AMS230 – Homework #1

Zayd Hammoudeh

April 19, 2018

Name: Zayd Hammoudeh
Course Name: AMS230
Assignment Name: Homework #1
Due Date: April 18, 2018
Student Discussions: I discussed the problems with the following students. All write-ups were prepared separately and independently.

-

Exercise #1
Exercise 2.1 in Nocedal and Wright.

Compute the gradient, $\nabla f(x)$, and Hessian, $\nabla^2 f(x)$ of the Rosenbrock function

$$f(x) = 100(x_2 - x_1^2)^2 + (1 - x_1)^2.$$

The gradient $\nabla f(x)$ is:

$$\begin{aligned}\nabla f(x) &= \begin{bmatrix} \frac{\partial f(x)}{\partial x_1} \\ \frac{\partial f(x)}{\partial x_2} \end{bmatrix} \\ &= \begin{bmatrix} -400x_1(x_2 - x_1^2) - 2(1 - x_1) \\ 200(x_2 - x_1^2) \end{bmatrix} \\ &= \boxed{\begin{bmatrix} 400x_1^3 - 400x_1x_2 + 2x_1 - 2 \\ 200x_2 - 200x_1^2 \end{bmatrix}}.\end{aligned}$$

The Hessian $\nabla^2 f(x)$ equals:

$$\begin{aligned}\nabla^2 f(x) &= \begin{bmatrix} \frac{\partial^2 f(x)}{\partial x_1^2} & \frac{\partial^2 f(x)}{\partial x_1 \partial x_2} \\ \frac{\partial^2 f(x)}{\partial x_2 \partial x_1} & \frac{\partial^2 f(x)}{\partial x_2^2} \end{bmatrix} \\ &= \boxed{\begin{bmatrix} 1200x_1^2 - 400x_2 + 2 & -400x_1 \\ -400x_1 & 200 \end{bmatrix}}\end{aligned}$$

(a) Show that $x^* = (1, 1)^{\mathbf{T}}$ is the only local minimizer of the function, and that the Hessian at that point is positive definite.

To determine the minimizer(s) (if any), set the gradient equal to 0. Therefore,

$$\begin{bmatrix} 400x_1^3 - 400x_1x_2 + 2x_1 - 2 \\ 200x_2 - 200x_1^2 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}.$$

Therefore, $x_2 = x_1^2$. If we substitute this into the first equation, we get:

$$\begin{aligned} 400x_1^3 - 400x_1x_2 + 2x_1 - 2 &= 0 \\ 400x_1^3 - 400x_1x_1^2 + 2x_1 - 2 &= 0 \\ 2x_1 - 2 &= 0 \\ x_1 &= 1 \end{aligned}$$

It is clear then that the only root is $\boxed{(1, 1)^{\mathrm{T}}}$.

The Hessian matrix $\nabla^2 f(x^*)$ of the Rosenbrock function equals:

$$\nabla^2 f(x^*) = \begin{bmatrix} 802 & -400 \\ -400 & 200 \end{bmatrix}.$$

Hessian matrix, $\nabla^2 f(x^*)$, is positive definite (PD) if all of its eigenvalues are positive. These are found via:

$$\begin{aligned} |A - \lambda I| = \begin{vmatrix} 802-\lambda & -400 \\ -400 & 200-\lambda \end{vmatrix} &= 0 \\ 160400 - 1002\lambda + \lambda^2 - 160000 &= 0 \\ 400 - 1002\lambda + \lambda^2 &= 0 \end{aligned}$$

Solving for λ (in Matlab), the eigenvalues are $\boxed{\lambda_1 \approx 0.3994}$ and $\boxed{\lambda_2 \approx 1001.6}$. Therefore, $\nabla^2 f(x^*)$ is PD.

Exercise #2

Exercise 2.9 in Nocedal and Wright.

Consider the function $f(x_1, x_2) = (x_1 + x_2^2)^2$. At the point $x^{\mathbf{T}} = (1, 0)$, we consider the search direction $p^{\mathbf{T}} = (-1, 1)$. Show that p is a descent direction.

Any vector p is a descent direction if $p \cdot \nabla f(x_1, x_2) < 0$. The gradient of $f(x_1, x_2)$ equals:

$$\begin{aligned}\nabla f(x_1, x_2) &= \begin{bmatrix} \frac{\partial f(x_1, x_2)}{\partial x_1} \\ \frac{\partial f(x_1, x_2)}{\partial x_2} \end{bmatrix} \\ &= \begin{bmatrix} 2(x_1 + x_2^2) \\ 4x_2(x_1 + x_2^2) \end{bmatrix}\end{aligned}$$

Therefore, the gradient at $(1, 0)^{\mathrm{T}}$ equals $(2, 0)^{\mathrm{T}}$. If the inner product of the gradient and p are negative, then p is a descent direction. Since

$$p \cdot \nabla f(x) = \begin{bmatrix} -1 \\ 1 \end{bmatrix} \cdot \begin{bmatrix} 2 \\ 0 \end{bmatrix} = -1 \cdot 2 + 1 \cdot 0 = \boxed{-2 < 0},$$

p is a descent direction.

(a) Find all minimizers of the equation (2.10).

Equation (2.10) in the text states:

$$\min_{\alpha > 0} f(x_k + \alpha p_k).$$

Substituting into f as described above, we get:

$$\begin{aligned}f(\vec{x} + \alpha \vec{p}) &= \left((x_1 + \alpha p_1) + (x_2 + \alpha p_2)^2\right)^2 \\ &= ((1 - 1 \cdot \alpha) + (0 + \alpha \cdot 1)^2)^2 \\ &= (\alpha^2 - \alpha + 1)^2.\end{aligned}$$

To determine the minimizers, take the derivative of the above function with respect to α and set the equation equal to zero as shown below.

$$\begin{aligned}\frac{\partial f(x + \alpha p)}{\partial \alpha} &= 2(2\alpha - 1)(\alpha^2 - \alpha + 1) \\ (2\alpha - 1)(\alpha^2 - \alpha + 1) &= 0\end{aligned}$$

The only real minimizer is at $\alpha^* = \frac{1}{2}$. There are two imaginary minimizers at $\frac{1 \pm i\sqrt{3}}{2}$.

Exercise #3
Exercise 3.6 in Nocedal and Wright.

Consider the steepest descent method with exact line searches applied to the convex quadratic function (3.24). Using the properties given in this chapter, show that if the initial point is such that $x_0 - x^*$ is parallel to an eigenvector of Q, then the steepest descent method will find the solution in one step.

For the convex quadratic function, define $\nabla f_k = Q(x_k - x^*)$. Similarly define the weighted norm as $\|x\|_Q^2 = x^{\mathrm{T}}Qx$. Then Nocedal defines equation (3.28) as:

$$\|x_{k+1} - x^*\|_Q^2 = \left\{1 - \frac{(\nabla^{\mathrm{T}} f_k \nabla f_k)^2}{(\nabla^{\mathrm{T}} f_k Q \nabla f_k)(\nabla^{\mathrm{T}} f_k Q^{-1} \nabla f_k)}\right\} \|x_k - x^*\|_Q^2.$$

Recall that for an eigenvector, u, of matrix A it holds that $Au = \lambda u$ where λ is the associated eigenvalue. Similarly, for A^{-1}, the inverse of A, it holds that $A^{-1}u = \frac{1}{\lambda}u$. Using x_0 as defined in the question, we get,

$$\begin{aligned}
\|x_1 - x^*\|_Q^2 &= \left\{1 - \frac{(\nabla^{\mathrm{T}} f_0 \nabla f_0)^2}{(\nabla^{\mathrm{T}} f_0 Q(Q(x_0 - x^*)))(\nabla^{\mathrm{T}} f_0 Q^{-1}(Q(x_0 - x^*))}\right\} \|x_0 - x^*\|_Q^2 \\
&= \left\{1 - \frac{(\nabla^{\mathrm{T}} f_0 \nabla f_0)^2}{(\nabla^{\mathrm{T}} f_0 Q\lambda(x_0 - x^*))(\nabla^{\mathrm{T}} f_0 Q(Q^{-1}(x_0 - x^*))}\right\} \|x_0 - x^*\|_Q^2 \\
&= \left\{1 - \frac{(\nabla^{\mathrm{T}} f_0 \nabla f_0)^2}{\lambda(\nabla^{\mathrm{T}} f_0 Q(x_0 - x^*))(\nabla^{\mathrm{T}} f_0 Q\frac{1}{\lambda}(x_0 - x^*))}\right\} \|x_0 - x^*\|_Q^2 \\
&= \left\{1 - \frac{(\nabla^{\mathrm{T}} f_0 \nabla f_0)^2}{(\nabla^{\mathrm{T}} f_0 \nabla f_0)(\nabla^{\mathrm{T}} f_0 \nabla f_0)}\right\} \|x_0 - x^*\|_Q^2 \\
&= \{1 - 1\} \|x_0 - x^*\|_Q^2 = 0
\end{aligned}$$

Since Q is positive definite, then for any non-zero vector it holds that $x^{\mathrm{T}}Qx > 0$. Therefore, if $\|x_1 - x^*\|_Q^2 = 0$, then $x_1 - x^* = 0$ meaning $x_1 = x^*$ meaning the system has converged, i.e., found the solution.

Exercise #4

Consider the steepest descent method with exact line searches applied to the convex quadratic function $f(x) = \frac{1}{2}x^{\mathbf{T}}Qx - b^{\mathbf{T}}x$, where Q is symmetric and positive definite. Show that the search direction at step $k+1$ is always orthogonal to the search direction at step k, i.e., $p_k^{\mathbf{T}}p_{k+1} = 0$ for all k.

In exact line search, the step size α is chosen such that:

$$\alpha_k = \arg\min_{\alpha} f(x_k + \alpha p_k).$$

Therefore, taking the derivative of f_{k+1} with respect to α yields:

$$\frac{\partial f_{k+1}}{\partial \alpha_k} = 0.$$

By the chain rule, this becomes,

$$\frac{\partial f_{k+1}}{\partial x_{k+1}} \cdot \frac{\partial x_{k+1}}{\partial \alpha_k} = 0.$$

Knowing that $x_{k+1} = x_k + \alpha_k p_k$, we can rearrange the terms to prove the required statement:

$$\begin{aligned} \nabla^{\mathrm{T}} f_{k+1} p_k &= 0 \\ p_{k+1}^{\mathrm{T}} p_k &= 0 \\ p_k^{\mathrm{T}} p_{k+1} &= 0. \end{aligned}$$

Exercise #5

Program Algorithm 3.5 (Line-Search) and Algorithm 3.6 (zoom) in the text book. For the selection of trial step length α_i in Algorithm 3.5 and α_j in Algorithm 3.6, you can use bisection.

Implementation was written in Python. Source code included at the end of this submission.

Exercise #6

Consider the problem of minimizing:

$$f(x_1, x_2) = (cx_1 - 2)^4 + x_2^2(cx_1 - 2)^2 + (x_2 + 1)^2,$$

where c is a nonzero parameter.

(a) Compute $\nabla f(x)$ and $\nabla f^2(x)$, and find the optimal solution.

$$\begin{aligned}\nabla f(x) &= \begin{bmatrix} \frac{\partial f(x)}{\partial x_1} \\ \frac{\partial f(x)}{\partial x_2} \end{bmatrix} \\ &= \boxed{\begin{bmatrix} 4c(cx_1 - 2)^3 + 2cx_2^2(cx_1 - 2) \\ 2x_2(cx_1 - 2)^2 + 2(x_2 + 1) \end{bmatrix}}\end{aligned}$$

$$\begin{aligned}\nabla^2 f(x) &= \begin{bmatrix} \frac{\partial^2 f(x)}{\partial x_1^2} & \frac{\partial^2 f(x)}{\partial x_1 \partial x_2} \\ \frac{\partial^2 f(x)}{\partial x_2 \partial x_1} & \frac{\partial^2 f(x)}{\partial x_2^2} \end{bmatrix} \\ &= \boxed{\begin{bmatrix} 12c^2(cx_1 - 2)^2 + 2c^2x_2^2 & 4cx_2(cx_1 - 2) \\ 4cx_2(cx_1 - 2) & 2(cx_1 - 2)^2 + 2 \end{bmatrix}}.\end{aligned}$$

The gradient, $\nabla f(x)$, above was set equal to $\vec{0}$ and the roots were solved for in Maple. It returned root $\boxed{\left(\frac{2}{c}, -1\right)^{\mathrm{T}}}$. To verify it is a local minimum, we need to substitute this into the Hessian. Therefore,

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

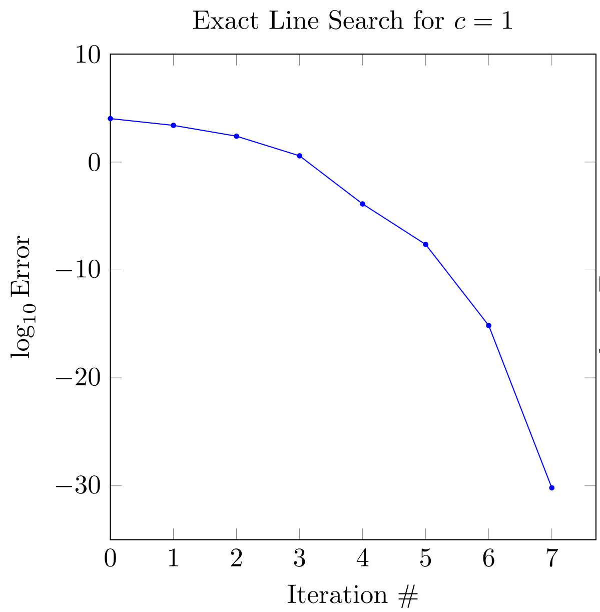
This matrix has eigenvalues $2c^2$ and 2. Since $c \neq 0$ by definition, this matrix is positive definition for all valid values of c.

(b) Program the steepest descent method (using your program from Problem 5 to find the step length). Use your program to numerically solve this problem under two cases: i) $c = 1$, ii) $c = 10$. Compare the convergence in these two cases.

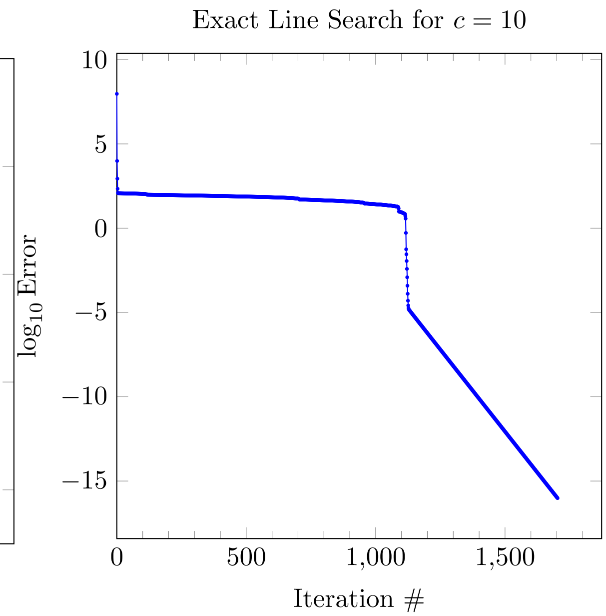
Table 1 lists the parameter values for this part of the problem. Figures 1a and 1b show the results of exact line search with the steepest descent method with for $c = 1$ and $c = 10$ respectively. Note that for $c = 1$, the result converged in only 7 iterations. In contrast for $c = 10$, convergence took 1,703 iterations.

Table 1: Parameters used for the exact line search of problem #6

Parameter	Value
c_1	0.1
c_2	0.9
tol	10^{-16}
x_0^{T}	$[10, 10]$



(a) Problem #6 – $c = 1$ Steepest Descent Exact Line Search



(b) Problem #6 – $c = 10$ Steepest Descent Exact Line Search

Figure 1: Exact line search steepest descent optimization for two values of c

(c) Using Theorem 3.4, explain why as c increases the convergence of the steepest descent method deteriorates.

For a matrix A with sorted eigenvalues in ascending order $\lambda_1, \ldots, \lambda_n$, the condition number is defined as:

$$\kappa(A) = \frac{\lambda_n}{\lambda_1}.$$

In general, as κ of the Hessian increases, the system has worse conditioning and takes more steps to converge. For this problem $\kappa(\nabla^2 f) = c^2$. Therefore, the convergence rates deteriorates quadratically with c.