialize the method with $x_0 = (0, 0, 0)^T$ and $B_0 = I$. Use an exact line search.

Implementing the symmetric rank-one quasi-Newton method to solve the above problem with the given initial guesses, we find convergence after three iterations with the following values for the optimal point x_*, $f(x_*)$, and $\nabla f(x_*)$:

```
x0 =

   -2.0000
    1.0000
   -1.0000

>> gradf(x0)

ans =

   1.0e-14 *

         0
    0.0444
    0.1776

>> f(x0)

ans =

  -13.0000
```

2. Apply the BFGS quasi-Newton method to solve

$$\text{minimize} \quad f(x) = \frac{1}{2}x^T Q x - c^T x$$

with

$$Q = \begin{pmatrix} 5 & 2 & 1 \\ 2 & 7 & 3 \\ 1 & 3 & 9 \end{pmatrix} \quad \text{and} \quad c = \begin{pmatrix} -9 \\ 0 \\ -8 \end{pmatrix}$$

Initialize the method with $x_0 = (0, 0, 0)^T$ and $B_0 = I$. Use an exact line search.

Implementing the BFGS algorithm, we find convergence after three iterations with the following values for the optimal point x_*, $f(x_*)$, and $\nabla f(x_*)$:

```
x0 =

   -2.0000
    1.0000
   -1.0000

>> gradf(x0)

ans =

   1.0e-14 *

         0
    0.0888
    0.1776

>> f(x0)

ans =

  -13.0000
```

4. Let C be a symmetric matrix of rank one. Prove that C must have the form $C = \gamma w w^T$, where γ is a scalar and w is a vector of norm one.

Proof: Let C be a symmetric $n \times n$ matrix of rank one. Since C is a rank one matrix, every row is a scalar multiple of one row of C. Call this the ith row of C and let $a_1, a_2, \cdots, a_n$ be scalars so that

$$C = \begin{pmatrix} a_1c_{i1} & a_1c_{i2} & \cdots & a_1c_{in} \\ a_2c_{i1} & a_2c_{i2} & \cdots & a_2c_{in} \\ \vdots & \vdots & & \vdots \\ a_ic_{i1} & a_ic_{i2} & \cdots & a_ic_{i2} \\ \vdots & \vdots & & \vdots \\ a_nc_{i1} & a_nc_{i2} & \cdots & a_nc_{in} \end{pmatrix}$$

Notice that we may rewrite the above expression for C as

$$C = \mathbf{a}\mathbf{c}^T$$

with

$$\mathbf{a} = \begin{pmatrix} a_1 \\ a_2 \\ \vdots \\ a_n \end{pmatrix} \quad \mathbf{c} = \begin{pmatrix} c_{i1} \\ c_{i2} \\ \vdots \\ c_{in} \end{pmatrix}$$

And since C is symmetric, we have

$$\begin{aligned} C^T &= (\mathbf{a}\mathbf{c}^T)^T \\ &= \mathbf{c}\mathbf{a}^T \\ &= \mathbf{a}\mathbf{c}^T \end{aligned}$$

Now, let $u \in \mathrm{Im}(C)$ consider the following:

$$\begin{aligned} Cu &= (\mathbf{a}\mathbf{c}^T)u \\ &= (\mathbf{c}^T u)\mathbf{a} \end{aligned}$$

Let $k = \mathbf{c}^T u$ so that $Cu = k\mathbf{a}$. Now consider

$$\begin{aligned} C^T u &= (\mathbf{c}\mathbf{a}^T)u \\ &= (\mathbf{a}^T u)c \end{aligned}$$

Let $l = \mathbf{a}^T u$ so that $C^T u = l\mathbf{c}$ and since $C^T = C$, we have

$$l\mathbf{c} = k\mathbf{a}$$

Since $u \in \mathrm{Im}(C)$, we have $k, l \neq 0$ and so $\mathbf{a} = \alpha\mathbf{c}$ with $\alpha = l/k$. Then

$$C = \alpha\mathbf{c}\mathbf{c}^T$$

Now let $w = \frac{\mathbf{c}}{\|\mathbf{c}\|}$ so that $\|w\| = 1$ and $\gamma = \|\mathbf{c}\|^2\alpha$. We finally have

$$C = \gamma w w^T$$

which is what we sought to show.