Solution to Homework #2

1. (#2.9) Let \( \hat{f}(z) = f(x) \) with \( x = Sz + c \). Then

\[
\frac{\partial \hat{f}}{\partial z_i} = \sum_{j=1}^{n} \frac{\partial f}{\partial x_j} \frac{\partial x_j}{\partial z_i}.
\]

Since \( x = Sz + c \), then \( \frac{\partial x_j}{\partial z_i} = S_{ji} \) and so

\[
\frac{\partial \hat{f}}{\partial z_i} = \sum_{j=1}^{n} S_{ji} \frac{\partial f}{\partial x_j} = (S^T \nabla f(x))_i.
\]

Thus \( \nabla \hat{f}(z) = S^T \nabla f(x) \). In a similar fashion, we get that

\[
\frac{\partial^2 \hat{f}}{\partial z_i \partial z_j} = \sum_{k=1}^{n} S_{kj} \frac{\partial S^T \nabla f(x)}{\partial x_k} = (S^T \nabla^2 f(x) S)_{ij}.
\]

So \( \nabla^2 \hat{f}(z) = S^T \nabla^2 f(x) S \).

2. (#3.10) Let \( q(\alpha) \) be a quadratic function that matches the given data. Then

\[
\phi_q(\alpha) = \left( \frac{\phi(\alpha_0) - \phi(0) - \alpha_0 \phi'(0)}{\alpha_0^2} \right) \alpha^2 + \phi'(0) \alpha + \phi(0).
\]

Then \( \phi_q(0) = \phi(0), \phi'_q(0) = \phi'(0) \) and

\[
\phi_q(\alpha_0) = \phi(\alpha_0) - \phi(0) - \alpha_0 \phi'(0) + \phi'(0) \alpha_0 + \phi(0) = \phi(\alpha_0).
\]

Since \( \phi_q \) is quadratic and matches the 3 values as determined above, it is the quadratic interpolator of this data.

Suppose \( \phi(\alpha_0) > \phi(0) + c_1 \alpha_0 \phi'(0) \) for some \( c_1 < 1 \). Then we have

\[
\frac{\phi(\alpha_0) - \phi(0) - \alpha_0 \phi'(0)}{\alpha_0^2} > \frac{\phi(0) + c_1 \alpha_0 \phi'(0) - \phi(0) - \alpha_0 \phi'(0)}{\alpha_0^2} = \frac{(c_1 - 1) \alpha_0 \phi'(0)}{\alpha_0^2} > 0,
\]

since \( \phi'(0) < 0 \). So the function \( \phi_q \) has positive curvature and is minimized by

\[
\alpha_1 = \frac{1}{2} \frac{-\alpha_0^2 \phi'(0)}{\phi(\alpha_0) - \phi(0) - \alpha_0 \phi'(0)}.
\]

Now, assuming \( \alpha_0 \) still does not satisfy the descent condition, we have that

\[
\alpha_1 < \frac{1}{2} \frac{-\alpha_0^2 \phi'(0)}{\phi(0) + c_1 \alpha_0 \phi'(0) - \phi(0) - \alpha_0 \phi'(0)} = \frac{-\alpha_0^2 \phi'(0)}{2(c_1 - 1) \alpha_0 \phi'(0)} = \frac{\alpha_0}{2(1 - c_1)}.
\]

3. Let \( \phi(\alpha) = (\alpha - 1)^2 \) so \( \phi'(\alpha) = 2(\alpha - 1) \).

(a) The Wolfe Conditions are: (1) \( \phi(\alpha) \leq \phi(0) + c_1 \alpha \phi'(0) \) and (2) \( \phi'(\alpha) \geq c_2 \phi'(0) \), with \( 0 < c_1 < c_2 < 1 \). For the given \( \phi \) the conditions become (using \( x \) for \( \alpha \)): (1) \( (x - 1)^2 \leq 1 - 2c_1 x \) and (2) \( 2(x - 1) \geq -2c_2 \). Using the fact that \( x > 0 \) and simple algebra, these simplify to (1) \( x \leq 2(1 - c_1) \) and (2) \( x \geq 1 - c_2 \). Thus the range of \( x \) values which satisfy...
the Wolfe Conditions is \(1 - c_2 \leq x \leq 2(1 - c_1)\). The condition that this interval be non-empty is that \(1 - c_2 \leq 2(1 - c_1)\) or \(c_2 \geq 2c_1 - 1\). However if we enforce the condition that \(c_1 < c_2 < 1\) then \(2c_1 < c_2 + 1\) and thus the condition hold automatically.

Now, the global minimizer of \(\phi\) is \(x = 1\) and clearly \(1 \geq 1 - c_2\) for all \(c_2 > 0\). But, \(1 \leq 2(1 - c_1)\) only when \(c_1 \leq 1/2\). Often you will see this restriction \((c_1 \leq 1/2)\) added when the method uses some form of quadratic approximation so that if the exact minimizer of the model or of \(\phi\) is the minimizer of the function, it can be found.

(b) The Goldstein Condition can be broken into two parts: (1) \(\phi(\alpha) \geq \phi(0) + (1 - c)\alpha\phi'(0)\) and (2) \(\phi(\alpha) \leq \phi(0) + c\alpha\phi'(0)\), with \(0 < c < \frac{1}{2}\). Using the given \(\phi\), we have \((x - 1)^2 \geq 1 - 2x(1 - c)\) and \((x - 1)^2 \leq 1 - 2cx\). Simplifying each of these we get \(x \geq 2c\) and \(x \leq 2(1 - c)\). Thus the range of values is \(2c \leq \alpha \leq 2(1 - c)\), which is non-empty as long as \(c \leq \frac{1}{2}\), and, in this case, also includes \(\alpha = 1\).

4. Let \(n\) be a positive integer and set \(f(x) = \sum_{i=1}^{n} f_i(x)^2\) where

\[
f_i(x) = n - \sum_{j=1}^{n} (\cos x_j + i(1 - \cos x_i) - \sin x_i).
\]

For this form of \(f\), \(\nabla f_i(x) = 2 \sum_{i=1}^{n} f_i(x) \nabla f_i(x)\) and \(\nabla^2 f(x) = 2 \sum_{i=1}^{n} \nabla f_i(x) \nabla f_i(x)^T + f_i \nabla^2 f_i(x)\) (Note the use of the outer product). So we need \(\nabla f_i\) and \(\nabla^2 f_i\). First rewrite \(f_i\):

\[
f_i(x) = n(1 - i) + nv_i - \sigma,
\]

where \(v_i = i \cos x_i + \sin x_i\) and \(\sigma = \sum_{j=1}^{n} \cos x_j\). Letting \(u_i = -i \sin x_i + \cos x_i\) we get

\[
(\nabla f_i)_k = nu_i \delta_{ik} + \sin x_k \quad \text{and} \quad (\nabla^2 f_i)_{k,l} = \cos x_k \delta_{kl} - nv_i \delta_{ik} \delta_{il},
\]

where \(\delta_{ij} = 1\) if \(i = j\) and \(0\) otherwise. Then we have \(\nabla f_i = s + nu_i e_i\) and \(\nabla^2 f_i = C - nv_i e_i e_i^T\) where \(s = (\sin x_1, \ldots, \sin x_n)\), \(C\) is a diagonal matrix with \(C_{kk} = \cos x_k\) and \(e_i\) is the \(i\)th column of \(I\). Thus

\[
\nabla f = 2 \sum_{i=1}^{n} f_i \nabla f_i
\]

\[
= 2 \sum_{i=1}^{n} f_i (s + nu_i e_i)
\]

\[
= 2s \sum_{i=1}^{n} f_i + 2n \sum_{i=1}^{n} u_i f_i e_i
\]

and

\[
\nabla^2 f = 2 \sum_{i=1}^{n} \nabla f_i \nabla f_i^T + f_i \nabla^2 f_i
\]

\[
= 2 \sum_{i=1}^{n} (s + nu_i e_i)(s + nu_i e_i)^T + f_i (C - nv_i e_i e_i^T)
\]

\[
= 2nss^T + 2n \sum_{i=1}^{n} u_i (e_i s^T + se_i^T) + 2n^2 \sum_{i=1}^{n} u_i^2 e_i e_i^T + 2C \sum_{i=1}^{n} f_i - 2n \sum_{i=1}^{n} v_i f_i e_i e_i^T.
\]
To express these more efficiently, let $F = \sum_{i=1}^{n} f_i$, $v(w_i)$ be the vector with entries $w_i$ and $\Delta(w_i)$ be the diagonal matrix with entries $w_i$. Then we have

$$\nabla f = 2Fs + 2n\nu(u_i f_i) \quad \text{and} \quad \nabla^2 f = 2nss^T + 2n\sum_{i=1}^{n} u_i(e_i s^T + se_i^T) + 2n^2 \Delta(u_i^2) + 2FC - 2n\Delta(v_i f_i).$$

In this form, it is fairly easy to compute the value for $n = 3$ with $x_0 = (1/3, 1/3, 1/3)$, and we get

$$\nabla f(x_0) = \begin{pmatrix} 5.2412 \\ 3.0263 \\ 1.4597 \end{pmatrix} \quad \text{and} \quad \nabla^2 f(x_0) = \begin{pmatrix} 7.0740 & 2.4255 & 1.7832 \\ 2.4255 & -2.9290 & 1.1409 \\ 1.7832 & 1.1409 & -7.2054 \end{pmatrix}.$$ 

5. Let $A = \begin{pmatrix} 2 & -1 \\ -1 & 2 \end{pmatrix}$. Define $f(x) = \frac{1}{2}x^T Ax$ for $x \in \mathbb{R}^2$, with unique minimizer $x^* = (0, 0)^T$. 

(a) If SD is to converge in one step we at least need $x_0 - t\nabla f(x_0)$ to intersect $x^*$ for some $t > 0$. If we set $x_0 = (x, y)^T$, then we have $(x, y)^T - tA(x, y)^T = (x - t(2x - y), y - t(-x + 2y))$. Setting each component equal to 0, we get that $t = x/(2x - y) = y/(-x + 2y)$, thus $(x, y)$ must satisfy $-x^2 + 2xy = 2xy - y^2$ or $x^2 = y^2$. The only such pair that also satisfy $x^2 + y^2 = 1$ are $(\beta, \beta), (\beta, -\beta), (-\beta, \beta)$ and $(-\beta, -\beta)$ where $\beta = 1/\sqrt{2}$. With Exact Line Search from these starting values, we get $\alpha = -x_0^TA^2x_0/(x_0^TA^3x_0) = 1$ or $1/3$ depending on $x_0$ and this is exactly the same as the value for $t$ for the same pairs. Thus these are the only 4 starting values that converge in 1 step.

(b) For Newton’s Method, the direction is $p = -(\nabla^2 f(x_0))^{-1}\nabla f(x_0)$. For this $f$, we get $p = -(A)^{-1}Ax_0 = -x_0$. Thus with a step-length of 1, we have $x_+ = x_0 + (1)(-x_0) = 0$. Thus for all starting values $x_0$ with $\|x_0\|_2 = 1$ Newton’s method converges in 1 step.

(c) (Bonus) The answer is that there are no starting values that converge to the exact answer in a $k$ steps for any $k > 1$. There are many ways to come up with the answer, but probably the easiest way to see it is to try to construct a 2-step solution. We know from part (a) that the 2nd step has to come from a point $x_1 = (a, b)^T$ where $a^2 = b^2$. We also know from the properties of SD that the step from $x_0$ to $x_1$ is along a direction perpendicular to the direction from $x_1$ to $x^*$. Since the line from $x_1$ to $x^*$ has slope $\pm 1$, the slope of the line from $x_0$ to $x_1$ must have slope $\mp 1$. However, the only points at which the gradient produces a line with slope $\mp 1$ are exactly the points we found in part (a), and those points converge in 1 step.