Homework assignment #1

1. *Gradient descent and nondifferentiable functions.* We work out the details of the example on page 1-5 of the lecture notes.

   (a) Let $\gamma > 1$. Show that the function

   $$f(x_1, x_2) = \begin{cases} \sqrt{x_1^2 + \gamma x_2^2} & |x_2| \leq x_1 \\ \dfrac{x_1 + \gamma |x_2|}{\sqrt{1+\gamma}} & \text{otherwise} \end{cases}$$

   is convex. You can do this, for example, by verifying that

   $$f(x_1, x_2) = \sup \{x_1 y_1 + \sqrt{\gamma} x_2 y_2 \mid y_1^2 + y_2^2 \leq 1,\ y_1 \geq 1/\sqrt{1+\gamma}\}.$$

   Note that f is unbounded below. (Take $x_2 = 0$ and let x_1 go to $-\infty$.)

   (b) Consider the gradient descent algorithm applied to f, with starting point $x^{(0)} = (\gamma, 1)$ and an exact line search. Show that the iterates are

   $$x_1^{(k)} = \gamma \left(\frac{\gamma - 1}{\gamma + 1} \right)^k, \qquad x_2^{(k)} = \left(-\frac{\gamma - 1}{\gamma + 1} \right)^k.$$

   Therefore $x^{(k)}$ converges to $(0, 0)$.

2. 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.

3. *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. Clearly, with this choice of H it is generally impossible to satisfy the secant condition $Hs = y$, where

$$y = \nabla f(x^{(k-1)}) - \nabla f(x^{(k-2)}), \qquad s = x^{(k-1)} - x^{(k-2)},$$

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:

$$\hat{t} = \underset{t}{\operatorname{argmin}} \|(1/t)s - y\|_2^2.$$

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

$$\tilde{t} = \underset{t}{\operatorname{argmin}} \|s - ty\|_2^2.$$

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 $\mathbf{dom}\, f = \mathbf{R}^n$, 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.)