Multiplier Stabilization Applied to Two-Stage Stochastic Programs

Clara Lage* Claudia Sagastizábal† Mikhail Solodov‡

November 4, 2018

Abstract

In many mathematical optimization 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, often end up with different optimal multipliers. From the perspective of a decision maker, this variability makes the price signals less reliable and, hence, less useful. We address this issue for a particular family of linear and quadratic programs by proposing a solution procedure that, among all possible optimal multipliers, systematically yields the one with the smallest norm. The approach, based on penalization techniques of nonlinear programming, amounts to a regularization in the dual of the original problem. As the penalty parameter tends to zero, convergence of the primal sequence and, more critically, of the dual is shown under natural assumptions. The methodology is illustrated on a battery of two-stage stochastic linear programs.

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

Keywords: multiplier stability, dual regularization, penalty method, stochastic programming, two-stage, empirical approximations

*ENGIE, 1 pl. S. de Champlain, 92930 Paris La Défense, France. CIFRE contract with Université de Paris Sorbonne, France; joint PhD with IMPA – Instituto de Matemática Pura e Aplicada, Brazil. Email: clara.lage@engie.com.

†IMECC - UNICAMP, Rua Sergio Buarque de Holanda, 651, 13083-859, Campinas, SP, Brazil. Email: sagastiz@unicamp.br. Adjunct Researcher. Partly supported by CNPq Grant 303905/2015-8 and by FAPERJ.

‡IMPA – Instituto de Matemática Pura e Aplicada, Estrada Dona Castorina 110, Jardim Botânico, Rio de Janeiro, RJ 22460-320, Brazil. Email: solodov@impa.br. Research of this author is supported in part by CNPq Grant 303724/2015-3 and by FAPERJ Grant 203.052/2016.

1 Introduction and motivation

Parametric optimality studies how the solution set of a nonlinear programming problem (NLP) behaves when subject to perturbations in the objective function and/or in the constraints. For the latter, it is possible to identify circumstances in which solutions vary in a Lipschitzian-like manner with respect to right-hand side perturbations.

Studies of this type are mostly concerned with optimal values and with *primal* solution sets. In this work, we are more interested in understanding how the *dual* solutions to certain NLPs behave when the constraints' right-hand side varies. Our motivation stems from the fact that Lagrange multipliers give the rates of change of the optimal value with respect to such perturbations. Each multiplier signals the marginal effect of raising or lowering the value of the corresponding constraint.

The interpretation of Lagrange multipliers as marginal costs (or shadow prices in the parlance of Linear Programming) has plenty of useful applications. When the Lagrange multiplier vector is not unique, the price components are related to rates of change of the subderivatives of the optimal value function. If there is a full set of prices attached to certain perturbation parameter, it is natural to ask the following important question:

Is it possible to devise a solution methodology that provides the minimal-norm multiplier?

The economic interest of such question is clear, since the mechanism would systematically yield the smallest possible price.

We provide an answer to this question for a particular family of linear and quadratic programs. To define and justify a solution procedure that, among all possible optimal multipliers, provides the smallest one, we combine some variational analysis considerations and penalization techniques of nonlinear programming.

The approach is illustrated on two-stage stochastic linear programming problems:

$$\left\{\begin{array}{lll} \min & \langle c, x_1\rangle + \mathbb{E}\left[\langle q, x_2(\xi)\rangle\right] & \\ \text{s.t.} & x_1 \geq 0\,, x_2(\xi) \geq 0 & \text{a.e. } \xi \\ & Tx_1 + Wx_2(\xi) = h & \text{a.e. } \xi \qquad (\leftrightarrow \pi(\xi))\,, \end{array}\right. \tag{1.1}$$

where $\xi = (q, h, T, W)$ represents the data uncertainty and the involved vectors and matrices have adequate dimensions specified in Section 4 below. In this modelling paradigm, decisions are taken independently of future observations, on the basis of the uncertainty realization, which reveals all at once; see [SDR09, Chapter 2]. The continuous probability distribution of ξ is often approximated by scenarios generated by Monte-Carlo simulation, as in the sample average approximation approach [KSH02].

Even though two-stage formulations are widely used, the paradigm remains computationally challenging in applications and has given rise to a vast literature on dedicated numerical solvers, most notably related with decomposition methods [SW69], [BL88], [Rus99], and more recently [ZPR00], [OSS11], [Fáb+15], [AM16].

Two-stage models are well suited to situations in which the output of interest is the (deterministic) first-stage solution. Another useful output is the expected value (of some components) of the multiplier in the affine constraint, denoted in (1.1) above by $\pi(\xi)$. In data-driven decision-making the corresponding mean multiplier gives a price signal that helps defining business strategies. Having this goal in mind, the fact that different samples can make the price signal vary wildly is a serious handicap. Our proposal addresses this issue by systematically providing the minimal-norm price signal, thus making the indicator more reliable for the decision maker.

The rest of the paper is organized as follows. In Section 2, we support our computational approach to building multiplier estimates converging to minimal-norm optimal ones by some variational analysis considerations. We also introduce an apparently new condition which allows to prove boundedness of the set of Lagrange multipliers associated with some part of the constraints of the problem, while allowing the other multipliers to be unbounded. In Section 3, we prove convergence of the proposed multiplier estimates (obtained from an exterior penalty scheme) to the ones of minimal norm, for linear or quadratic programs satisfying some natural assumptions. In Section 4 we explain how the methodology can be applied to two-stage stochastic linear programming problems, and present numerical results that confirm the interest of the approach. The work ends with some concluding remarks.

Notation We mostly follow the notation in [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., $\partial^\infty f(x) = N_{\mathrm{dom}\, f}(x)$; see [RW98, Proposition 8.12].

2 A Variational Analysis perspective

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

$$\left\{ \begin{array}{lll} \min & f(x) & \\ \text{s.t.} & x \in X & \\ & Ax - b = 0\,, & \end{array} \right. \quad \text{where} \quad \begin{array}{l} f(x) \text{ is finite valued, convex, and } C^1 \\ X \text{ is a closed convex set, and} \\ b \in \{ y \mid y = Ax,\ x \in X \}\,. \end{array} \tag{2.1}$$

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). Then, it is well known that uniqueness of Lagrange multipliers associated to a solution $\bar{x}$ of problem (2.1) is implied by the linear independence of gradients of the constraints active at $\bar{x}$, and is equivalent to the so-called strict Mangasarian-Fromovitz condition (see, e.g., [Sol10], [IS14, Sections 1.1, 1.2.4]). Note that in (2.1) either of these assumptions subsumes that the matrix A is of full rank. There are many important applications for which this rarely holds in practice. The less stringent Mangasarian-Fromovitz constraint qualification (MFCQ) is equivalent to having a nonempty compact set of Lagrange multipliers. However, MFCQ still subsumes that A has full rank.

Thus, if A is not of full rank, the multipliers associated to the equality constraint in (2.1) 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 minimum norm. The idea is to consider a sequence of problems that penalize the equality constraint in (2.1), depending on a parameter $\beta > 0$. Given a (primal) solution to the penalized problem, we then construct an explicit multiplier proxy/estimate, which we denote by π^β. Specifically, we solve

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

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

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

For the case when (2.1) is a linear or quadratic program, 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 3.

We note that approximating Lagrange multipliers in the setting of quadratic penalty methods by π^β is certainly not a new idea; see, e.g., [NW06, Chapter 17.1]. However, in the literature convergence of π^β is 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 3.2 below. In Section 3, 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 somewhat different motivation and insight for the multiplier estimates π^β, by specializing some results of Variational Analysis [RW98] to our setting. To this aim, 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:

$$\mathbb{R}^m \ni v \hookleftarrow \theta^\beta(v) := \sup_{y \in \mathbb{R}^m} \left\{ \langle v, y \rangle - \frac{1}{2}\beta\|y\|^2 \right\} = \begin{cases} \dfrac{1}{2\beta}\|v\|^2 & \text{if } \beta > 0 \\ \delta_{\{0\}}(v) & \text{if } \beta = 0\,. \end{cases} \tag{2.2}$$

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

$$\text{if } \beta > 0, \quad \text{for all } v \in \mathbb{R}^m = \operatorname{dom}\theta^\beta\,, \;\; \partial\theta^\beta(v) = \left\{\tfrac{1}{\beta}v\right\} \text{ and } \partial^\infty\theta^\beta(v) = \{0\} \tag{2.3}$$

while

$$\text{if } \beta = 0, \quad \text{for } v = 0 = \operatorname{dom}\theta^0\,, \;\; \partial\theta^0(v) = \mathbb{R}^m \text{ and } \partial^\infty\theta^0(v) = \mathbb{R}^m\,. \tag{2.4}$$

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

$$\min_{x \in \mathbb{R}^n} f^\beta(x, u) := f(x) + \delta_X(x) + \theta^\beta(Ax - b + u)\,, \tag{2.5}$$

noting that writing (2.5) with $\beta = 0$ and $u = 0$ yields our original problem (2.1).

When $\beta > 0$, some x^β is optimal in (2.5) if and only if

$$x^\beta \in X,\ \mu^\beta \in N_X(x^\beta), \quad \nabla f(x^\beta) + \mu^\beta + A^\top \pi^\beta = 0\,, \tag{2.6}$$

where, for $\bar{u} \in \mathbb{R}^m$ given,

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

is the unique *extended Lagrange multiplier* in [RW98].

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 (2.1) means that

$$0 = A^\top \pi + \mu, \quad \mu \in N_X(\bar{x}) \quad \Rightarrow \quad \pi = 0 \in \mathbb{R}^m \text{ and } \mu = 0 \in \mathbb{R}^n\,. \tag{2.7}$$

If $\bar{x}$ satisfies MFCQ (2.7), there exists a classical Lagrange multiplier $\bar{\pi} \in \mathbb{R}^m$, not necessarily unique, satisfying (2.6) written with $\beta = 0$ and $(x^\beta, \pi^\beta, \mu^\beta) = (\bar{x}, \bar{\pi}, \bar{\mu})$. Condition (2.6) is also sufficient for $\bar{x}$ to be optimal for (2.5) written with $\beta = 0$ and $\bar{u} := b - A\bar{x}$.

The *parametric version of Fermat rule* in [RW98, Example 10.12], analyzing (2.5) from a Variational Analysis perspective, condenses the key ingredients relating extended Lagrange multipliers to the marginal rate of change of the optimal value in (2.1), when considered as a function of the right-hand side perturbation of the affine constraint.

Proposition 2.1 (Perturbation functions and extended Lagrange multipliers)**.** *Associated to* (2.1)*, consider the parametric optimization problems* (2.5) *with penalties* (2.2)*, 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 $\bar{u} \in \operatorname{dom} p^\beta = \mathbb{R}^m$, with gradient $\nabla p^\beta(\bar{u}) = \pi^\beta$.

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

$$\partial p^0(\bar{u}) = \{\bar{\pi} \in \mathbb{R}^m : \text{ (2.6) } \textit{holds written with } (x^\beta, \pi^\beta, \mu^\beta) = (\bar{x}, \bar{\pi}, \bar{\mu})\}\,.$$

Proof. The perturbed function $f^\beta(x, u)$ is convex on (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

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

for any $x^\beta \in P^\beta(\bar{u})$ and $\bar{u} \in \operatorname{dom} p^\beta = \operatorname{dom} \theta^\beta$. Together with (2.3) and (2.4), this gives $\partial^\infty p^\beta(\bar{u}) = Y^\infty(\bar{u}) = \partial^\infty \theta^\beta(Ax^\beta - \bar{u})$, and

$$\partial p^\beta(\bar{u}) = Y(\bar{u}) = \left\{\pi^\beta : (x^\beta, \mu^\beta, \pi^\beta) \text{ satisfies (2.6)}\right\} ,$$

as claimed. □

The above characterization of π^β, obtained from the penalty scheme as extended Lagrange multiplier, motivates from the Variational Analysis point of view the choice of π^β as the multiplier estimates. In Section 3, 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 the MFCQ condition (2.7) and the fact that it is equivalent to the set of multipliers being nonempty and bounded. Consider the following condition at $\bar{x}$ feasible in (2.1).

$$0 = A^\top \pi + \mu, \quad \mu \in N_X(\bar{x}) \quad \Rightarrow \quad \mu = 0. \tag{2.8}$$

Clearly, (2.8) is a weaker condition than (2.7). In particular, as we show in Theorem 2.1 below, (2.8) 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 (2.8) is *not* a constraint qualification, i.e., it does not imply (by itself) that for a solution $\bar{x}$ of problem (2.1) the multiplier set is nonempty. An alternative, equivalent, formulation of condition (2.8) is

$$\operatorname{Im} A^\top \cap N_X(\bar{x}) = \{0\}. \tag{2.9}$$

Theorem 2.1 (On boundedness of multipliers)**.** *Let $\bar{x}$ be any feasible point in* (2.1)*. Then the following statements are equivalent:*

(i) *Condition* (2.8) *holds at $\bar{x}$.*

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

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

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

$$\bar{g} + A^\top \pi^k + \mu^k = 0, \quad \mu^k \in N_X(\bar{x}), \tag{2.10}$$

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 (2.10) 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 (2.8).

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 $t > 0$,

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

As $\tilde{\mu} \neq 0$, it follows that $\|\bar{\mu} + t\tilde{\mu}\| \to +\infty$ as $t \to +\infty$, i.e., the set $S_{\bar{g}}$ is unbounded. $\square$

We emphasize that, being weaker than MFCQ, condition (2.8) is certainly not restrictive (assuming that the existence of Lagrange multipliers is given or follows from some other considerations).

3 A Nonlinear Programming computational perspective

Consider now the following *quadratic programming* problem:

$$\left\{\begin{array}{lll} \min & f(x) & \\ \text{s.t.} & x \in X & \\ & Ax - b = 0\,, & \end{array}\right. \quad \begin{array}{ll} & f(x) := \langle g, x\rangle + \frac{1}{2}\langle x, Hx\rangle \\ \text{where} & X := \{x \in \mathbb{R}^n \mid x \geq 0\}\,, \text{ and} \\ & b \in \{y \mid y = Ax,\ x \in X\}\,, \end{array} \tag{3.1}$$

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 (3.1) is a particular instance of (2.1), and we can use the constructs in Section 2 for some motivations. In that case, fixing $\bar{u} = 0$, for $\beta > 0$ from (2.2) and (2.5) we have

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

and Proposition 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 (3.1):

$$\begin{cases} \min & f(x) + \frac{1}{2\beta}\|Ax - b\|^2 \\ \text{s.t.} & x \in X\,. \end{cases} \tag{3.2}$$

The multiplier estimate is then given by

$$\pi^\beta = \frac{Ax^\beta - b}{\beta}. \tag{3.3}$$

The usual penalty method (see, e.g., [FM68]) solves subproblems (3.2) above for a sequence of decreasing penalty parameters $0 < \beta_{k+1} < \beta_k$, tending to zero. We want to study how the multiplier estimates (3.3) for the equality constraints in (3.1) behave along the sequence of solving the penalized (sub)problems (3.2). In what follows, we shall show that, under reasonable assumptions, the generalized multipliers π^β converge to the minimal-norm multiplier $\hat{\pi}$; see (3.12) below for a formal definition.

We start with some standard facts on (primal) convergence of penalty methods [FM68]. In particular, they do not depend on our specific setting of (3.1), 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 (3.2).

Theorem 3.1 (Primal convergence of generic penalty methods)**.** *If* $0 < \beta_{k+1} < \beta_k$ *and* x^k *is a (global) solution of* (3.2) *for* $\beta = \beta_k$ *for each* k*, then*

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

If, in addition, $\beta_k \to 0$ *as* $k \to \infty$ *and the optimal value of problem* (3.1) *is finite, then every accumulation point of* $\{x^k\}$ *is a (global) solution of* (3.1). □

Note that this result refers to *global* solutions of subproblems. This is standard, and also not an issue when the problem (3.1) 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 (3.1) has a solution, by the Frank–Wolfe Theorem [FW56].

However, it is important to emphasize that the general convergence result in Theorem 3.1 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, therefore, will be to prove when the generated sequence $\{x^k\}$ is guaranteed to be 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 (3.2) for $\beta = \beta_k$. Define

$$\pi^k := \frac{1}{\beta_k}(Ax^k - b). \tag{3.4}$$

The assertion below is standard; see, e.g., [NW06, Theorem 17.2]. As Theorem 3.1 above, it does not depend on the setting of problem (3.1). I.e., it 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 3.2 (Dual convergence of the quadratic penalty method)**.** *In* (3.1)*, let $X = \mathbb{R}^n$. Let $\bar{x}$ be any accumulation point of $\{x^k\}$, where x^k is a solution of* (3.2) *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* (3.1) *this means that A has full rank).*

Then $\bar{x}$ is a stationary point of (3.1) *and the subsequence $\{\pi^{k_j}\}$ defined by* (3.4) *converges to the unique Lagrange multiplier $\bar{\pi}$ associated to $\bar{x}$.* □

Note, however, that Theorem 3.2 again implicilty 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 critical cone of (3.1) at a given stationary point $\bar{x}$ is defined by

$$K(\bar{x}) := \operatorname{Ker} A \cap \{d \in \mathbb{R}^n \mid \langle H\bar{x} + g, d\rangle \leq 0,\ d_i \geq 0 \text{ for } i \text{ s.t. } \bar{x}_i = 0\}. \tag{3.5}$$

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

$$\langle Hd, d\rangle > 0 \quad \text{for all } d \in K(\bar{x}) \setminus \{0\}. \tag{3.6}$$

When H is positive semidefinite, the solution set of problem (3.1) is convex. Since (3.6) implies that $\bar{x}$ is a strict (thus isolated) minimizer, it means that in the convex case it must be unique. In particular, when f is linear, i.e., $H = 0$, condition (3.6) can only hold when $K(\bar{x}) = \{0\}$. It can be seen that this means that $\langle g, d\rangle > 0$ for all feasible directions at $\bar{x}$. This is equivalent to saying that $\bar{x}$ is the unique solution of the linear program (3.1).

Note also that since $K(\bar{x}) \subset \operatorname{Ker} A$, the following is also a second-order sufficient optimality condition (as it implies (3.6)):

$$\langle Hd, d\rangle > 0 \quad \text{for all } d \in \operatorname{Ker} A \setminus \{0\}. \tag{3.7}$$

However, unlike (3.6), condition (3.7) is an assumption on H and A which does not depend on $\bar{x}$. Note that (3.7) does not require H to be positive semidefinite, and thus the objective function f in (3.1) can be non-convex.

Theorem 3.3 (Conditions for primal convergence)**.** *Suppose that one of the following two items holds:*

1. *Condition* (3.7) *is satisfied.*
2. *Matrix H is positive semidefinite and condition* (3.6) *holds for the solution $\bar{x}$ of* (3.1).

Then for any *sequence of parameters $\beta_k \to 0$ (even not necessarily monotone), any sequence $\{x^k\}$ generated by the penalty scheme* (3.2) *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 (3.1)*. 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 $\|x^k\| \to \infty$. 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 (3.2), it holds that

$$Hx^k + g + \frac{1}{\beta_k}A^\top(Ax^k - b) - \mu^k = 0, \quad x^k \geq 0,\ \mu^k \geq 0,\ \left\langle \mu^k, x^k \right\rangle = 0. \tag{3.8}$$

Note that $\left\langle \mu^k, z^k \right\rangle = 0$. 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 (3.1). Since $\tilde{x} \in X$, $A\tilde{x} - b = 0$, and x^k is a solution of (3.2), it holds that

$$f(\tilde{x}) \geq f(x^k) + \frac{1}{2\beta_k}\|Ax^k - b\|^2 \geq f(x^k). \tag{3.9}$$

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

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

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

$$0 \geq \frac{1}{2}\left\langle Hz, z \right\rangle . \tag{3.10}$$

Since $0 \neq z \in \operatorname{Ker} A$, this immediately gives a contradiction if the condition (3.7) holds.

Suppose now H is positive semidefinite and (3.6) holds for the solution $\bar{x}$ of (3.1), which is unique in this case. Since $\bar{x}$ is in particular feasible, from (3.9) 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 (3.5). Now (3.10) again gives a contradiction with (3.6).

We conclude that $\{x^k\}$ is bounded. The other assertions follow from the general results about penalty methods in Theorem 3.1 (and other considerations above). $\square$

Having established when there is primal convergence of solutions of the penalized subproblems (3.2), we now analyze the asymptotic behavior of the dual sequence $\{\pi^k\}$ defined by (3.4).

Recall that for a solution $\bar{x}$ of problem (3.1) the set of associated Lagrange multipliers (π, μ) is characterized by the following system:

$$H\bar{x} + g + A^\top \pi - \mu = 0, \quad \bar{x} \geq 0, \ \mu \geq 0, \ \langle \mu, \bar{x} \rangle = 0. \tag{3.11}$$

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 (3.11) for the given $\bar{x}$ and $\bar{\mu}$, i.e., the (unique) solution of

$$\min \ \frac{1}{2}\|\pi\|^2 \quad \text{s.t.} \quad H\bar{x} + g + A^\top \pi - \bar{\mu} = 0. \tag{3.12}$$

We have the following.

Theorem 3.4 (Convergence of the multipliers estimates)**.** *Let $\beta_{k+1} < \beta_k$ for all k, and $\beta_k \to 0$. Let the assumptions of Theorem 3.3 hold (i.e., one of the two items thereby). Let $\bar{x}$ be any accumulation point of the sequence $\{x^k\}$ (which is bounded by Theorem 3.3), $x^{k_j} \to \bar{x}$ as $j \to \infty$. Let condition* (2.8) *hold at $\bar{x}$.*

Then the sequence $\{\mu^{k_j}\}$ is bounded. Moreover, for any accumulation point $\bar{\mu}$ of $\{\mu^{k_j}\}$, the corresponding subsequence $\{\pi^{k_j}\}$, defined by (3.4)*, converges to $\hat{\pi}$, the minimal-norm solution of* (3.12)*. The point $(\bar{x}, \hat{\pi}, \bar{\mu})$ is a primal-dual solution of* (3.1)*.*

Proof. Under the assumptions of Theorem 3.3, it follows that $\{x^k\}$ is bounded. Recalling the subproblem KKT conditions (3.8) and using the definition (3.4) of π^k, we have that

$$Hx^k + g + A^\top \pi^k - \mu^k = 0, \quad x^k \geq 0, \ \mu^k \geq 0, \ \left\langle \mu^k, x^k \right\rangle = 0. \tag{3.13}$$

Let $\{x^{k_j}\} \to \bar{x}$ as $j \to \infty$. We first prove that the sequence $\{\mu^{k_j}\}$ is bounded. The argument is similar to the first part of the proof of Theorem 2.1. Suppose it is not, i.e., (3.13) holds with $\|\mu^{k_j}\| \to +\infty$ (possibly passing onto a subsequence). We can assume, passing onto a further subsequence if necessary, that

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

Denote $u^{k_j} = A^\top \pi^{k_j}/\|\mu^{k_j}\| \in \operatorname{Im} A^\top$. Dividing the equality in (3.13) 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 (3.13) 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 (2.8).

Once $\{\mu^{k_j}\}$ is bounded, the first equality in (3.13) implies that $\{A^\top \pi^{k_j}\}$ is bounded as well. Passing onto a further subsequence if necessary, let $\{x^{k_j}\} \to \bar{x}$, $\{\mu^{k_j}\} \to \bar{\mu}$, $\{A^\top \pi^{k_j}\} \to a$ 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 the subspace is closed, it holds that $a \in \operatorname{Im} A^\top A$, i.e., $a = A^\top \bar{\pi}$ for some $\bar{\pi} \in \operatorname{Im} A$.

Passing onto the limit in (3.13) as $j \to \infty$, we then have that

$$H\bar{x} + g + A^\top \bar{\pi} - \bar{\mu} = 0,\ \bar{\pi} \in \operatorname{Im} A, \quad \bar{x} \geq 0,\ \bar{\mu} \geq 0,\ \langle \bar{\mu}, \bar{x} \rangle = 0. \tag{3.14}$$

We next show that there exists only one $\bar{\pi} \in \operatorname{Im} A$ which satisfies the left equality in (3.14) 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 (3.14), 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 (3.12) 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 (3.12) 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

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

(To see this, assume the contrary, i.e., exists $\{u^k\}$ such that $Au^k \neq 0$ and $\|A^\top A u^k\|/\|Au^k\| \to 0$. Passing onto a 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 $(\pi^{k_j} - \hat{\pi}) \in \operatorname{Im} A$, for each j there exists some b^j such that $\pi^{k_j} - \hat{\pi} = Ab^j$. Then,

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

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

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 section 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, *via* the so-called recourse functions; see (4.3) below.

4 Application to Two-stage Stochastic Linear Programming

For a random variable ξ with finite support, we consider a two-stage stochastic linear program with complete random recourse where uncertainty appears only in the second-stage costs and the right-hand side vector. I.e., in (1.1) we have $(q, h) = (q(\xi), h(\xi))$, which leads to

$$\left\{ \begin{array}{lll} \min & f(x) & \\ \text{s.t.} & x_1 \geq 0\,, x_2(\xi) \geq 0 & \text{a.e. } \xi \\ & Tx_1 + Wx_2(\xi) = h(\xi) & \text{a.e. } \xi\,, \end{array} \right. \quad \text{where} \quad f(x) := \langle c, x_1 \rangle + \mathbb{E}\left[\langle q(\xi), x_2(\xi) \rangle\right] .$$

In this problem $x_i \in \mathbb{R}^{n_i}$ for $i = 1, 2$, the right-hand side $h(\cdot) \in \mathbb{R}^m$, and the matrices T and W have orders $n_1 \times m$ and $n_2 \times m$, respectively. Realizations $\{\omega^1, \ldots, \omega^S\}$, with respective probability $p^1, \ldots, p^S$, define the scenarios $\xi^s := \xi(\omega^s)$ for $s = 1, \ldots, S$, and yield the following linear programming problem:

$$\left\{ \begin{array}{lll} \min & \langle c, x_1 \rangle + \displaystyle\sum_{s=1}^{S} p^s \langle q^s, x_2^s \rangle & \\ \text{s.t.} & x_1 \geq 0\,, x_2^s \geq 0 & \text{for } s = 1, \ldots, S \\ & Tx_1 + Wx_2^s = h^s & \text{for } s = 1, \ldots, S\,. \end{array} \right. \tag{4.1}$$

Decomposition by scenarios is achieved by introducing the recourse functions

$$\mathbb{Q}(x_1; \xi^s) := \left\{ \begin{array}{ll} \min & \langle q^s, x_2 \rangle \\ \text{s.t.} & x_2 \geq 0 \\ & Wx_2 = h^s - Tx_1 \end{array} \right. = \left\{ \begin{array}{ll} \max & \langle \pi, h^s - Tx_1 \rangle \\ & \\ \text{s.t.} & W^\top \pi \leq q^s\,. \end{array} \right. \tag{4.2}$$

and writing (4.1) in the equivalent two-level formulation

$$\left\{ \begin{array}{ll} \min & \langle c, x_1 \rangle + \displaystyle\sum_{s=1}^{S} p^s\, \mathbb{Q}(x_1; \xi^s) \\ \text{s.t.} & x_1 \geq 0\,. \end{array} \right. \tag{4.3}$$

We note, in passing, that the assumption of complete recourse ensures the recourse functions $\mathbb{Q}(\cdot;\xi^s)$ are well-defined for each scenario s. The two equivalent formulations given for such recourse functions in (4.2) correspond to a primal and dual views (left and right, respectively). The dual view, in particular, motivated our proposal. Specifically, to "control" the multipliers, a sensible strategy would be to add a regularizing term to the objective function of the dual in (4.2). I.e,

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

$$\text{consider} \quad \mathbb{Q}^\beta(x_1;\xi^s) \;=\; \begin{cases} \max & \langle \pi, h^s - Tx_1\rangle - \frac{\beta}{2}\|\pi\|^2 \\ \text{s.t.} & W^\top \pi \leq q^s\,. \end{cases}$$

The quadratic term, making solution of the dual problem (optimal multiplier of the primal) unique, helps preventing oscillations, and somehow stabilizes the output of the overall process. The relation with our initial setting (3.1) is explained below.

4.1 The problem to be solved

To cast (4.1) as a particular instance of (3.1), it suffices to define the vectors

$$x := (x_1, x_2^1, \ldots, x_2^S) \in \mathbb{R}^{n_1+n_2S}\,, g := (c, p^1q^1, \ldots, p^Sq^S) \in \mathbb{R}^{n_1+n_2S}\,, b := (h^1, \ldots, h^S) \in \mathbb{R}^{mS}\,,$$

as well as the matrices $H = 0 \in \mathbb{R}^{n_1+n_2S} \times \mathbb{R}^{n_1+n_2S}$, and

$$A := \begin{bmatrix} T & W & 0 & \ldots & 0 \\ T & 0 & W & \ddots & \vdots \\ T & \vdots & \ddots & \ddots & 0 \\ T & 0 & \ldots & 0 & W \end{bmatrix} \in \mathbb{R}^{mS} \times \mathbb{R}^{n_1+n_2S}\,.$$

Given $\beta > 0$, it is not difficult to derive the penalized subproblems (3.2) for (4.1):

$$\begin{cases} \min & \langle c, x_1\rangle + \sum_{s=1}^{S} p^s \left\{ \langle q^s, x_2^s\rangle + \dfrac{1}{2\beta}\|h^s - Tx_1 - Wx_2^s\|^2 \right\} \\ \text{s.t.} & x_1 \geq 0\,, x_2^s \geq 0 \quad \text{for } s = 1, \ldots, S\,. \end{cases}$$

Regarding the two-level formulation, the penalization above amounts to replacing $\mathbb{Q}$ in (4.3) by the following recourse function

$$\mathbb{Q}^\beta(x_1;\xi^s) := \begin{cases} \min & \langle q^s, x_2\rangle + \frac{1}{2\beta}\|h^s - Tx_1 - Wx_2\|^2 \\ \text{s.t.} & x_2 \geq 0\,, \end{cases} \tag{4.4}$$

whose dual formulation is

$$\mathbb{Q}^\beta(x_1;\xi^s) = \begin{cases} \max & \langle \pi, h^s - Tx_1\rangle - \frac{\beta}{2}\|\pi\|^2 \\ \text{s.t.} & W^\top \pi \le q^s\,. \end{cases}$$

These primal and dual views for the second-stage problem highlight the double role of the β-term: a penalization in the primal becomes a regularization in the dual.

4.2 Numerical Results

The fact that the dual second-stage problem is perturbed by a term " $-\beta\|\pi\|^2$" evidently changes the dual and primal solutions, when compared to the original problem (4.1). Our goal is to keep close the original marginal cost, and at the same time, decrease its variance.

Theorem 3.4 describes theoretically what we should expect as behavior of the mean value of regularized price signals in terms of the original optimization problem. In this section we give numerical examples illustrating the main features of our approach.

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, we measured the variance and the mean value of regularized and non-regularized stochastic problems. In addition, in (4.6) below we created a single index that measures the joint dynamics of variance reduction and distance to the original dual solution set, as β goes to zero. The index is used to make the performance profile in Subsection 4.2.2, when running the methodology on a battery of problems in the literature.

4.2.1 Price Signal Analysis on an Illustrative Example

We first consider a simple instance that can be solved analytically for checking, and which illustrates well our theoretical results (the satisfaction of condition (2.8) and convergence to the minimal-norm price (3.12)). Take $S = 2$ equiprobable scenarios and let $n_1 = n_2 = 2$. The first-stage cost $c \in \mathbb{R}^2$ and second-stage costs are deterministic $q^1 = q^2 = q \in \mathbb{R}^2$. The technology and recourse matrices in (4.1) 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 $h^1 := (1,1,1)^\top$ and $h^2 := (1,0,3)^\top$.

Working out the algebra, it can be seen that the feasible set in (4.1) is completely determined by the first component of x_1, denoted by $y \ge 0$ below. Specifically,

$$x \text{ is feasible in (4.1) if and only if, for some } y \ge 0\,,$$
$$x_1 := \left(y, \tfrac{3}{4}(1+y)\right)^\top, \quad x_2^1 := \left(\frac{1}{2}(1-y), \frac{1}{4}(1+y)\right)^\top, \quad x_2^2 := \left(\frac{1}{2}(1-y), \frac{1}{4}(5+y)\right)^\top.$$

These relations result in the following one-dimensional problem, equivalent to (4.1):

$$\min_{y\geq 0}\Big(c_1+\frac{3}{4}c_2-\frac{1}{2}q_1+\frac{1}{4}q_2\Big)y+\frac{3}{4}c_2+\frac{1}{2}q_1+\frac{3}{4}q_2\,.$$

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

$$c_1\geq -\frac{3}{4}c_2+\frac{1}{2}q_1-\frac{1}{4}q_2\,. \tag{4.5}$$

The optimal value for the primal variable is $\bar{x}_1 := (0,\frac{3}{4})^\top\,,\quad \bar{x}_2^1 := (\frac{1}{2},\frac{1}{4})^\top\,,\quad \bar{x}_2^2 := (\frac{1}{2},\frac{5}{4})^\top\,.$ 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 (2.7) is satisfied, consider its equivalent formulation (2.9). Suppose $\bar{\mu}=-\alpha e_1\in \operatorname{Im}A^\top$. For condition (2.9) 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 $s:=\Big(4,3,-2,1,-2,1\Big)^\top$, this means that $\langle\bar{\mu},e^1\rangle=-4\alpha$ and forces $\alpha=0$, showing satisfaction of (2.9) and (2.8), as claimed.

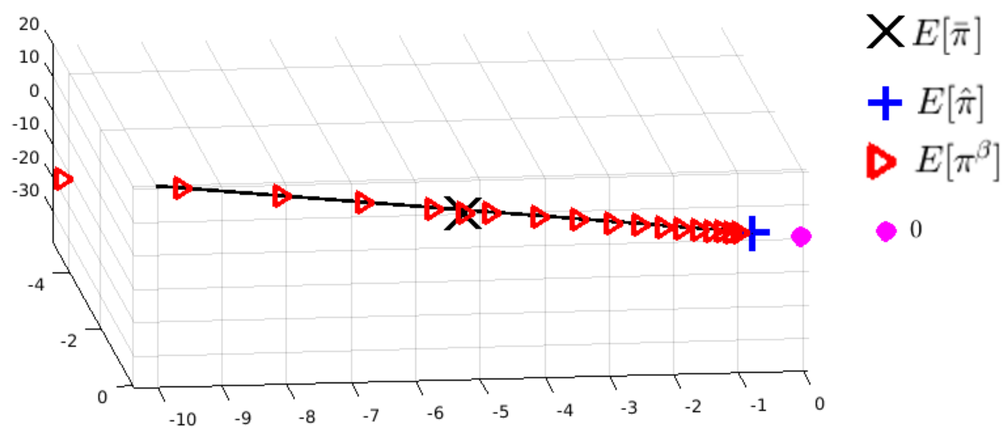

Figure 1: Unbounded set of optimal multipliers in mean (the line), the element with minimal norm (the plus sign), the mean multiplier found for $\beta=0$ (the cross), the mean multiplier estimates for different values of $\beta>0$ (the triangles), and origin (the dot)

Take $q_1=Q=-q_2$ for some Q and $c_2=0$ and any $c_1\geq Q$ (so that (4.5) is satisfied). Optimal Lagrange multipliers must solve the system

$$A^\top\pi=-g+\bar{\mu}\,,\ \text{with } g=\Big(c_1,0,\frac{Q}{2},-\frac{Q}{2},\frac{Q}{2},-\frac{Q}{2}\Big)^\top\ \text{ and } \bar{\mu}=-\alpha e_1 \text{ with } \alpha\geq 0.$$

After some algebraic manipulations, the unbounded set of optimal multipliers is:

$$\mathcal{L} := \left\{ \bar{\pi} = t\frac{c_1}{2}\Big(1, -4, 2, -3, 4, 2\Big)^\top \mid t \geq 1 \right\},$$

and, hence, $t = 1$ gives the minimal-norm element $\hat{\pi}$ for which $\mathbb{E}\left[\hat{\pi}\right] = \frac{c_1}{2}\Big(-1, 0, 2\Big)^\top$.

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 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 β tends to zero. Finally, the cross displays $\mathbb{E}\left[\bar{\pi}\right]$, the mean multiplier found when solving (4.1), whose norm is larger than the minimal one.

4.2.2 Combined index of variance and mean value

Performance profiles [DM02] are useful tools to benchmark different methods on a fair basis. For a battery of two-stage stochastic linear programming problems we compare the expected value of the multipliers obtained as follows:

- when solving (4.1) in its two-level formulation (4.3) with recourse function (4.2), by means of 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 (4.3) with recourse function (4.4).

All the tests were ran 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 (4.1) with 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 the vector $\frac{c}{2}$. 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

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

and define the following index c_P^β for problem p and regularization parameter $\beta \geq 0$:

$$c_P^\beta := \frac{\left\|\mathrm{Var}\left[\mathbb{E}\left[\pi_P^\beta\right]\right]\right\|}{\left\|\mathrm{Var}\left[\pi_P^{\text{best}}\right]\right\|} + \left(1 - \frac{\left\|\mathbb{E}\left[\pi_P^\beta\right]\right\|}{\left\|\mathbb{E}\left[\pi_P^{\text{best}}\right]\right\|}\right). \tag{4.6}$$

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 2.

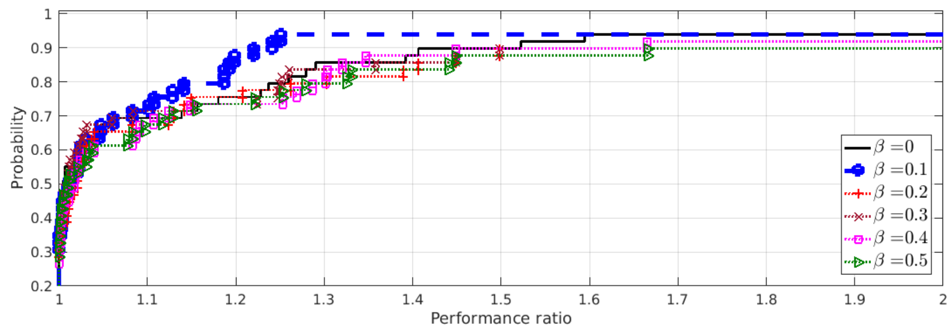

Figure 2: Combined gains in expected value and variance for the dual variable.

In terms of the combined index, and as expected, 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 cercles) 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 3 a performance profile of the first-stage variable, measured now with the index

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

defined for $\beta > 0$ (in our approach comparing variances is not sound, as there is no “stabilization" of the primal variables). It can be seen in the graph that the best value for β in the dual performance profile in Figure 2, that is $\beta = 0.1$ (dashed line with cercles), behaves reasonably well in the primal variable.

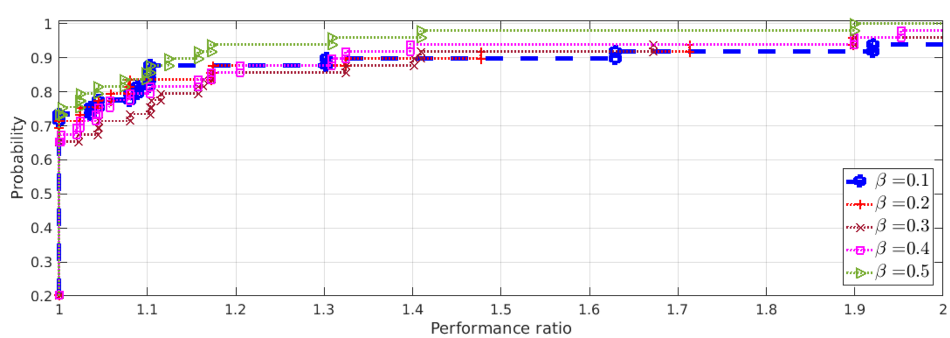

Figure 3: Progression of expected value of first-stage variable

Concluding Remarks

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 proof-of-concept computational experience 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 (4.6) 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 is to develop a quantitave stability analysis, along the lines of [LRX14], but on the dual variables; see also [DR00], [Röm03].

Acknowledgements

The authors are grateful to W. de Oliveira for providing the Matlab code for the test functions employed for the performance profiles.

References

[AM16] W. van Ackooij and J. Malick. "Decomposition algorithm for large-scale two-stage unit-commitment". *Annals of Operations Research* 238.1–2 (Mar. 2016), pp. 587–613.

[BL88] J. R. Birge and F. Louveaux. "A multicut algorithm for two-stage Stochastic Linear Programs". *European Journal of Operational Research* 34.3 (1988), pp. 384–392.

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

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

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

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

[Fáb+15] C. Fábián, C. Wolf, A. Koberstein, and L. Suhl. "Risk-Averse Optimization in Two-Stage Stochastic Models: Computational Aspects and a Study". *SIAM Journal on Optimization* 25.1 (2015), pp. 28–52.

[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". *Naval Research Logistics Quarterly* 3.1–2 (1956), pp. 95–110.

[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.

[KSH02] A. J. Kleywegt, A. Shapiro, and T. Homem-de-Mello. "The sample average approximation method for stochastic discrete optimization". *SIAM J. Optim.* 12.2 (2001/02), 479–502 (electronic).

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

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

[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". *SIAM Journal on Optimization* 21 (2 2011), pp. 517–544.

[Röm03] W. Römisch. "Stability of Stochastic Programming Problems". In: *Stochastic Programming*. Vol. 10. Handbooks in Operations Research and Management Science. Elsevier, 2003, pp. 483–554.

[Rus99] A. Ruszczyński. "Some advances in decomposition methods for stochastic linear programming". *Ann. Oper. Res.* 85 (1999), pp. 153–172.

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

[SDR09] A. Shapiro, D. Dentcheva, and A. Ruszczyński. *Lectures on Stochastic Programming: Modeling and Theory*. Ed. by M. P. Society. MPS–SIAM Series on Optimization. Philadelphia: SIAM - Society for Industrial and Applied Mathematics, 2009, p. 436.

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

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

[SW69] R. van Slyke and R.-B. Wets. “L-shaped linear programs with applications to optimal control and stochastic programming”. *SIAM Journal of Applied Mathematics* 17 (1969), pp. 638–663.

[ZPR00] G. Zakeri, A. B. Philpott, and D. M. Ryan. “Inexact cuts in Benders decomposition”. *SIAM J. Optim.* 10.3 (2000), 643–657 (electronic).