DATA 609 - Homework 3: Convex Sets

Eddie Xu

Instructions

Please submit a .qmd file along with a rendered pdf to the Brightspace page for this assignment. You may use whatever language you like within your qmd file, I recommend python, julia, or R.

Problem 1 (cvx-book 2.12):

Which of the following sets are convex? For each case give the reason(s) why or why not

a. A slab, i.e., a set of the form $\{x \in \mathbb{R}^n \mid \alpha \leq \mathbf{a}^T\mathbf{x} \leq \beta\}$.
b. A rectangle, i.e., a set of the form $\{x \in \mathbb{R}^n \mid \alpha_i \leq x_i \leq \beta_i, i = 1, \cdots, n\}$. A rectangle is sometimes called a hyperrectangle when n > 2.
c. A wedge, i.e., $\{\mathbf{x} \in \mathbf{R}^n \mid \mathbf{a_1}^T\mathbf{x} \leq \mathbf{b}_1, \mathbf{a}_2^T \leq \mathbf{b}_2\}$
d. The set of points closer to a given point than a given set, i.e., $\{\mathbf{x} \mid \|\mathbf{x} - \mathbf{x}_0\|^2 \leq \|\mathbf{x} - \mathbf{y}\|^2$ for all $y \in S\}$ where $S \subset \mathbb{R}^n$.
e. The set of points closer to one set than another, i.e., $\{\mathbf{x} \mid dist(\mathbf{x}, S) \leq dist(\mathbf{x}, T)\}$, where S, T are subsets of $\mathbf{R}^N$, and $dist(x, S) = \inf\{\|\mathbf{x} - \mathbf{z}\|^2 | \mathbf{z} \in S\}$.

Problem 1 Solution

a. The slab is a convex set because it is an intersection of two halfspaces.
b. The rectangle is a convex set because the set contains a finite intersection of halfspaces.
c. The wedge set is a convex set because like part a, it is an intersection of two halfspaces.
d. The set of points is a convex set because the set can be expressed as an intersection of halfspaces where $S = \{x \mid \|x - x_0\|_2 \leq \|x - y\|_2\}$ for fixed y
e. The set of points is not a convex set because for $x | dist(x, S) \leq dist(x, T)$, both subset S and T are not convex sets when $x \in R | x \leq -1/2$ or $x \geq 1/2$

Problem 2 (cvx-book 2.15):

Some sets of probability distributions. Let x be a real-valued random variable with probability distribution

Which of the following conditions are convex in $\mathbf{p}$? (That is, for which of the following conditions is the set of $\mathbf{p}$ in P that satisfy the condition convex?) For each case give the reason(s) why or why not.

a. The set of all $\mathbf{p}$ where the expectation of the function $f(x)$ is between two limits: $\alpha \leq Ef(x) \leq \beta$, $Ef(x) = \sum_{i=1}^{n} p_i f(a_i)$. Here $f(x)$ is a function from $\mathbb{R}$ to $\mathbb{R}$.
b. The set of all $\mathbf{p}$ such that the probability that $\mathbf{prob}(x \leq \alpha) \leq \beta$
c. The set of all $\mathbf{p}$ such that the expectation of $|x|^3$ is greater than a given constant α times the expectation of $|x| : E|x^3| \leq \alpha E|x|$
d. The set of all $\mathbf{p}$ such that the expectation of x^2 is less than a given constant α: $Ex^2 \leq \alpha$

Problem 2 Solution

a. $Ef(x) = \sum_{i=1}^{n} p_i f(a_i)$ can be defined as a convex set since the constraint is equivalent to two linear inequalities in the probabilities p_i.
b. The set of all $\mathbf{p}$ such that the probability that $\mathbf{prob}(x \leq \alpha) \leq \beta$ can be defined as a convex set because the constraint is equivalent to a linear inequality in the probabilities p_i.
c. The set of all $\mathbf{p}$ such that the expectation of $|x|^3$ is greater than a given constant α times the expectation of $|x|$: $E|x^3| \leq \alpha E|x|$ can be defined as a convex set because the constraint is equivalent to a linear inequality in the probabilities p_i.
d. The set of all $\mathbf{p}$ such that the expectation of x^2 is less than a given constant α: $Ex^2 \leq \alpha$ can be defined as a convex set because the constraint is equivalent to a linear inequality in the probabilities p_i.

Problem 3: Bounded Value Least Squares for Wine Mixing

We have seen several examples so far in the couse where we would like to have inequality constraints on the decision variable for our least squares problem, for example to prevent non-sensical solutions like spending a negative amount of money on advertising, limiting the total investment in certain types of assets, or perhaps bounding the value of a statistical coefficient to a certain range. Non-negative least squares is a type of least squares problem where the decision variables $\mathbf{x} \geq 0$, and Bounded-Value Least Squares allows for more general constraints.

This type of least-squares problem needs to be solved algorithmically, and we will use it to get our first practice using the `CVX` software package. You should install `CVXPY`, `CVXR`, or a flavor of `CVX` compatible with whatever software you are using to solve the problem and use `CVX` to solve this problem.

The problem is one of finding a mixture of wines which achieves certain chemical characteristics. I have attached a dataset which contains data on the chemical composition of 6 different wines (the dataset originates from kaggle but is reduced for our purposes). Each wine is described according to 11 chemical characteristics, including `alcohol`, `residual sugar`, `chlorides`, etc. I have also provided data for the chemical composition for a target wine.

- wine_data.csv
- target.csv

The goal of this problem is to find the blend of wines which has chemical characteristics closest to the target wine.

Concretely, you are solving for weights $\mathbf{p}$. The concentration of chemical i in wine j is given by the matrix C_{ij}, and the concentration in the blended wine is:

$$c_{blend,i} = \sum_{j=1}^{6} C_{ij} p_j,$$

so that the overall concentration vector in the blend satisfies:

$$\mathbf{c}_{blend} = C\mathbf{p}$$

The vector $\mathbf{p}$ is a discrete probability distribution, meaning that all entries are non-negative and must sum to 1 (you can't add negative wine). The range of each chemical varies greatly, so our objective function should incorporate a penalty that is weighted according to the magnitude of the value in the target function:

$$\min_{\mathbf{p}} \sum_{i=1}^{11} \left(\frac{c_i - c_{blend,_i}}{c_i} \right)^2$$

Implement this least squares optimization problem using CVX and determine the optimal blend of wines to match the target.

```
# load packages
import numpy as np
import pandas as pd
import cvxpy as cp

# extract data
wine_url = 'https://media.githubusercontent.com/media/georgehagstrom/DATA609Spring2025/refs/I
target_url = 'https://media.githubusercontent.com/media/georgehagstrom/DATA609Spring2025/refs

wine_data = pd.read_csv(wine_url)
```

```
target_data = pd.read_csv(target_url)

# extract the relevant data and define the decision variables
C = wine_data.values
target = target_data.values.flatten()
p = cp.Variable(6, nonneg=True)

# define the objective function
objective = cp.Minimize(cp.norm(np.diag(C) @ p - target))

# solve the problem
problem = cp.Problem(objective)
problem.solve()

# get the optimal weights for the wines
optimal_weights = p.value
print(f"The optimal wine blend weights using convex optimization are: {optimal_weights}")
```

```
The optimal wine blend weights using convex optimization are: [0.02638134 3.18943341 4.38723
```