Citation:
Rockafellar, R.T. and Sun, J. 2020. Solving Lagrangian variational inequalities with applications to stochastic programming. Mathematical Programming. 181 (2): pp. 435-451. http://doi.org/10.1007/s10107-019-01458-0

SOLVING LAGRANGIAN VARIATIONAL INEQUALITIES WITH APPLICATIONS TO STOCHASTIC PROGRAMMING

***R. Tyrrell Rockafellar,*[1] *Jie Sun*[2]**

Abstract

Lagrangian variational inequalities feature both primal and dual elements in expressing first-order conditions for optimality in a wide variety of settings where “multipliers” in a very general sense need to be brought in. Their stochastic version relates to problems of stochastic programming and covers not only classical formats with inequality constraints but also composite models with nonsmooth objectives. The progressive hedging algorithm, as a means of solving stochastic programming problems, has however focused so far only on optimality conditions that correspond to variational inequalities in primal variables alone. Here that limitation is removed by appealing to a recent extension of progressive hedging to multistage stochastic variational inequalities in general.

Keywords: *stochastic variational inequality problems, stochastic programming problems, Lagrangian variational inequalities, Lagrange multipliers, progressive hedging algorithm, proximal point algorithm, composite optimization*

March 20, 2019

[1]University of Washington, Department of Mathematics, Box 354350, Seattle, WA 98195-4350;
E-mail: *rtr@uw.edu*, URL: www.math.washington.edu/∼rtr/mypage.html
[2]Curtin University, Department of Mathematics and Statistics, Bentley, Australia 6102;
E-mail: *jie.sun@curtin.edu.au*, URL: oasisapps.curtin.edu.au/staff/profile/view/Jie.Sun

1 Introduction

Variational inequalities provide a effective modeling scheme for conditions that express optimality or equilibrium. Lagrangian variational inequalities emphasize a primal-dual structure in this respect which allows Lagrange multipliers to be brought in and handled explicitly. This paper is devoted to showing how Lagrangian variational inequalities, articulated in a stochastic setting, can furnish broader ways of formulating and solving problems of convex stochastic programming, whether single-stage or multistage. The particular aim within this is expanding the progressive hedging algorithm of [11] to take advantage of Lagrange multipliers as dual variables alongside of the usual primal variables.

For simplicity and in line with our computational aims, we keep to a finite-dimensional framework. When it comes to stochastics, that will mean limiting ourselves to probability spaces based on only finitely many scenarios, which was the case originally for the progressive hedging algorithm as well.

The *variational inequality* problem associated with a nonempty closed convex set $C \subset I\!R^n$ and a continuous mapping $F : C \rightrightarrows I\!R^n$ is to

$$\text{find } \bar{x} \in C \text{ such that } -F(\bar{x}) \in N_C(\bar{x}), \tag{1.1}$$

where $N_C(\bar{x})$ is the *normal cone* to C at $\bar{x}$ in the sense of convex analysis [4], namely the closed convex cone defined by

$$v \in N_C(\bar{x}) \iff \bar{x} \in C \text{ and } v{\cdot}(x - \bar{x}) \leq 0 \text{ for all } x \in C. \tag{1.2}$$

An immediate connection with optimization comes from the problem of minimizing a continuously differentiable function f_0 over C, since the first-order necessary condition for local optimality in that is the case of (1.1) in which $F = \nabla f_0$. This is moreover a sufficient condition for global optimality when f_0 is convex. But the connections between variational inequalities and optimality can be much richer than just this.

The *Lagrangian variational inequality* problem for a pair of nonempty closed convex sets $X \subset I\!R^n$ and $Y \subset I\!R^m$ and a continuously differentiable function L on $X \times Y$ is to

$$\text{find } (\bar{x}, \bar{y}) \in X \times Y \text{ such that } -\nabla_x L(\bar{x}, \bar{y}) \in N_X(\bar{x}), \quad \nabla_y L(\bar{x}, \bar{y}) \in N_Y(\bar{y}). \tag{1.3}$$

This is indeed a variational inequality problem in the mold of (1.1), although trageting $(\bar{x}, \bar{y})$ instead of just $\bar{x}$, as can be seen through the rule that

$$N_{X \times Y}(\bar{x}, \bar{y}) = N_X(\bar{x}) \times N_Y(\bar{y}) \tag{1.4}$$

by taking

$$F(x, y) = (\nabla_x L(x, y), -\nabla_y L(x, y)). \tag{1.5}$$

The classical case of this builds on minimizing $f_0(x)$ over $x \in C$ by introducing a constraint representation for C such as

$$x \in C \iff x \in X \text{ and } (f_1(x), \ldots, f_m(x)) \in (-\infty, 0]^s \times [0, 0]^{m-s} \tag{1.6}$$

with $f_0, f_1, \ldots, f_m$ being continuously differentiable. The associated Lagrangian function is

$$\begin{array}{c} L(x, y) = f_0(x) + y_1 f_1(x) + \cdots + y_m f_m(x) \text{ for } x \in X \\ \text{and } y = (y_1, \ldots, y_m) \in Y = [0, \infty)^s \times (-\infty, \infty)^{m-s}, \end{array} \tag{1.7}$$

and the Lagrangian variational inequality (1.3) then gives the corresponding Karush-Kuhn-Tucker necessary conditions for local optimality, which are sufficient in the convex programming case where

f_i is convex for $i = 0, 1, \ldots, s$ and affine for $i = s+1, \ldots, m$. It was the recognition of this formulation by Robinson in 1979 [3] that brought variational inequality problems (although he called them "generalized equations") into the mainstream of optimization from their origins in territory of partial differential equations, cf. [2].

Lagrangian variational inequality problems have diverse applications beyond this classical case, however. They can cover many other situations, for instance in nonsmooth composite optimization, where the y vector isn't merely tied to constraints. In general, as long as the continuously differentiable function$L$ on $X \times Y$ has $L(x, y)$ concave with respect to $y \in Y$ for each $x \in X$, there is an optimization problem in x directly associated with it:

$$\text{minimize } \; f(x) \; \text{ over } \; x \in X \; \text{ for } \; f(x) = \sup_{y \in Y} L(x, y). \tag{1.8}$$

(Here $f(x)$ might take on ∞, so an implicit constraint, beyond $x \in X$, is that x should belong to $\{\, x \,|\, f(x) < \infty \}$.) Typically the variational inequality in (1.3) will serve as a first-order necessary condition for local optimality under a constraint qualification and, on the other hand, as a sufficient condition for global optimality when, in addition $L(x, y)$ is convex in $x \in X$ for each $y \in Y$. For example, in generalization of (1.7), the function

$$\begin{gathered} L(x, y) = f_0(x) + y_1 f_1(x) + \cdots + y_m f_m(x) - k(y) \; \text{ for } \; x \in X \\ \text{for some choice of a differentiable convex function } k \text{ on } Y \end{gathered} \tag{1.9}$$

serves as the Lagrangian for the problem

$$\begin{aligned} &\text{minimize } \; f_0(x) + \theta(f_1(x), \ldots, f_m(x)) \; \text{ over } \; x \in X, \\ &\text{where } \; \theta(u_1, \ldots, u_m) = \sup_{y \in Y} \{\, y_1 u_1 + \ldots y_m u_m - k(y) \}. \end{aligned} \tag{1.10}$$

Observe that the classical case (1.7) corresponds to the Y indicated there along with $k(y) \equiv 0$, but the nonclassical possibilities can cover penalty functions, max expressions, and a lot more. Background on this can be found in [9] and [12, Chapter 11].

The variational inequalities associated with *stochastic programming* problems, in which optimization proceeds in discrete stages under constraints of nonanticipativity on how decisions can respond to information, will be described in the next section. They are more complicated because of the nonanticipativity, but can have aspects reflecting the optimization cases of (1.1) and (1.3). The Lagrangian forms allied with (1.3) will especially be our focus. They will be cast as *stochastic variational inequalities* in the pattern developed in [13] which adapts to allowing "informations prices" to serve as Lagrange multiplier of another sort.

Methods for solving variational inequalities have especially relied on taking avantage of monotonicity when available. A variational inequality problem (1.1) is *monotone* when the mapping F is monotone on C, which means that

$$(F(x') - F(x)) \cdot (x' - x) \geq 0 \; \text{ for all } \; x, x' \in C, \tag{1.11}$$

and is true in particular when $F = \nabla f_0$ for a convex function f_0. Then the set-valued mapping

$$T : x \mapsto F(x) + N_C(x) \tag{1.12}$$

is in fact *maximal* monotone in the sense that its graph can't be enlarged without upsetting monotonicity, i.e., it's impossible to have a pair (x^*, v^*) satisfying $(v^* - F(x)) \cdot (x^* - x) \geq 0$ for all $x \in C$

unless $x^* \in C$ and $v^* = F(x^*)$. The Lagrangian variational inequality problem enjoys such monotonicity when $L(x, y)$ is convex with respect to $x \in X$ and concave with respect to $y \in Y$, which is true of course in the convex programming version in particular.

Solving (1.1) corresponds with respect to (1.12) to finding some $\bar{x}$ such that $0 \in T(\bar{x})$. The proximal point algorithm [7] is one of the most basic approaches to doing that when T is maximal monotone. When executed in the Lagrangian framework, it comes out as a "multiplier method" of augmented Lagrangian type, as shown in [8]. That will have a role in our efforts later, on the side, but our main goal will be to expand the capabilities of the progressive hedging algorithm [11] for solving stochastic programming problems so as to iterate also on Lagrange multipliers.

Our contribution in this direction will be kept to the convex case of stochastic programming with its underlying monotonicity. Already in [10] we have shown that the progressive hedging algorithm can be generalized to work not just with minimization but in solving monotone stochastic variational inequality problems very broadly. The key here will be specializing that to Lagrangian variational inequalities and determining how that plays out in optimization-based iterations.

2 Stochastic variational inequalities and progressive hedging

The special feature of stochastic programming in contrast to other areas of optimization is the structure of one or more stages, in which information may be revealed, and the extra constraints that may impose on how decisions can be made. This structure carries over to stochastic variational inequalities as well and must be reviewed before we can proceed further.

The information structure put to use here will be based on there being finitely many scenarios, each with its own probability, but the particular scenario being followed is only known in part, until the very end. It's popular to express this in terms of a so-called scenario tree with branching probabilities, however our formulation (ultimately equivalent) will follow that in [13] and [10].

We consider a finite collection Ξ of *scenarios* ξ, each having an assigned probability $\pi(\xi) > 0$, with these probabilities adding of course to 1; this furnishes the elementary discrete probability space that underlies the developments. In allowing for one or more stages $k = 1, \ldots, N$ of scenario-influenced decision-making, we denote by $\mathcal{L}_n$ the (finite-dimensional) linear space consisting of all mappings

$$x(\cdot) : \xi \in \Xi \mapsto x(\xi) = (x_1(\xi), \ldots, x_N(\xi)) \in \Pi_{k=1}^N I\!R^{n_k} = I\!R^n \tag{2.1}$$

and furnish it with the expectational inner product

$$\langle w(\cdot), x(\cdot) \rangle = \sum\nolimits_{\xi \in \Xi} \pi(\xi) \sum\nolimits_{k=1}^{n} w_k(\xi) {\cdot} x_k(\xi) = E_\xi[w(\xi) {\cdot} x(\xi)]. \tag{2.2}$$

Here $w_k(\xi){\cdot}x_k(\xi)$ refers to the uaual inner product between the vectors $w_k(\xi)$ and $x_k(\xi)$ in $I\!R^{n_k}$, and likewise $w(\xi){\cdot}x(\xi)$ refers to the usual inner product in $I\!R^n$. But the inner product (2.2) makes $\mathcal{L}_n$ into a Hilbert space with a different norm than the Euclidean norm, which would have 1 in place of the probabilities, and also a different meaning for orthogonality — as is essential in what follows.

For information structure we regard scenarios as having the form

$$\xi = (\xi_1, \ldots, \xi_N) \ \text{ with } \ \xi_i \in \Xi_k, \ \text{ so that } \ \Xi \subset \Pi_{k=1}^N \Xi_i, \tag{2.3}$$

with the interpretation that ξ_k is the aspect of ξ that becomes known in decision stage k *after* the decision in that stage has been finalized.

This structure leads to the important constraint of *nonanticipativity* on a mapping $x(\cdot)$ as a *decision policy*, namely that

$$x_k(\xi) \ \text{ only depends on } \ (\xi_1, \ldots, \xi_{k-1}). \tag{2.4}$$

By introducing

$$\mathcal{N}_n = \{\, x(\cdot) \in \mathcal{L}_n \mid \text{ such that (2.4) holds} \,\} \tag{2.5}$$

as the *nonanticipativity subspace* of $\mathcal{L}_n$, we can express this constraint by $x(\cdot) \in \mathcal{N}_n$.

For the purpose of understanding the concept of a stochastic variational inequality problem that will be central here, consider now, in combination with this information structure, an underlying constraint $x(\cdot) \in \mathcal{C}$ for a nonempty closed convex set $\mathcal{C} \subset \mathcal{L}_n$. Specifically this is to be of the kind where

$$x(\cdot) \in \mathcal{C} \iff x(\xi) \in C(\xi) \;\text{ for all }\; \xi \in \Xi, \tag{2.6}$$

with each $C(\xi)$ being a nonempty closed convex set in $I\!R^n$. (Later, more details about $C(\xi)$ will be of interest, but they aren't needed for now.) Consider further a continuous mapping $\mathcal{F} : \mathcal{L}_n \to \mathcal{L}_n$ of the kind where

$$\mathcal{F}(x(\cdot)) = v(\cdot) \iff F(x(\xi), \xi) = v(\xi) \;\text{ for all }\; \xi \in \Xi, \tag{2.7}$$

with each $F(\cdot, \xi)$ being a continuous mapping from $I\!R^n$ into $I\!R^n$.[3] The problem

$$\text{find } \;\bar{x}(\cdot) \in \mathcal{C} \cap \mathcal{N}_n \;\text{ such that }\; -\mathcal{F}(\bar{x}(\cdot)) \in N_{\mathcal{C}\cap\mathcal{N}_n}(\bar{x}(\cdot)) \tag{2.8}$$

is the *stochastic variational inequality problem in basic form* associated with $\mathcal{C}$ and $\mathcal{F}$ in the terminology of [13]. It is assumed here that $\mathcal{C} \cap \mathcal{N}_n \neq \emptyset$.

This is truly a variational inequality in the pattern of (1.1), although placed in a more advanced context. It is of monotone type when $\mathcal{F}$ is monotone, which corresponds to each of the mappings $F(\cdot, \xi)$ being monotone. *Such monotonicity is assumed in what follows.*

But what does the normal cone condition in (2.8) with respect to the closed convex set $\mathcal{C} \cap \mathcal{N}_n \subset \mathcal{L}_n$ really say? For this, the orthogonal complement $\mathcal{M}_n$ of the subspace $\mathcal{N}_n$ with respect to the inner product (2.2) has to be brought in:

$$\mathcal{M}_n = \mathcal{N}_n^{\perp} = \{\, w(\cdot) \in \mathcal{L}_n \mid \langle w(\cdot), x(\cdot) \rangle = 0 \;\text{ for all }\; x(\cdot) \in \mathcal{N}_n \}. \tag{2.9}$$

It was established in [13] that a sufficient condition of having $-\mathcal{F}(x(\cdot)) \in N_{\mathcal{C}\cap\mathcal{N}_n}(x(\cdot))$ is the existence of some $w(\cdot) \in \mathcal{M}_n$ such that $-F(x(\xi), \xi) - w(\xi) \in N_{C(\xi)}(x(\xi))$ for all ξ, and this is moreover a necessary condition under a constraint qualification,[4] This leads to recasting the problem in (2.8) as

$$\begin{array}{l} \text{find } \;\bar{x}(\cdot) \in \mathcal{N}_n \;\text{ for which }\; \exists\, \bar{w}(\cdot) \in \mathcal{M}_n \;\text{ such that} \\ -\mathcal{F}(\bar{x}(\cdot)) - \bar{w}(\cdot) \in N_{\mathcal{C}}(\bar{x}(\cdot)), \;\text{ this being equivalent to} \\ -F(\bar{x}(\xi), \xi) - \bar{w}(\xi) \in N_{C(\xi)}(\bar{x}(\xi)) \;\text{ for all }\; \xi \in \Xi, \end{array} \tag{2.10}$$

which is the *stochastic variational inequality problem in extensive form* associated with $\mathcal{C}$ and $\mathcal{F}$ in the terminology of [13].

The striking feature in (2.10) is the decomposition into a separate variational inequality in the space $I\!R^n$ for each scenario ξ, but one which relies not just on $C(\xi)$ and $F(\cdot, \xi)$ as ingredients, but also on the assistance of an auxiliary vector $w(\xi)$. The progressive hedging algorithm, which will be explained shortly, is designed to take advantage of this decomposition.

[3] We are involved with a set-valued random variable $\xi \mapsto C(\xi)$ and a function-valued random variable $\xi \mapsto F(\cdot, \xi)$ as problem data, but in our elementary setting of discrete probability, that perspective isn't essential.

[4] A simple criterion is the existence of some $\hat{x}(\cdot) \in \mathcal{N}_n$ such that $\hat{x}(\xi)$ belongs to the relative interior of $C(\xi)$ for every ξ, but nothing at all is required if the sets $C(\xi)$ are polyhedral. In the finite-dimensional context here, the principles of convex analysis behind this are enough, and nothing about nonanticipativity of the mapping $\xi \mapsto C(\xi)$ is needed.

The subspace $\mathcal{M}_n$ that enters (2.10) from (2.9) has the probabilistic description that

$$w(\cdot) \in \mathcal{M}_n \iff E_{\xi \,|\, \xi_1,\ldots,\xi_{k-1}}[w_k(\xi)] = 0 \text{ for } k = 1, \ldots, N, \tag{2.11}$$

where the expectation is the conditional expectation over the remaining possibilities for the scenario ξ, given that the portion $(\xi_1, \ldots, \xi_{k-1})$ is already known when a decision has to be fixed for stage k. (Note that this means for $k = 1$ that $x_1(\xi)$ must be the *same* for all scenarios k.) The linear transformations

$$P_{\mathcal{N}_n} = \text{ projection onto } \mathcal{N}_n, \qquad P_{\mathcal{M}_n} = \text{ projection onto } \mathcal{M}_n, \qquad P_{\mathcal{N}_n} + P_{\mathcal{M}_n} = I, \tag{2.12}$$

are easily executed in a numerical context and provide vital computational tools.

Progressive Hedging Algorithm for General Stochastic Variational Inequalities [10]. *Under the assumption of monotonicity, the iterations, indexed by $\nu = 1, \ldots,$ utilize a parameter value $r > 0$ and current elements $x^\nu(\cdot) \in \mathcal{N}_n$ and $w^\nu(\cdot) \in \mathcal{M}_n$. For each $\xi \in \Xi$, a vector $\widehat{x}^\nu(\xi)$ is obtained by solving the variational inequality*

$$-F^\nu(\widehat{x}^\nu(\xi), \xi) \in N_{C(\xi)}(\widehat{x}^\nu(\xi)), \quad \textit{where } \; F^\nu(x, \xi) = F(x, \xi) + w^\nu(\xi) + r[x - x^\nu(\xi)]. \tag{2.13}$$

The function $\widehat{x}^\nu(\cdot) \in \mathcal{L}_n$ thereby determined is projected onto $\mathcal{N}_n$ and $\mathcal{M}_n$ to get the updates

$$x^{\nu+1}(\cdot) = P_{\mathcal{N}_n}(\widehat{x}^\nu(\cdot)), \qquad w^{\nu+1}(\cdot) = w^\nu(\cdot) + rP_{\mathcal{M}_n}(\widehat{x}^\nu(\cdot)). \tag{2.14}$$

Note that because of the proximal term $r[x - x^\nu]$ in (2.13) the monotonicity assumed for $F(\cdot, \xi)$ makes $F^\nu(\cdot, \xi)$ strongly monotone from $I\!R^n$ into itself. That ensures the existence of a unique solution $\widehat{x}^\nu(\xi)$ to the subproblem in question. The rule for getting $w^{\nu+1}(\cdot)$ can be posed more simply by appealing to the relationship at the end of (2.12), according to which $P_{\mathcal{M}_n}(\widehat{x}^\nu(\cdot)) = \widehat{x}^\nu(\cdot) - P_{\mathcal{N}_n}(\widehat{x}^\nu(\cdot))$ with $P_{\mathcal{N}_n}(\widehat{x}^\nu(\cdot))$ being $x^{\nu+1}(\cdot)$:

$$w^{\nu+1}(\xi) = w^\nu(\xi) + r[\widehat{x}^\nu(\xi) - x^{\nu+1}(\xi)] \text{ for all } \xi \in \Xi. \tag{2.15}$$

We derived this version of the progressive hedging algorithm in [10] by following the pattern in the original version for stochastic programming in [11]. But it relates also to the method of partial inverses of Spingarn [14], although that would only yield here the case of $r = 1$. Ultimately, however, all this goes back to the proximal point algorithm being applied in a special way, and it inherits the various convergence properties of that algorithm — except for a different role for r. In constrast to the parameter in the proximal point algorithm, having r either to high or too low might detract from performance. Anyway, as long as a solution to (2.10) exists, as holds under a constraint qualification when a solution to (2.8) exists,[5] the sequence of pairs $(x^\nu(\cdot), w^\nu(\cdot))$ is sure to converge to some particular solution pair $(\bar{x}(\cdot), \bar{w}(\cdot))$ and to do so in such a manner that the expression

$$r||x^\nu(\cdot) - \bar{x}(\cdot)||^2 + r^{-1}||w^\nu(\cdot) - \bar{w}(\cdot)||^2 \tag{2.16}$$

keeps decreasing, where the norm in $\mathcal{L}_n$ is the one associated with the inner product (2.2). (In the absence of existence, the norm of this pair tends to ∞.) Conditions are also available under which a linear convergence rate is assured, but such known details are not the focus here.

[5]A simple condition for that is the boundedness of the sets $C(\xi)$, but in this monotone case of a variational inequality many criteria in terms of growth conditions are also available; see [12, Chapter 12].

3 Lagrangian-type progressive hedging in stochastic programming

In the original version of the progressive hedging algorithm in [11], the mapping $\mathcal{F}$ in the stochastic variational inequality was the gradient of a convex objective function defined as an expectations:

$$\mathcal{F}(x(\cdot)) = \nabla\mathcal{F}_0(x(\cdot)) \text{ for } \mathcal{F}_0(x(\cdot)) = E_\xi[f_0(x(\xi),\xi)] = \sum\nolimits_{\xi\in\Xi)} \pi(\xi) f_0(x(\xi),\xi), \tag{3.1}$$

where $f_0(\cdot,\xi)$ is a continuously differentiable convex function and

$$\nabla\mathcal{F}_0(x(\cdot)) = v(\cdot) \iff \nabla f_0(x(\xi),\xi) = v(\xi) \text{ for all } \xi\in\Xi. \tag{3.2}$$

In that case, in which $F(x,\xi) = \nabla f_0(x,\xi)$, solving the variational inequality subproblems in (2.13) can be cast as solving optimization subproblems:

$$\widehat{x}^\nu(\xi) = \operatorname*{argmin}_{x\in C(\xi)} \Big\{ f_0(x,\xi) + w^\nu(\xi)\cdot x + \frac{r}{2}||x - x^\nu(\xi)||^2 \Big\}, \tag{3.3}$$

because the function being minimized is strongly convex with its gradient equal to $F(x,\xi) + w^\nu + r[x - x^\nu]$.

This is well and good, but it takes no advantage of any possible details in the specification of the convex set $C(\xi)$. There might, for example, be a constraint structure like

$$x \in C(\xi) \iff x\in X(\xi) \text{ and } (f_1(x,\xi),\dots,f_m(x,\xi)) \in (-\infty,0]^s \times [0,0]^{m-s} \tag{3.4}$$

in emulation of (1.6) with respect to sets $X(\xi)$ and functions $f_i(\cdot,\xi)$. Presumably, in solving the optimization subproblems in (3.3), Lagrange multipliers $y_i(\xi)$ would come up, but there would be no coordination of them from iteration to iteration.

Instead, we envision operating in a Lagrangian format in which current multipliers $y^\nu(\xi)$ get updated to $y^{\nu+1}(\xi)$ and converge to multipliers $\bar{y}(\xi)$. Moreover we propose to pursue this in the general format of Lagrangian functions and their associated optimization problems in (1.8) so as to provide a far broader modeling scheme for stochastic programming than has been available up to now. Although we will soon be interested in exploring features like how constraints such as in (3.4) might be broken down into batches that evolve in the decision stages $k = 1,\dots,N$, and the nonclassical counterparts to that, we begin with these ideas in a less cluttered and more abstract framework.

Limiting ourselves here anyway to modes of *convex* optimization, we introduce nonempty closed convex sets $X(\xi) \subset I\!R^n$ and $Y(\xi) \subset I\!R^m$ and let

$$\mathcal{X} = \{\, x(\cdot) \,|\, x(\xi) \in X(\xi),\, \forall\xi\}, \qquad \mathcal{Y} = \{\, y(\cdot) \,|\, y(\xi) \in Y(\xi),\, \forall\xi\}. \tag{3.5}$$

In this situation, where $x(\cdot) \in \mathcal{L}_n$, we similarly think of $y(\cdot)$ as belonging to the space $\mathcal{L}_m$ consisting of all mappings

$$y(\cdot) : \xi \to y(\xi) = (y_1(\xi),\dots,y_N(\xi)) \in I\!R^{m_1} \times \cdots \times I\!R^{m_N} = I\!R^m. \tag{3.6}$$

Likewise, in parallel to the nonanticipativity subspace $\mathcal{N}_n \subset \mathcal{L}_n$ and its complement $\mathcal{M}_n$ we will have the nonanticipativity subspace $\mathcal{N}_m \subset \mathcal{L}_n$ and its complement $\mathcal{M}_m$.

Next we introduce continuously differentiable functions

$$L(\cdot,\cdot,\xi) \text{ on } X(\xi)\times Y(\xi) \text{ such that } L(x,y,\xi) \text{ is convex in } x \text{ and concave in } y \tag{3.7}$$

and define

$$\Lambda(x(\cdot),y(\cdot)) = E_\xi[L(x(\xi),y(\xi),\xi)] = \sum_{\xi\in\Xi}\pi(\xi)L(x(\xi),y(\xi),\xi) \text{ for } x(\cdot)\in\mathcal{X},\ y(\cdot)\in\mathcal{Y}. \tag{3.8}$$

In this way we have a continuously differentiable convex-concave function Λ on the convex product set $\mathcal{X}\times\mathcal{Y}$. In formulating a corresponding problem of stochastic programming, however, nonanticipativity also has to be brought in: we have to consider the Lagrangian not just on $\mathcal{X}\times\mathcal{Y}$ but on the closed convex set $(\mathcal{X}\cap\mathcal{N}_n)\times(\mathcal{Y}\cap\mathcal{N}_m)$.

We arrive in this way at the following generalized Lagrangian format for a stochastic programming problem of convex type:

$$\text{minimize } \varphi(x(\cdot)) \text{ over } x(\cdot)\in\mathcal{X}\cap\mathcal{N}_n \text{ for } \varphi(x(\cdot)) = \sup_{y(\cdot)\in\mathcal{Y}\cap\mathcal{N}_m}\Lambda(x(\cdot),y(\cdot)). \tag{3.9}$$

for which the Lagrangian first-order optimality condition will have the saddle point form

$$-\nabla_{x(\cdot)}\Lambda(\bar{x}(\cdot),\bar{y}(\cdot))\in N_{\mathcal{X}\cap\mathcal{N}_n}(\bar{x}(\cdot)), \qquad \nabla_{y(\cdot)}\Lambda(\bar{x}(\cdot),\bar{y}(\cdot))\in N_{\mathcal{Y}\cap\mathcal{N}_m}(\bar{y}(\cdot)) \tag{3.10}$$

This looks quite different, but all we are really dealing with is the path from a basic variational inequality problem as in (1.1) to a Lagrangian version as in (1.3) by way of (1.4) and (1.5), when played out in the context of a stochastic variational inequality (2.8). Here $\mathcal{C}$, comprised of sets $C(\xi)$, is replaced by $\mathcal{X}\times\mathcal{Y}$, comprised of sets $X(\xi)\times Y(\xi)$, while the associated mapping that takes the form

$$\mathcal{F}(x(\cdot),y(\cdot)) = (\nabla_{x(\cdot)}\Lambda(x(\cdot),y(\cdot)), -\nabla_{y(\cdot)}\Lambda(x(\cdot),y(\cdot))) \tag{3.11}$$

arises from component mappings

$$F(x,y,\xi) = (-\nabla_x L(x,y,\xi), \nabla_y L(x,y,\xi)). \tag{3.12}$$

The analysis undertaken earlier in passing from the basic form for a stochastic variational in (2.8) to the expansive form in (2.10) is reflected then in passing from (3.10) to the problem

$$\begin{array}{l}\text{find } x(\cdot)\in\mathcal{N}_n,\ y(\cdot)\in\mathcal{N}_m, \text{ for which } \exists\,\bar{w}(\cdot)\in\mathcal{M}_n,\ \bar{z}(\cdot)\in\mathcal{M}_m, \text{ such that}\\ \quad -\nabla_{x(\cdot)}\Lambda(\bar{x}(\cdot),\bar{y}(\cdot))-\bar{w}(\cdot)\in N_{\mathcal{X}}(\bar{x}(\cdot)), \quad \nabla_{y(\cdot)}\Lambda(\bar{x}(\cdot),\bar{y}(\cdot))+\bar{z}(\cdot)\in N_{\mathcal{Y}}(\bar{y}(\cdot)),\\ \text{which is equivalent to having, for all scenarios } \xi\in\Xi,\\ \quad -\nabla_x L(\bar{x}(\xi),\bar{y}(\xi),\xi)-\bar{w}(\xi)\in N_{X(\xi)}(\bar{x}(\xi)), \quad \nabla_y L(\bar{x}(\xi),\bar{y}(\xi),\xi)+\bar{z}(\xi)\in N_{Y(\xi)}(\bar{y}(\xi)).\end{array} \tag{3.13}$$

Just as before in the relationship between (2,8) and (2.10), we have the fact that (3.13) is always sufficient for getting a solution to the condition in (3.10), and is moreover necessary under a constraint qualification.[6]

No doubt it would help now in understanding this to look at examples of how our scheme of a stochastic convex problem of optimization and its Lagrangian first-order condition works out in classical cases like the Lagrangian for the constraints in (3.4). But rather than get into that right away, with all its variants and commentary, we prefer to explain next how progressive hedging will be realized in this wider framework.

Although the progressive hedging algorithm in [11] was originally developed only for optimization without the explicit intervention of Lagrange multipliers, we are able now, in passing by way of our recent extension of progressing hedging to monotone stochastic variational inequalities in [10] as reviewed above, to present the new Lagrangian variational inequality form of the algorithm.

[6]Constraint qualification for this would be the existence of some $\widetilde{x}(\cdot)\in\mathcal{N}_n$ and $\widetilde{y}(\cdot)\in\mathcal{N}_m$ such that $\widetilde{x}(\xi)\in\operatorname{ri}X(\xi)$ and $\widetilde{y}(\xi)\in\operatorname{ri}Y(\xi)$. This isn't needed when $X(\xi)$ and $Y(\xi)$ are polyhedral.

Progressive Hedging for Lagrangian Stochastic Variational Inequalities. *Under the assumption of monotonicity, the iterations, indexed by $\nu = 1, \ldots$, utilize a parameter value $r > 0$ and current elements $x^\nu(\cdot) \in \mathcal{N}_n$, $y^\nu(\cdot) \in \mathcal{N}_m$, as well as $w^\nu(\cdot) \in \mathcal{M}_n$, $z^\nu(\cdot) \in \mathcal{N}_m$. For each $\xi \in \Xi$, a vector pair $(\widehat{x}^\nu(\xi), \widehat{y}^\nu(\xi))$, is obtained by solving the Lagrangian variational inequality*

$$-\nabla_x L^\nu(\widehat{x}^\nu(\xi), \widehat{y}^\nu(\xi), \xi) \in N_{X(\xi)}(\widehat{x}^\nu(\xi)), \qquad \nabla_y L^\nu(\widehat{x}^\nu(\xi), \widehat{y}^\nu(\xi), \xi) \in N_{X(\xi)}(\widehat{x}^\nu(\xi)), \tag{3.14}$$

where

$$L^\nu(x, y, \xi) = L(x, y, \xi) + w^\nu(\xi){\cdot}x + \frac{r}{2}||x - x^\nu(\xi)||^2 - z^\nu(\xi){\cdot}y - \frac{r}{2}||y - y^\nu(\xi)||^2. \tag{3.15}$$

Updates are obtained then by

$$\begin{array}{ll} x^{\nu+1}(\cdot) = P_{\mathcal{N}_n}(\widehat{x}^\nu(\cdot)), & w^{\nu+1}(\cdot) = w^\nu(\cdot) + rP_{\mathcal{M}_n}(\widehat{x}^\nu(\cdot)), \\ y^{\nu+1}(\cdot) = P_{\mathcal{N}_m}(\widehat{y}^\nu(\cdot)), & z^{\nu+1}(\cdot) = z^\nu(\cdot) + rP_{\mathcal{M}_n}(\widehat{y}^\nu(\cdot)). \end{array} \tag{3.16}$$

The choice of L^ν in (3.15) captures, in function terms, the replacement of the gradient-based mapping $F(x, y, \xi) = (-\nabla_x L(x, y, \xi), \nabla_y L(x, y, \xi))$ in (3.11) by

$$F^\nu(x, y, \xi) = F(x, y, \xi) + (w^\nu(\xi), z^\nu(\xi)) + r[(x, y) - (x^\nu(\xi), y^\nu(\xi))], \tag{3.17}$$

as demanded by the execution pattern of progressive hedging in solving a general stochastic variational inequality (with monotonicity).

This version of progressive hedging inherits from the one for general stochastic variational inequalities in the preceding section the property that, as long as a solution exists, the sequence of iterates $(x^\nu(\cdot), y^\nu(\cdot), w^\nu(\cdot), z^\nu(\cdot))$ will converge to some particular solution $(\bar{x}(\cdot), \bar{y}(\cdot), \bar{w}(\cdot), \bar{z}(\cdot))$, moreover with the expression

$$r||(x^\nu(\cdot), y^\nu(\cdot)) - (\bar{x}(\cdot), \bar{y}(\cdot)||^2 + r^{-1}||(w^\nu(\cdot), z^\nu(\cdot)) - (\bar{w}(\cdot), \bar{z}(\cdot))||^2 \tag{3.18}$$

decreasing (unless a solution has already been reached). However, instead of forcing the x part and the y part to obey the same parameter, the bond between them can be relaxed, as shown next.

Theorem 1 (double parameterization)**.** *The Lagrangian version of the progressive hedging algorithm still works if, for some $s > 0$ different from r, the function in (3.15) is replaced by*

$$L^\nu(x, y, \xi) = L(x, y, \xi) + w^\nu(\xi){\cdot}x + \frac{s}{2}||x - x^\nu(\xi)||^2 - z^\nu(\xi){\cdot}y - \frac{r}{2}||y - y^\nu(\xi)||^2. \tag{3.19}$$

and the w-update rule in (3.16) is changed to

$$w^{\nu+1}(\cdot) = w^\nu(\cdot) + sP_{\mathcal{M}_n}(\widehat{x}^\nu(\cdot)). \tag{3.20}$$

In that case convergence takes place with the different expression

$$s||x^\nu(\cdot) - \bar{x}(\cdot)||^2 + r||y^\nu(\cdot) - \bar{y}(\cdot)||^2 + s^{-1}||w^\nu(\cdot) - \bar{w}(\cdot)||^2 + r^{-1}||z^\nu(\cdot) - \bar{z}(\cdot)||^2 \tag{3.21}$$

decreasing instead of the expression (3.18).

Proof. This modification is achieved by a change of variable from $x(\cdot)$ to $\widetilde{x}(\cdot) = tx(\cdot)$ and $w(\cdot)$ to $\widetilde{w}(\cdot) = t^{-1}x(\cdot)$ with $t = \sqrt{s/r}$. This corresponds to replacing $L(x, y, \xi)$ by

$$\widetilde{L}(\tilde{x}, y, \xi) = L(t^{-1}\tilde{x}, y, \xi) \tag{3.22}$$

and $L^\nu(x,y,\xi)$ accordingly by

$$\widetilde{L}^\nu(\widetilde{x},y,\xi) = \widetilde{L}(\widetilde{x},y,\xi) + \widetilde{w}^\nu(\xi){\cdot}\widetilde{x} + \frac{r}{2}||\widetilde{x} - \tilde{x}^\nu(\xi)||^2 - z^\nu(\xi){\cdot}y - \frac{r}{2}||y - y^\nu(\xi)||^2 \tag{3.23}$$

while $X(\xi)$ is changed to

$$\widetilde{X}(\xi) = t^{-1}X(\xi). \tag{3.24}$$

Note that executing the algorithm with these modified ingredients just amounts to executing it as before, except with the alterations in (3.19) and (3.20), inasmuch as $rt^2 = s$.

The algorithm with the modified ingredients generates sequences $(\widetilde{x}^\nu(\cdot), y^\nu(\cdot))$ and $(\widetilde{w}^\nu(\cdot), z^\nu(\cdot))$ which converge to solution elements $(\bar{\widetilde{x}}(\cdot), \bar{y}(\cdot))$ and $(\bar{\widetilde{w}}(\cdot), \bar{z}(\cdot))$, moreover with the expression

$$r||(\widetilde{x}^\nu(\cdot), y^\nu(\cdot)) - (\bar{\widetilde{x}}(\cdot), \bar{y}(\cdot))||^2 + r^{-1}||(\widetilde{w}^\nu(\cdot), z^\nu(\cdot)) - (\bar{\widetilde{w}}(\cdot), \bar{z}(\cdot))||^2$$

always decreasing. In reversing the change of variables, we see that this expression converts to

$$r||tx^\nu(\cdot), y^\nu(\cdot)) - (t\bar{x}(\cdot), \bar{y}(\cdot))||^2 + r^{-1}||(t^{-1}w^\nu(\cdot), z^\nu(\cdot)) - (t^{-1}\bar{w}(\cdot), \bar{z}(\cdot))||^2,$$

which can be written as

$$rt^2||x^\nu(\cdot) - \bar{x}(\cdot)||^2 + r||y^\nu(\cdot) - \bar{y}(\cdot)||^2 + r^{-1}t^{-2}||w^\nu(\cdot) - \bar{w}(\cdot)||^2 + r^{-1}||z^\nu(\cdot) - \bar{z}(\cdot)||^2$$

and thus identified with (3.21). □

Our statement of the algorithm needs further explanation, which we turn to next. It's one thing to propose solving Lagrangian variational inequality subproblems in the form of (3.14) as a key step in computations, but another thing to make clear how this can actually be carried out with optimization software.

Theorem 2 (solving Lagrangian VI subproblems by optimization)**.** *Solving the subproblem (3.14) mean determining the unique saddle point of $L^\nu(x,y,\xi)$ with respect to minimizing over $x \in X(\xi)$ and maximizing over $y \in Y(\xi)$:*

$$\widehat{x}^\nu(\xi) = \operatorname*{argmin}_{x\in X(\xi)} L^\nu(x, \widehat{y}^\nu(\xi), \xi), \qquad \widehat{y}^\nu(\xi) = \operatorname*{argmax}_{y\in Y(\xi)} L^\nu(\widehat{x}^\nu(\xi), y, \xi). \tag{3.25}$$

This can be achieved in terms of the function $u^\nu_r(x,\xi)$ and set $U^\nu_r(x,\xi)$ defined by

$$\begin{aligned} u^\nu_r(x,\xi) &= \max_{y\in Y(\xi)} \Big\{ L(x,y,\xi) - z^\nu(\xi){\cdot}y - \frac{r}{2}||y - y^\nu(\xi)||^2 \Big\},\\ U^\nu_r(x,\xi) &= \operatorname*{argmax}_{y\in Y(\xi)} \Big\{ L(x,y,\xi) - z^\nu(\xi){\cdot}y - \frac{r}{2}||y - y^\nu(\xi)||^2 \Big\}, \end{aligned} \tag{3.26}$$

in two steps, as follows:

$$\widehat{x}^\nu(x) = \operatorname*{argmin}_{x\in X(\xi)} \Big\{ u^\nu_r(x,\xi) + w^\nu(\xi){\cdot}x + \frac{s}{2}||x - x^\nu(\xi)||^2 \Big\}, \qquad \widehat{y}^\nu(\xi) = U^\nu_r(\widehat{x}^\nu(\xi), \xi), \tag{3.27}$$

where having s in place of r in the minimization formula for $\widehat{x}^\nu(\xi)$ reflects the double parameterization allowable through Theorem 1.

Proof. The normal cone conditions in (3.14) are the first-order optimality conditions for the minimization and maximization in question, and because of the convexity in x and concavity in y they are both necessary and sufficient. The proximal terms in (3.15) guarantee strong convexity and concavity in fact, so the max and min are attained and the saddle point is unique.

It's fundamental from saddle point theory that having $(\widehat{x}^\nu(\xi), \widehat{y}^\nu(\xi))$ be a saddle point as in (3.25) is equivalent (when a saddle point is known to exist, as here) to having

$$\begin{array}{l} \widehat{x}^\nu(\xi) \in \operatorname{argmin}_{x\in X(\xi)} \varphi^\nu(x,\xi) \;\text{ for }\; \varphi^\nu(x) = \max_{y\in Y(\xi)} L^\nu(x,y,\xi), \\ \widehat{y}^\nu(\xi) \in \operatorname{argmax}_{y\in Y(\xi)} \psi^\nu(y,\xi) \;\text{ for }\; \psi^\nu(y) = \min_{x\in X(\xi)} L^\nu(x,y,\xi). \end{array} \tag{3.28}$$

In this case we have through (3.26) that

$$\varphi^\nu(x,\xi) = u_r^\nu(x,\xi) + w^\nu(\xi){\cdot}x + \frac{s}{2}||x - x^\nu(\xi)||^2,$$

so that the first part of (3.28) is equivalent to the first part of (3.27) and correctly identifies $\widehat{x}^\nu(\xi)$. Then, though, without going back to (3.28), we can appeal to the second part of (3.25) to get $\widehat{y}^\nu(\xi)$, with the recognition that this reduces to the prescription in the second part of (3.27). □

The usefulness of the mode of calculation in Theorem 2 depends, of course, on the accessibility of the expressions in (3.26). That could depend on the circumstances of the problem and its Lagrangian formulation, and there's a lot of territory to explore in that direction. However, in the case of classical Lagrangians like (1.7) entering as ingredients, everything works well and introduces to stochastic programming the methodology of so-called multiplier methods [5, 1, 6]. That carries over more or less intact also to the kind of composite optimization structure covered in (1.9)–(1.10). This advanced topic in implementation will be taken up in the next section along with other details.

4 Multistage details and implementations

The concept of decisions $x_k(\xi)$ being made nonanticipatively in stages k as information about a scenario ξ evolves is fundamental in stochastic programming problems and stochastic variational inequalities. The generalized multiplier vectors $y_k(\xi)$ introduced in the preceding section were likewise subjected to nonanticipativity, but nothing has yet been said about how that might specifically be valuable in optimization modeling, or how Lagrangian functions themselves might be built up in stages.

A basic source of stage structure in Lagrangians is easy to understand. The decision $x_k(\xi)$ in stage k, while subject to the nonanticipativty condition of only responding to $\xi_1, \ldots, \xi_{k-1}$ could be constrained to belong to a set that could depend on those scenario elements and the decisions already made in earlier stages. That set could be expressed by a constraint system for which $y_k(\xi)$ gives the Lagrange multipliers. We will get back to this pattern after formulating a Lagrangian stage scheme with more versatility.

For this purpose we introduce for each stage $k = 1, \ldots, N$, nonempty closed convex sets

$$\begin{array}{c} X_k(\xi) \subset I\!\!R^{n_k}, \quad Y_k(\xi) \subset I\!\!R^{m_k}, \;\text{ depending only on }\; (\xi_1, \ldots, \xi_{k-1}), \\ X(\xi) = X_1(\xi) \times \cdots \times X_N(\xi), \quad Y(\xi) = Y_1(\xi) \times \cdots \times Y_N(\xi), \end{array} \tag{4.1}$$

and a continuously differentiable Lagrangian term

$$L_k(x_1, \ldots, x_k, y_k, \xi_1, \ldots, \xi_{k-1}) \;\text{ for }\; (x_1, \ldots, x_k) \in X_1(\xi) \times \cdots \times X_k(\xi), \quad y_k \in Y_k(\xi), \tag{4.2}$$

that is convex in the $(x_1, \ldots, x_k)$ and concave in y_k. We then define the overall Lagrangian, to serve in the role of (3.7), by

$$L(x,y,\xi) = \sum\nolimits_{k=1}^N L_k(x_1, \ldots, x_k, y_k, \xi_1, \ldots, \xi_{k-1}) \;\text{ on }\; X(\xi) \times Y(\xi). \tag{4.3}$$

As explained in Section 3, the monotone stochastic variational inequality problem specified in this way expresses optimality in a saddle point format for a convex stochastic programming problem derived from (3.8) and (3.9), but in this framework that problem can be described in more detail.

Theorem 3 (multistage Lagrangian model)**.** *The stochastic programming problem associated with the Lagrangian structure in (4.1)–(4.2)–(4.3) is to*

$$\begin{array}{c}\textit{minimize over } x(\cdot)\in\mathcal{N}_n, \textit{ subject to } x_k(\xi)\in X_k(\xi), \textit{ the objective function}\\ \varphi(x(\cdot))=E_\xi[\varphi_1(x_1(\xi),\xi)+\varphi_2(x_1(\xi),x_2(\xi),\xi)+\cdots+\varphi_N(x_1(\xi),\cdots,x_N(\xi),\xi)],\end{array}\tag{4.4}$$

where the functions $\varphi_k(\cdot,\xi)$ are convex, lower semicontinuous and possibly take on ∞, and depend only on $\xi_1,\ldots,\xi_{k-1}$:

$$\varphi_k(x_1,\ldots,x_{k-1},x_k,\xi_1,\ldots,\xi_{k-1})=\sup_{y_k\in Y_k(\xi_1,\ldots,\xi_{k-1})} L_k(x_1,\ldots,x_{k-1},x_k,y_k,\xi_1,\ldots,\xi_{k-1}).\tag{4.5}$$

In applying the progressive hedging algorithm in this case as prescribed in Theorem 2, the expressions in (3.26) have stage structure in terms of

$$\begin{array}{l}u^\nu_{k,r}(x_1,\ldots,x_k,\xi)=\max\limits_{y_k\in Y_k(\xi)}\Big\{L_k(x_1,\ldots,x_k,y_k,\xi)-z^\nu_k(\xi)\cdot y_k-\dfrac{r}{2}||y_k-y^\nu_k(\xi)||^2\Big\},\\ U^\nu_{k,r}(x_1,\ldots,x_k,\xi)=\operatorname*{argmax}\limits_{y_k\in Y_k(\xi)}\Big\{L_k(x_1,\ldots,x_k,y_k,\xi)-z^\nu_k(\xi)\cdot y_k-\dfrac{r}{2}||y_k-y^\nu_k(\xi)||^2\Big\}.\end{array}\tag{4.6}$$

The two steps in (3.27) then take the form that

$$\begin{array}{l}\widehat{x}^\nu(\xi)=\operatorname*{argmin}\limits_{x\in X(\xi)}\sum_{k=1}^N\Big[u^\nu_{k,r}(x_1,\ldots,x_k,\xi)+w^\nu_k(\xi)\cdot x_k+\dfrac{s}{2}||x_k-x^\nu_k(\xi)||^2\Big],\\ \widehat{y}^\nu_k(\xi)=U^\nu_{k,r}(\widehat{x}^\nu_1(\xi),\ldots,\widehat{x}^\nu_k(\xi),\xi)\ \textit{ for }\ k=1,\ldots,N.\end{array}\tag{4.7}$$

Proof. The reduction in (4.4) takes advantage of the fact that the Lagrangian in (4.3) is separable with respect to the components y_k of y, which enables the maximization to be carried out separately in each y_k. This is also behind the creation of separate expressions for each stage k in (4.6), which then leads from (3.27) to (4.7). The proximal terms ensure in each case that the max or min is attained, uniquely. □

This result can be specialized in a rich variety of ways. An important example is the case of classical constraints of convex programming type, as follows. For each stage $k=1,\ldots,N$ let the set $Y_k(\xi_k)$ be independent of ξ_k and given simply by

$$Y_k=[0,\infty)^{s_k}\times(-\infty,\infty)^{m_k-s_k}\subset I\!\!R^{m_k},\ \text{ with elements }\ y_k=(y_{k1},\ldots,y_{k,m_k}),\tag{4.8}$$

and let the Lagrangian term L_k have the form

$$\begin{array}{l}L_k(x_1,\ldots,x_k,y_k,\xi_1,\ldots,\xi_{k-1})=\\ \qquad f_{k0}(x_1,\ldots,x_k,\xi_1,\ldots,\xi_{k-1})+\sum_{i=1}^{m_k}y_{ki}f_{ki}(x_1,\ldots,x_k,\xi_1,\ldots,\xi_{k-1}),\end{array}\tag{4.9}$$

where $f_{ki}(\cdot,\xi_1,\ldots,\xi_{k-1})$ is a continuously differentiable convex function on $X_k(\xi)=X_k(\xi_1,\ldots,x_k)$ for $i=0,1,\ldots,s_k$ and an affine function for $i=s_k+1,\ldots,m_k$. Then the objective terms in (4.4) are

$$\varphi_k(x_1,\ldots,x_k,\xi_1,\ldots,\xi_{k-1})=\begin{cases}f_{k0}(x_1,\ldots,x_k,\xi_1\ldots,\xi_{k-1})\\ \qquad\text{if } x_k\in C_k(x_1,\ldots,x_{k-1},\xi_1,\ldots,\xi_{k-1})\\ \infty\quad\text{if } x_k\notin C_k(x_1,\ldots,x_{k-1},\xi_1,\ldots,\xi_{k-1}),\end{cases}\tag{4.10}$$

for the closed convex set $C_k(x_1,\dots,x_{k-1},\xi_1,\dots,\xi_{k-1})$ defined by

$$x_k \in C_k(x_1,\dots,x_{k-1},\xi_1,\dots,\xi_{k-1}) \iff \left\{\begin{array}{l} x_k \in X_k(\xi_1,\dots,\xi_{k-1}), \\ (f_{ki}(x_1,\dots,x_{k-1},x_k,\xi_1,\dots,\xi_{k-1}) \le 0 \;\text{ for }\; i=1,\dots,s_k, \\ (f_{ki}(x_1,\dots,x_{k-1},x_k,\xi_1,\dots,\xi_{k-1}) = 0 \;\text{ for }\; i=s_k+1,\dots,m_k. \end{array}\right. \tag{4.11}$$

The expressions in (4.6) specialize then to

$$\begin{aligned} u^\nu_{k,r}(x_1,\dots,x_k,\xi) = f_{k0}(x_1,\dots,x_k,\xi_1\dots,\xi_{k-1}) \\ + \sum_{i=1}^{m_k} \lambda_\le\Big(f_{ki}(x_1,\dots,x_k,\xi_1\dots,\xi_{k-1}) - z^\nu_{ki}(\xi),\, y^\nu_{ki}(\xi)\Big) \\ + \sum_{i=s_k+1}^{m_k} \lambda_=\Big(f_{ki}(x_1,\dots,x_k,\xi_1\dots,\xi_{k-1}) - z^\nu_{ki}(\xi),\, y^\nu_{ki}(\xi)\Big), \end{aligned} \tag{4.12}$$

where

$$\lambda_\le(a,b) = \left\{\begin{array}{l} ab+\frac{1}{2}ra^2 \;\text{ if }\; r^{-1}a+b\ge 0, \\ -\frac{1}{2}rb^2 \;\text{ if }\; r^{-1}a+b\le 0, \end{array}\right. \qquad \lambda_=(a,b) = ab+\tfrac{1}{2}ra^2, \tag{4.13}$$

and on the other hand

$$y_k = U^\nu_{k,r}(x_1,\dots,x_k,\xi) \iff y_{ik} = \left\{\begin{array}{l} \mu_\le\Big(f_{ki}(x_1,\dots,x_k,\xi_1\dots,\xi_{k-1} - z^\nu_{ki}(\xi), y^\nu_{ki}(\xi)\Big) \;\text{ for }\; i=1,\dots,s_k, \\ \mu_=\Big(f_{ki}(x_1,\dots,x_k,\xi_1\dots,\xi_{k-1} - z^\nu_{ki}(\xi), y^\nu_{ki}(\xi)\Big) \;\text{ for }\; i=s_k+1,\dots,m_k, \end{array}\right. \tag{4.14}$$

where

$$\mu_\le(a,b) = \max\{0, r^{-1}a+b\}, \qquad \mu_=(a,b) = r^{-1}a+b. \tag{4.15}$$

Here the function $u^\nu_{k,r}$ is an *augmented Lagrangian* associated with L_k, and the procedure in (4.7) is then a so-called *multiplier method* in convex programming [5, 6].

Beyond this example in a classical format for convex programming, one can introduce similar details in the case of composite optimization as expressed in (1.9)–(1.10) and get essentially the same pattern.

References

[1] Bertsekas, D. P., "On penalty and multiplier methods in constrained minimization," *SIAM J. on Control* **13** (1975).

[2] Lions, J. L., and Stampacchia, G., "Variational inequalities," *Communications in Pure and Applied Mathematics* (1967), 493–519.

[3] Robinson, S. M., "Generalized equations and their solutions, I. Basic theory, point-to-set maps and mathematical programming," *Mathematical Programming Study* **10** (1979), 128–141.

[4] Rockafellar, R. T., *Convex Analysis*, Princetion University Press, 1970.

[5] Rockafellar, R. T., "The multiplier method of Hestenes and Powell applied to convex programming," *Optimization Theory and Applications* **12** (1973), 555–562.

[6] Rockafellar, R. T., "Augmented Lagrangians and applications of the proximal point algorithm in convex programming," *Mathematics of Operations Research* **1** (1976), 97–116.

[7] Rockafellar, R. T., "Monotone operators and the proximal point algorithm." *SIAM J. on Control and Optimization* **14** (1976), 877–898.

[8] Rockafellar, R. T., "Augmented Lagrangians and applications of the proximal point algorithm in convex programming," *Mathematics of Operations Research* **1** (1976), 97–116.

[9] Rockafellar, R. T., "Extended nonlinear programming," in *Nonlinear Optimization and Related Topics* (G. Di Pillo and F. Giannessi, eds.), Kluwer, 1999, 381–399. Downloadable as article #175 from http://math.washington.edu/ rtr/mypage.html.

[10] Rockafellar, R. T., and Sun, J., "Solving monotone stochastic variational inequalities and complementarity problems by progressive hedging," *Mathematical Programming B* (2018)

[11] Rockafellar, R. T., and Wets, R. J-B, "Scenarios and policy aggregation in optimization under uncertainty." *Mathematics of Operations Research* **16** (1991), 119–147.

[12] Rockafellar, R. T., and Wets, R. J-B, *Variational Analysis*, No. 317 in the series *Grundlehren der Mathematischen Wissenschaften*, Springer-Verlag, 1998.

[13] Rockafellar, R. T., and Wets, R. J-B, "Stochastic variational inequalities: single-stage to multistage." *Mathematical Programming B* **165** (2017), 291–330.

[14] Spingarn, J. E., "Partial inverse of a monotone operator," *Applied Mathematics and Optimization* **10** (1983), 247–265.