Homework assignment #1

1. Let \( F(x) = Ax + b \) be an affine function, with \( A \) an \( n \times n \)-matrix. What properties of \( A \) correspond to the following conditions on \( F \)? Distinguish three cases: \( A \) is symmetric, \( A \) is skew-symmetric \( (A + A^T = 0) \), and a general non-symmetric \( A \).

   (a) **Monotonicity:**
   \[
   (F(x) - F(y))^T (x - y) \geq 0 \quad \forall x, y.
   \]

   (b) **Strict monotonicity:**
   \[
   (F(x) - F(y))^T (x - y) > 0 \quad \forall x, y \neq x.
   \]

   (c) **Strong monotonicity:**
   \[
   (F(x) - F(y))^T (x - y) \geq m \|x - y\|_2^2 \quad \forall x, y,
   \]
   where \( m \) is a positive constant.

   (d) **Lipschitz continuity:**
   \[
   \|F(x) - F(y)\|_2 \leq L \|x - y\|_2 \quad \forall x, y,
   \]
   where \( L \) is a positive constant.

   (e) **Co-coercivity:**
   \[
   (F(x) - F(y))^T (x - y) \geq \frac{1}{L} \|F(x) - F(y)\|_2^2 \quad \forall x, y,
   \]
   where \( L \) is a positive constant.

2. **Barzilai-Borwein step sizes.** The gradient update
   \[
   x^{(k)} = x^{(k-1)} - t_k \nabla f(x^{(k-1)})
   \]
   can be interpreted as a variable metric update
   \[
   x^{(k)} = x^{(k-1)} - H^{-1} \nabla f(x^{(k-1)})
   \]
   with a very simple choice \( H = (1/t_k)I \) for the approximate Hessian. Obviously, with this choice of \( H \) it is generally impossible to satisfy the secant condition
   \[
   Hs = y, \quad y = \nabla f(x^{(k-1)}) - \nabla f(x^{(k-2)}), \quad s = x^{(k-1)} - x^{(k-2)}
   \]
as in a quasi-Newton method. However, the variable metric interpretation suggests two possible choices of $t_k$, known as the Barzilai-Borwein step sizes. The first choice minimizes the residual $Hs - y$ in Euclidean norm:

$$\hat{t} = \arg\min_t \| (1/t)s - y \|_2^2.$$ 

The second choice minimizes the norm of $s - H^{-1}y$:

$$\tilde{t} = \arg\min_t \| s - ty \|_2^2.$$ 

Find expressions for $\hat{t}$ and $\tilde{t}$, assuming that $f$ is strictly convex (so that $s^Ty > 0$). Show that if $\nabla f(x)$ is Lipschitz continuous with constant $L > 0$ and $\text{dom } f = \mathbb{R}^n$, then $\hat{t} \geq 1/L$ and $\tilde{t} \geq 1/L$. (This follows from the first inequality on page 1-13 and the first inequality on page 1-15 of the lecture notes.)