Submodularity for Distributionally Robust Optimization

Supervisors: Prof. Angelos Gheorghiou and Prof. Rosario Paradiso

Gianluca Covini

October – December 2024

Abstract

This report explores recent advances in the integration of submodular functions within the framework of Distributionally Robust Optimization (DRO), a powerful approach for decision-making under uncertainty. After reviewing the theoretical foundations of DRO and submodularity, we focus on conditions under which their combination yields tractable formulations. In particular, we examine how structural properties of submodular functions—such as diminishing returns and analogies to convexity—can be leveraged to design DRO models that are solvable in polynomial time, even in the presence of complex ambiguity sets. The report summarizes key results from the literature and discusses their implications in real-world scenarios such as sensor placement and influence maximization. We conclude by outlining open questions and proposing new research directions that build on the synergy between DRO and submodular optimization.

Introduction

In the following report, we summarize the work done during a short internship conducted remotely under the supervision of Prof. Angelos Georghiou and Prof. Rosario Paradiso. The project consisted in a further study of recent results that exploit the properties of submodular functions to derive tractable formulations for distributionally robust optimization. In the end of this report, we suggest possible applications to sensor placement, leaving the details for further research.

Motivations

Optimization under uncertainty seeks to model decision-making processes more realistically than deterministic formulations, which often overlook variability in real-world systems. This is particularly relevant in fields such as energy systems—where future demand is uncertain—and logistics, where travel times and costs are often unpredictable. Two classical approaches to handling such uncertainty are Stochastic Optimization (SO), which minimizes the expected cost, and Robust Optimization (RO), which minimizes the worst-case cost.

Distributionally Robust Optimization (DRO) is a more recent and promising framework that lies between these two paradigms. Instead of relying on a single probability distribution (as in SO) or the worst-case realization (as in RO), DRO models uncertainty via an ambiguity set—a family of plausible probability distributions—and seeks to minimize the worst-case expected cost across this set. This yields a more nuanced and balanced approach, which is particularly appealing in situations where the true distribution is not precisely known but can be estimated or bounded.

Despite its modeling power, DRO problems are generally hard to solve. Optimizing over sets of distributions is computationally intractable in the general case, and tractability can only be achieved under specific structural assumptions—such as particular choices of ambiguity sets or objective functions.

One such structural property that has attracted increasing attention is submodularity. A submodular function exhibits diminishing returns: the incremental benefit of adding an element to a set decreases as the set grows. This property arises naturally in many real-world settings, such as sensor placement—where adding a sensor yields less marginal information if others are already placed—and cost-sharing scenarios. Moreover, submodular functions share key similarities with convex functions in terms of tractability: for instance, minimization of submodular functions over discrete domains can often be performed efficiently.

This makes the intersection between DRO and submodularity a particularly promising research direction. By combining the modeling robustness of DRO with the structural properties of submodular functions, one can obtain tractable formulations of complex decision problems under uncertainty. Such formulations can better capture the nature of real-world systems while remaining computationally feasible, offering both theoretical insights and practical impact.

Contributions

This report investigates the conditions under which the submodularity of the objective function yields tractable formulations of distributionally robust optimization (DRO) problems. The goal is to provide a systematic understanding of how structural assumptions—motivated by real-world applications—can be exploited to obtain efficient algorithms for otherwise intractable problems. The wide range of problems covered in the literature highlights the broad applicability and potential impact of this line of research.

Our contribution is twofold. First, we present a comprehensive and self-contained review of recent theoretical results at the intersection of submodular optimization and DRO. Each result is introduced in a general probabilistic-optimization framework and then illustrated through its main applications. Second, we propose potential research directions, including an original formulation for sensor placement under uncertainty, aimed at integrating submodularity into robust decision-making models.

The structure of this report is as follows. Chapter 1 reviews the probabilistic and optimization background, including detailed definitions of DRO formulations and submodularity. Chapter 2 presents the main tractability results from the literature, alongside key applications. Chapter 3 explores new applications, with a focus on sensor placement. Finally, Chapter 4 summarizes the findings and outlines possible directions for future research.

Chapter 1

Background

1.1 Distributionally Robust Optimization

1.1.1 Formulation

The problems we studied are Distributionally Robust Optimization (DRO) problems. These consist of a general class of optimization problems whose objective function includes a stochastic term as a random variable $\tilde{\boldsymbol{\xi}}$ whose distribution is unknown but considered to be in an ambiguity set $\mathcal{P}$. DRO aims to minimize with respect to a decision variable x the expected objective for the worst-case distribution $\mathbb{P} \in \mathcal{P}$.

Formally, a DRO problem can be formulated as

$$\inf_{\boldsymbol{x} \in \mathcal{X}} \sup_{\mathbb{P} \in \mathcal{P}} \mathbb{E}_{\mathbb{P}}[f(\boldsymbol{x}, \tilde{\boldsymbol{\xi}})] \tag{1.1}$$

where

- $\boldsymbol{x}$ is a decision vector in a set $\mathcal{X} \subseteq \mathbb{R}^N$;
- $\mathbb{P}$ is a probability measure from a set called *ambiguity set* $\mathcal{P}$;
- $\tilde{\boldsymbol{\xi}}$ is a random vector distributed as $\mathbb{P}$.

The problem consists in choosing $\boldsymbol{x}$ such that it minimizes the worst-case expectation of a random cost function f, whose randomness is contained in the random vector $\tilde{\boldsymbol{\xi}}$.

The generality of the DRO framework comes from the fact that the two main paradigms for optimization under uncertainty, i.e. Stochastic Optimization (SO) and Robust Optimization (RO), can be derived as DRO with special choice of ambiguity set $\mathcal{P}$. In particular, if the ambiguity set consists only of one element $\mathcal{P} = \{\mathbb{P}\}$ the problem simplifies to

$$\inf_{\boldsymbol{x} \in \mathcal{X}} \mathbb{E}_{\mathbb{P}}[f(\boldsymbol{x}, \tilde{\boldsymbol{\xi}})] \tag{1.2}$$

which corresponds to a *stochastic optimization* (SO) problem.

Another speical case for 1.1 occurs when $\mathcal{P} = \mathcal{P}(\Xi)$, i.e. $\mathcal{P}$ is the set of the probability measures $\mathbb{P}$ whose support is a closed bounded set Ξ. In this case we know that $\tilde{\boldsymbol{\xi}} \in \Xi$, so the worst case correspond to the probability measure that gives all mass to the worst-case realization in Ξ. The resulting problem is a *robust optimization* (RO) problem:

$$\inf_{\boldsymbol{x} \in \mathcal{X}} \max_{\boldsymbol{\xi} \in \Xi} f(\boldsymbol{x}, \boldsymbol{\xi}) \tag{1.3}$$

1.1.2 Complexity Results

The previous formulation gives a general framework for many practical problems. We are now interested in studying the computational complexity of the previous formulations. To do so, we will use the following common notation for problem complexity:

- **P**: is the class of problems solvable in a polynomial time in the number of inputs;
- **NP**: is the class of problems for which a solution can be verified in polynomial time;
- **#P**: is the class of problems whose task is to count the number of solutions to a problem which is NP;
- **#P − hard**: is the class of problems at least as difficult as any other problem in $\#P$;
- **#P − complete**: is the class of problems which are both in $\#P$ and $\#P - hard$.

The aim of this report is to present situations in which the DRO problem is tractable, i.e. solvable in polynomial time. We will present in the following section tractable formulations in relation with the use of submodular functions.

Firstly, we present the main known results not connected with the use a submodular or supermodular objective.

For the SO problem, it is solvable efficiently in the case of the following:

Proposition 1. *If $\mathcal{X}$ is a polyhedral and $\mathbb{P}$ is discrete, 1.2 can be reformulated as a linear program in size that grows linearly with the number of realization of $\tilde{\boldsymbol{\xi}}$.*

Proof. The SO problem can be rewritten in terms of realizations of the random vector $\tilde{\boldsymbol{\xi}}$, called *scenario*, as follows

$$\inf_{\boldsymbol{x}\in\mathcal{X}} \sum_{\boldsymbol{\xi} \text{ scenario}} f(\boldsymbol{x},\boldsymbol{\xi})\mathbb{P}[\tilde{\boldsymbol{\xi}} = \boldsymbol{\xi}]$$

where $\boldsymbol{\xi}$ are single realizations of $\tilde{\boldsymbol{\xi}}$. □

The most complex case is when all the components of $\tilde{\boldsymbol{\xi}}$ are independent because it leads to a number of scenario exponential in the dimension N. Therefore

Proposition 2. *Computing the expected cost in 1.2 is #P-hard when the r.v. are mutually independent. Thus also the SO problem 1.2 is #P-hard when the r.v. are mutually independent. This is true also if $\tilde{\boldsymbol{\xi}}$ is continuous.*

The result can be found in Hanasusanto et al. 2016.

The problem can also be approximated by a sample average:

$$\inf_{\boldsymbol{x}\in\mathcal{X}} \sum_{\substack{\boldsymbol{\xi} \text{ sampled} \\ \text{scenario}}} f(\boldsymbol{x},\boldsymbol{\xi})\,\mathbb{P}[\tilde{\boldsymbol{\xi}} = \boldsymbol{\xi}]$$

Regarding the RO problem, a tractability case is expressed in the following:

Proposition 3. *If $\mathcal{X}$ and Ξ are polyhedral, and f is affine in both $\boldsymbol{x}$ and $\boldsymbol{\xi}$, the problem 1.3 is solvable using linear optimization.*

Proof. Since we have an affine f, the problem is analogous to

$$\inf_{\boldsymbol{x}\in\mathcal{X}} \max_{\boldsymbol{\xi}\in\Xi} \mathbf{a}_k^t(\boldsymbol{x})\boldsymbol{\xi} + b_k(\boldsymbol{x}) \tag{1.4}$$

It can be rewritten in linear formulation, where the $\max_{\boldsymbol{\xi}\in\Xi}$ appears in the constraints. This maximization problem in the constraints can be reformulated using duality and be integrated as a constraint on each entry. For details, look at page 10 of Bertsimas et al. 2010.

The result is a linear program with a polynomial number of constraints and variables. □

There exist also other tractability results for DRO problems with special ambiguity sets. The first is the marginal distribution ambiguity set, defined as follows.

Definition 4. We call *marginal distribution ambiguity set*, the Frechét set $\mathcal{P}(\mathbb{P}_1, \ldots, \mathbb{P}_N)$ of N marginal distributions $\mathbb{P}_1, \ldots, \mathbb{P}_N$, where $(\mathbb{P}_i) = \Xi_i$.

Proposition 5. *For marginal distribution ambiguity set, if $\mathcal{X}$ is a polyhedron, marginals discrete, with finite support, f affine in both variables, then the DRO problem 1.1 is solvable using a polynomial size linear program.*

Proposition 6. *For marginal distribution ambiguity set, if $\mathcal{X}$ is a polyhedron, marginals continuous, f affine in both variables, then the DRO problem 1.1 is solvable using a convex program.*

Note that this ambiguity set contains the case of independent random variables which in this case is not #P-hard, but polynomial.

Another tractable ambiguity set is the following.

Definition 7. We call *moment ambiguity set*, the set $\mathcal{P}$ of probabilities with $\mathbb{E}_{\mathbb{P}}[\tilde{\boldsymbol{\xi}}] = \mu$ and $\mathbb{E}_{\mathbb{P}}[\tilde{\boldsymbol{\xi}}\tilde{\boldsymbol{\xi}}^t] = \mathbf{Q}$, with $\mathbf{Q}$ and μ fixed such as $\mathbf{Q} \succeq \mu\mu^{\mathbf{t}}$.

Proposition 8. *For moment ambiguity set, if supp$(\mathbb{P}) = \mathbb{R}^N$ or is an ellipsoid, and f is affine, the DRO problem 1.1 is solvable in polynomial time.*

Proposition 9. *For moment ambiguity set, if supp$(\mathbb{P})$ is a polyhedron, the DRO problem 1.1 is NP-hard to solve.*

Important ambiguity sets that we refer in the following sections are the *statistical distance based*. We do not provide complexity results but we present the definitions.

Definition 10. A Φ-*divergence* between two probability measures $\mathbb{P}$ and $\mathbb{Q}$ defined over Ω is

$$D_\Phi(\mathbb{P}||\mathbb{Q}) = \int_\Omega \Phi(\frac{d\mathbb{P}}{d\mathbb{Q}}(\omega))d\mathbb{Q}(\omega) \tag{1.5}$$

Two special cases are:

Definition 11. A *Kullback-Leibler divergence* between two probability measures $\mathbb{P}$ and $\mathbb{Q}$ is

$$D_{KL}(\mathbb{P}||\mathbb{Q}) = \int_\Omega \log(\frac{d\mathbb{P}}{d\mathbb{Q}}(\omega))d\mathbb{P}(\omega) \tag{1.6}$$

Definition 12. A χ^2 *divergence* between two probability measures $\mathbb{P}$ and $\mathbb{Q}$ is

$$D_{\chi^2}(\mathbb{P}||\mathbb{Q}) = \frac{1}{2}\int_{\Omega}(\frac{d\mathbb{P}}{d\mathbb{Q}}(\omega) - 1)^2 d\mathbb{Q}(\omega) \tag{1.7}$$

In general we can build

Definition 13. A Φ *divergence uncertainty set* around an empirical distribution $\hat{\mathbb{P}}_n$ is

$$\mathcal{P}_{\rho,n} := \{\mathbb{P} : D_{\chi^2}(\mathbb{P}||\hat{\mathbb{P}}_n) \leq \frac{\rho}{n}\} \tag{1.8}$$

where Δ_n is the n-dim simplex.

If $\hat{\mathbb{P}}_n$ corresponds to an empirical sample $Z_1, \ldots, Z_n$, the probabilities $\mathbb{P} \in \mathcal{P}_{\rho,n}$ correspond to a vector $\mathbf{p} \in \Delta_n$ and thus we can equivalently define

$$\mathcal{P}_{\rho,n} := \{\mathbf{p} \in \Delta_n : \frac{1}{2}\|n\mathbf{p} - \mathbf{1}\|_2^2\} \tag{1.9}$$

In section 2.4, we will use the following

Definition 14. A χ^2 *uncertainty set* around an empirical distribution $\hat{\mathbb{P}}_n$ is

$$\mathcal{P}_{\rho,n} := \{\mathbb{P} : D_{\chi^2}(\mathbb{P}||\hat{\mathbb{P}}_n) \leq \frac{\rho}{n}\} \tag{1.10}$$

where Δ_n is the n-dim simplex.

If $\hat{\mathbb{P}}_n$ corresponds to an empirical sample $Z_1, \ldots, Z_n$, the probabilities $\mathbb{P} \in \mathcal{P}_{\rho,n}$ correspond to a vector $\mathbf{p} \in \Delta_n$ and thus we can equivalently define

$$\mathcal{P}_{\rho,n} := \{\mathbf{p} \in \Delta_n : \frac{1}{2}\|n\mathbf{p} - \mathbf{1}\|_2^2\} \tag{1.11}$$

Definition 15. The *Wasserstein metric* between two probability measures $\mathbb{P}_1$ and $\mathbb{P}_2$ defined over Ξ is

$$d_W(\mathbb{P}_1, \mathbb{P}_2) := \inf\{\int_{\Xi}^{2} \|\tilde{\boldsymbol{\xi}}_1 - \tilde{\boldsymbol{\xi}}_2\|\mathbb{P}(\tilde{\boldsymbol{\xi}}_1, \tilde{\boldsymbol{\xi}}_2) : \mathbb{P} \in \mathcal{P}(\mathbb{P}_1, \mathbb{P}_2) \text{ Fréchet set}\} \tag{1.12}$$

where $\|\cdot\|$ is an arbitrary norm on $\mathbb{R}^N$

We therefore define the ambiguity set

Definition 16. A *Wasserstein ambiguity set* around an empirical distribution $\hat{\mathbb{P}}_n$ is

$$\mathcal{P}_{\rho,n} := \{\mathbb{P} : d_W(\mathbb{P}, \hat{\mathbb{P}}_n) \leq \frac{\rho}{n}\} \tag{1.13}$$

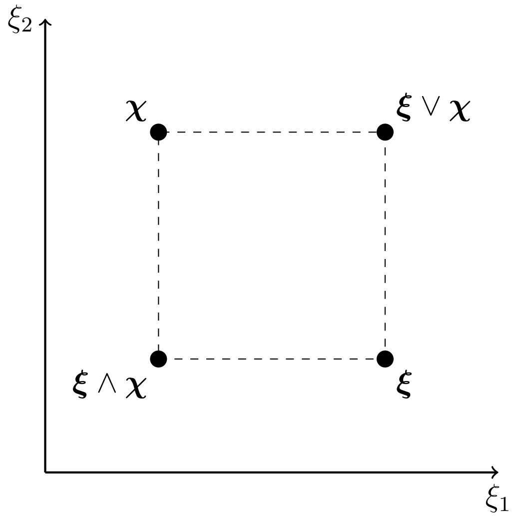
1.2 Submodularity

The main aim of this report is to present the existing tractability results for DRO problems with submodular objective. Submodularity is a property of a function which combines interesting practical insights with relevant optimization properties. We first give the general definition and then discuss these characteristics.

Definition 17. A function $f : \prod_{i=1,\ldots,N} \Xi_i \longrightarrow \mathbb{R}$ is *submodular* if:

$$f(\boldsymbol{\xi}) + f(\boldsymbol{\chi}) \geq f(\boldsymbol{\xi} \wedge \boldsymbol{\chi}) + f(\boldsymbol{\xi} \vee \boldsymbol{\chi}), \quad \forall \boldsymbol{\xi}, \boldsymbol{\chi} \in \prod_{i=1,\ldots,N} \Xi_i \tag{1.14}$$

The key idea of submodular functions is that of *diminishing marginal returns*, i.e. combining values leads to lower benefits. In particular, taking the two extremes (min and max) leads to a lower function value. With $N = 2$, we can give a geometric interpretation. Let suppose $\Xi = \mathbb{R}^2$, or a rectangle in $\mathbb{R}^2$. With reference to figure 1.2 below, if f is submodular, the sum of the values of the function on the top-left and bottom-right of a rectangle is greater or equal than the sum of the bottom-left and top-right corner.



We give this analogous definition of supermodular functions.

Definition 18. A function f is *supermodular* if $-f$ is submodular. Equivalently,

$$f(\boldsymbol{\xi}) + f(\boldsymbol{\chi}) \leq f(\boldsymbol{\xi} \wedge \boldsymbol{\chi}) + f(\boldsymbol{\xi} \vee \boldsymbol{\chi}), \quad \forall \boldsymbol{\xi}, \boldsymbol{\chi} \in \prod_{i=1,\ldots,N} \Xi_i \tag{1.15}$$

There exist some characterizations of submodularity function:

Proposition 19. *If $\Xi = \{0,1\}^N$. Let $S \subseteq \{1,\ldots,N\}$ and $\xi_i = \mathbb{1}_{i \in S}$. We define $f(S) := f(\boldsymbol{\xi})$. f is submodular if*

$$f(S) + f(T) \geq f(S \cap T) + f(S \cup T), \quad \forall S, T \subseteq \{1,\ldots,N\} \tag{1.16}$$

Example 1. $f(S) = |S|$, with $S = \{1,2\}$, $T = \{2,3,4\}$.

The key idea of diminishing returns is indeed that the marginal gain of adding an element in a set reduces as the set grows larger. This is clear stating the following:

Proposition 20. *The characterisation in 1.16 is equivalent to*

$$f(A \cup \{i\}) - f(A) \geq f(B \cup \{i\}) - f(B), \quad \forall A \subset B \subseteq \{1,\ldots,N\}, \quad i \in \{1,\ldots,N\} \setminus B \tag{1.17}$$

Proof. Let us first see that 1.16 $\Longrightarrow$ 1.17.
We take $S = A \cup \{i\}$ and $T = B$

$$\begin{aligned} f(A \cup \{i\}) + f(B) &\geq f(A \cup B \cup \{i\}) + f((A \cup \{i\}) \cap B) \\ &= f(B \cup \{i\}) + f(A) \end{aligned}$$

Now let us suppose 1.17.
If $S \subseteq T$, it is obvious, since

$$f(S \cap T) + f(S \cup T) = f(S) + f(T)$$

So we can suppose that $X_m := T \setminus S = \{v_1, \ldots, v_m\}$. Let us call $X_i := \{v_1, \ldots, v_i\}$ and $X_0 = \emptyset$. For hypothesis we have that

$$f((A \cap B) \cup X_i \cup \{v_{i+1}\}) - f((A \cap B) \cup X_i) \geq f(A \cup X_i \cup \{v_{i+1}\}) - f(A \cup X_i)$$

We sum for $i = 1, \ldots, n-1$ and we obtain

$$f((A \cap B) \cup X_n) - f(A \cap B) \geq f(A \cup X_n) - f(A)$$

which is, for definition of X_n

$$f(A) + f(B) \geq f(A \cup B) + f(A \cap B)$$

□

It is clear the interpretation of diminishing returns of 1.17.
We can give other characterisations of submodular functions:

Proposition 21. *If $\Xi = \mathbb{R}^N$ and f is differentiable, then f is submodular if*

$$\frac{\partial}{\partial \xi_i} f(\boldsymbol{\xi}) \geq \frac{\partial}{\partial \chi_i} f(\boldsymbol{\chi}), \quad \forall \boldsymbol{\xi} \leq \boldsymbol{\chi}, \forall i \in \{1, \ldots, N\} : \xi_i = \chi_i \tag{1.18}$$

In this case the interpretation is that, given the same value of f on a vector component, the gains grow slower if the other components of the vector are bigger.
Similarly

Proposition 22. *If $\Xi = \mathbb{R}^N$ and f is twice differentiable, then f is submodular if*

$$\frac{\partial^2}{\partial \xi_i \xi_j} f(\boldsymbol{\xi}) \leq 0, \quad \forall \boldsymbol{\xi} \in \mathbb{R}^N, \forall i \neq j \tag{1.19}$$

The concept is similar: if we increase the cost on one component, the cost along cannot increase faster. It is important to notice the difference with concavity given by the fact that 1.19 needs to hold only for $i \neq j$.
It is interesting to give some examples:

(*i*) $f(\boldsymbol{\xi}) = h(\mathbf{a}^{\mathbf{t}}\boldsymbol{\xi})$, where $\mathbf{a} \geq \mathbf{0}$ and $h : \mathbb{R} \longrightarrow \mathbb{R}$ is concave;

(*ii*) $f(\boldsymbol{\xi}) = \max(\xi_1, \ldots, \xi_N)$;

(*iii*) $f(\boldsymbol{\xi}) = -\prod_{i=1,\ldots,N} \xi_i$, where $\boldsymbol{\xi} \geq \mathbf{0}$;

(*iv*) $f(\boldsymbol{\xi}) = \mathbf{a}^{\mathbf{t}}\boldsymbol{\xi}$.

(*v*) univariate functions and separable functions.

The example (*i*) is of a concave function and the submodularity follows from it.
The example (*ii*) is of a convex function.
The example (*iii*) is a function neither convex or concave;
The example (*iv*) is of a function both convex and concave.
Submodular functions have various properties. First of all,

Proposition 23. *Let $f_1, \dots, f_n$ be submodular functions and $a_1, \dots, a_n \geq 0$, then $\sum_{i=1}^{n} a_i f_i$ is submodular, but $f_i \wedge f_j$ with $i \neq j$ is not necessarily submodular.*

Definition 24. We call *decomposable submodular functions* the submodular function derived as sum of submodular functions.

However, the key property of submodular functions which make them particularly interesting in the context of optimization is that they *well-behave* with respect to minimization. In particular, we can say that submodular functions on discrete domains behave similar to convex functions in continuous domains in terms of minimization.
This idea is given by the following theorem

Theorem 25. *If f is a submodular function with polynomial evaluation oracle, and Ξ_i are discrete and finite, the problem*

$$\inf_{\boldsymbol{\xi} \in \prod_{i=1,\dots,N} \Xi_i} f(\boldsymbol{\xi}) \tag{1.20}$$

is solvable in time polynomial in N, $\max_{i=1,\dots,N} |\Xi_i|$ and in the evaluation time of the oracle.

Note that this works only for minimization; maximization of submodular functions is, instead, in general NP-hard.
To understand the link between submodular and convex optimization we need to introduce the concept of comonotonic random vector.

Definition 26. Given a Fréchet set $\mathcal{P}(\mathbb{P}_1, \dots, \mathbb{P}_N)$ with $\tilde{\xi}_i \sim \mathbb{P}_i$ with *supp*$(\mathbb{P}_i) \subseteq \Xi_i$. Let F_i bet the cumulative distribution function of $\tilde{\xi}_i$, then the *comonotonic random vector* is the vector $\tilde{\boldsymbol{\xi}}^c$ with maximal positive dependence in the Fréchet set and is given by:

$$\tilde{\boldsymbol{\xi}}^c := (F_1^{-1}(U), \dots, F_N^{-1}(U)) \tag{1.21}$$

where U is a uniform r.v. in $[0, 1]$ and F_i^{-1} the generalized inverse cdf.
We denote with $\mathbb{P}^c$ the distribution of $\tilde{\boldsymbol{\xi}}^c$.

The comonotonic r.v. has important properties:

Proposition 27. *The following properties hold:*

(*i*) *$\tilde{\boldsymbol{\xi}}^c$ lies in a completely ordered subset of $\mathbb{R}^N$, i.e. if $\boldsymbol{\xi}_1$ and $\boldsymbol{\xi}_2$ are two realizations of $\tilde{\boldsymbol{\xi}}^c$, then $\boldsymbol{\xi}_1 \leq \boldsymbol{\xi}_2$ or $\boldsymbol{\xi}_1 \geq \boldsymbol{\xi}_2$;*

(*ii*) $\mathbb{P}^c(\tilde{\boldsymbol{\xi}}^c > \mathbf{t}) = \min_{\mathbf{i} \in \{\mathbf{1},\dots,\mathbf{N}\}} \mathbb{P}_{\mathbf{i}}(\tilde{\xi}_{\mathbf{i}} > \mathbf{t}_{\mathbf{i}})$ *and* $\mathbb{P}^c(\tilde{\boldsymbol{\xi}}^c \leq \mathbf{t}) = \min_{\mathbf{i} \in \{\mathbf{1},\dots,\mathbf{N}\}} \mathbb{P}_{\mathbf{i}}(\tilde{\xi}_{\mathbf{i}} \leq \mathbf{t}_{\mathbf{i}})$, $\forall \mathbf{t} \in \mathbb{R}^{\mathbf{N}}$*;*

(iii) $\mathbb{E}_{\mathbb{P}^c}[f(\tilde{\boldsymbol{\xi}}^c)] = \int_0^1 f(F_1^{-1}(t), \ldots, F_N^{-1}(t))dt.$

For (iii), we notice that if Ξ_i is discrete and finite the cardinality of the support of $\tilde{\boldsymbol{\xi}}^c$ is at most $\sum_{i=1}^N |\Xi_i|$ and thus the expectation is computable using a polynomial number of calls to the evaluations oracle.

Another important characterization of the comonotonic r.v. exists:

Proposition 28. *$\tilde{\boldsymbol{\xi}}$ is the comonotonic vector of $\mathcal{P}(\mathbb{P}_1, \ldots, \mathbb{P}_N)$ if and only if:*

(a) $\inf_{\mathbb{P}\in\mathcal{P}(\mathbb{P}_1,\ldots,\mathbb{P}_N)} \mathbb{E}_{\mathbb{P}}[f(\tilde{\boldsymbol{\xi}})] = \mathbb{E}_{\mathbb{P}^c}[f(\tilde{\boldsymbol{\xi}}^c)], \quad \forall f$ *submodular;*

(b) $\sup_{\mathbb{P}\in\mathcal{P}(\mathbb{P}_1,\ldots,\mathbb{P}_N)} \mathbb{E}_{\mathbb{P}}[f(\tilde{\boldsymbol{\xi}})] = \mathbb{E}_{\mathbb{P}^c}[f(\tilde{\boldsymbol{\xi}}^c)], \quad \forall f$ *supermodular;*

The connection with optimization of submodular functions is made through this result:

Proposition 29. *$f : \prod_{i=1,\ldots,N} \Xi_i \longrightarrow \mathbb{R}$ is submodular if and only if the functional $\inf_{\mathbb{P}\in\mathcal{P}(\mathbb{P}_1,\ldots,\mathbb{P}_N)} \mathbb{E}_{\mathbb{P}}[f(\tilde{\boldsymbol{\xi}})]$ is convex.*

The functional value is called *Choquet integral* or *Lovàsz extension.*

Theorem 30. *Therefore we can reformulate a minimization problem of submodular function as a convex optimization problem:*

$$\inf_{\boldsymbol{\xi}\in\prod_{i=1,\ldots,N}\Xi_i} f(\boldsymbol{\xi}) = \inf_{supp(\mathbb{P}_i)\subseteq\Xi_i,\forall i\in[N]} \inf_{\mathbb{P}\in\mathcal{P}(\mathbb{P}_1,\ldots,\mathbb{P}_N)} \mathbb{E}_{\mathbb{P}}[f(\tilde{\boldsymbol{\xi}})] = \inf_{supp(\mathbb{P}_i)\subseteq\Xi_i,\forall i\in[N]} \mathbb{E}_{\mathbb{P}^c}[f(\tilde{\boldsymbol{\xi}}^c)] \tag{1.22}$$

Proof. The second equality follows the characterization of the comonotonic vector. The first equality follows the fact that the optimal solution on the right-hand side is a Dirac measure and, thus, calculating the expectation is the same as calculate the value of the function. □

Chapter 2

Tractability of DRO Problems with Submodularity

In this Chapter, we present the main results from literature about tractable formulations of DRO problems which exploit in their formulation (in the objective or in the definition of the ambiguity sets) submodularity (or supermodularity). We will present four different approaches, followed by the discussions of some applications.

2.1 Polynomial Time Sharp Bound

In this section, we summarize the main results from Natarajan et al. 2023. In the paper a general ambiguity set is defined and a polynomial tractability of DRO affine bound for that ambiguity set is presented. Submodularity is exploited in the definition of the ambiguity set. The objective function is an affine function, therefore we consider a problem of the form

$$\inf_{\boldsymbol{x}\in\mathcal{X}} \sup_{\mathbb{P}\in\mathcal{P}} \mathbb{E}_{\mathbb{P}}[f(\boldsymbol{x},\tilde{\boldsymbol{\xi}}) := \max_{k\in\{1,\dots,K\}} (\mathbf{a_k^t}(\mathbf{x})\tilde{\boldsymbol{\xi}} + \mathbf{b_k}(\mathbf{x}))] \tag{2.1}$$

In particular, we consider the bound:

$$f^* = \sup_{\mathbb{P}\in\mathcal{P}} \mathbb{E}_{\mathbb{P}}[f(\tilde{\boldsymbol{\xi}}) := \max_{k\in\{1,\dots,K\}} (\mathbf{a_k^t}\tilde{\boldsymbol{\xi}} + \mathbf{b_k}))] \tag{2.2}$$

In the bound 2.2, we fix $\boldsymbol{x}$ and calculate the corresponding cost. Calculating f^* therefore means providing a bound on the optimal value of the DRO affine problem 3.3.

In the paper from Natarajan et al. 2023 it is studied f^* for the following ambiguity set:

$$\mathcal{P} = \{\mathbb{P} = \mathcal{P}(\Xi)|\mathbb{E}_{\mathbb{P}}[f_j(\tilde{\boldsymbol{\xi}})] \leq \gamma_j, \forall j \in \{1,\dots,J\}] \tag{2.3}$$

where γ_j are scalars and $f_j : \Xi \longrightarrow \mathbb{R}$ are functions.

Remark 31. The following assumptions are made for the ambiguity set:

(*i*) $\Xi = \prod_{i=1}^{N} \Xi_i$, with $\Xi : i \subset \mathbb{R}$ is discrete and finite;

(*ii*) f_j are submodular functions with polynomial time evaluation oracle.

Proposition 32. *Testing whether $\mathcal{P}$ with the assumptions from remark 31 is not empty is possible in polynomial time.*

The main result of the paper is

Theorem 33. *If the ambiguity set 2.3 follows the assumptions in remark 31. Then the bound f^* in 2.2 is computable in polynomial time.*

Proof. Since Ξ is discrete and finite, choosing a probability measure $\mathbb{P}$ in $\mathcal{P}$ corresponds in assigning a value to $p(\boldsymbol{\xi}) := \mathbb{P}(\tilde{\boldsymbol{\xi}} = \boldsymbol{\xi})$ for every $\boldsymbol{\xi} \in \Xi$, with $p(\boldsymbol{\xi}) \geq 0$ and $\sum_{\boldsymbol{\xi}\in\Xi} p(\boldsymbol{\xi}) = 1$. We can therefore reformulate 2.2 as a linear program with $p(\boldsymbol{\xi})$ as decision variables:

$$
\begin{aligned}
f^* = \text{maximize} \quad & \sum_{\boldsymbol{\xi}\in\Xi} \max_{k\in\{1,\dots,K\}} (\mathbf{a_k^t}\tilde{\boldsymbol{\xi}} + \mathbf{b_k})\mathbf{p}(\boldsymbol{\xi}) && (2.4)\\
\text{subject to} \quad & \sum_{\boldsymbol{\xi}\in\Xi} f_j(\boldsymbol{\xi})p(\boldsymbol{\xi}) \leq \gamma_j, \quad \forall j \in \{1,\dots,J\}, && (2.5)\\
& \sum_{\boldsymbol{\xi}\in\Xi} p(\boldsymbol{\xi}) = 1, && (2.6)\\
& p(\boldsymbol{\xi}) \geq 0, \quad \forall \boldsymbol{\xi} \in \Xi, && (2.7)
\end{aligned}
$$

We note that the number of constraints, excluding the nonegativity, is polynomial. The number of variables however is exponential, because if $\tilde{\boldsymbol{\xi}}$ have independent marginal, the number of variables, i.e. the number of possible realizations of the r.v., is exponential in N.

Therefore we build the dual problem:

$$
\begin{aligned}
f_d^* = \text{minimize} \quad & y_0 + \sum_{j=1}^{J} y_j\gamma_j && (2.8)\\
\text{subject to} \quad & y_0 + \sum_{j=1}^{J} y_j f_j(\boldsymbol{\xi}) \geq \max_{k\in\{1,\dots,K\}} (\mathbf{a_k^t}\boldsymbol{\xi} + \mathbf{b_k}), \quad \forall \boldsymbol{\xi} \in \boldsymbol{\Xi} && (2.9)\\
& y_j \geq 0, \quad \forall j \in \{1,\dots,J\} && (2.10)\\
& && (2.11)
\end{aligned}
$$

In the dual problem obviously the number of variables is polynomial and the number of constraints is exponential.

We have feasibility of the dual problem, since a solution is $y_j = 0\ \forall j \in \{1,\dots,J\}$ and $y_0 = \max_{k\in\{1,\dots,K\}}(\mathbf{a_k^t}\boldsymbol{\xi} + \mathbf{b_k})$.

Since it is linear, we can apply strong duality, which says that $f^* = f_d^*$. Therefore we look for f_d^*:

For the separation and optimization equivalence, optimizing the dual is equivalent to solving the following separation problem:

Given y_0 and $y_j \geq 0$ decide whether

$$
y_0 + \sum_{j=1}^{J} y_j f_j(\boldsymbol{\xi}) \geq \max_{k\in\{1,\dots,K\}} (\mathbf{a_k^t}\boldsymbol{\xi} + \mathbf{b_k}), \quad \forall \boldsymbol{\xi} \in \boldsymbol{\Xi} \tag{2.12}
$$

This is equivalent to check if

$$y_0 + \sum_{j=1}^{J} y_j f_j(\boldsymbol{\xi}) \geq \mathbf{a_k^t}\boldsymbol{\xi} + \mathbf{b_k}, \quad \boldsymbol{\xi} \in \Xi, \forall \mathbf{k} \in \{\mathbf{1}, \ldots, \mathbf{K}\} \tag{2.13}$$

And equivalently

$$y_0 - b_k + \min_{\boldsymbol{\xi} \in Xi}(\sum_{j=1}^{J} y_j f_j(\boldsymbol{\xi}) - bfa_k^t\boldsymbol{\xi}) \geq 0, \quad \forall k \in \{1, \ldots, K\} \tag{2.14}$$

Submodularity of f_j implies submodularity of $(\sum_{j=1}^{J} y_j f_j(\boldsymbol{\xi}) - bfa_k^t\boldsymbol{\xi})$. Thus it consists in solving K submodularity minimization problems, which require polynomial time as theorem 25 states.

□

We have an important corollary on the solvability of the DRO problem:

Theorem 34. *If $\mathcal{X}$ is a compact convex set with an efficient separarion oracle and the ambiguity set $\mathcal{P}$ satisfies the assumptions in remark 31, then the DRO affine problem 2.2 is solvable in polynomial time.*

Proof. We reformulate the dual problem with its dual as in proof of theorem 33:

$$\text{minimize} \quad y_0 + \sum_{j=1}^{J} y_j \gamma_j \tag{2.15}$$

$$\text{subject to} \quad y_0 + \sum_{j=1}^{J} y_j f_j(\boldsymbol{\xi}) \geq \max_{k \in \{1,\ldots,K\}} (\mathbf{a_k^t}\boldsymbol{\xi} + \mathbf{b_k}), \quad \forall \boldsymbol{\xi} \in \Xi \tag{2.16}$$

$$y_j \geq 0, \quad \forall j \in \{1, \ldots, J\} \tag{2.17}$$

$$\tag{2.18}$$

As before we reformulate in term of separation problem: given $y_j \geq 0, y_0$ and $\boldsymbol{x} \in \mathcal{X}$, we have to determine whether

$$y_0 - b_k(\boldsymbol{x}) + \min_{\boldsymbol{\xi} \in Xi}(\sum_{j=1}^{J} y_j f_j(\boldsymbol{\xi}) - bfa_k^t(\boldsymbol{x})\boldsymbol{\xi}) \geq 0, \quad \forall k \in \{1, \ldots, K\} \tag{2.19}$$

which can be done in polynomial time.

□

A generalization of the computable bound is stated in the following

Theorem 35. *Consider the bound* $\sup_{\mathbb{P} \in \mathcal{P}} \mathbb{E}_{\mathbb{P}}[\max_{k \in \{1,\ldots,K\}} g_k(\tilde{\boldsymbol{\xi}})]$ *where the ambiguity set $\mathcal{P}$ follows the assumptions in remark 31. If g_k is a supermodular function for each k with polynomial time evaluation oracle, then the bound is efficiently computable.*

Proof. The proof is analogous to the previous ones. □

Remark 36. There exists an efficient specialized algorithm to solve the decomposable submodular function minimization problem we have in the proof of theorem 33.

2.2 Applications of Polynomial Time Sharp Bound

2.2.1 Multimarginal Optimal Transport

Definition 37. Given N marginals $\mathbb{P}_1, \ldots, \mathbb{P}_N$, and a cost function $c(\xi_1, \ldots, \xi_N)$, where $\xi_i \sim \mu_i$. The *multimarginal optimal transport* (MMOT) problem consists in finding a measure $\pi \in \mathcal{P}(\mathbb{P}_1, \ldots, \mathbb{P}_N)$, Fréchet set of $\mathbb{P}_1, \ldots, \mathbb{P}_N$, which solves the following minimization problem:

$$\text{minimize} \quad \int c(\xi_1, \ldots, \xi_N) d\pi(\xi_1, \ldots, \xi_N) \tag{2.20}$$

In case that $\mathbb{P}_i$ are defined over Ξ_i finite and discrete, we can define the Fréchet class as an ambiguity set which satisfies the assumptions of remark 31.

Definition 38. Given a N-dimensional random vector $\tilde{\boldsymbol{\xi}} \sim \mathbb{P}$ and a N-dimensional random vector $\tilde{\boldsymbol{\chi}} \sim \mathbb{Q}$, we say that $\tilde{\boldsymbol{\xi}}$ is larger than $\tilde{\boldsymbol{\chi}}$ in the:

- *upper orthant* (UO) order if:

$$\mathbb{P}(\tilde{\boldsymbol{\xi}} > \mathbf{t}) \geq \mathbb{Q}(\tilde{\boldsymbol{\chi}} > \mathbf{t}), \quad \forall \mathbf{t} \in \mathbb{R}^N \tag{2.21}$$

- *lower orthant* (LO) order if:

$$\mathbb{P}(\tilde{\boldsymbol{\xi}} \leq \mathbf{t}) \geq \mathbb{Q}(\tilde{\boldsymbol{\chi}} \leq \mathbf{t}), \quad \forall \mathbf{t} \in \mathbb{R}^N \tag{2.22}$$

- *concordance* (or orthant) order if it is larger in both UO and LO orders

- *supermodular* (SM) order if:

$$\mathbb{E}_{\mathbb{P}}[f(\tilde{\boldsymbol{\xi}})] \geq \mathbb{E}_{\mathbb{Q}}[f(\tilde{\boldsymbol{\chi}})], \quad \forall \text{ supermodular } f \tag{2.23}$$

Proposition 39. *(i) For $N = 2, UO \iff LO \iff SM \iff$ concordance*

(ii) For $N \geq 3, SM \implies UO$ and $SM \implies LO$.

Remark 40. It is often chosen $\tilde{\boldsymbol{\chi}} \sim \mathbb{P}$ and $\tilde{\boldsymbol{\chi}} = \tilde{\boldsymbol{\xi}}^{\perp}$, the random vector in the Fréchet set $\mathcal{P}(\tilde{\xi}_1, \ldots, \tilde{\xi}_N)$ with independent components. In this case we use the terms *positive upper [lower] orthant dependence* (POUD [POLD]) and therefore *positive orthant dependence* (POD).

We now consider the two assumptions on the random variables we are considering:

(i) $\tilde{\xi}_i$ are discrete with probabilities given by $p_i(\xi_i) = \mathbb{P}(\tilde{\xi} = \xi_i)$, for $\xi_i \in \Xi_i$ finite set;

(ii) The bivariate distribution of $(\tilde{\xi}_i, \tilde{\xi}_j)$ is POD for $i \neq j$.

Under these assumptions we define the ambiguity set:

$$\mathcal{P} := \{\mathbb{P} \in \mathcal{P}(\prod_{i=1}^{N} \Xi_i) | \mathbb{P}(\tilde{\xi}_i = \xi_i) = p_i(\xi_i), \forall \xi_i \in \Xi_i, \forall i \in \{1, \ldots, N\}, \tag{2.24}$$

$$\mathbb{P}(\tilde{\xi}_i \geq \xi_i, \tilde{\xi}_j \geq \xi_j) \geq \sum_{\xi \geq \xi_i} p_i(\xi) \sum_{\xi \geq \xi_j} p_j(\xi), \forall \xi_i \in \Xi_i, \forall \xi_j \in \Xi_j, \ \forall i < j \in \{1, \ldots, N\}\}. \tag{2.25}$$

which corresponds to the ambiguity set of an instance of MMOT with upper dependence.

Remark 41. $\tilde{\boldsymbol{\xi}}^c$ and $\tilde{\boldsymbol{\xi}}^\perp$ both are included in $\mathcal{P}$, so there is no assumption on the dependence of 3 or more variables (on 2 it is assumed POD).

Remark 42. Since $\mathbb{1}_{\xi_i > t_i, \xi_j > t_j}(\xi_i, \xi_j)$ are supermodular, thus $\mathcal{P}$ satisfies the assumption of remark 31.

Proposition 43. *Taking as ambiguity set the $\mathcal{P}$ defined in 2.24 we have that the upper bound f^* in 2.2 is equivalent to a polynomial sized linear program.*

2.2.2 Moment Problem

As with moment problem we refer to problems where we define ambiguity sets fixing some of the moments of the probability measures.

Remark 44. We will make the following assumptions:

(*i*) Each random variable $\tilde{\xi}_i$ is discrete with support contained in a finite set Ξ_i. For each r.v., the first L moments are specified as $m_{i,l} = \mathbb{E}[\tilde{\xi}_i^l]$ for $l \in \{1, \ldots, L\}$;

(*ii*) Each pair of r.v. has a lower bound on the cross moment, i.e. $\mathbb{E}[\tilde{\xi}_i \tilde{\xi}_j] \geq Q_{i,j}$ for $i \neq j$.

Proposition 45. *Under the assumptions of remark 44 compute the DRO affine bound in 2.2 f^* is equivalent to a polynomial sized linear program.*

2.3 Price of Correlation

2.3.1 POC Framework

Another study of the relationship between submodularity and DRO is made in the paper Agrawal et al. 2012.

Remark 46. In the paper it is considered a DRO problem with a marginal ambiguity set and discrete distribution, i.e. a DRO in the form

$$\begin{aligned} g(x) := \text{maximize} \quad & \mathbb{E}_{\mathcal{D}}[f(x, S)] \\ \text{subject to} \quad & \sum_{S: i \in S} \mathbb{P}_{\mathcal{D}}(S) = p_i \quad \forall i \in V \end{aligned} \tag{2.26}$$

The paper introduces the concept:

Definition 47. For a certain DRO instance defined by $(f, V, \{p_i\})$. We call x_I the optimal decision assuming independent marginals and x_R the optimal DRO decision. We define as *price of correlations* (POC) the ratio

$$POC = \frac{g(x_I)}{g(x_R)} \tag{2.27}$$

The POC is a measure of the cost of solving the stochastic optimization problem for independent marginals instead of the DRO problem. In the paper there are shown some results for submodular and supermodular functions, but before we give the following definitions to introduce a cost-sharing scheme:

Definition 48. Given a set function $f(S)$, a *cost-sharing scheme* is a function $\chi(i,S)$, with $i \in S$, which specifies the share of i in S, i.e. how much the element i contributes to the total cost of $f(S)$.

Definition 49. A scheme χ is *β-budget balance* if, $\forall S$,

$$f(S) \geq \sum_{i \in S} \chi(i,S) \geq \frac{f(S)}{\beta} \tag{2.28}$$

Which means that the sum of the cost share in S is lower than the total cost, but higher than $1/\beta$ of the cost.

Definition 50. A scheme χ is *cross-monotonic* if, $\forall i \in S, S \subseteq T$,

$$\chi(i,S) \geq \chi(i,T) \tag{2.29}$$

which it means that if we add elements in the cost sharing scheme, the cost share of i does not increase.

Definition 51. A scheme χ is *weak η-summable* if $\forall S$ and $\forall \sigma_S$ permutation of $|S|$ elements,

$$\sum_{l=1}^{|S|} \chi(i_l, S_l) \leq \eta f(S) \tag{2.30}$$

where i_l is the l^{th} element and S_l is the set of the first l elements according to the ordering in σ_S.

Which it means that no matter how we order the elements, the sum of shared cost will not exceed $\eta f(S)$.

Theorem 52. *For any DRO instance $\{f, V, \{p_i\}\}$ with marginal ambiguity set, such that for all x, $f(x,S)$ is non decreasing in S and has a cost-sharing scheme which is β-budget balance, η-weak-summable and cross-monotonic, then*

$$POC \leq \min\{2\beta, \eta\beta(\frac{e}{e-1})\} \tag{2.31}$$

2.3.2 Submodularity POC

Considering a submodular set function h, we define the incremental cost-sharing scheme:

$$\chi(i,S) = h(S_i) - h(S_{i-1}) \tag{2.32}$$

where S_i is the set of the first i elements of S. The cost scheme defined simply takes the increase in the value of h by adding the element i as shared cost of i.

Proposition 53. *The cost-sharing scheme defined in 2.32 is 1-budget balance, cross-monotonic and 1-weak-summable.*

The proof is quite immediate from the definition.

From theorem 52, it follows

Corollary 54. *For any DRO instance $\{f, V, \{p_i\}\}$ with marginal ambiguity set and $f(x, S)$ non-decreasing and submodular in S for all feasible x,*

$$POC \leq \frac{e}{e-1} \tag{2.33}$$

Moreover,

Proposition 55. *For any DRO instance $\{f, V, \{p_i\}\}$ with marginal ambiguity set and $f(x, S)$ non-decreasing and submodular in S for all feasible x,*

$$POC = \frac{e}{e-1} \tag{2.34}$$

The result is particularly relevant because it says that we can approximate quite well the DRO problem for submodular functions and marginal discrete ambiguity set with the stochastic problem that assumes that all marginals are independent.

2.3.3 Supermodularity POC

When the cost function $f(x, S)$ is supermodular in S, we have that the POC can be even exponential.

Example 2. We consider a *two-stage minimum cost flow problem* with a single source s and n sinks $t_1, \dots, t_n$. We assume that each sink as a probability $p_i = 1/2$ to request a demand, and then an unit flow has to be sent from s to t_i. There are edges from u to t_i for every i with fixed capacity 1, and an edge from s to u, whose capacity needs to be purchased.

The cost of capacity x on the edge (s, u) is in the first stage

$$c^I(x) = \begin{cases} x, & x \leq n-1 \\ n+1, & x = n \end{cases} \tag{2.35}$$

and in the second stage it is

$$c^{II}(x) = 2^n x \tag{2.36}$$

Given x as the decision of the first stage, the cost of edges in the second stage to serve a set S of requests is

$$f(x, S) = c^I(x) + 2^n(|S| - x)^+ \tag{2.37}$$

which is supermodular.

The expected cost with independent demands is n, while the worst case distribution is the one with $\mathbb{P}(V) = \mathbb{P}(\varnothing) = 1/2$ and probability 0 to all other scenarios, which leads to a cost of $2^{n-1} + n - 1$.

It follows that $POC = \Omega(2^n)$.

In general it holds that

Proposition 56. *The worst-case distribution over S for a supermodular function f with marginals $p_1 \geq p_2 \geq \cdots \geq p_n$ is*

$$\mathbb{P}(S) = \begin{cases} p_n, & if S = S_n \\ p_i - p_{i+1}, & if S = S_i, 1 \leq i \leq n-1 \\ 1 - p_1, & if S = \varnothing \\ 0, & otherwise \end{cases} \tag{2.38}$$

A consequence is

Corollary 57. *The DRO problem with supermodular objective and marginals* $p_1, \ldots, p_n$ *can be formulated as*

$$\min_{x \in \mathcal{X}} p_n f(x, S_n) + \sum_{i=1}^{n-1} (p_i - p_{i+1}) f(x, S^i) + (1 - p_i) f(x, \varnothing) \tag{2.39}$$

which becomes tractable with convex optimization techniques if $f(x, S)$ is convex in x and $\mathcal{X}$ is convex.

2.3.4 Applications

An applications of the results above is stated in the paper Chen et al. 2020, where it is studied a distributionally robust model for influence maximization.

2.4 Distributionally Submodular Maximization

2.4.1 Formulation

In Staib et al. 2018, a different approach is adopted to address a different class of submodular distributionally robust optimization problems.

The DRO problem addressed is

$$\max_S \min_{\mathbb{P} \in \mathcal{P}_{\rho,n}} \mathbb{E}_{f \sim \mathbb{P}}[f(S)] = \max_S \min_{\mathbf{p} \in \mathcal{P}_{\rho,n}} \sum_{i=1}^{n} p_i f_i(S) \tag{2.40}$$

Where

- $f : 2^V \longrightarrow \mathbb{R}$ is a submodular set function, with f is distributed as $\mathbb{P}$
- f_i are i.i.d. samples from $\mathbb{P}$ and p_i corresponds which define the discrete empirical distribution $\hat{P}_n(f = f_i)$;
- $\mathcal{P}_{\rho,n}$ is a χ^2 uncertainty set of radius ρ and centered in $\hat{P}_n$.

Remark 58. We note some important differences with the problem of section 2.1:

- we aim to solve a maximization DR problem;
- f is intended only as set function;
- f is not necessarily affine;
- the ambiguity set is a χ^2 uncertainty set.

The paper firstly reformulates the problem as the relaxation:

$$\max_{\mathcal{D}} \min_{i \in \{1, \ldots, n\}} \mathbb{E}_{S \sim \mathcal{D}}[f_i(S)] \tag{2.41}$$

so we take an optimal distribution $\mathcal{D}$ over the sets S and look only at the worst-case sample f_i.

The paper then extends f with submodularity to obtain a function defined in a subset of $\mathbb{R}^N$ rather than a set function. In this way we also obtain a formulation more similar to the one in section 2.1:

Definition 59. A *multilinear extension* of f is a function $F : \mathcal{X} \subseteq [0,1]^N \longrightarrow \mathbb{R}$ defined as

$$F(\boldsymbol{x}) := \sum_{S \subseteq V} f(S) \prod_{i \in S} x_i \prod_{x_j \notin S} (1 - x_j) \tag{2.42}$$

As we see the extension F is the expected value of $f(S)$, given that each element of index i could belong to S independently with probability x_i.

Remark 60. F is defined over the convex hull $\mathcal{X}$ of indicator vector of feasible sets.

Proposition 61. *F is still submodular, in particular the following property of concavity along increasing directions holds:*

$$F(\boldsymbol{x} + \gamma \mathbf{e}_i) - F(\boldsymbol{x}) \geq F(\mathbf{y} + \gamma \mathbf{e}_i) - F(\mathbf{y}), \quad \forall \boldsymbol{x} \leq \mathbf{y}, \gamma > 0 \in \mathcal{X}, i \in \{1, \ldots, n\} \tag{2.43}$$

The paper therefore redefines the problem 2.40 as

$$\max_{\boldsymbol{x} \in \mathcal{X}} \min_{p \in \mathcal{P}_{\rho,n}} \sum_{i=1}^{n} p_i F_i(\boldsymbol{x}) \tag{2.44}$$

Remark 62. In this problem:

- $\mathcal{X}$ is convex;
- F_i are submodular but not necessarily affine;
- $\mathcal{P}_{\rho,n}$ is a ξ^2 ambiguity set of discrete probability distributions over continuous sets, represented as vectors over Δ_n.

Remark 63. The problems are two:

(i) How to solve efficiently 2.40 from 2.44;

(ii) How to efficiently solve 2.44.

For (i), the paper states the following

Proposition 64. *If $\boldsymbol{x}$ is an α-optimal solution to 2.44, then $\boldsymbol{x}$ induces a distribution $\mathcal{D}$ over subsets so that $\mathcal{D}$ is $(1 - 1/e)\alpha$-optimal for 2.40.*

This result is application of the price of correlation described in section 2.3. The idea is that, with the definition of F we are taking the set distribution $\mathcal{D}$ over the subsets S with independent marginals p_i of the single elements. The approximation with independent marginals for submodular functions has a POC of $e/(e-1)$, as shown in proposition 55.

Definition 65. We say that $\tilde{f}$ is an *α-optimal solution* for an optimization problem whose solution is f^* if it holds that

$$\tilde{f} \geq \frac{1}{\alpha} f^* \tag{2.45}$$

Remark 66. $\boldsymbol{x}$ induces a distribution in the sense that it gives the probability of each element to be in a set S, i.e. it induces the distribution

$$\mathbb{P}(S) = \prod_{i \in S} x_i \prod_{j \notin S} (1 - x_j) \tag{2.46}$$

We note that this distribution assumes marginals independent, i.e. that $\{x_i \in S\}$ and $\{x_j \in S\}$ are independent if $i \neq j$.

2.4.2 Algorithmic Solution

The problem (*ii*) from remark 63 is solved in the paper by presenting an algorithm.

The paper proposes a *Momentum Franck-Wolfe* (MFW) algorithm to solve the problem:

Remark 67. The Frank-Wolfe algorithm solves convex problems of the form

$$\text{minimize} \quad f(\boldsymbol{x}) \tag{2.47}$$
$$\text{subject to} \quad \boldsymbol{x} \in \mathcal{D} \tag{2.48}$$

with f and $\mathcal{D}$ convex.

The FW algorithm finds first the direction $s^{(t)}$ that minimizes the gradient with reference to the current point $x^{(t)}$ solving

$$\text{minimize} \quad \mathbf{s}^t \nabla f(\boldsymbol{x}^{(t)}) \tag{2.49}$$
$$\text{subject to} \quad \mathbf{s} \in \mathcal{D} \tag{2.50}$$

The solution $\mathbf{s}^{(t)}$ is then used to determine the next value of $\boldsymbol{x}$:

$$\boldsymbol{x}^{(t+1)} \longleftarrow \boldsymbol{x}^{(t)} + \alpha(\mathbf{s}_k - \boldsymbol{x}_k) \tag{2.51}$$

The algorithm proposed introduces a momentum in the Franck-Wolfe iteration Heuristically, the algorithm proposed, works as follow:

1. Based on the value of the latest solution $\boldsymbol{x}^{(t-1)}$, it computes $p^{(t)} := \text{argmin}_{p \in \mathcal{P}_{\rho,n}} \sum_{i=1}^{n} p_i F_i(\boldsymbol{x}^{(t-1)})$; remember we want to find $\boldsymbol{x}$ that then maximizes this quantity;
2. it calculates an approximation of the gradient of the objective selecting $c \leq n$, i.e. $\tilde{\nabla}^{(t)} := \frac{1}{c} \sum_{l=1}^{c} p_{i_l}^{(t)} \nabla F_{i_l}(\boldsymbol{x}^{(t-1)})$;
3. it moves the momentum towards $\tilde{\nabla}^{(t)}$, based on a step parameter ρ_t, i.e. $d^{(t)} := d^{(t-1)} + \rho_t(\tilde{\nabla}^{(t)} - d^{(t-1)})$;
4. it finds the best search direction in $\mathcal{X}$, with momentum $d^{(t)}$, i.e. $v^{(t)} := \text{argmax}_{v \in \mathcal{X}} \langle d^{(t)}, v \rangle$;
5. it moves the solution of $v^{(t)}$, i.e. $x^{(t)} := x^{(t-1)} + v^{(t)}/T$, where T is the time.

The main differences with a standard Frank-Wolfe is the use of a momentum $d^{(t)}$ which replaces the simple use of gradient. The momentum conserves part of the information from previous iterations in determining the best search direction.

Remark 68. The difference of FW and MFW with respect to gradient descent is that it solves an optimization problem instead of a simple projection. In our setting this problem is cheap to solve.

2.5 Applications of Submodular Maximization

2.5.1 Variance-regularization

The application presented in Staib et al. 2018 is *variance regularization*.

Definition 69. Let f be a stochastic set function and suppose we have the optimization problem:

$$\text{optimize} \quad f_{\mathbb{P}}(S) := \mathbb{E}_{f\sim\mathbb{P}}[f(S)] \tag{2.52}$$

Let $f_1, \ldots, f_n$ samples of $f \sim \mathbb{P}$ that form an empirical distribution $\hat{P}_n$. We call *variance-regularized* problem of 2.52, the following optimization problem:

$$\text{optimize} \quad f_{\hat{P}_n}(S) - C_1\sqrt{\frac{\mathrm{var}_{\hat{P}_n}(f(S))}{n}} \tag{2.53}$$

This problem arises when we have a limited number n of samples: in this case it is better to optimize directly the bias-variance trade-off instead of $f_{\hat{P}_n}$.

It is shown that the variance-regularized problem is equivalent to a DRO problem in the form as the one addressed in section 2.4, in particular three results state the equivalence of the variance reduction and a DRO problem:

Proposition 70 (Theorem 2.1 from Staib et al. 2018)**.** *Fix $\rho \geq 0$, and let $Z \in [0, B]$ be a random variable (i.e., $Z = f(S)$). Write $s_n^2 = Var_{\hat{P}_n}(Z)$ and let $OPT = \inf_{\tilde{P}\in P_{\rho,n}} \mathbb{E}_{\tilde{P}}[Z]$. Then*

$$\left(\sqrt{\frac{2\rho}{n}s_n^2} - \frac{2B\rho}{n}\right)_+ \leq \mathbb{E}_{\hat{P}_n}[Z] - OPT \leq \sqrt{\frac{2\rho}{n}s_n^2}. \tag{2.54}$$

Moreover, if $s_n^2 \geq \frac{2\rho(\max_i z_i - z_n)^2}{n}$, then

$$OPT = \mathbb{E}_{\hat{P}_n}[Z] - \sqrt{\frac{2\rho}{n}s_n^2}, \tag{2.55}$$

i.e., DRO is exactly equivalent to variance regularization.

Proposition 71 (Lemma 2.1 from Staib et al. 2018)**.** *Fix δ, and let S be a subset chosen to maximize $f_{\hat{P}_n}(S) - C_1\sqrt{Var_{\hat{P}_n}(f(S))/n}$. With probability at least $1-\delta$, the subset S satisfies*

$$f_P(S) \geq f_{\hat{P}_n}(S) - C_1\sqrt{Var_P(f(S))/n} - C_2/n, \tag{2.56}$$

where $C_1 \leq \sqrt{2\log(1/\delta)}$ and $C_2 \leq 2B\log(1/\delta)$.

Proposition 72 (Lemma 2.2 from Staib et al. 2018)**.** *Let $P_{\rho,n}$ be the χ^2 uncertainty set around the empirical distribution $\hat{P}_n$. If S is a solution to the problem $\max_S \min_{\tilde{P}\in P_{\rho,n}} \mathbb{E}_{\tilde{P}}[f(S)]$, then with high probability, it holds that*

$$f_P(S) \geq \min_{\tilde{P}\in P_{\rho,n}} \mathbb{E}_{\tilde{P}}[f(S)] - \frac{2B\rho}{n}. \tag{2.57}$$

2.5.2 Influence Maximization

Definition 73. Given a graph $G = (V, E)$, on which influence propagates. Edges can be active or inactive, i.e. edges which can or cannot spread the influence. The *influence maximization* problem consists in finding an initial seed set $S \subseteq V$ of influenced nodes to maximize the number of nodes subsequently influenced. A node is influenced if connected to an influenced node via an active edge.

The different diffusion model distinguish how an edge could be activated. In the most common ones the fact of an edge being active is stochastic, i.e. it induces a distribution on $f(S)$, the function which indicates the number of influenced nodes given the set seed S.

Trying to maximize the expectation of $f(S)$, the problem can be written as a submodular distributionally robust maximization problem. The only thing to observe is the fact that f is submodular, but that is quite natural thinking of the submodularity as the property of diminishing returns.

2.5.3 Facility Location

Definition 74. Given a set V of possible facility locations j and a number of demand points i, drawn from a distribution $\mathcal{D}$. The *facility location* problem consists in choosing a subset $S \subseteq V$ that covers the demand points as well as possible, i.e. maximizing a measure r_j^i, which expresses how well a point i is covered by the facility j.

The problem aims to solve the following optimization problem:

$$\text{maximize} \quad \mathbb{E}_{i\sim\mathcal{D}}[\max_{j\in S} r_j^i] \tag{2.58}$$

The function $f(S) = \max_{j\in S} r_j^i$ is submodular and the distribution $\mathcal{D}$ over i induces a distribution $\mathbb{P}$ over $f(S)$.

2.6 Supermodularity in Two-stage DRO

Another recent result in the topic comes from Long et al. 2023, which studies a tractable formulation of a two-stage distributionally robust optimization model, with a supermodular function representing the second stage cost.

Definition 75. The two-stage DRO problem is defined as follows:

$$\min\nolimits_{x\in\mathcal{X}}(a^t x + \sup_{\mathbb{P}\in\mathcal{D}} \mathbb{E}_{\mathbb{P}}[g(x,\tilde{z})]) \tag{2.59}$$

where $\tilde{z}$ is a random variable and g is defined as

$$\begin{aligned} g(x,z) = \min \quad & b^t y \\ \text{s.t.} \quad & Wx + Uy \geq Vz + v^0 \end{aligned} \tag{2.60}$$

x represents the first stage decision, y the second stage.

The ambiguity set used in the paper is a scenario-wise ambiguity set, defined as

Definition 76. We defined the scenario-wise ambiguity set as

$$\mathcal{D} = \left\{ \mathbb{P} \;\middle|\; \begin{array}{l} \mathbb{E}_{\mathbb{P}}[\tilde{z} \mid \tilde{k} = k] = \mu^k, \quad \forall k \in [K] \\ \mathbb{E}_{\mathbb{P}}[|\tilde{z}_i - \mu_i^k| \mid \tilde{k} = k] \leq \delta_i^k, \quad \forall k \in [K] \\ \mathbb{P}(\underline{z}^k \leq \tilde{z} \leq \overline{z}^k \mid \tilde{k} = k) = 1, \quad \forall k \in [K] \\ \mathbb{P}(\tilde{k} = k) = q_k, \quad q \in \mathcal{Q}, \forall k \in [K] \end{array} \right\}. \tag{2.61}$$

where $\tilde{k}$ is a random variable which can assume value in $[K]$ and represents a scenario which affects the distribution of $\tilde{z}$.

Remark 77. Some assumptions on $\mathcal{D}$ are made in the paper to avoid trivial cases.

Remark 78. If we fix a scenario $\tilde{k} = k$, we have an ambiguity set for $\tilde{z}$ defined by its mean, its mean absolute deviation and its support.

In the fixed scenario case we can state the worst-case marginal distribution distribution $\mathbb{P}_k^* \in \arg\sup_{\mathbb{P}^k \in \mathcal{D}^k} \mathbb{E}_{\mathbb{P}^k}[g(x, \tilde{z})]$ for each $\tilde{z}_i$.

Proposition 79.

$$P_k^*(z_i = w) = \begin{cases} \frac{\delta_{k,i}}{2(\mu_{k,i} - z_{k,i})} & \text{if } w = z_{k,i}, \\ 1 - \frac{\delta_{k,i}(z_{k,i} - \bar{z}_{k,i})}{2(\bar{z}_{k,i} - \mu_{k,i})(\mu_{k,i} - z_{k,i})} & \text{if } w = \mu_{k,i}, \\ \frac{\delta_{k,i}}{2(\bar{z}_{k,i} - \mu_{k,i})} & \text{if } w = \bar{z}_{k,i}. \end{cases} \tag{2.62}$$

where $\hat{\delta}_{k,i}$ *is defined as:*

$$\hat{\delta}_{k,i} = \min \left\{ \delta_{k,i}, \frac{2(\bar{z}_{k,i} - \mu_{k,i})(\mu_{k,i} - z_{k,i})}{\bar{z}_{k,i} - z_{k,i}} \right\}. \tag{2.63}$$

This distribution satisfies the mean and MAD constraints for each z_i under scenario k, providing the worst-case expected cost for each marginal component individually.

The exact joint distribution, however, remains unknown.

Proposition 80. *Given* $f : \mathbb{R}^n \longrightarrow \mathbb{R}$, *it is equivalent to say that*

(1) f *is supermodular*

(2) Given a certain random vector $\tilde{w}$ *and defined a set of probability measures* $\mathcal{P} = \{\mathbb{P} | \mathbb{P}(\tilde{w}_i = x_{ij}) = p_{ij}, j \in [m_i], i \in [n]\}$ *for fixed* x_{ij}, p_{ij}, m_i. *Then there exist a worst-case distribution* $\mathbb{P}^* \in \arg\sup_{\mathbb{P} \in \mathcal{P}} \mathbb{E}_{\mathbb{P}}[f(\tilde{w})]$, *such that the support of* $\tilde{w}$, $\mathcal{W}_{\mathbb{P}^*} = \{w \in \mathbb{R}^n | \mathbb{P}^*(\tilde{w} = w) > 0\}$, *forms a chain of at most* $(\sum_{i \in [n]}(m_i - 1) + 1)$ *points.*

Definition 81. A *chain* is a totally ordered subset of a partially ordered set.

The idea connected to supermodularity is that to obtain the worst case distribution we move mass from two point w', w'' to $w' \wedge w'', w' \vee w''$, which guarantees a higher expectation due to supermodularity, does not change the marginal distributions and introduces an ordering $(w' \wedge w'' \leq w' \vee w'')$. The idea to obtain the worst case distribution is to move all the mass in this way.

An algorithm is provided to find this worst case distribution $\tilde{w} = \tilde{z}$ in $\mathcal{D}^k$ by building this chained support, in particular if $g(x, z)$ is supermodular in z, we have that $\sup_{\mathbb{P}^k \in \mathcal{D}^k} \mathbb{E}[g(x, z)] = \sum_{i \in [2n+1]} p_i g(x, z^i)$. The algorithm provided outputs p and z^i in $O(n)$ time.

The idea of the algorithm is to move from $\underline{z}^k$ to $\overline{z}^k$ looking for a feasible chain subject to the marginal distribution from proposition 79. This also constitutes the support for the worst-case distribution.

The worst-case distribution so-defined is independent from the first stage decision x, but the result can be generalized reintroducing the scenarios k. In this case the problem becomes:

$$\sup_{\mathbb{P} \in \mathcal{D}} \mathbb{E}_{\mathbb{P}}[g(x, \tilde{z})] = \max_{q \in \mathcal{Q}} \sup_{\mathbb{P}_k \in \mathcal{D}_k, k \in [K]} \sum_{k \in [K]} q_k \mathbb{E}_{\mathbb{P}_k}[g(x, \tilde{z})] = \max_{q \in \mathcal{Q}} \sum_{k \in [K]} q_k \sup_{\mathbb{P}_k \in \mathcal{D}_k} \mathbb{E}_{\mathbb{P}_k}[g(x, \tilde{z})]. \tag{2.64}$$

by simply taking the expectation of a discrete and finite quantity of scenarios $[K]$.

We can reformulate $\mathbb{E}_{\mathbb{P}_k}[g(x, \tilde{z})]$ as before and then exploit the fact that $\mathcal{Q}$ is polyhedron by definition, to write the following

Theorem 82. *If $g(x, z)$ is supermodular in z the two-stage stochastic DRO problem is equivalent to the linear program*

$$\begin{aligned}
\min\ & a^t x + \nu^t l \\
s.t.\ & R_k^t l \geq \sum_{i\in[2n+1]} p_i^k b^t y^{k,i}, \quad k \in [K], \\
& Wx + Uy^{k,i} \geq Vz^{k,i} + v^0, \quad k \in [K],\ i \in [2n+1], \\
& l \geq 0, \quad x \in X.
\end{aligned} \tag{2.65}$$

Chapter 3

Applications to Sensor Placement

In this chapter, we present the *sensor placement problem* and give some hints on how exploiting submodularity could help solving this problem.

3.1 Sensor placement orienteering problem

We consider the *sensor placement orienteering problem* as in Paradiso et al. 2022 of the form

$$\max_{w \in \mathcal{W}} \min_{\bar{\xi} \in \Xi} \max_{y \in \mathcal{Y}} \min_{\xi \in \Xi(w, \bar{\xi})} \xi^t y \tag{3.1}$$

The decision process looks like this:

1. We first choose $w \in \mathcal{W}$ as the nodes in which to place a sensor, with $w_i = 1$ if we place a sensor in i.

   This decision is based on the worst-case scenario of a random variable $\bar{\xi}$, which satisfies some constraints $A\bar{\xi} \leq b$;

2. We then observe the variables elements $\bar{\xi}_i$ for which $w_i = 1$ and consider a new random variable ξ with the elements ξ_i fixed $= \bar{\xi}$ observed;

3. We choose $y \in \mathcal{Y}$ to minimize $\xi^t y$ with respect to the worst-case realization of ξ.

We would like to integrate submodularity in the problem to solve the distributionally robust problem efficiently through the techniques of the previous sections.

3.1.1 Existing Literature

We rapidly present two works which already exploit submodularity in problem of sensor placement.

Golovin and Krause 2011

In the literature, the connection between sensor placement and submodularity is often outlined. Often, such as in Golovin and Krause 2011 and in Krause and Guestrin 2007, it is connected to the fact that sensors cover a certain area and, since the areas covered by different sensors can overlap, the function which models the information gained by the sensors exhibits the property of diminishing returns. However, this seems not to be the case of our problem.

Li and Mehr 2023

In the work Li and Mehr 2023, the problem stated is similar to ours, namely (Equation (9) of the paper).

$$\begin{aligned} \max_{\mathcal{L}_m \subseteq \mathcal{L}} \quad & u(\mathcal{L}_m) \\ \text{s.t.} \quad & |\mathcal{L}_m| \leq k \end{aligned} \tag{3.2}$$

where $\mathcal{L}_m$ is a possible sensor placement among the possible configurations $\mathcal{L}$, and u a function which expresses the sensing performance of a configuration. The constraint is on the number of sensors.

However, u and the submodularity structure follows the specific definition of the network flow problem. Three different possible models are presented:

1. In the scenario where routing information is not available (section 5.1), the problem becomes a linear system $Af = b$ with the flows f we would like to know as unknown. The quantity that models the information is the rank(A).
2. With information on the network structure (section 5.2), maximize the information is intended as minimize the sum of variances of the estimation errors made by the sensors.
3. Lastly (section 5.3), the information is modeled as the number of origin-destination paths with a sensor on.

3.1.2 First approach

First approach to submodularity is to reformulate the problem 3.1 as in Natarajan et al. 2023.

In particular we would like a problem of the form

$$\inf_{y \in \mathcal{Y}} \sup_{\mathbb{P} \in \mathcal{P}} \mathbb{E}_{\xi \sim \mathbb{P}}[\max_{k \in \{1,\dots,K\}} (a_k^t(y)\xi + b_k(y))] \tag{3.3}$$

where a_k and b_k are affine in y and $\mathcal{P}$ is the following ambiguity set:

$$\mathcal{P} = \{\mathbb{P} = \mathcal{P}(\Xi) | \mathbb{E}_{\mathbb{P}}[f_j(\tilde{\boldsymbol{\xi}})] \leq \gamma_j, \forall j \in \{1, \dots, J\}] \tag{3.4}$$

where

i γ_j are scalars;

ii $f_j : \Xi \longrightarrow \mathbb{R}$ are submodular functions with polynomial time evaluation oracle;

iii $\Xi = \prod_{i=1}^N \Xi_i$, with $\Xi : i \subset \mathbb{R}$ is discrete and finite.

The idea is to model in this way only the second stage decision. We can reformulate the problem 3.1 in a distributionally robust way:

$$\max_{y \in \mathcal{Y}} \min_{\xi \in \Xi(w,\bar{\xi})} \xi^t y \quad \rightarrow \quad \max_{y \in \mathcal{Y}} \min_{\mathbb{P} \in \mathcal{Q}(w,\bar{\xi})} \mathbb{E}_{\xi \sim \mathbb{P}}[\xi^t y] \tag{3.5}$$

We note that $\xi^t y$ is in the form of $\max_{k \in \{1,\dots,K\}}(a_k^t(y)\xi + b_k(y))$ with $K = 1, a_k^t(y) = y$ and $b_k(y) = 0$.

We need to define $\mathcal{Q}$ to model the fact that $\xi \in \Xi(w, \bar{\xi})$. The set $\Xi(w, \bar{\xi})$ can be defined as

$$\Xi(w, \bar{\xi}) = \{\xi \in \mathbb{R}^N : \xi_i = \bar{\xi}_i \forall i \text{ such that } w_i = 1\} \tag{3.6}$$

$\mathcal{Q}$, which models the distribution that ξ can assume, includes all distributions with marginals i fixed for $w_i = 1$. In some sense, it is a Fréchet ambiguity set but with only some marginals fixed. The paper Natarajan et al. 2023 includes an example of marginal ambiguity set in formulation 3.4 in remark (b) of section 3 (page 12).

We can define

$$\mathcal{Q}(w, \bar{\xi}) = \{\mathbb{P} \in \mathcal{P}(\prod_{i \in [N]} \Xi_i) | \mathbb{E}_{\xi \sim \mathbb{P}}[\mathbb{1}_{\xi_i = \bar{\xi}_i}] \leq 1, \mathbb{E}_{\xi \sim \mathbb{P}}[-\mathbb{1}_{\xi_i = \bar{\xi}_i}] \leq -1, \forall i \text{ such that } w_i = 1\} \tag{3.7}$$

Note that $\mathbb{1}_{\xi_i = \bar{\xi}_i}$ is both submodular and supermodular since it is a univariate function.

There are two things to look at:

1. define $\prod_{i \in [N]} \Xi_i$;
2. address the fact that we have a maximization and not a minimization problem.

For 2. we can easily put a $-$ in front of the objective and the properties should remain the same:

$$\min_{y \in \mathcal{Y}} \max_{\mathbb{P} \in \mathcal{Q}(w, \bar{\xi})} \mathbb{E}_{\xi \sim \mathbb{P}}[-\xi^t y] \tag{3.8}$$

For 1., we need to consider the fact that the support of the distributions in $\mathcal{Q}(w, \bar{\xi})$ is in $\{\xi \in \mathbb{R}^N : A\xi \leq b\}$.

Since, we cannot write $\{\xi \in \mathbb{R}^N : A\xi \leq b\}$ in the form $\prod_{i \in [N]} \Xi_i$, the idea is to take Ξ_i as discrete segments (sufficiently big to meet the problem requirements), and then write the constraints $A\xi \leq b$ as probability constraint of the form

$$\mathbb{P}(A\xi \leq b) = 1 \text{ if } \xi \sim \mathbb{P} \tag{3.9}$$

To meet the formulation from Natarajan et al. 2023, we have to express 3.9 as the expectation of a submodular function.

The function $\tilde{f}(\xi) = A\xi - b$ is submodular, since it is linear. We can therefore add the constraint:

$$\mathbb{E}[\tilde{f}(\xi) = (A\xi - b)] \leq 0 \tag{3.10}$$

The ambiguity set therefore becomes

$$\mathcal{Q}(w, \bar{\xi}) = \{\mathbb{P} \in \mathcal{P}(\prod_{i \in [N]} \Xi_i) | \mathbb{E}_{\xi \sim \mathbb{P}}[\mathbb{1}_{\xi_i = \bar{\xi}_i}] \leq 1, \mathbb{E}_{\xi \sim \mathbb{P}}[-\mathbb{1}_{\xi_i = \bar{\xi}_i}] \leq -1, \forall i \text{ such that } w_i = 1,$$
$$\mathbb{E}_{\xi \sim \mathbb{P}}[\tilde{f}(\xi) = A\xi - b] \leq 0\} \tag{3.11}$$

where Ξ_i are discrete segments.

In the formulation

$$\min_{y \in \mathcal{Y}} \max_{\mathbb{P} \in \mathcal{Q}(w,\bar{\xi})} \mathbb{E}_{\xi \sim \mathbb{P}}[-\xi^t y] \tag{3.12}$$

this problem is therefore solvable in polynomial time using the results from Natarajan et al. 2023.

3.1.3 Second Approach

The second possible research direction is to write the entire sensor placement problem using submodularity. In fact, we can replace the decision vector w with a set S, where $S = \{i \colon w_i = 1\}$. We can thus write the set function:

$$f(S, \bar{\xi}) := \max_{y \in \mathcal{Y}} \min_{\xi \in \Xi(S,\bar{\xi})} \xi^t y \tag{3.13}$$

The idea is to use the submodularity property on the set S to prove tractability of the maximization problem, similarly at what is done in the paper Staib et al. 2018.

We take the idea from the example of **influence maximization** in section 2.1.

Remark 83. In the example, the uncertainty is on a different variable ($\mathcal{E}$), than the variable with respect of which we have submodularity (S), as in our case.

In the same way, we define

$$f_{SPO}(S) = f(S; \bar{\xi}) \tag{3.14}$$

and then consider the distribution induced by the uncertainty of $\bar{\xi}$ on the function f_{SOP} and obtain a distributionally robust maximization problem:

$$\max_{\substack{S \subseteq \mathcal{N} \\ |S| \leq B}} \inf_{\mathbb{Q} \in \mathcal{Q}} \mathbb{E}_{f \sim \mathbb{Q}}[f_{SPO}(S)] \tag{3.15}$$

which is analogous in the form to the problem (1) in the paper Staib et al. 2018.

However, to apply the tractability results, we need to know the following.

1. a f_{SPO} which is submodular in S;
2. an opportune ambiguity set $\mathcal{Q}$.

For 1., we may need to find a new formulation of the problem, from literature or from scratch, but we can first of all use the problem as we already formulated it, so

$$(f_{SPO}(S) =) f(S, \bar{\xi}) := \max_{y \in \mathcal{Y}} \min_{\xi \in \Xi(S,\bar{\xi})} \xi^t y \tag{3.16}$$

We want to prove diminishing returns in S, i.e.

$$f(S \cup \{i\}) - f(S) \geq f(T \cup \{i\}) - f(T), \quad \forall S \subset T \subseteq \mathcal{N}, \quad i \in \mathcal{N} \setminus T \tag{3.17}$$

Since f measures the rewards, this property, intuitively, means that the marginal gain of adding a new sensor diminishes as the number of sensors increases. For example, if I go from 0 to 1 sensor, I gain more than going from $n-1$ to n sensors (with $|\mathcal{N}| = n$).

It seems plausible, but it is not trivial to prove it and to prove that it is always like this (independently of S and T).

Equivalently, we could try to prove

$$f(S) + f(T) \geq f(S \cup T) + f(S \cap T), \quad \forall S, T \subseteq \mathcal{N} \tag{3.18}$$

which intuitively means that if we have two (independent) journeys, it is better to place sensors on S for the first and T on the second rather than $S \cup T$ for the first and $S \cap T$ for the second. It is still not trivial to prove...

Since S only affects the definitions on $\Xi(S, \bar{\xi})$ a possible idea could be to exploit the structure of $\Xi(S, \bar{\xi})$ and the property of the min.

How is $\Xi(S, \bar{\xi})$ defined? Intuitively, we can think of $\Xi(S, \bar{\xi})$ as a set of vectors ξ in $\mathbb{R}^N$, where the indices ξ_i for $i \in S$ are fixed and some correlations between the elements of ξ are fixed by the constraint $A\xi \leq b$.

Is therefore obvious that the cardinality of $\Xi(\cdot, \bar{\xi})$ is monotone, i.e.

$$\mid \Xi(S, \bar{\xi}) \mid \geq \mid \Xi(T, \bar{\xi}) \mid, \quad S \subseteq T \tag{3.19}$$

An analogous argument can be used to study supermodularity of the function $\mid \Xi(\cdot, \bar{\xi}) \mid$. If all ξ_i elements of ξ were independent, adding another element in S, i.e. considering $\Xi(S \cup \{i\}, \bar{\xi})$, would reduce the cardinality of $\Xi(S)$ independently of S (just fixes ξ_i and the others elements would continue a vary in the same way). Introducing the correlation in the constraint $A\xi \leq b$, we have that each index fixed ξ_i, also reduces the uncertain set for some of the other elements $\xi_j, j \neq i$.

If we consider the case in which we go from $\Xi(S, \bar{\xi})$ to $\Xi(S \cup \{i\}, \bar{\xi})$, we would have that the reduction in cardinality of the uncertain set is bigger if S is smaller, because bigger S means that the set of possible values of ξ_i is smaller and therefore fixing it will reduce less the cardinality of $\Xi(S, \bar{\xi})$. It follows the supermodularity of the cardinality in the form:

$$\mid \Xi(S \cup \{i\}, \bar{\xi}) \mid - \mid \Xi(S, \bar{\xi}) \mid \leq \mid \Xi(T \cup \{i\}, \bar{\xi}) \mid - \mid \Xi(T, \bar{\xi}) \mid, \quad S \subseteq T, i \notin S \cup T \tag{3.20}$$

From the supermodularity of the cardinality it easily follows the submodularity of the minimum by the fact that the minimum is monotonically not increasing in the set of definition.

It follows that $\tilde{f}_y(S; \bar{\xi}) := \min_{\xi \in \Xi(S, \bar{\xi})} \xi^t y$ is submodular in S, fixed y. However the pointwise maximum of submodular functions is not submodular, so $f(S; \bar{\xi} == \max_{y \in \mathcal{Y}} \min_{\xi \in \Xi(S, \bar{\xi})} \xi^t y$ is not guaranteed to be submodular.

Also the point 2. rises some difficulties, since the ambiguity set is not clear in our case.

Conclusion

This report has examined the use of submodularity in developing tractable formulations of distributionally robust optimization (DRO) problems. We first provided an overview of both DRO and submodularity, then presented the state-of-the-art on how submodular and supermodular properties can be exploited in this context. In particular, we focused on conditions under which such properties enable the design of efficient solution methods.

We found that the literature in this field is active but still developing, with several possible uses of submodularity—both in the definition of the objective function and in the construction of the ambiguity set. We also observed that supermodularity plays an important role, often providing complementary perspectives to submodular optimization. Alongside the presentation of theoretical and methodological results, we reviewed a range of applications, including some of the most well-known problems in operations research.

We concluded by suggesting potential applications to the sensor placement problem, outlining how it is formulated in the literature and how it might be addressed using existing results. This final chapter does not yet offer a complete and systematic approach to the problem, which naturally suggests directions for future research—starting with a full formulation of the sensor placement orienteering problem that leverages submodular or supermodular structure.

More broadly, this review opens avenues for a deeper exploration of submodularity in the context of DRO, bringing together results from a promising and dynamic area. This includes identifying new applications—both in simplified "toy" models and in real-world settings—and developing new methodologies that can unify and generalize existing isolated results.

In summary, this work provides a comprehensive review of the current state of submodularity in DRO. While substantial progress has been made, addressing the highlighted limitations and pursuing the proposed research directions could lead to a richer understanding of the interplay between these fields, ultimately enabling the development of more efficient and realistic methods for optimal decision-making under uncertainty.

Bibliography

Agrawal, Shipra, Yichuan Ding, Amin Saberi, and Yinyu Ye (Feb. 2012). "Price of Correlations in Stochastic Optimization". In: *Operations Research* 60.

Bertsimas, Dimitris, David B. Brown, and Constantine Caramanis (2010). *Theory and Applications of Robust Optimization*. arXiv: 1010.5445 [math.OC].

Chen, Louis, Divya Padmanabhan, Chee Chin Lim, and Karthik Natarajan (2020). "Correlation Robust Influence Maximization". In: *Advances in Neural Information Processing Systems*. Ed. by H. Larochelle, M. Ranzato, R. Hadsell, M.F. Balcan, and H. Lin. Vol. 33. Curran Associates, Inc., pp. 7078–7089.

Golovin, Daniel and Andreas Krause (Sept. 2011). "Adaptive submodularity: theory and applications in active learning and stochastic optimization". In: *J. Artif. Int. Res.* 42.1, 427–486.

Hanasusanto, Grani Adiwena, Daniel Kuhn, and Wolfram Wiesemann (2016). *A Comment on "Computational Complexity of Stochastic Programming Problems"*. Heidelberg.

Krause, Andreas and Carlos Guestrin (Jan. 2007). "Near-optimal Observation Selection using Submodular Functions." In: pp. 1650–1654.

Li, Ruolin and Negar Mehr (May 2023). "Submodularity of optimal sensor placement for traffic networks". In: *Transportation Research Part B: Methodological* 171, pp. 29–43.

Long, Daniel, Jin Qi, and Aiqi Zhang (Apr. 2023). "Supermodularity in Two-Stage Distributionally Robust Optimization". In: *Management Science* 70.

Natarajan, Karthik, Divya Padmanabhan, and Arjun Ramachandra (2023). *Distributionally robust optimization through the lens of submodularity*. arXiv: 2312.04890 [math.OC].

Paradiso, Rosario, Angelos Georghiou, Said Dabia, and Denise Tönissen (2022). *Exact and Approximate Schemes for Robust Optimization Problems with Decision Dependent Information Discovery*. arXiv: 2208.04115 [math.OC].

Staib, Matthew, Bryan Wilder, and Stefanie Jegelka (2018). *Distributionally Robust Submodular Maximization*. arXiv: 1802.05249 [cs.LG].