Problem 1 [BV 9.8]

By definition we have:

$$\begin{aligned}\Delta x_{nsd} &= \arg\min\{\nabla f(x)^T v \;\mid\; \|v\|_\infty \leq 1\} \\ &= \arg\min\{\nabla f(x)^T v \;\mid\; \max\{|v_1|, \dots, |v_n|\} \leq 1\} \\ &= \arg\min\{\frac{\partial f}{\partial x_1}(x)v_1 + \cdots + \frac{\partial f}{\partial x_n}(x)v_n \;\mid\; \max\{|v_1|, \dots, |v_n|\} \leq 1\}\end{aligned}$$

since each $-1 \leq v_i \leq 1$, if we choose $v_i = -\frac{\frac{\partial f}{\partial x_i}(x)}{|\frac{\partial f}{\partial x_i}(x)|}$ when $\frac{\partial f}{\partial x_i}(x) \neq 0$ and $v_i = 0$ otherwise, the objective function will be minimized. In other words, v_i is always the negative of the sign of the ith element of the gradient.

On the other hand, since the dual of l_∞-norm is l_1-norm, we have:

$$\begin{aligned}\Delta x_{sd} &= \|\nabla f(x)\|_1 \Delta x_{nsd} \\ &= \left(\sum_{i=1}^{n} |\frac{\partial f}{\partial x_i}(x)|\right) \Delta x_{nsd}\end{aligned}$$

where elements of Δx_{nsd} are in $\{-1, 0, 1\}$. This choice of steepest direction always points to a direction negative of the direction (sign) of gradient on each coordinate, unless the gradient is zero in that coordinate.

Problem 2 [BV 9.10]

(a)

$$\begin{aligned}f(x) &= \log(e^x + e^{-x}) \\ f'(x) &= \frac{e^x - e^{-x}}{e^x + e^{-x}} \\ f''(x) &= \frac{4}{(e^x + e^{-x})^2} \\ x^+ &= x - \frac{1}{f''(x)} f'(x)\end{aligned}$$

The Newton's updates with different initial points are reported below, it shows that in this case of fixed step size, a small deviation in the initial point could lead to non-convergence of the algorithm:

iteration (k)	0	1	2	3	4	5
$x^{(k)}$	1	-0.8134	0.4094	-0.0473	7.0602×10^{-5}	-2.3470×10^{-13}
$f(x^{(k)})$	1	0.9928	0.7747	0.6942	0.6931	0.6931

iteration (k)	0	1	2	3	4	5
$x^{(k)}$	1.1	-1.1285	1.2341	-1.6951	5.7153	-23021.3564
$f(x^{(k)})$	1.00468	1.2280	1.3154	1.7283	5.7153	23021.3564

(b)

$$\begin{aligned} f(x) &= -\log x + x \\ f'(x) &= -\frac{1}{x} + 1 \\ f''(x) &= \frac{1}{x^2} \\ x^+ &= x - \frac{1}{f''(x)} f'(x) \end{aligned}$$

In this case after a single iteration, the Newton's update get outside of the domain of function.

iteration (k)	0	1	2	3	4	5
$x^{(k)}$	3	-3	-	-	-	-
$f(x^{(k)})$	1.9013	not defined	-	-	-	-

Problem 3 [BV 9.11]

1. Gradient Method.

$$\nabla g(x) = \phi'(f(x)) \nabla f(x)$$

Since ϕ is increasing ($\phi'(f(x)) \geq 0$), then the gradient of g is always in the same direction as the gradient of f. Therefore both of them will have the same search direction and when the exact line search is used, minimizing both of the will have the same iterations.

2. Newton Method.

$$
\begin{aligned}
&\nabla^2 g(x)^{-1}\nabla g(x)\\
&= [\phi''\nabla f(x)\nabla f(x)^T + \phi'\nabla^2 f(x)]^{-1}\nabla g(x)\\
&= \left[\frac{1}{\phi'}\nabla^2 f^{-1} - \frac{1}{\phi'}\nabla^2 f^{-1}\phi''\nabla f(I + \frac{1}{\phi'}\nabla f^T\nabla^2 f^{-1}\phi''\nabla f)^{-1}\nabla f^T\frac{1}{\phi'}\nabla^2 f^{-1}\right]\nabla g(x)\\
&= \left[\frac{1}{\phi'}\nabla^2 f^{-1} - \frac{1}{\phi'}\nabla^2 f^{-1}\phi''\nabla f(I + \frac{\phi''}{\phi'}p)^{-1}\nabla f^T\frac{1}{\phi'}\nabla^2 f^{-1}\right]\nabla g(x)\\
&= \left[\frac{1}{\phi'}\nabla^2 f^{-1} - \frac{\phi''}{\phi'(\phi' + p\phi'')}\nabla^2 f^{-1}\nabla f\nabla f^T\nabla^2 f^{-1}\right]\phi'\nabla f\\
&= \nabla^2 f^{-1}\left[I - \frac{\phi''}{\phi' + p\phi''}\nabla f\nabla f^T\nabla^2 f^{-1}\right]\nabla f\\
&= \nabla^2 f^{-1}\left[\nabla f - \frac{\phi''}{\phi' + p\phi''}\nabla f\nabla f^T\nabla^2 f^{-1}\nabla f\right]\\
&= \nabla^2 f^{-1}\nabla f\left[I - \frac{\phi''}{\phi' + p\phi''}\nabla f^T\nabla^2 f^{-1}\nabla f\right]\\
&= \nabla^2 f^{-1}\nabla f\left[I - \frac{\phi''}{(\phi' + p\phi'')}p\right]\\
&= \nabla^2 f^{-1}\nabla f\left[\frac{\phi'}{\phi' + p\phi''}\right]
\end{aligned}
$$

where the second equality follows from Matrix Inversion Lemma (C.4.3) and $p := \nabla f^T\nabla^2 f^{-1}\phi''\nabla f$ is a nonnegative scalar, as f is a convex function. Since ϕ is increasing and convex, the coefficient $\frac{\phi'}{\phi' + p\phi''}$ is non-negative. Therefore both of f and g will have the same Newton search direction and when the exact line search is used, minimizing both of them will result in same iterations.
