Math 171B: Numerical Optimization: Nonlinear Problems

Instructor: Michael Holst

Spring Quarter 2015

Solutions for Homework Assignment #2

Exercise 2.1.* *Sketch $F(x) = 1/x - a$ for $a = 2$. Then derive a Newton iteration for computing the reciprocal of a positive real number a without performing division. Create a Matlab m-file to implement the algorithm, and use it with $x_0 = 0.3$ to approximate e^{-1}, where $e = 2.7182818284$, by calculating x_1, x_2 and x_3. Repeat for $x_0 = 1.0$.*

See the TA for the solution to this problem.

Exercise 2.2.* *A completed m-file newton.m has been placed on the class webpage.*

(a) *Download this function from the class webpage for use in this homework, and read it carefully so that you understand how it works.*

(b) *Use newton.m to find a zero of the function*

$$F(x) = \begin{pmatrix} x_1^2 + x_2^2 - 2 \\ (x_1 - \frac{1}{2})^2 + (x_2 - 1)^2 - \frac{9}{4} \end{pmatrix},$$

starting at the points $x_0 = (2, 3)^T$, $x_0 = (1, 3)^T$ and $x_0 = (1, 2 + 10^{-8})^T$. What rate of convergence do you observe? Comment on your results.

See the TA for the solution to this problem.

Exercise 2.3.* *An eigenvector x of an $n \times n$ matrix A satisfies $Ax = \lambda x$ for some scalar λ. The scalar λ is known as the eigenvalue of A corresponding to the eigenvector x.*

(a) *If x is an eigenvector of A, show that βx is also an eigenvector. What is the associated eigenvalue? Hence show that the unit vector $x/\|x\|$ is an eigenvector of A.*

If $Ax = \lambda x$, then clearly $A(\beta x) = \beta(Ax) = \beta(\lambda x) = \lambda(\beta x)$ holds, simply by the properties of scalar-vector and scalar-matrix multiplication, and by the definition of an eigenpair. Therefore, if x is an eigenvector then so is βx, and both have the same eigenvalue. Moreover, if x is an eigenvector, we know that $x \neq 0$, so that $\|x\| \neq 0$ (property of the norm), so that taking $\beta = 1/\|x\|$ is well-defined. Therefore, if x is an eigenvector, so is $x/\|x\|$, for the same eigevalue.

(b) *Define an iteration of Newton's method for the zero of the $n + 1$ equations*

$$(A - \lambda I)x = 0, \qquad x^T x = 1,$$

in the $n + 1$ unknowns (x, λ). Use the m-file newton.m to find an eigenvector and eigenvalue for the matrix

$$A = \begin{pmatrix} 4 & 2 & 1 \\ 2 & 3 & 0 \\ 1 & 0 & 1 \end{pmatrix}, \text{starting at} x_0 = \begin{pmatrix} \frac{1}{5} \\ -\frac{1}{5} \\ \frac{4}{5} \\ 1 \end{pmatrix}.$$

See the TA for the solution to this part of the problem.

Exercise 2.4. *One way to make Newton's method more robust for a given nonlinear equation $F(x) = 0$, where $F : \mathbb{R}^n \mapsto \mathbb{R}^n$, is to construct an associated minimization problem such that the minimum occurs at the solution to the original nonlinear equation. A standard choice for the function to minimize is:*

$$f(x) = \frac{1}{2}\|F(x)\|^2 = \frac{1}{2}F(x)^T F(x).$$

This is a real-valued function of several variables, and we can form its derivatives, as long as $F(x)$ is differentiable.

(a) *Show that the following is true:*

$$\nabla f(x) = F'(x)^T F(x).$$

We can either compute each component $\partial f(x)/\partial x_k$, $k = 1, \ldots, n$ of the gradient vector $\nabla f(x)$, or we can compute them all at once using the technique in class that was called *variational differentian*, which also goes by the name *directional* or *Gateaux* differentiation:

$$\begin{aligned}
(\nabla f(x), p) &= \frac{d}{dt} f(x + tp)\Big|_{t=0} \\
&= \frac{d}{dt}\frac{1}{2}(F(x + tp), F(x + tp))_{t=0} \\
&= \frac{1}{2}\frac{d}{dt}(F(x) + tF'(x)p + O(t^2)), F(x) + tF'(x)p + O(t^2)))_{t=0} \\
&= \frac{1}{2}\frac{d}{dt}(F(x) + tF'(x)p + O(t^2)), F(x) + tF'(x)p + O(t^2)))_{t=0} \\
&= \frac{1}{2}\frac{d}{dt}[(F(x), F(x)) + t(F(x), F'(x)p) + O(t^2) \\
&\quad + t(F'(x)p, F(x)) + t^2(F'(x)p, F'(x)p) + O(t^3)]_{t=0} \\
&= \frac{1}{2}[(F(x), F'(x)p) + (F'(x)p, F(x) + 2t(F'(x)p, F'(x)p) + O(t^2)]_{t=0} \\
&= \frac{1}{2}[(F(x), F'(x)p) + (F'(x)p, F(x)] \\
&= (F(x), F'(x)p) \\
&= (F'(x)^T F(x), p).
\end{aligned}$$

Therefore, we have that $\nabla f(x) = F'(x)^T F(x)$.

(b) *Now, recall that a* direction of decrease *y at x for such a real-valued function satisfies*

$$f(x + \lambda y) < f(x),$$

for some $\lambda > 0$. More over, a descent direction *y at x always satisfies $y^T \nabla f(x) < 0$. Show that a descent direction is always a direction of decrease. (*Hint: Taylor series.*)*

We did this in class, which was stated as a lemma. The proof was essentially just writing down the Taylor expansion:

$$f(x + \lambda y) = f(x) + \lambda f'(x)y + O(\lambda^2) = f(x) + \lambda[f'(x)y + O(\lambda)]. \tag{2.1}$$

If $f'(x)y = y^T \nabla f(x) < 0$, then for sufficient small (but positive) $\lambda > 0$, we can force:

$$f'(x)y + O(\lambda) < 0,$$

which in term forces:

$$\lambda[f'(x)y + O(\lambda)] < 0,$$

and so that finally equation (2.1) ensures that for this same λ,

$$f(x + \lambda y) < f(x).$$

(c) *Prove that the Newton direction $y = -F'(x)^{-1}F(x)$ at x is always a direction of decrease for $f(x)$ at x.*
(Hint: Show y is a descent direction and use (b).)

We just need to show that

$$f'(x)y = y^T\nabla f(x) = (\nabla f(x), y) < 0,$$

for our specific $f(x) = \|F(x)\|^2/2$, and our specific $y = -F'(x)^{-1}F(x)$. So let us check that:

$$\begin{aligned}
(\nabla f(x), y) &= (F'(x)^T F(x), -F'(x)^{-1}F(x)) \\
&= -(F'(x)^T F(x), F'(x)^{-1}F(x)) \\
&= -(F(x), F'(x)F'(x)^{-1}F(x)) \\
&= -(F(x), [F'(x)F'(x)^{-1}]F(x)) \\
&= -(F(x), F(x)) \\
&= -\|F(x)\|^2 \\
&< 0,
\end{aligned}$$

as long as $F(x) \neq 0$ (i.e., x is not a solution to $F(x) = 0$). So, we have show that the Newton direction y is a descent direction for this particular $f(x)$, and by our lemma in the previous part, this also means that the Newton direction y is a direction of decrease for this $f(x)$. That means we can do *back-tracking* (or *damping*) as described in class to make Newton's method more globally convergent.

(d) *If you don't solve the Newton equations exactly (e.g., you are left with some residual $r = -F(x) - F'(x)\delta \neq 0$), how does this effect the situation?*

We again just need to show that

$$f'(x)y = y^T\nabla f(x) = (\nabla f(x), y) < 0,$$

for our specific $f(x) = \|F(x)\|^2/2$, but now we have to use a potentially incorrect Newton direction: $y = -F'(x)^{-1}[F(x) - r]$. (From above, $r = -F(x) - F'(x)y$, and we have just solved for the direction y.) Let's just check things as we did earlier:

$$\begin{aligned}
(\nabla f(x), y) &= (F'(x)^T F(x), -F'(x)^{-1}[F(x) - r]) \\
&= -(F'(x)^T F(x), F'(x)^{-1}[F(x) - r]) \\
&= -(F(x), F'(x)F'(x)^{-1}[F(x) - r]) \\
&= -(F(x), [F'(x)F'(x)^{-1}][F(x) - r]) \\
&= -(F(x), F(x) - r) \\
&= -[\|F(x)\|^2 - (F(x), r)].
\end{aligned}$$

Since $\|F(x)\|$ is always non-negative when x is not a solution to $F(x) = 0$, the only thing we have to worry about is making sure that r is small enough so that:

$$\|F(x)\|^2 - (F(x), r) > 0,$$

when $F(x) \neq 0$. We can ensure this by enforcing the necessary and sufficient condition:

$$(F(x), r) < \|F(x)\|^2.$$

A simpler sufficient condition comes from enforcing the right-most inequality below:

$$(F(x), r) \leq \|F(x)\|\|r\| < \|F(x)\|^2,$$

which after division by $\|F(x)\|$ (again, for $F(x) \neq 0$) leads to:

$$\|r\| < \|F(x)\|. \tag{2.2}$$

If one solves the linear system in Newton iteration accurately enough so that (2.2) holds, then the back-tracking step for moving downhill in $f(x) = \|F(x)\|$ is still mathematically guaranteed to work.

Exercise 2.5. *Consider a vector-valued function F. The back-tracking step length criterion discussed in Professor Gill's notes (pages 62-63) enforces* sufficient descrease *in the norm of the nonlinear function. It requires that the reduction in $\|F(x)\|$ is not worse than μ times the reduction in $\|L_k(x)\|$, i.e.,*

$$\frac{\|F(x_k)\| - \|F(x_k + \alpha p_k)\|}{\|L_k(x_k)\| - \|L_k(x_k + \alpha p_k)\|} \geq \mu,$$

where μ is a pre-assigned constant such that $0 < \mu < 1$. Let p_k be the Newton step associated with a point x_k at which $F'(x_k)$ is nonsingular. Show that this condition on α_k is equivalent to the condition

$$\|F(x_k + \alpha_k p_k)\| \leq (1 - \alpha_k \mu)\|F(x_k)\|. \tag{2.3}$$

Hint: *Read pages 59-63 of Professor Gill's notes!*

The solution is contained entirely in the printed class notes, around page 62, above equation 2.6.4.

Exercise 2.6.* *Let $F : \mathbb{R}^n \to \mathbb{R}^n$ be continuously differentiable in $\mathbb{R}^n$. Write a* MATLAB *m-file that finds the zero of a function F using Newton's method with backtracking. I.e., you should modify the m-file* `newton.m` *from the class webpage so that it implements Algorithm 2.6.1 in Professor Gill's notes (page 63). Use $\mu = \frac{1}{4}$ to define the sufficient-decrease criterion, equation (2.3), in the backtracking algorithm. The step length α_k is chosen as the first member of the sequence $1, \frac{1}{2}, \frac{1}{4}, \frac{1}{8}, \ldots,$ that satisfies (2.3). The algorithm should be terminated when either $\|F(x_k)\| \leq 10^{-8}$, or 50 iterations are performed. The backtracking "*`while`*" or "*`for`*" loop must include a test that will terminate the loop if it is executed more than 20 times (this will keep you for burning a lot of CPU time if something goes wrong...) At each iteration, print k, α_k* and *$\|F(x_k)\|$.*

Use your m-file to find a zero the function

$$F(x) = \begin{pmatrix} e^{x_1}(4x_1^2 + 2x_2^2 + 4x_1x_2 + 6x_2 + 8x_1 + 1) \\ e^{x_1}(4x_2 + 4x_1 + 2) \end{pmatrix},$$

starting at $x_0 = (3,\, 0)^T$.

See the TA for the solution to this problem. One helpful tip is that in MATLAB, to call a function in another file, you need to put an "@" sign in front of the function name.