6.252 Nonlinear optimization, Recitation 2 (2/13/2015)

The problems are from the course textbook *Nonlinear programming* (2nd ed.), by D.P. Bertsekas.

**Problem 1.2.2** Let \( \beta \) be a nonnegative integer, and let \( f(x) = \|x\|^2 + \beta = (\sum_{i=1}^{n} x_i^2)^{1+\beta/2} \). \( f \) is continuously differentiable and for all \( x \in \mathbb{R}^n \)

\[
\nabla f(x) = (2 + \beta)\|x\|^\beta x.
\]

The iteration of the steepest descent method with constant step size \( s \) is:

\[
\begin{aligned}
x^{k+1} &= x^k - s\nabla f(x^k) \\
&= (1 - s(2 + \beta)\|x^k\|^\beta)x^k \\
&\quad \text{for } k = 0, 1, \ldots \tag{1}
\end{aligned}
\]

We assume that \( s > 0 \) and \( x^0 \neq 0 \) (otherwise the sequence becomes trivial, \( x_{k+1} = x_k \) for all \( k \)). Given the initial point \( x^0 \), we will show that the sequence \( (x^k) \) converges to 0 if and only if we have:

\[
1 - s(2 + \beta)\|x^0\|^\beta > -1. \tag{2}
\]

**Proof.**

- **Assume that (2) holds.** We are first going to show that \( \|x^k\| \) is a monotonically decreasing sequence, i.e., \( \|x^{k+1}\| < \|x^k\| \) for all \( k \in \mathbb{N} \). Let \( \gamma = 1 - s(2 + \beta)\|x^0\|^\beta \) and note that \(-1 < \gamma < 1 \) by assumption (2) and the fact that \( s(2 + \beta)\|x^0\|^\beta > 0 \). By definition of the iteration rule (1) we have \( \|x^1\| = \|\gamma\|\|x^0\| < \|x^0\| \). Using this we can show that \(-1 < 1 - s(2 + \beta)\|x^1\|^\beta < 1 \) since \( 1 - s(2 + \beta)\|x^1\|^\beta > 1 - s(2 + \beta)\|x^0\|^\beta = \gamma > -1 \). Thus by the iteration rule (1) we get that \( \|x^2\| = \|1 - s(2 + \beta)(\|x^1\|^\beta||x^1|| < ||x^1|| \). By iterating we can show that \( \|x^k\| < \|x^{k-1}\| \) for all \( k \), i.e., that \( \|x^k\| \) is a monotonically decreasing sequence. Thus \( \|x^k\| \) is a convergent sequence. Let \( \ell = \lim_{x \to +\infty} \|x^k\| \) and note that \( 0 < \ell < \|x^0\| \). Taking :math:`\|\cdot\|` from each side of the equality (1) and letting \( k \to +\infty \) we see that \( \ell \) has to satisfy the equality: \( \ell = |1 - s(2 + \beta)\ell^\beta|\ell \), i.e., \( \ell(1 - s(2 + \beta)\ell^\beta) = 0 \). This implies that either \( \ell = 0 \) or \( 1 - s(2 + \beta)\ell^\beta = -1 \). The latter is not possible because \( 1 - s(2 + \beta)\ell^\beta > 1 - s(2 + \beta)\|x^0\|^\beta > -1 \). Thus the only solution is \( \ell = 0 \) and thus this shows that \( \lim_{k \to +\infty} \|x^k\| = 0 \).

- **We will now show that if (2) does not hold then the sequence \( \|x^k\| \) does not converge to 0. In fact we will show that the sequence \( \|x^k\| \) is nondecreasing.** Let \( \gamma = 1 - s(2 + \beta)\|x^0\|^\beta \) and note that \( \gamma \leq -1 \).

From the iteration rule (1) note that we have \( \|x^1\| = |\gamma|\|x^0\| \geq \|x^0\| \). From there it follows that \( 1 - s(2 + \beta)\|x^1\|^\beta \leq 1 - s(2 + \beta)\|x^0\|^\beta \leq \gamma \leq -1 \), and thus \( \|x^2\| \geq \|x^1\| \). By iterating one can show that \( \|x^{k+1}\| \geq \|x^k\| \). Thus this shows that \( x^k \) does not converge to 0 (since \( x^0 \neq 0 \)).

\[\square\]

**Remark.** Note that, when \( \beta \neq 0 \), \( \nabla f \) is not Lipschitz continuous on \( \mathbb{R}^n \). Indeed consider for simplicity the case where \( x \) is a scalar variable (\( n = 1 \)). Then a simple calculation shows that the second derivative of \( f \) is:

\[f''(x) = (2 + \beta)(1 + \beta)\text{sign}(x)|x|^{\beta}.\]

If \( \beta \neq 0 \), we can make \( f''(x) \) arbitrarily large by taking \( |x| \) large, and thus \( f' \) is not Lipschitz continuous on \( \mathbb{R} \). Thus, when \( \beta \neq 0 \), the assumption of Proposition 1.2.3 does not apply for our function \( f \) here. Note however that if we restrict the domain of \( f \) to a compact subset \( X \) of \( \mathbb{R}^n \), then \( f'' \) is bounded on \( X \) and so \( f' \) is Lipschitz continuous on \( X \).

**Problem 1.2.12** Let \( x^0 \) be the starting point and we assume that \( v^T x^0 \neq 0 \) where \( v \) is an eigenvector of \( Q \) with a negative eigenvalue, i.e., \( Qv = \lambda v \) where \( \lambda < 0 \). The steepest descent method for \( f(x) = \frac{1}{2}x^T Qx \), with constant step size \( s > 0 \) is

\[
x^{k+1} = x^k - s\nabla f(x^k) = x^k - sQx^k.
\]

Note that

\[
v^T x^{k+1} = v^T x^k - sv^T Qx^k = v^T x^k - s(Qv)^T x^k = (1 - s\lambda)v^T x^k \tag{3}
\]

where we used the fact that \( Q \) is symmetric and that \( Qv = \lambda v \). From (3) it follows that for any \( k \)

\[
v^T x^k = (1 - s\lambda)^k v^T x^0.
\]

Since \( 1 - s\lambda > 1 \) (since \( \lambda < 0 \) and \( s > 0 \)) it follows that \( v^T x^k \) diverges which means that \( x^k \) also diverges.