1. (Polyak) In this problem we compare the convergence results for the conjugate gradient method (lecture 3) with the gradient method (lecture 1). We consider the minimization of a quadratic function

\[ f(x) = \frac{1}{2} x^T A x - b^T x \]

with \( A \) positive definite and \( \lambda_{\text{max}}(A) = L \). From the last expression on page 3-15 we have the following bound on the error after \( k \) iterations:

\[
2(f(x^{(k)}) - f^*) \leq \left( \sum_{i=1}^{n} \frac{d_i^2}{\lambda_i^2} \right) \inf_{\deg(q) \leq k, q(0)=1} \left( \max_{i=1,...,n} \lambda_i q(\lambda_i)^2 \right)
\]

\[
= \|x^*\|^2_2 \inf_{\deg(q) \leq k, q(0)=1} \left( \max_{i=1,...,n} \lambda_i q(\lambda_i)^2 \right).
\]  

(1)

(The second line follows from \( \|A^{-1}d\|_2 = \|QA^{-1}d\|_2 = \|A^{-1}b\|_2 \).) We will use Chebyshev polynomials to construct a polynomial \( q(\lambda) \) to bound the right-hand side of (1). The Chebyshev polynomial of degree \( m \) is denoted \( T_m \), and defined by the recursion

\[
T_0(t) = 1, \quad T_1(t) = t, \quad T_{m+1}(t) = 2tT_m(t) - T_{m-1}(t) \quad \text{for } m \geq 1.
\]

The following two properties will be needed.

- The Chebyshev polynomials of odd degree only contain odd powers of \( t \). The coefficient of \( t \) in \( T_{2k+1}(t) \) is \((-1)^k(2k + 1)\). For example,

\[ T_1(t) = t, \quad T_3(t) = 4t^3 - 3t, \quad T_5(t) = 16t^5 - 20t^3 + 5t, \quad \ldots. \]

- \(|T_m(t)| \leq 1 \) for \( t \in [-1, 1] \).

Verify that the polynomial

\[
q(t) = \frac{(-1)^k}{2k + 1} T_{2k+1}(\sqrt{t/L})
\]

is a polynomial of degree \( k \) and satisfies \( q(0) = 1 \). Use this polynomial in (1) to show that after \( k \) iterations of the conjugate gradient method (started at \( x^{(0)} = 0 \)),

\[
f(x^{(k)}) - f^* \leq \frac{L}{2(2k + 1)^2}\|x^{(0)} - x^*\|_2^2.
\]  

(2)
The corresponding result for the gradient method (page 1-22) with fixed step size \( t = 1/L \) is

\[
f(x^{(k)}) - f^* \leq \frac{L}{2k} \|x^{(0)} - x^*\|^2.
\]

While the bound (2) only holds for quadratic functions, the faster \( 1/k^2 \) convergence has motivated research on accelerated gradient methods.

2. For each of the following functions on \( \mathbb{R}^n \), explain how to calculate a subgradient at a given \( x \).

(a) \( f(x) = \sup_{0 \leq t \leq 1} p(t) \) where \( p(t) = x_1 + x_2 t + \cdots + x_n t^{n-1} \).

(b) \( f(x) = x_{[1]} + x_{[2]} + \cdots + x_{[i]} \) where \( x_{[i]} \) denotes the \( i \)th largest element of \( x \).

(c) \( f(x) = \|Ax - b\|_2 + \|x\|_2 \) where \( A \in \mathbb{R}^{m \times n} \).

(d) \( f(x) = \lambda_{\text{max}}(W + \text{diag}(x)) \) where \( W \in \mathbb{S}^n \).

(e) \( f(x) = \sup_{Ay \preceq b} x^T y \), where \( A \in \mathbb{R}^{m \times n} \) and the polyhedron defined by \( Ay \preceq b \) is nonempty and bounded.