CS345: Optimization
Homework 2: Solutions
Due Thursday, April 12

2.1

\[ \nabla f(x) = \begin{bmatrix}
-400(x_2 - x_1^2)x_1 - 2(1 - x_1) \\
200(x_2 - x_1^2)
\end{bmatrix}, \]

\[ \nabla^2 f(x) = \begin{bmatrix}
-400(x_2 - x_1^2) + 800x_1^2 + 2 & -400x_1 \\
-400x_1 & 200
\end{bmatrix}. \]

At \( x^* = (1,1) \), we have

\[ \nabla(x^*) = 0, \nabla^2 f(x^*) = \begin{bmatrix}
802 & -400 \\
-400 & 200
\end{bmatrix}. \]

To find the eigenvalues of the Hessian we solve \( \det(\nabla^2 f(x^*) - \lambda I) = 0 \). We get

\[(802 - \lambda)(200 - \lambda) - 160000 = \lambda^2 - 1002\lambda - 400,\]

which yields

\[ \lambda = \frac{1002 \pm \sqrt{1002^2 - 1600}}{2} \]

which are both positive values. Hence the Hessian is positive definite.

2.2 Since

\[ \nabla f(x) = \begin{bmatrix}
8 + 2x_1 \\
12 - 4x_2
\end{bmatrix}, \]

the only point satisfying first-order conditions is \( x = (-4,3) \). The Hessian (which is constant) is

\[ \nabla^2 f(x) = \begin{bmatrix}
2 & 0 \\
0 & -4
\end{bmatrix}. \]

It is obviously indefinite with a positive eigenvalue 2 and a negative eigenvalue \(-4\). Hence the stationary point \( x = (-4,3) \) is a saddle point.
2.3 We can write
\[ f_1(x) = \sum_{i=1}^{n} a_i x_i, \]
each \( x_i \) appears in just one of the \( n \) terms of this summation. Differentiation yields
\[ \frac{\partial f_1}{\partial x_i} = a_i, \quad i = 1, 2, \ldots, n, \]
so that
\[ \nabla f_1(x) = \begin{bmatrix} a_1 \\ a_2 \\ \vdots \\ a_n \end{bmatrix}, \quad \nabla^2 f_1 = 0. \]

For \( f_2 \), we have
\[ f_2(x) = \sum_{i=1}^{n} \sum_{j=1}^{n} a_{ij} x_i x_j, \]
where \( a_{ij} = a_{ji} \), for all \( i \) and \( j \), by symmetry of \( A \).
Let us compute the first term in the gradient, that is \( \partial f_2 / \partial x_1 \). We find that \( x_1 \) appears in \( 2n - 1 \) terms of this summation, that is,
\[ f_2(x) = a_{11} x_1^2 + a_{12} x_1 x_2 + \cdots + a_{1n} x_1 x_n + a_{21} x_2 x_1 + a_{23} x_3 x_1 + \cdots + a_{n1} x_n x_1 + \cdots \]
\[ = a_{11} x_1^2 + 2a_{12} x_1 x_2 + 2a_{13} x_1 x_3 + \cdots + 2a_{1n} x_1 x_n + \cdots. \]
Therefore we have
\[ \frac{\partial f_2}{\partial x_1} = 2a_{11} x_1 + 2a_{12} x_2 + \cdots + 2a_{1n} x_n, \]
which is the first element of the vector \( 2Ax \). Similar reasoning for the other elements of \( \nabla f_2 \) show that in fact \( \nabla f_2(x) = 2Ax \).
For second derivatives, we differentiate the above expression with respect to \( x_i \), for \( i = 1, 2, \ldots, n \), to obtain
\[ \frac{\partial^2 f_2}{\partial x_i \partial x_i} = 2a_{ii}, \quad i = 1, 2, \ldots, n, \]
which is twice the first row of the matrix \( A \). Similar reasoning for the other second partial derivatives yields that
\[ \nabla^2 f_2(x) = 2A. \]
2.5 Using a trig identity we find that

\[ f(x_k) = \left( 1 + \frac{1}{2^k} \right)^2 \left( \cos^2 k + \sin^2 k \right) = \left( 1 + \frac{1}{2^k} \right)^2, \]

from which it follows immediately that \( f(x_{k+1}) < f(x_k) \).

Let \( \theta \) be any point in \([0, 2\pi]\). We aim to show that the point \((\cos \theta, \sin \theta)\) on the unit circle is a limit point of \( \{x_k\} \).

From the hint, we can identify a subsequence \( \xi_{k_1}, \xi_{k_2}, \xi_{k_3}, \ldots \) such that \( \lim_{j \to \infty} \xi_{k_j} = \theta \). Consider the subsequence \( \{x_{k_j}\} \). We have

\[
\lim_{j \to \infty} x_{k_j} = \lim_{j \to \infty} \left( 1 + \frac{1}{2^k} \right) \left[ \begin{array}{c}
\cos k_j \\
\sin k_j
\end{array} \right] \\
= \lim_{j \to \infty} \left( 1 + \frac{1}{2^k} \right) \lim_{j \to \infty} \left[ \begin{array}{c}
\cos \xi_{k_j} \\
\sin \xi_{k_j}
\end{array} \right] \\
= \left[ \begin{array}{c}
\cos \theta \\
\sin \theta
\end{array} \right].
\]

2.6 We need to prove that “isolated local min” \( \Rightarrow \) “strict local min.” Equivalently, we prove the contrapositive: “not a strict local min” \( \Rightarrow \) “not an isolated local min.”

Suppose first that \( x^* \) is not even a local min. Then it is certainly not an isolated local min. Now suppose that \( x^* \) is a local min but that it is not strict. Let \( \mathcal{N} \) be any nbd of \( x^* \) such that \( f(x^*) \leq f(x) \) for all \( x \in \mathcal{N} \). Because \( x^* \) is not a strict local min, there is some other point \( x_{\mathcal{N}} \in \mathcal{N} \) such that \( f(x^*) = f(x_{\mathcal{N}}) \). Hence \( x_{\mathcal{N}} \) is also a local min of \( f \) in the neighborhood \( \mathcal{N} \) that is different from \( x^* \). Since we can do this for every neighborhood of \( x^* \) within which \( x^* \) is a local min, \( x^* \) cannot be an isolated local min.

2.7 Let \( x_0 \) and \( x_1 \) both be global minimizers of \( f \). We show that \( x_\alpha \overset{\text{def}}{=} \alpha x_0 + (1 - \alpha) x_1 \), any point on the line between \( x_0 \) and \( x_1 \), is also a global minimizer.

By convexity of \( f \), we have

\[ f(x_\alpha) \leq \alpha f(x_0) + (1 - \alpha) f(x_1) = f(x_0), \]
since $f(x_0) = f(x_1)$. But since $f(x_0)$ is the globally minimizing value of $f$, we also have that $f(x_0) \geq f(x_0)$. We conclude from these two inequalities that $f(x_0) = f(x_0)$, and hence that $x_0$ is also a global minimizer.

2.8 At $x = (1,0)$ we have
\[ \nabla f(x) = \begin{bmatrix} 2 \\ 0 \end{bmatrix}, \]
so that
\[ p^T \nabla f(x) = \begin{bmatrix} -1 \\ 1 \end{bmatrix}^T \begin{bmatrix} 2 \\ 0 \end{bmatrix} = -2 < 0, \]
and $p$ is a descent direction. We have
\[ f(x + \alpha p) = f(1 - \alpha, \alpha) = (1 - \alpha + \alpha^2)^2, \]
so that
\[ \frac{d}{d\alpha} f(x + \alpha p) = 2(1 - \alpha + \alpha^2)(-1 + 2\alpha). \]
We find that $(d/d\alpha)f(x + \alpha p)$ only when the second bracketed term is zero, that is, where $\alpha = 1/2$.

2.9 Note first that
\[ x_j = \sum_{i=1}^n S_{ji}z_i + s_j. \]
By the chain rule we have
\[ \frac{\partial}{\partial z_i} \tilde{f}(z) = \sum_{j=1}^n \frac{\partial f}{\partial x_j} \frac{\partial x_j}{\partial z_i} = \sum_{j=1}^n S_{ji} \frac{\partial f}{\partial x_j} = [S^T \nabla f(x)]_i. \]
For the second derivatives, we apply the chain rule again:
\[ \frac{\partial^2}{\partial z_i \partial z_k} \tilde{f}(z) = \frac{\partial}{\partial z_k} \sum_{j=1}^n S_{ji} \frac{\partial f(x)}{\partial x_j} \]
\[ = \sum_{j=1}^n S_{ji} \frac{\partial^2 f(x)}{\partial x_j \partial x_i} \frac{\partial x_i}{\partial z_k} \]
\[ = \sum_{j=1}^n S_{ji} \frac{\partial^2 f(x)}{\partial x_j \partial x_i} S_{ik} \]
\[ = [S^T \nabla^2 f(x) S]_{ki}. \]
2.12 Obviously $x_k \to 0$. For the rate expression, we have
\[
\frac{\|x_{k+1} - x^*\|}{\|x_k - x^*\|} = \frac{k}{k + 1} \to 1,
\]
so we cannot satisfy the definition of Q-linear convergence.

2.13
\[
\frac{\|x_{k+1} - x^*\|}{\|x_k - x^*\|^2} = \frac{(0.5)^{2k+1}}{(0.5)^{2k}} = (0.5)^{2k+1} = 1 < \infty.
\]
Hence the sequence is Q-quadratic.

2.14
\[
\frac{\|x_{k+1} - x^*\|}{\|x_k - x^*\|} = \frac{k!}{(k+1)!} = \frac{1}{k+1} \to 0
\]
so the convergence is Q-superlinear. For the Q-quadratic test, we have
\[
\frac{\|x_{k+1} - x^*\|}{\|x_k - x^*\|^2} = \frac{k!}{(k+1)!} = \frac{k!}{k+1} \to k(1) = \infty,
\]
So the convergence is not Q-quadratic.

2.15 For $k$ even, we have
\[
\frac{\|x_{k+1} - x^*\|}{\|x_k - x^*\|} = \frac{x_k/k}{x_k} = \frac{1}{k} \to 0,
\]
while for $k$ odd we have
\[
\frac{\|x_{k+1} - x^*\|}{\|x_k - x^*\|} = \frac{(1/4)^{2^{k-1}}}{x_k/k} = \frac{k(1/4)^{2^{k-1}}}{(1/4)^{2^{k-2}}} = k(1/4)^{2^{k-1}} \to 0,
\]
Hence we have
\[
\frac{\|x_{k+1} - x^*\|}{\|x_k - x^*\|} \to 0,
\]
so the sequence is Q-superlinear. The sequence is not Q-quadratic because for $k$ even we have
\[
\frac{\|x_{k+1} - x^*\|}{\|x_k - x^*\|^2} = \frac{x_k/k}{x_k^2} = \frac{1}{k^2} \to \infty.
\]
The sequence is however R-quadratic as it is majorized by the sequence
\[ z_k = (0.5)^{2k}, \; k = 1, 2, \ldots \] For even \( k \), we obviously have
\[ x_k = (0.25)^{2k} < (0.5)^{2k} = z_k, \]
while for \( k \) odd we have
\[ x_k < x_{k-1} = (0.25)^{2k-1} = ((0.25)^{1/2})^{2k} = (0.5)^{2k} = z_k. \]
A simple argument shows that \( z_k \) is Q-quadratic.