Homework assignment #2

1. Heavy-ball method [Polyak]. We consider a “two-step” variant of the gradient method:

\[ x_{k+1} = x_k - t \nabla f(x_k) + s(x_k - x_{k-1}), \quad k = 1, 2, \ldots, \]

with \( x_1 = x_0 \). The step sizes \( t \) and \( s \) are fixed. The term \( s(x_k - x_{k-1}) \) in the recursion is a momentum term added to suppress the typical zigzagging in the gradient method.

We examine the convergence of the method applied to a strictly convex quadratic function \( f(x) = \frac{1}{2}x^T Ax + b^T x + c \). The notation \( m \) and \( L \) will be used for the smallest and largest eigenvalues of the symmetric positive definite matrix \( A \):

\[ m = \lambda_{\min}(A) > 0, \quad L = \lambda_{\max}(A) \geq m. \]

(a) Verify that the iteration can be written as a linear recursion

\[ z_{k+1} = M z_k + q, \quad k = 1, 2, \ldots, \]

where

\[ z_k = \begin{bmatrix} x_k \\ x_{k-1} \end{bmatrix}, \quad M = \begin{bmatrix} (1 + s)I - tA & -sI \\ I & 0 \end{bmatrix}, \quad q = \begin{bmatrix} -tb \\ 0 \end{bmatrix}. \]

If the sequence converges, the limit \( z^* = M z^* + q \) is \( z^* = (-A^{-1}b, -A^{-1}b) \).

(b) The speed of convergence depends on the spectral radius \( \rho(M) \) of the matrix \( M \). (The spectral radius of a matrix is the largest absolute value of its eigenvalues.) If \( \rho(M) < 1 \), then the iterates \( z_k \) converge to \( z^* \). For large \( k \) the distance \( \|z_k - z^*\| \) decreases as \( \rho(M)^k \).

Express the eigenvalues of \( M \) in terms of the eigenvalues \( \lambda_1, \ldots, \lambda_n \) of \( A \). Show that \( \rho(M) = \sqrt{s} \) if

\[ s < 1, \quad \frac{(1 - \sqrt{s})^2}{m} \leq t \leq \frac{(1 + \sqrt{s})^2}{L}. \quad (1) \]

(c) Find \( s, t \) that minimize the spectral radius subject to the constraints (1). Show that for the optimal step sizes,

\[ \rho(M) = \frac{\sqrt{L - \sqrt{m}}}{\sqrt{L + \sqrt{m}}} = \frac{\sqrt{\gamma} - 1}{\sqrt{\gamma} + 1}, \]

where \( \gamma = L/m \). Compare this with the linear convergence rate of the gradient method on page 1.31 of the lecture notes.
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 $p(t) = x_1 + x_2t + \cdots + x_nt^{n-1}$.

(b) $f(x) = x_{[1]} + x_{[2]} + \cdots + x_{[k]}$ 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.