PRICE SIGNAL QUALITY
IN STOCHASTIC ENERGY OPTIMIZATION

Clara Macêdo Lage

Dissertation
jointly presented to
IMPA
Instituto nacional de Matemática Pura e Aplicada, in Brazil
for
a PhD degree in mathematics
and
Université de Paris I Panthéon-Sorbonne, in France
for
a PhD degree in applied mathematics

Advisors: Jean-Marc Bonnisseau (Paris I)
Mikhail Solodov (IMPA)

Coadvisor: Claudia Alejandra Sagastizábal (UNICAMP)

work developed under a CIFRE contract between Paris 1 and ENGIE

April 2020

Abstract

MACEDO LAGE, C. **Price signal quality in stochastic energy optimization**. 72 pp. Dissertation - joint PhD IMPA and Paris I, 2020

For long time, due to its complexity in terms of generation, distribution, and variety of sources, the management of energy has been a rich source of applied mathematical problems. In regions where hydro-generation is the main source of energy, the random components of rain and snow that arrive to the reservoirs in a given time period affect the interaction of a low-cost technology with other ones in the power system. Stochastic Optimization plays a central role in this context.

This work considers Energy Management problems from a Two-Stage Stochastic perspective, introducing a dual regularization that provides the minimal norm Lagrange Multiplier. This result is particularly useful, since Lagrange Multipliers can be understood as opportunity costs that provide price signals for the power system.

Several informative cases are presented to help understading the problem under consideration. An analytical example, combining optimization and probability, illustrates the mathematics involved. Numerical results on a large number of toy problems make it possible to test the impact of the approach on the price histograms. The proposal is assessed on real data comparing non-regularized and regularized problems in terms of its real interpretation for the Northern European region, therefore connecting practical and theoretical areas of knowledge in the Energy Optimization area.

AMS Subject Classification: Primary 90C15, 65K05 ; Secondary 90C25, 65K10, 46N10.

Keywords: multiplier stability, dual regularization, price signals, two-stage stochastic programming, energy optimization

Contents

Chapter 1

Introduction

1.1 Motivation

Energy optimization is, by nature, an essentially interdisciplinary area, where advanced mathematics, statistics and engineering cooperate. The hydro-generation model is one example of the success of such collaboration [PP91], [MMF17]. In fact, scientific articles in mathematics often use energy problems as illustrations of "real-life" applications; see, e.g., [Sha11], [DCW00], [ZM13], as just a few (of the numerous) examples.

This work considers stochastic optimization problems related to hydro-dominated electric systems. Such energy systems have some particularities that need to be carefully addressed. Two issues are especially important: first, hydro-energy has the particular capacity of being stored (held in reserve) in the form of water in the reservoirs. Second, hydro-power has low cost of generation when compared to other sources, such as coal or gas. Regarding the former, if we were restricted to one isolated period, it would be advantageous to use all the hydraulic potential to fulfill the demand, because the generated energy is practically free. However, considering the possibility of storage, this decision is far from optimal in long-time considerations. For example, in Brazil and in Northern Europe long-time planning of hydro-power management is absolutely essential for the proper functioning of the whole energy system.

Another characteristic of managing hydro-power plants is the necessity of considering different possibilities for the stream inflows, since their amount can deeply affect decisions to be taken. In this context, two-stage stochastic programs can be used to model uncertainty, always minimizing the expectation of the recourse cost at each time in the considered period. The uncertain data is represented by a sample with a set of scenarios, to which some probability is attached.

Typically, in energy generation problems, the objective function includes the cost of production for each power plant. An important constraint is demand satisfaction, that is, the agents have to generate enough energy to attend the demand through the considered time horizon. The growth rate of the overall cost with respect to the demand is the *marginal cost*, used as a *price signal*. Having different scenarios for the streamflow results in different generation schedules to attend the demand, and in different price signals. A common practice in the energy sector is to average the corresponding signals and use the resulting mean price to guide the given company's business strategies. Suppose the company has two managers, say in two different locations. Each manager determines price signals using a sample with the same number of scenarios, but not necessarily the *same scenarios*. The (more-or-less) common belief is that, if the scenario sets are sufficiently large, the multiplier empirical distributions obtained with both samples will be alike and, hence, the two averaged prices will be similar; in some sense, statistically the same. This is clearly desirable, as then our two managers are likely to make similar/consistent business decisions, as common sense dictates in this situation. The present study analyzes in detail this issue, from the point of view of computational solution of the problem and the resulting potential discrepancies. See Chapter 3 and Section 3.2.

When uncertainty is present in the problem, the price becomes a random output that appears to be highly sensitive to changes in the probability of scenarios. We propose a new regularization approach that stabilizes the price signal, a dual regularization. We compute the dual recourse function and add a regularizing term involving the square norm of the price signal. The new recourse function has some mathematical advantages, as it is a smooth function. As we see in chapter 4, its derivative is readily computable. We can also come back to the primal problem and interpret the future cost in terms of the original constraints, giving some intuitions about the meaning of the dual regularization proposed in this work.

Dealing with uncertainty means also dealing with imperfect approximations of a distribution and instability. In stochastic optimization the stability of primal variables with respect to distributions has already been studied numerically and theoretically ([Roe03b], [HBT18], [HW01]). The theoretical approaches usually establish the regularity of a function that links the distribution and the first-stage decision. Given their important role of providing price signals, in this work we rather perform a stability analysis of the *dual* variables.

While literature on stability of the primal decisions is vast ([Kum84], [KR92], [LXL11]), the only other study for dual solutions that we are aware of is [ZM13]. Most of the literature analyzes how the optimal (primal) decision behaves as a function of the probability distribution from which scenarios are sampled. The idea is to consider the different samples as being a perturbation of a probability distribution. From this viewpoint, stability amounts to showing some kind of weak Lipschitz property of the solution mapping, as a function of the perturbation. In particular, [KR92] considers an abstract linear normed space and shows that the solution mapping is locally upper-Lipschitz, and also differentiable (Theorem 7.2 therein). The work [LXL11] defines a special distance in the set of probabilities, and proves a similar local Lipschitz property for the primal solution mapping. In [Kum84], the local Lipschitz property is shown by means of generalized equations and the Implicit Function Theorem.

The work [ZM13] deserves a special comment, as it also deals with the issue of stabilizing price signals in energy problems, from a different perspective. The approach therein is employed to solve a short-term electricity production management problem that has 10^6 variables and 10^6 constraints. For a power system with 200 plants, the model covers 48 hours that are discretized in half-hour steps, see [Hec+10]. Because of the large scale and the practical necessity of solving the problem in a couple of hours at most, certain subproblems that arise when applying Lagrangian relaxation cannot be solved to optimality. The (inexact) oracle computing the problem data returns an approximation that causes instability in the dual solution, when considering consecutive time steps (electricity prices are not meant to jump in a bang-bang from one half hour to the next one). This phenomenon occurs even when the inaccuracy is small. The article [ZM13] proposes a regularization with respect to *total variation* of price signals that yields satisfactory results when a bundle method handles the inexactness.

In the context of this work, uncertainty is in perturbations of the right-hand side of some equality constraints of the stochastic optimization problem. We are looking at instability with respect to samples, and not with respect to time. Rather than controlling the accuracy of information used to iteratively determine optimal Lagrange multipliers, our aim is to control the variance of the perturbation induced by a sampling process.

1.2 Contributions

The PhD work reported in this manuscript was developed under CIFRE contract number 2016-1416, between Université de Paris I Panthéon-Sorbonne and ENGIE. Academically, this PhD is joint between IMPA and Université de Paris I. This (rare) combination (IMPA–Paris 1–ENGIE) was then favorable to appreciate and promote the combination of theory and practice, contributing to the interchange between Optimization and energy management areas. This manuscript aims to share contributions in the area of Stochastic Optimization and Energy Management.

The dual regularized future-cost function is analyzed using convex analysis and parametric optimization results. Theorem 4.1.2 describes the form of future-cost function in the dual regularized problem. This result can be useful for future works (numerically and theoretically) since it helps us to analytically and intuitively understand the behavior of the dual regularized problem.

The asymptotic results of section 4.2 are already published in [LSS19b] along with some numerical tests, here presented in section 5.1. These results establish the link between the dual regularized problem and the original one, and, most importantly, provide a way to find the minimal norm Price Signal in the multipliers set.

Modeling and computational work in chapter 5 are also original. It is different from [ZM13], where another dual regularization function is considered. The presented study reflects a comprehensive analysis of dual and primal variables in a complex "real-life" optimization problem of ENGIE.

1.3 Organization of the Manuscript

This manuscript is organized in six chapters. In chapter 2, we present some background material on results in convex and stochastic optimization theory that are used in the sequel. Particularly, we introduce the recourse function (see section 2.4) and the shadow prices from the point of view of a perturbation function (see section 2.5). This is an interesting interpretation for our proposal.

Chapter 3 presents the Energy Management problem and its notation. As our application is the Northern Eastern model, we focus in this particular energy system, even if the model is also used for other regions with important participation of hydro-power plants, including the Brazilian case. We also give some interesting examples of price signals. One of the examples, in section 3.2, is a simplification of the Energy Management problem in a *two-stage linear program*, and allows us to visualize the instability of price signals.

Chapter 4 contains the introduction to dual regularization and the main properties of the regularized recourse function. These properties have to do with asymptotic results for primal and dual solutions, that is, the description of the behavior of dual regularized solution sets when the regularized problem approaches the original one. Section 4.3 also gives a numerical example to illustrate the theory.

Finally, in chapter 5, we present numerical results for the dual regularized problem. These results come from different types of problems under consideration. In section 5.1, a set of "toy" problems is used to test the effect of regularization for different parameters. Section 5.2 describes results for a simple but very illustrative example of a recourse function, There, we show how the distribution of price signal changes with the regularization parameter. Section 5.3 analyzes the behavior of the regularized price signals on a simplification of the energy generation problem in Northern Europe, presented section 3.2.3. Finally, results on a "real-world" problem in a partnership with ENGIE is presented, using the Northern Europe energy system (see section 5.4).

Chapter 2

Background Material

Optimization under uncertainty is crucially important in many areas, such as finance ([PR07], [DP09]), logistics [FHR13], energy ([MMF17], [GG04], [Per+05]), petroleum industry [RHS10]; just to mention a few. All these applications call for a rigorous mathematical theory about optimization under uncertainty, ever evolving, and for stochastic algorithms to solve these problems in practice.

The expected value is the most used estimator of the impact that different possible realizations have in the objective function of the stochastic optimization problem. Possible reasons for that are the simple intuition of expectation and its good mathematical properties. Let (ξ, Ξ) be a random variable, and $F : \Omega \times E \to \mathbf{R}$ be the objective function. A classical formulation of an optimization problem under uncertainty is:

$$\min_{x \in X} \{f(x) = \mathbf{E}[F(x, \xi)]\}, \tag{2.1}$$

where $X \subset \Omega$ is the feasible set. Because of the *Law of Large Numbers* we know that the expected value can be approximated by a discretization of ξ's distribution. In real-life problems, this discretization in frequently a data set that can be artificially generated or comes from a database.

Formulation (2.1) implies that the decision variable x will be optimized without the knowledge of the value ξ. This framework characterizes *here-and-now* decisions, in contrast to *wait-and-see* decisions, common in *two-stage* stochastic problems (for a formal mathematical formulation; see section 2.1). In applications, the random variable ξ usually represents an event in the future, and x a decision made before the realization of this event.

Lagrange multipliers give useful information in the theory of optimization and applications, in general. In the energy sector, it is probably even more so, due to their interpretation as shadow prices. These multipliers associated with constraints in the feasible set X, estimate the rate of growth of the function f with respect to changes in the constraint right-hand side (when X is represented by equalities/inequalities, as is typical).

Since uncertainty is an intrinsic part of the type of problems under consideration, it brings up some questions about how the distribution of the random variable ξ affects the decision variable and Lagrange multipliers. The question is natural, and naturally important. Thus, there exist some publications in this direction: [DLR03], [**Implicit**], [GE07], particularly dealing with the primal variable. Perturbations in the dual variable with respect to changes in the right-hand side constraints are considered in [ZM13].

In this chapter, in section 2.1, we present the mathematical formulation of two-stage stochastic programs. Sections 2.3 and 2.4 describe the essentials of the Lagrange multiplier theory and how it applies to some properties of the *recourse function*. Finally, in section 2.5 we discuss the concept of *shadow prices*, which is important for understanding the theoretical and applied aspects of the Energy problem in eventual consideration.

2.1 Two-stage Stochastic Linear Programs

For decision vectors $x_1 \in \mathbb{R}^{n_1}$ and $x_2 \in \mathbb{R}^{n_2}$, a right-hand side vector $h \in \mathbb{R}^m$, matrices T and W of appropriate dimensions and linear costs $F_i \in \mathbb{R}^{n_i}$, $i = 1, 2$, consider the following linear programming (LP) problem

$$\begin{cases} \min\limits_{x_1,x_2} & \langle F_1, x_1 \rangle + \langle F_2, x_2 \rangle \\ \text{s.t.} & x_1 \geq 0\,, x_2 \geq 0 \\ & Tx_1 + Wx_2 = h\,. \end{cases} \tag{2.2}$$

If the decision in (2.2) is actually in two levels, the problem becomes

$$\begin{cases} \min\limits_{x_1} & \langle F_1, x_1 \rangle + \mathbb{Q}(x_1) \\ \text{s.t.} & x_1 \geq 0 \end{cases}$$

where

$$\mathbb{Q}(x_1) := \begin{cases} \min\limits_{x_2} & \langle F_2, x_2 \rangle \\ \text{s.t.} & x_2 \geq 0 \\ & Wx_2 = h - Tx_1 \end{cases} = \begin{cases} \max\limits_{\pi} & \langle \pi, h - Tx_1 \rangle \\ & \\ \text{s.t.} & W^\top \pi \leq F_2\,, \end{cases}$$

where π denotes Lagrange multiplier (dual variable) for the equality constraint in (2.2). The matrices T and W are called, respectively, technology and recourse. When h depends on $\xi \in \Xi$, a random variable, the problem becomes:

$$\begin{cases} \min\limits_{x_1,x_2(\xi)} & \langle F_1, x_1 \rangle + \mathbb{E}_P[\,\langle F_2, x_2(\xi) \rangle\,] \\ & \\ \text{s.t.} & x_2(\xi) \geq 0 \ x_1 \geq 0 \\ & Wx_2(\xi) = h(\xi) - Tx_1 \end{cases} \tag{2.3}$$

We are interested in dual formulations of stochastic two-stage programs, [EO75]. Since our work includes computations, we work with a finite set of scenarios to represent the uncertainty.

In the case where Ξ is a discrete set: $\Xi = (\xi^1, \xi^2, \dots, \xi^{\mathbb{S}})$, and the distribution of ξ is the probability vector $p = (p^1, p^2, \dots, p^{\mathbb{S}})$, such that $p^s > 0$ for $s \in \{1, \dots, \mathbb{S}\}$ and $\sum_{s=1}^{\mathbb{S}} p^s = 1$, the problem can be written as:

$$\begin{cases} \min\limits_{x_1, x_2(\xi^s)} & \langle F_1, x_1\rangle + \sum\limits_{s=1}^{\mathbb{S}} \langle F_2, x_2(\xi^s)\rangle \\ \text{s.t.} & x_2(\xi^s) \geq 0 \; x_1 \geq 0 \quad \forall s \in \{1, \dots, \mathbb{S}\} \\ & W x_2(\xi^s) = h(\xi^s) - T x_1 \quad \forall s \in \{1, \dots, \mathbb{S}\}. \end{cases} \tag{2.4}$$

Choosing a scenario ξ^s, the second-stage problem for this fixed scenario has the form:

$$\mathbb{Q}(x_1; \xi^s) := \begin{cases} \min\limits_{x_2} & \langle F_2, x_2\rangle \\ \text{s.t.} & x_2 \geq 0 \\ & W x_2 = h(\xi^s) - T x_1 \end{cases} = \begin{cases} \max\limits_{\pi} & \langle \pi, h(\xi^s) - T x_1\rangle \\ \text{s.t.} & W^\top \pi \leq F_2. \end{cases} \tag{2.5}$$

A primal-dual solution to problem (2.4) is denoted by $(\bar{x}, \bar{\pi})$, where the primal solution is the vector $\bar{x} = (\bar{x}_1, \bar{x}_2(\xi^1), \dots, \bar{x}_2(\xi^{\mathbb{S}}))$, and the dual solution is the vector $\bar{\pi} = (\bar{\pi}(\xi^1), \dots, \pi(\xi^{\mathbb{S}}))$. If we fix $\bar{x}_1$ the primal solution of the first-stage problem, the second-stage primal solution $\bar{x}_2(\xi^s)$ is also a solution to $\mathbb{Q}(\bar{x}_1, \xi^s)$.

Notation 2.1.1. *The second-stage solution to $\mathbb{Q}(\cdot, \xi^s)$ is denoted by $\bar{x}_2(\xi^s)$ or $\bar{x}_2^s$. The same for the right-hand side uncertainty and the dual solution, respectively denoted $h(\xi^s)$ or h^s and $\bar{\pi}(\xi^s)$ or π^s. When we want to clarify the dependence on x_1, the dual solution $\bar{\pi}$ is written as $\bar{\pi}_{x_1}^s$.*

Other constraints can also be included, depending on application. For matrices B_1 and B_2 with appropriate dimensions, we have that

$$\begin{cases} \min\limits_{x_1, x_2(\xi^s)} & \langle F_1, x_1\rangle + \mathbb{E}_P[\,\langle F_2, x_2(\xi^s)\rangle\,] \\ \text{s.t.} & x_2(\xi^s) \geq 0 \; x_1 \geq 0 \qquad \text{for } s \in \mathbb{S} \\ & W x_2(\xi^s) = h(\xi^s) - T x_1 \qquad \text{for } s \in \mathbb{S} \\ & B_2 x_2(\xi^s) \leq b_2 \qquad \text{for } s \in \mathbb{S} \\ & B_1 x_1 \leq b_1. \end{cases} \tag{2.6}$$

Again, choosing a scenario ξ^s fixes the right-hand side term to h^s and the second-stage problem for this fixed scenario has the form:

$$\mathbb{Q}(x_1;\xi^s) := \begin{cases} \min\limits_{x_2} & \langle F_2, x_2\rangle \\ \text{s.t.} & x_2 \geq 0 \\ & B_2 x_2 \leq b_2 \\ & W x_2 = h^s - T x_1 \end{cases} = \begin{cases} \max\limits_{\pi,\lambda} & \langle \pi, h^s - T x_1\rangle - \langle \lambda, b_2\rangle \\ \text{s.t.} & -B_2^\top \lambda + W^\top \pi \leq F_2 \\ & \lambda \geq 0. \end{cases} \tag{2.7}$$

The function $\mathbb{Q}$ is called *recourse function* or *future-cost function*, sometimes cost-to-go functions. To simplify notation, we denote A as follows:

$$A = \begin{bmatrix} T & W & 0 & \dots & 0 \\ T & 0 & W & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ T & 0 & 0 & \dots & W \end{bmatrix},$$

and the uncertainty $\xi = (\xi^1, \dots, \xi^{\mathbb{S}})$, $B = [B_1|B_2]$, and $b = (b_1, b_2)$. If $p^s,\ s \in \{1, \dots, \mathbb{S}\}$ represents the probabilities of scenarios $\xi^s,\ s \in \{1, \dots, \mathbb{S}\}$, $x = (x_1, x_2^1, ..., x_2^s)$, and $h(\xi) = (h^1, ..., h^s)$ we can rewrite problem (2.6) as:

$$\begin{cases} \min\limits_{x} & c^\top x \\ \text{s.t.} & x \in X \\ & Ax - h(\xi) = 0\,, \end{cases} \qquad \text{where} \quad \begin{array}{l} c := (F_1, p^1 F_2, \dots, p^{\mathbb{S}} F_2) \\ X := \{x \in \mathbb{R}^n \mid x \geq 0\ Bx \leq b\}, \text{ and} \\ h(\xi) \in \{y \mid y = Ax,\ x \in X\}\,. \end{array}$$

2.2 Multi-stage Stochastic Programs

When uncertainty reveals progressively, decisions need to be made in stages $t \in \{1, 2, \dots, \mathbb{T}\}$, for $\mathbb{T} > 2$, and we have a *Multi-stage Stochastic Program.*

The random variables ξ_t represent the elements necessary to model what happens between time $t-1$ and t. We would like to estimate how this random variable affects our perception of the future cost. The most common estimator, that we have already seen in *Two-Stage Stochastic Programs*, is the expected value.

In a Multi-stage Stochastic Program, $\xi_{[t]} = (\xi_1, \xi_2, \dots, \xi_t)$ is the history of the process up to time t. It is usually assumed that the problem is *Stage-wise Independent*, that is, that ξ_{t+1} is independent from $\xi_{[t]}$.

For each stage, we make a decision considering *Cost-to-go Function.* This function is given by:

$$\mathbb{Q}_t(x_{t-1};\xi_t) := \begin{cases} \min\limits_{x_t} & \langle c_t, x_t\rangle + \mathbb{E}\left[\mathbb{Q}_{t+1}(x_t)\right] \\ \text{s.t.} & x_t \geq 0 \\ & B_t x_{t-1} + A_t x_t = b_t\,. \end{cases}$$

The set of functions $\bar{x}_t = \bar{x}_t(\xi_{[t]})$, for $t \in \{1, \dots, \mathbb{T}\}$, is said to be an *implementable policy*, if it is sequence of decisions made assuming knowledge of past events only that satisfies the constraints.

Solving a Multi-stage Problem Computationally

When ξ_t is a discrete random variable $\xi_t \in \{\xi^1, \ldots, \xi^{\mathbb{S}}\}$, the most natural mathematical object used to model what happens in a multistage problem is a graph in form of tree. This graph, called *tree of scenarios*, is a tree in which each edge represents a scenario $\xi_t^s, s \in \{1, \ldots, \mathbb{S}\}$, each stage is a tree generation and for each node we associate a decision $x_t(\xi_{[t]})$ that is taken based on past events and evaluating the *cost-to-go function*.

From this point of view, if we want to estimate the future, we need to parse all possible paths in the future and compute the expected value. This estimation, even if rigorous, is very expensive in time and computer memory, which makes its execution impractical. Some algorithms, being the most known SDDP [PP91], have clever forms to explore the tree of scenarios without covering each possible path.

Rolling-Horizon Perspective: This is an alternative to covering the tree of scenarios. In this model, we divide the problem in a set of *two-stage problems*. For time one, we consider the first stage as time one and other stages as a second stage. Each possible path of the tree is a scenario $\xi_2^s, s \in \{1, \ldots, \mathbb{S}^{\mathbb{S}}\}$. We solve this two-stage problem and make a decision $x_1(\xi_{[1]})$. In time t, for each ξ_t as initial condition, the first-stage problem considers time t and the second stage includes $t' \in \{t+1, \ldots, \mathbb{T}\}$. Again, we make a decision $x_t(\xi_{[t]})$. This decision takes into account just the past events, and so it is an *implementable policy.* Once again, parsing the whole tree is not possible, because of the huge quantity of operations required. Usually, the rolling-horizon algorithm [GS12] considers just a small amount of paths per stage, as in the forward path of SDDP.

2.3 Duality and Lagrange Multipliers

As seen in the mathematical formulation above, one must look at primal and dual problems to connect to the recourse function. In this section, the basic theory of duality and Lagrange multipliers is briefly recalled, to set the stage for further developments.

Consider the standard optimization problem:

$$\begin{cases} \min\limits_{x} & f(x) \\ \text{s.t.} & h(x) = 0,\ g(x) \leq 0, \end{cases} \tag{2.8}$$

where $x \in \mathbb{R}^n$, $h : \mathbb{R}^n \to \mathbb{R}^l$ and $g : \mathbb{R}^n \to \mathbb{R}^k$. Denote

$$D = \{x : h(x) = 0, g(x) \le 0\}.$$

In this context, problem (2.8) is called *primal* problem. The *Lagrangian* of problem (2.8) is defined as $L : \mathbb{R}^n \times \mathbb{R}^l \times \mathbb{R}^k_+ \to \mathbb{R}$:

$$L(x, \pi, \mu) = f(x) + \langle \pi, h(x) \rangle + \langle \mu, g(x) \rangle .$$

Problem (2.8) is formally equivalent to

$$\min_{x \in D} \sup_{(\pi,\mu) \in \mathbb{R}^l \times \mathbb{R}^k_+} L(x, \pi, \mu).$$

The *dual* to the original problem (2.8) is given by (formally interchanging the min and sup operations):

$$\max_{(\pi,\mu)\in\Gamma} \inf_x L(x, \pi, \mu), \tag{2.9}$$

where

$$\Gamma := \{(\pi, \mu) \in \mathbb{R}^l \times \mathbb{R}^k_+ : \inf_x L(x, \pi, \mu) \ge -\infty\}.$$

When problem (2.8) has certain properties, the dual (2.9) can be easier to solve, and at the same time advance the understanding and solution of the original (primal) problem (2.8). And when f, h and g are linear, problem (2.9) is also linear, as in the case of the recourse function (2.5). In the linear case, primal and dual problems are just equivalent.

Let $\bar{x}$ be a solution of (2.8). The set of Lagrange multipliers associated with $\bar{x}$ is:

$$\Delta = \{(\pi, \mu) \in \mathbb{R}^l \times \mathbb{R}^k_+ : L'_x(\bar{x}, \pi, \mu) = 0,\ \langle \mu, g(\bar{x}) \rangle = 0\}. \tag{2.10}$$

According to the KKT Theorem (see, e.g., [IS05, Theorem 4.2.2]), under certain conditions, if $\bar{x}$ is a solution of (2.8), the set Δ is nonempty.

If $(\pi, \mu) \in \Delta$, we have that

$$-f'(x) = \sum_{i=1}^{l} \pi_i h'_i(\bar{x}) + \sum_{i=1}^{k} \mu_i g'_i(\bar{x}).$$

Also, under certain conditions (convexity and an appropriate constraint qualification), the dual (2.9) and primal (2.8) problems have the same optimal values.

In the case of our eventual application, problem (2.9) is the linear problem:

$$\mathbb{Q}(x_1;\xi^s) := \begin{cases} \min & \langle F_2, x_2 \rangle \\ \text{s.t.} & x_2 \geq 0 \\ & Wx_2 = h(\xi^s) - Tx_1 \end{cases} = \begin{cases} \max & \langle \pi, h(\xi^s) - Tx_1 \rangle \\ \text{s.t.} & W^\top \pi \leq F_2. \end{cases}$$

2.4 The Recourse Function

The recourse function $\mathbb{Q}$ in (2.5) is:

$$\mathbb{Q}(x_1;\xi^s) := \begin{cases} \max\limits_{\pi} & \langle \pi, h(\xi^s) - Tx_1 \rangle \\ \text{s.t.} & W^\top \pi \leq F_2. \end{cases}$$

The function $\mathbb{Q}(x_1, \xi^s)$ is well defined if the corresponding problem has non-empty feasible set for all scenarios ξ^s, a condition called *complete recourse*.

One can also consider the function $\mathbb{Q}$ in the domain D_1 of the variable x_1, defined in the first-stage problem. In (2.4), for example, this domain is: $D_1 = \{x_1 : x_1 \geq 0\}$. When $\mathbb{Q}(x_1, \xi^s)$ has non-empty feasible set for all $x_1 \in D_1$ and for almost every $\xi^s \in \Xi$, the recourse is said to be *relatively complete*.

The function $\mathbb{Q}$ can also be described by:

$$\mathbb{Q}(x_1;\xi^s) = \max_{\{\pi^1,\ldots,\pi^K\}} \left\langle \pi^k, h(\xi^s) - Tx_1 \right\rangle, \tag{2.11}$$

where $\{\pi^1, \ldots, \pi^K\}$ is the set of vertices of the feasible set $\Pi = \{W^\top \pi \leq F_2\}$. Defining $\psi(\pi, x_1) := \left\langle \pi^k, h(\xi^s) - Tx_1 \right\rangle$, we can write

$$\mathbb{Q}(x_1;\xi^s) = \max_{\{\pi^1,\ldots,\pi^K\}} \psi(\pi^k, h(\xi^s) - Tx_1). \tag{2.12}$$

Since ψ is convex and, hence, continuous, $\mathbb{Q}$ is also convex (see, e.g., [IS05, Theorem 3.4.60]). Note that $\psi(\pi, \cdot)$ is linear, implying that $\mathbb{Q}$ is also *piecewise affine*. As a consequence, the recourse function is convex but not everywhere differentiable.

Convex functions, at every point of the effective domain, have the associated subdifferential (a set of certain generalized derivatives).

Definition 2.4.1. *(Subdifferential) Let $f : U \subset \mathbb{R}^n \to \mathbb{R}$ be a convex real-valued function. The subdifferential of f at a point $x_0 \in U$, denoted by $\partial f(x_0)$, is given by*

$$\partial f(x_0) = \{\pi : f(x) - f(x_0) \geq \langle \pi, x - x_0 \rangle\}.$$

Being $\mathbb{Q}$ the maximum of linear functions, for a fixed x_1 we have that:

$$\mathbb{Q}(x_1, \xi^s) = \left\langle \bar{\pi}^s_{x_1}, h(\xi^s) - Tx_1 \right\rangle = \left\langle \bar{\pi}^s_{x_1}, h(\xi^s) \right\rangle - \left\langle T^\top \bar{\pi}^s_{x_1}, x_1 \right\rangle,$$

where $\bar{\pi}^s_{x_1}$ achieves the maximum objective value in formulation (2.11) for a fixed scenario ξ^s. For two feasible points x_1 and $\hat{x}_1$, replacing $\bar{\pi}^s_{x_1}$ for $\bar{\pi}^s_{\hat{x}_1}$, we have that:

$$\mathbb{Q}(x_1, \xi^s) \geq \left\langle \bar{\pi}^s_{\hat{x}_1}, h(\xi^s) - Tx_1 \right\rangle,$$

since $\mathbb{Q}(x_1, \xi^s)$ is a maximum. Then:

$$\mathbb{Q}(x_1, \xi^s) - \mathbb{Q}(\hat{x}_1, \xi^s) \geq \left\langle -T^\top \bar{\pi}^s_{\hat{x}_1}, x_1 \right\rangle - \left\langle -T^\top \bar{\pi}^s_{\hat{x}_1}, \hat{x}_1 \right\rangle = \left\langle -T^\top \bar{\pi}^s_{\hat{x}_1}, x_1 - \hat{x}_1 \right\rangle.$$

In particular,

$$-T^\top \bar{\pi}^s_{x_1} \in \partial \mathbb{Q}(x_1, \xi^s).$$

Summarizing,

Proposition 2.4.2. *For each scenario ξ^s, the recourse function $\mathbb{Q}(\cdot, \xi^s)$ is convex, and for the dual solution $\bar{\pi}^s_{x_1}$ it holds that $-T^\top \bar{\pi}^s_{x_1} \in \partial \mathbb{Q}(x_1)$.*

Remark 2.4.3. The set where $\mathbb{Q}$ is not differentiable is a closed (not necessarily compact) subset of $\mathbb{R}^{n_1}$ in which π is feasible and has the same objective value that the optimal vertex $\bar{\pi}^s : \langle \bar{\pi}^s, h(\xi^s) - Tx_1 \rangle = \langle \pi, h(\xi^s) - Tx_1 \rangle$.

For a fixed scenario ξ^s, and first-stage decision x_1, consider the Lagrangian of the problem $\mathbb{Q}(x_1, \xi^s)$, given by $L : \mathbb{R}^{n_2} \times \mathbb{R}^k \times \mathbb{R}^{n_2}_+ \to \mathbb{R}$:

$$L(x_2, \pi, \mu) = \langle F_2, x_2 \rangle + \langle -h(\xi^s) + Tx_1 + Wx_2, \pi \rangle - \langle \mu, x_2 \rangle.$$

Since $\mathbb{Q}(x_1, \xi^s)$ is a linear optimization problem, we can express its dual problem as a linear problem in terms of π, the first component of the dual variable. The other optimal component $\bar{\mu}$ can be expressed in terms of $\bar{\pi}$ and the constraints of the problem. The dual problem is:

$$\mathbb{Q}(x_1; \xi^s) = \begin{cases} \max\limits_{\pi} & \langle \pi, h(\xi^s) - Tx_1 \rangle \\ \text{s.t.} & W^\top \pi \leq F_2. \end{cases} \tag{2.13}$$

The following is standard from duality theory.

Proposition 2.4.4. *For a fixed x_1 and scenario ξ^s, it holds that*

$$\{\bar{\pi} : \langle \bar{\pi}, h(\xi^s) - Tx_1 \rangle = \mathbb{Q}(x_1, \xi^s)\} = \{\bar{\pi} : \exists \bar{\mu} \text{ such that } (\bar{\pi}, \bar{\mu}) \in \Delta\}.$$

2.5 Shadow Prices

Mathematically,the shadow price provides the rate of change of the objective function with respect to infinitesimal perturbations in the right-hand side of a problem's constraint. For our two-stage problem, the economical interpretation assesses the impact in the recourse function of small variations in the first stage or *here-and-now* decision. The i-th component of the shadow price represents the price of the i-th constraint in the model, that is, the value paid to keep satisfying it, also known as *marginal utility.*

Let $\nu : \mathbb{R}^n \to \bar{\mathbb{R}}$ be the value-function of an optimization problem:

$$\nu(u,v) := \begin{cases} \min\limits_{x} & f(x) \\ \text{s.t.} & \langle h_j, x\rangle - b_j = v_j \; j = \{1,\dots,K\}, \text{ (or } Hx = v), \\ & g_i(x) \le u_i, \; i = \{1,\dots,L\}, \text{ (or } g(x) \le u), \end{cases}$$

where $f : \mathbb{R}^n \to \mathbb{R}$, $g_i : \mathbb{R}^n \to \mathbb{R}$, $h_j \in \mathbb{R}$. The function ν measures variations of the objective function with respect to changes in the right-hand side constrains. We assume that the functions f and g_i are convex. Denote

$$D(u,v) := \{x : Hx = v, \; g(x) \le u\}.$$

Note that $D(u,v)$ can be empty at some points in a neighborhood of $(0,0)$. Thus, for ν to be finite, some assumption is needed. In any case, ν is a decreasing function with respect to u, since $u_i < u'_i$ implies that $D(u,v) \subseteq D(u',v)$. Also, ν is a convex function, as shown in the following proposition.

Proposition 2.5.1. *Suppose f and g_i, $i \in \{1,\dots,L\}$, are convex. Then ν is convex.*

Proof. See Theorem 3.3.3, chapter 3, [HL93]. □

Some properties of the value function ν are related to stability of the associated optimization problem around $(0,0)$. The sensitivity to changes in u can be measured by the behavior of $\nu(0,0) - \nu(u,0)$ for small values of $\|u\|$. Since ν is convex and it has finite value at $(0,0)$, it has a non-empty subdifferential set $\partial\nu(0,0)$. For any element $\pi \in \partial\nu(0,0)$, by definition we have that:

$$\nu(u,v) - \nu(0,0) \ge \langle \pi, (u,v)\rangle \,.$$

If we think of the function as *cost*, subdifferentials provide lower bounds for how much the cost would increase if some of the conditions in the model change. On the other hand, regarding upper bounds, we have the following definition:

Definition 2.5.2. *The primal problem $\nu(O,0)$ is said to be stable if there is $M > 0$ such that for all $u \neq 0$:*

$$\frac{\nu(0,0) - \nu(u,v)}{\|(u,v)\|} \le M.$$

It is desirable to be able to describe the whole subdifferential set of ν, since it is an important indicator. The *strong Slater condition* is equivalent to finite values of ν in a neighborhood of $(0,0)$.

Definition 2.5.3. *(Strong Slater Condition) If H is linear and g_i are convex, the Strong Slater condition is satisfied if the elements h_i are linearly independent and $\exists\, x \in D$ such that $g_i(x) < 0,\ i = 1, \ldots, N$.*

Proposition 2.5.4. *Under our standing assumptions, the strong Slater condition is necessary and sufficient for ν to be finite in a neighborhood of the origin.*

Proof. See Theorem 3.3.3, chapter 3, [HL93]. □

Finally, we can describe the whole set $\partial\nu(0,0)$ in terms of the Lagrange multipliers of the associated problem.

Theorem 2.5.5. *(Strong Duality) If the primal problem $\nu(0,0)$ satisfies the Strong Slater Condition, we have that:*

1. *The dual problem has an optimal solution, and the optimal values of the dual and primal problems are the same.*
2. *π is an optimal solution for the dual problem if and only if $-\pi$ is in the subdifferential of the perturbation function ν at $(u,v) = 0$.*

Proof. Since $D(0,0)$ is not empty and $\nu(0,0)$ is finite, the dual set is nonempty, see Theorem 5.2.19, [IS05], so the dual problem has an optimal solution, and the optimal values of the dual and primal problems are the same, see Theorem 5.2.18, chapter 5, [IS05].

For item 2 see Theorem 3.3.2, chapter 3, [HL93]. □

As a consequence of Proposition 2.5.4 and Theorem 2.5.5, under *Strong Slater Condition*, ν is differentiable if and only if the set of multipliers Δ is a singleton. In this case, we have that:

$$\nu'(0; e_i) = \pi_i,$$

where e_i is the i-th component of the canonical vector, and $\psi'(x; d)$ stands for the usual directional derivate of the function ψ at the point x in the direction d. So, informally speaking, the components of vector π say which components of constraints are more important in the growth of the objective function value.

In the case when f, g, and h are linear, π_i can be directly expressed as a solution of a linear problem, since the dual can be easily computed. Given a vector F and matrices H, G, we have that:

$$\nu(u,v) := \begin{cases} \min\limits_{x} & \langle F, x\rangle \\ \text{s.t.} & Hx = v, \\ & Gx \le u \end{cases} = \begin{cases} \max & \langle v, \pi\rangle - \langle u, \mu\rangle \\ \text{s.t.} & H^\top \pi - G^\top \mu = F, \end{cases} \tag{2.14}$$

and (π, μ) is the full price signal vector.

Chapter 3

On Price Signals for Energy Optimization

Modern energy markets involve a large number of technologies to generate electricity. Finding the best policy with lower prices is a challenging problem. We describe the main elements of the considered energy generation system. Related material, with variations in the problem to be solved that depends on a country and the type of the market, can be consulted in, for instance, ([Hec+10], [PP91], [SBV17]).

3.1 Energy Management Problems

Energy systems like the North European (Nordic) energy market, used in the numerical experiments in Chapter 5, involve several balancing zones, representing a country (say Finland) or a region in a country (say Norway 1 to Norway 5). Typically, intra-zone constraints are demand satisfaction and water-balance equations, while the overall balance of the system is achieved by exchanges between the zones, as represented by the diagram in Figure 3.1.

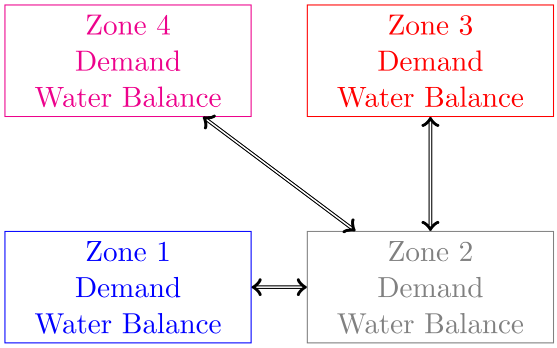
In our real-life application with ENGIE, we consider the problem of managing in an optimal manner the generation of an energy system as in Figure 3.1, over a time horizon of 12 months, with hourly discretization.

In Figure 3.1 not all zones are connected, as this is what happens in some energy systems, as the North European, considered in our numerical tests. Inside each zone, demand must be satisfied and reservoirs follow a dynamic represented by the water-balance equation. In order to write down the corresponding optimization problem, we use the notation described below.

3.1.1 Nomenclature

The notation for the different elements defining the optimization problem is given below.

Figure 3.1: Energy system with balancing zones



- **Sets:**
  - Scenarios $s \in \mathbb{S}$ each one with probability p^s, representing uncertainty ξ^s on water inflows, changing the hydro-power availability.
  - Time steps in the set $\{t \in \mathbb{T}\}$.
  - Balancing zones $\{z_l,\ l \in \mathbb{L}\}$.
  - A zone z_l has thermal power plants $\{i \in \mathbb{I}_l\}$ and hydro-plants $\{j \in \mathbb{J}_l\}$.
  - For each $l \in \mathbb{L}$, $\mathbb{F}_l$ is a set of zones connected with z_l, to import or export energy. When the zone has no connections, this set is empty.
- **Variables at time t:**
  - A hydro-plant $j \in \mathbb{J}_l$ has reservoir level v_j^t and spillage sp_j^t.
  - The generated energy of thermal and hydro-power plants gt_i^t and gh_j^t, respectively.
  - For each $l \in \mathbb{L}$ having nonempty set $\mathbb{F}_l$, and for all $l_1 \in \mathbb{F}_l$, the energy exchanged between zones z_l and z_{l_1} is $f_{l \leftrightarrow l_1}^t$.
  - A possible deficit in generation of zone z_l is represented by an artificial power plant in the set $\mathbb{I}_l$, with very high generating cost and large capacity.
- **Parameters:**
  - For the reservoir in hydro-plant $j \in \mathbb{J}_l$, the water inflow $\mathcal{I}_j^{t,s} = \mathcal{I}_j^s(\xi^s)$, noting that for $t = 1$ this is a deterministic value: $\mathcal{I}_j^{1,s} = \mathcal{I}_j^1$, the same for all scenarios s. The reservoir initial volume is v_j^0, its maximum and minimum levels are $\underline{v}_j$ and $\overline{v}_j$, respectively.

– For thermal power plant $i \in \mathbb{I}_l$ at time t, its maximum generation capacity $\overline{gt}_i^t$ and unit generation cost C_i^t.
– The hydro-cost of hydro-power plant $j \in \mathbb{J}_l$ is null, and its maximum capacity $\overline{gh}_j^t$.
– For zone z_l at time t, the deterministic demand $\mathcal{D}_l^t$.

3.1.2 Formulating the Problem

For a fixed scenario s, the deterministic formulation for the problem of interest is the following:

$$\left\{\begin{array}{lll}\displaystyle\min_{gt_i^t, v_j^t, f_{l\leftrightarrow l_1}^t, sp_j^t} & \displaystyle\sum_{t\in\mathbb{T}}\sum_{l\in\mathbb{L}}\sum_{i\in\mathbb{I}_l} C_i^t gt_i^t & \\ \text{s.t.} & \underline{v}_j^t \le v_j^t \le \overline{v}_j^t \text{ and } 0 \le sp_j^t, & j\in\mathbb{J}_l\,, l\in\mathbb{L}\,, t\in\mathbb{T} \\ & 0 \le gh_j^t \le \overline{gh}_j^t, \quad 0 \le gt_i^t \le \overline{gt}_i^t, & j\in\mathbb{J}_l\,, i\in\mathbb{I}_l\,, l\in\mathbb{L}\,, t\in\mathbb{T} \\ & 0 \le f_{l\leftrightarrow l_1}^t, & l\in\mathbb{L} : l_1\in\mathbb{F}_l\neq\emptyset\,, t\in\mathbb{T} \\ & v_j^t - v_j^{t-1} + gh_j^t + sp_j^t = \mathcal{I}_j^{t,s}, & j\in\mathbb{J}_l\,, l\in\mathbb{L}\,, t\in\mathbb{T} \\ & \displaystyle\sum_{j\in\mathbb{J}_l} gh_j^t + \sum_{i\in\mathbb{I}_l} gt_i^t + \sum_{l_1\in\mathbb{F}_l\neq\emptyset} f_{l\leftrightarrow l_1}^t = \mathcal{D}_l^t, & l\in\mathbb{L}\,, t\in\mathbb{T}\,. \end{array}\right. \tag{3.1}$$

Recall that the demand equation involves no variables $f_{l\leftrightarrow l_1}^t$ for zones z_l without exchanges, as in this case $E_l = \emptyset$. Note also that feasibility is ensured by the spillage and deficit (the artificial power plant generation), which act as slack variables in the equality constraints.

For easier reading, the stochastic version of the linear program (3.1) is given in an abstract format, more suitable for our developments. To this end, we adopt a two-stage approach to handle uncertainty, splitting the time steps into two sets, $\mathbb{T}_1 := \{t \in \mathbb{T} : t \le t_1\}$ and $\mathbb{T}_2 := \{t \in \mathbb{T} : t > t_1\}$. Until time t_1 all data is considered known, with scenarios s corresponding to right hand side uncertainty, relative to times in $\mathbb{T}_2$.

Variables with time index $t \le t_1$ define the first-stage decision vector

$$x_1 := \bigcup_{t\in\mathbb{T}_1} \left\{ (v_j^t, sp_j^t, gh_j^t)_{j\in\mathbb{J}_l}, (gt_i^t)_{i\in\mathbb{I}_l}, (f_{l\leftrightarrow l_1}^t)_{l_1\in\mathbb{F}_l\neq\emptyset}) : l \in \mathbb{L} \right\}, \tag{3.2}$$

which is of the “here-and-now” type. Since we consider that uncertainty reveals at time t_1, all variables with index $t > t_1$ are of the “wait-and-see” type and, hence, denoted by x_2^s for each scenario s:

$$x_2^s := \bigcup_{t\in\mathbb{T}_2} \left\{ (v_j^{t,s}, sp_j^{t,s}, gh_j^{t,s})_{j\in\mathbb{J}_l}, (gt_i^{t,s})_{i\in\mathbb{I}_l}, (f_{l\leftrightarrow l_1}^{t,s})_{l_1\in\mathbb{F}_l\neq\emptyset}) : l \in \mathbb{L} \right\}.$$

The objective function in (3.1) is likewise split, so that we have vectors

F_1 and F_2 of appropriate dimensions satisfying

$$\sum_{t\in\mathbb{T}_1}\sum_{l\in\mathbb{L}}\sum_{i\in\mathbb{I}_l} C_i^t gt_i^t = \langle F_1, x_1\rangle \quad \text{and} \quad \sum_{t\in\mathbb{T}_2}\sum_{l\in\mathbb{L}}\sum_{i\in\mathbb{I}_l} C_i^t gt_i^{t,s} = \langle F_2, x_2^s\rangle \,.$$

In a similar manner, the box constraints in (3.1), written for x_1 and x_2^s, are rewritten as

$$x_1 \le b_1 \quad \text{and} \quad x_2^s \le b_2\,, s\in\mathbb{S}\,,$$

taking appropriate vectors b_1 and b_2. Although not present in (3.1), explicit upper bounds for the spillage and exchanges (variables sp_j^t and $f_{l\leftrightarrow l_1}^t$) can be obtained from the water-balance and demand equality constraints.

Finally, notice that in (3.1) only the water-balance equations couple time steps. In particular, for $t = t_1 + 1$, this gives

$$v_j^{t_1+1,s} - v_j^{t_1} + gt_j^{t_1+1,s} + sp_j^{t_1+1,s} = \mathcal{I}_j^{t_1+1,s}, j\in\mathbb{J}_l\,, l\in\mathbb{L}\,,$$

an equality coupling components of x_2^s with components of x_1. As usual in two-stage stochastic programming, this relation is expressed as

$$Tx_1 + Wx_2^s = h^s\,,$$

where the vector $h^s = h(\xi^s)$ has components given by the right-hand side terms $\mathcal{I}_j^{t,s}$ and $\mathcal{D}_{l_1}^t$, and the technology and recourse matrices T and W have appropriate dimensions. Typically, these matrices are very sparse, as they involve few variables at once. For instance, for identity matrices Id of proper dimension, when $\mathbb{T} = 2$ they are as follows:

$$T := \begin{bmatrix}-Id & 0 & 0 & 0 & 0\end{bmatrix} \in \mathbb{R}^{5I}, \qquad W := \begin{bmatrix}Id & Id & 0 & 0 & 0\end{bmatrix} \in \mathbb{R}^{I}.$$

With the above notation, the two-stage stochastic programming formulation for our energy management problem is:

$$\begin{cases} \min\limits_{x_1,x_2^s} & \langle F_1, x_1\rangle + \sum\limits_{s\in\mathbb{S}} p^s \langle F_2, x_2^s\rangle & \\ \text{s.t.} & 0 \le x_1 \le b_1 & \\ & 0 \le x_2^s \le b_2 & \text{for } s\in\mathbb{S} \\ & Tx_1 + Wx_2^s = h^s & \text{for } s\in\mathbb{S}\,. \end{cases} \tag{3.3}$$

3.2 Examining the Reliability of Price Signals

The price signals given by the demand constraint correspond to components of the optimal multiplier associated with the last constraints in (3.3), with right-hand side vector h^s, for $s\in\mathbb{S}$. A common practice in the energy sector is to average those signals and use the resulting mean price to guide the company business strategies.

Suppose the company has two managers $m = 1$ and $m = 2$, say in different locations. Each manager determines price signals using a set $\mathbb{S}_m$ with the same number of scenarios, but not necessarily the same ones ($\mathbb{S}_1 \neq \mathbb{S}_2$ have the same cardinality). The usual belief is that, if the scenario sets $\mathbb{S}_m$ are sufficiently large, the multiplier empirical distributions obtained with $\mathbb{S}_1$ and with $\mathbb{S}_2$ will be similar and, hence, the two averaged prices will be similar too. This is clearly desirable, as then our two managers are likely to make similar/consistent business decisions. We next give some examples showing that this may not be the case in general.

3.2.1 Analytical Solution

Suppose in (2.3) we consider realizations of some continuous random variable $\xi \in \mathbb{R}$, with cumulative distribution function denoted by $\mathbb{P}$.

In the example, the right-hand side vector $h_s = \xi_s$ is a particular realization of the continuous variable. The second-stage vectors x_2^s have components $(x_2^+(\xi), x_2^-(\xi)) \in \mathbb{R}^2$, with respective scalar costs F_2^+ and F_2^-, satisfying $F_2^- \geq F_2^+ > 0$. We furthermore take $b_1\,, b_2 = +\infty$, $x_1 \in \mathbb{R}$ and let $T = 1$, $W = [1\ -1]$, so that the optimization problem is

$$\left\{\begin{array}{lll} \min\limits_{x_1, x_2^+, x_2^-} & F_1 x_1 + \mathbb{E}[F_2^+ x_2^+(\xi) + F_2^- x_2^-(\xi)] & \\ \text{s.t.} & x_1 \geq 0 & \\ & x_2^+(\xi) \geq 0\,, x_2^-(\xi) \geq 0 & \text{for a.e. } \xi \\ & x_1 + x_2^+(\xi) - x_2^-(\xi) = \xi & \text{for a.e. } \xi\,, \end{array}\right.$$

with $F_1 > 0$ and the feasible set assumed non-empty for a.e. ξ. Rewriting this problem in a two-level formulation,

$$\left\{\begin{array}{ll} \min\limits_{x_1} & F_1 x_1 + \mathbb{E}[\mathbb{Q}(x_1, \xi)] \\ \text{s.t.} & x_1 \geq 0 \end{array}\right. \quad , \quad \mathbb{Q}(x_1, \xi) := \left\{\begin{array}{ll} \min\limits_{x_2^+, x_2^-} & F_2^+ x_2^+ + F_2^- x_2^- \\ \text{s.t.} & x_2^+, x_2^- \geq 0 \\ & x_2^+ - x_2^- = \xi - x_1 \end{array}\right. \tag{3.4}$$

gives, by Linear Programming duality, that

$$\mathbb{Q}(x_1, \xi) := \left\{\begin{array}{ll} \max\limits_{\pi} & \pi(\xi - x_1) \\ \text{s.t.} & -F_2^- \leq \pi \leq F_2^+\,. \end{array}\right.$$

Therefore, the optimal Lagrange multiplier associated with the affine constraint in (3.4) is

$$\pi(x_1, \xi) := \left\{\begin{array}{cl} -F_2^- & \text{if } \xi - x_1 < 0 \\ \text{any element in } [-F_2^-, F_2^+] & \text{if } \xi - x_1 = 0 \\ F_2^+ & \text{if } \xi - x_1 > 0\,. \end{array}\right. \tag{3.5}$$

The recourse function can be written as

$$\mathbb{Q}(x_1, \xi) = F_2^- \max(x_1 - \xi, 0) + F_2^+ \max(\xi - x_1, 0),$$

yielding an explicit form for the expected value:

$$\mathbb{E}[\mathbb{Q}(x_1, \xi)] = F_2^- \mathbb{P}(\xi \le x_1) + F_2^+ \mathbb{P}(\xi \ge x_1) = F_2^+ + (F_2^- - F_2^+)\mathbb{P}(\xi \le x_1)\,.$$

Then problem (3.4) boils down to

$$\min_{x_1 \ge 0} F_1 x_1 + \mathbb{E}[\mathbb{Q}(x_1, \xi) = F_2^+ + \min_{x_1 \ge 0} F_1 x_1 + (F_2^- - F_2^+)\mathbb{P}(\xi \le x_1)\,,$$

The cumulative distribution $\mathbb{P}(\xi \le \cdot)$ is a non-decreasing function. Since, in addition, $F_1 > 0$ and $F_2^- \ge F_2^+$ by assumption, the minimizer is $\bar{x}_1 = 0$. The corresponding optimal price distribution is

$$\bar{\pi}(\xi) = \begin{cases} -F_2^- & \text{if } \xi < 0 \\ \text{any element in } [-F_2^-, F_2^+] & \text{if } \xi = 0 \\ F_2^+ & \text{if } \xi > 0\,. \end{cases}$$

For simplicity, let $F_2^- = F_2^+ = F_2$, and suppose that ξ has a symmetric probability distribution $\mathbb{P}$. Then the continuous price signal for (3.4) has the following mean and variance:

$$\mathbb{E}[\bar{\pi}(\xi)] = 0 \quad \text{and} \quad \mathbb{V}ar[\bar{\pi}(\xi)] = \mathbb{E}[\bar{\pi}(\xi)^2] = F_2^2\,. \tag{3.6}$$

However, if in our example Manager 1 samples only negative numbers while Manager 2 samples only positive numbers, then $\mathbb{S}_1 \subset \mathbb{R}_-$ and $\mathbb{S}_2 \subset \mathbb{R}_+$ will respectively result in the very different (and wrong) empirical signals

$$\begin{array}{llll} \forall s \in \mathbb{S}_1 & \bar{\pi}_1(\xi^s) = -F_2 & \Longrightarrow \mathbb{E}[\bar{\pi}_1] = -F_2 & \text{and} \quad \mathbb{V}ar[\bar{\pi}_1] = 0 \\ \forall s \in \mathbb{S}_2 & \bar{\pi}_2(\xi^s) = F_2 & \Longrightarrow \mathbb{E}[\bar{\pi}_2] = F_2 & \text{and} \quad \mathbb{V}ar[\bar{\pi}_2] = 0\,, \end{array} \tag{3.7}$$

no matter how large are the samples.

Of course, this example illustrates an extreme case and any intermediate situation between the most wrong one (as above) and "right" ones (with $\mathbb{S}_1$ and $\mathbb{S}_2$ containing the same number of positive and negative numbers) are possible. In this sense, the sampling method certainly has an impact. Nevertheless, the stochastic nature of the energy management problem (3.3) still remains, making the issue of producing reliable price signals a real concern.

The next examples show the high variability that the Lagrange multiplier output can exhibit, depending on the method and algorithm used.

3.2.2 Same Problem, Different Output from Different Solvers

Take 2 equiprobable scenarios and let $x_1 \in \mathbb{R}^2$ and $x_2 \in \mathbb{R}^3$. The first-stage cost $F_1 = (2,3)^\top$, while second-stage cost is deterministic $F_2 = (1,0,0)^\top$. The technology and recourse matrices in (2.4) are

$$T := \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \quad \text{and} \quad W := \begin{bmatrix} 1 & -1 & 0 \\ 1 & 0 & -1 \end{bmatrix}.$$

The uncertain right-hand side terms are $h^1 := (1,1)^\top$ and $h^2 := (2,0)^\top$.

For this small example, the feasible set of the dual problem can be computed explicitly. It is the simplex in Figure 3.2, where the Lagrange multiplier set for both scenarios is the edge in blue (if in color; otherwise it is the thicker line).

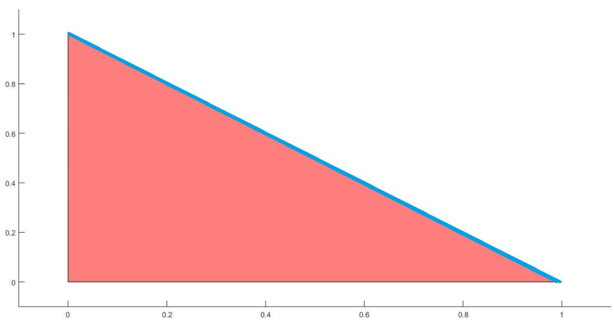


Figure 3.2: The feasible set of the dual problem (the triangle, red if in color), and the set of Lagrange multipliers (the edge in blue if in color; otherwise it is the thicker line)

In this simple case, we tested different algorithms: Simplex method for the linear deterministic problem (LP), the L-Shaped method (LS), and the Proximal Bundle method [LS97] (BM). The point is that they all compute different Lagrange multipliers:

Algorithm	(LP)	(LS)	(BM)
Lagrange multipliers	$(0,-1),(-1,0)$	$(-1,0),(-1,0)$	$(\frac{1}{2},\frac{1}{2}),(\frac{1}{2},\frac{1}{2})$

This is already problematic, by itself. The next consideration is that having as output a solution with the smallest norm is important when dealing with prices, since they are unique and, depending on the application, can have a special meaning. For this example, the minimal-norm Lagrange multiplier is:

$$\hat{\pi} = \Big(\frac{1}{2}, \frac{1}{2}\Big).$$

But out of the three options, only the Proximal Bundle Method provides this multiplier. Moreover, there is no guarantee that this will still be so for some other problem.

3.2.3 Same Problem, Different Price Distribution

We now consider some price signals obtained for two examples. The first one is an academic case, taken from [SH98]. The second one is the energy management problem from a real industrial application.

A simple two-stage linear program

The following tests were ran using Matlab 7.8.0 (R2009a), on an AMD Athlon II X2 240 computer with 2800 MHz, 2 GB RAM, with Ubuntu OS, using MOSEK optimization toolbox for Matlab; see http://www.mosek.com.

In this example we are going to observe how the Lagrange multiplier changes depending on the distribution of the scenarios considered in each sample. Scenarios are generated by a normal distribution $N(\mu, \sigma)$, where the mean value of μ and σ is 2.35 and 1.2, respectively. For each sample P_i, where $i \in \{1, 2, 3\}$, we generate 50 scenarios, and these scenarios define the discrete distribution of the random variable (ξ_i, Ξ), $i = \{1, 2, 3\}$.

The dimensions of W and T are 40×60. The Lagrange multiplier $\pi^s \in \mathbb{R}^{40}$, $s \in \{1, \ldots, 50\}$. $\mathbb{E}[\pi_i]$ is expected value of the component i with respect to the scenarios s. In the next table, we show the expected value for some components:

Sample	$\mathbb{E}[\pi_{33}]$	$\mathbb{E}[\pi_{12}]$
$\mathbb{S}_1$	56.5784	4.2821
$\mathbb{S}_2$	62.6715	2.012
$\mathbb{S}_3$	60.6757	4.6553

Table 3.1: Expected price signals for 50 scenarios in each sample

Component 33 is the larger component of the price signal for this problem, which is an important indicator in applications since it indicates the component in which variations have greater impact on the cost - or on the objective function. We see in table 3.1 that the expected values vary between samples. For component 33, the standard deviation with respect to samples is 3.11, which represents 5.5% of the expected value of the first sample. For component 12, which has the most important variation proportionally, the standard deviation is 1.4, which represents 33% of the expected value for the first component.

These values decrease slowly if we increase the quantity of scenarios in each sample:

Sample	$\mathbb{E}[\pi_{33}]$	$\mathbb{E}[\pi_{12}]$
$\mathbb{S}_1$	55.2371	4.2564
$\mathbb{S}_2$	52.8023	3.8833
$\mathbb{S}_3$	59.9722	4.3294

Table 3.2: Expected price signals for 100 scenarios in each sample

Considering the sparsity and low dimension of the problem, these results show that even simple two-stage problems can be sensitive to changes with respect to samples.

The Nordic Energy System

The industrial problem under consideration follows the schematic diagram given in Figure 3.1 to represent the 12 bidding zones in northern Europe. Zones are connected and can interchange energy. Each of these zones is a component of the price signal, denoted by π_i^s.

In this formulation we summarize the energy generation problem of a whole year in a two-stage problem with stages $t \in \{1, 2\}$. Stage 1 represents winter/autumn and stage 2 represents summer/spring. To do that, we consider the mean of the real data, originally organized by hour, for each part of the year, that is, the mean of demand, inflow and capacity. The other part of the data is maintained as before, as maximum level of reservoirs and flow, that were not originally dependent on time. This modeling makes the problem more tractable, and more directly comparable with the developed theory, since it is a two-stage model.

For this hydro-dominated system, the uncertainty ξ comes from inflows, the most important source of randomness (with a periodic nature, easier to forecast than wind). In Northern Europe inflows tend to be much larger in summer/spring than in winter/autumn, hence justifying our division of the time horizon in two separate stages. The most important variation of inflows occurs in Norway, even if other countries like Sweden have also a significant difference in the inflows of their bidding zones.

As in the previous example, we consider different samples, each one with 50 inflow scenarios. We had available a set of 80 real scenarios that could be used to solve the real problem, however, algorithms do not go through all the three of scenarios. In the case of SDDP, for example, just some branches of the tree are covered, because of the exporential nature of the graph. In Table 3.3 we see the behavior of the expected value of the price and its variance, computed with different samples for the bidding zones 1 and 4 in Norway.

We confirm that significant differences may appear, when using different samples. In the runs reported in Table 3.3, the most important differences are observed for NO4. The fact that prices are computed by solving a lin-

Sample	$\mathbb{E}[\pi_{NO1}]$	$\mathbb{VAR}[\pi_{NO1}]$	$\mathbb{E}[\pi_{NO4}]$	$\mathbb{VAR}[\pi_{NO4}]$
$\mathbb{S}_1$	28,16	51.43	27,34	40.67
$\mathbb{S}_2$	29,50	47.23	24,50	41.80
$\mathbb{S}_3$	27,98	51.70	27,20	43.76

Table 3.3: Expected price signals and variance for zones NO1 and NO4

ear programming problem has certainly an impact, because typical solvers for linear programs return a solution which is a vertex of the (polyhedral) feasible set.

In this experiment, prices tend to oscillate between three main values between 25 and 35, as shown in Figure 3.3, with histograms for the three samples.

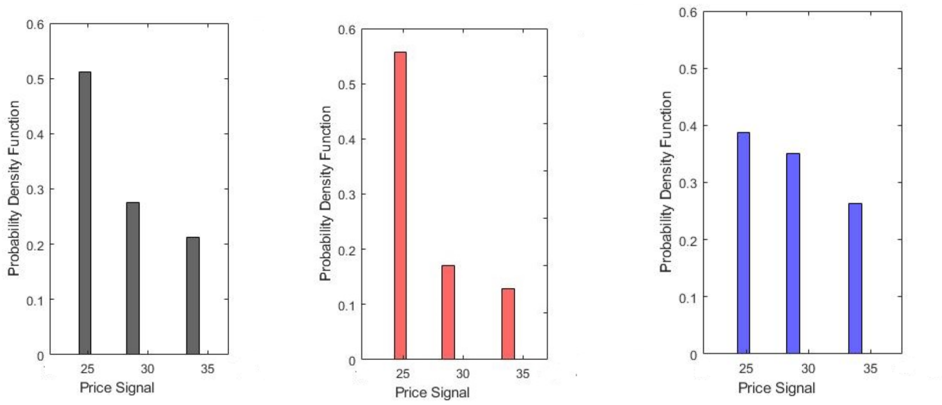


Figure 3.3: Price distribution of three samples with $\beta = 0$

The same behavior was observed with 15 different samples; we only report here three for brevity. If one needed to make a business decision based on these histograms, it would be difficult (basically unclear how) to choose which expected value or variance is the "right" one.

We could clearly increase the number of scenarios, however, beside the fact that there is no guarantee that multipliers will stabilize, it is not always possible in practice because of time and computational limitations.

Chapter 4

The Theory of Dual Regularized Price Signals

Dealing with uncertainty means also dealing with imperfect approximations of a distribution, and with the inherent instability. In stochastic optimization, stability of primal variables with respect to distributions has already been studied numerically and theoretically ([Roe03b], [HBT18], [HW01]). Theoretical approaches usually establish certain regularity properties of a function that links the distribution and the first-stage decision.

In this work, regularization is understood from the more computationally oriented perspective. In some ways, it can be (loosely) related to the recent developments in Machine Learning and Artificial Intelligence, with applications in fields like computer science, statistics and deep learning. In these applications, regularization usually aims at avoiding over-fitting (undesirable dependence on specific, possibly eventually not very representative, data) by penalizing complex solutions. Along these lines, in our case we also want the *marginal cost* to be less sensitive to the data input. Some previous works have successfully implemented *primal* regularization in stochastic optimization problems, see ([AOS19], [HS94]). Our focus is *dual*.

As seen in Chapter 2, in two-stage stochastic problems the recourse function is affine by parts, which means that it is not differentiable. The Lagrange multiplier gives elements in the subdifferential set of the recourse function:

$$-T^\top \pi_{x_1}(\xi) \in \partial_x \mathbb{Q}(x_1, \xi),$$

for all $x_1 \in X$ and a fixed $\xi \in E$. If we could make the recourse function $\mathbb{Q}$ differentiable, we would have a unique Lagrange multiplier π_{x_1}, that would then be a smoother function with respect to the first-stage decision x_1.

We introduce a dual regularization that adds a quadratic term to the recourse function, written in its dual formulation. Informally, the technique can be seen as a way to gain regularity/stability, giving a better control of the Lagrange multiplier computed. We show that considering a sequence of regularization parameters $\beta_k \to 0$, the corresponding sequence of regularized Lagrange multipliers π^{β_k} converges to a multiplier in the original set (when $\beta = 0$) and that it is the minimal-norm element in the set. The economic interest of this result is clear, since the mechanism would systematically yield the price with smallest possible norm. The theory passes through a variational analysis and non-linear computational analysis perspectives.

In this chapter we present the dual regularized two-stage problem and discuss its mains analytical properties. In section 4.1 we start with a simple but very illustrative example of regularized recourse function. We discuss the main properties of the regularized future-cost function. These properties are used in section 4.2, where we talk about the asymptotic behavior for primal and dual variables. Finally, in section 4.3 we give a numerical example to illustrate our results.

Most of the material presented below was published in the article [LSS19b].

4.1 Dual Regularization of Two-Stage Problems

The stabilization device proposed in [LSS19b] aims at producing multipliers with minimal norm. Like in the simple example above, consider the two-level reformulation of (3.3),

$$\begin{cases} \min\limits_{x_1} & \langle F_1, x_1 \rangle + \sum\limits_{s\in\mathbb{S}} p^s\, \mathbb{Q}^s(x_1) \\ \text{s.t.} & 0 \le x_1 \le b_1\,, \end{cases} \tag{4.1}$$

making use of the following recourse function:

$$\mathbb{Q}^s(x_1) := \begin{cases} \min\limits_{x_2} & \langle F_2, x_2 \rangle \\ \text{s.t.} & 0 \le x_2 \le b_2 \\ & W x_2 = h^s - T x_1 \end{cases} = \begin{cases} \max\limits_{\pi} & \langle \pi, h^s - T x_1 \rangle - \langle \lambda, b_2 \rangle \\ \text{s.t.} & -\lambda + W^\top \pi \le F_2 \\ & \lambda \ge 0\,, \end{cases} \tag{4.2}$$

where we use the short notation $h^s := h(\xi^s)$.

Stability of the dual variables is achieved by considering the following regularized recourse functions, depending on a parameter $\beta > 0$,

$$\mathbb{Q}^{\beta,s}(x_1) := \begin{cases} \max\limits_{\pi,\lambda} & \langle \pi, h^s - T x_1 \rangle - \langle \lambda, b_2 \rangle - \dfrac{\beta}{2}\|\pi\|^2 \\ \text{s.t.} & -\lambda + W^\top \pi \le F_2 \\ & \lambda \ge 0\,. \end{cases} \tag{4.3}$$

By construction, the solution $\bar{\pi}^{\beta,s}(x_1)$ of (4.3) is unique. Using once again duality, we have that

$$\mathbb{Q}^{\beta,s}(x_1) = \begin{cases} \min\limits_{x_2} & \langle F_2, x_2 \rangle + \dfrac{1}{2\beta}\|h^s - Wx_2 - Tx_1\|^2 \\ \text{s.t.} & 0 \leq x_2 \leq b_2 \,, \end{cases} \tag{4.4}$$

with solutions $\bar{x}_2^{\beta,s}(x_1)$ of (4.4) satisfying the relation

$$\bar{\pi}^{\beta,s}(x_1) = \frac{h^s - Wx_2^{\beta,s}(x_1) - Tx_1}{\beta}. \tag{4.5}$$

Notice that neither the λ-components solving (4.3) nor the second-stage primal minimizers $\bar{x}_2^{\beta,s}(x_1)$ are guaranteed to be unique.

Going back to the one-level formulation, instead of the linear programming problem (3.3), we shall solve a quadratic programming problem of the form

$$\begin{cases} \min\limits_{x_1, x_2^s} & \langle F_1, x_1 \rangle + \sum\limits_{s\in\mathbb{S}} p^s \Big(\langle F_2, x_2^s \rangle + \dfrac{1}{2\beta} \Big\| h^s - Wx_2^s - Tx_1 \Big\|^2 \Big) \\ \text{s.t.} & 0 \leq x_1 \leq b_1 \\ & 0 \leq x_2^s \leq b_2 \quad \text{for } s \in \mathbb{S} \,. \end{cases} \tag{4.6}$$

Revisiting the Analytical Case

In our simple illustrative problem in Subsection 3.2.1, the probability distribution is symmetric and $F_2^- = F_2^+ = F_2$. The stabilized version of (3.4) is

$$\begin{cases} \min\limits_{x_1} & F_1 x_1 + E[Q^\beta(x_1, \xi)] \\ \text{s.t.} & x_1 \geq 0 \,, \end{cases} \quad \text{for } Q^\beta(x_1, \xi) = \begin{cases} \max\limits_{\pi} & (\xi - x_1)\pi - \dfrac{\beta}{2}\pi^2 \\ & -F_2 \leq \pi \leq F_2 \,, \end{cases} \tag{4.7}$$

which yields the multiplier

$$\pi^\beta(x_1, \xi) := \begin{cases} -F_2 & \text{if } \xi - x_1 < -\beta F_2 \\ \dfrac{\xi - x_1}{\beta} & \text{if } \xi - x_1 \in [-\beta F_2, \beta F_2] \\ F_2^+ & \text{if } \xi - x_1 > \beta F_2 \,, \end{cases} \tag{4.8}$$

to be compared with the multipliers (3.5), computed for the initial problem.

Continuing with the actions of two managers, now solving the regularized problems, suppose β_k is sufficiently large for the inequality $x_1^k - \beta_k F_2 < 0$ to hold. Even if Manager 1 still samples only negative numbers, now the set $\mathbb{S}_1$ may contain scenarios for which $\xi^s - x_1^k \in [-\beta_k F_2, 0]$, with prices possibly larger than $-F_2$. Similarly, now Manager 2 can sample $\xi^s \in [0, x_1^k + \beta_k F_2]$ for some $s \in \mathbb{S}_2$, thus also considering prices smaller than F_2 when computing the mean. It is therefore likely that the empirical expected prices will be closer to the true mean.

The meaning of this example is not to suppose that one Manager will test just positive or negative scenarios in a reality, but to illustrate that in one sample negative or positive scenarios can predominate. We aim to intuitively show that, regardless of how scenarios are distributed, Managers 1 and 2 should have results closer to the mean. Clearly, when we increase the number of scenarios in each sample we approximate the real distribution, but the fact is that, in real problems, we do not know how close we are from the real distribution or the quantity of scenarios necessary to achieve a good approximation.

The result below formalizes this assertion. Recall from (3.6) that the moments for the price signal at the solution $\bar{x}_1 = 0$ are

$$\mathbb{E}[\bar{\pi}(\xi)] = 0 \quad \text{and} \quad \mathbb{V}ar[\bar{\pi}(\xi)] = F_2^2 \,.$$

Proposition 4.1.1. *Consider the simple problem* (3.4) *with minimizer* $\bar{x}_1 = 0$*, and let* $\bar{\pi}(\xi)$ *be an optimal multiplier associated with the affine constraint in the recourse function* $\mathbb{Q}(\bar{x}_1, \xi)$*. The following holds for a solution* $\left(x_1^k, \pi^k(\xi) = \pi^{\beta_k}(x_1^k, \xi)\right)$ *of the regularized problem* (4.7)*, written with* $\beta = \beta_k$*:*

$$\lim_{\beta_k \to 0} \mathbb{E}[\pi^k(\xi)] = \mathbb{E}[\bar{\pi}(\xi)] \quad \textit{with} \quad \mathbb{V}ar[\pi^k(\xi)] \le \mathbb{V}ar[\bar{\pi}(\xi)] \,. \tag{4.9}$$

Proof. It is convenient to introduce the short notation

$$\Gamma_-^k := \mathbb{P}(\xi - x_1^k \le -\beta_k F_2)\,, \text{ and } \Gamma_+^k := \mathbb{P}(\xi - x_1^k \ge \beta_k F_2)\,,$$

noting that, by the symmetry assumption and the fact that $x_1^k \to \bar{x} = 0$,

$$\lim_{\beta_k \to 0} \Gamma_+^k = \mathbb{P}(\xi \ge 0) = \frac{1}{2} = \mathbb{P}(\xi \le 0) = \lim_{\beta_k \to 0} \Gamma_-^k \,. \tag{4.10}$$

The average of prices (4.8) is

$$\mathbb{E}[\pi^k(\xi)] = F_2(\Gamma_+^k - \Gamma_-^k) + \int_{x_1^k - \beta_k F_2}^{x_1^k + \beta_k F_2} \frac{\xi - x_1^k}{\beta_k} \mathrm{d}\xi \,.$$

To compute the limit, first bound the integral as follows:

$$-\Gamma^k F_2 \le \int_{x_1^k - \beta_k F_2}^{x_1^k + \beta_k F_2} \frac{\xi - x_1^k}{\beta_k} \mathrm{d}\xi \le \Gamma^k F_2 \,,$$

for

$$\Gamma^k := \mathbb{P}(-\beta_k F_2 \le \xi - x_1^k \le \beta_k F_2)\,.$$

Then, using that $\lim_{\beta_k \to 0} \Gamma^k = 0$, passing to the limit as $\beta_k \to 0$ in the inequalities below

$$F_2(\Gamma_+^k - \Gamma_-^k - \Gamma^k) \leq \mathbb{E}[\pi^k] \leq F_2(\Gamma_+^k - \Gamma_-^k + \Gamma^k)\,, \tag{4.11}$$

yields, together with (4.10), that $\lim_{\beta_k \to 0} \mathbb{E}[\pi^k] = 0$, as claimed.

To get the variance expression, since $\mathbb{V}ar[\pi] = \mathbb{E}[\pi^2] - \mathbb{E}[\pi]^2$, we first compute the term

$$\mathbb{E}[\pi^k(\xi)^2] = F_2^2(\Gamma_+^k - \Gamma_-^k) + \int_{x_1^k - \beta_k F_2}^{x_1^k + \beta_k F_2} \frac{(\xi - x_1^k)^2}{\beta_k^2} \mathrm{d}\xi\,,$$

and bound again the integral, as follows:

$$-\Gamma^k F_2^2 \leq \int_{x_1^k - \beta_k F_2}^{x_1^k + \beta_k F_2} \frac{(\xi - x_1^k)^2}{\beta_k^2} \mathrm{d}\xi \leq \Gamma^k F_2^2\,.$$

Adding the negative of (4.11) gives

$$F_2(\Gamma_+^k - \Gamma_-^k - \Gamma^k) \leq \mathbb{E}[\pi^k(\xi)^2] - \mathbb{E}[\pi^k] F_2(\Gamma_+^k - \Gamma_-^k + \Gamma^k)\,.$$

Since $\mathbb{V}ar[Z] = \mathbb{E}[Z^2] - \mathbb{E}[Z]^2$ for any random variable Z, this means that

$$F_2(\Gamma_+^k - \Gamma_-^k - \Gamma^k) \leq \mathbb{V}ar[\pi^k] \leq F_2(\Gamma_+^k - \Gamma_-^k + \Gamma^k)\,.$$

Then, passing to the limit as $\beta_k \to 0$, gives the desired result. □

We next examine some properties of the regularized recourse function. In particular, we show that it is piecewise quadratic.

Proposition 4.1.2 (The regularized recourse function is piecewise quadratic)**.** *Given the regularized recourse function defined in* (4.3)*, consider the function* $p^s : \mathbb{R}^{n_1} \to \mathbb{R}^{n_2}$ *defined by*

$$p^s(x_1) := F_2 - \frac{1}{\beta} W^\top (\xi^s - T x_1)\,.$$

There exists a partition of $\mathbb{R}^{n_2}$ *into polyhedral sets* $\mathbb{U}^r$*,* $r = 1, \ldots, R$*, and a piecewise quadratic function*

$$\varphi(p) := \frac{1}{\beta} \left\langle (A^r)^\top A^r p, p \right\rangle \ \textit{for } p \in \mathbb{U}^r\,,$$

with A^r *injective, such that* $\mathbb{Q}^{\beta,s}(\cdot, \xi^s) = \Big(\varphi \circ p^s\Big)(\cdot)$.

Proof. We shall make use of the stability result [Ban+83, Theorem 5.5.2]. To this aim, consider $\mathbb{Q}^{\beta,s}$ in its formulation 4.4 as a parametric optimization problem

$$
\begin{aligned}
\mathbb{Q}^{\beta,s}(x_1) &= \min_{0\le x_2\le b_2} \langle F_2, x_2\rangle + \frac{1}{2\beta}\|h^s - Wx_2 - Tx_1\|^2 \\
&= \min_{0\le x_2\le b_2} \left\langle F_2 - \frac{1}{\beta}W^\top(\xi^s - Tx_1), x_2 \right\rangle + \frac{1}{2\beta}\|Wx_2\|^2 \\
&= \min_{0\le x_2\le b_2} \langle p^s, x_2\rangle + \frac{1}{2\beta}\|Wx_2\|^2 .
\end{aligned}
$$

Let $\bar{x}_2^s$ be a solution, so that

$$
\mathbb{Q}^{\beta,s}(x_1) = \langle p^s, \bar{x}_2^s\rangle + \frac{1}{2\beta}\|W\bar{x}_2^s\|^2 .
$$

Then, by [Ban+83, Theorem 5.5.2], the function $\varphi(p) := \langle \bar{x}_2^s, p\rangle - \frac{1}{2\beta}\|W\bar{x}_2^s\|^2$ is a *piecewise quadratic* function, that is, we can divide the domain of φ in polyhedral sets $\mathbb{U}^r$, $r \in \{1, 2, \ldots, R\}$, such that φ is quadratic on each one of them. In addition, for each polyhedral set $\mathbb{U}^r$, there exists a matrix E^r such that:

$$
\varphi(p) = \frac{1}{\beta}\|WE^r p\|^2 + \langle E^r p, p\rangle = \left\langle \left(\frac{1}{\beta}W^\top W + I\right) E^r p, p \right\rangle .
$$

It remains to exhibit an injective matrix A^r such that

$$
\left(\frac{1}{\beta}W^\top W + I\right) E^r p = \frac{1}{\beta}(A^r)^\top A^r p ,
$$

and for this we show that the left hand side can not be the zero vector. To prove the claim, we examine in [Ban+83, Theorem 5.5.2] the structure of $E^r p$ for any $p \in \mathbb{R}^{n_2}$. More specifically, the polygonal sets $\mathbb{U}^r$ are defined in terms of the regions of positivity of the Lagrange multipliers, so that in the following matrix:

$$
D^r := \begin{bmatrix} \frac{1}{\beta}W^\top W & I & I & 0 \\ I & 0 & 0 & I \\ I_1^r & I_2^r & J_1^r & J_2^r \end{bmatrix} ,
$$

the matrices I_1^r, I_2^r, J_1^r and J_2^r are diagonal, with components depending on such regions. As a result, $E^r p$ corresponds to the first n_2 coordinates of the vector $D^r(p, 0, 0)$. Let P_{n_2} be the projection on $\mathbb{R}^{n_2}$. We have that

$$
E^r p := P_{n_2} D^r(p, 0, 0) = \frac{1}{\beta}W^\top W p,
$$

and

$$
\left(\frac{1}{\beta}W^\top W + I\right) E^r p = \left(\frac{1}{\beta}W^\top W + I\right) \frac{1}{\beta}W^\top W p .
$$

As the left-hand side cannot be zero for non-zero vector, as (($W^\top W$ would then have negative eigenvalues), the function

$$\varphi(p) = \left\langle \left(\frac{1}{\beta} W^\top W + I \right) E^r p, p \right\rangle$$

has no constant directions, and it can be written in terms of an injective matrix A^r, as stated. $\square$

4.2 Asymptotic Properties of the Dual Regularization

In addition to smoothing the recourse function, an important conclusion of the results described below in the context of price signals is that, when $\beta \to 0$, the regularized dual variable (4.5) converges to the Lagrange Multiplier (for the equality constraint) associated to a minimizer in (4.2) that has *minimal norm*.

To cast (3.3) and (4.6) in the more general setting considered in [LSS19b], suppose the scenario set is $\mathbb{S} := \{1, \dots, S\}$ and recall that n_1 and n_2 are the respective dimensions of x_1, x_2^s defined in (3.2). We define the vectors

$$x := (x_1, x_2^1, \dots, x_2^{\mathbb{S}}) \in \mathbb{R}^n \,, \text{ where } n := n_1 + n_2 S \,,$$

$$g := (F_1, p^1 F_2, \dots, p^{\mathbb{S}} F_2) \in \mathbb{R}^n \,,$$

$$a := (h^1, \dots, h^{\mathbb{S}}) \in \mathbb{R}^{m\mathbb{S}} \,, b := (b_1, b_2, \dots, b_2) \in \mathbb{R}^n \,,$$

and the $mS \times n$ matrix

$$A := \begin{bmatrix} T & W & 0 & \dots & 0 \\ T & 0 & W & \ddots & \vdots \\ T & \vdots & \ddots & \ddots & 0 \\ T & 0 & \dots & 0 & W \end{bmatrix} , \tag{4.12}$$

where m is the dimension of h^s.

With this notation, problem (3.3) becomes

$$\begin{cases} \min & \langle g, x \rangle \\ \text{s.t.} & Ax = a \\ & 0 \le x \le b \,. \end{cases} \tag{4.13}$$

The multiplier associated with the constraint $Ax = a$ in (4.13) is denoted π, while the box-constraint multipliers are denoted by μ_0 and μ_b, respectively for the lower and upper bounds.

The penalization problem (4.6) writes down as follows:

$$\begin{cases} \min\limits_{x} & \langle g, x\rangle + \dfrac{1}{2\beta}\|Ax - a\|^2 \\ \text{s.t.} & 0 \le x \le b\,. \end{cases} \tag{4.14}$$

The theory in [LSS19b] considers more general (than (4.13)) quadratic objective functions, including possibly nonconvex. On the other hand, [LSS19b] deals with equality and nonnegativity (thus, lower bound) constraints, while here we also have upper bounds. Nevertheless, as shown in Theorem 4.2.7 below, the results from [LSS19b] can be adapted to the setting of (4.13).

4.2.1 A Variational Analysis Perspective

We mostly follow the notation of [RW98]. Points in $\mathbb{R}^n$ are considered as column vectors. The Euclidean inner product and norm are denoted by $\langle\cdot,\cdot\rangle$ and $\|\cdot\|$, respectively. The indicator function of a set S is denoted by $\delta_S(\cdot)$, i.e., this function is 0 for points in S and is $+\infty$ otherwise. If S is convex, then $N_S(x)$ stands for the normal cone of S at the point x. The unit ball centered at 0 is $\mathbb{B}$ and the identity matrix is I; in both cases the dimension is always clear from the context. For a proper convex function f, its subdifferential at x is denoted by $\partial f(x)$ while its horizon subdifferential at the point x is the normal cone of the function's domain, i.e., $\partial^\infty f(x) = N_{\operatorname{dom} f}(x)$; see [RW98, Proposition 8.12].

We are particularly interested in the Lagrange multiplier of the affine equality constraint of the following (feasible) optimization problem:

$$\begin{cases} \min & f(x) \\ \text{s.t.} & x \in X \\ & Ax - b = 0\,, \end{cases} \quad \text{where} \quad \begin{array}{l} f(x) \text{ is finite valued, convex, and } C^1 \\ X \text{ is a closed convex set, and} \\ b \in \{y : y = Ax,\ x \in X\}\,. \end{array} \tag{4.15}$$

While this is not essential for some of the subsequent considerations, we shall assume that X is defined by smooth convex inequalities, as is certainly the case in applications we have in mind. It is then well known that uniqueness of Lagrange multipliers associated to a solution $\bar{x}$ of problem (4.15) is implied by the linear independence of gradients of the constraints active at $\bar{x}$. This is in turn equivalent to the so-called strict Mangasarian-Fromovitz (MF) condition (see, e.g., [Sol10], [IS14, Sections 1.1, 1.2.4]). We emphasize that in (4.15) either of these assumptions implies that the matrix A is of full rank, a condition that does not hold in practice for many important applications. The less stringent MF constraint qualification (MFCQ), equivalent to having a nonempty compact set of Lagrange multipliers, also subsumes that A has full rank.

Thus, if A is not of full rank, the multipliers associated to the equality constraint in (4.15) are necessarily not unique. In fact, since MFCQ is violated in this case, the multiplier set is unbounded. This leads us to focus on devising a mechanism to identify/compute the multiplier that has the minimal norm. The idea is to consider a sequence of problems that penalize the equality constraint in (4.15), depending on a parameter $\beta > 0$. Given a (primal) solution to the penalized problem, we then construct an explicit multiplier estimate, which we denote by π^β. Specifically, we solve

$$\begin{cases} \min\limits_x & f(x) + \dfrac{1}{2\beta}\|Ax - b\|^2 \\ \text{s.t.} & x \in X\,, \end{cases}$$

for $\beta > 0$ to obtain x^β, and define as multiplier proxy

$$\pi^\beta := \frac{Ax^\beta - b}{\beta}.$$

For the case when (4.15) is a linear or quadratic program, including the two-stage stochastic linear programming problems considered in Section 2.1, we then exhibit some natural conditions which ensure that as $\beta \to 0$, the sequence of the constructed multiplier estimates π^β tends to the specific multiplier $\hat{\pi}$ of minimal norm. The precise details will be given in Section 4.2.2.

Approximating Lagrange multipliers in the setting of quadratic penalty methods is certainly not a new idea; see, e.g., [NW06, Chapter 17.1]. However, in the literature convergence results are established assuming linear independence of active gradients (as well as subsequential convergence of the primal sequence x^β), in which case the optimal multiplier is unique; see [NW06, Theorem 17.2] and Theorem 4.2.4 below. In Section 4.2.2, we give conditions under which x^β converges, and show convergence of π^β (to minimal-norm multiplier) without assuming the linear independence condition, thus covering a much more general case.

In this section, we give a different motivation and insight for the multiplier proxies by specializing some results of Variational Analysis [RW98] to our setting. We start with a fixed $\beta \geq 0$ and relate the estimates π^β with a particular instance of the *generalized Lagrange multiplier* rule [RW98, Example 10.8, p.429]. More precisely, given a scalar $\beta \geq 0$, consider the following penalties:

$$\mathbb{R}^m \ni v \hookleftarrow \theta^\beta(v) := \sup_{y \in \mathbb{R}^m} \left\{ \langle v, y \rangle - \frac{1}{2}\beta\|y\|^2 \right\} = \begin{cases} \dfrac{1}{2\beta}\|v\|^2 & \text{if } \beta > 0 \\ \delta_{\{0\}}(v) & \text{if } \beta = 0\,. \end{cases} \tag{4.16}$$

These (lsc, proper, convex) functions are a particular case of the piecewise linear-quadratic penalties in [RW98, Example 11.18, p. 497] (therein, θ^β corresponds to $\theta_{Y,B}$, written for $Y = \mathbb{R}^m$ and the, possibly zero, matrix

$B = \beta I$). The respective subdifferentials are:

$$\begin{array}{l} \text{if } \beta > 0, \quad \text{for all } v \in \mathbb{R}^m = \operatorname{dom} \theta^\beta\,, \\ \partial\theta^\beta(v) = \left\{\frac{1}{\beta}v\right\} \text{ and } \partial^\infty\theta^\beta(v) = \{0\}, \end{array} \tag{4.17}$$

while

$$\text{if } \beta = 0, \quad \text{for } v = 0 = \operatorname{dom}\theta^0\,, \quad \partial\theta^0(v) = \mathbb{R}^m \text{ and } \partial^\infty\theta^0(v) = \mathbb{R}^m\,. \tag{4.18}$$

The connection between penalties and dual variables (multipliers) is made clear when considering, for perturbation parameters $u \in \mathbb{R}^m$, the (unconstrained) parametric minimization problems

$$\min_{x\in\mathbb{R}^n} f^\beta(x,u) := f(x) + \delta_X(x) + \theta^\beta(Ax - b + u)\,, \tag{4.19}$$

noting that writing (4.19) with $\beta = 0$ and $u = 0$ yields our original problem (4.15).

When $\beta > 0$, some x^β is optimal in (4.19) if and only if

$$x^\beta \in X,\ \mu^\beta \in N_X(x^\beta), \quad \nabla f(x^\beta) + \mu^\beta + A^\top\pi^\beta = 0\,, \tag{4.20}$$

where, for $\bar{u} \in \mathbb{R}^m$ given, the unique *extended Lagrange multiplier* in [RW98] is

$$\pi^\beta := \frac{Ax^\beta - b + \bar{u}}{\beta}\,.$$

To consider the case when $\beta = 0$, recall that, in its dual formulation (see, e.g., [Sol10], [IS14, Sections 1.1, 1.2.4]), the MFCQ at a feasible point $\bar{x}$ of our (unperturbed) problem (4.15) means that

$$0 = A^\top\pi + \mu, \quad \mu \in N_X(\bar{x}) \quad \Rightarrow \quad \pi = 0 \in \mathbb{R}^m \text{ and } \mu = 0 \in \mathbb{R}^n\,. \tag{4.21}$$

If $\bar{x}$ satisfies MFCQ (4.21), there exists a classical Lagrange multiplier $\bar{\pi} \in \mathbb{R}^m$, not necessarily unique, satisfying (4.20) written with $\beta = 0$ and $(x^\beta, \pi^\beta, \mu^\beta) = (\bar{x}, \bar{\pi}, \bar{\mu})$. Condition (4.20) is also sufficient for $\bar{x}$ to be optimal for (4.19) written with $\beta = 0$ and $\bar{u} := b - A\bar{x}$.

The proposition below follows from the *parametric version of Fermat rule* in [RW98, Example 10.12], analyzing (4.19) from a Variational Analysis perspective, condensing the key ingredients relating extended Lagrange multipliers to the marginal rate of change of the optimal value in (4.15), when considered as a function of the right-hand side perturbation of the affine constraint.

Proposition 4.2.1 (Extended Lagrange multipliers)**.** *Associated to* (4.15)*, consider the parametric optimization problems* (4.19) *with penalties* (4.16)*, where $u = \bar{u} \in \mathbb{R}^m$ and $\beta \geq 0$ are fixed. Let the corresponding optimal value* *and solution set be given by*

$$p^\beta(u) := \inf_{x\in\mathbb{R}^n} f^\beta(x,u) \quad \textit{and} \quad P^\beta(u) := \arg\min_{x\in\mathbb{R}^n} f^\beta(x,u)\,.$$

The following holds.

(i) *When $\beta > 0$, the function p^β is convex, strictly differentiable at any point $\bar u \in \operatorname{dom} p^\beta = \mathbb{R}^m$, with gradient $\nabla p^\beta(\bar u) = \pi^\beta$.*

(ii) *When $\beta = 0$, the function p^0 is convex, strictly continuous at the point $\bar u = A\bar x - b = \operatorname{dom} p^0$, with subdifferential*

$$\partial p^0(\bar u) = \{\bar\pi \in \mathbb{R}^m : \text{ (4.20) } \textit{holds written with } (x^\beta,\pi^\beta,\mu^\beta) = (\bar x,\bar\pi,\bar\mu)\}\,. \tag{4.22}$$

Proof. The perturbed function $f^\beta(x,u)$ is convex in (x,u) and the finite-valued function f has full domain. In this situation, by [RW98, Example 10.8],

$$\partial f^\beta(\bar x,\bar u) = \nabla f(\bar x) + N_X(\bar x) + \partial\theta^\beta(A\bar x - \bar u), \quad \partial^\infty f^\beta(\bar x,\bar u) = \partial^\infty\theta^\beta(A\bar x - \bar u)\,.$$

Since f^β is convex (therefore regular) with our definitions, the Y-sets in [RW98, Theorem 10.13] satisfy the relations

$$Y(\bar u) = \{\pi : (0,\pi) \in \partial f^\beta(x^\beta,\bar u)\} \quad \text{and} \quad Y^\infty(\bar u) = \{\pi : (0,\pi) \in \partial^\infty f^\beta(x^\beta,\bar u)\}$$

for any $x^\beta \in P^\beta(\bar u)$ and $\bar u \in \operatorname{dom} p^\beta = \operatorname{dom}\theta^\beta$. Together with (4.17) and (4.18), this gives $\partial^\infty p^\beta(\bar u) = Y^\infty(\bar u) = \partial^\infty\theta^\beta(Ax^\beta - \bar u)$ and, as claimed,

$$\partial p^\beta(\bar u) = Y(\bar u) = \{\pi^\beta : (x^\beta,\mu^\beta,\pi^\beta) \text{ satisfies (4.20)}\}\,.$$

□

The above characterization, obtained from the penalty scheme as extended Lagrange multiplier, motivates from the Variational Analysis point of view the choice of the multiplier estimates. In Section 4.2.2, we shall show under which conditions such estimates converge to the minimal-norm multipliers. Among other things, we shall need for this the following result, which establishes boundedness of the set of Lagrange multipliers associated to some part of the constraints of the problem, while allowing the other multipliers to be unbounded. Apparently, this result is new.

Recall that the MFCQ condition (4.21) is equivalent to the set of multipliers being nonempty and bounded. Consider the following condition at $\bar x$ feasible in (4.15):

$$0 = A^\top\pi + \mu, \quad \mu \in N_X(\bar x) \quad \Rightarrow \quad \mu = 0. \tag{4.23}$$

Clearly, (4.23) is a weaker condition than (4.21). In particular, as we show in Theorem 4.2.2 below, (4.23) implies boundedness only for the μ-part of the multipliers, while the π-part can be unbounded. The condition in question can be interpreted as a "partial" MFCQ condition. However, note that (4.23) is *not* a constraint qualification, i.e., it does not imply (by itself) that for a solution $\bar{x}$ of problem (4.15) the multiplier set is nonempty. An alternative, equivalent, formulation of condition (4.23) is

$$\operatorname{Im} A^\top \cap N_X(\bar{x}) = \{0\}. \tag{4.24}$$

Theorem 4.2.2 (On boundedness of multipliers)**.** *Let $\bar{x}$ be any feasible point in* (4.15)*. Then the following statements are equivalent:*

(i) *Condition* (4.23) *holds at $\bar{x}$.*

(ii) *For any $\bar{g} \in \mathbb{R}^n$, the set*

$$S_{\bar{g}} := \{\bar{\mu} \in N_X(\bar{x}) : \exists\, \bar{\pi} \in \mathbb{R}^m \text{ s.t. } \bar{g} + A^\top \bar{\pi} + \bar{\mu} = 0\}$$

is bounded.

Proof. We shall show the equivalent assertion

$$\exists\, \tilde{\mu} \in \operatorname{Im} A^\top \cap N_X(\bar{x}),\ \tilde{\mu} \neq 0 \iff \exists\, \bar{g} \in \mathbb{R}^n \text{ such that } S_{\bar{g}} \text{ is unbounded.}$$

Assume first that for some $\bar{g}$ the set $S_{\bar{g}}$ is unbounded, i.e., there exists a sequence $\{(\pi^k, \mu^k)\}$ such that

$$\bar{g} + A^\top \pi^k + \mu^k = 0, \quad \mu^k \in N_X(\bar{x})\,, \tag{4.25}$$

with $\|\mu^k\| \to +\infty$. As $N_X(\bar{x})$ is a closed cone, we can assume, passing onto a subsequence if necessary, that

$$\mu^k / \|\mu^k\| \to \bar{\mu} \in N_X(\bar{x}),\ \bar{\mu} \neq 0\,.$$

Denote $u^k = -A^\top \pi^k / \|\mu^k\| \in \operatorname{Im} A^\top$. Dividing the equality in (4.25) by $\|\mu^k\|$ and passing onto the limit, it follows that

$$u^k = (\bar{g} + \mu^k)/\|\mu^k\| \to \bar{\mu}.$$

As $u^k \in \operatorname{Im} A^\top$, $u^k \to \bar{\mu}$, and $\operatorname{Im} A^\top$ is closed, we conclude that $\bar{\mu} \in \operatorname{Im} A^\top$. As it also holds that $\bar{\mu} \in N_X(\bar{x})$ and $\bar{\mu} \neq 0$, this contradicts (4.23).

Suppose now that there exists $0 \neq \tilde{\mu} \in N_X(\bar{x})$ such that $A^\top \tilde{\pi} + \tilde{\mu} = 0$ for some $\tilde{\pi}$. If for some $\bar{g}$ there is a pair $(\bar{\pi}, \bar{\mu})$ satisfying

$$\bar{g} + A^\top \bar{\pi} + \bar{\mu} = 0, \quad \bar{\mu} \in N_X(\bar{x}),$$

then, for any $t > 0$, it holds that $\bar{\mu} + t\tilde{\mu} \in N_X(\bar{x}) + N_X(\bar{x}) = N_X(\bar{x})$, since the cone in question is convex. Hence, for any $t > 0$,

$$\bar{g} + A^\top(\bar{\pi} + t\tilde{\pi}) + (\bar{\mu} + t\tilde{\mu}) = 0\,, \quad \bar{\mu} + t\tilde{\mu} \in N_X(\bar{x}).$$

Since $\tilde{\mu} \neq 0$, as $t \to +\infty$ we have $\|\bar{\mu}+t\tilde{\mu}\| \to +\infty$ and the set $S_{\bar{g}}$ is unbounded. □

We emphasize that, being weaker than MFCQ, condition (4.23) is certainly not restrictive (assuming that the existence of Lagrange multipliers is given or follows from some other considerations).

4.2.2 A Nonlinear Programming Computational Perspective

Consider now the following (linear or) *quadratic programming* problem:

$$\left\{\begin{array}{ll} \min & f(x) \\ \text{s.t.} & x \in X \\ & Ax - b = 0\,, \end{array}\right. \quad \text{where} \quad \begin{array}{l} f(x) := \langle g, x\rangle + \dfrac{1}{2}\langle x, Hx\rangle \\ X := \{x \in \mathbb{R}^n : x \geq 0\}\,, \text{ and} \\ b \in \{y : y = Ax,\ x \in X\}\,, \end{array} \tag{4.26}$$

where $g \in \mathbb{R}^n$ and H is an $n \times n$ matrix ($H = 0$ corresponding to linear programming). We note that in our developments below, H is not necessarily positive semidefinite, although it also might be.

When H is positive semidefinite, the convex problem (4.26) is a particular instance of (4.15), and we can use the constructs in Section 4.2.1 for some motivations. In that case, fixing $\bar{u} = 0$, for $\beta > 0$ from (4.16) and (4.19) we have

$$f^\beta(x, 0) = f(x) + \delta_X(x) + \frac{1}{2\beta}\|Ax - b\|^2\,,$$

and Proposition 4.2.1 characterizes the extended Lagrange multiplier as follows:

$$\pi^\beta = \frac{Ax^\beta - b}{\beta}, \qquad \text{for } x^\beta \in P^\beta(0)\,.$$

As a result, finding $x^\beta \in P^\beta(0)$ is equivalent to finding x^β, a solution to the following (partial) exterior penalization of problem (4.26):

$$\left\{\begin{array}{ll} \min & f(x) + \dfrac{1}{2\beta}\|Ax - b\|^2 \\ \text{s.t.} & x \in X\,. \end{array}\right. \tag{4.27}$$

Penalty methods (see, e.g., [FM68]) solve subproblems (4.27) for a sequence of decreasing penalty parameters $0 < \beta_{k+1} < \beta_k$, tending to zero. We want to study how the multiplier estimates for the equality constraints in (4.26) behave along the sequence of solving the penalized subproblems (4.27). We shall show that, under reasonable assumptions, the generalized multipliers π^β converge to the minimal-norm multiplier $\hat{\pi}$; see (4.37) below for a formal definition.

We start with some standard facts on (primal) convergence of penalty methods [FM68] that do not depend on the setting of (4.26), and can also use other forms of exterior penalties (not necessarily quadratic). But we shall keep this setting for the sake of not introducing extra notation. Define

$$F_k(x) := f(x) + \frac{1}{2\beta_k}\|Ax - b\|^2,$$

the objective function in (4.27).

Theorem 4.2.3 (Primal convergence of generic penalty methods)**.** *Let f be convex and x^k be a (global) solution of* (4.27) *for $\beta = \beta_k$ for each k, with $0 < \beta_{k+1} < \beta_k$. Then*

$$F_{k+1}(x^{k+1}) \geq F_k(x^k), \quad \|Ax^{k+1} - b\| \leq \|Ax^k - b\|, \quad f(x^{k+1}) \geq f(x^k). \tag{4.28}$$

If, in addition, $\beta_k \to 0$ as $k \to \infty$ and the optimal value of problem (4.26) *is finite, then every accumulation point of $\{x^k\}$ is a (global) solution of* (4.26)*.*

Note that this result refers to *global* solutions of subproblems. This is standard, and also not an issue when the problem (4.26) is convex. Another observation is that in the case of a quadratic program as ours, if f is bounded below on the feasible region (i.e., the optimal value is finite) then problem (4.26) has a solution, by the Frank–Wolfe Theorem [FW56].

However, it is important to emphasize that the general convergence result in Theorem 4.2.3 asserts optimality of accumulation points but does not say anything about their *existence*. It can thus be "vacuous" if the sequence is unbounded. Our first task will be to prove when the generated sequence $\{x^k\}$ is bounded. But before proceeding, we shall mention the following classical result on convergence of the multiplier estimates obtained from the quadratic penalty method. Let x^k be a solution of (4.27) for $\beta = \beta_k$. Define

$$\pi^k := \frac{1}{\beta_k}(Ax^k - b). \tag{4.29}$$

The assertion below is standard; see, e.g., [NW06, Theorem 17.2]. Like Theorem 4.2.3 above, it does not depend on the setting of problem (4.26) and can be easily extended to the case of general nonlinear objective function f and general nonlinear constraints, including inequality constraints. As this is not essential for our developments, we state the result for equality constraints only.

Theorem 4.2.4 (Dual convergence of the quadratic penalty method)**.** *In* (4.26)*, let $X = \mathbb{R}^n$. Let $\bar{x}$ be any accumulation point of $\{x^k\}$, where x^k is a* *solution of* (4.27) *for* $\beta = \beta_k$ *for each* k, $\{x^{k_j}\} \to \bar{x}$ *as* $j \to \infty$. *Let the linear independence constraints qualification hold at* $\bar{x}$ *(in the setting of* (4.26) *this means that* A *has full rank).*

Then $\bar{x}$ *is a stationary point of* (4.26) *and the subsequence* $\{\pi^{k_j}\}$ *defined by* (4.29) *converges to the unique Lagrange multiplier* $\bar{\pi}$ *associated to* $\bar{x}$.

Note, however, that Theorem 4.2.4 again implicitly assumes boundedness of $\{x^k\}$ (as it refers to its accumulation points), and requires the linear independence constraints qualification for convergence of the dual sequence. The latter, in particular, is not assumed in our setting.

For establishing boundedness of the primal sequence, we shall need the following conditions. Recall that the standard critical cone of (4.26) at a given stationary point $\bar{x}$ is defined by

$$K(\bar{x}) := \operatorname{Ker} A \cap \{d \in \mathbb{R}^n : \langle H\bar{x} + g, d\rangle \leq 0,\ d_i \geq 0 \text{ for } i \text{ s.t. } \bar{x}_i = 0\}\,. \tag{4.30}$$

In the case at consideration, the Hessian of the Lagrangian (for any point $(\bar{x}, \bar{\pi}, \bar{\mu})$) is the matrix H. Thus, the usual second-order sufficient optimality condition for $\bar{x}$ states that

$$\langle Hd, d\rangle > 0 \quad \text{for all } d \in K(\bar{x}) \setminus \{0\}. \tag{4.31}$$

When H is positive semidefinite, the solution set of problem (4.26) is convex. Since (4.31) implies that $\bar{x}$ is a strict (thus isolated) minimizer, the condition means that in the convex case the primal solution must be unique. In particular, when f is linear, i.e., $H = 0$, condition (4.31) holds if and only if $K(\bar{x}) = \{0\}$. It can be further seen that this means that $\langle g, d\rangle > 0$ for all feasible directions d at $\bar{x}$. This, in turn, is equivalent to saying that $\bar{x}$ is the unique solution of the linear program (4.26). Thus, for linear programming, including the two-stage stochastic linear programming setting in Section 2.1, the assumption (4.31) amounts to stating that the primal solution of the problem is unique.

Note also that since $K(\bar{x}) \subset \operatorname{Ker} A$, the following is also a second-order sufficient optimality condition (as it implies (4.31)):

$$\langle Hd, d\rangle > 0 \quad \text{for all } d \in \operatorname{Ker} A \setminus \{0\}. \tag{4.32}$$

However, unlike (4.31), condition (4.32) is an assumption on H and A which does not depend on $\bar{x}$. Note that (4.32) does not require H to be positive semidefinite, and thus the objective function f in (4.26) can be non-convex.

Theorem 4.2.5 (Conditions for primal convergence)**.** *Suppose that one of the following two items holds:*

1. *Condition* (4.32) *is satisfied.*

2. *The matrix H is positive semidefinite and* (4.31) *holds for the solution $\bar{x}$ of* (4.26) *(if $H = 0$, this amounts to* (4.26) *having a unique primal solution).*

Then for any *sequence of parameters $\beta_k \to 0$ (even not necessarily monotone), any sequence $\{x^k\}$ generated by the penalty scheme* (4.27) *is bounded.*

If also $\beta_{k+1} < \beta_k$ for all k, then each of the accumulation points of $\{x^k\}$ is a solution of (4.26)*. In particular, in the second case above, the whole sequence converges to the unique solution $\bar{x}$.*

Proof. We reason by contradiction: taking a subsequence if necessary, suppose that $\|x^k\| \to \infty$.

Define $z^k = x^k/\|x^k\|$. Again passing onto a subsequence if necessary, we can assume that $z^k \to z$, $z \neq 0$.

By the KKT optimality conditions for the subproblems (4.27), it holds that

$$Hx^k + g + \frac{1}{\beta_k} A^\top (Ax^k - b) - \mu^k = 0, \quad x^k \geq 0,\ \mu^k \geq 0,\ \langle \mu^k, x^k \rangle = 0. \tag{4.33}$$

Note that $\langle \mu^k, z^k \rangle = 0$. Thus, multiplying the first relation above by z^k yields

$$\left\langle Hx^k + g, z^k \right\rangle = \frac{1}{\beta_k} \left\langle A^\top (b - Ax^k), z^k \right\rangle .$$

Next, multiplying both sides of the latter equality by $\beta_k/\|x^k\|$, we conclude that

$$\beta_k \left\langle Hz^k, z^k \right\rangle + \frac{\beta_k}{\|x^k\|} \left\langle g, z^k \right\rangle = \frac{1}{\|x^k\|} \left\langle A^\top b, z^k \right\rangle - \|Az^k\|^2 .$$

As $\{z^k\}$ is bounded while $\|x^k\| \to \infty$ and $\beta_k \to 0$, passing onto the limit as $k \to \infty$ yields that $0 = \|Az\|^2$, i.e., $z \in \operatorname{Ker} A$.

Let $\tilde{x}$ be any feasible point in (4.26). Since $\tilde{x} \in X$, $A\tilde{x} - b = 0$, and x^k is a solution of (4.27), it holds that

$$f(\tilde{x}) \geq f(x^k) + \frac{1}{2\beta_k} \|Ax^k - b\|^2 \geq f(x^k). \tag{4.34}$$

Dividing this inequality by $\|x^k\|^2$, we obtain that

$$\frac{f(\tilde{x})}{\|x^k\|^2} \geq \frac{f(x^k)}{\|x^k\|^2} = \frac{1}{2} \langle Hz^k, z^k \rangle + \frac{1}{\|x^k\|} \langle g, z^k \rangle .$$

Passing onto the limit as $k \to \infty$ gives

$$0 \geq \frac{1}{2} \langle Hz, z \rangle. \tag{4.35}$$

Since $0 \neq z \in \operatorname{Ker} A$, this immediately gives a contradiction if the condition (4.32) holds.

Suppose now H is positive semidefinite and (4.31) holds for the solution $\bar{x}$ of (4.26), which is unique in this case. Since $\bar{x}$ is in particular feasible, from (4.34) written with $\tilde{x} = \bar{x}$, using also the convexity of f, we conclude that

$$f(\bar{x}) \geq f(x^k) \geq f(\bar{x}) + \left\langle \nabla f(\bar{x}), x^k - \bar{x} \right\rangle,$$

and hence,

$$0 \geq \left\langle \nabla f(\bar{x}), x^k - \bar{x} \right\rangle.$$

Dividing both sides above by $\|x^k\|$ and passing onto the limit, we conclude that $\langle \nabla f(\bar{x}), z \rangle \leq 0$. Since $z \geq 0$ is obvious (because $x^k \geq 0$), and recalling that $z \in \operatorname{Ker} A$, we obtain that $0 \neq z \in K(\bar{x})$; see (4.30). Now (4.35) again gives a contradiction with (4.31). We conclude that $\{x^k\}$ is bounded. The other assertions follow from the general results about penalty methods in Theorem 4.2.3 (and other considerations stated above). □

Having established when there is primal convergence of solutions of the penalized subproblems (4.27), we now analyze the asymptotic behavior of the dual sequence $\{\pi^k\}$ defined by (4.29).

Recall that for a solution $\bar{x}$ of problem (4.26) the set of associated Lagrange multipliers (π, μ) is characterized by the following system:

$$H\bar{x} + g + A^\top \pi - \mu = 0, \quad \bar{x} \geq 0,\ \mu \geq 0,\ \langle \mu, \bar{x} \rangle = 0. \tag{4.36}$$

To exhibit the specific dual behavior (dual limit) of the sequence $\{\pi^k\}$, denote by $\hat{\pi} = \hat{\pi}(\bar{x}, \bar{\mu})$ the minimal-norm element which solves (4.36) for the given $\bar{x}$ and $\bar{\mu}$, i.e., the (unique) solution of

$$\min \frac{1}{2}\|\pi\|^2 \quad \text{s.t.} \quad H\bar{x} + g + A^\top \pi - \bar{\mu} = 0. \tag{4.37}$$

We have the following.

Theorem 4.2.6 (Convergence of the multipliers estimates)**.** *Let $\beta_k \to 0$ and $\beta_{k+1} < \beta_k$ for all k. Let the assumptions of Theorem 4.2.5 hold. Let $\bar{x}$ be any accumulation point of the sequence $\{x^k\}$ (which is bounded by Theorem 4.2.5), $x^{k_j} \to \bar{x}$ as $j \to \infty$. Let condition* (4.23) *hold at $\bar{x}$.*

Then the sequence $\{\mu^{k_j}\}$ is bounded. Moreover, for any of its accumulation points $\bar{\mu}$, the subsequence $\{\pi^{k_j}\}$ defined by (4.29) *converges to $\hat{\pi}$, the minimal-norm solution of* (4.37)*. The point $(\bar{x}, \hat{\pi}, \bar{\mu})$ is a primal-dual solution of* (4.26)*.*

Proof. Under the assumptions of Theorem 4.2.5, it follows that $\{x^k\}$ is bounded. Recalling the subproblem KKT conditions (4.33) and using the definition (4.29) of π^k, we have that

$$Hx^k + g + A^\top \pi^k - \mu^k = 0, \quad x^k \geq 0,\ \mu^k \geq 0,\ \langle \mu^k, x^k \rangle = 0. \tag{4.38}$$

Let $\{x^{k_j}\} \to \bar{x}$ as $j \to \infty$. We first prove that the sequence $\{\mu^{k_j}\}$ is bounded. Similarly to the first part of the proof of Theorem 4.2.2, suppose by contradiction that (4.38) holds with $\|\mu^{k_j}\| \to +\infty$ (possibly passing onto a subsequence). We can assume, passing onto a further subsequence if necessary, that

$$\mu^{k_j}/\|\mu^{k_j}\| \to \bar{\mu} \geq 0,\ \bar{\mu} \neq 0\,. \tag{4.39}$$

Denote $u^{k_j} = A^\top \pi^{k_j}/\|\mu^{k_j}\| \in \operatorname{Im} A^\top$. Dividing the equality in (4.38) by $\|\mu^{k_j}\|$ and passing onto the limit as $j \to \infty$, it follows that

$$u^{k_j} = (\mu^{k_j} - Hx^{k_j} - g)/\|\mu^{k_j}\| \to \bar{\mu},$$

where boundedness of $\{x^{k_j}\}$ was taken into account. As $u^{k_j} \in \operatorname{Im} A^\top$, $u^{k_j} \to \bar{\mu}$, and $\operatorname{Im} A^\top$ is closed, we conclude that $\bar{\mu} \in \operatorname{Im} A^\top$. Obviously $\bar{x} \geq 0$ and, dividing the last two relations in (4.38) by $\|\mu^{k_j}\|$ and passing onto the limit, $\bar{\mu} \geq 0$, $\langle \bar{\mu}, \bar{x} \rangle = 0$. This means that $-\bar{\mu} \in N_X(\bar{x})$, where $X = \mathbb{R}^n_+$. As $\bar{\mu} \neq 0$ and $-\bar{\mu} \in \operatorname{Im} A^\top$, we obtain a contradiction with (4.23).

Once $\{\mu^{k_j}\}$ is bounded, the first equality in (4.38) implies that $\{A^\top \pi^{k_j}\}$ is bounded as well. Passing onto a further subsequence if necessary, we can assume that $\{x^{k_j}\} \to \bar{x}$, $\{\mu^{k_j}\} \to \bar{\mu}$, $\{A^\top \pi^{k_j}\} \to a$ as $j \to \infty$.

Taking any point $\tilde{x}$ such that $A\tilde{x} - b = 0$, we observe that

$$\pi^k = \frac{1}{\beta_k}(Ax^k - b) = \frac{1}{\beta_k}A(x^k - \tilde{x}) \in \operatorname{Im} A.$$

Thus, $A^\top \pi^k \in \operatorname{Im} A^\top A$. Because this subspace is closed, we have $a \in \operatorname{Im} A^\top A$, i.e., $a = A^\top \bar{\pi}$ for some $\bar{\pi} \in \operatorname{Im} A$.

Passing onto the limit in (4.38) as $j \to \infty$, we then have that

$$H\bar{x} + g + A^\top \bar{\pi} - \bar{\mu} = 0,\ \bar{\pi} \in \operatorname{Im} A, \quad \bar{x} \geq 0,\ \bar{\mu} \geq 0,\ \langle \bar{\mu}, \bar{x} \rangle = 0. \tag{4.40}$$

We next show that there exists only one $\bar{\pi} \in \operatorname{Im} A$ which satisfies the left equality in (4.40) for the given $\bar{x}$ and $\bar{\mu}$. Let $\tilde{\pi}$ be any other element in $\operatorname{Im} A$ such that $H\bar{x} + g + A^\top \tilde{\pi} - \bar{\mu} = 0$. Subtracting this equality from the first one in (4.40), we conclude that

$$(\bar{\pi} - \tilde{\pi}) \in \operatorname{Ker} A^\top, \quad (\bar{\pi} - \tilde{\pi}) \in \operatorname{Im} A,$$

As $\operatorname{Ker} A^\top = (\operatorname{Im} A)^\perp$, it follows that $\bar{\pi} = \tilde{\pi}$, i.e., the element with the properties under consideration is unique. Observe further that the solution $\hat{\pi}$ of (4.37) satisfies those properties: it exists, is unique, and $H\bar{x} + g + A^\top \hat{\pi} - \bar{\mu} = 0$. Further, by the optimality condition for (4.37) it holds that there exists some λ such that $\hat{\pi} + A\lambda = 0$, i.e., $\hat{\pi} \in \operatorname{Im} A$. As we have shown that such an element is unique, it follows that $\bar{\pi} = \hat{\pi}$.

In particular, $\{A^\top \pi^{k_j}\} \to a = A^\top \bar{\pi}$ now means that $\{A^\top(\pi^{k_j} - \hat{\pi})\} \to 0$ as $j \to \infty$. Finally, we show that this implies that $\pi^{k_j} \to \hat{\pi}$ (recall that $(\pi^{k_j} - \hat{\pi}) \in \operatorname{Im} A$).

To that end, recall that for any matrix A there exists $\gamma > 0$ such that

$$\|A^\top Au\| \geq \gamma \|Au\| \quad \text{for all } u.$$

(To see this, assume the contrary, i.e., that there exists $\{u^k\}$ such that $Au^k \neq 0$ and $\|A^\top Au^k\|/\|Au^k\| \to 0$. Passing onto a further subsequence, if necessary, $Au^k/\|Au^k\| \to v \in \operatorname{Im} A$, $v \neq 0$, $A^\top v = 0$. This gives a contradiction, since $\operatorname{Ker} A^\top \cap \operatorname{Im} A = \{0\}$.)

As $(\pi^{k_j} - \hat{\pi}) \in \operatorname{Im} A$, there exists some b^j such that $\pi^{k_j} - \hat{\pi} = Ab^j$. Then,

$$\begin{aligned} \|A^\top(\pi^{k_j} - \hat{\pi})\| &= \|A^\top Ab^j\| \geq \gamma \|Ab^j\| \\ &= \gamma \|\pi^{k_j} - \hat{\pi}\|, \end{aligned}$$

implying the assertion, since the left-hand side tends to zero as $j \to \infty$. $\square$

The development above is for quadratic functions and nonnegativity constraints. We next adapt it to problem (4.13), which has upper bounds as well.

The set X in (4.13) is given by

$$X = \{x \in \mathbb{R}^n : 0 \leq x \leq b\}.$$

Then, the normal cone at $\bar{x} \in X$ has the form

$$N_X(\bar{x}) = \{\mu \in \mathbb{R}^n : \mu_i \leq 0, \text{ if } x_i = 0, \ \mu_i \geq 0 \text{ if } x_i = b_i, \ \mu_i = 0, \text{ if } 0 < x_i < b_i\}.$$

For a solution $\bar{x}$ of problem (4.13), the set of associated Lagrange multipliers (π, μ_0, μ_b) is characterized by the following system:

$$\begin{array}{c} g + A^\top \pi - \mu_0 + \mu_b = 0, \quad \bar{x} \geq 0, \ \mu_0 \geq 0, \ \langle \mu_0, \bar{x} \rangle = 0, \\ \bar{x} \leq b, \ \mu_b \geq 0, \ \langle \mu_b, \bar{x} - b \rangle = 0. \end{array} \tag{4.41}$$

Denote by $\hat{\pi} = \hat{\pi}(\bar{x}, \bar{\mu}_0, \bar{\mu}_b)$ the minimal-norm element which solves (4.41) for the given $\bar{x}$ and $(\bar{\mu}_0, \bar{\mu}_b)$, i.e., the (unique) solution of

$$\min \frac{1}{2}\|\pi\|^2 \quad \text{s.t.} \quad g + A^\top \pi - \bar{\mu}_0 + \bar{\mu}_b = 0. \tag{4.42}$$

Theorem 4.2.7 (Primal and dual convergence in the setting of problem (4.13)). *Let the triplet (x^k, μ_0^k, μ_b^k) denote the optimal primal and dual solutions (Lagrange multipliers) to* (4.14)*, written with $\beta = \beta_k$, and let $\pi^k :=$*

$(Ax^k - b)/\beta_k$.

If $\beta_{k+1} < \beta_k$ for all k, and $\beta_k \to 0$, the following holds:

(i) The primal sequence $\{x^k\}$ is bounded and any of its accumulation points is a solution to (4.13).

(ii) Let $\{x^{k_j}\} \to \bar{x}$ as $j \to \infty$. If the condition (4.24) *holds, then $\{\mu_0^{k_j}\}$ and $\{\mu_b^{k_j}\}$ are bounded. Moreover, for any accumulation point $(\bar{\mu}_0, \bar{\mu}_b)$ of the subsequence $\{(\mu_0^{k_j}, \mu_b^{k_j})\}$, the corresponding subsequence $\{\pi^{k_j}\}$ converges to $\hat{\pi}$, the solution of* (4.42).

The point $(\bar{x}, \hat{\pi}, \bar{\mu}_0, \bar{\mu}_b)$ is a primal-dual solution of (4.13).

Proof. Item (i) is immediate: Every sequence $\{x^k\}$ generated by the method is automatically bounded (and thus has a convergent subsequence), because the set $X = \{0 \le x \le b\}$ is compact. Then, every accumulation point of $\{x^k\}$ is a solution of (4.13), by Theorem 4.2.3.

We proceed to Item (ii). By KKT conditions for (4.14), we have that

$$\begin{gathered} g + \frac{1}{\beta_k} A^\top (Ax^k - a) - \mu_0^k + \mu_b^k = 0, \\ x^k \ge 0,\ \mu_0^k \ge 0,\ \langle \mu_0^k, x^k \rangle = 0, \quad x^k \le b,\ \mu_b^k \ge 0,\ \langle \mu_b^k, x^k - b \rangle = 0. \end{gathered} \tag{4.43}$$

Recalling the definition of π^k and setting $\mu^k := \mu_b^k - \mu_0^k$, we obtain that

$$g + A^\top \pi^k + \mu^k = 0. \tag{4.44}$$

Let $\{x^{k_j}\} \to \bar{x}$ as $j \to \infty$. We next prove that the sequence $\{(\mu_0^{k_j}, \mu_b^{k_j})\}$ is bounded. To that end, using the second line in (4.43), first observe the following:

$$\begin{gathered} (\mu_0^k)_i > 0 \ \Rightarrow\ x_i^k = 0 \ \Rightarrow\ (\mu_b^k)_i = 0,\ \mu_i^k = -(\mu_0^k)_i < 0, \\ (\mu_b^k)_i > 0 \ \Rightarrow x_i^k = b \ \Rightarrow (\mu_0^k)_i = 0,\ \mu_i^k = (\mu_b^k)_i > 0. \end{gathered} \tag{4.45}$$

From those relations, it is obvious that $\{(\mu_0^{k_j}, \mu_b^{k_j})\}$ is bounded if and only if $\{\mu^{k_j}\}$ is bounded.

Next, similarly to the proof of Theorem 4.2.6, suppose by contradiction that (4.43) (and thus (4.44)) hold with $\|\mu^{k_j}\| \to +\infty$. Passing onto a subsequence, if necessary, let $\{\mu^{k_j}/\|\mu^{k_j}\|\} \to \bar{\mu} \neq 0$. Denote $u^{k_j} = -A^\top \pi^{k_j}/\|\mu^{k_j}\| \in \operatorname{Im} A^\top$. Dividing the equality in (4.44) by $\|\mu^{k_j}\|$ and passing onto the limit as $j \to \infty$, it follows that

$$u^{k_j} = (g + \mu^{k_j})/\|\mu^{k_j}\| \to \bar{\mu} \neq 0.$$

As $u^{k_j} \in \operatorname{Im} A^\top$, $u^{k_j} \to \bar{\mu}$, and $\operatorname{Im} A^\top$ is closed, we conclude that $\bar{\mu} \in \operatorname{Im} A^\top$.

Observe now that from (4.45) it follows that

$$\bar{\mu}_i < 0 \;\Rightarrow\; \bar{x}_i = 0 \quad \text{and} \quad \bar{\mu}_i > 0 \;\Rightarrow\; \bar{x}_i = b.$$

This shows that $\bar{\mu} \in N_X(\bar{x})$. As $\bar{\mu} \neq 0$ and $\bar{\mu} \in \operatorname{Im} A^\top$, we obtain a contradiction with (4.24) (or (4.23)).

It follows that $\{\mu^{k_j}\}$ is bounded. And as already observed from (4.45), this means that $\{\mu_0^{k_j}\}$ and $\{\mu_b^{k_j}\}$ are bounded.

The the proof that the corresponding subsequence $\{\pi^{k_j}\}$ converges to $\hat{\pi}$, the solution of (4.42), is analogous to that in Theorem 4.2.6.

□

Some comments are in order about the relations between the statements in Theorem 4.2.7 and [LSS19b]. The fact that all accumulation points of $\{x^k\}$ are solutions to (4.13) is a standard property of exterior penalty methods, see [LSS19b, Thm. 3.1] and accompanying comments. However, the *existence* of accumulation points (i.e., boundedness of $\{x^k\}$) is not automatic. In [LSS19b, Thm. 3.3], boundedness of the primal sequence was established under certain assumptions. Note that in the setting of problem (4.13) we do not need any assumptions, as (4.14) has box-constraints (and so its feasible set is compact). The role of the *partial* Mangasarian–Fromovitz condition (4.23) or (4.24) is to ensure boundedness of the multipliers associated to box-constraints, while allowing the set of multipliers associated to the equality constraints to be unbounded (and in particular, allow for the matrix A to be not of full rank).

The results presented so far provide a constructive answer to our initial question, on how to devise a solution methodology yielding the minimal-norm multiplier. The mechanism is applied in the next chapter to an important class of stochastic optimization problems, with linear objective function and affine constraints, and where uncertainty is dealt with by sample average approximations in two stages.

4.3 Price Signal Analysis on an Illustrative Example

The price signals given by the demand constraint correspond to components of the optimal multiplier associated with the last constraints in (3.3), with right-hand side vector h^s, for $s \in \mathbb{S}$. A common practice in the energy sector is to average those signals and use the resulting mean price to guide the company business strategies.

We now give a simple example that can be solved analytically for checking, and which illustrate well our theoretical results (the satisfaction of condition (4.23) and convergence to the minimal-norm price (4.37)). Take 2 equiprobable scenarios and let $n_1 = n_2 = 2$. The first-stage cost $(F_{11}, F_{12}) \in$

$\mathbb{R}^2$ and second-stage costs are deterministic $F_2^1 = F_2^2 = F_2 \in \mathbb{R}^2$, with components F_{21} and F_{22}. The technology and recourse matrices are

$$T := \begin{bmatrix} 1 & 0 \\ 0 & 1 \\ 0 & 0 \end{bmatrix} \quad \text{and} \quad W := \begin{bmatrix} 2 & 0 \\ 1 & -1 \\ 1 & 2 \end{bmatrix},$$

so $m = 3$. The uncertain right-hand side terms are given by $h^1 := (1, 1, 1)^\top$ and $h^2 := (1, 0, 3)^\top$.

Working out the algebra shows that the feasible set in (4.1)-(4.2) is completely determined by the first component of x_1, denoted by $y \geq 0$ below. Specifically,

$$\begin{gathered} x \text{ is feasible if and only if, for some } y \geq 0\,, \\ x_1 := \left(y, \frac{3}{4}(1+y)\right)^\top,\ x_2^1 := \left(\frac{1}{2}(1-y), \frac{1}{4}(1+y)\right)^\top, \\ x_2^2 := \left(\frac{1}{2}(1-y), \frac{1}{4}(5+y)\right)^\top, \end{gathered}$$

and, therefore, the following one-dimensional problem is equivalent to (4.1)-(4.2):

$$\min_{y \geq 0} \left(F_{11} + \frac{3}{4}F_{12} - \frac{1}{2}F_{21} + \frac{1}{4}F_{22}\right)y + \frac{3}{4}F_{12} + \frac{1}{2}F_{21} + \frac{3}{4}F_{22}\,.$$

Its optimal solution is $\bar{y} = 0$, as long as

$$F_{11} \geq -\frac{3}{4}F_{12} + \frac{1}{2}F_{21} - \frac{1}{4}F_{22}\,. \tag{4.46}$$

The optimal value for the primal variable is

$$\bar{x}_1 := \left(0, \frac{3}{4}\right)^\top,\ \bar{x}_2^1 := \left(\frac{1}{2}, \frac{1}{4}\right)^\top,\ \bar{x}_2^2 := \left(\frac{1}{2}, \frac{5}{4}\right)^\top.$$

To compute the optimal value for the multiplier, recall that any normal element $\bar{\mu} \in N_X(\bar{x})$ has all of its components null, except for the first one, because $\langle \bar{\mu}, \bar{x} \rangle = 0$. Therefore,

$$\bar{\mu} = -\alpha e_1 \quad \text{for some } \alpha \geq 0\,,$$

where $e_j \in \mathbb{R}^6$ is the j-th canonical vector (all components are zero except the j-th, equal to 1). To check that (4.21) is satisfied, consider its equivalent formulation (4.24). Suppose $\bar{\mu} = -\alpha e_1 \in \operatorname{Im} A^\top$. For condition (4.24) to hold, for any $\nu \in \operatorname{Ker} A$ we must have that $-\alpha\langle e_1, \nu \rangle = 0$, because the subspaces $\operatorname{Im} A^\top$ and $\operatorname{Ker} A$ are orthogonal. Since the latter (one-dimensional) subspace is generated by the vector $s := \left(4, 3, -2, 1, -2, 1\right)^\top$, we have that $\langle \bar{\mu}, e^1 \rangle =$

-4α, forcing $\alpha = 0$. It is then seen that (4.24) and (4.23) hold, as claimed.

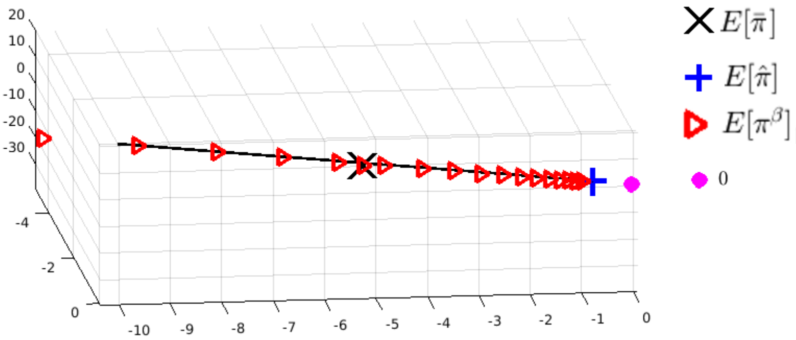


Figure 4.1: Unbounded set of optimal multipliers in mean (the line), the element with minimal norm (the plus sign), the mean multiplier found for $\beta = 0$ (the cross), the mean multiplier estimates for different values of $\beta > 0$ (the triangles), and origin (the dot).

Take $F_{21} = F = -F_{22}$ for some F, $F_{12} = 0$, and any $F_{11} \geq F$ (so that (4.46) is satisfied). Optimal Lagrange multipliers must solve the system

$$A^\top \pi = -g + \bar{\mu}\,,$$

with

$$g = \left(F_{11}, 0, \frac{F}{2}, -\frac{F}{2}, \frac{F}{2}, -\frac{F}{2}\right)^\top \quad \text{and } \bar{\mu} = -\alpha e_1 \text{ for } \alpha \geq 0\,.$$

After some algebraic manipulations, the unbounded optimal multiplier set is:

$$\mathcal{L} := \left\{\bar{\pi} = t\frac{F_{11}}{2}\Big(1, -4, 2, -3, 4, 2\Big)^\top \mid t \geq 1\right\}.$$

Hence, $t = 1$ gives the minimal-norm element, for which

$$\mathbb{E}\left[\hat{\pi}\right] = \frac{F_{11}}{2}\Big(-1, 0, 2\Big)^\top.$$

Applying our approach with several decreasing values of β gives the multiplier estimates $\pi^\beta \in \mathbb{R}^6$ with the mean $\mathbb{E}\left[\pi^\beta\right] \in \mathbb{R}^3$. The line in Figure 4.1 shows a portion of the mean optimal multiplier set,

$$\left\{t\frac{c_1}{2}\Big(-1, 0, 2\Big)^\top \mid t \geq 1\right\}\,.$$

The dot represents the origin in $\mathbb{R}^3$, the plus sign $\mathbb{E}\left[\hat{\pi}\right]$, the minimal-norm multiplier in mean value, to which the mean values $\mathbb{E}\left[\pi^\beta\right]$, represented with triangles, converge as $\beta \to 0$. The cross displays $\mathbb{E}\left[\bar{\pi}\right]$, the mean multiplier found when solving (3.3), whose norm is larger than the minimal one.

Numerical results for a large group of academic two-stage linear programs, and for a real-life case of the Nordic power system, are given in the next chapter.

Chapter 5

Computational Experience

Theorem 4.2.6 describes theoretically the behavior of the mean value of regularized price signals in terms of the original optimization problem. The numerical examples below illustrate the main features of our approach.

In this chapter we perform tests using a set of toy problems, in Section 5.1, measuring the trade-off between proximity to the original solution and variability of the output. Section 5.2 continues with a numerical assessment of the analytical example presented in Section 3.2.1. Results for the Nordic Energy system, in a simplified form, and for the real-life instance of ENGIE, are reported in Sections 5.3 and 5.4, respectively.

5.1 Results on a Battery of Academic Instances

Perturbing the dual second-stage problem with a term "$-\beta\|\pi\|^2$" changes the solutions, with respect to the original problem (4.1)-(4.2). Our goal is to keep close the original marginal cost, and at the same time, decrease its variance. The contents of this section is part of the article [LSS19b].

Theoretical results in terms of variance are not simple. It is not always true that regularization reduces variance, but it happens for a large amount of problems. To make sure we are going in the right direction, for our performance profiles [DM02] we created an index that measures the joint dynamics of variance reduction and distance to the original dual solution set, as β tends to zero.

For a battery of two-stage stochastic linear programming problems we compare the expected value of the multipliers obtained as follows:

- when solving problem (3.3) in its two-level formulation (4.1) with recourse function (4.2), by a proximal bundle method [LS97]; see also [Bon+06, Ch. 10.3]; and
- with our proposal, i.e., solving, for decreasing values of β, several instances of (4.1) with regularized recourse function (4.3).

All the tests were run in Matlab R2016, on an Intel Core i5 computer with 2.4 GHz, 4 cores and 4 GB RAM, running under Ubuntu 18.04.1 LTS and using Gurobi 5.6 optimization toolbox for Matlab.

The battery comprises 50 problems, for which 10 independent instances, each one with 50 scenarios, were created. The considered two-stage stochastic problems are of the form (3.3) with $b_2 = +\infty$ and uncertainty only on the right hand-side $h \in \mathbb{R}^m$, independently and normally distributed. The expectation and standard deviation of the considered distribution is problem-dependent and proportional to $c/2$. The problem dimension ranges are $n_1 \in \{20, 40, 60\}$, $n_2 \in \{30, 60, 90\}$, and $m \in \{20, 40, 60\}$. For full details we refer to [Dea06]; see also [OSS11].

The test has a total of 500 runs, labeled $P = 1, \ldots, 500$. With the purpose of doing a performance profile, we compute

$$\left\|\mathbb{E}\left[\pi_P^{\text{best}}\right]\right\| := \arg\min\left\{\|\mathbb{E}\left[\pi_P^{\beta}\right]\|\colon \beta \in \{0, 0.1, 0.2, \ldots, 0.5\}\right\},$$

and define, for problem p and parameter $\beta \geq 0$, the following index:

$$c_P^{\beta} := \frac{\left\|\text{Var}\left[\mathbb{E}\left[\pi_P^{\beta}\right]\right]\right\|}{\left\|\text{Var}\left[\pi_P^{\text{best}}\right]\right\|} + \left(1 - \frac{\left\|\mathbb{E}\left[\pi_P^{\beta}\right]\right\|}{\left\|\mathbb{E}\left[\pi_P^{\text{best}}\right]\right\|}\right). \tag{5.1}$$

c_P^{β} is a compromise between a small variance and a small mean relative to the π_P^{β} with smaller mean, that here is defined as the better one. For $\pi_P^{\beta} = \pi_P^{\text{best}}$, $c_P^{\beta} = 1$ but other values of β can have a smaller value of c_P^{β} due to a smaller variance.

Here, π_P^0 corresponds to $\bar{\pi}$ in our previous notation, while π_P^{β} is the Lagrange multiplier computed for problem P with regularization parameter $\beta \geq 0$. The corresponding performance profile is given in Figure 5.1.

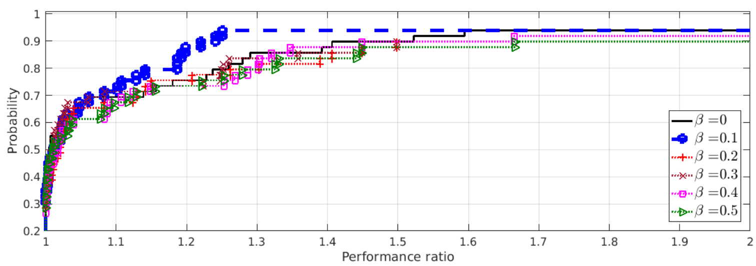


Figure 5.1: Combined gains in expected value and variance for the dual variable.

As expected, in terms of the combined index, the multiplier of the original problem ($\beta = 0$) performs worse, confirming the empirical observation that in general $Var[\pi_P^{\beta}] \leq Var[\pi_P^0]$. For this set of runs, the value $\beta = 0.1$ (dashed line with circles) seems to give a good compromise between stability of the mean multiplier, and approximation of the minimal-norm multiplier.

For completeness, we present in Figure 5.2 a performance profile of the first-stage variable, measured now with the index

$$\tilde{c}_P^\beta := \left(1 - \frac{\left\|\mathbb{E}\left[{x_1}_P^\beta\right]\right\|}{\left\|\mathbb{E}\left[{x_1}_P^0\right]\right\|}\right),$$

defined for $\beta > 0$ (in our approach comparing variances is not sound, as there is no "stabilization" of the primal variables). The graph shows that the best value for β in the dual performance profile in Figure 5.1 ($\beta = 0.1$, dashed line with circles), also behaves reasonably well in the primal variable.

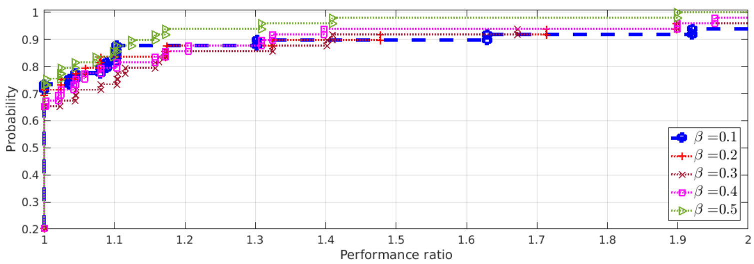


Figure 5.2: Progression of expected value of first-stage variable

5.2 Revisiting the Analytical Case

For the simple problem (3.4), an explicit expression of the prices is available in (3.5) and (4.8), respectively for $\beta = 0$ and $\beta > 0$.

$\xi \sim \mathcal{N}(0, 10)$, the normal distribution

$F_2 = F_1 = 5$

$\Xi = \{\xi^1, \ldots, \xi^{\mathbb{S}}\}$ is a sample, and $S = 200$

Ξ^n, $n \in \{1, \ldots, N\}$ are different samples, $n = 40$

For the first test we fix $\beta = 1$, and compute the numerical value of the multipliers $\bar{\pi}$ and $\bar{\pi}^\beta$. The respective histograms are shown in Figure 5.3.

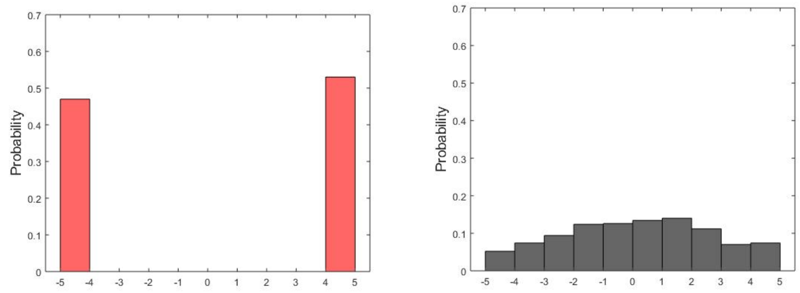


Figure 5.3: Non-regularized (left) and regularized (right) price signal distributions

The values observed in the histograms are consistent with the theory: on the left, the values of $\bar{\pi}$ oscillate between -5 and 5, (that is, $-F_2$ and F_2); on the right, the distribution of $\bar{\pi}^\beta$ is smoother.

To see the impact of regularization on the expected value and variance, we repeated the same test with 40 different samples. Recall that the analytical result for the non-regularized moments in (3.6) is

$$\mathbb{E}[\bar{\pi}(\xi)] = 0 \quad \text{and} \quad \mathbb{V}ar[\bar{\pi}(\xi)] = \mathbb{E}[\bar{\pi}(\xi)^2] = F_2^2\,.$$

The distribution of $\bar{\pi}^\beta$ is computed using $\bar{x}_1^\beta$, the regularized primal solution for the one-level regularized problem (4.7). Clearly, results cannot match exactly the theoretical ones because we are using a finite sample normally distributed, instead of the continuous normal distribution. The expected value and variance for the 40 samples are reported in Figure 5.4.

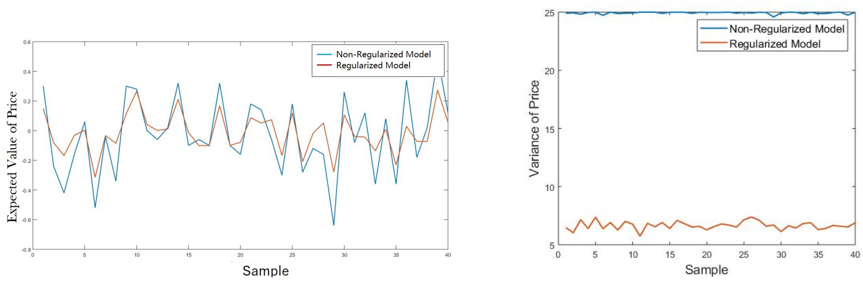


Figure 5.4: Expected value and variance for 40 samples

On the left plot in Figure 5.4 we confirm that the expected value is around zero, with the non-regularized model exhibiting a higher variability. On the right plot, we see that when $\beta = 0$ the variance turns around $25 = 5^2 = F_2^2$, but it is considerably smaller for the regularized model.

Our final graph in Figure 5.5 displays the variance of the expected value for different values of the regularization parameter. The plot can help choosing the "best" value of β, taking into account the trade-off between lowering variance and not getting too far from the true expected value ($\beta = 0$).

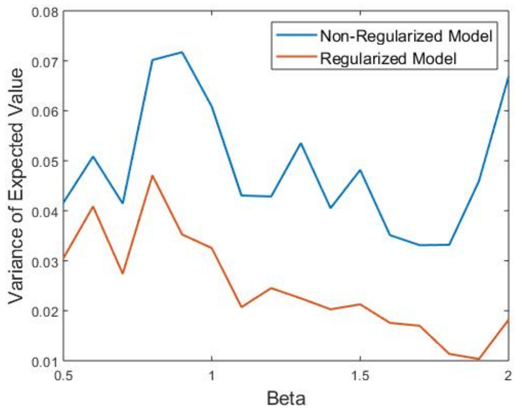


Figure 5.5: Variance of the expected value for different values of β

5.3 Simplified Nordic Energy System

In section 3.2.3 we presented a simplification of the energy generation problem in Northern Europe. For that example, we now verify the behavior of the regularized problem. The multistage problem is represented by a two-stage stochastic program. The first stage represents winter and autumn, and the second stage represents the seasons of spring and summer that follow.

The results reported in Table 3.3 for $\beta = 0$, illustrate well the difficulties in basing business decisions on the price signal distribution of one given sample. We again consider three samples $\Xi_i := \{\xi_i^1, \ldots, \xi_i^{\mathbb{S}}\}$ for $i = 1, 2, 3$ and $\mathbb{S} = 80$ scenarios, and compute the price signals with the regularized approach, for three different values of $\beta > 0$.

Table 5.1 shows how the mean price signal got stabilized with the regularization, where the values vary much less than in Table 3.3. On the other hand, differences between the expected value of the regularized and non-regularized prices represent just 5% of the original value.

Sample	$\mathbb{E}[\pi^\beta],\ \beta = 8$	$\mathbb{E}[\pi^\beta],\ \beta = 12$	$\mathbb{E}[\pi^\beta],\ \beta = 25$
$\mathbb{S}_1$	27,16	27,34	27,8
$\mathbb{S}_2$	27,2	27,33	27,76
$\mathbb{S}_3$	26,98	26,34	26,43

Table 5.1: Expected price signals for three values of β

The beneficial effect of the regularization is also clear in the the histograms in Figure 5.6, to be compared with those in Figure 3.3, with prices ranging between 25 and 35. The colors of the histograms in Figure 5.6 is meant for the different samples $\mathbb{S}_1, \mathbb{S}_2$ and $\mathbb{S}_3$. We observe that distributions get progressively smoother as β increases.

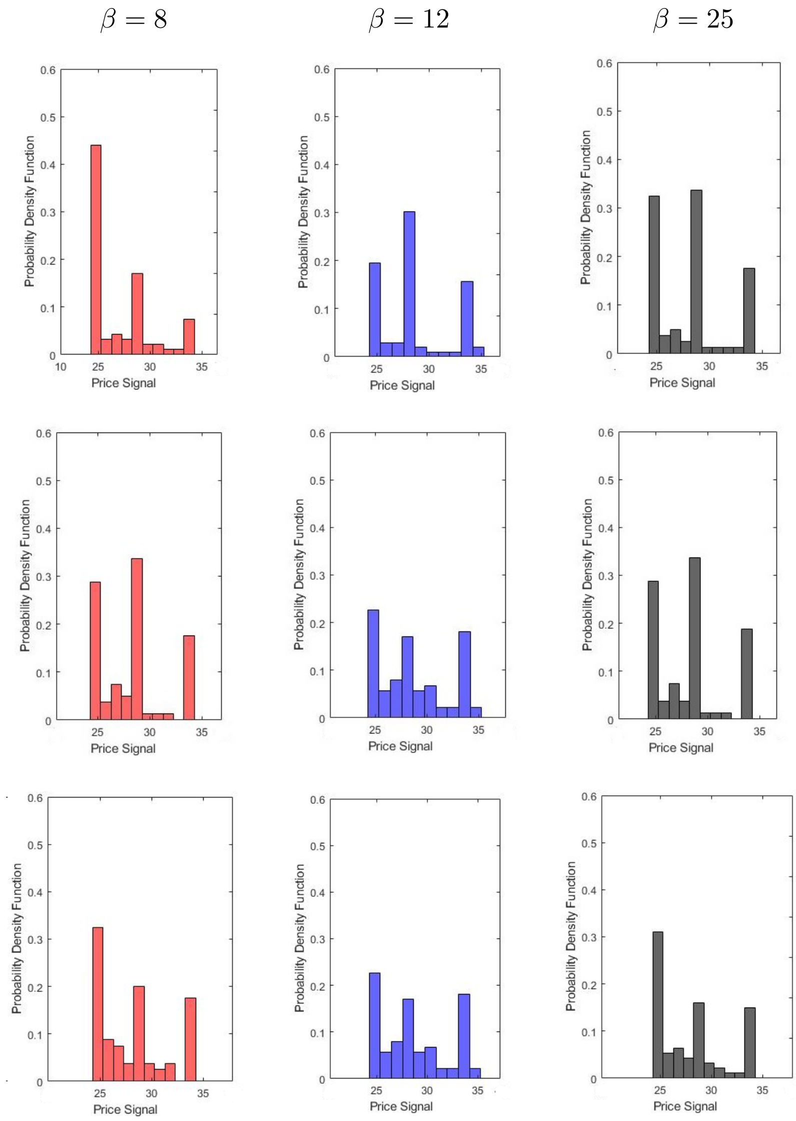


Figure 5.6: Price distribution of three samples for different values of β

5.4 Realistic System

The Nordic energy system in section 3.1 is composed of 12 bidding zones. Norway, with 5 zones, has the largest percentage of hydro-energy generation, which amounts to an equivalent of 95.2% of its demand. Other countries have several different sources of energy, Sweden has a large nuclear generation, and Denmark wind generation amounted to half of its demand in 2014. As in the diagram in Figure 3.1, some zones were incorporated to handle imports and exports.

The benchmark uses real and estimated data for historical inflows, generation costs and capacity of each power plant in the system, importation costs, minimum and maximum level of reservoirs and maximum flow between zones.

Our goal is to compare the price signals obtained with the multistage original problem and with the regularized approach running in a rolling-horizon mode. We simulate different scenarios as input data and examine the Wasserstein distance between the histograms of the respective price signals. We expect to obtain similar expected values, but a smoother distribution.

A similar measure for the quality of optimal decisions is considered by K. Hoyland and S.W. Wallace, who proposed in [HW01] a method to build a scenario tree that preserves certain essential statistical properties. The different trees generated with their approach maintain the optimal value of the considered optimization problem. The rationale is that if the relevant statistical information is captured by the method, the result should not vary too much with the data input. Along these lines, J. Higle and S. Sen show in [HS94] that primal regularization does not affect the statistical properties of the solution and the number of iterations required by algorithms based on cutting planes. In our tests, solving the regularized model required slightly more iterations.

In the numerical assessment that follows, we analyze the price distribution as well as the impact of regularization on the reservoirs levels, the hydro-generation, and the exchanges between zones. We also compare the output with the one provided by the well-known *Stochastic Dual Dynamic Programming* (SDDP) method [PP91].

5.4.1 Rolling-horizon Methodology, Simulation, and Data

The time horizon of one year was discretized in $t = 1, \dots, 8760$ hours. At the first hour of each week, the inflow uncertainty of the whole week becomes known. Since the year has 54 weeks, this defines a multistage structure of uncertainty, that we cast in our two-stage setting as follows.

We put in place a *rolling-horizon* mode, in which we solve $w = 1, \dots, 53$ two-stage problems derived from (3.1) in Section 3.1. In the w-th two-stage problem, the decision variables of the w-th week are considered in the first stage. The uncertainty of the remaining $w + 1, \dots, 54$ weeks is revealed at once, at the end of the week w and, hence, the corresponding decision variables are considered in the second stage. Since one week has 168 hours, the time horizon of the w-th problem covers $\mathbb{T}^w := (54 - w + 1)168$ hours. The output of the w-th two-stage problem provides input for the problem

$w+1$, similarly to [GS12]; see also [Bis+17].

With this mechanism, the first-stage components of the decision vector of problem w are in fact decision variables for the w-th week. Accordingly, if $\bar{x}_1^{\beta;w}$ denotes the first-stage component of a solution obtained for the w-th two-stage problem, then the policy

$$\left\{\bar{x}_1^{\beta;w} : w = 1, \ldots, 52\right\}$$

is implementable, as defined in Section 2.2.

This is the primal policy assessed in out-of-sample simulations of the system operation, after the optimization phase. There is also a dual implementable policy, formed by the cuts of the successive expected recourse functions at a solution, say

$$\left\{ \text{ the cuts for } \mathbb{E}\left[\mathbb{Q}^{\beta,s;w}(\bar{x}_1^{\beta;w})\right]; w = 1, \ldots, 53\right\} .$$

More details on the data for the benchmark is given below.

- $\mathbb{L} := 30$ balancing zones.
- $\mathbb{J} := 21$ hydro-power plants.
- $\mathbb{I} := 224$ power plants.
- Generation costs are from ENGIE's data base. We assume in this work that hydro-power plants have zero production cost.
- The inflow uncertainty was generated using a log normal distribution calibrated with historical data.

The benchmark compares the performance of SDDP against RH and RRH, respectively corresponding to the rolling-horizon non-regularized and regularized two-stage models.

5.4.2 The Problem to be Solved

In this section we define mathematically the problem solved in a rolling horizon. In the next sections we shall use this formulation as a reference to explain our simulations.

Following the notation introduced in section 3.1 to formulate the energy management problem, for the w-th week we let

$$\mathbb{T}_w = \{t \ : \ 7(w-1) < t \leq 7w\}, \text{ and } A_w = \{t \ : \ 7w < t\}.$$

We write the problem for the w-th week as follows:

$$\left\{\begin{array}{lll}\min & \displaystyle\sum_{t\in\mathbb{T}_w}\sum_{l\in\mathbb{L}}\sum_{i\in\mathbb{I}_l} C_i^t gt_i^t + \mathbb{E}[\mathbb{Q}_w^\beta(v_j^{7w}, \mathcal{I}_j^s)] & \\ \text{s.t.} & \underline{v}_j^t \le v_j^t \le \overline{v}_j^t \text{ and } 0 \le sp_j^t, & j\in\mathbb{J}_l\,, l\in\mathbb{L}\,, t\in\mathbb{T}_w \\ & 0\le gh_j^t \le \overline{gh}_j^t, \quad 0 \le gt_i^t \le \overline{gt}_i^t, & j\in\mathbb{J}_l\,, i\in\mathbb{I}_l\,, l\in\mathbb{L}\,, t\in\mathbb{T}_w \\ & 0\le f_{l\leftrightarrow l_1}^t, & l\in\mathbb{L} : l_1\in\mathbb{F}_l\neq\emptyset\,, t\in\mathbb{T}_w \\ & v_j^t - v_j^{t-1} + gh_j^t + sp_j^t = \mathcal{I}_j^t, & j\in\mathbb{J}_l\,, l\in\mathbb{L}\,, t\in\mathbb{T}_w \\ & \displaystyle\sum_{j\in\mathbb{J}_l} gh_j^t + \sum_{i\in\mathbb{I}_l} gt_i^t + \sum_{l_1\in\mathbb{F}_l\neq\emptyset} f_{l\leftrightarrow l_1}^t = \mathcal{D}_l^t, & l\in\mathbb{L}\,, t\in\mathbb{T}_w\,. \end{array}\right.$$

where $v_j^{7(w-1)}$ is the reservoir level decision that had been made in the $(w-1)$-th week. Here, $\mathcal{I}_j^t$ is random, but it does not depend on the scenario s. The sequence $(\mathcal{I}_{j,w}^t)$, where $t\in\mathbb{T}_w$ is a path in the tree of scenarios.

The regularized recourse functions defined in the second stage are

$$\mathbb{Q}_w^\beta(v_j^{7w}, \mathcal{I}_j^s) :=$$

$$\left\{\begin{array}{llll}\min & \displaystyle\sum_{t\in A_w}\sum_{l\in\mathbb{L}}\sum_{i\in\mathbb{I}_l} C_i^t gt_i^t & \multicolumn{2}{l}{+\tfrac{1}{2\beta}\|v_j^{7n+1} - v_j^{7w} + gh_j^{7w+1} + sp_j^{7w+1} - \mathcal{I}_j^{7w+1,s}\|^2} \\ \text{s.t.} & \underline{v}_j^t \le v_j^t \le \overline{v}_j^t \text{ and } 0 \le sp_j^t, & j\in\mathbb{J}_l\,, l\in\mathbb{L}, & t\in A_w \\ & 0\le gh_j^t \le \overline{gh}_j^t, \quad 0 \le gt_i^t \le \overline{gt}_i^t, & j\in\mathbb{J}_l\,, i\in\mathbb{I}_l\,, l\in\mathbb{L}, & t\in A_w \\ & 0\le f_{l\leftrightarrow l_1}^t, & l\in\mathbb{L} : l_1\in\mathbb{F}_l\neq\emptyset, & t\in A_w \\ & v_j^t - v_j^{t-1} + gh_j^t + sp_j^t = \mathcal{I}_j^{t,s} \;\leftrightarrow \pi_j^{t,s}, & j\in\mathbb{J}_l\,, l\in\mathbb{L},, & t\in A_w \\ & \displaystyle\sum_{j\in\mathbb{J}_l} gh_j^t + \sum_{i\in\mathbb{I}_l} gt_i^t + \sum_{l_1\in\mathbb{F}_l\neq\emptyset} f_{l\leftrightarrow l_1}^t = \mathcal{D}_l^t, & l\in\mathbb{L}\,, & t\in A_w\,, \end{array}\right.$$

where $\mathcal{I}_j^{t,s}$ is random and depends on the scenario s. When $\beta = 0$, the constraint in the objective function moves to the feasible set.

In this notation the sequence $\left\{\bar{x}_1^{\beta;w} : w = 1,\ldots,52\right\}$ corresponds to the reservoir level $\bar{x}^{\beta;w} = (v_j^{7w})$, while the cuts for the function $\mathbb{E}\left[\mathbb{Q}^{\beta,s;w}(\bar{x}_1^{\beta;w})\right]$ are expressed in terms of the couple $(\bar{\pi}_{j,k}^{7w+1}, \delta_{it}^w)$, for each iteration k, where $\bar{\pi}_k^{7n+1,k} > 0$, and:

$$\bar{\pi}_{j,it}^{7w+1,k} = \frac{1}{S}\sum_s \pi_{j,it}^{7w+1,k,s},$$

$$\delta_k^w = \frac{1}{S}\sum_s \mathbb{Q}_{w,k}(v_j^{7w}, \mathcal{I}^s) - \sum_j \bar{\pi}_{j,k}^{7w+1,k} v_j^{7w}.$$

5.4.3 Optimization Phase

The optimization part of the experiment uses an in-sample set with 30 inflow scenarios and takes $\beta := 7000$ for RRH (below we explain that with this value, the magnitude of the violation incurred by RRH is less then 1% for the whole system). The SDDP method takes one scenario randomly in the forward pass and all the 30 scenarios in the backward pass. The rolling-horizon variants take the same forward scenario, for the different weeks $w = 1, 2, \ldots$, including the 30 scenarios from week $w + 1$ to week 48 in the second stage of the w-th two-stage problem.

The aggregate reservoir management (adding all the hydro-plants), is shown in Figure 5.7, where colors correspond to the different weeks. The values are normalized with respect to the maximum system capacity.

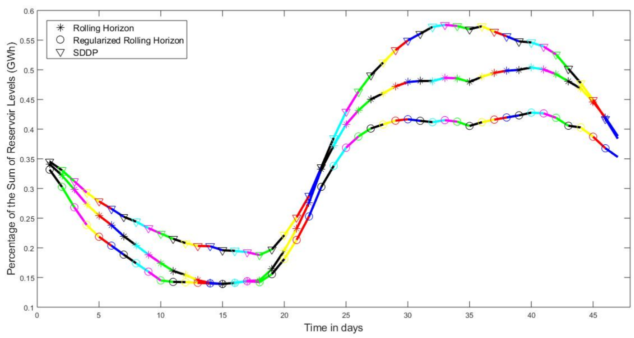


Figure 5.7: Aggregate reservoir management - optimization phase

The levels of reservoirs with the rolling-horizon modes are lower, with the RRH using the most of water. Since SDDP sees a larger portion of the scenario tree in the backward pass (not only the tail of the weeks $w+1, \ldots, 54$), SDDP water management is more conservative. Regarding the comparison between RRH and RH, note that the inequality $\mathbb{Q}^{\beta,s}(\cdot) \leq \mathbb{Q}^{s}(\cdot)$ always holds (the feasible set defining the former is included in the one defining the latter). Having an under-estimation of the future-cost function, RRH tends to be less conservative than RH, keeping less water in the reservoirs.

Since RRH does not consider the water-balance equations, in Figure 5.8 we examine the gap

$$v_j^t - v_j^{t-1} + gh_j^t + sp_j^t - \mathcal{I}_j^{t,s}$$

over the first 48 weeks for $s = 1, \ldots, 12$ scenarios (different colors in the figure), for two hydro-plants in Norway.

We note that the system exploits most the possibility of not satisfying the water-balance equation in the beginning of the year, when inflows are smaller. In order to determine the real extent of the violation, relative to

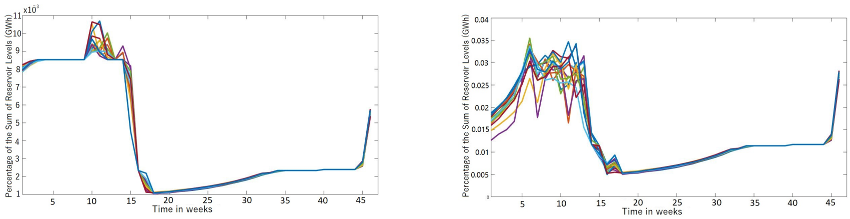


Figure 5.8: Absolute violation of water balance - Hydro-plants 1 and 3 in Norway (left and right)

the total hydro-capacity, we use the expression for the multiplier proxy

$$\pi^{\beta}(\xi) = \frac{Tx_1 + Wx_2(\xi) - h(\xi)}{\beta},$$

to estimate the gap by $\beta\pi^{\beta}(\xi)$. Since $\beta = 7000$ and the price signals are about 10^2, the magnitude of the violation incurred by RRH is of order 10^5. This is consistent with the graphs in Figure 5.8. The whole hydro-capacity being about 10^7, the gap is less then 1% for the whole system. The mean violation, in relative values, over the 30 scenarios and all the hydro-power plants, is shown in Figure 5.9 for the first 48 weeks.

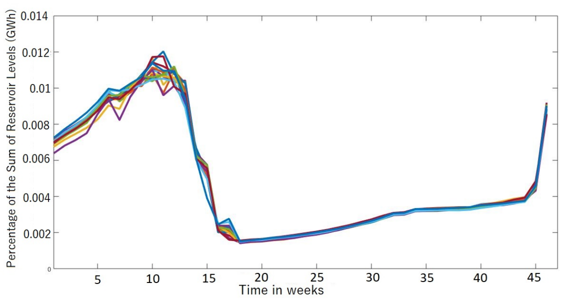
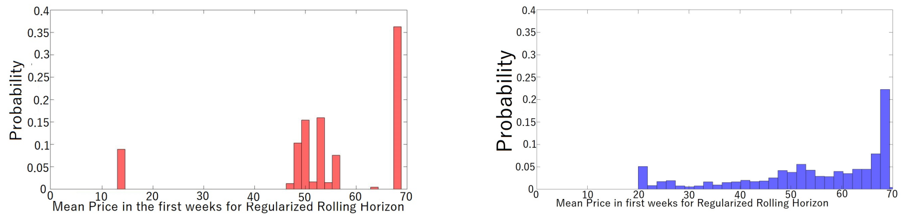


Figure 5.9: Relative mean violation of water balance in the whole system

5.4.4 Primal Simulation

The simulation over 200 out-of-sample scenarios compares the performance of RH and RRH in terms of distribution of the price signals in Figure 5.10. In this simulation we keep the fist stage decisions: $(v_j^7, v_j^{14}, \ldots, v_j^{7\times 52})$ that come from optimization part and simulate the cost-to-go function with new scenarios $(\mathcal{I}_j^t)$, $t \in A_w$, in the w-th week.

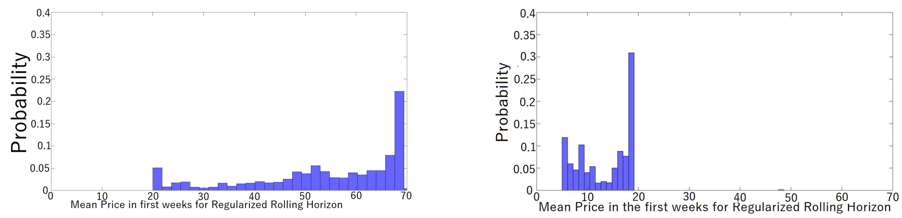
It is not possible to benchmark the output with SDDP because this (multistage) method lacks a primal policy. The comparison with SDDP is done in the next section, when assessing the dual policies.



RH RRH with $\beta = 7000$

Figure 5.10: Mean price signals of the first 25 weeks, simulated with primal policy of RH and RRH

Once again, we observe that the regularized price signal has a smooth distribution, and RRH is less susceptible than RH to variations of different samples. The right graph in Figure 5.10 is repeated on the left in Figure 5.11, to contrast the difference in RRH's price distribution when increasing the regularization parameter (on the right, $\beta = 100000$).



RRH with $\beta = 7000$ RRH with $\beta = 100000$

Figure 5.11: Mean price signals, simulated with RRH primal policy with two different values for β

In the right histogram in Figure 5.11 the shift to the left indicates a reduction in the price signals. This is in agreement with Table 5.1, reporting results from the optimization phase: with the primal policy, the higher β, the lower the reservoir levels. Table 5.2 shows the expected value and variance of the rolling-horizon variants.

	RH	RRH ($\beta = 7000$)	RRH ($\beta = 100000$)
Mean Value	61.22	54.93	13.8
Standard Deviation	16.2	14.31	11.34

Table 5.2: Price signal mean and deviation for one primal simulation

In order to evaluate the variability of the approaches under different 15 samples, in Table 5.3 we measure the variation of the distributions using the Wasserstein distance. The figures in Table 5.3 show a clear drop in the standard deviation for RRH, reflected also in the Wasserstein distance.

	RH	RRH	SDDP ($\beta = 300$)
Mean (Samples)	57.93	53.28	67.95
Standard Deviation (Samples)	38.48	16.38	16.40
Wasserstein Distance	28.16	5.64	15.78

Table 5.3: Price signal mean and deviation over different primal policies, first 25 weeks

Our final Figure 5.12, with the level of reservoir NO2 in Norway shows a typical behavior, observed for all the hydro-plants, with RH exhibiting a more erratic management of the water and keeping lower levels, when compared to RRH. The reason why we see apparently just one blue line is that all paths have close first-stage decisions.

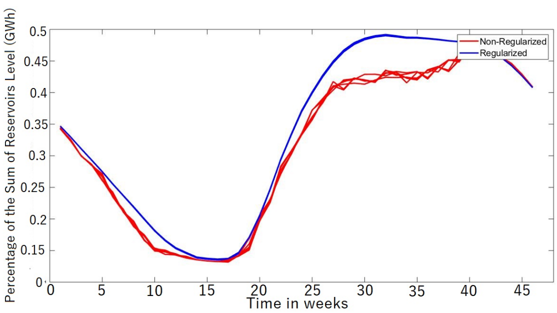
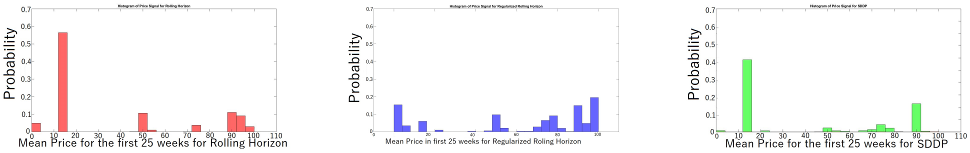


Figure 5.12: Water management of NO2 with primal simulation

5.4.5 Dual Simulation

As explained, by using the cuts for the future-cost functions obtained at the optimization phase, we can include SDDP in the comparisons.

Repeating for the dual simulation the calculation of price signal distributions done with the primal simulation in Figure 5.10 we obtain the output in Figure 5.13. We note that prices vary between 0 and 100 for all approaches, with both SDDP and RH concentrating prices mostly in two extreme values (about 15 and about 90).



RH RRH SDDP

Figure 5.13: Mean price signals of the first 25 weeks with dual simulation

The variability of the approaches under different samples, as in Table 5.3, but with the dual policy, is reported in Table 5.4 where we observe once again more stability for RRH. Contrary to SDDP, the regularization approach RRH was able to reduce the distance between histograms while keeping the expected value close to the one with RH.

	SDDP	RH	RRH ($\beta = 7000$)
Mean (Samples)	68.2	57.93	58.08
Standard Deviation (Samples)	34.5	38.48	30.1
Wasserstein Distance	6.11	5.64	3.46

Table 5.4: Price signal mean and deviation over different dual policies, first 25 weeks

We finish our analysis comparing in Figure 5.14 different paths of the first-stage primal decision that is, the level of reservoirs. Each path consists of a different sequence of scenarios. Each line represents a path and each color a different algorithm.

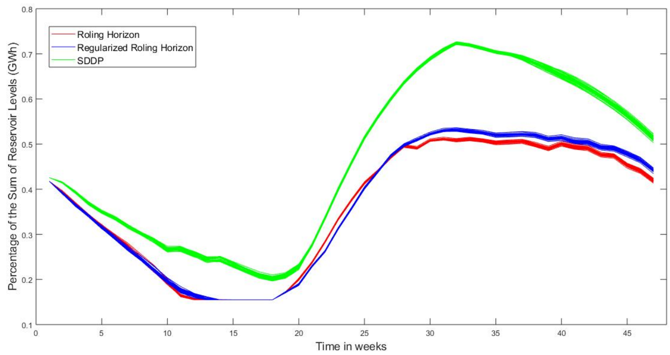


Figure 5.14: Reservoir dynamics

We observe a similar behavior as with the primal simulation for RH and RRH. For some paths, the curve of the reservoir level in SDDP is distant from the mean curve, as with SDDP prices vary the most. This is a consequence of the prudence of SDDP that, when confronted to a sequence of more favorable scenarios, is forced to a change, with respect to the initial conservative perspective.

Chapter 6

General Conclusions and Future Work

In many applications dual variables are an important output of the solving process, due to their role as price signals. When dual solutions are not unique, different solvers or different computers, even different runs in the same computer if the problem is stochastic, end up with different price indicators. Even though all of such values are correct, the fact that the obtained dual variable can vary among many possibilities makes unreliable any economic analysis based on marginal prices. We have presented an approach that yields reliable indicators, by providing the minimal-norm multiplier. Our computational experience, both proof-of-concept and on a real-life problem of ENGIE, shows the benefits of the methodology for two-stage stochastic linear programs.

The best choice for the penalization/regularization parameter β is clearly problem dependent. Somewhat similarly to the solution concept called *compromise decision* in [SL16], but adopting a dual point-of-view, the performance index proposed in (5.1) aims at measuring bias and variance in multiple replications of sampling-based approximations of two-stage stochastic programs. We observe empirically that our approach yields a significant reduction in the variance of the dual solutions (optimal Lagrange multipliers).

A topic of on-going research, not presented in this dissertation because it is not yet finalized, refers to how smoothness properties of the piecewise quadratic regularized recourse function translates into stability of the price signal. This analysis of quantitative stability is close in spirit to the one in [LRX14], but on the dual variables; see also [DR00], [Roe03b].

Bibliography

[AOS19] W. Ackooij, W. de Oliveira, and Y. Song. "On level regularization with normal solutions in decomposition methods for multistage stochastic programming problems". In: *Comput Optim Appl* 74.1 (2019), pp. 1–42.

[Ban+83] B. Bank, J.Guddat, D. Klatte, B. Kummer, and K. Tammer. *Non-Linear Parametric Opitmization.* Birkhauser Basel, 1983.

[Bis+17] A. Bischi, L. Taccari, E. Martelli, E. Amaldi, G. Manzolini, P. Silva, S. Campanari, and E. Macchi. "A rolling-horizon optimization algorithm for the long term operational scheduling of cogeneration systems". In: *Energy* (Dec. 2017).

[Bon+06] J. Bonnans, J. Gilbert, C. Lemaréchal, and C. Sagastizábal. *Numerical Optimization. Theoretical and Practical Aspects.* Universitext. Second edition, xiv+490 pp. Berlin: Springer-Verlag, 2006.

[Dea06] I. Déak. "Two-Stage Stochastic Problems with Correlated Normal Variables: Computational Experiences". In: *Annals of Operations Research* 142.1 (Feb. 2006), pp. 79–97.

[DR00] D. Dentcheva and W. Römisch. "Differential Stability of Two-Stage Stochastic Programs". In: *SIAM Journal on Optimization* 11.1 (2000), pp. 87–112.

[DM02] E. Dolan and J. Moré. "Benchmarking optimization software with performance profiles". In: *Mathematical Programming* 91 (2002), pp. 201–213.

[DLR03] A. L. Dontchev, A. S. Lewis, and R. T. Rockafellar. "The radius of metric regularity". In: *Transactions of the American Mathematical Society* 2 (2003), pp. 493–517.

[DCW00] J. Dupačová, G. Consigli, and S. W. Wallace. "Scenarios for Multistage Stochastic Programs". In: *Annals of Operations Research* 100.1 (Dec. 2000), pp. 25–53.

[DP09] J. Dupačová and J. Polivka. “Asset-liability management for czech pension funds using stochastic programming”. In: *J. Ann Oper Res* 165.1 (2009), pp. 5–28.

[EO75] M. J. Eisner and P. Olsen. “Duality for Stochastic Programming Interpreted as L. P. in Lp-Space”. In: *SIAM Journal on Applied Mathematics* 28.4 (1975), pp. 779–792.

[FHR13] B. Fhoula, A. Hajji, and M. Rekik. “Stochastic dual dynamic programming for transportation planning under demand uncertainty”. In: *2013 International Conference on Advanced Logistics and Transport* (2013), pp. 550–555.

[FM68] A. Fiacco and G. McCormick. *Nonlinear Programming: Sequential Unconstrained Minimization Techniques*. New York: John Wiley & Sons, Inc., 1968.

[FW56] M. Frank and P. Wolfe. “An algorithm for quadratic programming”. In: *Naval Research Logistics Quarterly* 3.1–2 (1956), pp. 95–110.

[GE07] A. Gibbs and F. Edward. “On choosing and bounding probability metrics”. In: *International Statistical Review* 70.3 (2007), pp. 419–435.

[GG04] V. Goel and I. E. Grossmann. “A stochastic programming approach to planning of offshore gas field developments under uncertainty in reserves”. In: *Computers & Chemical Engineering* 28.8 (2004), pp. 1409–1429.

[GS12] V. Guigues and C. Sagastizábal. “The value of rolling-horizon policies for risk-averse hydro-thermal planning”. In: *European Journal of Operational Research* 217.1 (2012), pp. 129–140.

[Hec+10] G. Hechme-Doukopoulos, S. Brignol-Charousset, J. Malick, and C. Lemaréchal. “The short-term electricity production management problem at EDF”. In: *Optima Newsletter - Mathematical Optimization Society* 84 (Oct. 2010), pp. 2–6.

[HBT18] L. Hellemo, P. I. Barton, and A. Tomasgard. “Decision-dependent probabilities in stochastic programs with recourse”. In: *Computational Management Science* 15.3 (Oct. 2018), pp. 369–395.

[HS94] J. L. Higle and S. Sen. “Finite master programs in regularized stochastic decomposition”. In: *Mathematical Programming* 67.1 (Oct. 1994), pp. 143–168.

[HL93] J.-B. Hiriart-Urruty and C. Lemaréchal. *Convex Analysis and Minimization Algorithms I*. Berlin Heidelberg: Springer-Verlag, 1993.

[HW01] K. Hoyland and S. Wallace. "Generating scenario trees for multistage decision problems". In: *Management Science* 47 (2001), pp. 295–307.

[IS05] A. Izmailov and M. Solodov. *Otimização Volume 1*. Rio de Janeiro: IMPA, 2005.

[IS14] A. F. Izmailov and M. V. Solodov. *Newton-type methods for optimization and variational problems*. Springer Series in Operations Research and Financial Engineering. Springer, Cham, 2014.

[KR92] A. J. King and R. T. Rockafellar. "Sensitivity analysis for nonsmooth generalized equations". In: *Mathematical Programming* 55.1 (Apr. 1992), pp. 193–212.

[Kum84] B. Kummer. "Generalized equations: Solvability and regularity". In: *Sensitivity, Stability and Para Analysis*. Berlin, Heidelberg: Springer Berlin Heidelberg, 1984, pp. 199–212.

[LSS19b] C. Lage, C. Sagastizábal, and M. Solodov. "Multiplier Stabilization Applied to Two-Stage Stochastic Programs". In: *Journal of Optimization Theory and Applications* 183.1 (Oct. 2019), pp. 158–178.

[LS97] C. Lemaréchal and C. Sagastizábal. "Variable metric bundle methods: from conceptual to implementable forms". In: *Mathematical Programming* 76 (1997), pp. 393–410.

[LRX14] Y. Liu, W. Römisch, and H. Xu. "Quantitative Stability Analysis of Stochastic Generalized Equations". In: *SIAM Journal on Optimization* 24.1 (2014), pp. 467–497.

[LXL11] Y. Liu, H. Xu, and G.-H. Lin. "Stability Analysis of Two-Stage Stochastic Mathematical Programs with Complementarity Constraints via NLP Regularization". In: *SIAM Journal on Optimization* 21 (2011), pp. 669–705.

[MMF17] V. L. de Matos, D. P. Morton, and E. C. Finardi. "Assessing policy quality in a multistage stochastic program for long-term hydrothermal scheduling". In: *Ann Oper Res* (2017), pp. 253–713.

[NW06] J. Nocedal and S. Wright. *Numerical Optimization*. Springer, New York, 2006.

[OSS11] W. Oliveira, C. Sagastizábal, and S. Scheimberg. "Inexact Bundle Methods for Two-Stage Stochastic Programming". In: *SIAM Journal on Optimization* 21 (2 2011), pp. 517–544.

[Per+05] M. V. Pereira, S. Granville, M. H. C. Fampa, R. Dix, and L. A. Barroso. "Strategic bidding under uncertainty: a binary expansion approach". In: *IEEE Transactions on Power Systems* 20.1 (Feb. 2005), pp. 180–188.

[PP91] M. V. F. Pereira and L. M. V. G. Pinto. "Multi-stage stochastic optimization applied to energy planning". In: *Mathematical Programming* 52.1 (May 1991), pp. 359–375.

[PR07] G. C. Pflug and W. Römisch. *Modeling, Measuring and Managing Risk*. World Scientific Publishing Co. Pte. Ltd., 2007.

[RHS10] G. P. Ribas, S. Hamacher, and A. Street. "Optimization under uncertainty of the integrated oil supply chain using stochastic and robust programming". In: *International Transactions in Operational Research* 17.6 (2010), pp. 777–796.

[RW98] R. Rockafellar and R. J.-B. Wets. *Variational Analysis*. Heidelberg, Berlin, New York: Springer Verlag, 1998.

[Roe03b] W. Römisch. "Stability of Stochastic Programming Problems". In: *Stochastic Programming*. Vol. 10. Handbooks in Operations Research and Management Science. Elsevier, 2003, pp. 483–554.

[SL16] S. Sen and Y. Liu. "Mitigating Uncertainty via Compromise Decisions in Two-Stage Stochastic Linear Programming: Variance Reduction". In: *Operations Research* 64.6 (2016), pp. 1422–1437.

[SH98] A. Shapiro and T. Homem-de-Mello. "A simulation-based approach to two-stage stochastic programming with recourse". In: *Math. Program.* 81.1 (1998), pp. 301–325.

[Sha11] A. Shapiro. "Analysis of stochastic dual dynamic programming method". In: *European Journal of Operational Research* 209.1 (2011), pp. 63–72.

[Sol10] M. V. Solodov. "Constraint qualifications". In: *Wiley Encyclopedia of Operations Research and Management Science*. Ed. by J. J. Cochran et al. John Wiley & Sons, Inc., 2010.

[SBV17] A. Street, A. Brigatto, and D. M. Valladão. "Co-Optimization of Energy and Ancillary Services for Hydrothermal Operation Planning Under a General Security Criterion". In: *IEEE Transactions on Power Systems* 32.6 (Nov. 2017), pp. 4914–4923.

[ZM13] S. Zaourar and J. Malick. "Prices stabilization for inexact unit-commitment problems". In: *Mathematical Methods of Operations Research* 78.3 (Dec. 2013), pp. 341–359.