Nonlinear Optimization (18799 B, PP)
IST-CMU PhD course
Spring 2013
Instructor: `jxavier@isr.ist.utl.pt`
TA: `augustos@andrew.cmu.edu`

Important: The homework is due February, 21.

Homework 1

Instructions: read sections 2.1, 2.2, and 2.3 of [1].

Problem A. *(Markov chains)* Consider a discrete-time (homogeneous) Markov chain $(X_t)_{t\geq 0}$, with state space $\mathcal{A} = \{\alpha_1, \alpha_2, \ldots, \alpha_n\} \subset \mathbb{R}$. Let $P = (P_{ij})$ be the $n \times n$ transition matrix. Thus, the (i, j)th entry of P is the probability of switching from α_i to α_j:

$$P_{ij} = \mathbb{P}(X_{t+1} = \alpha_j \mid X_t = \alpha_i), \quad t \geq 0.$$

Note that $P \geq 0$ and each of its rows sums to one. The matrix P is fixed throughout this problem.

Let the row vector π be a distribution for the initial state, i.e., $\pi_i = \mathbb{P}(X_0 = i)$, for $i = 1, \ldots, n$. Naturally, π is in the probability simplex $\Delta = \{x \in \mathbb{R}^n \,:\, x \geq 0, 1^\top x = 1\}$.

It is well known that π, jointly with P, determine the distribution of X_t (state of chain at time t) for all $t \geq 0$. More precisely, the distribution of X_t is given by the row vector πP^t.

We let $\mathbb{E}_\pi(\cdot)$ denote the expectation operator; note that we made explicit the dependence on the initial distribution. For example, the mean value of the random variable X_t, denoted by

$$\mu_t := \mathbb{E}_\pi(X_t),$$

is a function of π. In fact, it is easy to see that $\mu_t = \pi P^t \alpha$, where $\alpha := (\alpha_1, \ldots, \alpha_n)$. Similarly, the variance of X_t, denoted by

$$\mathrm{var}_\pi(X_t) = \mathbb{E}_\pi\left((X_t - \mu_t)^2\right),$$

is a function of π.

We are interested in the set

$$S = \{\pi \in \Delta \,:\, \mathrm{var}_\pi(X_t) \geq \epsilon, \text{ for all } t \geq 0\}, \tag{1}$$

for a given $\epsilon > 0$. In words, any initial distribution $\pi \in S$ imposes a minimum of "fluctuation" of the chain state, for all t. The main goal of this problem is to show that S is convex (assume it is non-empty).

We proceed by small steps.

(a) Let $c_1, c_2, d \in \mathbb{R}$ be given. Show that the set

$$B = \left\{(x_1, x_2) \in \mathbb{R}^2 \,:\, c_1 x_1 + c_2 x_2 - x_1^2 \geq d\right\}$$

is convex.
Hint: you may find useful the inequality $ab \leq (a^2 + b^2)/2$, for $a, b \in \mathbb{R}$.

(b) Assume that the set

$$C = \left\{x \in \mathbb{R}^n \,:\, c^\top x - x^\top a a^\top x \geq d\right\}$$

is convex for a given $a, c \in \mathbb{R}^n$ and $d \in \mathbb{R}$.
Let Q be an $n \times n$ orthogonal matrix, i.e., $QQ^\top = Q^\top Q = I_n$, where I_n is the identity matrix. Show that the set

$$D = \left\{x \in \mathbb{R}^n \,:\, c^\top Q^\top x - x^\top Q a a^\top Q^\top x \geq d\right\}$$

is convex.

(c) Let $c_1, c_2, a_1, a_2, d \in \mathbb{R}$ be given. Show that the set

$$E = \left\{ (x_1, x_2) \in \mathbb{R}^2 \,:\, c_1 x_1 + c_2 x_2 - (a_1 x_1 + a_2 x_2)^2 \geq d \right\}$$

is convex.

Hint: reduce to part (a), by using part (b).

(d) Let $c = (c_1, c_2, 0, \ldots, 0) \in \mathbb{R}^n$ and $a = (a_1, a_2, 0, \ldots, 0) \in \mathbb{R}^n$ be given. Show that the set

$$F = \left\{ x \in \mathbb{R}^n \,:\, c^\top x - x^\top a a^\top x \geq d \right\}$$

is convex.

Caution: the set F lies in $\mathbb{R}^n$, whereas E is in $\mathbb{R}^2$.

(e) Prove that the set S in (1) is convex.

Hint: use parts (b) and (d), and represent $\mathrm{var}_\pi\left(X_t\right)$ as a quadratic function of π.

Problem B. *(Martingales)* Let X and Y be discrete random variables taking values in the alphabet $\mathcal{A} = \{\alpha_1, \alpha_2, \ldots, \alpha_n\}$. Let the $n \times n$ matrix $P = (P_{ij})$ represent their joint probability mass distribution:

$$P_{ij} = \mathbb{P}\left(X = \alpha_i, Y = \alpha_j\right).$$

Note that P is a nonnegative matrix and *all* its entries sum to one, i.e., P belongs to the set

$$S = \left\{ X \in \mathbb{R}^{n \times n} \,:\, X \geq 0, 1^\top X 1 = 1 \right\}. \tag{2}$$

Usual probability constructs are a function of P. For example, the conditional probability

$$\mathbb{P}(Y = \alpha_j \,|\, X = \alpha_i) = \frac{\mathbb{P}(Y = \alpha_j, X = \alpha_i)}{\mathbb{P}(X = \alpha_i)} = \frac{P_{ij}}{\sum_{k=1}^n P_{ik}}$$

is a function of P. Similary, the conditional expectation

$$\mathbb{E}\left(Y \,|\, X = \alpha_i\right) = \sum_{j=1}^{n} \alpha_j \mathbb{P}(Y = \alpha_j \,|\, X = \alpha_i)$$

is a function of P. (Contrary to problem A, the dependence on P is now hidden in the notation: we could have written $\mathbb{E}_P\left(Y \,|\, X = \alpha_j\right) \ldots$)

(a) Show that the set S is convex.

(b) Show that the set

$$T = \{ P \in S \,:\, \mathbb{E}\left(Y \,|\, X = \alpha_i\right) = \alpha_i, \text{ for } i = 1, \ldots, n \}$$

is convex.

Note: an ordered pair of random variables (X, Y) with the property in T is said to be a martingale.

References

[1] S. Boyd and L. Vandenberghe. *Convex optimization.* Cambridge University Press, 2004.