Homework assignment #3

1. *Relaxation method for linear inequalities* [Agmon, Motzkin, and Schoenberg]. We consider the problem of solving a set of linear inequalities $a_i^T x \leq b_i$, $i = 1, \ldots, m$. This is a special case of the problem on page 3.12 of the lecture notes, with

$$C_i = \{x \mid a_i^T x \leq b_i\}, \quad i = 1, \ldots, m.$$

We assume that the inequalities are strictly feasible, and that $a_i \neq 0$ for all i.

As in the lecture notes, we denote by $f_i(x)$ the Euclidean distance of x to C_i, and by $f(x)$ the maximum of $f_1(x)$, ..., $f_m(x)$:

$$f_i(x) = \max\{0, \frac{a_i^T x - b_i}{\|a_i\|_2}\}, \qquad f(x) = \max\{f_1(x), \ldots, f_m(x)\}.$$

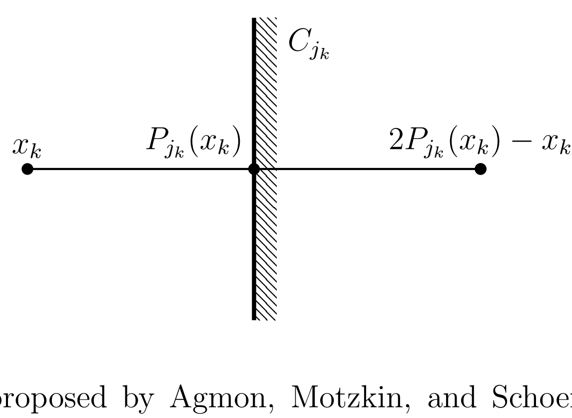
The Euclidean projection of x on the halfspace C_i is denoted by $P_i(x)$:

$$P_i(x) = x - f_i(x) \frac{a_i}{\|a_i\|_2}.$$

The subgradient method with step size $t_k = \lambda f(x_k)$ uses the iteration

$$x_{k+1} = x_k + \lambda(P_{j_k}(x_k) - x_k) \quad \text{where } j_k = \underset{i=1,\ldots,m}{\operatorname{argmax}} \frac{a_i^T x_k - b_i}{\|a_i\|_2}, \tag{1}$$

until x_k is feasible. The constant $\lambda \in (0, 2]$ is an algorithm parameter. If $\lambda = 1$, the new point x_{k+1} is the projection of x_k on the halfspace C_{j_k} farthest from x_k. If $\lambda = 2$, the new point x_{k+1} is the reflection of x_k through the boundary hyperplane of C_{j_k}.



The algorithm was proposed by Agmon, Motzkin, and Schoenberg in 1954. Other variants, with different rules to select j_k (for example, cyclic or random), have also been studied. In the neural network literature, the recursion is known as the perceptron learning algorithm for training linear classifiers.

Motzkin and Schoenberg showed that for $\lambda \in (0, 2)$ the algorithm either finds a solution in a finite number of iterations or converges to a point in the boundary of $C = \cap_{i=1,\ldots,m} C_i$. For $\lambda = 2$ they showed that the algorithm finds a solution in a finite number of iterations. The following is an outline of the proof with some short questions to complete.

(a) Show that the projection $P_i(x)$ on the halfspace C_i satisfies the property

$$\|z - P_i(x)\|_2 \leq \|z - x\|_2 \quad \text{for all } z \in C_i.$$

Use this to show that the iterates (1) satisfy

$$\|z - x_{k+1}\|_2 \leq \|z - x_k\|_2 \quad \text{for all } z \in C.$$

(b) We use the result in part (a) to show that the sequence x_k converges.

A first consequence of (a) is that the iterates x_k are bounded. A standard result from analysis says that every bounded sequence has at least one limit point (a limit of a converging subsequence). To show that the entire sequence converges we show that there is at most one limit point. Consider any $z \in C$. From part (a) the distances $\|x_k - z\|_2$ form a nonincreasing sequence of nonnegative numbers. Therefore this sequence converges to a limit, which we denote by $r(z) = \lim_{k\to\infty} \|x_k - z\|_2$. Every limit point of the sequence x_k must lie on the sphere $\{x \mid \|x - z\|_2 = r(z)\}$. Now suppose $\hat{x}$ and $\tilde{x}$ are two distinct limit points of the sequence x_k. Since $\|\hat{x} - z\|_2 = \|\tilde{x} - z\|_2 = r(z)$, the point z is at the same distance from $\hat{x}$ and $\tilde{x}$. This is true for any $z \in C$. Explain why this contradicts the assumption that the inequalities are strictly feasible, *i.e.*, the polyhedron C has nonempty interior.

(c) Let $\bar{x}$ be the limit of x_k. We show that $\bar{x} \in C$. Verify that the iteration (1) satisfies

$$f(x_k) = \frac{\|x_{k+1} - x_k\|_2}{\lambda}.$$

Since x_k converges, $\lim_{k\to\infty} f(x_k) = 0$. Since the function f is continuous, $f(\bar{x}) = \lim_{k\to\infty} f(x_k) = 0$. Hence $\bar{x} \in C$.

(d) In the last part of the problem we show that if $\lambda = 2$, then $x_k \in C$ after a finite number of iterations. We prove this by contradiction. Suppose $x_k \notin C$ for all k, and let j_k be the index of the halfspace selected in iteration k of (1). Verify that

$$\frac{|a_{j_k}^T \bar{x} - b_{j_k}|}{\|a_{j_k}\|_2} \leq \frac{\|x_{k+1} - x_k\|_2}{2} + \frac{|a_{j_k}^T (\bar{x} - x_k)|}{\|a_{j_k}\|_2}.$$

The left-hand side is the distance of $\bar{x}$ to $H_{j_k} = \{x \mid a_{j_k}^T x = b_{j_k}\}$. The right-hand side converges to zero as $k \to \infty$. Since j_k is chosen from a finite set $\{1, \ldots, m\}$, we must have $\bar{x} \in H_{j_k}$ for all k after some finite number of iterations K. Show that this implies that $\|\bar{x} - x_{k+1}\|_2 = \|\bar{x} - x_k\|_2$ for all $k \geq K$, and therefore $\|\bar{x} - x_k\|_2$ remains constant for $k \geq K$. This contradicts the assumption that x_k is an infinite non-constant sequence with limit $\bar{x}$.

2. For each f, find the subdifferential ∂f, the conjugate f^*, the subdifferential of the conjugate ∂f^*, and verify graphically that ∂f and ∂f^* are inverses.

   (a) $f(x) = \exp(|x|)$.

   (b) $f(x) = -\sqrt{1 - x^2}$ with domain $[-1, 1]$.

   (c) The Huber penalty

   $$f(x) = \begin{cases} x^2/2 & |x| \leq 1 \\ |x| - 1/2 & |x| > 1. \end{cases}$$

   (d) $f(x) = \max\{0, |x| - 1\}$.