UNIVERSIDAD DE CHILE
FACULTAD DE CIENCIAS FÍSICAS Y MATEMÁTICAS
DEPARTAMENTO DE INGENIERÍA MATEMÁTICA

APPLICATIONS OF STOCHASTIC AND BILEVEL OPTIMIZATION TO NETWORK PROBLEMS

TESIS PARA OPTAR AL GRADO DE DOCTOR EN CIENCIAS DE LA INGENIERÍA
MENCIÓN MODELACIÓN MATEMÁTICA

GIANFRANCO LIBERONA HENRÍQUEZ

PROFESOR GUÍA:
DAVID SEBASTIÁN SALAS VIDELA

PROFESOR CO-GUÍA:
RAFAEL CORREA FONTECILLA

MIEMBROS DE LA COMISIÓN:
DIDIER AUSSEL
SEBASTIÁN DONOSO FUENTES
LUDOVIC MONTASTRUC
EMILIO VILCHES GUTIÉRREZ

Este trabajo ha sido financiado por Beca Doctorado Nacional 2014, CMM ANID BASAL ACE210010 y CMM ANID BASAL FB210005

SANTIAGO DE CHILE
2022

Aplicaciones de Optimización Estocástica y Binivel a Problemas de Redes

Resumen

El presente trabajo corresponde a una Tesis de Doctorado, elaborada en el Departamento de Ingeniería Matemática de la Universidad de Chile, para obtener el grado de Doctor en Ciencias de la Ingeniería, Mención en Modelamiento Matemático.

Las siguientes páginas están compuestas por dos partes principales. En la primera de ellas se trabaja con optimización binivel, donde se define un indicador denominado *Valor Esperado de Información Compartida* para medir si es conveniente o no para el líder de un juego binivel *compartir* información con el seguidor del mismo. Este indicador es aplicado al modelamiento de la reubicación de conductores de una plataforma de *ride-hailing*, donde una formulación del problema es construida usando técnicas de optimización binivel. En este trabajo, se modela a la empresa duela de la plataforma de *ride-hailing* como el líder del problema, y al conjunto de conductores de la misma como un seguidor.

La segunda parte trabaja el *diseño óptimo* de ecoparques industriales, que consisten en comunidades de negocios de manufactura y servicios ubicados en una propiedad común, cuyos miembros buscan mejores desempeños a nivel medioambiental, económico y social a través de la colaboración en el manejo de temas medioambientales y de recursos. En este contexto, definimos un nuevo indicador, denominado *resiliencia* de un ecoparque industrial, que mide el número de escenarios *suficientemente buenos* que pueden ser obtenidos en la operación diaria del parque, cuando se considera incertidumbre en ella. En este trabajo, restricciones física y económicas para la operación diaria del ecoparque son consideradas, y la introducción del concepto de resiliencia es contrastado con herramientas clásicas en diseño, tales como eficiencia y costos de inversión.

En ambos trabajos, simulaciones numéricas son consideradas para comparar los resultados obtenidos. Por una parte, en el problema de *ride-hailing*, nuestras simulaciones sugieres que efectivamente es conveniente para el líder del problema (los dueños de la plataforma) compartir información con el seguidor (vale decir, el conjunto de conductores usando la aplicación). Por otra parte, en el diseño y operación de parques industriales, nuestro resultado más prometedor consiste en que, minimizando costos de inversión para el diseño, pero pidiendo un nivel mínimo de resiliencia deseado, podemos obtener importantes reducciones en inversión, reduciendo la performance del ecoparque en una cantidad muy pequeña.

Applications of Stochastic and Bilevel Optimization to Network Problems

Abstract

This work corresponds to a PhD Thesis elaborated at the Department of Mathematical Engineering of the University of Chile for obtaining the Degree of Doctor in Engineering Sciences, Mention in Mathematical Modeling.

The following pages are composed by two main parts. The first one deals with bilevel optimization, where an indicator called the *Expected Value of Shared Information* is defined in order to measure whether is convenient or not for the leader of a bilevel problem to *share* information with its follower. This indicator is then applied to the modeling of reallocation of drivers in *ride-hailing* platforms, where a formulation of the problem is constructed using bilevel optimization techniques. In this work, we model the ride-hailing enterprise as the leader of the problem, and the group of drivers as a follower.

The second part deals with the *optimal design* of eco-industrial parks, which consist in communities of manufacturing and service businesses located together on a common property, whose members seek enhanced environmental, economic, and social performance through collaboration in managing environmental and resource issues. In this context, we define a new indicator, called the *resilience* of an eco-industrial park, which measures the number of *good enough* scenarios that can be obtained on the eco-industrial park daily operation, when uncertainty is considered. In this work, physical and economical constraints for daily operation of an eco-industrial park are considered, and the introduction of the resilience is contrasted with classical tools in design, such as efficiency and investment costs.

In both works, numerical simulations are considered in order to compare the obtained results. On one hand, in the ride-hailing problem, our simulations suggest that it is indeed convenient for the leader of the problem (that is, the ride-hailing platform owner) to share information with the follower (the drivers using the application). On the other hand, in the eco-industrial park design and operation problem, our most promising result consists in that minimizing investment costs for the design, but asking for a desired level of resilience, we can obtain important investment deductions, reducing the overall performance of the eco-industrial park in a very small amount.

A Rosa, Stefano, Jorge, Mabel y Tilosa, mis pilares.

Acknowledgements

I would like to thank every person that, in one way or another, has contributed to make this accomplishment possible. First off, to my beloved parents Jorge and Mabel, and my godmather Tilosa, from who I learned very early in my childhood the importance and beauty of knowledge, and teaching it to others. I would also like to thank my brother for being the best partner-in-crime in my life, I love you and I am very proud of the person that you have become. Last, but of course not least, my wife Rosa has my eternal gratitude, thanks for believing in me, dealing with my frustrations, celebrating my accomplishments and sharing your life with me. All of my nuclear family are teachers, and I'm deeply grateful for that, because I have in all of you the better examples of what I want to do for the rest of my life.

I would also like to thank all of the people working at the Department of Mathematical Engineering (Facultad de Ciencias Físicas y Matemáticas, Universidad de Chile), for the opportunity and trust given in order to continue my studies after graduation, and after a couple of difficulties over the past years. In particular, to every person working there, including secretaries (I have to give a special shout-out for Silvia Mariano here), former and current directors along the way and the Academic staff. It would have been impossible to even begin (and of course, restart) this process if they were not there to support me in this long journey.

I am also indebted to ANID (National Agency for Research and Development, formerly known as CONICYT). It was thanks to this National Doctorate scholarship that it was possible for me to start this studies.

Let me now dedicate some words to the members of the Committee. I would like to thank all the members for their time and disposition to be a part of this process. In particular, I would like to thank Ludovic Montastruc for receiving me last February at Toulouse ENSIACET. His support and collaboration in the development of the ecoindustrial parks results present in this work were of utmost value. To Didier Aussel for being early available and enthused to form part of these Committe. Also, to Emilio and Sebastián, who gladly agreed with no hesitation to be part of the Committee.

To Rafael Correa, I am eternally in debt for the never-ending hours of teaching work as his T.A. at FCFM, but moreover, for giving me the greatest chance, being part of the creation and placement of the Universidad de O'Higgins in my native city, Rancagua. Working here, and making a real and huge impact on our country, has been the job of a lifetime, and I will remember every second of it forever.

I have also the obligation and honor to thank David Salas, who I have the luck of call my best friend, for allowing me to restart my PhD studies under his direction, for his constant support and advice during the various stages of this Thesis. I am very looking forward for continue working with you, and through these words, I congratulate you and Camila again for Violeta.

In my life, I have had the opportunity of working with tons of very valuable persons. I specially remember my time at CIAE, working for the ARPA team, traveling through the country and working with math and science teachers. I would like to thank in this space to Patricio, Cristián, Teresa, Constanza, Nicole, Marcela, Mayra and deeply to Carmen Luz, who where my colleagues for many years and from who I learned a lot. Very much of my teaching practices, have to do with the elements that I saw in all of you. Similarly, I would like to thank Felipe Célèry and Jaime Ortega, for giving me my first jobs as a T.A., and letting me know that this is my path in life. I learned a lot of things from both of you, and I hope to live up to the challenge of following your steps.

Finally, I would like to thank all my friends and colleagues at Universidad de O'Higgins. As I said before, this has been the job of a lifetime, and I will be forever grateful for having the chance of getting to know everyone here. Marcello, Javier, JM, Paloma, Nicolás, Karina, Fernando, Soledad, Sergio, Muriel, Cristóbal and everyone else (I know that I will not be able to write all these important names): it has been a real honor to work alongside all of you for the past 6 years, and getting to know each other. Specially, I would like to thank Carlos Pérez, as I see him as my "great-master". Carlos, I learned a lot of things, work-and-life speaking, and I hope I could help you on our first years here. Last, but not least, to my beloved team (former and actual members) at the "Dirección de Gestión Académica". Marta, Carla, Diego, Gonzalo, Monserrat, Felipe and Christian, you have had my back for all of this work (knowing it or not), and for that, I am forever in debt with you.

Table of Content

Introduction

This work corresponds to a PhD Thesis elaborated at the Department of Mathematical Engineering of the University of Chile for obtaining the Degree of Doctor in Engineering Sciences, Mention in Mathematical Modeling.

The following pages are composed by two main parts. The first one deals with bilevel optimization, where an indicator called the *Expected Value of Shared Information* is defined in order to measure whether is convenient or not for the leader of a bilevel problem to *share* information with its follower. This indicator is then applied to the modeling of reallocation of drivers in *ride-hailing* platforms. The second part deals with the optimal design of *eco-industrial parks*, where a new indicator, called the *resilience* of an ecopark is defined. In this second part, the physical and economical constraints for daily operation of an ecopark are considered, and the introduction of the resilience is contrasted with classical tools in design, such as efficiency and investment costs. In both works, numerical simulations are considered in order to compare the obtained results.

In this introduction, we summarize the main contributions of the thesis, which are detailed in Chapters 3 and 4. In Chapter 1, we enunciate the preliminary results on Optimization, Convex Analysis and Probability that will be needed in order to detail foundational elements of Bilevel and Two-Stage Programming in Chapter 2. If a reader wants to have a general overview of this work, reading this Introduction would suffice for such a goal. In contrast, for a full in-depth reading of this work, this Introduction can be skipped, as it content is largely revisited on the main Chapters 3 and 4.

Allocation Problems in Ride-Hailing Platforms

Ride-hailing consists on a form of transportation service delivered using platforms (such as Uber and Lyft), usually through smart-phone applications, where riders connect with drivers. Here, we are interested in the way information affects the relation between a ride-hailing company and its drivers.

The abstract model that will be used to describe the interaction between a ride-hailing company and its drivers is the following optimistic parametric bilevel programming problem

$$
\varphi(z_1, z_2) = \begin{cases} \min\limits_{x,y} & \theta(x, y, z_1, z_2) \\ \text{s.t.} & \begin{cases} x \in X \\ y \text{ solves } \begin{cases} \min\limits_{y} f(x, y, z_1, z_2) \\ \text{s.t. } y \in Y(x), \end{cases} \end{cases} \end{cases}
$$

where $(z_1, z_2) \in Z = Z_1 \times Z_2 \subset \mathbb{R}^{k_1} \times \mathbb{R}^{k_2}$ are the parameters. Here, the function $\varphi : Z \to \mathbb{R}$ is called the *value function* of this problem. For each pair $(z_1, z_2) \in Z$, the leader aims to minimize the loss function $\theta : \mathbb{R}^n \times \mathbb{R}^m \times \mathbb{R}^{k_1} \times \mathbb{R}^{k_2} \to \mathbb{R}$. He or she only controls the first variable $x \in X \subset \mathbb{R}^n$, which we call the leader's decision. The set of admissible leader's decisions $X \subset \mathbb{R}^n$ is fixed. In a similar way, for each pair $(z_1, z_2) \in Z$ and each leader's decision $x \in X$, the follower aims to minimize the loss function $f : \mathbb{R}^n \times \mathbb{R}^m \times \mathbb{R}^{k_1} \times \mathbb{R}^{k_2} \to \mathbb{R}$. He or she only controls the second variable $y \in Y(x) \subset \mathbb{R}^m$, which we call the follower's decision. The set of admissible decisions $Y(x) \subset \mathbb{R}^m$ depends on the leader's decision x, inducing a set-valued map $Y : X \rightrightarrows \mathbb{R}^m$.

Given some hypothesis for this problem, thoroughly described in Chapter 3, the existence of solutions is ensured. In this setting, uncertainty is formalized as random variables $\zeta = (\zeta_1, \zeta_2) : \Omega \to Z_1 \times Z_2$ and $\xi = (\xi_1, \xi_2) : \Omega \to Z_1 \times Z_2$, over a probability space $(\Omega, \Sigma, \mathbb{P})$. These variables determine the parameters (z_1, z_2).

For this type of problems, we introduce the *Shared Wait-and-See (SWS)*. This indicator, which is given by the formula

$$
SWS := \mathbb{E}_\zeta(\varphi) = \int_\Omega \varphi(\zeta_1(\omega), \zeta_2(\omega)) d\mathbb{P}(\omega),
$$

which captures the expected value for a leader that has perfect information and shares it with the follower. Comparing with the classic (see, e.g., [111, 21]) Wait-and-See (WS) indicator, we define the *Expected Value of Shared Information* (EVSI), that measures the value of sharing information in the context of Stackelberg games. This indicator is relevant in problems where both agents, the leader and the follower, must make their decisions prior to some uncertain event in a non-full cooperative nor-full competitive scheme, such as the ride-hailing problem.

In this approach, we consider the following model: a driver associated with a ride-hailing company that has not been matched with a passenger must decide whether to keep searching for a match around his or her current location, or to move to another one within the city. We can model the different locations as a finite set of zones, $I = \{1, \ldots, n\}$, connected as a directed graph.

If the driver is in the ith zone, his or her reallocation decision will depend on five factors: 1) the vector of marginal prices fixed by ride-hailing company, $p = (p_i \ : \ i \in I)$; 2) the vector of previously matched drivers who will arrive to each node (and will become available at that node), $y = (y_i \ : \ i \in I)$; 3) the vector of demands of each zone $d = (d_i \ : \ i \in I)$; 4) the marginal costs of moving to another zone, $\alpha_i = (\alpha_{ij} \ : \ j \in I)$; and 5) the vector of previously unmatched drivers that will be at each node, $x = (x_i \ : \ i \in I)$.

In this work, we model the situation where drivers can communicate between them outside the ride-hailing platform, and they can coordinate their allocation. Thus, we model all unmatched drivers as a single new follower, who aims to maximize the social welfare of all drivers. Then, for a given price vector p, the aggregated allocation problem is posed as follows:

$$F(p) := \begin{cases} \max_v & \sum_{i=1}^n cp_i \mathbb{E}_\xi \left[\min(x_i + y_i, d_i)\right] - \sum_{i\neq j} \alpha_{ij} v_{ij} \\ s.t & \begin{cases} v \geq 0 \\ \sum_{k\neq j} v_{jk} \leq x_{0,j}, \quad \forall j \in I \end{cases} \end{cases},$$

where $c \in (0,1)$ is the fraction of the ride price that the driver gets, and v_{ij} models the number of unmatched drivers that move from zone i to zone j. The first sum at the objective function represents the expected revenue of all drivers, while the second sum stands for the aggregated costs of reallocation.

The distribution of ξ, which models the belief over y and d_0, must reflect the fact that drivers have not access to the data of the ride-hailing company. On the one hand, we will model the vector of previously matched drivers who will arrive to each node as an uniformly distributed random variable. This distribution represents the lack of information for the unmatched drivers about the matched ones. On the other hand, we will assume that drivers perceive the distribution of the nominal demand as a discrete one, considering $m \in \mathbb{N}$ feasible scenarios, $\{\omega_1, \dots, \omega_m\}$. Thus, we can write

$$\begin{aligned} \mathbb{E}_\xi \left[p_i \min(x_i + y_i, d_i)\right] &= -\mathbb{E}_\xi(p_i \max(-x_i - y_i, -d_i)) \\ &= -\sum_{k=1}^m p_i \mathbb{E}_y[\max(-x_i - y_i, -d_{i,k})] \cdot \mathbb{P}(\omega_k), \end{aligned} \tag{1}$$

where we define the discrete expression $d_{i,k} = d_i(\omega_k)$. Therefore, the follower will deal with a discrete version of its original problem $F(p)$, given by

$$F_m(p) := \begin{cases} \min_v & c\sum_{i=1}^n \sum_{k=1}^m p_i \mathbb{E}_y[\max(-x_i - y_i, -d_{i,k})] \cdot \mathbb{P}(\omega_k) + \sum_{i\neq j} \alpha_{ij} v_{ij} \\ s.t & \begin{cases} -v \leq 0 \\ \sum_{j\neq i} v_{ij} - x_{0,i} \leq 0, \quad \forall i \in I \end{cases} \end{cases}. \tag{2}$$

Now, the ride-hailing company must decide the price vector p. The company does not necessarily know the exact value of the demand vector d, but it knows the vector y of occupied drivers. Since the company aims to maximize its revenues, it must solve the following bilevel programming problem:

$$L(y) := \begin{cases} \max_{p,v} & \sum_{i=1}^n (1-c)p_i \cdot \mathbb{E}_\zeta[\min(x_i + y_i, d_i)] \\ s.t & \begin{cases} p_i \in [p_{i,\min}, p_{i,\max}], \quad \forall i \in I \\ v \text{ solves } F_m(p). \end{cases} \end{cases} \tag{3}$$

Here ζ is the random variable that models the belief of the leader about the behavior of the nominal demand. The distribution of ζ should be multivariate normal-like distribution around a nominal value $\bar{d}_0 = (\bar{d}_{0,i} \ : \ i \in I)$. Observe that the objective function of the leader is partially cooperative with the follower, in the sense that both of them share the term of aggregated revenues of the fleet.

In order to compute the EVSI, MPCC reformulations to single bilinear optimizations are obtained. This technique is classic in Bilevel Programming (see, e.g., [32]). Our main results, in relation to this announced reformulations, are the following:

Theorem 1 *For any given value of the random vector $\zeta = (y, d_0)$, the Wait-and-See problem associated to the leader's problem (3.13) is equivalent (in the sense of global solutions) to its MPCC reformulation given by*

$$\max_{p,v,\lambda,\gamma} \sum_{i=1}^{n} (1-c)p_i \cdot \min(x_i + y_i, d_i)$$

$$s.t \begin{cases} p_i \in [p_{i,\min}, p_{i,\max}], \quad \forall i \in I \\ \sum_{j \neq i} v_{ij} - x_{0i} \leq 0, \quad \forall i \in I \\ \displaystyle\sum_{k=1}^{m} (p_i \beta_{i,k} - p_j \beta_{j,k}) + \alpha_{ij} - \lambda_{ij} + \gamma_i = 0, \quad \forall i \neq j \in I \\ \gamma_i (\sum_{j \neq i} v_{ij} - x_{0i}) = 0, \quad \forall i \in I \\ \lambda_{ij} v_{ij} = 0, \quad \forall i \neq j \in I \\ v \geq 0, \gamma, \lambda \geq 0, \end{cases}$$

where the coefficients $\{\beta_{i,k} \ : \ i \in I, k \in K\}$ are given by

$$\beta_{i,k} := \begin{cases} 0 & \text{if } d_{i,k} - x_i \leq 0 \\ c\mathbb{P}(\omega_k) \frac{x_i - d_{i,k}}{\bar{y}} & \text{if } 0 \leq d_{i,k} - x_i \leq \bar{y} \\ -c\mathbb{P}(\omega_k) & \text{if } d_{i,k} - x_i \geq \bar{y}. \end{cases}$$

Furthermore, the multipliers $\gamma = (\gamma_i \ : \ i \in I)$ and $\lambda = (\lambda_{ij} \ : \ i \neq j \in I)$ verify

$$0 \leq \gamma_i \leq 2mp_{\max} \quad \text{and} \quad 0 \leq \lambda_{ij} \leq 4mp_{\max},$$

where $p_{\max} = \max_{i \in I} \{p_{i,\max}\}$.

Theorem 2 *For any given value of the random vector $\zeta = (y, d_0)$, the Shared-Wait-and-See problem associated to the leader's problem (3.13) is equivalent (in the sense of global solutions) to its MPCC reformulation given by*

$$\max_{p,v,\lambda,\gamma,\beta} \sum_{i=1}^{n} (1-c)p_i \cdot \min(x_i + y_i, d_i)$$

$$s.t. \begin{cases} p_i \in [p_{i,\min}, p_{i,\max}], \quad \forall i \in I \\ \sum_{j \neq i} v_{ij} - x_{0i} \leq 0, \quad \forall i \in I \\ c(\beta_i - \beta_j) + \alpha_{ij} - \lambda_{ij} + \gamma_i = 0, \quad \forall i \neq j \\ \gamma_i (\sum_{j \neq i} v_{ij} - x_{0i}) = 0, \quad \forall i \in I \\ \lambda_{ij} v_{ij} = 0, \quad \forall i \neq j \in I \\ v \geq 0, \gamma, \lambda \geq 0, \end{cases}$$

where the variables $\{\beta_i \ : \ i \in I\}$ verify that

$$\beta_i \in \begin{cases} \{0\} & \text{if } d_i - x_i < y_i \\ \{p_i\} & \text{if } d_i - x_i > y_i \\ [0, p_i] & \text{if } d_i - x_i = y_i \end{cases}.$$

Furthermore, the multipliers $\gamma = (\gamma_i \ : \ i \in I)$ *and* $\lambda = (\lambda_{ij} \ : \ i \neq j \in I)$ *verify that*

$$0 \leq \gamma_i \leq 2p_{\max} \quad \textit{and} \quad 0 \leq \lambda_{ij} \leq 4p_{\max},$$

where $p_{\max} = \max_{i \in I}\{p_{i,\max}\}$.

Using both theorems, we are able to compute the EVSI for randomly generated data, that supports the idea that sharing information with the drivers, might be beneficial for the ride-hailing platform.

Resilient Design of Eco-Industrial Parks

An Eco-Industrial Park (EIP), as defined in [76], consists in a community of manufacturing and service businesses located together on a common property. Member businesses seek enhanced environmental, economic, and social performance through collaboration in managing environmental and resource issues. In this work, as it is usual in the literature, we model the EIP community as a central authority in charge of the design of the park at a first stage, and of optimizing the interactions within the members in the daily operation during its lifetime.

A canonical example of EIP corresponds to the modeling of water exchange networks (see, e.g., [23, 102, 108] and the references therein), where each participant of the EIP needs to consume fresh water for its industrial processes, and to send away partially contaminated water. In parallel, there is a central authority of the EIP, which is in charge of design the park and operate it afterwards, following some criteria that reflects environmental, economic and/or social benefits.

This work falls into the context of Two-Stage problems, which are optimization problems where a single-agent must take decisions before some random events occur, and other decisions (also called recourse actions) afterwards. In this sense, Recourse programs can be seen as a particular case of Stochastic Bilevel programming, where the same agent is solving the upper-level and lower-level problem.

A general two-stage problem can be stated as

$$\min_{x \in X} \ f_1(x) + \mathbb{E}[Q(x,\xi)],$$

where X is a nonempty (usually compact) subset of $\mathbb{R}^n$, $f_1 : \mathbb{R}^n \to \mathbb{R}$ is a real-valued (usually continuous) function, and $Q(x,\xi)$ is obtained as the optimal value of the so called *second-stage* problem

$$\begin{cases} \min\limits_{y} & f_2(x,y,\xi) \\ \text{s.t.} & y \in Y(x,\xi). \end{cases}$$

Here, $f_2 : \mathbb{R}^n \times \mathbb{R}^m \times \mathbb{R}^p \to \mathbb{R}$ and $Y : \mathbb{R}^n \times \mathbb{R}^d \rightrightarrows \mathbb{R}^m$ is a set-valued function. This representation shows the sequence of events, where the first-stage decisions x are taken with uncertainty about future realizations of ξ, and after that, some corrective actions y can be taken.

In order to design an EIP, an important uncertain quantity must be considered: the mass load production of contaminant for every participant, $\xi = (M_1, \ldots, M_n)$ is uncertain at the design stage, because each process has daily unpredictable variations. These variations are only revealed during the daily operation of the EIP, and of course, they can be different every time. The operation variables of the park are given by: 1) the fresh water consumption of each agent, given by the vector $z = (z_i \ : \ i \in I)$; 2) the exchange water matrix $F = (F_{ij} \ : \ i \neq j \in I)$; and 3) the discharge of each agent to the sink node, given by the vector $O = (O_i \ : \ i \in I)$.

A valid operation is then given by values of (z, F, O) satisfying the following operation constraints:

1. **Water Mass Balance:** for every participant $i \in I$, the total inlet flux must coincide with the total outlet flux.

$$z_i + \sum_{k \neq i} F_{ki} = \sum_{j \neq i} F_{ij} + O_i.$$

   At the sink node, there is no balance constraint.

2. **Contaminant Mass Balance:** For every participant $i \in I$, the total inlet contaminant mass must coincide with the total outlet contaminant mass, that is,

$$M_i + \sum_{k \neq i} C_{k,\text{out}} F_{ki} = C_{i,\text{out}} \left(O_i + \sum_{j \neq i} F_{ij} \right).$$

   The mass is computed from the fluxes F thanks to the optimality assumption that the outlet concentration is always attained.

3. **Inlet/Outlet Concentration Constraints:** for every participant $i \in I$, we have that

$$\sum_{k \neq i} C_{k,\text{out}} F_{ki} \leq C_{i,\text{in}} \left(z_i + \sum_{k \neq i} F_{ki} \right).$$

   The above inequality is the inlet concentration constraint expressed in terms of contaminant mass.

4. **Positivity of Fluxes:** all the fluxes in the EIP must be non-negative, that is,

$$F_{ij} \geq 0, \quad \forall i \neq j \in I \qquad \text{and} \qquad z_i, O_i \geq 0, \quad \forall i \in I.$$

5. **Boundedness of exchanges:** all the fluxes in the EIP must be within the capacities given by the vector L, that is,

$$F_{ij} \leq L_{ij}, \quad \forall i \neq j \in I.$$

The central authority's goal at the daily operation is to minimize the global fresh water consumption. Nevertheless, this is not necessarily aligned with the individual participants' interests, which is to minimize their operational costs. Hence, jointly with the already detailed physical constraints, we must add economical ones for the daily operation of every participant, considering the principle of *individual rationality*: any enterprise will take part of the EIP only if this participation is economically convenient (see, e.g., [61]). Therefore, we will add a constraint for the model in order to tackle the *individual rationality* of every participant, which ensures that the operational costs for every agent are less than the *stand-alone* operation, that is,

$$\mathrm{Cost}_i(F, z) \leq \mathrm{SA}_i(M_i).$$

An option to solve this two-stage problem, is to obtain or define first an analytical expression for $Q(L, \xi)$. However, this is not always possible. Instead, we will solve this problem using the implicit expression of $Q(L, \xi)$ as the optimal value of the second-stage problem. This general approach is known as *stochastic optimization with recourse* [21, 111, 75].

On a first approach, is natural to consider the averaged fresh water consumption costs of the EIP as the objective function to minimize. This correspond to expected value $A\mathbb{E}[Q(L, \xi)]$, where A is the lifetime factor, and allows us to control some kind of "average day" on its long-term operation. However, this does not necessarily give us an optimal EIP considering other indicators, such as robustness in face of uncertainty.

In this context, we define in Chapter 4 the *resilience* of an EIP, which measures the number of *good* scenarios from an economic point of view, for the EIP daily operation. In order to quantify this goal, we introduce here what we call the $(1-\alpha)$-level of *goodness* for an EIP as

$$\mathrm{G}_\alpha(L, \xi) \doteq Q(L, \xi) - \alpha \mathrm{SA}(\xi),$$

where $\mathrm{SA}(\xi) = \sum_{i \in I} \mathrm{SA}_i(M_i)$ is the total net cost, if all the agents worked on *stand-alone* operation. If $\mathrm{G}_\alpha \leq 0$, it means that the EIP operation is better than the stand-alone one. Hence, we define the *resilience* of the EIP as

$$\mathrm{Res}_\alpha(L) = \mathbb{P}[\mathrm{G}_\alpha(L, \xi) \leq 0]$$

We consider this functional as a part of the objective function in the design stage, or as a *chance constraint* at the same stage. Also, investment constraints are considered, then, the problem of optimal design can be formulated as follows:

$$\mathcal{P} = \begin{cases} \min_L & w_1\langle c, L\rangle + w_2\mathbb{E}[Q(L,\xi)] - w_3\mathrm{Res}_\alpha(L) \\ \text{s.t.} & L \geq 0 \\ & c_1(\langle c, L\rangle - B) \leq 0 \\ & c_2\left(\mathrm{Res}_\alpha(L) - \beta\right) \geq 0 \end{cases},$$

where $c_1, c_2 \in \{0, 1\}$, allowing us to consider or dismiss budget and resilience constraints. Similarly, w_1, w_2, w_3 are non-negative weights, indicating the priority of each element of the objective function.

In order to compute some calculations and comparisons between these two configurations, we consider the *Sample Average Approximation* reformulation of the Two-Stage problem, and numerical results are obtained. The SAA formulation of $\mathcal{P}$ is given by

$$\widehat{\mathcal{P}}_0 = \begin{cases} \min\limits_{L} & w_1\langle c, L\rangle + w_2\hat{q}_N(L) - w_3\widehat{\mathrm{Res}}_\alpha(L) \\ \text{s.t.} & L \geq 0 \\ & c_1(\langle c, L\rangle - B) \leq 0 \\ & c_2(\widehat{\mathrm{Res}}_\alpha(L) - \beta) \geq 0 \end{cases},$$

where $\hat{q}_N$ and $\widehat{\mathrm{Res}}_\alpha$ are the empirical values of $\mathbb{E}[Q(L,\xi)]$ and $\mathrm{Res}_\alpha(L)$, respectively, for a given sample $\{\hat{\xi}_1, \ldots, \hat{\xi}_N\}$. By introducing auxiliary variables (z, F, y), we can write the following MIP formulation

$$\widehat{\mathcal{P}}_1 = \begin{cases} \min\limits_{L,z,F,y} & w_1\langle c, L\rangle + \dfrac{w_2}{N}\sum\limits_{m=1}^{N}\sum\limits_{i=1}^{n} z_i^m - \dfrac{w_3}{N}\sum\limits_{m=1}^{N} y_m \\ \text{s.t.} & L \geq 0 \\ & c_1(\langle c, L\rangle - B) \leq 0 \\ & c_2\left(\dfrac{1}{N}\sum\limits_{m=1}^{N} y_m - \beta\right) \geq 0 \\ & (z^m, F^m) \in X(L, \hat{\xi}^m), \ \forall m \in [N] \\ & \sum_{i=1}^{n} z_i^m - \alpha SA(\hat{\xi}^m) \leq \mathrm{SA}(\hat{\xi}^m)(1 - y_m), \quad \forall m \in [N] \\ & y \in \{0,1\}^N, \end{cases}$$

where $X(L, \hat{\xi}^m)$ stands for the feasible set of the second-stage problem, for a given L and $\hat{\xi}^m$. Our main result in this section states the following.

Theorem 3 *If L^* is an optimal solution of Problem $\widehat{\mathcal{P}}_0$, then there exists (z^*, F^*, y^*) such that (L^*, z^*, F^*, y^*) is an optimal solution of Problem $\widehat{\mathcal{P}}_1$. Conversely, if (L^*, z^*, F^*, y^*) is an optimal solution of Problem $\widehat{\mathcal{P}}_1$, then L^* is an optimal solution for Problem $\widehat{\mathcal{P}}_0$. In both cases, one has that*

$$w_1\langle c, L^*\rangle + w_2\hat{q}_N(L^*) - w_3\widehat{Res}_\alpha(L^*) = w_1\langle c, L^*\rangle + \frac{w_2}{N}\sum_{m=1}^{N} z^{*m} - \frac{w_3}{N}\sum_{m=1}^{N} y_m^*.$$

Chapter 1

Preliminaries on Variational Analysis and Probablity

In this chapter, we will briefly describe the preliminary ideas and concepts on Optimization and Probability that will be widely used in the development of this work.

1.1 Optimization and Convex Analysis

Let us consider the optimization problem

$$\begin{cases} \min\limits_{x \in \mathbb{R}^n} & f(x) \\ \text{s.t.} & g_i(x) \leq 0, \ i \in I \\ & h_j(x) = 0, \ j \in J \end{cases} \tag{1.1}$$

and define $X = \{x \in \mathbb{R}^n : \ g_i(x) \leq 0, \ h_j(x) = 0, \ \forall i \in I, \ j \in J\}$.

In this section, we aim to recall what are the first-order optimality conditions of (1.1), by means of the well-known *Karush-Kuhn-Tucker* (KKT) equations. First, we recall a basic first-order result for problems like (1.1).

Theorem 1.1 ([106, Theorem 6.12]) *Consider the problem*

$$\min_{x \in X} f(x),$$

where f is a differentiable function. If x^* is a local optimal point, then

$$0 \in \nabla f(x^*) + N_X(x^*),$$

where $N_X(x^*) = \{v \in \mathbb{R}^n \ : \ \langle v, x - x^* \rangle \leq 0, \ \forall x \in X\}$ is the normal cone to X at x^*.

Remark When X is convex, the necessary condition from the last theorem can be written as

$$\langle \nabla f(x^*), x - x^* \rangle \geq 0, \ \forall x \in X.$$

This means that the linearization of f, given by $l(x) \doteq f(x^*) + \langle \nabla f(x^*), x - x^* \rangle$, attains its minimum over X at x^*. If f is also convex, then this condition becomes sufficient for global optimality for x^*.

Definition 1.2 (Active Inequality Constraints) *If $x \in \mathbb{R}^n$ is a feasible point of* (1.1)*, we define the set*

$$I(x) \doteq \{i \in I \ : \ g_i(x) = 0\}$$

as the set of active constraints *at x.*

Definition 1.3 (Lagrangian Function) *We call the* Lagrangian *of Problem* (1.1) *to the function*

$$\mathcal{L}(x, \mu, \lambda) = f(x) + \sum_{i \in I} \mu_i g_i(x) + \sum_{j \in J} \lambda_j h_j(x).$$

The vectors μ, λ are called the Lagrangian multipliers*.*

Using these concepts, we can now define the Karush-Kuhn-Tucker conditions as follow.

Definition 1.4 (KKT Conditions) *Let us consider problem* (1.1)*, with continuously differentiable functions f and g_i, for every $i \in I$. The conditions*

$$\begin{aligned} \nabla_x \mathcal{L}(x, \mu, \lambda) &= 0 \\ \mu \geq 0, \ \mu^T g(x) &= 0 \end{aligned} \tag{1.2}$$

are called the KKT conditions*, and (x^*, μ^*, λ^*) that satisfies them, is called a* KKT *point.*

We will now enunciate the famous Karush-Kuhn-Tucker Theorem, which states that, under some *Constraint Qualifications*, we can relate the KKT conditions to the normal cone $N_X(x)$, and therefore, to the the first-order optimality conditions given in Theorem 1.1.

Definition 1.5 (Constraint Qualifications) *Consider Problem* (1.1) *and $\tilde{x} \in \mathbb{R}^n$ a feasible point.*

1. *We say that Problem* (1.1) *satisfies the* Linear Independence Constraint Qualification (LICQ) *at $\tilde{x}$ if the set*

$$\{\nabla g_i(\tilde{x}) \ : i \in I(\tilde{x})\} \cup \{\nabla h_j(\tilde{x}) \ : \ j \in J\}$$

   is linearly independent.

2. *We say that Problem* (1.1) *verifies the* Mangarasarian-Fromovitz Constraint Qualification (MFCQ) *at $\tilde{x}$, if the set*

$$\{\nabla h_j(\tilde{x}) \ : \ j \in J\}$$

   is linearly independent, and there exists $d \in \mathbb{R}^n$ such that $\nabla g_i(\tilde{x})^T d < 0$ for all active constraints $i \in I(x)$, and $\nabla h_j(\tilde{x})^T d = 0$ for all $j \in J$.

Similarly, if f and g_i are convex functions for every $i \in I$, we say that Problem (1.1) *satisfies the* Slater's Constraint Qualification *if there exists $\tilde{x} \in \mathbb{R}^n$ such that $g_i(\tilde{x}) < 0, \forall i \in I$ and $h_j(\tilde{x}) = 0$ for all $j \in J$.*

Theorem 1.6 (Karush-Kuhn-Tucker Theorem) *Let $x^* \in X$. If any of the Contraint Qualifications detailed in Definition* (1.5) *is verified, the normal cone $N_X(x^*)$ can be written as*

$$N_X(x^*) = \left\{ \sum_{i \in I} \mu_i \nabla g_i(x^*) + \sum_{j \in J} \lambda_j \nabla h_j(x^*) \ : \ \mu \geq 0, \ \mu^T g(x^*) = 0 \right\}$$

In particular, if x^* is a local minimizer of Problem (1.1), *then there exists Lagrangian multipliers (μ^*, λ^*) such that (x^*, μ^*, λ^*) is a KKT point.*

1.1.1 Set-valued analysis

In this section, we aim to detail the tools that we need, in order to tackle *set-valued* functions (from now on, *multifunctions*) and its main properties.

Let X and Y be two non-empty sets, and let us denote by $\mathcal{P}(Y)$ the power set of Y. A *multifunction* (also known as *correspondence*) F is a function $F : X \to \mathcal{P}(Y)$, that is, a function which, for every $x \in X$, assigns $F(x) \subseteq Y$. We denote such a multifunction

$$F : X \rightrightarrows Y.$$

We now define two associated ideas of continuity for a multifunction, given that X and Y are treated as metric spaces. From now on, we say that $F : X \rightrightarrows Y$ is *closed-valued* if $F(x) \subseteq Y$ is closed, for every $x \in X$.

Definition 1.7 (Upper Semicontinuity) *A multifunction F is called* upper semicontinuous *at $x_0 \in X$ if, for each neighborhood G of $F(x_0)$ in Y, there exists a neiborhood U of x_0 in X such that*

$$F(x) \subset G, \quad \forall x \in U.$$

If F is upper semicontinuous at every $x \in X$, we simply say that F is upper semicontinuous.

In general, checking upper semicontinuity for a multifunction is a hard task. Therefore, it is relevant to have some characterizations of these concept, easier to work with. We will establish a relation between *closedness* of F and its upper semicontinuity.

Definition 1.8 (Closed Multifunction) *We say that $F : X \rightrightarrows Y$ is* closed *if its* graph

$$\mathrm{gph}(F) = \{(x, y) \ : \ x \in X, \ y \in F(x)\}$$

is closed in the product space $X \times Y$.

The following result, displays a strong relationship between the definitions listed before.

Theorem 1.9 *Let X, Y be metric spaces, Y being compact. Then, $F : X \rightrightarrows Y$ is upper semicontinuous and closed-valued in X if, and only if, F is closed.*

Therefore, for $x \in X$, $F(x)$ is closed and F is upper semicontinuous in x if, and only if, for every $(x_k)_{k\in\mathbb{N}} \subseteq X$ such that $x_k \to x$, $(y_k)_{k\in\mathbb{N}} \subseteq Y$ such that $y_k \to x_0$ and $y_k \in F(x_k)$ for every $k \in \mathbb{N}$, then $y_0 \in F(x_0)$.

Of course, the idea of *lower semicontinuity* is also important to consider, as we will define afterwards the notion of *continuous* multifunctions.

Definition 1.10 (Lower Semicontinuity) *A multifunction F is called* lower semicontinuous at $x_0 \in X$ *if, for each open set $G \subset Y$ for which*

$$F(x_0) \cap G \neq \varnothing,$$

there exists a neighborhood U of x_0 such that

$$F(x) \cap G \neq \varnothing, \quad \forall x \in U.$$

If F is lower semicontinuous at every $x \in X$, we simply say that F is lower semicontinuous.

Similarly to closedness we can establish a sequential characterization of lower semicontinuity. More precisely, F is lower semicontinuous at $x_0 \in X$ if, and only if, for every sequence $(x_k)_{k\in\mathbb{N}}$ such that $x_k \to x_0$ and for every point $y \in F(x_0)$, there exists a sequence $(y_k)_{k\in\mathbb{N}}$ such that $y_k \to y$ and $y_k \in F(x_k)$, $\forall k \geq k_0 \in \mathbb{N}$. We say that F is *continuous* if it is upper and lower semicontinuous simultaneously. With this notion of continuity, it is possible to establish the following important theorem of stability of optimization problems, known as the *Maximum Principle.*

Theorem 1.11 (Maximum Principle [60, Theorem 2.3.1]) *Let X, Y be metric spaces, $f : X \times Y \to \mathbb{R}$ a continuous function, and $K : X \rightrightarrows Y$ a non-empty-valued, compact-valued and continuous multifunction in X. Then,*

1. *The function $\varphi : X \to \mathbb{R}$ given by*

$$x \mapsto \varphi(x) = \max\{f(x,y) \ : \ y \in K(x)\}$$

   is continuous in X.

2. *The multifunction $\Phi : X \rightrightarrows Y$ given by*

$$x \mapsto \Phi(x) = \{y \in K(x) \ : \ f(x,y) = \varphi(x)\}$$

   is upper semicontinuous in X.

Another version of the maximum principle can be found in [5], where the continuity of the marginal function φ is decoupled in two semicontinuity results. Here we state one of those results, which will be useful in the sequel.

Theorem 1.12 ([5, Theorem 1.4.16]) *Let X, Y be metric spaces, $f : X \times Y \to \mathbb{R}$ an upper semicontinuous function, and $K : X \rightrightarrows Y$ a non-empty-valued, compact-valued and upper semicontinuous multifunction in X. Then, the marginal function $\varphi : X \to Y$ given by*

$$\varphi(x) = \sup_{y \in K(x)} f(x, y)$$

is upper semicontinuous.

1.1.2 Convex Subdifferential

When f is not necessarily differentiable, we can resort to what is called *convex subdifferential analysis*. For this section, let $f : \mathbb{R}^n \to \mathbb{R}$ be a convex function (which is known to be continuous, see e.g. [105]).

Definition 1.13 (Subdifferential) *The* subdifferential *of f at $x \in \mathbb{R}^n$, is the set*

$$\partial f(x) \doteq \{v \in \mathbb{R}^n \;:\; f(y) \geq f(x) + \langle v, y - x \rangle, \;\forall y \in \mathbb{R}^n\}.$$

A vector $v \in \partial f(x)$ is called a subgradient *of f at x.*

Several *calculus rules* for the subdifferential have been developed in the literature. Specifically, we recall the following sum rule:

Theorem 1.14 ([8, Corollary 16.48]) *If $f, g : \mathbb{R}^n \to \mathbb{R}$ are convex functions. Then,*

$$\partial(f + g)(x) = \partial f(x) + \partial g(x).$$

When we do not have differentiability for the objective function f in a minimization problem, we still have the following result, similar to Theorem (1.1).

Theorem 1.15 (Optimality relative to a Set, [106, Theorem 8.15]) *Let us consider the problem*

$$\min_{x \in X} f(x),$$

where f is a convex function and X is a convex set. Then, x^* is an optimal point if, and only if,

$$0 \in \partial f(x^*) + N_X(x^*),$$

where $N_X(x^*)$ is the normal cone (1.1)*.*

In particular, if x^* verifies any of the Constraint Qualifications from Definition (1.5)*, then there exists a Lagrange multiplier λ^* such that*

$$\begin{aligned} \partial_x \mathcal{L}(x^*, \lambda^*) = \partial f(x^*) + \textstyle\sum_{i \in I} \mu_i^* \nabla g_i(x) + \sum_{j \in J} \lambda_j \nabla h_j(x) &\ni 0 \\ \lambda^* \geq 0, \;\; (\lambda^*)^T g(x) &= 0 \end{aligned} \tag{1.3}$$

1.2 Probability

1.2.1 Random Variables and Functions

Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space, and $\mathcal{B}(\mathbb{R}^\ell)$ be the Borel sigma algebra of $\mathbb{R}^\ell$. Recall that a mapping $V : \Omega \to \mathbb{R}^\ell$ is said to be *measurable* if for any Borel set $B \in \mathcal{B}(\mathbb{R}^\ell)$, its inverse image

$$V^{-1}(B) = \{w \in \Omega \ : \ V(w) \in B\}$$

is $\mathcal{F}$-measurable.

Definition 1.16 (Random Vectors and Variables) *A mapping $V(w)$ from the probability space $(\Omega, \mathcal{F}, \mathbb{P})$ into $\mathbb{R}^\ell$ is called a* random vector*. If $\ell = 1$, we call V a* random variable*.*

Since, in general, in this work we will deal with random variables which are given as optimal values of optimization problems, we need to consider random variables that can take $\pm\infty$ values. We will call the measurable functions $Z : \Omega \to \overline{\mathbb{R}}$ as *extended* random variables.

Definition 1.17 (Expected Value) *The* expected value *or* expectation *of an extended random variable $Z : \Omega \to \overline{\mathbb{R}}$ is defined by the integral*

$$\mathbb{E}[Z] \doteq \int_\Omega Z(w)\ d\mathbb{P}(w).$$

If Z can only take a finite or at most countable number of different values, let us say $\{z_i\}_{i\in\mathbb{N}}$, it is said to be *discrete*. In that case,

$$\mathbb{E}[Z] = \sum_{i\in\mathbb{N}} z_i \cdot \mathbb{P}[Z = z_i]$$

Definition 1.18 (Random Function) *We say that $F : \mathbb{R}^n \times \Omega \to \overline{\mathbb{R}}$ is a* random function *if, for every fixed $x \in \mathbb{R}^n$, $F(x, \cdot)$ is $\mathcal{F}$-measurable.*

For a random function $F(x, w)$, we can define its corresponding *expected value function*

$$f(x) \doteq \mathbb{E}[F(x, w)] = \int_\Omega F(x, w)\ d\mathbb{P}(w). \tag{1.4}$$

We can also see $F(\cdot, w)$ as an extended real valued function, for every $w \in \Omega$.

1.2.2 The Law of Large Numbers and Sample Average Functions

Let us consider a sequence $\{\xi^j\}_{j\in\mathbb{N}}$ of $\xi^j = \xi^j(w)$ random vectors, defined on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$. In this section, we will treat ξ^j as an *element* of its support $\Xi \subseteq \mathbb{R}^\ell$, equipped with its Borel sigma algebra, $\mathcal{B}(\mathbb{R}^\ell)$.

Definition 1.19 (Identical Distribution) *The random vectors* $\{\xi^j\}_{j\in\mathbb{N}}$ *are* identically distributed, *if each* ξ^j *has the same probability distribution on* $(\Xi, \mathcal{B}(\mathbb{R}^\ell))$. *Furthermore, if all* ξ^j *are independent, we will say that they are* independent identically distributed, *or from now on,* iid.

Remark If $\{\xi^j\}_{j\in\mathbb{N}}$ are identically distributed and $F : \Xi \to \mathbb{R}$ is a measurable function, then $\{F(\xi^j)\}_{j\in\mathbb{N}} \subseteq \mathbb{R}^\ell$ are also identically distributed. Hence, the value of $\mathbb{E}[F(\xi^j)]$ is the same for every $j \in \mathbb{N}$.

Theorem 1.20 (Classical Law of Large Numbers (LLN)) *If* $\{\xi^j\}_{j\in\mathbb{N}}$ *are identically distributed, then*

$$\frac{1}{N}\sum_{j=1}^{N} F(\xi^j) \to \mathbb{E}[F(\xi^1)],$$

where the convergence is with probability 1 *as* $N \to \infty$. *That is,*

$$\mathbb{P}\left[\lim_{N\to\infty} \frac{1}{N}\sum_{j=1}^{N} F(\xi^j) = \mathbb{E}[F(\xi^1)]\right] = 1.$$

Let us consider a random function $F : X \times \Xi \to \mathbb{R}$, where $X \subseteq \mathbb{R}^n$ in non-empty and $\xi = \xi(w)$ is a random vector supported on Ξ. Let us suppose that

$$f(x) = \mathbb{E}[F(x, \xi)]$$

is well defined and finite-valued for every $x \in X$. Let $\{\xi^j\}_{j\in\mathbb{N}}$ be an *iid* sequence of random vectors having the same distribution as ξ.

Definition 1.21 (Sample Average Function) *For every* $N \in \mathbb{N}$, *we define the* sample average functions *of* f, *given by*

$$\hat{f}_N(x) \doteq \frac{1}{N}\sum_{j=1}^{N} F(x, \xi^j). \tag{1.5}$$

Observe that, for every fixed $x \in X$, the LLN holds, which means that $\hat{f}_N$ converges pointwisely to $f(x)$. This is known as *pointwise* LLN. Unfortunately, this convergence is usually not enough to guarantee good properties in the context of stochastic optimization. This is why we recall the notion of uniform LLN.

Definition 1.22 (Uniform LLN) *We say that* $\hat{f}_N(x)$ *converges to* $f(x)$ *with probability 1* uniformly *on* X *if*

$$\sup_{x\in X}\left|\hat{f}_N(x) - f(x)\right| \to 0, \tag{1.6}$$

with probability 1 as $N \to \infty$.

Let us recall that $F(x, \xi)$ is *dominated* by another integrable function g, if $\mathbb{E}[g(\xi)] < +\infty$ and for every $x \in X$, the inequality

$$|F(x, \xi)| \leq g(\xi)$$

holds, with probability 1. The following theorems give us sufficient conditions to guarantee the desired uniform LLN.

Theorem 1.23 *Let $X \subseteq \mathbb{R}^n$ be compact and non-empty, and suppose that: 1) for every $x \in X$, $F(\cdot, \xi)$ is continuous at x for a.e. $\xi \in \Xi$; 2) $F(x, \xi)$ is dominated by an integrable function; and 3) the sample $\{\xi^j\}$ is* iid. *Then, the expected value function $f(x)$ is finite-valued and continuous on X, and*

$$\hat{f}_N(x) \to f(x)$$

with probability 1, uniformly on X.

Theorem 1.24 *Suppose that the random function $F(x, \xi)$ is also convex, and $X \subseteq \mathbb{R}^n$ compact and non-empty. Suppose that the expectation function $f(x)$ is finite-valued on a neighborhood of X and the* pointwise *LLN holds for every x in that neighborhood. Then,*

$$\hat{f}_N(x) \to f(x)$$

with probability 1, uniformly on X.

1.2.3 Monte Carlo Sampling Methods

Let $f : [0,1]^n \to \mathbb{R}$ be an integrable function (with respect to the usual Lebesgue measure on $[0,1]^n$). When working with numerical integration, the task of computing

$$I(f) = \int_{[0,1]^n} f(x)dx$$

can not be obtained analytically in general, since f could not have an explicit primitive function; or, if it does exist, the expression might fail to be computable.

In order to tackle this difficulty, we recall in the following what is called the *Monte Carlo* method. Here, we try to approximate the integral I by an expression of the form

$$I_N(f) \doteq \frac{1}{N} \sum_{j=1}^{N} f(x_j), \tag{1.7}$$

where $\{x_1, \ldots, x_N\} \subseteq [0,1]^n$.

Definition 1.25 (Integration Error) *Given $f : [0,1]^n \to \mathbb{R}$ and a set $\{x_1, \ldots, x_N\}$, we define the* integration error *as*

$$e(f, \{x_1, \ldots, x_N\}) = I(f) - I_N(f) \tag{1.8}$$

The main idea is to pick N independent and uniformly distributed random variables $X_1, \dots, X_N$ in $[0,1]^n$, choose a realization for every one of them, and check the expected value for the resulting error. In other words, use I_N as an statistical estimator for I, writing now

$$\hat{I}_N(f) \doteq \frac{1}{N}\sum_{j=1}^{N} f(X_j).$$

Considering f as a random variable on the probability space $([0,1]^n, \mathcal{B}(\mathbb{R}^\ell), \lambda_n)$, where λ_n is the Lebesgue measure; we can obtain that, using the linearity of the expected value,

$$\mathbb{E}\left[\hat{I}_N(f)\right] = \frac{1}{N}\sum_{j=1}^{N}\mathbb{E}[f] = \mathbb{E}[f] = I,$$

and therefore, $\hat{I}_N(f)$ is an unbiased estimator of $I(f)$. Moreover, Theorem 1.20 ensures that

$$\mathbb{P}\left[\lim_{N\to\infty}\hat{I}_N(f) = I(f)\right] = 1.$$

To bound the estimation error of $\hat{I}_N(f)$, we need to control its variance.

Definition 1.26 (Variance and Standard Deviation) *The* variance *of f is given by*

$$\mathbb{V}[f] \doteq \int_{[0,1]^n}\left(f(x) - \int_{[0,1]^n} f(y)dy\right)^2 dx, \tag{1.9}$$

and its standard deviation *is given by* $\sigma[f] \doteq \sqrt{\mathbb{V}[f]}$.

Theorem 1.27 *Let $f \in L_2([0,1]^N)$ and $X_1, \dots, X_N$ be independent and uniformly distributed random variables. Then, for every $N \in \mathbb{N}$, we have*

$$\mathbb{V}[\hat{I}_N(f)] = \frac{\mathbb{V}[f]}{N}.$$

Note that

$$\mathbb{V}[\hat{I}_N(f)] = \mathbb{E}[(\hat{I}_N(f) - I(f))^2] = \mathbb{E}[e^2(f, \{X_1, \dots, X_N\})]$$

Therefore, the following result follows directly from Jensen's inequality and Theorem (1.27).

Corollary 1.28

$$\mathbb{E}\left[|e(f, \{X_1, \dots, X_N\})|\right] \leq \sqrt{\mathbb{E}[e^2(f, \{X_1, \dots, X_N\})]} = \frac{\sigma[f]}{\sqrt{N}}.$$

For more details on the Monte Carlo method, its properties and extensions, we refer to [25, 92].

Chapter 2

Elements of Bilevel Programming and Uncertainty

In this chapter we review the key elements of Bilevel programming and Optimization that are used to develop the scientific contributions of this thesis. While both fields are quite broad, we will try to keep the exposition as concise and self-contained as possible. We present both subjects together since optimization with recourse can be viewed as a particular case of bilevel programming with uncertainty, where the upper and lower level share the same objective function.

2.1 Deterministic Bilevel Programming

Bilevel optimization tackles a special kind of programs, where one problem is *nested* within another, forcing some variables to be the solution of a second optimization problem. This allows bilevel optimization to model *hierarchical* decision processes.

Historically, the first study related to bilevel optimization can be found in the seminal work of Heinrich von Stackelberg [123] in the field of game theory. This work from 1934 studies the economic equilibrium of a duopoly, under the particular condition where one of the enterprises "somehow knows" the decision process of the other one. The model captures the possible asymmetry between "big and small" firms. The larger firm is known as the *leader* and it makes its decision first, taking into account the future reaction of the smaller firm, known as the *follower*. The follower then reacts optimally in a second stage.

While this first model was focus on traditional economics, the proposed hierarchical structure, nowadays known as bilevel programming or Stackelberg games, derived in many applications such as: toll-setting problems [26]; structural optimization [30]; human resources models [13] and water resources allocation [28], to mention a few.

In this section, we will review the main tools needed for its mathematical formulation, the existence of solutions for bilevel programs, and reformulations to single-level nonlinear problems.

2.1.1 Mathematical Formulation

We start with the formal definition of a bilevel optimization problem, as presented in [10].

Definition 2.1 (Bilevel Programming problem) *A general Bilevel Optimization or Bilevel Programming problem can be stated as*

$$\begin{cases} \min\limits_{x,y} & F(x,y) \\ s.t. & x \in X,\ y \in S(x) \end{cases} \tag{2.1}$$

where $S(x)$ is the set of optimal solutions of the problem

$$\begin{cases} \min\limits_{y \in Y} & f(x,y) \\ s.t. & g(x,y) \leq 0 \end{cases} \tag{2.2}$$

Here, $X \in \mathbb{R}^n$, $Y \in \mathbb{R}^m$, $F, f : X \times Y \to \mathbb{R}$, and $g : X \times Y \to \mathbb{R}^\ell$. The problem formed by the equations (2.1) is usually called the *upper-level* or *leader's* problem and the equations (2.2) form the so-called *lower-level* or *follower's* problem. In some cases, we can add to the upper-level an extra constraint $G(x,y) \leq 0$, where $G : X \times Y \to \mathbb{R}^p$, called the *coupling constraint*. However, in the context of this work, we will maintain the setting of Definition 2.1.

The names leader and follower come from the setting of Stackelberg games, where both problems (2.1)-(2.2) are solved by different agents within the hierarchical structure. That is, the leader decides x first, by solving (2.1) and anticipating the decision y to be rational, that is, to respect the inclusion $y \in S(x)$. Then, the follower reacts, deciding y by solving the parametric problem (2.2).

Definition 2.2 (Shared Constraint Set and Inducible Region) *For the bilevel programming problem* (2.1)-(2.2)*, we define the following sets:*

1. *The set*
$$Z = \{(x,y) \in X \times Y \ : \ g(x,y) \leq 0\}$$
is called the shared constraint set*, and its projection onto the x-space is denoted by*
$$Z_x = \{x \in X \ : \ \exists y,\ (x,y) \in \ Z\}.$$

2. *The set*
$$\mathcal{F} = \{(x,y) \in X \times Y \ : \ x \in Z_x,\ y \in S(x)\} = \operatorname{gph}(S),$$
is called the inducible region *or* bilevel feasible set.

In this work, just as in (2.1), the *optimistic* version of the bilevel problem is considered, where the leader is able to optimize over the lower-level outcome $y \in S(x)$, whenever this set fails to be a singleton. However, this is not the only alternative to deal with the situation where the optimal response of the follower is not unique.

Another popular approach is the *pessimistic* version of the problem, which is defined as

$$\begin{array}{ll} \min\limits_{x \in X} & \max\limits_{y \in S(x)} F(x,y) \\ \text{s.t.} & x \in X. \end{array} \tag{2.3}$$

The main difference here is that the leader and the follower have an *adversarial* (but still rational) behaviour. In this case, the leader must be prepared for every rational decision of the follower, hence, for any decision $y \in S(x)$ that optimizes the follower's objective function. The pessimistic approach assumes that the follower will select the response $y \in S(x)$ that harms the leader the most, leading to the robust formulation (2.3).

Example Let $S : [0, 10] \rightrightarrows [0, 8]$ with

$$\text{gph}(S) = \text{co}\{(0,4),\ (8,0),\ (8,8),\ (10,1),\ (10,5)\}$$

and consider the leader's problem given by

$$\min_{x} \{y :\ y \in S(x),\ x \in [0,10]\}. \tag{2.4}$$

Let $\varphi^o(x) = \min_y y$ and $\varphi^p(x) = \max_y y$. The following figure shows the difference between these approaches (φ^o in red and φ^p in blue).

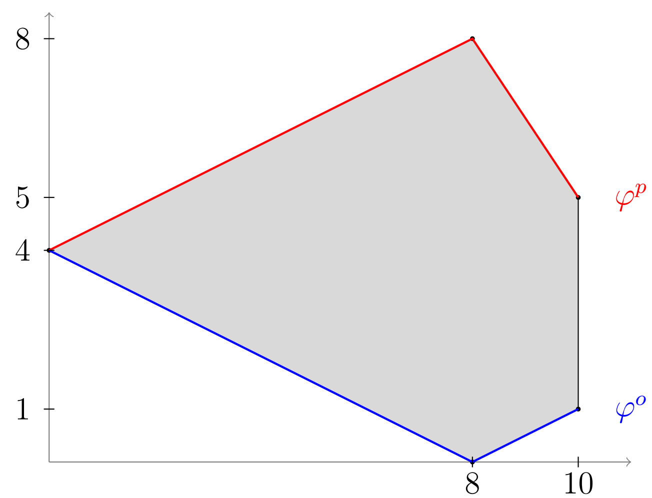

Figure 2.1: Difference between approaches (2.1) and (2.3).

On the one hand, for the optimistic approach, the optimal decision of the leader is $x = 8$ with $y = \varphi^o(x) = 0$. On the other hand, the optimal decision for the pessimistic approach is $x = 0$ with $y = \varphi^p(x) = 4$. ♢

Remark If the lower-level solution is unique for all $x \in Z_x$, both the pessimistic and the optimistic variants of the bilevel problem coincide.

Some other approaches have been proposed in the literature, such as moderate approach [1, 62, 3], selection approach (see, e.g., [34, pp. 6]), and Intermediate/Bayesian approach [83, 109]. Nevertheless, the most relevant in the literature are the pessimistic and optimistic ones. For a more in-depth revision of the pessimistic approach for Stackelberg games, we refer to [126, 72, 73].

2.1.2 Existence of Solutions

Bilevel programs can be studied as non-linear optimization problems. Therefore, a natural approach is to analyze its existence of solutions by applying Weierstrass Theorem. The following theorem, shows the simplest result in this way, deduced from the well-known Maximum Principle (see Chapter 1). Since this theorem is hard to find in the literature in its presented form, we include a short proof.

Theorem 2.3 *For Problem* (2.1)-(2.2)*, let us consider the set-valued map* $K : X \rightrightarrows Y$ *given by*

$$K(x) = \{y \ : \ g(x, y) \leq 0\},$$

and suppose that: 1) F *is lower-semicontinuous; 2)* f *and* g *are continuous; 3)* X *and* Y *are compact sets and; 4)* $K(x)$ *is continuous in the sense of multifunctions (see Chapter 1). Then, Problem* (2.1)-(2.2) *admits a solution.*

Proof. Given that Y is a compact set and g is continuous, $K : X \rightrightarrows Y$ is closed-valued, and therefore compact-valued. Since f is continuous, we deduce directly that

$$S(x) = \arg\min_{y}\{f(y) \ : \ y \in K(x)\}$$

is closed and nonempty for every $x \in X$. Furthermore, we can apply the Maximum Principle (1.11) and conclude that S is upper-semicontinuous, and hence closed. Now, since $\mathrm{gph}(S) \subset X \times Y$, we conclude that it is compact.

Finally, considering that the upper-level problem (2.1) can be written as

$$\left\{\begin{array}{ll} \min\limits_{x,y} & F(x, y) \\ \text{s.t.} & (x, y) \in \mathrm{gph}(S) \end{array}\right. , \tag{2.5}$$

and applying a last time the Weierstrass Theorem, given that F is a lower-semicontinuous function over a compact set, we conclude that Problem (2.1)-(2.2) has a solution. □

In general, upper semicontinuity for K is easy to obtain. In fact, continuity of g suffices for having this result (see Chapter 1). Nonetheless, lower semicontinuity is harder to get. In some cases, we can assume that $K(x) = Y$, which verifies lower semicontinuity trivially. This case consist in bilevel programs where the leader's decisions only affect the follower's objective function. Another favorable case where lower semicontinuity is also verified is when K is given by linear constraints, as stated in the following result.

Theorem 2.4 *If $K(x)$ is given by linear constraints, that is, when*

$$g(x, y) = Ax + By - b$$

(where A, B and b are two matrices and a vector of appropiated dimensions), then K is lower semicontinuous. Therefore, in this context, Problem (2.1)-(2.2) *admits a solution.*

2.1.3 Single-Level Karush-Kuhn-Tucker Reformulation

Starting with the seminal work of [43], a popular way used to solve a bilevel problem consists in re-formulate it into a single-level optimization problem. There are at least two approaches in order to do so: consider the *optimal value function* of the problem, or reformulate it using the Karush-Kuhn-Tucker equations for the lower-level problem.

Let us first discuss the value function alternative. Defining the optimal value function for the lower-level problem

$$\varphi(x) \doteq \min_{y \in Y}\{f(x, y) \ : \ g(x, y) \leq 0\}, \tag{2.6}$$

we can re-formulate the general optimistic bilevel problem (2.1)-(2.2) as

$$\begin{cases} \min\limits_{x \in X, y \in Y} & F(x, y) \\ \text{s.t.} & g(x, y) \leq 0 \\ & f(x, y) \leq \varphi(x) \end{cases} \tag{2.7}$$

which is now a single-level problem. However, we have a major set of difficulties working with the function φ. This function can be evaluated, but in order to do so, we have to solve, for the given x, the lower-level problem. Moreover, in general, this function does not have an algebraic expression.

There is a lot of work in the study of optimal value functions, also known as *marginal functions* and its difficulties. As can be seen in [105, 47, 24] to name a few, this functions are deeply non-smooth, so generalized derivatives of various kinds are used to study their properties. In later works from B. S. Mordukhovich [89, 88, 90], significant progress was made in this area.

Now, we will discuss the alternative using Karush-Kuhn-Tucker equations, which is the one we use in this thesis. Let us suppose that the lower level problem is parametric convex, that is, for all $x \in X$, the set $K(x)$ is given by $K(x) = \{y \ : \ g(x, y) \leq 0\}$, every function $g_i(x, \cdot)$ is convex for each $i = 1, \ldots, \ell$, and $f(x, \cdot)$ is convex. Then, the following direct application of the celebrated Karush-Kuhn-Tucker theorem holds (see, e.g., [32, 10]).

Theorem 2.5 (First-order optimality for the lower-level problem) *Consider the bilevel problem* (2.1)-(2.2) *with parametric convex lower level. Suppose $f(x, \cdot)$ is differentiable. Then*

$$y \in S(x) \iff 0 \in \nabla_y f(x, y) + N_{K(x)}(y). \tag{2.8}$$

Moreover, if $g : X \times Y \to \mathbb{R}^\ell$ is continuously differentiable, and $K(x)$ satisfies any Constraint Qualification (see Chapter 1), then we can write

$$y \in S(x) \iff \exists \lambda \in \mathbb{R}^\ell \text{ such that } \begin{cases} \nabla_y f(x,y) + \sum_i \lambda_i \nabla_y g_i(x,y) = 0 \\ 0 \leq \lambda \perp g(x,y) \leq 0. \end{cases} \tag{2.9}$$

When our problem is reformulated using its Karush-Kuhn-Tucker equations given by (2.9), we can reduce the overall bilevel optimization problem to a single-level constrained optimization problem. This new problem is the *mathematical program with complementarity constraints* (MPCC) reformulation, and is given by

$$\begin{cases} \min\limits_{x,y} & F(x,y) \\ \text{s.t.} & \nabla_y f(x^*,y^*) + \sum_{i=1}^{\ell} \lambda_i \nabla_y g_i(x^*,y^*) = 0 \\ & \lambda \geq 0,\ \lambda^T g(x^*,y^*) = 0 \\ & x \in X \end{cases} \tag{2.10}$$

MPCC problems are a research field in itself, beyond its applications to bilevel programming. For an in-depth revision on the details of this kind of problems, we refer to [80]. There is an active field of work onto optimally solving MPCC problems, considering different strategies as can be seen in [45, 58, 40] to name a few.

At this point, a natural question arises: is the MPCC reformulation (2.10) equivalent to (2.1)-(2.2), in the sense of sharing the same optimal solutions? Let us start by looking at the global solutions of both problems.

Theorem 2.6 (See [33, Theorem 2.1]) *Let (x^*, y^*) be a global optimal solution of* (2.1)-(2.2). *If* (2.2) *is convex and satisfies LICQ for x^*, then the point (x^*, y^*, λ^*) is a global optimal solution of* (2.10), *for every*

$$\lambda \in \Lambda(x^*,y^*) = \left\{ \lambda \geq 0 \ : \ \nabla_y f(x^*,y^*) + \lambda^T \nabla_y g(x^*,y^*) = 0,\ \lambda^T g(x^*,y^*) = 0 \right\}.$$

Under some additional hypothesis, the converse is also true, as stated in the following result.

Theorem 2.7 (See [33, Theorem 2.3]) *Let (x^*, y^*, λ^*) be a global optimal solution of* (2.10) *and let* (2.2) *be convex. If the lower-level problem satisfies LICQ for all $x \in X$, then (x^*, y^*) is a global optimal solution of* (2.1)-(2.2).

Once revised the global optimal solutions behaviour, we can now consider the relationship between the local optimal values of (2.1)-(2.2) and (2.10). As it turns out, the local optima of (2.10) is not necessarily local optima of (2.1)-(2.2). In [33], additional contraint qualifications are considered, in order to obtain equivalence for local optima.

Definition 2.8 (Slater's and Constant Rank CQs) *For the bilevel problem* (2.1)-(2.2), *we consider the following constraint qualifications:*

1. ***Slater's CQ:*** *There exists* $\bar{y}(x)$ *such that* $g_i(x, \bar{y}(x)) < 0$, *for* $i = 1, \ldots, \ell$.

2. ***Constant Rank CQ (at*** (x^*, y^*)***):*** *There exists an open neighborhood* V *of* (x^*, y^*) *such that for each* $I \subseteq \{j : g_j(x^*, y^*) = 0\}$, *the family of gradients* $\{\nabla_y g_j(x, y) : j \in I\}$ *has the same rank on* V.

The main results of [33] can be summarized in the following theorem.

Theorem 2.9 (See [33, Theorem 3.2, Corollary 3.3]) *Let the problem* (2.2) *be convex, Slater's CQ be satisfied at* x^*, *and* (x^*, y^*, λ^*) *be a local optimal solution of* (2.10), *for all* $\lambda^* \in \Lambda(x^*, y^*)$. *Then,* (x^*, y^*) *is a local optimal solution of problem* (2.1)-(2.2) *too. Additionally, if Constant Rank CQ is assumed, local optimality of* (x^*, y^*, λ^*) *for all* vertices $\lambda^* \in \Lambda(x^*, y^*)$, *implies local optimality of* (x^*, y^*).

2.1.4 Stochastic Bilevel Programming

Uncertainty has been recently introduced in bilevel programming models, defining what we call *stochastic bilevel programs.*

In the classical (one-level) optimization context, there are mainly two ways to address uncertainty: *stochastic optimization*, where the uncertain parameter is modeled by probabilistic scenarios observed by random variables (see, e.g. [21, 64]); and *robust optimization* (see, e.g. [14, 17, 15]), where the range of the uncertain parameter values is defined by a set, over which the worst-case scenarios must be considered. This same two roads have been developed for bilevel optimization problems.

The *sources* of uncertainty in bilevel programming can be separated mainly in two kinds: *data* and *decision.*

Data uncertainty is seen when, for example, the lower-level agent only has limited access to data, or its data is inaccurate. Let represent this, for a leader's decision x, and a specific realization of the uncertainty ξ, with the set of optimal follower's decisions

$$S(x, \xi) = \begin{cases} \min\limits_{y \in Y} & f(x, y) \\ \text{s.t.} & g(x, y) \leq z(\xi), \end{cases} \tag{2.11}$$

where $z(\xi) \in \mathbb{R}^\ell$ represents the uncertainty, for the given realization ξ. Popular approaches are the two following.

- Uncertainty is assumed to take values in a given set Ξ. Following a *robust* approach, we can consider the worst-case scenario for the uncertainty realizations; i.e.,

$$\begin{array}{ll} \min\limits_{x \in X} & \max\limits_{\xi \in \Xi} F(x, y) \\ \text{s.t.} & x \in X, \; y \in S(x, \xi). \end{array} \tag{2.12}$$

- On a more *stochastic* approach, we assume that the uncertainty can be tackled considering its distributions. Hence, we can consider solving the problem in a probabilistic sense, by optimizing the expected value

$$\begin{aligned} \min_{x,y} \quad & \mathbb{E}_\xi[F(x,y)] \\ \text{s.t.} \quad & x \in X,\ y \in S(x,\xi). \end{aligned} \tag{2.13}$$

In both of these cases, we are considering a specific timing,

$$x \text{ is decided } \to \xi \text{ is revealed } \to y \text{ is decided considering } (x,\xi), \tag{2.14}$$

which is: the leader makes a *here-and-now* decision (without information about the realization of the uncertainty), then uncertainty realizes, and finally the follower decides in a *wait-and-see* scenario.

However, other timings can be taken into account. For example, considering problems of the form

$$\begin{aligned} \min_{x,y} \quad & F(x,y) \\ \text{s.t.} \quad & x \in X,\ y \in S(x,\xi), \end{aligned} \tag{2.15}$$

where now $S(x,\xi)$ has the form

$$S(x,\xi) = \begin{cases} \min\limits_{\bar{y}\in Y} & f(x,\bar{y}) \\ \text{s.t.} & g(x,\bar{y}) \leq z(\xi),\ \forall \xi \in \Xi \end{cases}. \tag{2.16}$$

That is, when we consider some kind of *fixed* scenario for y (representing e.g. an average situation), given that in this case, the follower also makes its decision *before* the uncertainty realizes. Hence, we are considering the timing

$$x \text{ is decided } \to y \text{ is decided considering } x \to \xi \text{ is revealed.} \tag{2.17}$$

Data uncertainty can also happen in the upper-level problem's data and even in its objective function.

On a different scope, uncertainty can be also taken into account at the decision making. *Decision uncertainty* refers to cases where any of the agents face uncertainties regarding the decision of the other ones. Although this kind of uncertainty is not considered in this work, we refer to [11, 18] for extensive reviews on the subject.

2.2 Same Agent: Two-Stage Problems

Two-Stage problems, also known as Recourse Programs, are optimization problems where a single-agent must take some decisions before some random events occur, and other decisions (also called recourse actions) afterwards. In this sense, Recourse programs can be seen as a particular case of Stochastic Bilevel programming, where the same agent is solving the upper-level and lower-level problem.

Therefore, we have two kinds of actions in this context: There is the decision vector $x \in \mathbb{R}^n$ taken *before* the uncertainty, modeled as a random vector $\xi(\omega)$, occurs. We call them *first-stage* or *here-and-now* decisions. Then, *after* the uncertainty data reveals itself, there is a new set of decisions y to be made, for which the decision-making process takes into account both, the first-stage decision and the realization of the random vector. These are called the *second-stage* or *wait-and-see* decisions. This chain of events can be then modeled as

$$x \text{ is decided } \to \xi(\omega) \text{ is revealed } \to y \text{ is decided considering } (x, \xi(\omega)). \tag{2.18}$$

The first references of the two-stage stochastic linear program with recourse can be found in [9, 31]. Specifically in [31], George B. Dantzig proposed that linear programming methods could be extended to include the case of uncertain information, such as the problem of optimal allocation of a carrier fleet to airline routes to meet an anticipated demand distribution.

Recourse programs can be easily used in order to model different situations. First off, the uncertainty may represent a limited number of well studied scenarios of a specific event. This way of using the uncertainty has been used, for example, in overbooking and revenue management [2, 121]. Another modeling option, consists on using the randomness to represent uncertainties that recur frequently on a short-term basis, over a longer time horizon. Then, it would be desirable to compute a mean over a lot of possible values of this daily operation, so that the expected value will match closely. This kind of models have been applied in optimal design of non-conventional renewable energy systems [127, 51].

Some relevant applications of Two-Stage programming are, to name a few, stochastic vehicle routing [122], stochastic networks and stochastic facility locations problems [54] have been mostly treated as a natural extension of the stochastic transportation problem with simple recourse [128, 55]. An important extension of Recourse Programs is called *Multistage Stochastic Programs*, which has allowed a more realistic treatment of the dynamics or sequential structures of decision problems. Approaches to portfolio management have become the cornerstone of the contemporary financial applications and have contributed also to modeling and software development for multistage stochastic programs. See, e.g. [84, 36].

Nowadays, one popular result is financial applications of stochastic programming. Another application areas contains planning and allocation of resources [68, 70, 71], energy production and transmission [49], production planning and optimization of technological processes, logistics problems [99, 46] and telecommunications [103, 104].

In this section, we summarize the key elements of Two-Stage Stochastic Programming. For a more extensive primer in the field, we refer to [75] and [111], and a list of applications, achievements and unsolved problems which can be found in [39, 124].

2.2.1 Mathematical Formulation

Let $(\Omega, \Sigma, \mathbb{P})$ a probability space and let $\xi : \Omega \to \mathbb{R}^p$ be a random vector. A general two-stage problem can be stated as

$$\min_{x \in X} \; f_1(x) + \mathbb{E}[Q(x, \xi)], \tag{2.19}$$

where X is a nonempty (usually compact) subset of $\mathbb{R}^n$, $f_1 : \mathbb{R}^n \to \mathbb{R}$ is a real-valued (usually continuous) function, and $Q(x, \xi)$ is obtained as the optimal value of the so called *second-stage* problem

$$\begin{cases} \min\limits_{y} & f_2(x, y, \xi) \\ \text{s.t.} & y \in Y(x, \xi). \end{cases} \tag{2.20}$$

Here, $f_2 : \mathbb{R}^n \times \mathbb{R}^m \times \mathbb{R}^p \to \mathbb{R}$ and $Y : \mathbb{R}^n \times \mathbb{R}^d \rightrightarrows \mathbb{R}^m$ is a set-valued function. This representation shows the sequence of events, where the first-stage decisions x are taken with uncertainty about future realizations of ξ, and after that, some corrective actions y can be taken. The first-stage decisions are made taking future effects into account, which are measured by the expected value of the function $Q(x, \cdot)$, also known as the *cost-to-go* function. The main difficulty in stochastic programming lies then in the computation of the expected value $\mathbb{E}[Q(x, \xi)]$. Therefore, it is interesting to study the main properties on this function, in order to get its optimal values.

Definition 2.10 (Carathéodory Function) *We say that a function $F : \mathbb{R}^d \times \Omega \to \mathbb{R}$ is a Carathéodory function if*

1. *$F(z, \cdot)$ is measurable $z \in \mathbb{R}^d$,*
2. *$F(\cdot, \omega)$ is continuous for almost every $\omega \in \Omega$.*

Proposition 2.11 *If $(x, y, \omega) \mapsto f_2(x, y, \xi(\omega))$ is a Carathéodory function, then the optimal value function $Q(x, \cdot)$ is measurable. Moreover, if $Q(\cdot, \xi)$ is continuous for a.e. $\xi \in \Omega$, then $Q(x, \xi)$ is also a Carathéodory function.*

The previous proposition is important in order to prove the following theorem.

Theorem 2.12 *Let $\mathcal{M} = \mathcal{L}_p(\Omega, \mathcal{F}, P)$ with $p \in [1, +\infty]$. The general two-stage problem* (2.19)+(2.20) *is equivalent to the problem*

$$\begin{cases} \min\limits_{x \in \mathbb{R}^n,\ y \in \mathcal{M}} & f_1(x) + \mathbb{E}[f_2(x, y(\xi), \xi)] \\ s.t. & x \in X,\ \ y(\xi) \in Y(x, \xi)\ \ a.e.\ \ \xi \in \Omega. \end{cases} \tag{2.21}$$

in the sense that optimal values of problems (2.19) *and* (2.21) *are equal to each other, provided that the optimal value of* (2.21) *is finite. Moreover, $(\bar{x}, \bar{y})$ is an optimal solution of problem* (2.21) *if and only if $\bar{x}$ is a solution of* (2.19) *and $\bar{y} = \bar{y}(\xi)$ is an optimal solution of* (2.20).

If the set X is closed and convex, and for every $\xi \in \Omega$ the function

$$\bar{f}_2(x, y, \xi) = \begin{cases} f_2(x, y, \xi) & \text{if } y \in Y(x, \xi) \\ +\infty & \text{otherwise} \end{cases} \tag{2.22}$$

is convex in $(x, y) \in \mathbb{R}^{n+m}$, we say that the problem (2.19)+(2.20) is convex. In this context, the optimal value function $Q(x, \xi)$ is also convex [111, Section 2.3.2], and therefore (2.19) is a convex problem.

In our work, we will focus on a particular class of Two-Stage Linear Programs, which have the general form

$$\begin{cases} \min\limits_{x \in X} & c^T x + \mathbb{E}[Q(x, \xi)] \\ \text{s.t.} & A_1 x \leq b_1, \end{cases} \tag{2.23}$$

where $c \in \mathbb{R}^n$, $b_1 \in \mathbb{R}^m$ are known vectors, $A_1 \in \mathcal{M}_{m \times n}(\mathbb{R})$ and $Q(x, \xi)$ is the cost-to-go function of a problem of the form

$$Q(x, \xi) = \begin{cases} \min\limits_{y \in Y} & q^T y \\ \text{s.t.} & A_2 x + B_2 y \leq b_2(\xi). \end{cases} \tag{2.24}$$

Here, A_2 and B_2 (known as the *recourse* matrix) are fixed matrices, and the parameter $b_2(\xi)$ is the data of the second-stage problem, where the randomness is taken into account. The problem (2.23)+(2.24) is known in this context as a *Two-Stage Stochastic Linear Program with Fixed Recourse.*

Definition 2.13 *A random variable ξ is said to be* discrete *if the set of values it can take (also called* support*) has either a finite or an infinite but countable number of elements.*

We now present some basic properties when ξ is a discrete random variable. This is an important class of random variables, because are widely used in applications, either directly or through sampling of a continuous distribution.

Definition 2.14 (Feasibility Sets)

- *We define the* first-stage *feasibility set as*

$$K_1 = \{x \;:\; A_1 x \leq b_1\}. \tag{2.25}$$

- *For a given ξ, we define the* elementary second-stage *feasibility set as*

$$K_2(\xi) = \{x \;:\; y \text{ exists s.t. } A_2 x + B_2 y \leq b_2(\xi)\}. \tag{2.26}$$

- *When ξ is discrete with support Ξ, we define the* second-stage *feasibility set as*

$$K_2 = \bigcap_{\xi \in \Xi} K_2(\xi). \tag{2.27}$$

Theorem 2.15

1. *For any given realization of ξ, $K_2(\xi)$ is a convex polyhedron.*

2. *When ξ is finite and discrete, K_2 is a convex polyhedron.*

Definition 2.16 (Polyhedral Function) *An extended real valued function $f : \mathbb{R}^n \to \bar{\mathbb{R}}$ is called* polyhedral *if it is proper convex and lower semicontinuous, its domain is a convex closed polyhedron, and f is piece-wise linear on its domain.*

Proposition 2.17 *For problem* (2.24)*, for any given realization of ξ, the function $Q(\cdot, \xi)$ is convex. Moreover, when ξ is finite and discrete, $Q(\cdot, \xi)$ is polyhedral.*

Proposition 2.18 *If ξ is discrete and finite, and $\mathbb{E}[Q(\cdot, \xi)]$ has a finite value in at least one point x, then $\mathbb{E}[Q(\cdot, \xi)]$ is polyhedral in K_2.*

When ξ is not a discrete random variable, we may now re-define K_2 as,

$$K_2 = \{x \ : \ Q(x, \xi) < \infty\} \tag{2.28}$$

and our first definition (for the discrete case) can be re-named as in [21] as the *possibility interpretation* of K_2

$$K_2^p = \bigcup_{\xi \in \Xi} K_2(\xi) \tag{2.29}$$

Intuitively, a first-stage decision x belongs to K_2^p if, for all *possible* values of ξ, a feasible second-stage decision y can be taken.

Theorem 2.19 *If ξ has finite second moments,*

1. *The sets K_2 and K_2^p coincide.*

2. *$\mathbb{E}[Q(\cdot, \xi)]$ is a Lipschitz convex function, and is finite in K_2.*

Theorem 2.20 *Suppose that ξ has finite second moments and K is bounded. Then, if problem* (2.23) *has a finite optimal value, it is attained for some $x \in \mathbb{R}^n$.*

2.2.2 The Sample Average Approximation Method

The Sample Average Approximation (SAA), firstly shown in [67], allows one to tackle two-stage problems through the use of sampling and optimization Monte-Carlo based methods, in order to be able to use deterministic programming tools to solve them. In this section, we will briefly visit how the Sample Average Approximation Method works. For recent surveys, please refer to [56, 111].

Let us consider a minimization problem to be the expected value of a function, that is,

$$\min_{x \in X} q(x) = \mathbb{E}[Q(x, \xi)], \tag{2.30}$$

where $x \in \mathbb{R}^n$ and $\xi \in \mathbb{R}^m$, $X \subseteq \mathbb{R}^n$ is a fixed set and $\xi = \xi(\omega)$ is a random vector. We denote by $\Xi \in \mathbb{R}^m$ the support of the probability distribution of ξ, that is, the smallest closed set of $\mathbb{R}^m$ such that $\mathbb{P}(\xi \in \mathbb{R}^m \setminus \Xi) = 0$.

If we have a set of i.i.d. random vectors with the same distribution as ξ, say $\{\xi^1, \xi^2, \ldots, \xi^N\}$ (normally called a *sample*) of size N, we can estimate $q(x)$ by averaging the values of $Q(x, \xi^j)$, with $j = 1, \ldots, N$. This leads to the main idea of the Sample Average Approximation (SAA) method, where we replace $q(x)$ with its average approximation

$$q_N(x) \doteq \frac{1}{N} \sum_{j=1}^{N} Q(x, \xi^j). \tag{2.31}$$

From a statistics approach, since each ξ^i has the same distribution as ξ, it is clear that

$$\mathbb{E}\left[q_N(x)\right] = q(x), \tag{2.32}$$

and then, $q_N(x)$ is an unbiased estimator for $q(x)$. Then, given a realization $\{\hat{\xi}^1, \hat{\xi}^2, \ldots, \hat{\xi}^N\}$ of the sample, we can define

$$\hat{q}_N(x, \hat{\xi}^1, \ldots, \hat{\xi}^N) \doteq \frac{1}{N} \sum_{j=1}^{N} Q(x, \hat{\xi}^j), \tag{2.33}$$

which is now a deterministic expression, so it can be tackled with standard optimization tools, in order to get an approximated solution of (2.30) by computing the solution of

$$\min_{x \in X} \hat{q}_N(x, \hat{\xi}^1, \ldots, \hat{\xi}^N) = \frac{1}{N} \sum_{j=1}^{N} Q(x, \hat{\xi}^j), \tag{2.34}$$

The optimal value $\hat{\nu}_N$ and an optimal solution $\hat{x}_N$ of the problem (2.34) will be considered statistical estimators of their counterparts of the original problem (2.30), ν^* and x^* respectively. Naturally, when using this technique, some questions arise

1. Is this method consistent? That is, do the solution of the SAA version of the problem converge in some way to the original problem solution?

2. Can we give any guarantees about a solution obtained by this method? Are the obtained solutions any good?

Statistical Properties of SAA Estimators

Definition 2.21 (Consistency) *Let $\hat{S}_N$ be the set of optimal solutions of* (2.34) *and S the set of optimal solutions of the original problem* (2.30). *We say that*

1. *$\hat{\nu}_N$ is a* consistent *estimator of ν^* if $\hat{\nu}_N \to \nu^*$ with probability one, as $N \to \infty$.*

2. *$\hat{x}_N$ is* consistent *if dist$(\hat{x}_N, S) = 0$ with probability one, as $N \to \infty$.*

Theorem 2.22 *Suppose that for almost every $\xi \in \Xi$ the function $Q(\cdot,\xi)$ is convex, the function $q(\cdot)$ is lower semicontinuous and its domain has a non-empty interior, and the LLN holds pointwise. Then, $\hat{q}_N$ epiconverges to q with probability one.*

Corollary 2.23 *Suppose that for almost every $\xi \in \Xi$ the function $Q(\cdot,\xi)$ is convex and the LLN holds pointwise. Let $C \subseteq \mathbb{R}^n$ a compact set such that q is finite valued on a neighborhood of C. Then $\hat{q}_N$ epiconverges to f uniformly on C, that is*

$$\sup_{x\in C} |\hat{q}_N(x) - q(x)| \to 0 \quad \textit{with probability one as } N \to \infty.$$

Theorem 2.24 *Suppose that: 1) Q is random lower semicontinuous; 2) For a.e. $\xi \in \Xi$ the function $Q(\cdot,\xi)$ is convex; 3) X is closed and convex, 4) q is lower semicontinuous; 5) The set of optimal solution S of the problem* (2.19) *is non-empty and bounded; and 6) The LLN holds pointwise.*

Then, $\hat{\nu}_N \to \nu^*$ and

$$\sup_{x\in \hat{S}_N} \ dist(x,S) \to 0 \quad \textit{with probability one.}$$

as $N \to \infty$.

Assessing Solution Quality

Lets denote a candidate solution of the original problem as $\hat{x} \in X$. One of the approaches for assessing solution quality is to bound the candidate solution's optimality gap, for which we can use a method called the *Multiple Replications Procedure* (MRP) as presented in [82] .

The optimality gap for $\hat{x}$ can be computed as $f(\hat{x}) - \nu^*$. The value of ν^* is not known, but can be bound as in [94] by the bias result

$$\mathbb{E}[\nu_N] \leq \nu^*.$$

Therefore, an *upper bound* on the optimality gap of $\hat{x}$ (from now on, *point estimator of the optimality gap* of $\hat{x}$), can be estimated via

$$\mathcal{G}_N(\hat{x}) = f_N(\hat{x}) - \nu_N. \tag{2.35}$$

When viewed as an estimator of the optimality gap, $\mathcal{G}_N(\hat{x})$ is biased, $\mathbb{E}[\mathcal{G}_N(\hat{x})] \geq f(\hat{x}) - \nu^*$.

In order to compute this estimator, we can use the same i.i.d. random variables $\xi^1,\ldots,\xi^N$ from the distribution of ξ for both terms in (2.35). That is, given a realization $\{\hat{\xi}^1,\hat{\xi}^2,\ldots\}$ of the random vector ξ, we compute

$$\hat{\mathcal{G}}_N(\hat{x}) = \hat{f}_N(\hat{x},\hat{\xi}^1,\ldots,\hat{\xi}^N) - \hat{\nu}_N(\hat{\xi}^1,\ldots,\hat{\xi}^N),$$

where the notation $\hat{\nu}_N(\hat{\xi}^1, \ldots, \hat{\xi}^N)$ emphasizes that this quantity corresponds to the optimal value of the approximated problem, for the particular realization $\{\hat{\xi}^1, \ldots, \hat{\xi}^N\}$. The use of the same observations in both terms of $\hat{\mathcal{G}}_N$ results in variance reduction.

As shown in [110], $\mathcal{G}_N(\hat{x})$ is typically not asymptotically normal, complicating statistical inference. This difficulty can be avoided employing a "batch-means" approach, commonly used in the simulation literature. Thinking of this approach as an algorithm, the process goes as follows:

1. First off, a number of "batches" $N_{\mathcal{G}}$ is decided, in order to compute a confidence interval for the point estimator of the optimality gap. As mentioned in [56], tipical values for $N_{\mathcal{G}}$ move between 20 or 30.

2. Observations $\{\xi^{k1}, \xi^{k2}, \ldots, \xi^{kN}\}$, for $k = 1, 2, \ldots, N_{\mathcal{G}}$ are generated; and these are averaged to obtain a point estimator of the optimality gap

$$\bar{\mathcal{G}}(\hat{x}) = \frac{1}{N_{\mathcal{G}}} \sum_{k=1}^{N_{\mathcal{G}}} \mathcal{G}_N^k(\hat{x}),$$

   where $\mathcal{G}_N^k$ is computed following (2.35), using the $k-$th batch of observations.

3. We compute the sample variance as

$$S_{\mathcal{G}}^2 = \frac{1}{N_{\mathcal{G}} - 1} \sum_{k=1}^{N_{\mathcal{G}}} \left(\mathcal{G}_N^k(\hat{x}) - \bar{\mathcal{G}}(\hat{x})\right)^2.$$

4. An approximate $(1-\gamma)$-level confidence interval estimator on the optimality gap of $\hat{x}$ is given by

$$\left[0,\ \bar{\mathcal{G}}(\hat{x}) + \frac{z_\gamma S_{\mathcal{G}}}{\sqrt{N_{\mathcal{G}}}}\right],$$

   where z_γ denotes a $(1-\gamma)$-quantile from a standard Normal distribution.

2.2.3 Chance constraints in two-stage problems

Chance constrained optimization appeared firstly in [29] and [86]. It is a probabilistic way of handling uncertainty, by using probability tools in order to tackle inequality constraints. This kind of optimization is a relatively robust approach, often difficult to solve.

Consider a chance constrained problem of the form

$$\begin{cases} \min\limits_{x \in X} & f(x) \\ \text{s.t.} & \mathbb{P}[C(x,\xi) \leq 0] \geq 1 - \gamma, \end{cases} \tag{2.36}$$

where $X \subseteq \mathbb{R}^n$ is a closed deterministic set, $f : \mathbb{R}^n \to \mathbb{R}$ is continuous, $\gamma \in (0,1)$ is some significance level and C is a Carathéodory function.

Problems like (2.36) have been extensively studied. Both from theoretical and computational points of view, it is recognized that chance constrained problems are hard to treat [100, 64], mainly for two primary reasons. First, the function

$$\varphi : x \mapsto \varphi(x) = \mathbb{P}[C(x,\xi) \leq 0]$$

is hard to compute, as this requires multidimensional integration techniques. Therefore, just checking the feasibility of a solution is, in general, difficult. Second, the feasible region defined by a chance constraint generally is not convex, even if X is convex and C is convex in x.

Some recent contributions for this kind of problem have been carried on two different approaches. One is to discretize the function φ and consequently solve the obtained combinatorial problem [78, 79, 35]; and the other one is to employ convex approximations of chance constraints [93].

Chance constraints have been used in various applications including supply chain management [69], production planning [91], optimization of chemical processes [53] and water quality management [116]. For an in-depth theoretical background, see [100].

Chance Constraints and the SAA Method

As shown in [111] (Chapter 5, Section 5.7), a way of solving (2.36) is using a SAA approach. First, note that we can write the probability constraint as

$$\mathbb{P}[C(x,\xi) \leq 0] = \mathbb{E}[\mathbb{1}_{(-\infty,0]}(C(x,\xi))]$$

and estimate this by the corresponding sample average approximation

$$\hat{P}_N(x) = \frac{1}{N}\sum_{i=1}^{N} \mathbb{1}_{(0,\infty)}(C(x,\hat{\xi}^i)). \tag{2.37}$$

Proposition 2.25 ([111, Proposition 5.29]) *Let $C(x,\xi)$ be a Carathéodory function. Then, the functions $\hat{P}_N$ and $\mathbb{P}[C(x,\xi) \leq 0]$ are upper semicontinuous. Moreover, assume that for every $x \in X$ it holds that*

$$\mathbb{P}[\xi \in \Xi \ : \ C(x,\xi) = 0] = 0, \tag{2.38}$$

that is, $C(x,\xi) \neq 0$ with probability one. Then, $\mathbb{P}[C(\cdot,\xi) \leq 0]$ is continuous on X and $\hat{P}_N$ converges to $\mathbb{P}[C(x,\xi) \leq 0]$ with probability one, uniformly on any compact subset of X.

Theorem 2.26 ([111, Proposition 5.30]) *Suppose that in the problem* (2.36) *we have: 1) X is compact; 2) f is continuous; 3) C is a Carathéodory function; 4) the sample $\{\xi^1, \xi^2, \ldots, \xi^N\}$ is iid.; and 5) the following condition holds: there is an optimal solution $\bar{x}$ such that for every $\varepsilon > 0$ there is $x \in X$ with $\|x - \bar{x}\| \leq \varepsilon$ and $\mathbb{P}[C(x,\xi) > 0] < \gamma$. Then, $\hat{\nu}_N \to \nu^*$ and*

$$\sup_{x \in \hat{S}_N} \ dist(x, S) \to 0 \quad \textit{with probability one.}$$

2.3 The Value of Information

When dealing with uncertainty in optimization problems, different levels of information about it can be considered. For a given random variable ξ, one can know: its expected value $\bar{\xi} = \mathbb{E}[\xi]$; its distribution which gives us access to the map $x \mapsto \mathbb{E}[\varphi(x,\xi)]$, for any function φ; and even its future realizations, $\xi(\omega)$, which would allow us to optimize $\varphi(x, \xi(\omega)$ in a deterministic fashion.

Usually, each level of information requires to invest more resources in order to obtain it. Thus, it is important to be able to measure the gain that one might obtain by improving them. We finish this Chapter my quickly summarize the different indicators that allows us to do so.

2.3.1 The Expected Value of Perfect Information

First appearances on the concept of Expected Value of Perfect Information (from now on, EVPI) are seen in the context of decision analysis [101]. This quantifies how much a decision maker would pay, in order to obtain complete information about the future.

Consider (2.23) and define the problem

$$\begin{cases} \min\limits_{x \in X} & \varphi(x,\xi) = c^T x + Q(x,\xi) \\ \text{s.t.} & A_1 x \leq b_1, \end{cases} \tag{2.39}$$

associated with a specific scenario ξ and $Q(x,\xi)$ is the cost-to-go function of the problem (2.24). Assuming that for all $\xi \in \Xi$, there exists at least one $x \in \mathbb{R}^n$ such that $\varphi(x,\xi) < \infty$, it implies the existence of at least one optimal solution, let us denote it $x^*(\xi)$.

Assuming that we have the ability to find these decisions for every ξ, we are able to compute the expected value of φ.

Definition 2.27 (Wait-and-See/Here-and-Now Solutions) *We define the Wait-and-See Solution (from now on, WS) as*

$$WS = \mathbb{E}_\xi[\min_x \varphi(x,\xi)] = \int_\Omega \varphi(x^*(\xi(\omega)), \xi(\omega)) d\mathbb{P}(\omega), \tag{2.40}$$

and the Stochastic Solution (from now on, STO) as

$$STO = \min_x \mathbb{E}_\xi[\varphi(x,\xi)] = \min_x \int_\Omega \varphi(x(\xi(\omega)), \xi(\omega)) d\mathbb{P}(\omega), \tag{2.41}$$

Definition 2.28 (EVPI) *The EVPI is defined as*

$$EVPI = STO - WS. \tag{2.42}$$

The EVPI measures the expected gain of knowing the value of ξ, without uncertainty with respect to STO.

2.3.2 The Value of the Stochastic Solution

When STO is too hard to solve, a simpler idea is to solve the problem obtained by replacing all uncertainty by the expected value of the random variables that model it.

Definition 2.29 *The Expected Value Problem is defined as*

$$EV = \min_x \varphi(x, \bar{\xi}), \tag{2.43}$$

where $\bar{\xi} = \mathbb{E}(\xi)$ *denotes the expected value for the random variable.*

Definition 2.30 (Expected result of using EV) *Let* $x^*(\bar{\xi})$ *be an optimal solution for* (2.43). *The Expected result of using the EV solution (from now on, EEV) is defined as*

$$EEV = \mathbb{E}_\xi[\varphi(x^*(\bar{\xi}), \xi)] = \int_\Omega \varphi(x^*(\bar{\xi}(\omega)), \xi(\omega)) d\mathbb{P}(\omega). \tag{2.44}$$

Definition 2.31 (Value of Stochastic Solution) *The Value of Stochastic Solution (from now on, VSS) is defined as*

$$VSS = EEV - STO.$$

The VSS measures the gain of solving STO despite the difficulties.

The relationship between all of the above indicators, can be summarized in the following propositions.

Proposition 2.32 *For any stochastic program, we have that*

$$WS \leq STO \leq EEV,$$

and therefore,

$$EVPI \geq 0, \;\; VSS \geq 0.$$

Proposition 2.33 *For stochastic programs with fixed recourse and fixed objective coefficients, we have that*

$$EV \leq WS,$$

and therefore,

$$EVPI \leq EEV - EV, \;\; VSS \leq EEV - EV.$$

Chapter 3

Allocation Problems in Ride-Hailing Platforms

Ride-hailing consists on a form of transportation service delivered using platforms, usually through smart-phone applications, where riders connect with drivers. Unlike ride-sharing, the vehicle used in ride-hailing is not necessarily shared among multiple riders for each trip.

The growing popularity of ride-hailing companies, such as Uber and Lyft, has changed the way we move around the city. There is a new relation between passengers and drivers, which now interact through this new third party. Several new problems have arisen from this context, such as spatio-temporal pricing [20], reallocation of resources [7, 52], or online matching [41] (see, e.g., [125, 16] for some recent surveys). Here, we are interested in the way information affects the relation between a ride-hailing company and its drivers.

To understand this relation, let us describe the general framework we are set in, which is motivated by recent literature [125, 16, 20]. First, a city can be understood as a network of interconnected locations, to which drivers are allocated. At every given time, new passengers appear in the locations, requesting a ride. The ride-hailing company then matches each passenger with a driver in the same location, and receives a compensation proportional to the cost of the ride. While the compensation can be assumed to be constant, the company has the liberty to adapt prices, generating different fares depending on the location and time.

Of course, pricing affects the demand. But more interesting for us, spatial pricing (different fares between locations) can induce reallocation of unmatched drivers. Indeed, a particularity of the Ride-sharing companies is that they do not employ drivers, but rather they consider drivers as independent operators using the matching service. Thus, drivers are free to reallocate themselves whenever they consider it convenient.

Some key elements for an unmatched driver to decide whether to change location or not are the following: the available ride fares, the costs of reallocation, the number of demanded rides at each location, and the number of previously matched drivers at each location (who can be matched to other passengers as soon as they finish their previous rides). The first two are known information for the drivers, but the demand which has random variations (exogenous uncertainty) and the previously matched drivers (endogenous uncertainty) are not. On the other hand, at each stage of reallocation and matching, the ride-hailing company can forecast the exogenous uncertainty, and has all the information available for the endogenous one.

The goal of this chapter is to assess whether the ride-hailing company might benefit from sharing its information with the unmatched drivers. This behavior is observed nowadays, where ride-hailing companies provide some demand information to the drivers beyond spatial pricing (see, e.g., the Uber's driver-app description in [120]).

The notion of sharing information has been studied before, for example in the context of supply chains [107, 74], network restoration [112] and pricing problems [114], where different parties on each context, can be benefited from sharing information between them. Here, we propose to study the problem of value of information through the lens of Stackelberg games (see, e.g., [32, 34]).

On one hand, the company acts as the leader, deciding the spatial prices. On the other hand, the drivers act as followers, solving a stochastic allocation equilibrium problem. By considering the demand to be uncertain for both, the company and the drivers, we derive a model where all agents must decide their actions in a *here-and-now* fashion, that is, prior to the revelation of the nature. In this context, to evaluate the value of perfect information, we consider two indicators: the classic *Expected Value of Perfect Information (EVPI)*, defined for example in [21, Chapter 4], which measures the impact of forecasting on the leader's benefits; and a new indicator that we call the *Expected Value of Shared Information (EVSI)*, which measures the impact, again for the leader, of forecasting and then sharing the perfect information with the drivers.

3.1 Information indicators for general stochastic bilevel problems

In the following, we present the abstract model that we will use in the sequel to describe the interaction between a ride-hailing company and its drivers. The setting fits into the general framework of stochastic bilevel optimization, but with the particular property that randomness is revealed after the followers' decision, differing from the usual concept of the problem (see, e.g., [27]).

Here-and-now Stochastic bilevel problems

We consider the following optimistic parametric bilevel programming problem

$$\varphi(z_1, z_2) = \begin{cases} \min\limits_{x,y} \ \theta(x, y, z_1, z_2) \\ \text{s.t.} \ \begin{cases} x \in X \\ y \text{ solves } \begin{cases} \min\limits_{y} f(x, y, z_1, z_2) \\ \text{s.t. } y \in Y(x), \end{cases} \end{cases} \end{cases} \tag{3.1}$$

where $(z_1, z_2) \in Z = Z_1 \times Z_2 \subset \mathbb{R}^{k_1} \times \mathbb{R}^{k_2}$ are the parameters. Here, the function $\varphi : Z \to \mathbb{R}$ is the value function of Problem (3.1). For each pair $(z_1, z_2) \in Z$, the leader aims to minimize the loss function $\theta : \mathbb{R}^n \times \mathbb{R}^m \times \mathbb{R}^{k_1} \times \mathbb{R}^{k_2} \to \mathbb{R}$. He or she only controls the first variable $x \in X \subset \mathbb{R}^n$, which we call the leader's decision. The set of admissible leader's decisions $X \subset \mathbb{R}^n$ is fixed.

Similarly, for each pair $(z_1, z_2) \in Z$ and each leader's decision $x \in X$, the follower aims to minimize the loss function $f : \mathbb{R}^n \times \mathbb{R}^m \times \mathbb{R}^{k_1} \times \mathbb{R}^{k_2} \to \mathbb{R}$. He or she only controls the second variable $y \in Y(x) \subset \mathbb{R}^m$, which we call the follower's decision. The set of admissible decisions $Y(x) \subset \mathbb{R}^m$ depends on the leader's decision x, inducing a set-valued map $Y : X \rightrightarrows \mathbb{R}^m$. The range of Y is contained in an ambient set $\bar{Y} \subset \mathbb{R}^m$, that is,

$$R(Y) := \bigcup_{x \in X} Y(x) \subset \bar{Y}.$$

In what follows, we consider the following (standard) assumptions over Problem (3.1):

(A.1) The sets X and $\bar{Y}$ are nonempty, convex and compact, and Z is nonempty and closed.

(A.2) The loss functions θ and f are continuous.

(A.3) The loss function f is convex in the second variable.

(A.4) The set-valued map Y has nonempty convex compact values, and it is both upper and lower semicontinuous.

Under this framework, which is fairly general, one can ensure the existence of solutions of the parametric Problem (3.1). This existence result is classic in the literature (see, e.g., [32]), and we recall it for completeness.

Lemma 3.1 *Assume hypotheses **(A.1)-(A.4)**. Then, for each $(z_1, z_2) \in Z$, Problem* (3.1) *admits at least one solution. Furthermore, the value function $\varphi : Z \to \mathbb{R}$ is lower semicontinuous.*

Proof. To simplify notation, let us write $z = (z_1, z_2) \in Z$. Let us denote $S : X \times Z \rightrightarrows \bar{Y}$ given by

$$S(x, z) = \arg\min_{y \in Y(x)} f(x, y, z)$$

Using once more the Maximum Principle (Theorem 1.11), we conclude that S is upper-semicontinuous, and given that f is continuous, it is also closed-valued. Now, by Theorem 1.9, we obtain that $\mathrm{gph}(S)$ is closed.

Let us define $K : Z \rightrightarrows X \times \bar{Y}$ given by

$$K(z) = \{(x, y) \in X \times \bar{Y} \ : \ (x, y, z) \in \mathrm{gph}(S)\}.$$

Noting that $K(z)$ is compact for every z, since it is closed and a subset of $X \times \bar{Y}$, and that Problem (3.1) can be written as

$$\varphi(z) = \min_{x,y} \{\theta(x, y, z) \ : \ (x, y) \in K(z)\},$$

we deduce from Weierstrass theorem that Problem (3.1) has a solution for every fixed $z \in Z$.

Since $\mathrm{gph}(K) = \mathrm{gph}(S)$ and K is compact-valued, Theorem 1.9 entails that K is upper-semicontinuous. Moreover, since θ is continuous, we can apply Theorem 1.12 to conclude that $z \mapsto \sup_{(x,y)\in K(z)} -\theta(x, y, z)$ is upper semicontinuous. The result follows by noting that

$$\varphi(z) = \inf_{(x,y)\in K(z)} \theta(x, y, z) = - \sup_{(x,y)\in K(z)} -\theta(x, y, z),$$

which is therefore lower semicontinuous. □

In our setting, uncertainty is formalized as random variables $\zeta = (\zeta_1, \zeta_2) : \Omega \to Z_1 \times Z_2$ and $\xi = (\xi_1, \xi_2) : \Omega \to Z_1 \times Z_2$, over a probability space $(\Omega, \Sigma, \mathbb{P})$. These variables determine the parameters (z_1, z_2) in Problem (3.1). Specifically,

- The leader knows the first parameter z_1, which is the realization of a random variable ζ_1.
- The variable ζ_2 determines the value of z_2 for the leader, and its value is unknown for the leader.
- Both parameters (z_1, z_2) are unknown for the follower, and are given by the variables ξ_1 and ξ_2, respectively.

We allow the leader and the follower to have different beliefs about z_1 and z_2, and thus the distribution of $\zeta = (\zeta_1, \zeta_2)$ might differ from the distribution of $\xi = (\xi_1, \xi_2)$.

With this model, the follower solves a *here-and-now* problem, considering the leader's decision x as a parameter and the value of $z = (\xi_1(\omega), \xi_2(\omega))$ as uncertain. Thus, for each leader's decision x, the follower is solving the problem

$$D(x) = \begin{cases} \min_{y} \mathbb{E}\left[f\left(x, y, \xi_1(\omega), \xi_2(\omega)\right)\right] \\ \text{s.t. } \; y \in Y\left(x\right). \end{cases} \tag{3.2}$$

We refer by $D(x)$ to both Problem (3.2) and its solution set. From the leader's perspective, for each decision vector x, the follower's optimal response is a deterministic point $y \in D(x)$. Thus, the leader is solving a *parametric here-and-now* problem, simultaneously with the follower, which is given by:

$$STO(z_1) := \begin{cases} \min_{x,y} \; \mathbb{E}\left[\theta\left(x, y, z_1, \zeta(\omega)\right)\right] \\ \text{s.t.} \; \begin{cases} x \in X \\ y \in D(x). \end{cases} \end{cases} \tag{3.3}$$

We set the Stochastic value of Problem (3.3) as

$$STO = \mathbb{E}[STO(\zeta_1)] = \int_{\Omega} STO(\zeta_1(\omega)) d\mathbb{P}(\omega), \tag{3.4}$$

which is given by the averaged cost for the leader, according to the distribution of ζ_1, for an optimal policy $x : Z_1 \to \mathbb{R}^n$. In what follows, we will assume that all expectations are well-defined and finite.

Measuring the value of information: Expected Value of Shared Information

In stochastic bilevel optimization, Problem (3.3) has not got too much attention, since its formulation fits into the setting of stochastic optimization with recourse (see, e.g., [21, 111]). This reduction follows from the fact that the random parameter z_2 is revealed after the follower's decision, leading to a sequential structure of the form

$$\text{Leader decides } x \rightarrow \text{Follower decides } y \rightarrow (z_1, z_2) \text{ is revealed.}$$

In contrast, stochastic bilevel optimization, as understood in the literature (see, e.g., [27]), focuses on bilevel programs where the random parameter is revealed before the follower's decision, inducing a *here-and-now* problem for the leader, and a *wait-and-see* problem for the follower, that is,

$$\text{Leader decides } x \rightarrow (z_1, z_2) \text{ is revealed} \rightarrow \text{Follower decides } y.$$

In this second case, when one aims to assess the expected value of perfect information (EVPI), it is enough to do it for the leader only. Indeed, the follower already has the perfect information. Thus, one can follow the standard developments of, e.g., [21, Chapter 4] or [111, Chapter 2].

Similarly, in our context, we propose to set the EVPI as a measure for the leader only. The proposal is motivated by the interpretation of a bilevel problem as a Stackelberg game: the bilevel optimization problem is the problem of the leader only, who is able to anticipate the follower's best response [57].

Definition 1 (EVPI) For Problem (3.3), we define the *Wait-and-See Value (WS)* as the expected value of the value function

$$\psi(z_1, z_2) := \begin{cases} \min\limits_{x,y} & \theta\,(x, y, z_1, z_2) \\ \text{s.t.} & \begin{cases} x \in X \\ y \in D(x), \end{cases} \end{cases} \tag{3.5}$$

that is,

$$WS := \mathbb{E}_\zeta(\psi) = \int_\Omega \psi(\zeta(\omega))d\mathbb{P}(\omega). \tag{3.6}$$

The *Expected Value of Perfect Information (EVPI)* is then defined as

$$EVPI := STO - WS,$$

which is always non-negative, and measures the gain of the leader under perfect information.

While the above definition is consistent with the interpretation of Problem (3.3) as a parametric problem with recourse, the leader has another option in Stackelberg games under perfect information: he or she might share this information with the follower. This alternative comes from the fact that the follower is an independent agent, which reacts to new information. In order to measure the value of sharing perfect information, we introduce the following definition:

Definition 2 (EVSI) For Problem (3.3), we define the *Shared Wait-and-See Value (SWS)* as the expected value of the value function φ from the parametric problem (3.1) that is,

$$SWS := \mathbb{E}_\zeta(\varphi) = \int_\Omega \varphi(\zeta_1(\omega), \zeta_2(\omega))d\mathbb{P}(\omega), \tag{3.7}$$

The *Expected Value of Shared Information (EVSI)* is then defined as

$$EVSI := WS - SWS,$$

which measures the gain or loss of the leader having perfect information and sharing it with the follower.

When evaluating the value of perfect information for the leader, both WS and SWS should be computed, since SWS is, in general, neither a lower bound nor an upper bound of WS nor STO, as the following example shows.

Example Let us consider only an exogenous random event ξ to be a fair Bernoulli trial and the indicator functions δ_0 and δ_1 given by

$$\delta_i(\xi) = \begin{cases} 1 & \text{if } \xi = i \\ 0 & \text{otherwise,} \end{cases} \qquad \text{for } i = 0, 1.$$

Let the leader's decision set to be $X = [0, 1]$ and the follower's decision set to be $Y(x) = [0, 1]$, for all $x \in X$. Let the follower's loss function to be

$$f(y, \xi(\omega)) := y^2\delta_0(\xi(\omega)) + (1 - y)^2\delta_1(\xi(\omega)).$$

Assume that $\zeta = \xi$ and consider two possible loss functions for the leader:

$$\begin{aligned} \theta_+(x, y, \xi(\omega)) &= \frac{1}{2}\big(x^2\delta_0(\xi(\omega)) + (1 - x)^2\delta_1(\xi(\omega))\big) + f(x, y, \xi(\omega)) \\ \theta_-(x, y, \xi(\omega)) &= \frac{1}{2}\big(x^2\delta_0(\xi(\omega)) - (1 - x)^2\delta_1(\xi(\omega))\big) - f(x, y, \xi(\omega)) \end{aligned}$$

We consider then two problems of the form of Problem (3.3):

$\displaystyle \min_{x,y} \ \mathbb{E}\left(\theta_+(x, y, \xi(\omega))\right)$ s.t. $\begin{cases} x \in [0, 1] \\ y \text{ solves } \begin{cases} \min_y \mathbb{E}\left(f\left(y, \xi(\omega)\right)\right) \\ \text{s.t. } \ y \in [0, 1]. \end{cases} \end{cases}$	$\displaystyle \min_{x,y} \ \mathbb{E}\left(\theta_-(x, y, \xi(\omega))\right)$ s.t. $\begin{cases} x \in [0, 1] \\ y \text{ solves } \begin{cases} \min_y \mathbb{E}\left(f\left(y, \xi(\omega)\right)\right) \\ \text{s.t. } \ y \in [0, 1]. \end{cases} \end{cases}$
Plus Case	Minus Case

It is easy to see that the optimal solutions for the leader and the follower are very similar. Observe too that the decision of the leader has no influence on the follower's decision, but the objective of the follower is directly included in the leader's objective. The set of optimal decisions is easy to determine. In the stochastic case, where no information is given to either of the players, we have $x^* = y^* = 0.5$. On the other hand, if the players have access to the perfect information, we have $x_0^* = y_0^* = 0$ for $\xi = 0$, and $x_1^* = y_1^* = 1$ for $\xi = 1$. Thus we can compute the values of STO, WS (as in [21, Chapter 4]) and SWS, displayed in Table 3.1. $\diamondsuit$

	Plus Case	Minus Case
STO	0.375	-0.125
WS	0.25	-0.25
SWS	0	0

Table 3.1: Values of STO, WS and SWS

The above example is quite simple but very illustrative. First, for the Minus Case it shows that the SWS is not necessarily a lower bound of WS, nor even of STO. However, for the Plus Case it shows that sharing information could be beneficial, beyond having perfect information. The intuition behind this example is simple. On the one hand, the Plus Case is collaborative: the leader wants to collaborate with the follower, since the leader is losing what the follower loses as well.

On the other hand, the Minus Case is adversarial: the leader is against the follower, since what the follower loses translates in gains for the leader. However, in practical situations, the collaboration or competition between the leader and the follower might not be so clear.

3.2 The Ride-Hailing Bilevel Model

Let us consider the following situation: at a certain moment, a driver associated with a ride-hailing company that has not been matched with a passenger must decide whether to keep searching for a match around his or her current location, or to move to another one within the city. We can model the different locations as a finite set of zones, $I = \{1, \dots, n\}$, connected as a directed graph.

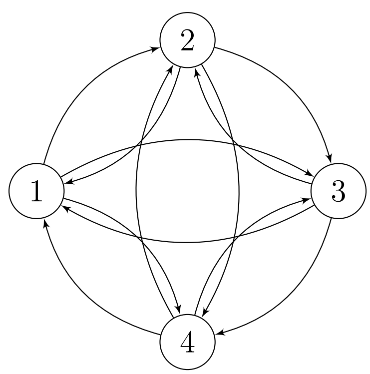

Figure 3.1: Directed Graph Modeled for $n = 4$ zones.

If the driver is in the ith zone, his or her reallocation decision will depend on five factors:

1. The vector of marginal prices fixed by ride-hailing company, $p = (p_i \ : \ i \in I)$.

2. The vector of previously matched drivers who will arrive to each node (and will become available at that node), $y = (y_i \ : \ i \in I)$.

3. The vector of demands of each zone $d = (d_i \ : \ i \in I)$.

4. The marginal costs of moving to another zone, $\alpha_i = (\alpha_{ij} \ : \ j \in I)$. Of course, $\alpha_{ii} = 0$.

5. The vector of previously unmatched drivers that will be at each node, $x = (x_i \ : \ i \in I)$.

The demand on each zone $i \in I$, depends on two factors: the marginal price p_i and a random variable which models the uncertain variation. In this work, we model this dependency as a nominal value $d_{0,i}(\omega)$ which represents the demand for the minimal price $p_{i,\min}$, multiplied by a linear discount factor depending on the price:

$$d_i = d_i(p_i, \omega) = d_{0,i}(\omega)\left(1 - \delta \frac{p_i - p_{i,\min}}{p_{i,\max} - p_{i,\min}}\right) \tag{3.8}$$

Since each driver has limited observability about the other drivers, the value of $y = (y_i \ : \ i \in I)$ is uncertain, even though it is known information for the ride-hailing company. Thus, from the drivers' perspective $y = y(\omega)$ is also a random variable. Then, by setting $\xi(\omega) = (y(\omega), d_0(\omega))$, each driver must solve the following optimization problem:

$$\max_{j \in I} \mathbb{E}_\xi \left[p_j \min\left(\frac{d_j}{x_j + y_j}, 1\right)\right] - \alpha_{ij}, \tag{3.9}$$

where $x_j + y_j$ is the amount of available drivers in zone j, and the value $\min\left(\frac{d_j}{x_j+y_j}, 1\right)$ represents the probability of being matched in zone j: if $d_j \geq x_j + y_j$, then the driver will be matched. On the other hand, if $d_j < x_j + y_j$, the probability of being matched coincides with $d_j/(x_j + y_j)$, assuming that in such a case, all passengers will be matched.

We will model the situation where drivers can communicate between them outside the ride-hailing platform, and they can coordinate their allocation. Thus, we model all unmatched drivers as a single new follower, who aims to maximize the social welfare of all drivers. Such situation has been recently studied in [118]. We will assume that only unmatched drivers report to this central decision-maker, while matched drivers become unavailable. Thus, the follower must decide the allocation of unmatched drivers $x = (x_i \ : \ i \in I)$ while the vector $y = (y_i \ : \ i \in I)$ is uncertain.

Now, to model the decision process of the single follower, let us assume that there is an amount of N_0 drivers unmatched, with initial allocation $x_0 = (x_{0,1}, \ldots, x_{0,n}) \in \mathbb{R}^n$. Let us define the variable v_{ij} as the amount of unmatched drivers who will change from zone i to zone j, and let v be the matrix that collects all this information. In this context, $v_{ii} = 0$ for all $i \in I$. Then, we can compute a reallocation x in terms of the displacement matrix v simply as:

$$x_j(v) = x_{0,j} + \sum_{i \neq j} v_{ij} - \sum_{k \neq j} v_{jk}, \quad \forall j \in I. \tag{3.10}$$

Then, for a given price vector p, the aggregated allocation problem is posed as follows:

$$F(p) := \begin{cases} \max_v & \sum_{i=1}^n c p_i \mathbb{E}_\xi\left[\min(x_i + y_i, d_i)\right] - \sum_{i \neq j} \alpha_{ij} v_{ij} \\ s.t & \begin{cases} v \geq 0 \\ \sum_{k \neq j} v_{jk} \leq x_{0,j}, \quad \forall j \in I \end{cases} \end{cases},$$

where $c \in (0, 1)$ is the fraction of the ride price that the driver gets.

The distribution of ξ, which models the belief over y and d_0, must reflect the fact that drivers have not access to the data of the ride-hailing company. On the one hand, we model the vector of previously matched drivers who will arrive to each node as an uniformly distributed random variable, that is, $y_j \sim U(0, \bar{y})$ where $\bar{y}$ is a constant value for all nodes. This distribution represents the lack of information for the unmatched drivers about the matched ones. On the other hand, we assume that drivers perceive the distribution of the nominal demand as a discrete one, considering $m \in \mathbb{N}$ feasible scenarios. Thus, we can write

$$\begin{aligned} \mathbb{E}_\xi\left[p_i \min(x_i + y_i, d_i)\right] &= -\mathbb{E}_\xi(p_i \max(-x_i - y_i, -d_i)) \\ &= -\sum_{k=1}^{m} p_i \mathbb{E}_y[\max(-x_i - y_i, -d_{i,k})] \cdot \mathbb{P}(\omega_k), \end{aligned} \tag{3.11}$$

where we define the discrete expression $d_{i,k} = d_i(\omega_k)$. Therefore, the follower will deal with a discrete version of its original problem $F(p)$, given by

$$F_m(p) := \begin{cases} \min_v & c\sum_{i=1}^{n}\sum_{k=1}^{m} p_i \mathbb{E}_y[\max(-x_i - y_i, -d_{i,k})] \cdot \mathbb{P}(\omega_k) + \sum_{i \neq j} \alpha_{ij} v_{ij} \\ s.t & \begin{cases} -v \leq 0 \\ \sum_{j \neq i} v_{ij} - x_{0,i} \leq 0, \quad \forall i \in I \end{cases} \end{cases}. \tag{3.12}$$

Now, the ride-hailing company must decide the price vector p. The company does not necessarily know the exact value of the demand vector d, but it knows the vector y of occupied drivers. Since the company aims to maximize its revenues, it must solve the following bilevel programming problem:

$$L(y) := \begin{cases} \max_{p,v} & \sum_{i=1}^{n} (1-c) p_i \cdot \mathbb{E}_\zeta[\min(x_i + y_i, d_i)] \\ s.t & \begin{cases} p_i \in [p_{i,\min}, p_{i,\max}], \quad \forall i \in I \\ v \text{ solves } F_m(p). \end{cases} \end{cases} \tag{3.13}$$

Here ζ is the random variable that models the belief of the leader about the behavior of the nominal demand. The distribution of ζ should be multivariate normal-like distribution around a nominal value $\bar{d}_0 = (\bar{d}_{0,i} \ : \ i \in I)$.

Observe that Problem (3.13) fits the setting of Problem (3.3), by considering $z_1 = y$ as the endogenous uncertainty known by the leader, and $z_2 = d_0$ as the exogenous uncertainty. Since ride-hailing companies already invest in demand forecasting (see, e.g, [119, 81]), we want to know what the policy should be about this information: should the perfect information be kept private or should it be shared with the drivers?

3.3 Reformulation to Single Bilinear Optimization

In this section, we will focus in how to compute WS and SWS for Problem (3.13), and then obtaining the EVSI. Here, we assume that the leader already has perfect information, and thus, the values of $\zeta = (y, d_0)$ are known for it. In what follows, we identify the scenario set $\{\omega_1, \ldots, \omega_m\}$ with the set of indexes $K = \{1, \ldots, m\}$.

In order to compute the EVSI, our approach is to follow a Monte-Carlo estimation: first, we consider a sampling $(y^1, d_0^1), \ldots, (y^T, d_0^T)$ of the random parameters (y, d_0), accordingly to the leader's distributions. Then, to compute the empirical expectations of the value functions ψ given by (3.5) for the WS, and φ given by (3.1) for the SWS. Finally, we compute the EVSI as

$$EVSI = \frac{1}{T}\sum_{t=1}^{T} \varphi(y^t, d_0^t) - \frac{1}{T}\sum_{t=1}^{T} \psi(y^t, d_0^t). \tag{3.14}$$

Thus, our problem is reduced to compute the value functions $\psi(y^t, d_0^t)$ and $\varphi(y^t, d_0^t)$ for each sample (y^t, d_0^t). Our technique is, for both values, to reformulate the corresponding bilevel programming problems into single level bilinear problems. In both cases, we replace the corresponding follower's problem by its Karush-Kuhn-Tucker (KKT) conditions and consider the associated multipliers as new variables. This approach, known as the Mathematical Programming with Complementarity Constraints (MPCC) reformulation, is quite popular in the literature and can be applied whenever the follower's problem satisfies a constraint qualification (see, e.g., [34, Chapter 3] and the references therein).

In this section, we show that the reformulations we obtain through this technique have two main properties: firstly, they preserve global solutions in the sense that a pair (x, y) of leader-follower decision variables is a global solution of a bilevel program if and only if there exists a multiplier u such that (x, y, u) is a global solution of the MPCC reformulation; and secondly, the MPCC reformulations can be rewritten as mixed-integer bilinear problems.

3.3.1 Constraint Qualifications of the lower-level

Before studying the single-level reformulations for WS and SWS, we will study the regularity properties of the feasible set of Problem $F_m(p)$, defined in (3.12). To do so, we consider the following notation and definitions. First off, we identify $\mathbb{R}^{n(n-1)}$, which is the space of decision variables of the follower, with the subspace V of $n \times n$ square matrices with 0-entries in the diagonal. For every $i, j \in \{1, \ldots, n\}$ with $i \neq j$, we denote by e_{ij} as the $n \times n$ matrix given by

$$e_{ij}(a, b) = \begin{cases} 1 & \text{if } a = i \text{ and } b = j, \\ 0 & \text{otherwise.} \end{cases} \tag{3.15}$$

and we denote by $e_{i\bullet} = \sum_{j\,:\,j\neq i} e_{ij}$, which is the $n \times n$ matrix with 1-entries in the ith row (except for the entry (i, i)), and 0 otherwise. Similarly, we set $e_{\bullet j} = \sum_{i\,:\,i\neq j} e_{ij}$ which is the $n \times n$ matrix with 1-entries in the jth column (except for the entry (j, j)), and 0 otherwise. We now state the following key lemma.

Lemma 3.2 *Assume that the initial allocation vector x_0 is strictly positive (i.e. $x_{0,i} > 0$ for each $i \in \{1, \ldots, n\}$). Then, the feasible set of the followers' problem, which is given by*

$$\left\{ v \in V \ : \begin{array}{c} v \geq 0 \\ \sum_{j\neq i} v_{ij} \leq x_{0,i}, \quad \forall i \in I \end{array} \right\}$$

satisfies Slater's CQ, and it satisfies (LICQ) an every point.

Proof. First, we define $x_{0,\min} = \min\{x_{0,1}, \ldots, x_{0,n}\}$. We claim that

$$v_{ij} = \frac{x_{0,\min}}{2(n-1)}$$

is a Slater point. In fact, it is clear that $v_{ij} > 0$ for every $i \neq j$, and furthermore,

$$\sum_{j \neq i} v_{ij} = \sum_{j \neq i} \frac{x_{0,\min}}{2(n-1)} = (n-1)\frac{x_{0,\min}}{2(n-1)} = \frac{x_{0,\min}}{2} < x_{0,i}, \quad \forall i \in I.$$

Thus, the claim is verified and this finishes the first part of the proof. Now, let us show that (LICQ) is verified at every point. For every $i \in I$, and every $j \neq i$, let $g_{ij}(v) = -v_{ij}$ and let $h_i(v) = \sum_{j \neq i} v_{ij} - x_{0,i}$. Then, this set can be written as

$$\left\{ v \in V \ : \ \begin{array}{ll} g_{ij}(v) \leq 0, & \forall i \neq j \\ h_i(v) \leq 0, & \forall i \in I \end{array} \right\}$$

Now, suppose that there exists v^* in this set, not satisfying (LICQ). It is not hard to see that $\nabla g_{ij}(v^*) = -e_{ij}$ and $\nabla h_i(v^*) = e_{i\bullet}$. Thus, since $\{\nabla g_{ij}(v^*) \ : \ i \neq j\}$ is linearly independent, there must be $i_0 \in I$ such that $h_{i_0}(v^*) = 0$, and such that $\nabla h_{i_0}(v^*) = e_{i_0\bullet}$ is a linear combination of the gradients of the other active constraints. However, this is only possible if $g_{i_0 j}$ is active at v^* for every $j \neq i_0$, which would mean that

$$\sum_{j \neq i_0} v_{i_0 j} = x_{0,i_0} \quad \text{and} \quad v_{i_0 j} = 0, \quad \forall j \neq i_0,$$

which is a contradiction since $x_{0,i_0} \neq 0$. This finishes the proof. □

3.3.2 Reformulation of Wait-and-See

Recall that we want to solve Problem (3.5), which in this context is given by

$$\psi(y, d_0) := \begin{cases} \min\limits_{p,v} \ \sum\limits_{i=1}^{n} (1-c)p_i \cdot \mathbb{E}_\zeta[\max(-x_i - y_i, -d_i)] \\ \text{s.t.} \ \begin{cases} p_i \in [p_{i,\min}, p_{i,\max}], & \forall i \in I \\ v \text{ solves } F_m(p). \end{cases} \end{cases} \tag{3.16}$$

Based on (3.11), for each scenario $k \in K$ and each location $i \in I$, we set (recalling that $y_i \sim U(0, \bar{y})$) a function $\phi_{i,k}$ as follows:

$$\begin{aligned} \phi_{i,k}(x_i) &= \mathbb{E}_{y_i}\left[\max(d_{i,k} - x_i - y_i, 0)\right] \\ &= \frac{1}{\bar{y}} \int_0^{\bar{y}} \max(d_{i,k} - x_i - y_i, 0) dy \\ &= \begin{cases} 0 & \text{if } d_{i,k} - x_i \leq 0 \\ \frac{(d_{i,k} - x_i)^2}{2\bar{y}} & \text{if } 0 \leq d_{i,k} - x_i \leq \bar{y} \\ (d_{i,k} - x_i) - \frac{\bar{y}}{2} & \text{if } d_{i,k} - x_i \geq \bar{y} \end{cases} \end{aligned} \tag{3.17}$$

Then, it is not hard to see that the follower's problem $F_m(p)$ can be written as

$$F_m(p) := \begin{cases} \min_v & c\sum_{i=1}^{n}\sum_{k=1}^{m} p_i(\phi_{k,i}(x_i) - d_{i,k}) \cdot \mathbb{P}(\omega_k) + \sum_{i\neq j} \alpha_{ij}v_{ij} \\ s.t & \begin{cases} -v \leq 0 \\ \sum_{j\neq i} v_{ij} - x_{0i} \leq 0, \quad \forall i \in I. \end{cases} \end{cases} \tag{3.18}$$

Then, we can state the following proposition.

Theorem 3.3 *For any given value of the random vector $\zeta = (y, d_0)$, the Wait-and-See problem associated to the leader's problem (3.13) is equivalent (in the sense of local and global solutions) to its MPCC reformulation given by*

$$\max_{p,v,\lambda,\gamma} \quad \sum_{i=1}^{n}(1-c)p_i \cdot \min(x_i + y_i, d_i)$$
$$s.t \begin{cases} p_i \in [p_{i,\min}, p_{i,\max}], \quad \forall i \in I \\ \sum_{j\neq i} v_{ij} - x_{0i} \leq 0, \quad \forall i \in I \\ \sum_{k=1}^{m}(p_i\beta_{i,k} - p_j\beta_{j,k}) + \alpha_{ij} - \lambda_{ij} + \gamma_i = 0, \quad \forall i \neq j \in I \\ \gamma_i(\sum_{j\neq i} v_{ij} - x_{0i}) = 0, \quad \forall i \in I \\ \lambda_{ij}v_{ij} = 0, \quad \forall i \neq j \in I \\ v \geq 0, \gamma, \lambda \geq 0, \end{cases} \tag{3.19}$$

where the coefficients $\{\beta_{i,k} \ : \ i \in I, k \in K\}$ are given by

$$\beta_{i,k} := \begin{cases} 0 & \textit{if } d_{i,k} - x_i \leq 0 \\ c\mathbb{P}(\omega_k)\frac{x_i - d_{i,k}}{\bar{y}} & \textit{if } 0 \leq d_{i,k} - x_i \leq \bar{y} \\ -c\mathbb{P}(\omega_k) & \textit{if } d_{i,k} - x_i \geq \bar{y}. \end{cases} \tag{3.20}$$

Furthermore, the multipliers $\gamma = (\gamma_i \ : \ i \in I)$ and $\lambda = (\lambda_{ij} \ : \ i \neq j \in I)$ verify

$$0 \leq \gamma_i \leq 2mp_{\max} \quad \textit{and} \quad 0 \leq \lambda_{ij} \leq 4mp_{\max}, \tag{3.21}$$

where $p_{\max} = \max_{i\in I}\{p_{i,\max}\}$.

Proof. The equivalence between the bilevel problem (3.13) and the MPCC reformulation follows from Lemma 3.2. Indeed, since Slater's CQ is verified, and LICQ implies Constant Rank CQ, the desired conclusion follows from [33, Theorem 3.2 and Corollary 3.3]. Thus, it is enough to show that problem (3.19) coincides with the MPCC reformulation (3.13).

Based on (3.17) we can compute the partial derivatives of $\phi_{i,k}$ as

$$\frac{\partial}{\partial x_i}\phi_{i,k} = \begin{cases} 0 & \text{if } d_{i,k} - x_i \leq 0 \\ \frac{x_i - d_i(\omega_k)}{\bar{y}} & \text{if } 0 \leq d_{i,k} - x_i \leq \bar{y} \\ -1 & \text{if } d_{i,k} - x_i \geq \bar{y} \end{cases}$$

Let $f(v)$ be the objective function for $F_m(p)$. By a mild application of the chain rule, considering the definition of x_r in Equation (3.10) we can compute its partial derivative as

$$\begin{aligned}
\frac{\partial f}{\partial v_{ij}}(v) &= \frac{\partial}{\partial v_{ij}}\left(c\sum_{r=1}^{n}\sum_{k=1}^{m} p_r\phi_{r,k}(x_r)\mathbb{P}(\omega_k) - c\sum_{r=1}^{n}\sum_{k=1}^{m} p_r d_r(\omega_k)\mathbb{P}(\omega_k) + \sum_{r\neq s}\alpha_{rs}v_{rs}\right)\\
&= \frac{\partial}{\partial v_{ij}}\left(c\sum_{k=1}^{m} p_i\phi_{i,k}(x_i)\mathbb{P}(\omega_k) + c\sum_{k=1}^{m} p_j\phi_{j,k}(x_j)\mathbb{P}(\omega_k) + \alpha_{ij}v_{ij}\right)\\
&= c\sum_{k=1}^{m}(p_i\frac{\partial}{\partial x_i}\phi_{i,k}(x_i)\mathbb{P}(\omega_k) - p_j\frac{\partial}{\partial x_j}\phi_{j,k}(x_j)\mathbb{P}(\omega_k)) + \alpha_{ij}\\
&= \sum_{k=1}^{m}(p_i\beta_{i,k} - p_j\beta_{j,k}) + \alpha_{ij},
\end{aligned}$$

where the coefficients $\{\beta_{i,k} \ : \ i \in I, k \in J\}$ are defined as in (3.20). Hence, the KKT equations for $F_m(p)$ have the form

$$\sum_{k=1}^{m}(p_i\beta_{i,k} - p_j\beta_{j,k}) + \alpha_{ij} - \lambda_{ij} + \gamma_i = 0, \quad \forall i,j \in \{1,\dots,n\},\ i\neq j$$

with the complementarity constraints

$$\begin{aligned}
\lambda_{ij}v_{ij} = 0, \quad \forall i,j \in \{1,\dots,n\},\ i\neq j.\\
\gamma_i\left(\sum_{j\neq i} v_{ij} - x_{0i}\right) = 0, \quad \forall i \in \{1,\dots,n\}.
\end{aligned} \tag{3.22}$$

Putting all together, we get that the MPCC reformulation of (3.13) is given by (3.19).

Lastly, we compute the multiplier bounds. Fix a feasible price vector p and let v^* be an optimal point of $F_m(p)$. Let (λ, γ, β) be a feasible tuple of multipliers for problem (3.19). Then, for the ith coordinate, we have that

$$\sum_{k=1}^{m}(p_i\beta_{i,k} - p_j\beta_{j,k}) + \alpha_{ij} - \lambda_{ij} + \gamma_i = 0, \quad \forall j \neq i.$$

We have two possible scenarios:

- If $\gamma_i = 0$, then

$$\lambda_{ij} = \sum_{k=1}^{m}(\beta_{i,k} - \beta_{j,k}) + \alpha_{ij} \leq 2mp_{\max}, \quad \forall j \in \{1,\dots,n\}.$$

- If $\gamma_i \neq 0$, then the second complementary equation of (3.22) implies that $x_{0i} = \sum_{j\neq i} v^*_{ij}$. Since $x_{0i} \neq 0$, this also implies that there exists a value of j such that $v^*_{ij} \neq 0$, in which case $\lambda_{ij} = 0$ using the first complementary equation of (3.22). Hence, we conclude that

$$\begin{gathered}
\textstyle\sum_{k=1}^{m}(\beta_{i,k} - \beta_{j,k}) + \alpha_{ij} + \gamma_i = 0\\
\Longrightarrow\\
\gamma_i = \textstyle\sum_{k=1}^{m}(\beta_{j,k} - \beta_{i,k}) - \alpha_{ij} \leq 2mp_{\max},
\end{gathered}$$

and so we can compute

$$\lambda_{ij} = \sum_{k=1}^{m} (\beta_{i,k} - \beta_{j,k}) + \gamma_i + \alpha_{ij} \leq 4mp_{\max}, \quad \forall j \in \{1, \ldots, n\}.$$

Regardless the case, we conclude that

$$\lambda_{ij} \in [0, 4mp_{\max}], \quad \gamma_i \in [0, 2mp_{\max}],$$

finishing our proof. □

This result allows us to compute the Wait-and-See value by sampling $\zeta = (y, d_0)$ and solving (3.19). To do so, we will follow the classic big-M strategy, which seems to be first introduced in the context of bilevel optimization in [44]. Even though computing a sufficiently large M is hard in general [65], the above proposition has already provided the needed bounds in (3.21). Hence, defining $M = 4mp_{\max}$, we proceed as follows:

1. For each pair $i \neq j \in I$, we introduce a boolean variable $z_{ij} \in \{0, 1\}$ and replace the constraint $\lambda_{ij} v_{ij} = 0$ by

$$\begin{aligned} &-Mz_{ij} \leq \lambda_{ij} \leq Mz_{ij} \\ &-M(1 - z_{ij}) \leq v_{ij} \leq M(1 - z_{ij}) \end{aligned} \tag{3.23}$$

2. For each $i \in I$, we introduce a boolean variable $w_i \in \{0, 1\}$ and replace the constraint $\gamma_i(\sum_{j \neq i} v_{ij} - x_{0i}) = 0$ by

$$\begin{aligned} &-Mw_i \leq \gamma_i \leq Mw_i \\ &-M(1 - w_i) \leq \sum_{j \neq i} v_{ij} - x_{0,i} \leq M(1 - w_i) \end{aligned} \tag{3.24}$$

The last numerical consideration involves the additional constraints which are used to tackle the term $\beta_{i,k}$, which are given piecewise linear functions. We first define a constant

$$C \geq \max\{d_{0,i}(\omega_k) \ : \ i \in I, k \in J\} + N_0,$$

which is an upper bound of $|d_i(\omega_k) - x_i|$ for every $i \in I$ and every scenario ω_k. Then, for each term $\beta_{i,k}$, we define three integer variables $a_{i,k}, b_{i,k}, c_{i,k} \in \{0, 1\}$, three continuous variables $r_{i,k}$, $s_{i,k}$, $t_{i,k}$, and we replace (3.20) by the set of constraints

$$\begin{cases} a_{i,k} + b_{i,k} + c_{i,k} = 1 \\ d_{i,k} - x_i = r_{i,k} + s_{i,k} + t_{i,k} \\ -Ca_{i,k} \leq r_{i,k} \leq 0 \\ 0 \leq s_{i,k} \leq \bar{y} b_{i,k} \\ \bar{y} c_{i,k} \leq t_{i,k} \leq Cc_{i,k} \\ \beta_{i,k} = -c\mathbb{P}(\omega_k) \left(\frac{s_{i,k}}{\bar{y}} + c_{i,k} \right) \end{cases} \tag{3.25}$$

The above replacement works as follows:

- If $a_{i,k} = 1$, then $b_{i,k} = c_{i,k} = s_{i,k} = t_{i,k} = 0$. Hence, $d_{i,k} - x_i = r_{i,k} \leq 0$, and $\beta_{i,k} = 0$.
- If $b_{i,k} = 1$, then $a_{i,k} = c_{i,k} = r_{i,k} = t_{i,k} = 0$. Hence, $d_{i,k} - x_i = s_{i,k} \in [0, \bar{y}]$, and $\beta_{i,k} = -c\mathbb{P}(\omega_k)\frac{s_{i,k}}{\bar{y}} = c\mathbb{P}(\omega_k)\frac{(x_i - d_{i,k})}{\bar{y}}$.
- If $c_{i,k} = 1$, then $a_{i,k} = b_{i,k} = r_{i,k} = s_{i,k} = 0$. Hence, $d_{i,k} - x_i = t_{i,k} \geq \bar{y}$, and $\beta_{i,k} = -c\mathbb{P}(\omega_k)$.

The final problem we solve for each $\zeta = (y, d_0)$ is then given by

$$\begin{aligned} \max \quad & \sum_{i=1}^{n} (1-c)p_i \cdot \min(x_i + y_i, d_i) \\ s.t. \quad & \begin{cases} p_i \in [p_{i,\min}, p_{i,\max}], \quad \forall i \in I \\ \sum_{k=1}^{m} (p_i \beta_{i,k} - p_j \beta_{j,k}) + \alpha_{ij} - \lambda_{ij} + \gamma_i = 0, \quad \forall i \neq j \in I, \\ -M z_{ij} \leq \lambda_{ij} \leq M z_{ij} \\ -M(1 - z_{ij}) \leq v_{ij} \leq M(1 - z_{ij}) \\ -M w_i \leq \gamma_i \leq M w_i \\ -M(1 - w_i) \leq \sum_{j \neq i} v_{ij} - x_{0i} \leq M(1 - w_i) \\ a_{i,k}, b_{i,k}, c_{i,k} \in \{0, 1\}, \quad \forall i \in I, \forall k \in K \\ a_{i,k} + b_{i,k} + c_{i,k} = 1, \quad \forall i \in I, \forall k \in K \\ d_{i,k} - x_i = r_{i,k} + s_{i,k} + t_{i,k}, \quad \forall i \in I, \forall k \in K \\ -C a_{i,k} \leq r_{i,k} \leq 0, \quad \forall i \in I, \forall k \in J \\ 0 \leq s_{i,k} \leq \bar{y} b_{i,k}, \quad \forall i \in I, \forall k \in K \\ \bar{y} c_{i,k} \leq t_{i,k} \leq C c_{i,k}, \quad \forall i \in I, \forall k \in K \\ \beta_{i,k} = -c\mathbb{P}(\omega_k)\left(\frac{s_{i,k}}{\bar{y}} + c_{i,k}\right), \quad \forall i \in I, \forall k \in K. \end{cases} \end{aligned} \tag{3.26}$$

3.3.3 Reformulation of Shared-Wait-and-See

When the leader shares the value of the vector d_0 with the follower, we must consider this information in the objective function. Since prices are also parameters, the demand vector d becomes fixed (given by (3.8)) and known by the follower. The leader then must solve

$$\varphi(y, d_0) := \begin{cases} \max_{p,v} & \sum_{i=1}^{n} (1-c)p_i \cdot \min(x_i + y_i, d_i) \\ s.t. & \begin{cases} p_i \in [p_{i,\min}, p_{i,\max}], \quad \forall i \in I \\ v \text{ solves } F(p). \end{cases} \end{cases} \tag{3.27}$$

where $F(p)$ is given by

$$F(p) := \begin{cases} \min_v & \sum_{i=1}^{n} c p_i \max(-x_i - y_i, -d_i) + \sum_{i \neq j} \alpha_{ij} v_{ij} \\ s.t. & \begin{cases} -v \leq 0 \\ \sum_{j \neq i} v_{ij} - x_{0i} \leq 0, \quad \forall i \in I \end{cases} \end{cases} \tag{3.28}$$

In this scenario, we can state the following proposition.

Theorem 3.4 *For any given value of the random vector $\zeta = (y, d_0)$, the Shared-Wait-and-See problem associated to the leader's problem (3.13) is equivalent (in the sense of global solutions) to its MPCC reformulation given by*

$$\max_{p,v,\lambda,\gamma,\beta} \quad \sum_{i=1}^{n}(1-c)p_i \cdot \min(x_i + y_i, d_i)$$
$$s.t. \quad \begin{cases} p_i \in [p_{i,\min}, p_{i,\max}], \quad \forall i \in I \\ \sum_{j\neq i} v_{ij} - x_{0i} \leq 0, \quad \forall i \in I \\ c(\beta_i - \beta_j) + \alpha_{ij} - \lambda_{ij} + \gamma_i = 0, \quad \forall i \neq j \\ \gamma_i(\sum_{j\neq i} v_{ij} - x_{0i}) = 0, \quad \forall i \in I \\ \lambda_{ij} v_{ij} = 0, \quad \forall i \neq j \in I \\ v \geq 0, \gamma, \lambda \geq 0, \end{cases} \tag{3.29}$$

where the variables $\{\beta_i \ : \ i \in I\}$ verify that

$$\beta_i \in \begin{cases} \{0\} & \textit{if } d_i - x_i < y_i \\ \{p_i\} & \textit{if } d_i - x_i > y_i \\ [0, p_i] & \textit{if } d_i - x_i = y_i \end{cases}. \tag{3.30}$$

Furthermore, the multipliers $\gamma = (\gamma_i \ : \ i \in I)$ and $\lambda = (\lambda_{ij} \ : \ i \neq j \in I)$ verify that

$$0 \leq \gamma_i \leq 2p_{\max} \quad \textit{and} \quad 0 \leq \lambda_{ij} \leq 4p_{\max}, \tag{3.31}$$

where $p_{\max} = \max_{i\in I}\{p_{i,\max}\}$.

Proof. The equivalence between the bilevel problem (3.27) and the MPCC reformulation follows as in the proof of Theorem 3.3. Thus, it is enough to show that problem (3.29) coincides with the MPCC reformulation of problem (3.27).

Fix a pair of multipliers (λ, γ). As the objective function for the follower is non-differentiable this time, the Fermat condition within the Karush-Kuhn-Tucker equations is given by the inclusion $0 \in \partial\mathcal{L}(v)$, where

$$\mathcal{L}(v) = \sum_{i=1}^{n} cp_i \max\{-x_i - y_i, -d_i\} + \sum_{i\neq j} \alpha_{ij} v_{ij} - \langle \lambda, v\rangle + \sum_{i\in I} \gamma_i (\sum_{j\neq i} v_{ij} - x_{0i}).$$

Furthermore, as all the involved functions in this formula are convex and continuous we can compute the required subdifferential as a sum of separated subdifferentials (see Chapter 1). If we call $\Gamma(t) = \partial(\max\{0, \cdot\})(t)$, it is clear that

$$\Gamma(t) = \begin{cases} \{0\} & \text{if } t < 0 \\ \{1\} & \text{if } t > 0 \\ [0, 1] & \text{if } t = 0 \end{cases}.$$

Therefore, by the convex subdifferential chain rule (see, e.g., [8, Chapter 16]),

$$\begin{aligned}
\partial\left(\sum_{i=1}^{n} cp_i \max\{-x_i - y_i, -d_i\}\right) &= \sum_{i=1}^{n} cp_i \partial\left(\max\{0, -x_i - y_i + d_i\}\right) \\
&= \sum_{i=1}^{n} cp_i \Gamma(-x_i - y_i + d_i)\nabla_v(-x_i) \\
&= \sum_{i=1}^{n} cp_i \Gamma(-x_i - y_i + d_i)(e_{i\bullet} - e_{\bullet i})
\end{aligned}$$

With this formula in mind, the inclusion $0 \in \partial\mathcal{L}(v)$ is equivalent to the existence of a vector $\beta \in \mathbb{R}^n$ such that $\beta_i \in p_i\Gamma(-x_i - y_i + d_i)$ for every $i \in I$, and such that

$$0 = \sum_{i=1}^{n} c\beta_i(e_{i\bullet} - e_{\bullet i}) + \nabla\left(\sum_{i\neq j} \alpha_{ij}v_{ij} - \langle\lambda, v\rangle + \sum_{i\in I}\gamma_i(\sum_{j\neq i} v_{ij} - x_{0i})\right).$$

The above vector equation can be equivalently written as the set of equations

$$0 = c(\beta_i - \beta_j) + \alpha_{ij} - \lambda_{ij} + \gamma_i, \quad \forall i \neq j.$$

where $\{\beta_i \ : \ i \in I\}$ are new variables verifying the inclusion (3.30).

The complementary equations are still the same as in (3.22). Thus, putting all together, we conclude that the MPCC reformulation of (3.27) is indeed given by (3.29).

Finally, similar to the Wait-and-See case, we can prove the following:

- If $\gamma_i = 0$, then
$$\lambda_{ij} = c\beta_i - c\beta_j + \alpha_{ij} \leq 2p_{\max}, \quad \forall j \in \{1, \ldots, n\}.$$
- If $\gamma_i \neq 0$, then the second complementary equation implies that $x_{0i} = \sum_{j\neq i} v^*_{ij}$. Since $x_{0i} \neq 0$, this also implies that there exists a value of j such that $v^*_{ij} \neq 0$, in which case $\lambda_{ij} = 0$ using the first complementary equation. Hence, we conclude that
$$c\beta_i - c\beta_j + \alpha_{ij} + \gamma_i = 0 \implies \gamma_i = c\beta_j - c\beta_i - \alpha_{ij} \leq 2p_{\max},$$
and so we can compute
$$\lambda_{ij} = c\beta_i - c\beta_j + \gamma_i + \alpha_{ij} \leq 4p_{\max}, \quad \forall j \in \{1, \ldots, n\}.$$

Regardless the case, we conclude that

$$\lambda_{ij} \in [0, 4p_{\max}], \quad \gamma_i \in [0, 2p_{\max}],$$

finishing our proof. □

This proposition allows us to replicate the big-M strategy used with the Wait-and-See value, regarding the complementarity constraints. In the Shared-Wait-and-See case, one last numerical consideration involves the additional constraints used to tackle the β_i terms. We define now

$$C = p_{\max}(\max\{d_{0,i}(\omega_k) \ : \ i \in I, k \in J\} + N_0)$$

and two additional boolean variables,

$$s_i, t_i \in \{0,1\} \ , \quad s_i + t_i \leq 1$$

so at most, one of them gets the value 1. Then, we add the following constraints:

$$\beta_i \leq p_{i,\max}(1 - s_i) \ , \quad x_i + y_i - d_i \geq -C(1 - s_i)$$

$$p_i - \beta_i \leq p_{i,\max}(1 - t_i) \ , \quad x_i + y_i - d_i \leq C(1 - t_i)$$

$$x_i + y_i - d_i \leq C(s_i + t_i) \ , \quad x_i + y_i - d_i \geq -C(s_i + t_i).$$

Here we have three feasible scenarios:

- If $s_i = 1$, then $t_i = 0$, the first set of equations leads to $\beta_i = 0$, and the other ones leave $x_i + y_i - d_i$ able to get positive values.
- If $t_i = 1$, then $s_i = 0$, the second set of equations leads to $\beta_i = p_i$, and the other ones leave $x_i + y_i - d_i$ able to get negative values.
- If $s_i = t_i = 0$, the third set of equations leads to $x_i + y_i - d_i = 0$, and the other ones leave $\beta_i \in [0, p_{i,\max}]$.

The final problem we solve for each $\zeta = (y, d_0)$ is then given by

$$\begin{aligned} \max \quad & \sum_{i=1}^{n} (1-c) p_i \cdot \min(x_i + y_i, d_i) \\ s.t \quad & \begin{cases} p_i \in [p_{i,\min}, p_{i,\max}], \quad \forall i \in I \\ c(-\beta_i + \beta_j) + \alpha_{ij} - \lambda_{ij} + \gamma_i = 0, \quad \forall i \in I, \ \forall j \neq i \\ -M z_{ij} \leq \lambda_{ij} \leq M z_{ij} \\ -M(1 - z_{ij}) \leq v_{ij} \leq M(1 - z_{ij}) \\ -M w_i \leq \gamma_i \leq M w_i \\ -M(1 - w_i) \leq \sum_{j \neq i} v_{ij} - x_{0i} \leq M(1 - w_i) \\ x_i = x_{0,i} + \sum_{j \neq i} v_{ji} - \sum_{k \neq i} v_{ik}, \quad \forall i \in I \\ \beta_i \leq p_{i,\max}(1 - s_i), \quad \forall i \in I \\ x_i + y_i - d_i \geq -C(1 - s_i), \quad \forall i \in I \\ p_i - \beta_i \leq p_{i,\max}(1 - t_i), \quad \forall i \in I \\ x_i + y_i - d_i \leq C(1 - t_i), \quad \forall i \in I \\ x_i + y_i - d_i \leq C(s_i + t_i) \quad \forall i \in I \\ x_i + y_i - d_i \geq -C(s_i + t_i), \quad \forall i \in I \\ s_i + t_i \leq 1, \quad \forall i \in I \\ z_{ij} \in \{0,1\}, \ w_i, s_i, t_i \in \{0,1\}, \quad \forall i \in I \end{cases} \end{aligned} \tag{3.32}$$

Remark It is worth to notice that both MPCC reformulations (3.26) and (3.32) are mixed-integer bilinear programming problems. While in general even the simplest bilevel programming problems are NP-hard [12], mixed-integer bilinear problems can be handled efficiently by some commercial solvers, like `Gurobi` [50]. Of course, the treatment of these problems is heuristic, but with good results in practice.

3.4 Numerical Results

In order to compare the WS and SWS indicators and test the previously presented formulations, we designed some numerical experiments with artificial data. We consider $I = \{1, 2, 3, 4\}$ (four connected zones) who simulate four different communes of the Metropolitan Region of Chile (Santiago, Renca, Maipu, La Florida), $m = 3$ (low, normal and high demand scenarios) and $c = 0.75$ (75% of the ride fare is taken by the driver). Equation (3.8), that represents the connection between the ride price on each zone and the respective effective demand, is considered with $\delta = 0.9$, and the ride prices are delimited on each zone with $p_{i,\min} = 2.5$ USD and $p_{i,\max} = 12.5$ USD. The costs α_{ij} are modeled considering the actual distance between the four communes and the price of gasoline in May 2021, i.e. 1.10 USD.

Distances (km)	1	2	3	4
1	-	11	17	20
2	11	-	22	33
3	17	22	-	18
4	20	33	18	-

Table 3.2: Distances considered between zones.

A total fleet of $N_0 = 1000$ unmatched drivers is considered. For every $i \in I$, we generated 10 uniformly distributed values $h_{0,i}$ in the $[0, 1]$ interval. From those values, we built samples for the nominal demand given by

$$\bar{d}_{0,i} = D_0 h_{0,i}, \quad \forall i \in I.$$

For each of these scenarios, we simulated a hundred samplings for the triangular distributed nominal demand $d_{0,i} \sim \text{Tri}(0.7\bar{d}_{0,i}, 1.3\bar{d}_{0,i})$. These simulations are combined with the following scenarios for the demand and the previously matched drivers:

1. We consider the aggregated demand coefficient D_0 as a function of the total fleet of unmatched drivers: that is, $D_0 = PN_0$, with $P = 1, 2, 3, 4, 5$. This leads to the construction of the samplings for $(d_{0,i} : i \in I)$.

2. The quantity of previously matched drivers $y_j \sim U(0, \bar{y})$, with $\bar{y}$ as a proportion of the total fleet of unmatched drivers, $\bar{y} = QN_0$, with $Q = 1/4, 1/2, 3/4$.

Therefore, we consider a total of 15 different scenarios as a combination of the values for D_0 and $\bar{y}$. Then, for each scenario, we consider 10 different nominal values for the vector $\bar{d}_0$ (given by the generated vectors $(h_{0,j})_{j \in I}$), and for each of those nominal values, we compute solutions for the WS and SWS reformulated problems for 100 different samplings of the pairs (y, d_0). Thus, for each scenario, we solved 1000 samples.

For each sample of (y, d_0), we solved problems (3.26) and (3.32) using `Gurobi v9.1.2` as solver [50], and `Julia v1.6.2` as programming language [19], with the extra constraint of having integer fluxes between the zones. Our results are displayed in Figures 3.2, 3.3 and 3.4, which are organized as follow: each figure represents 5 scenarios, given by a fixed value of Q and the five values of P. The plotted values are the mean values from the 1000 samples for the corresponding scenario: in blue dots, the values for the Wait-and-See problem (3.26); in red triangles, the values for the Shared-Wait-and-See problem (3.32).

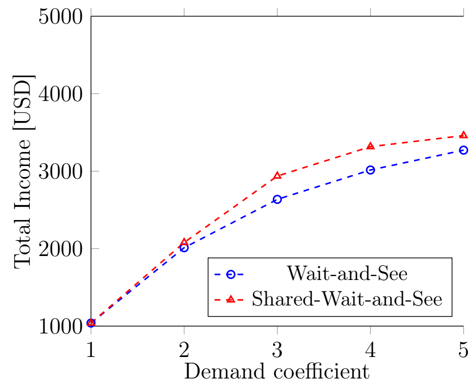

Figure 3.2: WS and SWS Comparison: $N_0 = 1000$, $\bar{y} = N_0/4$

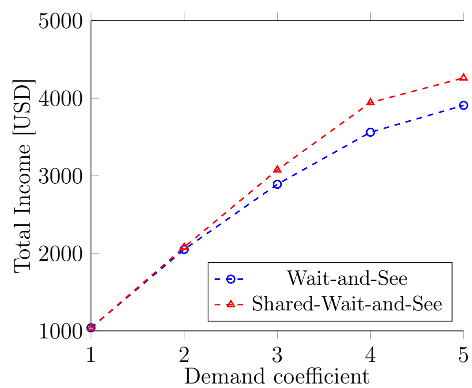

Figure 3.3: WS and SWS Comparison: $N_0 = 1000$, $\bar{y} = N_0/2$

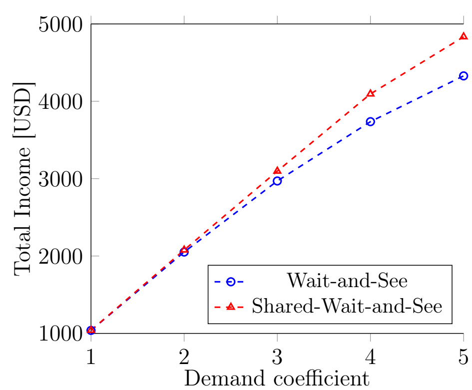

Figure 3.4: WS and SWS Comparison: $N_0 = 1000$, $\bar{y} = 3N_0/4$

As we can see in Figure 3.2, the SWS solution generates a greater total income for the scenarios with $\bar{y} = N_0/4$, especially in the middle scenarios, when the aggregated demand coefficient does not reach extremely low or extremely high values. This behavior seems to be the same as the floating population of previously matched drivers increases, that is in Figures 3.3 and 3.4, where $\bar{y} = N_0/2$ and $\bar{y} = 3N_0/4$ respectively. It seems that the $EVSI$, which is given by $SWS - WS$, is positive and has a concave behavior: when the aggregated demand is too small or too large, the $EVSI$ seems to be zero, reaching its maximum in a middle value.

Another effect we observe in our results is that pricing along is not enough to modify the behavior of the drivers. Their beliefs about the chances of getting a ride are also influential. Thus the revealing of such information has an important effect in the final reallocation.

As an illustrative example, we show how different the WS and the SWS solutions are for a particular scenario, given by one of the numerical experiments we did. In this scenario, $\bar{y} = 750$ and $P = 5$, and the values for x_0, y and d_0 are the ones described in Table 3.3.

Zone	x_0	y	d_0
1	107	444.55	510
2	283	296.68	1945
3	399	420.20	1010
4	211	568.07	1535

Table 3.3: Particular scenario with $\bar{y} = 750$ and $P = 5$.

The WS and SWS solutions are presented in Figures 3.5 and 3.6, respectively. The missing edges in each graph have 0-flow. The value functions for each solution are $\psi(y, d_0) = 3935.4$ USD and $\varphi(y, d_0) = 4635.4$ USD.

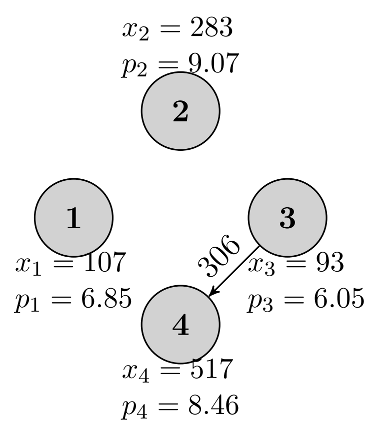

Figure 3.5: Graph with perfect information (Wait-and-see solution). Optimal value: 3935.4 USD.

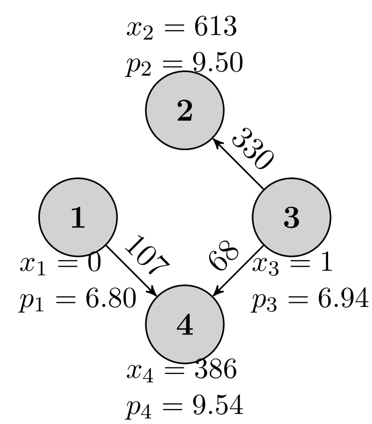

Figure 3.6: Graph with shared information (Shared wait-and-see solution). Optimal value: 4635.4 USD.

We can appreciate that the revelation of information produces a huge change in the solution. For this scenario, Zone 2 has a huge demand, but drivers simply do not know it since this value is vastly different of what they usually observe. With the second reallocation (Figure 3.6), the ride-hailing company can increase the prices in zones 2 and 4.

Chapter 4

Resilient Design of Eco-Industrial Parks

4.1 Overview of Eco-Industrial Parks Optimal Design Problems

An Eco-Industrial Park (EIP), as defined in [76], is a community of manufacturing and service businesses located together on a common property. Member businesses look for environmental, economic, and social performance through collaboration in managing environmental and resource issues. In this work, as it is usual in the literature, we model the EIP community as a central authority in charge of the design of the park at a first stage, and of optimizing the interactions within the members in the daily operation during its lifetime.

An example of EIP corresponds to the modeling of water exchange networks (see, e.g., [23, 102, 108] and the references therein). In this model, each participant of the EIP needs to consume fresh water for its industrial processes, and to send away partially contaminated water. In parallel, there is a central authority of the EIP, which is in charge of design the park and operate it afterwards, following some criteria that reflects environmental, economic and/or social benefits. The EIP design problem has already been studied under the setting of deterministic optimization, as reviewed in [22]. In this section, we aim to describe the model that we are going to use, in order to approach the design problem considering uncertainty. Most part of the water exchange model we present here is well-known in the literature, and we are based mainly on [102, 108].

During the last decades, uncertainty has gained major attention in the vast majority of the engineering and scientific research fields due to the need to deal with both the market volatility and the variation related to the operating conditions.

According to the works of Oberkampf et al. in [96], uncertainty can classified into three main categories, namely aleatory uncertainty or variability, epistemic uncertainty or reducible and error. The first one refers to inherent variation of the physical environment under consideration while the second one is related to any lack of knowledge or information. Finally, error is defined as a recognizable deficiency during the modelling or simulation phases that is not due to the lack of knowledge. According to this definition, the fluctuation of input data or operating conditions, as in this research work, can be directly related to the so called aleatory or stochastic uncertainty.

The approach to integrate uncertainty in modeling and simulation has been discussed in detail in [95] and several applications with different methodologies are available in literature. One of the possible applications in the process industry is design for flexibility that exploits both deterministic [115] and stochastic flexibility [98] indexes for various applications such as process units design [38] and optimal scheduling [37].

4.2 The Eco-Industial Park Bilevel Model

A water exchange network can be modeled as a simple directed graph $G = (V, \vec{E})$ with some specific conditions. First, the set $V = \{0, 1, \ldots, n\}$ represents the n agents/enterprises participating of the EIP, and 0 is an extra *sink node*. We will denote $I = \{1, \ldots, n\}$ to describe the set of agents of the park, and so $V = I \cup \{0\}$. The sink node 0 receives all the residual water exiting the park as final waste. Every agent is connected to the sink node, that is, $(i, 0) \in \vec{E}$ for each $i \in I$, and 0 is not connected back to any participant, that is, $(0, i) \notin \vec{E}$ for each $i \in I$.

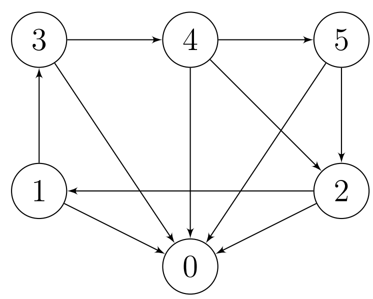

Figure 4.1: Example of $G = (V, \vec{E})$ for 5 agents.

In this model, we consider flux capacities $L = (L_{ij} \ : \ i, j \in I)$ [ton/h] for the connections between every pair of participants $i, j \in I$. For a pair $i, j \in I$, L_{ij} is non-negative and it determines how much water can be sent from i to j. We assume that the capacities of the connections $(i, 0)$ are unbounded. If $L_{ij} = 0$, then $(i, j) \notin \vec{E}$. Therefore, each EIP is uniquely defined by its capacity vector L.

At the design stage, the design problem is to decide the vector L. When L is decided, we implicitly solve the problem of deciding the connections in $\vec{E}$.

4.2.1 Physical operation model

After the design stage, for a given EIP with capacities $L = (L_{ij})$, there is a set of operational constraints that every participant $i \in I$ must follow. These constraints must hold every day of the park's operation, during the whole lifetime horizon.

Each day, each enterprise $i \in I$ produces a mass load of contaminant M_i [kg/h] that needs to be diluted, considering a maximum inlet outlet concentration $C_{i,\text{out}}$ [mg/L]. To do so, the enterprise can buy fresh water z_i [ton/h] and receive partially polluted water from other enterprises. We denote by F_{ki} [ton/h] the water sent from participant $k \in I$ to participant i.

We will assume that the operation of every participant is optimal, in the sense that their maximal outlet concentration is always attained and no excess of fresh water is consumed (see, e.g., [108, 102]).

Finally, after diluting M_i, participant i must send away the polluted water, by either sending part of their polluted water to another participant $j \in I$, through the flux F_{ij} [ton/h], or by discharging residual polluted water to the sink node through the flux O_i [ton/h].

Also, for an exchange to be valid, every participant can accept polluted water with a maximum inlet concentration $C_{i,\text{in}}$ [mg/L]. This constraint is given by the process description of each participant. We assume that all inlet fluxes, including the fresh water, are mixed before entering the process. Thus, the inlet concentration constraint is evaluated in the mixed inlet flux.

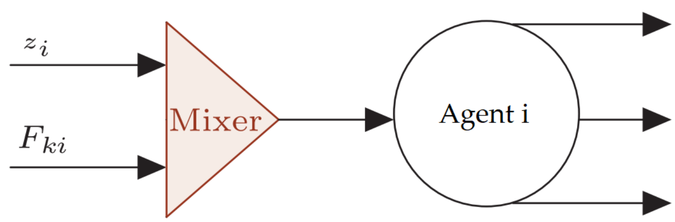

Figure 4.2: Water Mixture Description for Agent i. Original Figure from [108].

To sum up, the operation variables of the park are given by: 1) the fresh water consumption of each agent, given by the vector $z = (z_i \;:\; i \in I)$; 2) the exchange water matrix $F = (F_{ij} \;:\; i \neq j \in I)$; and 3) the discharge of each agent to the sink node, given by the vector $O = (O_i \;:\; i \in I)$. A valid operation is then given by values of (z, F, O) satisfying the following operation constraints:

1. **Water Mass Balance:** for every participant $i \in I$, the total inlet flux must coincide with the total outlet flux.

$$z_i + \sum_{k \neq i} F_{ki} = \sum_{j \neq i} F_{ij} + O_i. \tag{4.1}$$

   At the sink node, there is no balance constraint.

2. **Contaminant Mass Balance:** For every participant $i \in I$, the total inlet contaminant mass must coincide with the total outlet contaminant mass, that is,

$$M_i + \sum_{k \neq i} C_{k,\text{out}} F_{ki} = C_{i,\text{out}} \left(O_i + \sum_{j \neq i} F_{ij} \right). \tag{4.2}$$

   The mass is computed from the fluxes F thanks to the optimality assumption that the outlet concentration is always attained.

3. **Inlet/Outlet Concentration Constraints:** for every participant $i \in I$, we have that

$$\sum_{k \neq i} C_{k,\text{out}} F_{ki} \leq C_{i,\text{in}} \left(z_i + \sum_{k \neq i} F_{ki} \right). \tag{4.3}$$

   The above inequality is the inlet concentration constraint expressed in terms of contaminant mass.

4. **Positivity of Fluxes:** all the fluxes in the EIP must be non-negative, that is,

$$F_{ij} \geq 0, \quad \forall i \neq j \in I \qquad \text{and} \qquad z_i, O_i \geq 0, \quad \forall i \in I. \tag{4.4}$$

5. **Boundedness of exchanges:** all the fluxes in the EIP must be within the capacities given by the vector L, that is,

$$F_{ij} \leq L_{ij}, \quad \forall i \neq j \in I. \tag{4.5}$$

4.2.2 Economical constraints for participation

The central authority's goal at the daily operation is to minimize the global fresh water consumption. Nevertheless, this is not necessarily aligned with the individual participants' interests, which is to minimize their operational costs. Hence, jointly with the already detailed physical constraints, we must add economical ones for the daily operation of every participant, considering their *individual rationality*; a well-known principle : any enterprise will take part of the EIP only if this participation is economically convenient (see, e.g., [61]).

Consider the operation of an isolated agent $i \in I$ (i.e. no participation in an EIP). This agent should operate by *stand-alone* conditions, where all the water needed to dilute the mass load of contaminant should be bought fresh, and all the water waste should then be discharged to the sink. This means that all the fresh water z_i and the discharged waste O_i, are given by the same following expression:

$$z_i = \frac{M_i}{C_{i,\text{out}}} \qquad \text{and} \qquad O_i = \frac{M_i}{C_{i,\text{out}}}. \tag{4.6}$$

Then, if we denote by c as the marginal cost for consuming fresh water, and by d the marginal cost for the discharging water, the daily stand-alone operational cost is given by

$$\text{SAO}_i(M_i) = (c + d)\text{SA}_i(M_i), \tag{4.7}$$

with $SA_i(M_i) = \dfrac{M_i}{C_{i,\text{out}}}$.

Now, let us consider the operation of an agent $i \in I$ connected to other participants in an EIP. Now, this agent has two additional options for its operation: it can partially replace fresh water consumption by receiving partially polluted water from other participants, and it also can send some of its own water waste to other agents as partially polluted water. Hence, the daily operational cost in this new scenario is given by

$$\mathrm{Cost}_i(F, z) = e\left(\sum_{j\neq i} F_{ij} + \sum_{k\neq i} F_{ki}\right) + dO_i + cz_i, \tag{4.8}$$

where e models a marginal costs for using the connections of the park, which is co-paid for the sending and receiving participants.

After this discussion, we add an additional constraint for the model in order to tackle the *individual rationality* of every participant, which ensures that the operational costs for every agent are less than the *stand-alone* operation, that is,

$$\mathrm{Cost}_i(F, z) \leq (c + d)\mathrm{SA}_i(M_i). \tag{4.9}$$

This economic constraint was first introduced in [108] to validate models with a central operation under the assumption of selfish agents. It was then used in [6] with the same purposes. While here we are not necessarily considering selfish agents, constraint (4.9) is still needed to model rational (cooperative) agents that follow the central operation.

4.2.3 Uncertainty and Two-stage model

In order to design the EIP, we have to consider a new element: the mass load production of contaminant of every participant, $\xi = (M_1, \ldots, M_n)$ is uncertain at the design stage, because each process has daily unpredictable variations. These variations are only revealed during the daily operation of the EIP, and of course, they can be different every time. Thus, after ξ is revealed, we can define for a given capacity vector L, the daily operation problem $Q(L, \xi)$ as

$$Q(L,\xi) = \begin{cases} \displaystyle\min_{(z,F)} & Z := \sum_{i=1}^{n} z_i \\ \text{s.t.} & z_i + \sum_{k\neq i} F_{ki} = O_i + \sum_{j\neq i} F_{ij} \\ & M_i + \sum_{k\neq i} C_{k,\mathrm{out}} F_{ki} = C_{i,\mathrm{out}}\left(O_i + \sum_{j\neq i} F_{ij}\right) \\ & \sum_{k\neq i} C_{k,\mathrm{out}} F_{ki} \leq C_{i,\mathrm{in}}\left(z_i + \sum_{k\neq i} F_{ki}\right) \\ & \mathrm{Cost}_i(F, z) \leq (c + d)\mathrm{SA}_i(M_i) \\ & 0 \leq F \leq L \\ & z \geq 0, \end{cases} \tag{4.10}$$

It is well known that $\xi \mapsto Q(L, \xi)$ is a measurable random function (see, e.g. [111, Chapter 2]) and therefore, its expected value is well defined.

For L and ξ fixed, we define $X(L, \xi)$ as the feasible set of Problem (4.10), which is given by linear constraints only. That is,

$$X(L,\xi) := \left\{ (z,F) \ : \ \begin{array}{l} z_i + \sum_{k\neq i} F_{ki} = O_i + \sum_{j\neq i} F_{ij} \\ M_i + \sum_{k\neq i} C_{k,\text{out}} F_{ki} = C_{i,\text{out}} \left(O_i + \sum_{j\neq i} F_{ij} \right) \\ \sum_{k\neq i} C_{k,\text{out}} F_{ki} \leq C_{i,\text{in}} \left(z_i + \sum_{k\neq i} F_{ki} \right) \\ \text{Cost}_i(F,z) \leq (c+d)\text{SA}_i(M_i) \\ 0 \leq F \leq L \\ z \geq 0. \end{array} \right\} \tag{4.11}$$

This is the general setting of a two-stage problem: minimize a cost function, which depends on the design variables L (decided here-and-now), taking into account the optimal operation of every day, which is given by the parametric problem (4.10).

An option to solve this two-stage problem, is to obtain or define first an analytical expression for $Q(L,\xi)$. See, e.g.[38] for an application. However, this is not always possible. Instead, we will solve this problem using the implicit expression of $Q(L,\xi)$ as the optimal value of the problem (4.10), leading to a stochastic optimization problem with recourse (see Chapter 2). Using this tool, we can avoid the determination of the analytical expression for Q, and work directly with the optimal value obtained in the daily operation stage.

Therefore, we can consider for the first stage, the minimization of an objective function of the form

$$\langle \kappa, L \rangle + \mathcal{R}[Q(L,\xi)], \tag{4.12}$$

where κ models investment costs and $\mathcal{R}$ is a operator that takes into account the stochasticity of the model, and computes what is called a *risk measure*. These kind of measure aim to characterize the uncertain value of $Q(L,\xi)$ by two relevant characteristics, its mean (to measure the expected outcome) and its risk or dispersion (to measure its uncertainty).

This kind of functions allow us to go beyond classic computations (such as the expected value) and consider risk-averse design (see, e.g. [111, Chapter 6]), in order to get, for example, an EIP that is *resilient* in some sense to difficult or undesired conditions during its lifetime. Other works (see, e.g. [48]) discuss these ideas in different contexts.

4.3 Optimization Criteria: Performance vs. Resilience

Once defined $Q(L,\xi)$, we can tackle the optimal design problem: to decide the best capacity vector L, taking into account the uncertain operational cost-to-go $Q(L,\xi)$.

On a first approach, is natural to consider the averaged fresh water consumption costs of the EIP as the objective function to minimize. This correspond to expected value $A\mathbb{E}[Q(L,\xi)]$, where A is the lifetime factor, and allows us to control some kind of "average day" on its long-term operation. However, this does not necessarily give us an optimal EIP considering other indicators, such as robustness in face of uncertainty.

Another possible approach is to design an EIP that admits *good-enough* operations in most of the uncertain scenarios, for example by fixing an admissible level of improvement on the total expendings for the daily operation, and look for a configuration that allows us to get the greatest possible number of scenarios where this improvement is achieved.

In order to quantify this goal, we introduce here what we call the $(1-\alpha)$-level of *goodness* for an EIP as

$$\mathrm{G}_\alpha(L,\xi) \doteq Q(L,\xi) - \alpha \mathrm{SA}(\xi), \tag{4.13}$$

where $\mathrm{SA}(\xi) = \sum_{i\in I} \mathrm{SA}_i(M_i)$ is the total net cost, if all the agents worked on *stand-alone* operation. If $\mathrm{G}_\alpha \leq 0$, it means that the EIP operation is better than the stand-alone one. Hence, we define the *resilience* of the EIP as

$$\mathrm{Res}_\alpha(L) = \mathbb{P}[\mathrm{G}_\alpha(L,\xi) \leq 0] \tag{4.14}$$

In order to maximize the number of *good* scenarios, we can maximize the value of $\mathrm{Res}_\alpha(L)$ (correspondingly minimizing $-\mathrm{Res}_\alpha(L)$) as a part of the objective function. Using this idea we can obtain more *resilient* designs, that ensure us that most of the scenarios work well for the imposed improvement level.

Hence, an unifying formulation that considers the average fresh water consumption, the *resilient* design and the investment costs, is given by:

$$\begin{cases} \min_L & w_1\langle \kappa, L\rangle + w_2\mathbb{E}[Q(L,\xi)] - w_3\mathrm{Res}_\alpha(L) \\ \text{s.t.} & L \geq 0 \end{cases}$$

where $w = (w_1, w_2, w_3) \geq 0$ is a specific weight vector, which allows us to prioritize the criteria. Therefore, our general two-stage problem is given by the design problem (4.17) in the first stage, and the daily operation (4.10) in the second one.

Finally, at the constraints level in the design stage, similarly to the objective function construction, we can add criteria for the design of the EIP. Consider for this, the following models:

- **Budget constraints**: For a budget $B \geq 0$ [$], a budget constraint is given by

  $$\langle \kappa, L\rangle \leq B \tag{4.15}$$

  we can control the total budget for the design of the EIP, combined with a minimization of its resilience and/or its fresh water consumption.

- **Minimal Desired Resilience**: considering the definition (4.14), this can be also considered as a constraint, writing for example

  $$\mathrm{Res}_\alpha(L) \geq \beta. \tag{4.16}$$

  There's a lot of work in this area, known as *chance constraints optimization*, where, as discussed in Chapter 2, the value of a probability function is fixed over a desired quantity.

Therefore, we get the general problem,

$$\mathcal{P} = \begin{cases} \min_L & w_1\langle\kappa, L\rangle + w_2\mathbb{E}[Q(L,\xi)] - w_3\mathrm{Res}_\alpha(L) \\ \text{s.t.} & L \geq 0 \\ & c_1(\langle\kappa, L\rangle - B) \leq 0 \\ & c_2(\mathrm{Res}_\alpha(L) - \beta) \geq 0 \end{cases}, \tag{4.17}$$

where $c_1, c_2 \in \{0,1\}$ will be used in order to eventually consider budget or resilience constraints, as recently defined.

4.4 Reformulation under SAA Method

Considering the elements detailed in Section 2.2.2, we can now reformulate our two-stage problem using the Sample Average Approximation Method.

Given a sample $\{\xi^1, \xi^2, \ldots, \xi^N\}$ for the mass load production of contaminant, where each $\xi^m = (M_1^m,\ M_2^m,\ \ldots,\ M_n^m)$ for $m = 1, \ldots, N$, we can compute the estimator for the expected value of Q as

$$q_N(L) = \frac{1}{N}\sum_{m=1}^{N} Q(L, \xi^m), \tag{4.18}$$

and given a realization $\{\hat{\xi}^1, \hat{\xi}^2, \ldots, \hat{\xi}^N\}$ of this sample, we can define

$$\hat{q}_N(L) = \frac{1}{N}\sum_{m=1}^{N} Q(L, \hat{\xi}^m), \tag{4.19}$$

which, as discussed before, is now a deterministic expression.

For $\mathrm{Res}_\alpha(L)$, as it is defined in (4.14) by a probability, it can also be written as an expected value, considering the corresponding indicator function for $G_\alpha(L,\xi)$. Namely,

$$\mathrm{Res}_\alpha(L) = \mathbb{P}[G_\alpha(L,\xi) \leq 0] = \mathbb{E}\left[\mathbb{1}_{(-\infty,0]}[G_\alpha(L,\xi)]\right].$$

Here, for a set $A \subset \mathbb{R}$, $\mathbb{1}_A$ stands for the indicator function of A, that is,

$$\mathbb{1}_A(t) = \begin{cases} 1 & \text{if } x \in A. \\ 0 & \text{otherwise.} \end{cases} \tag{4.20}$$

Following the same previously considered realization of the sample, we can define in this case

$$\widehat{\mathrm{Res}}_\alpha(L) = \frac{1}{N}\sum_{m=1}^{N} \mathbb{1}_{(-\infty,0]}[\mathrm{G}_\alpha(L, \hat{\xi}^m)]. \tag{4.21}$$

With all of this in mind, we define the initial reformulation of the Problem (4.17)

$$\widehat{\mathcal{P}}_0 = \begin{cases} \min\limits_L & w_1\langle\kappa, L\rangle + w_2\hat{q}_N(L) - w_3\widehat{\mathrm{Res}}_\alpha(L) \\ \text{s.t.} & L \geq 0 \\ & c_1(\langle\kappa, L\rangle - B) \leq 0 \\ & c_2(\widehat{\mathrm{Res}}_\alpha(L) - \beta) \geq 0 \end{cases}, \tag{4.22}$$

where $c_1, c_2 \in \{0,1\}$ will be used in order to eventually consider/remove budget or resilience constraints, as defined in (4.15) and (4.16) respectively.

In order to compute the objective function for this problem, two main difficulties arise: we do not have an explicit expression for $Q(L, \hat{\xi}^m)$ at any realization $\hat{\xi}^m$, and the indicator functions that define $\widehat{\mathrm{Res}}_\alpha(L)$ do not have good properties for optimization in general (they are not convex nor continuous).

To avoid computation of $Q(L, \hat{\xi}^m)$, we create variables (z^m, F^m) that, a posteriori, will represent the optimal operation of the ecopark at the second-stage, given the realization $\hat{\xi}^m$. Similarly, to avoid computation of $\mathbb{1}_{(-\infty,0]}[G_\alpha(L, \hat{\xi}^m)]$, we will include a binary variable y^m that represents the value of the indicator function. The idea is to formulate a mixed linear problem that, a posteriori, will verify

$$Q(L, \hat{\xi}^m) = \sum_i z_i^m \quad \text{and} \quad \mathbb{1}_{(-\infty,0]}[G_\alpha(L, \hat{\xi}^m)] = y_m.$$

For dealing with the computation of $\hat{q}_N$, we consider for every $m \in [N] = \{1, \dots, N\}$, $(z^m, F^m) \in X(L, \hat{\xi}^m)$ and

$$Z^m = \sum_{i=1}^n z_i^m.$$

The sum of every Z^m will be used as a replacement of $\hat{q}_N(L)$ in our final reformulations of Problem (4.22). Now, in order to compute the indicator functions needed at the resilience term, we define an additional binary variable $y \in \{0,1\}^N$ that works as follows: our objective function is reformulated as

$$f(L, y) := w_1 \langle \kappa, L \rangle + \frac{w_2}{N} \sum_{m=1}^N Z^m - \frac{w_3}{N} \sum_{m=1}^N y_m, \tag{4.23}$$

where y_m verifies, for every realization $\hat{\xi}^m$, the additional constraint

$$\mathrm{G}_\alpha(L, Z^m, \hat{\xi}^m) \leq \mathrm{SA}(\hat{\xi}^m)(1 - y_m), \tag{4.24}$$

where $\mathrm{G}_\alpha(L, Z^m, \hat{\xi}^m) = Z^m - SA(\hat{\xi}^m)$, is an extension of the goodness function G_α that evaluates how good is the operation (z^m, F^m) for the realization $\hat{\xi}^m$. This works as follows: if $y_m = 1$, then (4.24) implies that $\mathrm{G}_\alpha(L, Z^m, \hat{\xi}^m) \leq 0$, or in other words,

$$y_m = 1 \implies \mathbb{1}_{(-\infty,0]}[\mathrm{G}_\alpha(L, Z^m, \hat{\xi}^m)] = 1.$$

If $y_m = 0$, then (4.24) implies that

$$\mathrm{G}_\alpha(L, Z^m, \hat{\xi}^m) \leq \mathrm{SA}(\hat{\xi}^m),$$

which becomes a non-active constraint, as we can see that

$$\begin{aligned}
\mathrm{G}_\alpha(L, Z^m, \hat{\xi}^m) &= Z^m - \alpha \mathrm{SA}(\hat{\xi}^m) \\
&\leq \mathrm{SA}(\hat{\xi}^m) - \alpha \mathrm{SA}(\hat{\xi}^m) \\
&= (1 - \alpha) \mathrm{SA}(\hat{\xi}^m) \\
&\leq \mathrm{SA}(\hat{\xi}^m).
\end{aligned}$$

Before stating our main result for this section, we define the following reformulation of problem (4.17), considering the additional constraint (4.24).

$$\widehat{\mathcal{P}}_1 = \begin{cases} \min\limits_{L,z,F,y} & w_1\langle \kappa, L\rangle + \dfrac{w_2}{N}\displaystyle\sum_{m=1}^{N} Z^m - \dfrac{w_3}{N}\sum_{m=1}^{N} y_m \\ \text{s.t.} & L \geq 0 \\ & c_1(\langle \kappa, L\rangle - B) \leq 0 \\ & c_2\left(\dfrac{1}{N}\displaystyle\sum_{m=1}^{N} y_m - \beta\right) \geq 0 \\ & (z^m, F^m) \in X(L, \hat{\xi}^m),\ \forall m \in [N] \\ & \mathrm{G}_\alpha(L, Z^m, \hat{\xi}^m) \leq \mathrm{SA}(\hat{\xi}^m)(1 - y_m),\ \forall m \in [N] \\ & y \in \{0,1\}^N \end{cases} \tag{4.25}$$

Theorem 4.1 *If L^* is an optimal solution of Problem (4.22), then there exists (z^*, F^*, O^*, y^*) such that $(L^*, z^*, F^*, O^*, y^*)$ is an optimal solution of Problem (4.25). Conversely, if $(L^*, z^*, F^*, O^*, y^*)$ is an optimal solution of Problem (4.25), then L^* is an optimal solution for Problem (4.22). In both cases, one has that*

$$w_1\langle c, L^*\rangle + w_2\hat{q}_N(L^*) - w_3\widehat{Res}_\alpha(L^*) = w_1\langle c, L^*\rangle + \frac{w_2}{N}\sum_{m=1}^{N} z^{*m} - \frac{w_3}{N}\sum_{m=1}^{N} y^*_m. \tag{4.26}$$

Proof. If L^* is an optimal solution of Problem (4.22), we can define, for each $m \in [N]$, (z^{*m}, F^{*m}, O^{*m}) as the optimal solution of the second-stage problem (4.10), for L^* and the realization $\hat{\xi}^m$. Then, one has that

$$Z^{*m} = \sum_{i=1}^{n} z_i^{*m} = Q(L^*, \hat{\xi}^m).$$

We can also define $y^* = (y^*_m)$ as

$$y^*_m = \mathbb{1}_{(-\infty,0]}[\mathrm{G}_\alpha(L^*, Z^{*m}, \hat{\xi}^m)].$$

Then, it follows that

$$\hat{q}_N(L) = \sum_{m=1}^{N} Z^{*m}, \quad \widehat{\mathrm{Res}}_\alpha(L) = \frac{1}{N}\sum_{m=1}^{N} y_m,$$

and that constraint (4.24) is verified for each $m \in [N]$.

Let us suppose that Problem (4.25) had a different optimal solution, let us say $(\bar{L}, \bar{z}, \bar{F}, \bar{O}, \bar{y})$. As we discussed before, if $\bar{y}_m = 1$, then $G_\alpha(\bar{L}, \bar{Z}, \hat{\xi}^m) \leq 0$, implying that

$$\frac{1}{N}\sum_{m=1}^{N} \bar{y}_m \leq \widehat{\mathrm{Res}}_\alpha(\bar{L}).$$

Moreover, $\frac{1}{N}\sum_{m=1}^{N}\bar{Z}_m \geq \hat{q}_N(\bar{L})$. Thus, we can write:

$$\begin{aligned}
w_1\langle c, L^*\rangle + w_2\hat{q}_N(L^*) - w_3\widehat{\text{Res}}_\alpha(L^*) &= w_1\langle c, L^*\rangle + \frac{w_2}{N}\sum_{m=1}^{N} Z^{*m} - \frac{w_3}{N}\sum_{m=1}^{N} y_m^* \\
&> w_1\langle c, \bar{L}\rangle + \frac{w_2}{N}\sum_{m=1}^{N} \bar{Z}^m - \frac{w_3}{N}\sum_{m=1}^{N} \bar{y}_m \\
&\geq w_1\langle c, \bar{L}\rangle + w_2\hat{q}_N(\bar{L}) - w_3\text{Res}_\alpha(\bar{L}),
\end{aligned}$$

which is a contradiction, given that L^* was an optimal solution for Problem (4.22). Therefore, we conclude that $(L^*, z^*, F^*, O^*, y^*)$ is optimal for Problem (4.25), and moreover, (4.26) holds.

Conversely, if $(L^*, z^*, F^*, O^*, y^*)$ is an optimal solution of Problem (4.25), we claim that L^* is feasible for (4.22). Clearly, L^* is nonnegative and it verifies the budget constraint whenever $c_1 = 1$. We only need to show that, whenever $c_2 = 1$, L^* also verifies the resilience constraint $\widehat{\text{Res}}_\alpha(L^*) \geq \beta$. Note that, we always have that

$$G_\alpha(L^*, Z^{*m}, \hat{\xi}^m) = Z^{*m} - \text{SA}(\hat{\xi}^m) \geq Q(L^*, \hat{\xi}^m) - \text{SA}(\hat{\xi}^m) = G_\alpha(L^*, \hat{\xi}^m).$$

Then, constraint (4.24) yields that

$$\frac{1}{N}\sum_{m=1}^{N} y_m^* \leq \frac{1}{N}\sum_{m=1}^{N} \mathbb{1}_{(-\infty,0]} G_\alpha(L^*, Z^{*m}, \hat{\xi}^m) \leq \frac{1}{N}\sum_{m=1}^{N} \mathbb{1}_{(-\infty,0]} G_\alpha(L^*, \hat{\xi}^m) = \widehat{\text{Res}}_\alpha(L^*).$$

Thus, since $\frac{1}{N}\sum_{m=1}^{N} y_m^* \geq \beta$, the conclusion follows. The claim is then proved.

Now, we will show that optimality of $(L^*, z^*, F^*, O^*, y^*)$ entails that

$$\mathbb{1}_{(-\infty,0]}[\text{G}_\alpha(L^*, Z^*, \hat{\xi}^m)] = y_m^*, \quad \forall m \in \{1, \ldots, N\}.$$

Indeed, if this were not the case, we would have that at least one $j \in \{1, \ldots, N\}$ such that $\mathbb{1}_{(-\infty,0]}[\text{G}_\alpha(L^*, Z^*, \hat{\xi}^j)] \neq y_j^*$. That situation only happens if $\text{G}_\alpha(L^*, Z^*, \hat{\xi}^j) \leq 0$ and $y_j^* = 0$. However, that is not possible, because in that case, we could define $\tilde{y}$ such that

$$\tilde{y}_m = \begin{cases} y_m^* & \text{if } m \neq j \\ 1 & \text{if } m = j \end{cases},$$

which is a contradiction, because $(L^*, z^*, F^*, O^*, \tilde{y})$ would be a better point that $(L^*, z^*, F^*, O^*, y^*)$. Noting that for each $m \in [N]$, optimality of $(L^*, z^*, F^*, O^*, y^*)$ yields that $Z^{*m} = Q(L^*, \hat{\xi}^m)$, we deduce that (4.26) holds.

To finish, let us suppose that the optimal solution of Problem (4.22) is not L^*, but instead is another solution, $\tilde{L}$. In this case, using our previous development, there would exist $(\tilde{z}, \tilde{F}, \tilde{O}, \tilde{y})$ such that $(\tilde{L}, \tilde{z}, \tilde{F}, \tilde{O}, \tilde{y})$ would be feasible and, thanks to (4.26), it would be a better solution of Problem (4.25). This contradicts the optimality of $(L^*, z^*, F^*, O^*, y^*)$, finishing the proof. □

Remark Up to now, we do not know if Problem (4.22) is a good approximation of Problem (4.17) in the case $c_2 = 1$, that is, when resilience constraints are active. If one wants to apply Theorem 2.26, a variational stability condition must be verified at the optimal solution of Problem (4.17). In our case, we don't know if such a condition holds. Thus, a theoretical study is required, which is a perspective of this work. Nonetheless, the numerical results of the next section are very promising in this line.

4.5 Numerical Experiments

In this section, we show our main numerical results using the SAA approach in order to solve the two-stage problem of the EIP design and operation.

For our simulations, we consider water exchange networks between 4, 10 and 15 agents. For each network, we have to define the inlet and outlet concentrations for every agent, and a *nominal* value of contaminant mass loaded by each participant, $\bar{M}_i$. Using this nominal values as means, we work with three different distributions (two of them symmetrical).

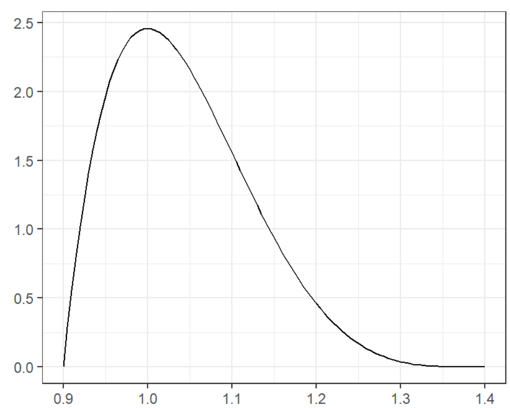

Figure 4.3: Beta Distribution, $a = 2$, $b = 5$.

Specifically, we consider:

1. Uniform Distribution, with range $[a, b] = [0.9\bar{M}_i, 1.1\bar{M}_i]$.
2. Normal Distribution, with $\mu = \bar{M}_i$ and $\sigma = 0.1\bar{M}_i$.
3. Beta Distribution, centered in $\bar{M}_i$ with $a = 2$, $b = 5$.

The Beta Distribution with these chosen parameters, has the shape seen in Figure 4.3.

Agent	C_{in} [mg/L]	C_{out} [mg/L]	Nominal Mass Load (kg/h)
1	0	200	7.00
2	100	500	22.40
3	200	650	62.55
4	0	200	2.00

Table 4.1: First Study Case, $n = 4$ participants.

Agent	C_{in} [mg/L]	C_{out} [mg/L]	Nominal Mass Load (kg/h)
1	25	80	2.00
2	25	90	2.88
3	25	200	4.00
4	50	100	3.00
5	50	800	30.00
6	400	800	5.00
7	400	600	2.00
8	0	100	1.00
9	50	300	20.00
10	150	300	6.50

Table 4.2: Second Study Case, $n = 10$ participants.

Agent	C_{in} [mg/L]	C_{out} [mg/L]	Nominal Mass Load (kg/h)
1	30	100	7.50
2	0	200	6.00
3	50	100	5.00
4	80	800	30.00
5	400	800	4.00
6	20	100	2.50
7	50	100	2.20
8	80	400	5.00
9	100	800	30.00
10	400	1000	4.00
11	30	60	2.00
12	25	50	2.00
13	25	75	5.00
14	50	800	30.00
15	100	900	13.00

Table 4.3: Third Study Case, $n = 15$ participants.

The smaller network is considered in order to compare the results for this approach, with the one proposed by [117] on their previous work. The bigger academic examples considered are the following, based on EIPs suggested by [108] and [6] respectively.

The marginal costs $c = 0.13$ [USD/ton] for consuming fresh water, $d = 0.22$ [USD/ton] for the discharging water and $e = 0.01$ [USD/ton] for using the connections of the park, are considered for every simulation.

All the simulations where implemented in a computer with an Intel Core(TM) i7-10700F processor, running at 2.90GHz, with 16 GB of RAM, running Windows 10 Pro; with Julia v1.6.1 programming language, using Gurobi v9 as solver.

4.5.1 Sensitivity Analysis for the Sample Size

First, we perform a sensitivity analysis of the sample size N, in order to choose a value that ensures a good optimality gap, without increasing too highly the CPU time. For this, we consider different values of N, considering the Beta distribution for every simulation, in order to get a comparable set of data.

Sample Size N	CPU time (s)	Opt.Gap (%)
500	3.64	0.09
1000	14.07	0.08
2000	29.85	0.05
3000	45.49	0.04
4000	66.68	0.04
5000	80.34	0.03
10000	178.75	0.02
20000	419.22	0.02
30000	662.95	0.02

Table 4.4: $n = 4$ agents, increasing sample size, batch size = 30.

Sample Size N	CPU time (s)	Opt.Gap (%)
500	954.97	0.07
1000	5424.30	0.05
2000	10735.23	0.05
3000	16145.14	0.04
4000	21295.41	0.03
5000	26741.33	0.03
10000	135487.75	0.02
20000	248945.71	0.02
30000	389463.57	0.01

Table 4.5: $n = 10$ agents, increasing sample size, batch size = 30.

As we can see considering the two smaller EIPs, fixing $N = 5000$ gives us a good compromise between solution quality and CPU time.

4.5.2 Sensitivity Analysis for α

In this part, we will present a sensitivity analysis for α, which measures the desired costs reduction comparing to the *stand-alone* global operation of the EIP (see equation (4.13)).

It is clear that, if we demand a higher costs reduction, the obtained resilience will decrease, because less tested scenarios will achieve the desired level. In order to show this, we analyze the problem of maximizing resilience, considering cost penalization, for different values of α.

α	Res_α (%)
0.78	0.00
0.79	0.06
0.80	0.98
0.81	7.13
0.82	26.94
0.83	56.99
0.84	89.21
0.85	98.86
0.86	99.99
0.87	100.00

Table 4.6: $n = 4$ agents, increasing α, batch size = 30.

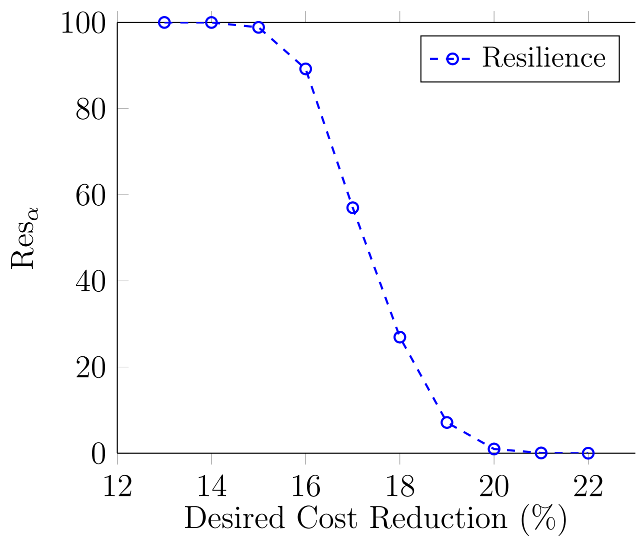

Figure 4.4: Obtained Resilience for EIP, $n = 4$, increasing desired Cost Reduction

α	Res_α (%)
0.64	0.12
0.65	3.00
0.66	15.36
0.67	47.86
0.68	77.86
0.69	94.93
0.70	97.99
0.71	99.21
0.72	99.86
0.73	100.00

Table 4.7: $n = 10$ agents, increasing α, batch size = 30.

We see that, around the value $\alpha = 0.83$ for the smaller EIP, we have a very sensible change for the obtained resilience. This value differs, according to the studied EIP, as we have $\alpha = 0.68$

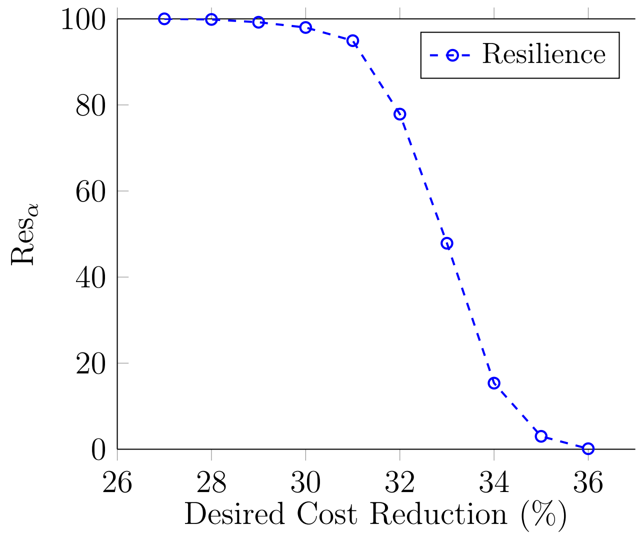

Figure 4.5: Obtained Resilience for EIP, $n = 10$, increasing desired Cost Reduction

α	Res_α (%)
0.60	0.06
0.61	0.85
0.62	1.03
0.63	6.29
0.64	28.71
0.65	62.86
0.66	84.99
0.67	92.87
0.68	98.13
0.69	98.45

Table 4.8: $n = 15$ agents, increasing α, batch size $=$ 30.

for the $n = 10$ EIP critical value and $\alpha = 0.65$ for the bigger one.

4.5.3 Efficiency vs. Resilience

In this section, we compare the results of solving the following two different versions of (4.17).

$$\mathcal{P}_1 = \begin{cases} \min_L & w_1\langle \kappa, L\rangle + w_2\mathbb{E}[Q(L,\xi)] \\ \text{s.t.} & L \geq 0 \end{cases} \tag{4.27}$$

$$\mathcal{P}_2 = \begin{cases} \min_L & w_1\langle \kappa, L\rangle - w_3\text{Res}_\alpha(L) \\ \text{s.t.} & L \geq 0 \end{cases} \tag{4.28}$$

Problem (4.27) corresponds to (4.17) considering $w_3 = 0$ and (4.28) corresponds to (4.17) considering $w_2 = 0$. This allows us to compare the efficiency and the resilience as optimization

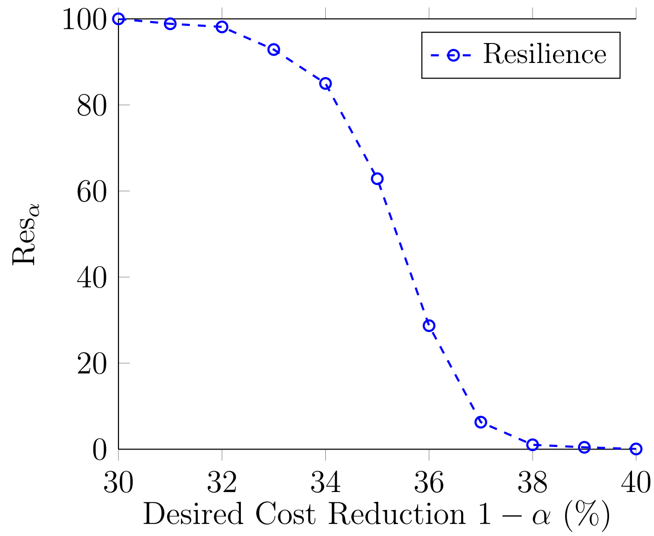

Figure 4.6: Obtained Resilience for EIP, $n = 15$, increasing desired Cost Reduction

criteria.

Both experiments consider penalization over the investment costs term. By doing this, the investment costs are not too high, and actually constructing connections is feasible.

The results are given in the following tables, for the $n = 4$ and $n = 10$ EIPs.

M_i distrib.	α	$Q(L,\xi)$ [ton/h]	Res_α (%)	EIP size	Opt.Gap (%)	CPU time (s)
Beta	0.83	160.61	56.90	84.09	0.03	80.24
Uniform	0.83	154.03	59.26	75.35	0.02	79.78
Normal	0.83	154.01	55.71	82.63	0.03	80.78

Table 4.9: $n = 4$ agents, penalized costs and efficiency, sample = 5000, batch size = 30.

M_i distrib.	α	$Q(L,\xi)$ [ton/h]	Res_α (%)	EIP size	Opt.Gap (%)	CPU time (s)
Beta	0.83	162.44	61.11	63.15	0.04	289.08
Uniform	0.83	155.26	64.36	53.53	0.03	275.64
Normal	0.83	156.05	58.95	62.90	0.03	267.95

Table 4.10: $n = 4$ agents, penalized costs and resilience, sample = 5000, batch size = 30.

M_i distrib.	α	$Q(L,\xi)$ [ton/h]	Res_α (%)	EIP size	Opt.Gap (%)	CPU time (s)
Beta	0.83	175.31	100.00	164.27	0.02	26741.33
Uniform	0.83	175.34	100.00	164.36	0.03	25465.91
Normal	0.83	168.56	100.00	165.33	0.02	24712.66

Table 4.11: $n = 10$ agents, penalized costs and efficiency, sample = 5000, batch size = 30.

4.5.4 Budget Constraints

In this part, our goal is to compare the EIPs designs, when budget constraints are considered, instead of having the investment costs on the objective function. In particular, we are

M_i distrib.	α	$Q(L, \xi)$ [ton/h]	Res_α (%)	EIP size	Opt.Gap (%)	CPU time (s)
Beta	0.83	217.37	100.00	138.75	0.02	90549.33
Uniform	0.83	208.92	100.00	139.36	0.02	89447.31
Normal	0.83	208.44	100.00	141.33	0.03	91290.13

Table 4.12: $n = 10$ agents, penalized costs and resilience, sample = 5000, batch size = 30.

interested in analyze the obtained resilience in two cases: when maximizing efficiency, or maximizing resilience.

In the following tables, we alter the optimal budget obtained in the previous simulations by a percentage of itself, varying from 80% to 120%.

Budget	Res_α (%)
80%	45.48
90%	54.70
95%	56.82
100%	57.06
105%	57.06
110%	57.10
120%	57.10

Table 4.13: $n = 4$ agents, $\alpha = 0.83$, increasing budget.

Budget	Res_α (%)
80%	66.59
90%	74.61
95%	75.32
100%	77.17
105%	77.19
110%	77.22
120%	77.23

Table 4.14: $n = 10$ agents, $\alpha = 0.68$, increasing budget.

Budget	Res_α (%)
80%	53.52
90%	59.16
95%	60.33
100%	62.77
105%	63.19
110%	63.24
120%	63.31

Table 4.15: $n = 15$ agents, $\alpha = 0.65$, increasing budget.

4.5.5 Resilience Constraints

In this last analysis part, our goal is to compare the EIPs designs, when resilience constraints are considered. Particularly, in this case we will be interested in minimizing the investment costs, without maximizing the efficiency of the EIP.

For the maximum efficiency EIP in the $n = 4$ case, we have that the obtained resilience for $\alpha = 0.83$ is $\mathrm{Res}_{0.83} = 61.6\%$, with an investment of 89.1. Considering only the minimization of investment, with increasing asked resilience, we get the following results

Res_α (%)	Budget
53	53.44
54	54.96
55	56.47
56	57.78
57	58.65
58	59.29
59	59.84
60	60.35

Table 4.16: $n = 4$ agents, $\alpha = 0.83$, increasing β.

4.6 Discussion of the results

The results of Section 4.5.2 shows us that, when dealing with resilience, there exists some kind of *critical value* for the parameter α, at which is not possible to build connections between participants that gets us the desired cost reduction.

Looking at the results obtained in Section 4.5.3 we can observe that designing a resilient EIP, instead of designing a more efficient one, is cheaper in terms of investment costs. This makes sense, as we ask for a minimum of *good enough* scenarios instead of trying to maximize the efficiency of the whole EIP when dealing with resilience instead. Nevertheless, the CPU time is higher (in between 3 or 4 times) when dealing with this indicator. The increase in CPU time is expected, since Problem 4.28 is more difficult due to the integer variables $y \in \{0,1\}^N$ used to model the risk measure $\widehat{\mathrm{Res}}_\alpha(L)$.

According to the experiments developed in Section 4.5.4, we see that, past from the total reference budget, we are not able to increase significantly the obtained resilience. For every value of α, we can reach a *limit resilience*, not depending on the considered budget to do so. At the network configuration, we see that, adding more budget to the simulations, gives us bigger, but not more resilient EIPs.

From our point of view, the most important results are the ones obtained in the last set of experiments, as they seem as a very promising result: minimizing investment but asking for a desired level of resilience, gives us important investment deductions, as we will further discuss in the Conclusions of this Thesis.

4.7 Future Work

From this starting point, a lot of work can be done at designing eco-industrial parks that can be able to *endure* different uncertain scenarios, such as the developed here, where *resilience* is defined as an indicator that measures the capability of an EIP to have a minimum level of *good* operating scenarios, looked from an economical point of view.

As a first natural extension of this work, we could include more uncertainty to the model, adding to the operational time horizon, a probability that, at the end of any given period (for example, yearly), a number of participants could *leave* the EIP. It would be desirable that, given this event, the ecopark could still maintain its operations going.

Another extension, considering the works on EIP design by [108, 6], non-cooperative models can be added to the model, e.g., blind-input, control-input or de-regulated exchange markets. In this last scenario, adding *decision uncertainty* between the participants adds a challenging, but also interesting component.

Conclusions

During the development of Chapters 1 and 2, we focused our attention on the main tools that we needed to work with Bilevel and Stochastic Optimization, this being: Bilebel Programming, Two-Stage Problems and Value of Information in this context. The first important point was to enunciate and study the theorems and previous results that we needed to apply in two different problems: the Allocation Problem in Ride-Hailing Platforms (Chapter 3), studying the value of *sharing* information between enterprises and drivers; and the Resilient Design of Eco-Industrial Parks (Chapter 4) using the Two-Stage programming approach, in order to model the EIP design and daily operation. In both problems, we tackle *uncertainty* as a very important property of the studied problems.

In Chapter 3, we presented a new indicator, the *Expected Value of Shared Information*, that allows us to measure the value of sharing information in the context of Stackelberg games. This indicator is relevant in problems where both agents, the leader and the follower, must make their decisions prior to some uncertain event.

As an application, we studied the value of sharing information in the context of ride-hailing companies, considering the demand as the uncertain information. We studied the problem of allocation: prior to the reveal of the demand, the ride-hailing company (leader) decides the spatial prices, while the unmatched drivers (the followers) decide their allocations.

Our numerical results, coming from simulations with randomly generated data, strongly suggests that sharing information is beneficial for the leader. This conclusion is coherent with what is observed nowadays, where ride-hailing companies provide some demand information to the drivers beyond spatial pricing.

As a first work dealing with this new indicator in the context of ride-hailing companies, several simplifying assumptions where made: 1) we studied only the one-stage problem; 2) we simplified the drivers' equilibrium problem into a single welfare optimization problem; 3) we worked with artificially generated data; and 4) we assumed that the leader had access to a perfect forecast of the demand. Motivated by the promising results we obtained here, we aim to improve all these aspects in future works.

In Chapter 4, the results of the last set of experiments are, from our point of view, the most important ones. Here we can conclude that, minimizing investment but asking for a desired level of resilience, we can obtain investment deductions up to a 25%, only losing 1% of resilience.

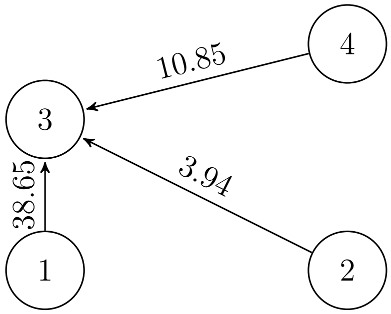

Figure 4.7: EIP at $\mathrm{Res}_\alpha = 53\%$

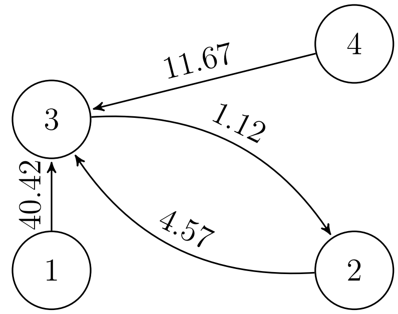

Figure 4.8: EIP at $\mathrm{Res}_\alpha = 56\%$

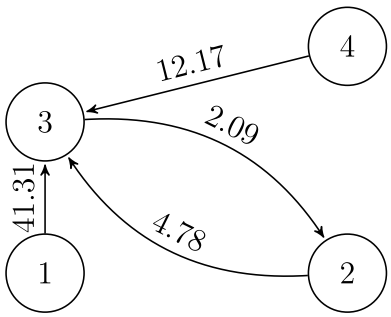

Figure 4.9: EIP at $\mathrm{Res}_\alpha = 59\%$

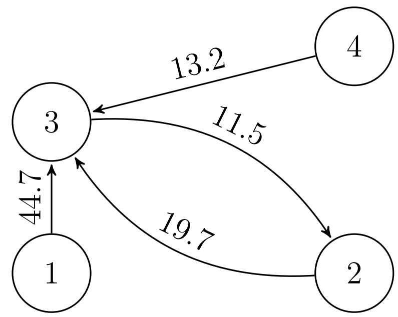

Figure 4.10: Eff. EIP, $\mathrm{Res}_\alpha = 61.6\%$

Considering bigger EIPs, the results are still promising, considering that the investment cost is still reduced by at least 25% in both examples, only losing 2% of resilience. At the $n = 10$ and $n = 15$ EIPs, we also need less connections between participants (downgrade from 40 to 32 at the $n = 10$ EIP, and from 66 to 57 at the $n = 15$ EIP).

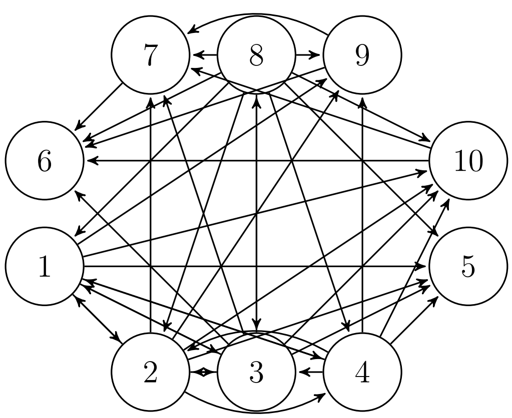

Figure 4.11: Most Efficient $n = 10$ EIP, $\mathrm{Res}_\alpha = 77.17\%$

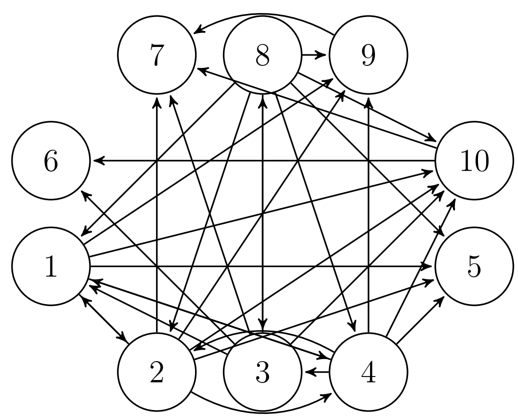

Figure 4.12: $n = 10$ EIP at $\mathrm{Res}_\alpha = 75\%$, minimizing investment.

Bibliography

[1] A. Aboussoror and P. Loridan. Strong-weak Stackelberg problems in finite-dimensional spaces. *Serdica Math. J.*, 21(2):151–170, 1995.

[2] A. Al-Bazi, E. Üney, and A. Abu-Monshar. Developing an overbooking fuzzy-based mathematical optimization model for multi-leg flights. *Transportation Research Procedia*, 43:165–177, 12 2019.

[3] M. Alves, C. Antunes, and J. Costa. New concepts and an algorithm for multiobjective bilevel programming: optimistic, pessimistic and moderate solutions. *Operational Research*, pages 1–34, 2019.

[4] Z. Artstein and R. Wets. Consistency of minimizers and the SLLN for stochastic programs. *J. Convex Anal.*, 2(1-2):1–17, 1995.

[5] J.-P. Aubin and H. Frankowska. *Set-valued analysis*, volume 2 of *Systems & Control: Foundations & Applications*. Birkhäuser Boston, Inc., Boston, MA, 1990.

[6] D. Aussel, K. Cao Van, and D. Salas. Optimal design of exchange water networks with control inputs in eco-industrial parks, 2022. (Preprint arXiv:2204.09863).

[7] S. Balseiro, D. Brown, and C. Chen. Dynamic pricing of relocating resources in large networks. *Management Science*, 2020.

[8] H. Bauschke and P. Combettes. *Convex analysis and monotone operator theory in Hilbert spaces*. CMS Books in Mathematics/Ouvrages de Mathématiques de la SMC. Springer, Cham, second edition, 2017. With a foreword by Hédy Attouch.

[9] E. M. L. Beale. On minimizing a convex function subject to linear inequalities. *Journal of the Royal Statistical Society. Series B (Methodological)*, 17(2):173–184, 1955.

[10] Y. Beck and M. Schmidt. A gentle and incomplete introduction to bilevel optimization, 2021.

[11] Y. Beck and M. Schmidt. A robust approach for modeling limited observability in bilevel optimization. *Operations Research Letters*, (49(5)):752 – 758, 2021.

[12] O. Ben-Ayed and C. E. Blair. Computational Difficulties of Bilevel Linear Programming. *Operations Research*, 38(3):556–560, 1990.

[13] H. Ben-Gal, I. Forma, and G. Singer. A flexible employee recruitment and compensation model: A bi-level optimization approach. *Computers & Industrial Engineering*, 165:107916, 2022.

[14] A. Ben-Tal, T. And, and A. Nemirovski. Robust convex optimization. *Mathematics of Operations Research - MOR*, 23, 11 1998.

[15] A. Ben-Tal, L. Ghaoui, and A. Nemirovski. *Robust Optimization*. Princeton University Press, 08 2009.

[16] S. Benjaafar and M. Hu. Operations management in the age of the sharing economy: what is old and what is new? *Manufacturing & Service Operations Management*, 22(1):93–101, 2020.

[17] D. Bertsimas, D.B. Brown, and C. Caramanis. Theory and applications of robust optimization. *SIAM Review*, 53(3):464–501, 2011.

[18] M. Besançon, M. Anjos, and L. Brotcorne. Complexity of near-optimal robust versions of multilevel optimization problems. *Optimization Letters*, 15:1–14, 11 2021.

[19] J. Bezanson, A. Edelman, S. Karpinski, and V. B. Shah. Julia: A fresh approach to numerical computing. *SIAM review*, 59(1):65–98, 2017.

[20] K. Bimpikis, O. Candogan, and D. Saban. Spatial pricing in ride-sharing networks. *Operations Research*, 67(3):744–769, 2019.

[21] J. R. Birge and F. Louveaux. *Introduction to Stochastic Programming*. Springer Series in Operations Research and Financial Engineering. Springer, July 1997.

[22] M. Boix, L. Montastruc, C. Azzaro-Pantel, and S. Domenech. Optimization methods applied to the design of eco-industrial parks: a literature review. *Journal of Cleaner Production*, 87:303–317, 2015.

[23] M. Boix, L. Montastruc, L. Pibouleau, C. Azzaro-Pantel, and S. Domenech. Industrial water management by multiobjective optimization: from individual to collective solution through eco-industrial parks. *Journal of Cleaner Production*, 22(1):85–97, 2012.

[24] J.F. Bonnans and A. Shapiro. Perturbation analysis of optimization problems. In *Springer Series in Operations Research*, 2000.

[25] H. Brass and K. Petras. Quadrature theory: The theory of numerical integration on a compact interval. 2011.

[26] L. Brotcorne, M. Labbé, P. Marcotte, and G. Savard. A bilevel model for toll optimization on a multicommodity transportation network. *Transportation Science*, 35:345–358, 11 2001.

[27] J. Burtscheidt and M. Claus. Bilevel linear optimization under uncertainty. In S. Dempe and A. Zemkoho, editors, *Bilevel Optimization: Advances and Next Challenges*, pages

485–511. Springer International Publishing, Cham, 2020.

[28] H. Calvete, C. Galé, J. Iranzo, and P. Mateo. A decision tool based on bilevel optimization for the allocation of water resources in a hierarchical system. *International Transactions in Operational Research*, 02 2021.

[29] A. Charnes and W. W. Cooper. Chance-constrained programming. *Management Science*, 6(1):73–79, 1959.

[30] S. Christiansen, M. Patriksson, and L. Wynter. Stochastic bilevel programming in structural optimization. *Structural and Multidisciplinary Optimization*, 21:361–371, 07 2001.

[31] G. Dantzig. Linear programming under uncertainty. *Management Science*, 1(3-4):197–206, 1955.

[32] S. Dempe. *Foundations of bilevel programming.* Springer Science & Business Media, 2002.

[33] S. Dempe and J. Dutta. Is bilevel programming a special case of a mathematical program with complementarity constraints? *Math. Program.*, 131(1-2, Ser. A):37–48, 2012.

[34] S. Dempe, V. Kalashnikov, G. A. Pérez-Valdés, and N. Kalashnykova. *Bilevel programming problems. Theory, algorithms and applications to energy networks*. Energy Systems. Springer, Heidelberg, 2015.

[35] D. Dentcheva, A. Prékopa, and A. Ruszczyński. Concavity and efficient points of discrete distributions in probabilistic programming. *Mathematical Programming, Series B*, 89:55–77, 01 2000.

[36] D. Dentcheva and A. Ruszczyński. *Portfolio optimization with stochastic dominance constraints.* 09 2003.

[37] A. Di Pretoro, F. D'Iglio, and F. Manenti. Optimal cleaning cycle scheduling under uncertain conditions: A flexibility analysis on heat exchanger fouling. *Processes*, 9(1), 2021.

[38] A Di Pretoro, L. Montastruc, F. Manenti, and X. Joulia. Flexibility analysis of a distillation column: Indexes comparison and economic assessment. *Computers & Chemical Engineering*, 124:93–108, 2019.

[39] J. Dupačová. Applications of stochastic programming: Achievements and questions. *European Journal of Operational Research*, 140:281–290, 02 2002.

[40] J.-P. Dussault, M. Haddou, and T. Migot. The new butterfly relaxation method for mathematical programs with complementarity constraints. In *Optimization, variational analysis and applications*, volume 355 of *Springer Proc. Math. Stat.*, pages 35–67. Springer, Singapore, [2021] ©2021.

[41] O. El Housni, V. Goyal, O. Hanguir, and C. Stein. Matching drivers to riders: A two-stage robust approach, 2021. arXiv:2011.03624.

[42] A. Ferguson and G. Dantzig. The allocation of aircraft to routes-an example of linear programming under uncertain demand. *Management Science*, 3(1):45–73, 1956.

[43] J. Fortuny-Amat and B. McCarl. A representation and economic interpretation of a two-level programming problem. *Journal of the Operational Research Society*, 32(9):783–792, 1981.

[44] J. Fortuny-Amat and B. McCarl. A representation and economic interpretation of a two-level programming problem. *J. Oper. Res. Soc.*, 32(9):783–792, 1981.

[45] M. Fukushima, Z.-Q. Luo, and J.-S. Pang. A globally convergent sequential quadratic programming algorithm for mathematical programs with linear complementarity constraints. *Comput. Optim. Appl.*, 10(1):5–34, 1998.

[46] R. Garrido, P. Lamas, and F. J. Pino. A stochastic programming approach for floods emergency logistics. *Transportation Research Part E: Logistics and Transportation Review*, 75:18–31, 2015.

[47] J. Gauvin and F. Dubeau. *Differential properties of the marginal function in mathematical programming*, pages 101–119. Springer Berlin Heidelberg, Berlin, Heidelberg, 1982.

[48] R. Giahi, C. MacKenzie, and C. Hu. Design optimization for resilience for risk-averse firms. *Computers & Industrial Engineering*, 139:106122, 2020.

[49] N. Gröwe-Kuska, K. Kiwiel, M. Nowak, W. Roemisch, and I. Wegner. *Power Management in a Hydro-Thermal System under Uncertainty by Lagrangian Relaxation*, pages 39–70. 01 2002.

[50] LLC Gurobi Optimization. Gurobi optimizer reference manual, 2021.

[51] D. Han and J. H. Lee. Two-stage stochastic programming formulation for optimal design and operation of multi-microgrid system using data-based modeling of renewable energy sources. *Applied Energy*, 291:116830, 2021.

[52] L. He, Z. Hu, and M. Zhang. Robust repositioning for vehicle sharing. *Manufacturing & Service Operations Management*, 22(2):241–256, 2020.

[53] R. Henrion, P. Li, A. Moller, M. Steinbach, M. Wendt, and G. Wozny. Stochastic optimization for operating chemical processes under uncertainty. *Online Optimization of Large Scale Systems*, 01 2001.

[54] R. Hochreiter and G. Pflug. Financial scenario generation for stochastic multi-stage decision processes as facility location problems. *Annals OR*, 152:257–272, 03 2007.

[55] K. Holmberg. *Stochastic transportation and location problemsStochastic Transportation*

and Location Problems, pages 2546–2551. 01 2001.

[56] T. Homem-de Mello and G. Bayraksan. Monte carlo sampling-based methods for stochastic optimization. *Surv. Oper. Res. Manag. Sci.*, 19(1):56–85, 2014.

[57] M. Hu and M. Fukushima. Multi-leader-follower games: models, methods and applications. *J. Oper. Res. Soc. Japan*, 58(1):1–23, 2015.

[58] X. M. Hu and D. Ralph. Convergence of a penalty method for mathematical programming with complementarity constraints. *J. Optim. Theory Appl.*, 123(2):365–390, 2004.

[59] X. M. Hu and D. Ralph. Using EPECs to model bilevel games in restructured electricity markets with locational prices. *Oper. Res.*, 55(5):809–827, 2007.

[60] T. Ichiishi. *Game theory for economic analysis*. Economic Theory, Econometrics, and Mathematical Economics. Academic Press, Inc. [Harcourt Brace Jovanovich, Publishers], New York, 1983.

[61] M. O. Jackson. *Mechanism Theory*. 2014. Available at SSRN: https://ssrn.com/abstract=2542983 or http://dx.doi.org/10.2139/ssrn.2542983.

[62] S. Jia, Z. Wan, Y. Feng, and G. Wang. New partial cooperation model for bilevel programming problems. *Journal of Systems Engineering and Electronics*, 22(2):263–266, 2011.

[63] R. Kadadevaramath, J. Chen, B. Latha, and K. Rameshkumar. Application of particle swarm intelligence algorithms in supply chain network architecture optimization. *Expert Syst. Appl.*, 39(11):10160–10176, sep 2012.

[64] P. Kall and S.W. Wallace. *Stochastic Programming*. Wiley Interscience Series in Systems and Optimization. Wiley, 1995.

[65] T. Kleinert, M. Labbé, F. Plein, and M. Schmidt. Technical note—there's no free lunch: on the hardness of choosing a correct big-M in bilevel optimization. *Oper. Res.*, 68(6):1716–1721, 2020.

[66] T. Kleinert and M. Schmidt. Why there is no need to use a big-m in linear bilevel optimization: A computational study of two ready-to-use approaches, 2020. Preprint at optimization-online.org: 2020-10-8065.

[67] A. Kleywegt, A. Shapiro, and T. Homem-de Mello. The sample average approximation method for stochastic discrete optimization. *SIAM Journal on Optimization*, 12(2):479–502, 2002.

[68] X. Kong, D. Wang, Y. Wang, and N. Wang. A stochastic programming method for resources allocation in education of higher vocational college. *Journal of Physics: Conference Series*, 1637:012090, 09 2020.

[69] M. Lejeune and A. Ruszczyński. An efficient trajectory method for probabilistic production-inventory-distribution problems. *Operations Research*, 55:378–394, 04 2007.

[70] Y. Li, G. Huang, and X. Chen. Multistage scenario-based interval-stochastic programming for planning water resources allocation. *Stochastic Environmental Research and Risk Assessment*, 23:781–792, 08 2009.

[71] Y. Li, G. Huang, Y. Huang, and H. Zhou. A multistage fuzzy-stochastic programming model for supporting sustainable water-resources allocation and management. *Environmental Modelling & Software*, 24(7):786–797, 2009.

[72] J. Liu, Y. Fan, Z. Chen, and Y. Zheng. Pessimistic bilevel optimization: A survey. *International Journal of Computational Intelligence Systems*, 11:725–736, 2018.

[73] J. Liu, Y. Fan, Z. Chen, and Y. Zheng. *Methods for Pessimistic Bilevel Optimization*, pages 403–420. Springer International Publishing, Cham, 2020.

[74] M. Liu, B. Dan, S. Zhang, and S. Ma. Information sharing in an e-tailing supply chain for fresh produce with freshness-keeping effort and value-added service. *European Journal of Operational Research*, 290(2):572–584, 2021.

[75] F. Louveaux and J. Birge. *Two-stage stochastic programs with recourse*, pages 3959–3961. Springer US, Boston, MA, 2009.

[76] E. Lowe. *Eco-industrial Park Handbook for Asian Developing Countries Report to Asian Development Bank*. 10 2001.

[77] J. Lu, G. Feng, S Shum, and K. Lai. On the value of information sharing in the presence of information errors. *European Journal of Operational Research*, 294(3):1139–1152, 2021.

[78] J. Luedtke and S. Ahmed. A sample approximation approach for optimization with probabilistic constraints. *SIAM Journal on Optimization*, 19:674–699, 06 2008.

[79] J. Luedtke, S. Ahmed, and G. Nemhauser. An integer programming approach for linear programs with probabilistic constraints. volume 122, pages 247–272, 01 2010.

[80] Z.-Q. Luo, J.-S. Pang, and D. Ralph. *Mathematical Programs with Equilibrium Constraints*. Cambridge University Press, 1996.

[81] Lyft Engineering. Making cohort-based long-term forecasts at lyft. `https://eng.lyft.com/making-long-term-forecasts-at-lyft-fac475b3ba52`, 2019. (accessed on November 29th, 2021).

[82] W-K. Mak, P. Morton, and R.K. Wood. Monte carlo bounding techniques for determining solution quality in stochastic programs. *Operations Research Letters*, 24(1):47–56, 1999.

[83] L. Mallozzi and J. Morgan. *Hierarchical Systems with Weighted Reaction Set*, pages

271–282. Springer US, Boston, MA, 1996.

[84] H. Markowitz. Portfolio selection. *The Journal of Finance*, 7(1):77–91, 1952.

[85] L. Mellouk, M. Boulmalf, A. Aaroud, K. Zine-Dine, and D. Benhaddou. Genetic algorithm to solve demand side management and economic dispatch problem. *Procedia Computer Science*, 130:611–618, 2018. The 9th International Conference on Ambient Systems, Networks and Technologies (ANT 2018) / The 8th International Conference on Sustainable Energy Information Technology (SEIT-2018) / Affiliated Workshops.

[86] B. Miller and H. Wagner. Chance constrained programming with joint constraints. *Operations Research*, 13(6):930–945, 1965.

[87] L. Montastruc, M. Boix, L. Pibouleau, C. Azzaro-Pantel, and S. Domenech. On the flexibility of an eco-industrial park (eip) for managing industrial water. *Journal of Cleaner Production*, 43:1–11, 2013.

[88] B.S. Mordukhovich. *Variational Analysis and Generalized Differentiation II: Applications*. Grundlehren der mathematischen Wissenschaften. Springer Berlin Heidelberg, 2006.

[89] B.S. Mordukhovich. *Variational Analysis and Generalized Differentiation I: Basic Theory*. Grundlehren der mathematischen Wissenschaften. Springer Berlin Heidelberg, 2012.

[90] B.S. Mordukhovich. Variational analysis of marginal functions with applications to bilevel programming. *J. Optimization Theory and Applications*, 152:557–586, 03 2012.

[91] M. Murr and A. Prékopa. Solution of a product substitution problem using stochastic programming. 01 2000.

[92] T. Müller-Gronbach, E. Novak, and K. Ritter. *Monte Carlo-Algorithmen.* Springer Berlin, Heidelberg, 20129.

[93] A. Nemirovski and A. Shapiro. Convex approximations of chance constrained programs. *SIAM Journal on Optimization*, 17:969–996, 01 2006.

[94] V. Norkin, G. Pflug, and A. Ruszczyński. A branch and bound method for stochastic global optimization. *Mathematical Programming, Series B*, 83:425–450, 11 1998.

[95] W. Oberkampf, S. DeLand, B. Rutherford, K. V. Diegert, and K. Alvin. Error and uncertainty in modeling and simulation. *Reliability Engineering & System Safety*, 75(3):333–357, 2002.

[96] W. Oberkampf, J. Helton, and K. Sentz. Mathematical representation of uncertainty. *Proceedings of the 19th AIAA Applied Aerodynamics Conference*, 2012.

[97] B. Pagnoncelli, S. Ahmed, and A. Shapiro. Sample average approximation method for chance constrained programming: Theory and applications. *Journal of Optimization*

Theory and Applications, 142:399–416, 08 2009.

[98] E. Pistikopoulos and T. Mazzuchi. A novel flexibility analysis approach for processes with stochastic parameters. *Computers & Chemical Engineering*, 14(9):991–1000, 1990.

[99] W. B. Powell and H. Topaloglu. Stochastic programming in transportation and logistics. In *Stochastic Programming*, volume 10 of *Handbooks in Operations Research and Management Science*, pages 555–635. Elsevier, 2003.

[100] A. Prékopa. *Stochastic Programming.* Mathematics and Its Applications. Springer Netherlands, 2013.

[101] H. Raiffa and R. Schlaifer. *Applied statistical decision theory.* Harvard University Graduate School of Business Administration (Division of Research); Bailey & Swinfen., 1961.

[102] M. Ramos, M. Boix, D. Aussel, L. Montastruc, and S. Domenech. Water integration in eco-industrial parks using a multi-leader-follower approach. *Computers and Chemical Engineering*, 87:190–207, 2016.

[103] M. Riis and J. Lodahl. A bicriteria stochastic programming model for capacity expansion in telecommunications. *Mathematical Methods of Operations Research*, 56:83–100, 08 2002.

[104] M. Riis, A. Skriver, and J. Lodahl. Deployment of mobile switching centers in a telecommunications network: A stochastic programming approach. *Telecommunication Systems*, 26:93–109, 05 2004.

[105] R.T. Rockafellar. Marginal values and second-order necessary conditions for optimality. *Math. Program.*, 26:245–286, 1983.

[106] R.T. Rockafellar and R. Wets. *Variational Analysis.* Springer Verlag, Heidelberg, Berlin, New York, 1998.

[107] H. Ruud, M. Zied Babai, J. Bokhorst, and A. Syntetos. Revisiting the value of information sharing in two-stage supply chains. *European Journal of Operational Research*, 270(3):1044–1052, 2018.

[108] D. Salas, K. Cao Van, D. Aussel, and L. Montastruc. Optimal design of exchange networks with blind inputs and its application to eco-industrial parks. *Computers & Chemical Engineering*, 143:107053, 2020.

[109] D. Salas and A. Svensson. Existence of solutions for deterministic bilevel games under a general bayesian approach, 2020.

[110] A. Shapiro. Monte carlo sampling methods. *Handbooks in Operations Research and Management Science*, 10, 12 2003.

[111] A. Shapiro, D. Dentcheva, and A. Ruszczyński. *Lectures on Stochastic Programming*,

volume 9 of *MOS-SIAM Series on Optimization*. Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA; Mathematical Optimization Society, Philadelphia, PA, second edition, 2014. Modeling and theory.

[112] T. Sharkey, B. Cavdaroglu, H. Nguyen, J. Holman, J. Mitchell, and J. Wallace. Interdependent network restoration: On the value of information-sharing. *European Journal of Operational Research*, 244(1):309–321, 2015.

[113] L. Sun, R. Teunter, M. Babai, and G. Hua. Optimal pricing for ride-sourcing platforms. *European Journal of Operational Research*, 278(3):783–795, 2019.

[114] Z. Sun, A. Hupman, and A. Abbas. The value of information for price dependent demand. *European Journal of Operational Research*, 288(2):511–522, 2021.

[115] R. Swaney and I. Grossmann. An index for operational flexibility in chemical process design. part i: Formulation and theory. *Aiche Journal*, 31:621–630, 1983.

[116] A. Takyi and B. Lence. Surface water quality management using a multiple-realization chance constraint method. *Water Resources Research*, 35(5):1657–1670, 1999.

[117] R. Tan and D. Cruz. Synthesis of robust water reuse networks for single-component retrofit problems using symmetric fuzzy linear programming. *Computers & Chemical Engineering*, 28(12):2547–2551, 2004.

[118] M. Tripathy, J. Bai, and H. Sebastian (Seb) Heese. Driver collusion in ride-hailing platforms. *Decision Sciences*, n/a(n/a).

[119] Uber Engineering. Forecasting at uber: An introduction. `https://eng.uber.com/forecasting-introduction/`, 2018. (accessed on November 29th, 2021).

[120] Uber Technologies Inc. Introducing the new driver app, your partner on the road. `https://www.uber.com/cl/en/drive/driver-app/`, 2021. (accessed on November 29th, 2021).

[121] G. van Ryzin and R. H. Smith. Coordinating overbooking and capacity control decisions on a network. 2004.

[122] B. Verweij, S. Ahmed, A. Kleywegt, G. Nemhauser, and A. Shapiro. The sample average approximation method applied to stochastic routing problems: A computational study. *Computational Optimization and Applications*, 24:289–333, 02 2003.

[123] H. von Stackelberg. *Marktform und Gleichgewicht*. Die Handelsblatt-Bibliothek "Klassiker der Nationalökonomie". J. Springer, 1934.

[124] S.W. Wallace, W.T. Ziemba, Society for Industrial, and Applied Mathematics. *Applications of Stochastic Programming*. MPS-SIAM Series on Optimization. Society for Industrial and Applied Mathematics, 2005.

[125] H. Wang and H. Yang. Ridesourcing systems: A framework and review. *Transportation*

Research Part B: Methodological, 129:122–155, 2019.

[126] W. Wiesemann, A. Tsoukalas, P. Kleniati, and B. Rustem. Pessimistic bilevel optimization. *SIAM Journal on Optimization*, 23(1):353–380, 2013.

[127] R. Wijekoon. Integration of non-conventional renewable energy based electricity generation into sri lanka grid – generation planning perspective. 10 2016.

[128] A. C. Williams. A stochastic transportation problem. *Operations Research*, 11(5):759–770, 1963.