Notes on Stochastic Programming

Chuwen Zhang[1] and Jingyuan Yang[1]

[1]Stochastic Programming Reading Group

February 22, 2022

1 Preface

This monograph records the reading notes for the subject "stochastic programming" starting from Spring 2022. The goal is to increase the authors' familiarity in this fascinating field.

2 Mathematical Background

The part is based on [Shapiro et al., 2014], [Birge and Louveaux, 2011] and so forth.

2.1 Basic Convex Analysis

Firstly, we introduce some definitions.

K_C	recession cone to set C
C^0	polar cone to set C
C^*	dual cone to set C
$\mathcal{N}_C$	normal cone to set C
$\mathcal{R}_C$	radial cone to set C
$\mathcal{T}_C$	tangent cone to set C

Table 1: A summary of convex sets

The Figure 1 give a clue to these special sets.

Understanding the definition of the cones is not an easy task, which requires materials in a set of books for convex analysis in different views.

Definition 2.1 (The Minkowski Summation). *If $\emptyset \neq A, B \subset X$, the Minkowski sum of A and B is $A + B := \{a + b \mid a \in A, b \in B\}$. Moreover,*

(i) if $x \in X, \lambda \in R$ and $\emptyset \neq \Gamma \subset R$, then $x + A := A + x := A + \{x\}, \Lambda \cdot A = \{\gamma a \mid \lambda \in \Lambda, a \in A\}$ and $\lambda A := \{\lambda\} \cdot A$.

(ii) We shall consider that $A + \emptyset = \emptyset$ and $\lambda \cdot \emptyset = \emptyset \cdot A = \emptyset$.

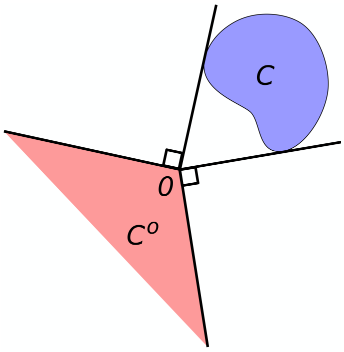
Definition 2.2 (The spaces in general vector spaces). *In a vector space X, the linear (subspace), affine space, cone, and convex set are, subspace of X that is closed under $(\cdot)$ combinations.*

Definition 2.3 (The equivalent definitions of the $(\cdot)$ hulls). *The two definitions are equivalent.*

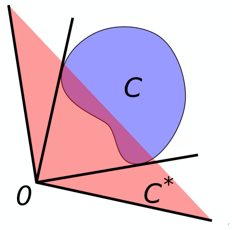
(i) The $(\cdot)$ hull of $A \subset X$ is the $(\cdot)$ combinations of $\{x\} \subseteq A$

(ii) The $(\cdot)$ hull of A can be represented as the following,

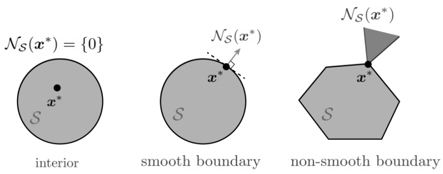
$$(\cdot)(A) = \bigcap_{V} \{V \subset X \mid A \subseteq V, V(\cdot)\} \tag{1}$$



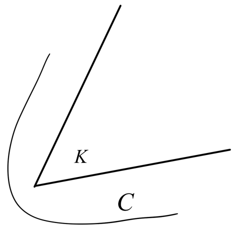
(a) The polar cone of C



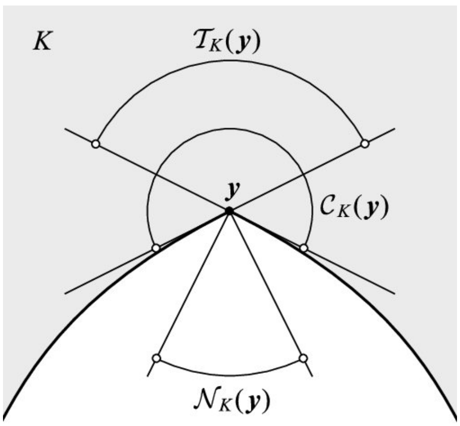
(b) The dual cone of C



(c) The normal cone of C



(d) The recession cone of C



(e) Another example

Figure 1: Illustration of the cones

Definition 2.4 (The basic convex sets). *(i) The recession cone K_C of the set C is formed by vectors h such that for any $x \in C$ and any $t > 0$ it follows that $x + t \cdot h \in C$.*

(ii) The polar cone,

$$C^0 = \{y : \langle y, x \rangle \leq 0 \quad \forall x \in C\} \tag{2}$$

is the opposite of dual cone.

(iii) The normal cone of C is,

$$\mathcal{N}_C(x) = \{y : \langle y, z - x\rangle \leq 0, \forall x \in C\} \tag{3}$$

(iv) The radial cone $\mathcal{R}_C(x_0) = \bigcup_{t>0}\{t(C - x_0)\}$.

(v) The tangent cone $\mathcal{T}_C(x_0)$ is called the tangent cone to C at x_0

Theorem 2.5 (The theorems for the recession cone)**.** *(i) The K_C is convex if C is convex, and is closed if the set C is closed.*

(ii) The convex set C is bounded iff. $K_C = \{0\}$

(iii) We have $C = K_C + C$ by the Minkowski sum.

(iv) If C is nonempty, closed, and convex, then

$$K_C = \bigcap_{t>0} t(C - a), \forall a \in C$$

Remark 2.6. *The case where C is a polyhedron. Consider polyhedron $C = \{x : Ax \leq b\}$, based on definition,*

- *The recession cone is $K_c = \{x : Ax \leq 0\}$. Compare to the null space of A, i.e., $K_c \supseteq \mathbf{Null}(A)$*

Theorem 2.7 (Farkas Lemma)**.** *Consider $Ax = b$, then either of the following in True:*

(i) There exists $x \geq 0$ such that $Ax = b$

(ii) There exists y such that $y^T A \geq 0, b^T y < 0$

Farkas Lemma is very important in the Linear programming and convex optimization.

Proof. (0) Method 0, Direct proof. (1) Method 1, using LP duality. But actually, LP duality is derived from Farkas Lemma. (2) Method 2, using Fourier–Motzkin elimination. (3) Method 3, use the separating hyperplane theorem of a convex set, see general Farkas Theorem □

Theorem 2.8 (General Farkas Theorem)**.** *Let $f, g_1, \ldots, g_m : R^n \to R$ be convex functions, $C \subseteq R^n$ a convex set, and let us assume that the Slater condition holds for $g_1, \ldots, g_m$; i.e., there exists an $\bar{x} \in \operatorname{rel\,int} C$ such that $g_j(\bar{x}) < 0, j = 1, \ldots,$ m. The following two statements are equivalent:*

(i) The system below is not solvable,

$$\begin{aligned} f(x) &< 0 \\ g_j(x) &\leq 0, \quad j = 1, \ldots, m \\ x &\in C \end{aligned}$$

(ii) There are $y_1, \ldots, y_m \geq 0$ such that

$$f(x) + \sum_{i=1}^{m} y_j g_j(x) \geq 0$$

for all $x \in C$.

Proof. To prove this, one consider the set $\{(u, \boldsymbol{v}) : f(x) < u, \boldsymbol{g}(x) \leq \boldsymbol{v}\}$, then show that $(0, \mathbf{0})$ does not in the set by the separating argument. □

Remark 2.9. *There are many ways to understand the Farkas lemma and the Farkas theorem. One way to see this is that the Farkas is to answer the question,*

$$g(x) \geq 0 \Rightarrow f(x) \geq 0?$$

If true, can we do this by linear aggregation?

3 Two-Stage Problems

Introduction

To understand the difficulties of multistage (integer) programs, we first look at the structural properties of the value functions.

Consider the two-stage problem,

$$\begin{aligned} \min_{x\in\mathcal{X}} \; & c^T x + \mathcal{Q}(x) \\ \text{s.t. } & Ax \leq b, x \in \mathcal{X} \end{aligned}$$

where the second stage recourse, the value function $\mathcal{Q}(x)$ is defined as the expectation with a recourse variable y.

$$\begin{aligned} \mathcal{Q}(x) &= \mathbb{E}Q(x, \boldsymbol{\xi}) \\ & \boldsymbol{\xi}(\omega) = (T, W, h)(\omega) \\ & T(\boldsymbol{\omega})x + W(\boldsymbol{\omega})y = h(\boldsymbol{\omega}), y \geq 0, \forall \boldsymbol{\omega} \\ Q(x, \boldsymbol{\xi}) &= \min_{y\in\mathcal{Y}} q^T y \end{aligned}$$

Remark 3.1.

- *Assume random variable $\boldsymbol{\omega}$ living on some probability space,* $(\Omega, \mathcal{F}, P)$
- *The first stage variable x is made before a realization of $\boldsymbol{\xi}$, i.e., the "here-and-now" ones.*
- *The second stage variable y, to be precise, should be $\boldsymbol{y}(\omega)$ that actually depends on $\boldsymbol{\omega}$*

Analysis on the value function

We first assume (x, y) are continuous, for example, $\mathcal{X}, \mathcal{Y}$ are convex. Define the function the support function $s_q(\chi)$ of set $\Pi(q)$, we notice,

$$\begin{aligned} \chi &= h - Tx \\ \Pi(q) &\doteq \left\{\pi : W^\top \pi \leq q\right\} \\ s_q(\chi) &\doteq \inf\left\{q^\top y : Wy = \chi, \quad y \geq 0\right\} \\ \Rightarrow s_q(\chi) &= \sup_{\pi\in\Pi(q)} \pi^\top \chi \end{aligned}$$

Obviously,

Theorem 1. *The value function $Q(x, \boldsymbol{\xi}) = s_q(\chi)$, furthermore*

1. *Q is a homogeneous polyhedron function supporting* $\Pi(q)$
2. *Also, the subdifferential of Q could also be defined.*

$$\partial Q\left(x_0, \boldsymbol{\xi}\right) = -T^\top \mathcal{D}\left(x_0, \boldsymbol{\xi}\right)$$

where

$$\mathcal{D}(x, \boldsymbol{\xi}) = \pi^* \doteq \arg\max_{\pi\in\Pi(q)} \pi^\top (h - Tx)$$

4 Analysis on Value Function

Now we inspect the existence of Q. Firstly, we introduce the definition,

Definition 2. Classification of recourse. The recourse problem is,

- i (Fixed) if the matrix $W(\boldsymbol{\omega}) = W$ is fixed (not random).
- ii (Complete) if the system $\{y : Wy = \chi, y \geq 0\}$ has a solution for every χ
- iii (Simple) if both complete and fixed
- iv (Relatively complete) relatively complete if for every feasible x, the feasible set of the second-stage problem is nonempty for almost everywhere (a.e.) $\boldsymbol{\omega} \in \Omega$.

For a recourse to be complete, $\Pi(q)$ must be bounded, i.e., the recession cone $\Pi(0) = \{0\}$.

References

[Birge and Louveaux, 2011] Birge, J. R. and Louveaux, F. (2011). *Introduction to Stochastic Programming.* Springer Series in Operations Research and Financial Engineering. Springer-Verlag New York, 2 edition.

[Shapiro et al., 2014] Shapiro, A., Dentcheva, D., and Ruszczyński, A. (2014). *Lectures on stochastic programming: modeling and theory.* SIAM.