Department of Mathematics, University of Utah

**Introduction to Optimization**

**MATH 5770/6640, ME EN 6025 – Section 001 – Fall 2021**

**Homework 4 Solutions**

**Newton's method and convex sets**

**Due November 2, 2021**

---

Submit your homework assignment as a scanned copy **ON CANVAS**, to the `Homework 4` assignment.

Some of the exercises below are computational. The book problems are explained in Matlab. You need not use Matlab to complete the assignment; numerical simulation with any programming language is acceptable.

Text: *Introduction to Nonlinear Optimization*, Amir Beck,

| | | |
|---|---|---|
| Exercises: | # | 5.2, |
| | | 6.2, |
| | | 6.7, |
| | | ~~6.20~~ |
| Extra: | | P1, |
| | | P2, |

**5.2.** Conside the Freudenstein and Roth test function,

$$f(\boldsymbol{x}) = f_1(\boldsymbol{x})^2 + f_2(\boldsymbol{x})^2, \qquad \boldsymbol{x} \in \mathbb{R}^2,$$

where

$$f_1(\boldsymbol{x} = -13 + x_1 + ((5 - x_2)x_2 - 2)x_2,$$
$$f_2(\boldsymbol{x} = -29 + x_1 + ((x_2 + 1)x_2 - 14)x_2.$$

(i) Show that the function $f$ has three stationary points. Find them and prove that one is a global minimizer, one is a strict local minimum and the third is a saddle point.

(ii) Use MATLAB to employ the following three methods on the problem of minimizing $f$:

1. the gradient method with backtracking and parameters $(s, \alpha, \beta) = (1, 0.5, 0.5)$.
2. the hybrid Newton's method with parameters $(s, \alpha, \beta) = (0.5, 0.5)$.
3. damped Gauss-Newton's method with a backtracking line search strategy with parameters $(s, \alpha, \beta) = (1, 0.5, 0.5)$.

All the algorithms should use the stopping criteria $\|\nabla f(\boldsymbol{x})\| \leq 10^{-5}$. Each algorithm should be employed four times on the following four starting points: $(-50, 7)^T$, $(20, 7)^T$, $(20, -18)^T$, $(5, -10)^T$. For each of the four strating points, compare the number of iterations and the point to which each method converged. If a method did not converge, explain why.

**Solution**: We first compute the stationary points of $f$. In preparation for this, we compute the gradient of $f_1$ and $f_2$:

$$\nabla f_1(\boldsymbol{x}) = \begin{pmatrix} 1 \\ -3x_2^2 + 10x_2 - 2 \end{pmatrix}, \qquad \nabla f_2(\boldsymbol{x}) = \begin{pmatrix} 1 \\ 3x_2^2 + 2x_2 - 14 \end{pmatrix}.$$

Stationary points of $f$ are values of $\boldsymbol{x}$ satisfying,

$$\nabla f(\boldsymbol{x}) = 2f_1(\boldsymbol{x})\nabla f_1(\boldsymbol{x}) + 2f_2(\boldsymbol{x})\nabla f_2(\boldsymbol{x}) = \mathbf{0}.$$

Simplifying the above expression, we must find values $(x_1, x_2)$ satisfying,

$$\begin{pmatrix} f_1(\boldsymbol{x}) + f_2(\boldsymbol{x}) \\ f_1(\boldsymbol{x})[-3x_2^2 + 10x_2 - 2] + f_2(\boldsymbol{x})[3x_2^2 + 2x_2 - 14] \end{pmatrix} = \mathbf{0}. \tag{1}$$

The first component of the vector above implies:

$$f_1(\boldsymbol{x}) + f_2(\boldsymbol{x}) = 0 \implies x_1 = 21 + x_2(8 - 3x_2), \tag{2}$$

so that at any stationary point, $x_1$ must have this value. Then at this value of $x_1$, we have

$$x_1 = 21 + x_2(8 - 3x_2) \implies \begin{cases} f_1(\boldsymbol{x}) & = 8 + x_2(-x_2^2 + 2x_2 + 6) \\ f_2(\boldsymbol{x}) & = -8 - x_2(-x_2^2 + 2x_2 + 6) \end{cases} \tag{3}$$

I.e., at a staionary point $\boldsymbol{x}$ we also have $f_1(\boldsymbol{x}) = -f_2(\boldsymbol{x})$. Then the second component of (1) implies:

$$\begin{aligned} f_1(\boldsymbol{x})\frac{\partial f_1}{\partial x_2} + f_2(\boldsymbol{x})\frac{\partial f_2}{\partial x_2} &= 0 \\ \overset{f_1=-f_2}{\implies} f_1(\boldsymbol{x})\left[\frac{\partial f_1}{\partial x_2} - \frac{\partial f_2}{\partial x_2}\right] &= 0 \\ f_1(\boldsymbol{x})\left[-3x_2^2 + 10x_2 - 2 - 3x_2^2 - 2x_2 + 14\right] &= 0. \\ f_1(\boldsymbol{x})\left[-3x_2^2 + 4x_2 + 6\right] &= 0. \end{aligned}$$

Thus, to compute stationary points, either $-3x_2^2 + 4x_2 + 6 = 0$, or $f_1(\boldsymbol{x}) = 0$. The first, quadratic, equation yields solutions,

$$-3x_2^2 + 4x_2 + 6 = 0 \implies x_2 = \frac{2}{3}\left(1 \pm \sqrt{\frac{11}{2}}\right).$$

And enforcing $f_1 = 0$ along with (3) implies,

$$-8 - x_2(-x_2^2 + 2x_2 + 6) = 0 \implies x_2 = 4.$$

where the root at $x_2 = 4$ is found by, e.g., by trial and error or by graphing the expression. With these three values of $x_2$, we compute three critical points using the expression for $x_1$ in (2):

$$\begin{aligned} &\text{SP1} : (x_1, x_2) = (5, 4) \\ &\text{SP2} : (x_1, x_2) = \left(15 + \frac{8}{3}(1 + \sqrt{11/2}), \frac{2}{3}(1 + \sqrt{11/2})\right) \\ &\text{SP3} : (x_1, x_2) = \left(15 + \frac{8}{3}(1 - \sqrt{11/2}), \frac{2}{3}(1 - \sqrt{11/2})\right). \end{aligned}$$

Note that SP1 was found by enforcing $f_1(\boldsymbol{x}) = 0$, and since $f_1 = -f_2$ at stationary points, then at SP1 we have $f(\boldsymbol{x}) = f_1(\boldsymbol{x})^2 + f_2(\boldsymbol{x})^2 = 0 + 0 = 0$. Note that $f$ itself is non-negative, so that $f \geq 0$ always holds. Therefore, SP1 is a global minimum.

To classify SP2 and SP3, we require the Hessian, which, after some simplification, takes the form,

$$\nabla^2 f(\boldsymbol{x}) = 2\begin{pmatrix} 2 & \frac{\partial f_1}{\partial x_2} + \frac{\partial f_2}{\partial x_2} \\ \frac{\partial f_1}{\partial x_2} + \frac{\partial f_2}{\partial x_2} & \left(\frac{\partial f_1}{\partial x_2}\right)^2 + \left(\frac{\partial f_2}{\partial x_2}\right)^2 + f_1 \frac{\partial^2 f_1}{\partial x_2^2} + f_2 \frac{\partial^2 f_2}{\partial x_2^2} \end{pmatrix}$$

Note that at any stationary point we have $f_1 = -f_2$, and at SP2 and SP3 we also have $\frac{\partial f_1}{\partial x_2} - \frac{\partial f_2}{\partial x_2} = 0$. Using these simplifications in the Hessian, we have:

$$\nabla^2 f(\boldsymbol{x}) \overset{\text{SP2,SP3}}{=} 2\begin{pmatrix} 2 & 2\frac{\partial f_1}{\partial x_2} \\ 2\frac{\partial f_1}{\partial x_2} & 2\left(\frac{\partial f_1}{\partial x_2}\right)^2 + f_1\left(\frac{\partial^2 f_1}{\partial x_2^2} - \frac{\partial^2 f_2}{\partial x_2^2}\right) \end{pmatrix}$$

To determine the signs of the eigenvalues of this matrix, we first compute the determinant:

$$\begin{aligned} \frac{1}{4}\det H &\overset{\text{SP2,SP3}}{=} 4 f_1\left(\frac{\partial^2 f_1}{\partial x_2^2} - \frac{\partial^2 f_2}{\partial x_2^2}\right) \\ &= -8 f_1(\boldsymbol{x})(3x_2 + 2). \end{aligned}$$

One can verify directly that $f_1(\boldsymbol{x}) > 0$ at both critical points. From the expressions for $x_2$ in SP2, SP3, we have $3x_2 + 2 < 0$ for SP3, but $3x_2 + 2 > 0$ at SP2. Thus, $\det \nabla^2 f > 0$ at SP3 (meaning that both eigenvalues have the same sign), but $\det \nabla^2 f < 0$ at SP2 (meaning that one eigenvalue is positive, and one is negative). Therefore, SP2 is a saddle point.
That SP3 is a local minimum can be determined by evaluting the trace of $\nabla^2 f$ at SP3:

$$\frac{1}{2}\mathrm{Tr}\nabla^2 f = 2 + 2\left(\frac{\partial f_1}{\partial x_2}\right)^2 + f_1\left(\frac{\partial^2 f_1}{\partial x_2^2} - \frac{\partial^2 f_2}{\partial x_2^2}\right) > 0.$$

Thus, since both the determinant and trace are positive at SP3, then the Hessian there is positive-definite, so that SP3 is a local minimum.
We now run the three algorithms described with the given initial conditions, and report the iteration counts to termination, and identify the stationary points to which the methods converged. The stationary points along with the initialization locations are shown in Figure 1.

| | $\boldsymbol{x}_0 = (-50, 7)^T$ | $\boldsymbol{x}_0 = (20, 7)^T$ | $\boldsymbol{x}_0 = (20, -18)^T$ | $\boldsymbol{x}_0 = (5, -10)^T$ |
|---|---|---|---|---|
| 2-5 Gradient descent | SP3, 6491 iterations | SP3, 6265 iterations | SP3, 6320 iterations | SP1, 7119 iterations |
| Hybrid Newton | SP1, 9 iterations | SP1, 9 iterations | SP3, 17 iterations | SP3, 14 iterations |
| Damped Newton | SP1, 9 iterations | SP1, 9 iterations | SP3, 17 iterations | SP3, 14 iterations |

Table 1: Stationary point identification and number of iterations until convergence for each method with each starting location. All methods converged successfully.

Python code associated to this problem is available in the Git repo `https://github.com/akilnarayan/2021Fall-Optimization-homework4`, in particular the script `problem_5.2.py`.
Optimization results are summarized in Table 1. We observe that gradient descent takes *far* more iterations than either Newton variant. In this case, both Newton variants behaved in nearly identical ways. All methods converged to a staionary point. We observe that gradient descent takes so many iterations that are not optimal in terms of direction taken, and so an

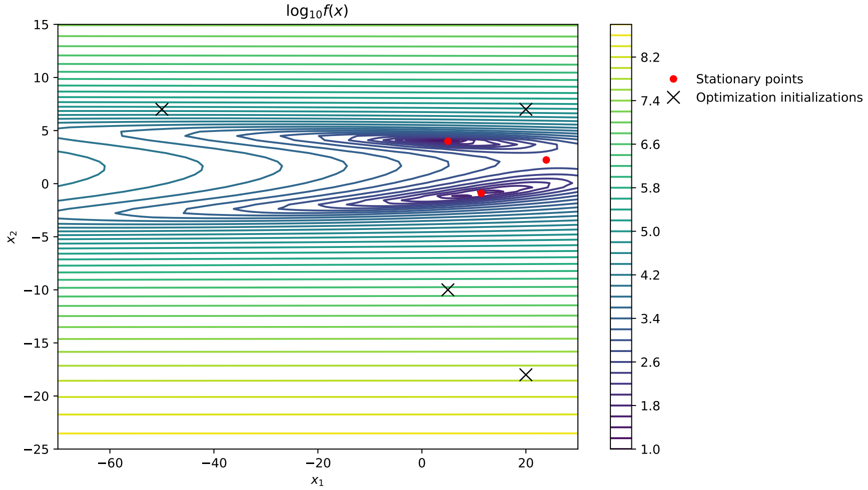

Figure 1: Contour plot of $\log_{10} f$ in Problem 5.2, overlayed with locations of the stationary points of $f$ along with the 4 initialization locations for the optimization algorithms.

initialization $\boldsymbol{x}_0 = (-50, 7)^T$ that is closer to the global minimum SP1 actually converges to the local minimum SP3 that is far away from the starting location.

**6.2.** Give an example of two convex sets $C_1$, $C_2$ whose union $C_1 \cup C_2$ is not convex.

**Solution**: Let $C_1, C_2 \subset \mathbb{R}$ with $C_1 = [0,1]$ and $C_2[2,3]$. Being line segments, both $C_1$ and $C_2$ are convex sets, but $C_1 \cup C_2 = [0,1] \cup [2,3]$ is not convex since $1.5 = \frac{1}{2}1 + \frac{1}{2}2$, and $1, 2 \in C_1 \cup C_2$, but $1.5 \notin C_1 \cup C_2$.

**6.7.** Let $C$ be a convex set. Prove that cone$(C)$ is a convex set.

**Solution**: Let $C \subset \mathbb{R}^n$, and let $\boldsymbol{x}, \boldsymbol{y} \in \text{cone}(C)$ and $\eta \in [0,1]$ be arbitrary. We must show that $\eta \boldsymbol{x} + (1-\eta)\boldsymbol{y} \in \text{cone}(C)$. By definition, we have that there is some collection of vectors $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_k \in \text{cone}(C)$ and some $\boldsymbol{\lambda} \in \mathbb{R}^k_+$ such that

$$\boldsymbol{x} = \sum_{j=1}^{k} \lambda_j \boldsymbol{x}_j.$$

Similarly, there must be some vectors $\boldsymbol{y}_1, \ldots, \boldsymbol{y}_\ell \in \text{cone}(C)$ and some $\boldsymbol{\mu} \in \mathbb{R}^\ell_+$ such that

$$\boldsymbol{y} = \sum_{j=1}^{\ell} \mu \boldsymbol{y}_j.$$

Then the convex combination $\eta \boldsymbol{x} + (1-\eta)\boldsymbol{y}$ is given by,

$$\eta \boldsymbol{x} + (1-\eta)\boldsymbol{y} = \sum_{j=1}^{k} (\eta \lambda_j) \boldsymbol{x}_j + \sum_{j=1}^{\ell} ((1-\eta)\mu_j) \boldsymbol{y}_j. \tag{4}$$

Now, for $j = 1, \ldots, k + \ell$, define vectors $\boldsymbol{z}_j$ by,

$$\boldsymbol{z}_j = \left\{ \begin{array}{rl} \boldsymbol{x}_j, & 1 \leq j \leq k \\ \boldsymbol{y}_{j-k}, & k+1 \leq j \leq k + \ell, \end{array} \right.$$

and define scalars $\tau_j$ by

$$\tau_j = \left\{ \begin{array}{rl} \eta\lambda_j, & 1 \leq j \leq k \\ (1-\eta)\mu_{j-k}, & k+1 \leq j \leq k + \ell. \end{array} \right.$$

Then (4) can be written as,

$$\eta \boldsymbol{x} + (1-\eta)\boldsymbol{y} = \sum_{j=1}^{k+\ell} \tau_j \boldsymbol{z}_j,$$

where $\boldsymbol{z}_j \in \operatorname{cone}(C)$ and $\tau_j \geq 0$ for all $j$. Thus, $\eta \boldsymbol{x} + (1-\eta)\boldsymbol{y} \in \operatorname{cone}(C)$ and we have shown that cone$(C)$ is convex.

**P1.** For each of the following statements, either prove that it is true, or give a counterexample showing that it is false in general.

**a.** If $C_1$ and $C_2$ are convex, then $C_1 \cup C_2$ is convex.
**b.** If $C_1$ and $C_2$ are (not necessarily convex) cones, then $C_1 \cup C_2$ is a (not necessarily convex) cone.
**c.** Consider the set of points $\boldsymbol{x}$ defined by finitely many linear inequalities, i.e., the set of points $\boldsymbol{x}$ defined by $\boldsymbol{A}\boldsymbol{x} \leq \boldsymbol{b}$, where $\boldsymbol{A}$ and $\boldsymbol{b}$ are an arbitrary matrix and vector, respectively, of conforming size, and the inequality is true elementwise. Then this set is convex.
**d.** A convex set $C$ is bounded.
**e.** A convex set $C$ is closed.
**f.** (**6000-level students only**) If $C_1$ and $C_2$ are convex sets, then conv$(C_1)\cup$conv$(C_2) =$ conv$(C_1 \cup C_2)$.

**Solution**:

**a.** This is false. Take $C_1, C_2 \subset \mathbb{R}$ with $C_1 = [0, 1]$ and $C_2 = [2, 3]$. Then the point 1.5 is on a line segment connecting points $1, 2 \in C_1 \cup C_2$, but $1.5 \notin C_1 \cup C_2$.
**b.** This is true. Let $x \in C_1 \cup C_2$ and $\lambda \geq 0$. We seek to show $\lambda x \in C_1 \cup C_2$. Since $x \in C_1 \cup C_2$, then either $x \in C_1$ or $x \in C_2$. If $x \in C_1$, then $\lambda x \in C_1$ since $C_1$ is a cone. If $x \in C_2$, then $\lambda x \in C_2$ since $C_2$ is a cone. Thus, $\lambda x \in C_1$ or $x \in C_2$ holds, so that $\lambda x \in C_1 \cup C_2$.
**c.** This is true. Suppose $\boldsymbol{x} \in \mathbb{R}^n$ and let $\boldsymbol{A}$ be $m \times n$. Then $\boldsymbol{A}\boldsymbol{x} \leq \boldsymbol{b}$ is true if all $m$ of the inequalities,

$$\boldsymbol{a}_i^T \boldsymbol{x} \leq b_i, \qquad\qquad i = 1, \ldots, m,$$

are true, where $\boldsymbol{a}_i^T$ is the $i$th row of $\boldsymbol{A}$, and $b_i$ is the $i$th component of $\boldsymbol{b}$. We know that the set

$$C_i := \left\{ \boldsymbol{x} \in \mathbb{R}^n \;\middle|\; \boldsymbol{a}_i^T \boldsymbol{x} \leq b_i \right\},$$

is a half-space in $\mathbb{R}^n$ and is hence convex. Thus,

$$\left\{ \boldsymbol{x} \in \mathbb{R}^n \;\middle|\; \boldsymbol{A}\boldsymbol{x} \leq \boldsymbol{b} \right\} = \cap_{i=1}^m C_i,$$

and since each $C_i$ is convex, then the right hand side, being an intersection of convex sets, is itself a convex set.

**d.** This is false. In $\mathbb{R}$, the set $C = \mathbb{R}$ is unbounded but convex.
**e.** This is false. In $\mathbb{R}$, the set $C = (0, 1)$ is an open (not closed) set, but is convex.
**f.** This is false. In $\mathbb{R}$, let $C_1 = [0, 1]$ and $C_2 = [2, 3]$. Then $\text{conv}(C_1) = C_1$ and $\text{conv}(C_2) = C_2$, so that

$$\begin{aligned} \text{conv}(C_1) \cup \text{conv}(C_2) &= [0, 1] \cup [2, 3] \\ \text{conv}(C_1 \cup C_2) &= \text{conv}([0, 1] \cup [2, 3]) = [0, 3], \end{aligned}$$

and these two are clearly not equal.

**P2.** (**6000-level students only**) Consider the set of matrices in $\mathbb{R}^{n\times n}$ given by,

$$S_+(n) \coloneqq \left\{ \boldsymbol{A} \in \mathbb{R}^{n\times n} \;\middle|\; \boldsymbol{A} = \boldsymbol{A}^T \text{ and } \boldsymbol{A} \succeq \boldsymbol{0} \right\}.$$

Prove that $S_+(n)$ is a convex cone.

**Solution**: Let $\boldsymbol{A}, \boldsymbol{B} \in S_+(n)$. Then by definition, for every $\boldsymbol{x} \in \mathbb{R}^n$, we have,

$$\boldsymbol{x}^T \boldsymbol{A} \boldsymbol{x} \geq 0, \qquad\qquad \boldsymbol{x}^T \boldsymbol{B} \boldsymbol{x} \geq 0.$$

To show that $S_+(n)$ is a cone, let $\lambda \geq 0$ be arbitrary. Then for any $\boldsymbol{x} \in \mathbb{R}^n$, $\lambda \boldsymbol{A}$ satisfies,

$$\boldsymbol{x}^T (\lambda \boldsymbol{A}) \boldsymbol{x} = \lambda \boldsymbol{x}^T \boldsymbol{A} \boldsymbol{x} \overset{\lambda\geq 0}{\geq} 0,$$

so that $\lambda \boldsymbol{A} \in S_+(n)$. Thus $S_+(n)$ is a cone.

To show that $S_+(n)$ is convex, let $\lambda \in [0, 1]$. Then for any $\boldsymbol{x} \in \mathbb{R}^n$, $\lambda \boldsymbol{A} + (1 - \lambda)\boldsymbol{B}$ satisfies,

$$\boldsymbol{x}^T (\lambda \boldsymbol{A} + (1 - \lambda)\boldsymbol{B}) \boldsymbol{x} = \lambda \boldsymbol{x}^T \boldsymbol{A} \boldsymbol{x} + (1 - \lambda) \boldsymbol{x}^T \boldsymbol{B} \boldsymbol{x} \overset{\lambda, 1-\lambda \geq 0}{\geq} 0,$$

so that $\lambda \boldsymbol{A} + (1 - \lambda)\boldsymbol{B} \in S_+(n)$. Thus $S_+(n)$ is convex.