DATA130026 Optimization
Assignment 9
Due Time: at the beginning of the class, May 11, 2023

1. Let f be a convex and continuous differentiable function over $\mathbb{R}^n$. For a fixed $x \in \mathbb{R}^n$, define the function
$$g_x(y) = f(y) - \nabla f(x)^T y.$$
Suppose ∇f is L Lipschitz continuous, i.e.,
$$\|\nabla f(x) - \nabla f(y)\| \leq L\|x - y\|, \forall x, y \in \mathbb{R}^n.$$
   (a) Prove that x is a minimizer of g_x over $\mathbb{R}^n$.
   (b) Show that for any $x, y \in \mathbb{R}^n$,
$$g_x(x) \leq g_x(y) - \frac{1}{2L}\|\nabla g_x(y)\|^2.$$
   (c) Show that for any $x, y \in \mathbb{R}^n$,
$$f(x) + \nabla f(x)^T (y - x) + \frac{1}{2L}\|\nabla f(x) - \nabla f(y)\|^2 \leq f(y).$$

2. Let $F(x) = Ax + b$ be an affine function, with A an $n \times n$-matrix. What properties of the matrix A correspond to the following conditions (a)-(e) on F? Suppose that A is symmetric, so $F(x)$ is the gradient of a quadratic function
   (a) Monotonicity:
$$(F(x) - F(y))^T (x - y) \geq 0, \ \forall x, y.$$
   (b) Strict monotonicity:
$$(F(x) - F(y))^T (x - y) > 0, \ \forall x, y.$$
   (c) Strong monotonicity (for the Euclidean norm):
$$(F(x) - F(y))^T (x - y) \geq m\|x - y\|_2^2, \ \forall x, y,$$
   where m is a positive constant.
   (d) Lipschitz continuity (for the Euclidean norm):
$$\|F(x) - F(y)\|_2 \leq L\|x - y\|_2, \ \forall x, y,$$
   where L is a positive constant.
   (e) Co-coercivity (for the Euclidean norm):
$$(F(x) - F(y))^T (x - y) \geq \frac{1}{L}\|F(x) - F(y)\|_2^2, \ \forall x, y,$$
   where L is a positive constant.

3. (Heavy-ball method [Polyak])[**Only required for DATA130026h.01.**] 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 $s(x_k - x_{k-1})$ 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 $f(x) = (1/2)x^T Ax + b^T x + c$. 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

$$z_{k+1} = Mz_k + q, \quad k = 1, 2, \ldots$$

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 $z^\star = Mz^\star + q$ is $z^\star = (-A^{-1}b, -A^{-1}b)$.

(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 $z^\star$. For large k the distance $\|z_k - z^\star\|$ 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

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

(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 $\gamma = L/m$.