Homework 8

This homework is due at 11 PM on October 25, 2023.

Submission Format: Your homework submission should consist of a single PDF file that contains all of your answers (any handwritten answers should be scanned).

1. **Simple Constrained Optimization Problem**

   Consider the optimization problem

$$\min_{x_1, x_2 \in \mathbb{R}} \quad f(x_1, x_2) \tag{1}$$

$$\text{s.t.} \quad 2x_1 + x_2 \geq 1, \tag{2}$$

$$x_1 + 3x_2 \geq 1, \tag{3}$$

$$x_1 \geq 0,\ x_2 \geq 0. \tag{4}$$

   (a) Make a sketch of the feasible set.

   For each of the following objective functions, give the optimizer or set of optimizers, as well as the corresponding optimal value.

   (b) $f(x_1, x_2) = x_1 + x_2$.

   (c) $f(x_1, x_2) = -x_1 - x_2$.

   (d) $f(x_1, x_2) = x_1$.

(e) $f(x_1, x_2) = \max\{x_1, x_2\}$.

(f) $f(x_1, x_2) = x_1^2 + 9x_2^2$.

2. Fun with Hyperplanes

In this problem we work with hyperplanes, which are key components of linear programming as well as future topics such as support vector machines.

(a) Sketch the affine hyperplane $\mathcal{H} \doteq \{\vec{x} \in \mathbb{R}^2 \mid \begin{bmatrix} 1 & 1 \end{bmatrix} \vec{x} = 2\}$.

(b) Let $\vec{c} \in \mathbb{R}^n$ be nonzero, and let $\mathcal{H} \doteq \{\vec{x} \in \mathbb{R}^n \mid \vec{c}^\top \vec{x} = 0\}$. Show that $\mathcal{H}$ is a linear subspace of $\mathbb{R}^n$. What is $\dim(\mathcal{H})$?

(c) Let $\vec{c} \in \mathbb{R}^n$ be nonzero, and let $\mathcal{H} \doteq \{\vec{x} \in \mathbb{R}^n \mid \vec{c}^\top \vec{x} = 0\}$. Suppose $\vec{x}_\star \in \mathbb{R}^n$ is on one side of the hyperplane, i.e., $\vec{c}^\top \vec{x}_\star > 0$. Give any vector which is on the other side of the hyperplane but not on the hyperplane itself.

(d) Let $\vec{c} \in \mathbb{R}^n$ be nonzero, and let $\vec{x}_0 \in \mathbb{R}^n$ be arbitrary. Let $\mathcal{H} \doteq \{\vec{x} \in \mathbb{R}^n \mid \vec{c}^\top (\vec{x} - \vec{x}_0) = 0\}$. Suppose $\vec{x}_\star \in \mathbb{R}^n$ is on one side of this affine hyperplane. Give any vector which is on the other side of $\mathcal{H}$ but not on $\mathcal{H}$ itself.

(e) Let $\vec{x}_0 \in \mathbb{R}^n$ be arbitrary. For a nonzero vector $\vec{c} \in \mathbb{R}^n$, let $\mathcal{H}(\vec{c}) \doteq \{\vec{x} \in \mathbb{R}^n \mid \vec{c}^\top (\vec{x} - \vec{x}_0) = 0\}$. Show that $\vec{0} \in \mathcal{H}(\vec{c})$ for every nonzero $\vec{c} \in \mathbb{R}^n$ if and only if $\vec{x}_0 = \vec{0}$.

3. Quadratic inequalities

Consider the set S defined by the following inequalities:

$$(x_1 \geq -x_2 + 1 \text{ and } x_1 \leq 0) \text{ or } (x_1 \leq -x_2 + 1 \text{ and } x_1 \geq 0)\,. \tag{5}$$

To be more precise,

$$S_1 = \{\vec{x} \in \mathbb{R}^2 \mid x_1 \geq -x_2 + 1, x_1 \leq 0\}; \tag{6}$$

$$S_2 = \{\vec{x} \in \mathbb{R}^2 \mid x_1 \leq -x_2 + 1, x_1 \geq 0\}; \tag{7}$$

$$S = S_1 \cup S_2. \tag{8}$$

(a) Draw the set S. Is it convex?

(b) Show that the set S, can be described via a single quadratic inequality of the form $q(\vec{x}) := \vec{x}^\top A\vec{x} + 2\vec{b}^\top \vec{x} + c \leq 0$, for some $A = A^\top \in \mathbb{R}^{2\times 2}$, $\vec{b} \in \mathbb{R}^2$ and $c \in \mathbb{R}$, i.e., S can be written as $S = \{\vec{x} \in \mathbb{R}^2 \mid q(\vec{x}) \leq 0\})$. Find $A, \vec{b}, c$.
Hint: Can you combine the constraints to make one quadratic constraint?

(c) What is the convex hull of S?

(d) We will now consider some convex optimization problems over S_1 that illustrate the role of the constraints in an optimization problem. For each of the following optimization problems find the optimal point, $\vec{x}^*$. Describe the constraints that are active in attaining the optimal value. *Hint: Suppose that there exists a point $\vec{x}$ such that $\nabla f(\vec{x}) = 0$. From the first order characterization of a convex function $\vec{x}$ would be an optimum value for f subject to no constraints. If $\vec{x}$ is not in the constraint set S_1, then the optimum point must be on the boundary of the set, i.e. it satisfies at least one of the constraints defining S_1 with equality.*

  i. Minimize $f(\vec{x}) = (x_1 + 1)^2 + (x_2 - 3)^2$ subject to $\vec{x} \in S_1$.
  ii. Minimize $f(\vec{x}) = (x_1 + 2)^2 + (x_2 - 2)^2$ subject to $\vec{x} \in S_1$.
  iii. Minimize $f(\vec{x}) = x_1^2 + x_2^2$ subject to $\vec{x} \in S_1$.

4. Gradient Descent Algorithm

Given a continuous and differentiable function $f : \mathbb{R}^n \to \mathbb{R}$, the gradient of f at any point $\vec{x}$, $\nabla f(\vec{x})$, is orthogonal to the level curve of f at point $\vec{x}$, and it points in the increasing direction of f. In other words, moving from point $\vec{x}$ in the direction $\nabla f(\vec{x})$ leads to an increase in the value of f, while moving in the direction of $-\nabla f(\vec{x})$ decreases the value of f. This idea leads to an iterative algorithm to minimize the function f: the gradient descent algorithm.

(a) Consider $f(x) = \frac{1}{2}(x-2)^2$, and assume that we use the gradient descent algorithm:

$$x_{k+1} = x_k - \eta \nabla f(x_k) \quad \forall k \geq 0, \tag{9}$$

with some random initialization x_0, where $\eta > 0$ is the step size of the algorithm. Write $(x_k - 2)$ in terms of $(x_0 - 2)$, and show that x_k converges to 2, which is the unique minimizer of f, when $\eta = 0.2$.

(b) What is the condition that the step size η must satisfy, to ensure that the gradient descent algorithm converges to 2 from all possible initializations in $\mathbb{R}$? What happens if we choose a larger step size?

(c) Now assume that we use the gradient descent algorithm to minimize $f(\vec{x}) = \frac{1}{2}\left\|A\vec{x} - \vec{b}\right\|_2^2$ for some $A \in \mathbb{R}^{m \times n}$ and $\vec{b} \in \mathbb{R}^m$, where A has full column rank. First compute $\nabla f(\vec{x})$. Note that $(A^\top A)^{-1}A^\top \vec{b}$ is the solution to the least-squares problem, and $(\vec{x}_k - (A^\top A)^{-1}A^\top \vec{b})$ can be thought of as the error at time k relative to the solution. Write $(\vec{x}_k - (A^\top A)^{-1}A^\top \vec{b})$ in terms of $(\vec{x}_0 - (A^\top A)^{-1}A^\top b)$.

(d) Now consider $f(\vec{x}) = \frac{1}{2}\left\|A\vec{x} - \vec{b}\right\|_2^2 + \frac{1}{2}\lambda \left\|\vec{x}\right\|_2^2$ for some $A \in \mathbb{R}^{m \times n}$, $\vec{b} \in \mathbb{R}^m$, and scalar $\lambda > 0$, where A has full column rank. Suppose we solve this problem via gradient descent with step-size $\eta = \frac{1}{\sigma_1^2 + \lambda}$, where σ_1 is the maximum singular value of A. Show the gradient descent converges.

5. Gradient Descent Algorithm, continued

Consider $g : \mathbb{R}^n \to \mathbb{R}$, $g(\vec{x}) = \frac{1}{2}\vec{x}^\top M\vec{x} - \vec{x}^\top\vec{b} + c$, where $\vec{b} \in \mathbb{R}^n$, $c \in \mathbb{R}$, and M is a symmetric positive definite matrix, i.e., $M \in \mathbb{S}^n_{++}$.

(a) Write the update rule for the gradient descent algorithm

$$\vec{x}_{k+1} = \vec{x}_k - \eta\nabla g(\vec{x}_k), \tag{10}$$

where η is the step size of the algorithm, and bring it into the form

$$(\vec{x}_{k+1} - \vec{x}_\star) = P_\eta(\vec{x}_k - \vec{x}_\star), \tag{11}$$

where $P_\eta \in \mathbb{R}^{n\times n}$ is a matrix that depends on η. Find $\vec{x}_\star$ and P_η in terms of M, $\vec{b}$, c, and η. *NOTE*: $\vec{x}_\star$ is a minimizer of g.

(b) Write a condition on the step size η and the matrix M that ensures convergence of $\vec{x}_k$ to $\vec{x}_\star$ for every initialization of $\vec{x}_0$.

(c) Assume that all the eigenvalues of M are distinct. Let η_m denote the largest stepsize that ensures convergence for all initializations $\vec{x}_0$, based on the condition computed in part (b).

Does there exist an initialization $\vec{x}_0 \neq \vec{x}_\star$ for which the algorithm converges to the minimum value of g for certain values of the step size η that are larger than η_m?

Justify your answer. *HINT: The question asks if such initializations exist; not whether it is practical to find them.*