CrossMark

Stochastic variational inequalities: single-stage to multistage

R. Tyrrell Rockafellar[1] · **Roger J-B Wets[2]**

Abstract Variational inequality modeling, analysis and computations are important for many applications, but much of the subject has been developed in a deterministic setting with no uncertainty in a problem's data. In recent years research has proceeded on a track to incorporate stochasticity in one way or another. However, the main focus has been on rather limited ideas of what a stochastic variational inequality might be. Because variational inequalities are especially tuned to capturing conditions for optimality and equilibrium, stochastic variational inequalities ought to provide such service for problems of optimization and equilibrium in a stochastic setting. Therefore they ought to be able to deal with multistage decision processes involving actions that respond to increasing levels of information. Critical for that, as discovered in stochastic programming, is introducing nonanticipativity as an explicit constraint on responses along with an associated "multiplier" element which captures the "price of information" and provides a means of decomposition as a tool in algorithmic developments. That idea is extended here to a framework which supports multistage optimization and equilibrium models while also clarifying the single-stage picture.

✉ R. Tyrrell Rockafellar
rtr@uw.edu

Roger J-B Wets
rjbwets@ucdavis.edu

[1] Department of Mathematics, University of Washington, Seattle, WA 98195-4350, USA

[2] Department of Mathematics, University of California-Davis, Davis, CA 95615, USA

 Springer

1 Introduction

Solving a "variational inequality" is a problem that originated in an infinite-dimensional setting with partial differential operators subjected to one-sided constraints inside domains or on their boundaries. Independently the same concept, viewed as solving a "generalized equation," arose in the finite-dimensional context of optimality conditions in nonlinear programming. Special versions called "complementarity problems" gained attention even earlier on those lines. See e.g. [5,8,17].

The theory has since undergone enormous development from every angle, but largely in deterministic mode. More recently there have been efforts to extend it to models where random variables enter into the data. A challenge is to understand then what a "stochastic variatonal inequality" might properly be in general. Efforts so far have mainly been directed at formulations proposed in a rather limited setting. Our purpose here is to address the subject more broadly and develop a formulation that can cover vastly more territory. Roughly speaking, previous approaches have their motivations in special kinds of "single-stage" modeling in optimization and equilibrium, at best, whereas our appoach will also encompass "multistage" models that allow for response to increasing levels of information. Without such elaboration the dynamics that are essential to stochastic decision processes, whether by one agent or a multiplicity in a gamelike setting, can't be captured.

Response to information is like a "policy" in dynamic programming and thus inevitably involves a search for the right functions of some kind. A basic issue is how that may conveniently be modeled. The tried-and-true technology in advanced stochastics is that of a probability space supplied with various "information fields" with respect to which a response function can be constrained to be "measurable." Here we follow an easier pattern by tying information to finitely many scenarios which progressively diverge from each other. That lets us avoid the complications of infinite-dimensional function spaces so as to present the modeling ideas and examples without getting into distracting technicalities.[1]

The feature that distinguishes and drives our approach, even for single-stage modeling, is the treatment of *nonanticipativity of response*. Nonanticipativity is the requirement that decisions can't be based on information not yet known, and it is inescapable in determining "policies" that are actually implementable. Our treatment, in a pattern that goes back to the foundations of stochastic programming, cf. [22,23], formulates nonanticipativity explicitly as a constraint on response functions which can be dualized by "multipliers." Such nonanticipativity multipliers, which enable

---

[1] Of course a full probability setting is important to have available in the long run, but we plan to deal with that in a follow-up article.

decomposition into a separate problem for each scenario, have been understood in stochastic programming as furnishing the shadow price of information. They were the basis there of one of the most effective solution methods, the "progressive hedging algorithm" of [24]. They might well be useful also in solving stochastic variational inequality problems, but we don't get into that here.

The rest of this introduction is devoted to the current state of affairs in research on stochastic variational inequalities, which requires first recalling key concepts and examples in a nonstochastic setting. The examples help to demonstrate the breadth of variational inequality applications that needs to be encompassed when stochasticity is incorporated.

Our different approach to single-stage stochastic variational inequalities, in contrast to the approaches up to now, is laid out in Sect. 2. The multistage extension of it follows in Sect. 3. Section 4 offers examples of applications involving expectation functions and constraints. Potential applications even to nonconvex constraints are indicated in Sect. 5.

Background in nonstochastic variational inequalities Most simply, in a standard deterministic framework in $\mathbb{R}^n$ to start with, a *variational inequality* condition, or *generalized equation* is a relation of the form

$$- F(x) \in N_C(x) \tag{1.1}$$

in which F is a continuous single-valued mapping from $\mathbb{R}^n$ to $\mathbb{R}^n$ and $N_C(x)$ is the normal cone to a nonempty closed convex set $C \subset \mathbb{R}^n$ at the point $x \in C$, defined by

$$v \in N_C(x) \iff x \in C \text{ and } \langle v, x' - x \rangle \leq 0 \text{ for all } x' \in C. \tag{1.2}$$

The *problem* associated with the variational inequality (1.1) is to find an x that satisfies it. Here $\langle \cdot, \cdot \rangle$ denotes usual inner product in $\mathbb{R}^n$ for now, but it will be important later to understand that some other expression obeying the axioms of an inner product could serve just as well, since this comes up in adaptations to stochastic structure. Passing to a different inner product changes the normal cone and thus the meaning of the variational inequality.

By putting $-F(x)$ in place of v in (1.2), one arrives at the system of inequalities behind the "variational inequality" name for (1.1). On the other hand, by taking $C = \mathbb{R}^n$ one gets the reduction of (1.1) to the vector equation $F(x) = 0$ which underlies the name "generalized equation." The case of a *complementarity* problem corresponds to C being the nonnegative orthant $\mathbb{R}^n_+$. Rich applications set up as complementarity problems can be seen in [10], but variational inequalities provide more flexibility and easier connection with a variety of modular structures which can be assembled to cover a given application.[2] Although they can often be reduced to complementarity

---

[2] Generalizations beyond (1.1) are available. The set-valued normal cone mapping N_C is the subdifferential mapping $\partial \delta_C$ associated with the indicator δ_C, which is a closed proper convex function. A natural step therefore is to replace N_C in (1.1) by the subdifferential mapping ∂f for any closed proper convex function f on $\mathbb{R}^n$ (which goes back to the earliest days of the subject) or perhaps any set-valued mapping T that, like such ∂f, is maximal monotone. It is possible also to take F to be set-valued, built out of other

and solved that way in principle, there is a huge literature for solving variational inequality problems directly, for instance by minimizing some residual or working with an equivalent nonsmooth equation.

Other instances of (1.1) relate to optimization models and deserve a brief review because they help with motivation. The *elementary optimization* case concerns the first-order necessary condition for minimizing a continuously differentiable function $f(x)$ over C (cf. [25, 6.12]), namely:

$$-\nabla f(x) \in N_C(x). \tag{1.3}$$

This fits (1.1) with F being the gradient mapping ∇f. Of course C might be specified by a constraint system like

$$C = \{ x \in X \mid G(x) \in K \} \text{ with } G(x) = (g_1(x), \ldots, g_m(x)) \tag{1.4}$$

for a closed convex set $X \subset \mathbb{R}^n$, a closed convex cone $K \subset \mathbb{R}^m$, and continuously differentiable functions g_i. Formulas of variational analysis can take over then to provide details about $N_C(x)$. Under a constraint qualification, having $v \in N_C(x)$ for C as in (1.4) corresponds to having

$$v = \sum_{i=1}^m y_i \nabla y_i g_i(x) + u \text{ for some } u \in N_X(x) \text{ and some}$$
$$y = (y_1, \ldots, y_m) \in Y \text{ with } G(x) \in N_Y(y), \text{ where } Y = K^*. \tag{1.5}$$

(Here K^* denotes the polar of K. See [25, 6.14, 6.15] for detailed development of Lagrange multipliers in this mode.) In terms of the Lagrangian function

$$L(x, y) = f(x) + \sum_{i=1}^m y_i g_i(x) \tag{1.6}$$

the combination of (1.3) and (1.4) takes the appealing form

$$-\nabla_x L(x, y) \in N_X(x), \qquad \nabla_y L(x, y) \in N_Y(y), \tag{1.7}$$

which nicely covers the Karush–Kuhn–Tucker conditions of nonlinear programming in particular.[3] The double condition (1.7) can be written equivalently as

$$-H(z) \in N_Z(z) \text{ for } z = (x, y), \quad Z = X \times Y, \quad H(z) = (\nabla_x L(x, z), -\nabla_y L(y, z)) \tag{1.8}$$

Footnote 2 continued
subdifferential mappings, say. Despite the genuine interest in such generalizations and their eventual importance in applications, the fundamental version in (1.1) will be our touchstone here.

[3] Such a variational inequality representation of the KKT conditions was the "generalized equation" which inspired Robinson in his pioneering work [17]. In [7] the term "generalized equation" is applied also to the case of (1.1) in which N_C is replaced by a more general set-valued mapping.

and thus it actually comprises a single variational inequality. We will speak of (1.7) as a *Lagrangian* variational inequality — even when L has a different form than (1.6) (as long as L is continuously differentiable).[4] This example also signals something that might be overlooked at first but needs to be kept in mind: the solution to a variation inequality problem might well involve not just "decision elements" (in the language we fall back on for convenience) but also "multipliers" tied to constraints.

Another source of interest in variational inequalities is their capability of representing "equilibrium." Already in (1.7) with general L we have an expression of equilibrium for a two-person zero-sum game in which one player wishes to minimize $L(x, y)$ with respect to $x \in X$ while the other player wishes to maximize $L(x, y)$ with respect to $y \in Y$. Equilibrium of Nash type in a game with players $i = 1, \ldots, m$ can similarly be represented. Suppose player i wishes choose $x_i \in C_i$ so as to minimize $f_i(x_i, x_{-i})$, where (in the standard notation of game theory) x_{-i} stands for the choices of the other players. Suppose C_i is a closed convex set in $\mathbb{R}^{n_i}$ and f_i is continuously differentiable. Then the first-order optimality condition for player i is $-\nabla_{x_i} f_i(x_i, x_{-1}) \in N_{C_i}(x_i)$, and an equilibrium $x = (x_1, \ldots, x_m)$ among the players is expressed by

$$-F(x) \in N_C(x) \text{ for } C = C_1 \times \cdots \times C_m \text{ and}$$
$$F(x) = (\nabla_{x_1} f_1(x_1, x_{-1}), \ldots, \nabla_{x_m} f_1(x_m, x_{-m})). \tag{1.9}$$

This is the *elementary equilibrium* case of a variational inequality.[5] Other versions of equilibrium can likewise be set up as instances of (1.1). For example, this has been carried out in [14] for a classical economic model of equilibrium in prices, supply and demand.

A solution x to (1.1) is sure to exist, in particular, when C is bounded. The set of all solutions, if any, is always closed, and in the *monotone* case of (1.1), where F is monotone relative to C, meaning that

$$\langle F(x') - F(x), x' - x \rangle \geq 0 \text{ for all } x, \quad x' \in C, \tag{1.10}$$

it must also be convex. Under *strict* monotonicity, which requires the inequality in (1.10) to be strict unless $x = x'$, there can be at most one solution. For this and more see [7, Theorems 2A.1 and 2F.1].

Monotone variational inequalities relate closely to *convex* optimization and its methodology as illustrated by (1.3) when $f(x)$ is convex with respect to $x \in C$ and by (1.6) when $L(y, z)$ is convex in $y \in Y$ for each $z \in Z$, but concave in $z \in Z$ for each $y \in Y$. For monotone variational inequalities, other criteria than the boundedness of C are available for the existence of solutions (see [25, Chapter 12]), and the library of

---

[4] For other L it can characterize first-order optimality in an amazingly large class of problems which can even be "nonsmooth"; see [20], [25, 11.46–11.47].

[5] The equilibrium in (1.9) is a true Nash equilibrium when $f_i(x_i, x_{-i})$ is convex in x_i, so that first-order optimality in its local sense coincides with global optimality. However, equilibrium is an apt term even without the convexity, since it's hardly reasonable to burden agents with mastering global minimization in a context where the actions of competitors render perceptions local at best anyway.

algorithms for solving the problem is much richer. A solution strategy that doesn't rely on monotonicity is expressing (1.1) as the nonsmooth equation $P_C(x - F(x)) - x = 0$, where P_C is the projection onto C, and applying an algorithm in that context. Another approach is in [6], which introduces Newton-like iterations in terms of subproblems that are "linearized" variational inequalities. Additional background in that area can be found in the book [7].

As in the examples above, the rules obeyed by normal cones to convex sets are valuable in putting variational inequalities together. Besides the "multiplier rule" (1.5) for the constraint system (1.4) there is the product rule that

$$N_{C_1 \times \cdots \times C_m}(x_1, \ldots, x_m) = N_{C_1}(x_1) \times \cdots \times N_{C_m}(x_m) \tag{1.11}$$

which directly entered game model (1.9) and, before that, the identification of the Lagrangian variational inequality (1.7) with (1.8). Another rule utilized in reaching (1.7) was that

for a closed convex cone K and its polar K^*: $y \in N_K(u) \iff u \in N_{K^*}(y)$.
$$\tag{1.12}$$

Still another rule, concerning a intersection of closed convex sets C_1 and C_2, hasn't yet been needed but will be important later:

$$N_{C_1 \cap C_2}(x) \supset N_{C_1}(x) + N_{C_2}(x) = \left\{ v_1 + v_2 \,\middle|\, v_1 \in N_{C_1}(x), \quad v_2 \in N_{C_2}(x) \right\},$$
$$\text{and equality holds if the sets } C_i \text{ are polyhedral or have ri } C_1 \cap \text{ri } C_2 \neq \emptyset,$$
$$\tag{1.13}$$

where ri marks the relative interior of a convex set. A reference for this is [18, Corollary 23.8.1].

Uncertainty and how to think about it Extensions to allow for stochasticity in the formulation of a variational inequality can be portrayed as involving dependence on elements ξ of a probability space Ξ. In this paper we focus on Ξ being a finite set of "scenarios" ξ, each having a nonzero probability $p(\xi)$ as assigned by a function p with $\sum_{\xi \in \Xi} p(\xi) = 1$. Either F or C, or both, can then taken to be scenario-dependent as $F(x, \xi)$ and $C(\xi)$, but what exactly should be made of that in *problem formulation* in light of examples such as above?

A central question is whether ξ is supposed to be known before, or only after, x is finalized. If before, we are faced with a collection of individual variational inequality conditions, one for each ξ, and can consider ξ-dependent solutions $x(\xi)$ to them:

$$- F(x(\xi), \xi) \in N_{C(\xi)}(x(\xi)) \text{ for all } \xi \in \Xi. \tag{1.14}$$

We are engaged then with a *response function*

$$x(\cdot) : \xi \mapsto x(\xi) \text{ with } \xi \in \Xi, \ x(\xi) \in \mathbb{R}^n, \tag{1.15}$$

that acts on the information provided by knowing ξ. Such a function may be useful in a larger picture, but (1.14) is not what people have in mind as a possible formulation of a (single) "stochastic variational inequality." For that, the focus is on the opposite situation, where an x has to be fixed in advance of knowing ξ but still should cope to some degree with the uncertainty in $F(x, \xi)$ and $C(\xi)$.

That situation is easier to think about when we just have a fixed set C instead of $C(\xi)$ and all the uncertainty is in $F(x, \xi)$. It is tempting then to take the *expected value* (EV) approach and, following [11–13,26,27] and others, study the condition

$$- \overline{F}(x) \in N_C(x) \text{ for } \overline{F}(x) = E_\xi[F(x, \xi)]. \tag{1.16}$$

as a "stochastic variational inequality" which specializes (1.1) to expectational structure. Motivation for (1.16) comes in particular from the fact that

$$\overline{F}(x) = \nabla \bar{g}(x) \text{ when } F(x, \xi) = \nabla_x g(x, \xi) \text{ and } \bar{g}(x) = E_\xi[g(x, \xi)], \tag{1.17}$$

in which case (1.16) corresponds to first-order optimality in minimizing $\bar{g}$ over C.

An alternative which has received widespread attention in allowing for uncertain $C(\xi)$ and not relying on expectations, is looking for an x such that

$$- F(x, \xi) \in N_{C(\xi)}(x) \text{ for all } \xi \in \Xi, \tag{1.18}$$

which entails $x \in \bigcap_{\xi \in \Xi} C(\xi)$. This is the starting point for the *expected residual minimization* (ERM) approach followed in [1–4,9,15,16,28,29]. That line of research, however, has to face the fact that the existence of a solution x is highly unlikely—which suggests looking instead for an "approximate solution" in some sense. The reason why an x satisfying (1.18) is hardly imaginable can be seen from the examples of variational inequalities reviewed above. In the elementary optimization case with $F(x, \xi) = \nabla_x g(x, \xi)$, one is demanding in (1.18) that x satisfy the first-order optimality condition for minimization of $g(\cdot, \xi)$ over $C(\xi)$ simultaneously for every ξ. Ordinarily, no one looks for an x that solves two different optimization problems at the same time, much less one that solves all problems in a possibly large collection indexed by $\xi \in \Xi$. In the Lagrangian case of a variational inequality, one would be asking for not just the optimal solution x but also the associated Lagrange multiplier vector y in the problem to be the same for all ξ.

In reacting to that by settling for an x that only "approximately" solves all the individual variational inequalities in (1.18) simultaneously (one for each ξ), researchers have turned to minimizing the expectation of some "residual" quantity as a random variable with respect to ξ, hence the ERM name. A natural justification would be that there is a single underlying nonstochastic variational inequality with "noisy data."[6]

---

[6] An analogy with overdetermined systems of linear equations can be considered, where a single solution x ought to exist but the equations disagree slightly because of measurement errors in the coefficients, rendering their simultaneous solution impossible. One might think similarly of collection of linear programming problems, say, which are identical except for "coefficient noise" and look for an approximate common solution.

But it doesn't seem quite right to refer to solving (1.18) as a solving a (single) stochastic variational inequality.

Unaddressed by the EV and ERM approaches is the prospect of a stochastic variational inequality being a model that covers optimization and equilibrium problems in situations where decisions have to interact dynamically with the availability of information. The information may be tied to scenarios based on the realizations of some random variables, and when more is known those realizations, opportunities for recourse decisions might need to be provided, perhaps in a number of stages.

2 Single-stage modeling with nonanticipativity constraints

There is a way around the impasse over uncertain $C(\xi)$ which can also open the door later to more complex dynamics than just fixing one x before observing one ξ. It had a prominent role in the theory of stochastic programming (cf. [22,23]) and in particular in the development of one of the main solution methods in that subject, the "progressive hedging algorithm" in [24].

Nonanticipativity of response The key idea, even if at first it seems only to bring unnecessary hardship, is to think of the "solution" to be targeted not as a vector x but as a response function $x(\cdot) : \xi \to x(\xi)$, and then to *constrain* that function to always give the same response, i.e., to be a constant function, thus furnishing a single element x of $\mathbb{R}^n$ in the end after all. This condition on $x(\cdot)$, saying that foreknowledge of ξ can't be a basis for response, is *nonanticipativity*. The advantage gained by such a constraint formulation is that a "multiplier" can be attached, which can be a powerful tool in both modeling and computation.

In order to explain that better, we have to pass from $\mathbb{R}^n$ to a space of functions, namely

$$\mathcal{L}_n = \mathcal{L}_n(\Xi, p) = \text{ the collection of all functions } x(\cdot) : \Xi \to \mathbb{R}^n, \qquad (2.1)$$

but this is not as bothersome as might be imagined. Since there are only finitely scenarios $\xi \in \Xi$, each with a probability $p(\xi) > 0$, those elements could be indexed from 1 to s, say, and the responses $x(\xi) \in \mathbb{R}^n$ could thereby be lined up into a "supervector" in $(\mathbb{R}^n)^s$. No doubt this would be the best tactic when it comes to numerical work, but not necessarily for a theory that promotes insights and structural understanding.

Anyway, although $\mathcal{L}_n$ is effectively identified with $(\mathbb{R}^n)^s$ by the "supervector" tactic, a distinction in the inner product comes up. We don't want to operate in $\mathcal{L}_n$ with the inner product that would transfer from $(\mathbb{R}^n)^s$ through that identification, but rather with the *expectational* inner product

$$\langle v(\cdot), x(\cdot) \rangle = E_\xi[\langle v(\xi), x(\xi) \rangle] = \sum_{\xi \in \Xi} p(\xi) \langle v(\xi), x(\xi) \rangle \qquad (2.2)$$

with associated norm $||x(\cdot)|| = (E_\xi[\langle x(\cdot), x(\cdot) \rangle])^{1/2}$; here $\langle x(\xi), w(\xi) \rangle$ is the usual inner product between two vectors in $\mathbb{R}^n$.

Proceeding in that framework with our ingredients $F(x, \xi)$ and $C(\xi)$, we introduce the nonempty closed convex subset set $\mathcal{C}$ of $\mathcal{L}_n$ defined by

$$\mathcal{C} = \big\{\, x(\cdot) \in \mathcal{L}_n \,\big|\, x(\xi) \in C(\xi) \text{ for all } \xi \in \Xi \,\big\} \tag{2.3}$$

and the linear subspace $\mathcal{N}$ of $\mathcal{L}_n$ defined by

$$\mathcal{N} = \big\{\, x(\cdot) \in \mathcal{L}_n \,\big|\, x(\xi) \text{ is the same for all } \xi \in \Xi \,\big\}. \tag{2.4}$$

This is the *nonanticipativity* subspace; the nonanticipativity constraint can be expressed by $x(\cdot) \in \mathcal{N}$.

Where we were previously thinking about a vector x belonging to the intersection of all the sets $C(\xi)$ as in (1.18), we are now thinking about a function $x(\cdot)$ in $\mathcal{C} \cap \mathcal{N}$. That amounts to the same thing from one angle, but it provides a really different platform for a variational inequality, one which is better able to draw on historical advances in stochastic optimization.

To formulate a variational inequality in $\mathcal{L}_n$ with respect to the nonempty closed convex set $\mathcal{C} \cap \mathcal{N}$ in parallel mode to (1.1), we also need a continuous mapping $\mathcal{F} : \mathcal{L}_n \to \mathcal{L}_n$. We get it from the vectors $F(x, \xi)$ by[7]

$$\mathcal{F}(x(\cdot)), \text{ for } x(\cdot) \in \mathcal{L}_n, \text{ is the function in } \mathcal{L}_n \text{ that takes } \xi \in \Xi \text{ to } F(x(\xi), \xi) \in \mathbb{R}^n. \tag{2.5}$$

The continuous dependence of $F(x, \xi)$ on $x \in \mathbb{R}^n$ makes the mapping $\mathcal{F}$ be continuous from $\mathcal{L}_n$ to $\mathcal{L}_n$.

Definition 2.1 (*SVI basic form, single-stage*) With respect to $F(x, \xi)$ and $C(\xi)$, the condition

$$- \mathcal{F}(x(\cdot)) \in N_{\mathcal{C} \cap \mathcal{N}}(x(\cdot)), \tag{2.6}$$

is the associated *single-stage stochastic variational inequality in basic form*.

Clearly (2.6), entailing $x(\cdot) \in \mathcal{C} \cap \mathcal{N}$, fits the definition of a variational inequality as originally presented except for being articulated in $\mathcal{L}_n$ instead of $\mathbb{R}^n$. Its precise meaning depends on the normal cones to the convex set $\mathcal{C} \cap \mathcal{N}$ in $\mathcal{L}_n$, which can be calculated by rules of convex analysis. The important thing is that such analysis, taken up next, leads to an alternative expression of (2.6) which incorporates multiplier elements.

Dualization The nonanticipativity constraint $x(\cdot) \in \mathcal{N}$ can be dualized in a way that invokes the elements $w(\cdot)$ of another linear subspace of $\mathcal{L}_n$, defined by

$$\mathcal{M} = \big\{\, w(\cdot) \in \mathcal{L}_n \,\big|\, E[w(\cdot)] = 0 \,\big\}. \tag{2.7}$$

---

[7] Note that, in this kind of notation, $\mathcal{F}(x(\cdot))(\xi)$ could be used for $F(x(\xi), \xi)$, but $\mathcal{F}(x(\xi))$ wouldn't make any sense, since $\mathcal{F}$ acts on elements of $\mathcal{L}_n$, not on vectors in $\mathbb{R}^n$.

It's easy to verify that $\mathcal{M}$ is the orthogonal complement of $\mathcal{N}$ with respect to the expectational inner product (2.2):

$$w(\cdot) \in \mathcal{M} \iff E_\xi \langle x(\xi), w(\xi) \rangle = 0 \text{ for all } x(\cdot) \in \mathcal{N},$$
$$x(\cdot) \in \mathcal{N} \iff E_\xi \langle x(\xi), w(\xi) \rangle = 0 \text{ for all } w(\cdot) \in \mathcal{M}. \tag{2.8}$$

The functions $w(\cdot)$ in $\mathcal{M}$ enter as *nonanticipativity multipliers* in the following condition, which will partner closely with our "basic" variation inequality (2.6).

Definition 2.2 (*SVI extensive form, single-stage*) With respect to $F(x, \xi)$ and $C(\xi)$, the condition

$$x(\cdot) \in \mathcal{N} \text{ and there exists } w(\cdot) \in \mathcal{M} \text{ such that}$$
$$-F(x(\xi), \xi) - w(\xi) \in N_{C(\xi)}(x(\xi)) \text{ for all } \xi \in \Xi. \tag{2.9}$$

is the associated *single-stage stochastic variational inequality in extensive form*.

Although the condition in this definition might not seem to warrant being called a (single) variational inequality, the designation is justified because of the following theorem, which says that (2.9) is essentially just another way of expressing (2.6).

Theorem 2.3 (basic-extensive equivalence, single-stage). *If $x(\cdot)$ solves* (2.9)*, then $x(\cdot)$ solves* (2.6)*. Conversely, if $x(\cdot)$ solves* (2.6)*, then $x(\cdot)$ is sure also to solve* (2.9) *if*

$$\text{there exists some } \hat{x}(\cdot) \in \mathcal{N} \text{ such that } \hat{x}(\xi) \in \text{ri } C(\xi) \text{ for all } \xi \in \Xi. \tag{2.10}$$

This constraint qualification is unnecessary if the sets $C(\xi)$ are all polyhedral.

Proof We are dealing in (2.6) with normals to the intersection of two closed convex sets, $\mathcal{C}$ and $\mathcal{N}$, and can apply the calculus rule in (1.13) to get a handle on $N_{\mathcal{C} \cap \mathcal{N}}(x(\cdot))$. By this rule,

$$N_{\mathcal{C} \cap \mathcal{N}}(x(\cdot)) \supset N_{\mathcal{C}}(x(\cdot)) + N_{\mathcal{N}}(x(\cdot)) = \{ v(\cdot) + w(\cdot) \mid v(\cdot) \in N_{\mathcal{C}}(x(\cdot)),$$
$$w(\cdot) \in N_{\mathcal{N}}(x(\cdot)) \} \text{ always}, \tag{2.11}$$

and under further conditions this inclusion becomes an equation. For $x(\cdot) \in \mathcal{C}$ in (2.3), the elements of $N_{\mathcal{C}}(x(\cdot))$ are by definition the functions $v(\cdot) \in \mathcal{L}_n$ such that

$$0 \geq \langle v(\cdot), x(\cdot) \rangle = \sum_{\xi \in \Xi} p(\xi) \langle v(\xi), x'(\xi) - x(\xi) \rangle \text{ for all } \xi \in \Xi \text{ and } x'(\cdot) \in \mathcal{C}. \tag{2.12}$$

Because $p(\xi) > 0$, this is equivalent to having, for each ξ,

$$0 \geq \langle v(\xi), x'(\xi) - x(\xi) \rangle \text{ for all } x'(\xi) \in C(\xi),$$

or in other words, $v(\xi) \in N_{C(\xi)}(x(\xi))$. Thus,

$$N_{\mathcal{C}}(x(\cdot)) = \big\{ v(\cdot) \in \mathcal{L}_n \,\big|\, v(\xi) \in N_{C(\xi)}(x(\xi)) \text{ for all } \xi \in \Xi \big\}. \qquad (2.13)$$

To determine the elements $w(\cdot)$ of $N_{\mathcal{N}}(x(\cdot))$, the definition of the normal cone can be applied as in (2.12) with $w(\cdot)$ and $\mathcal{N}$ in place of $v(\cdot)$ and $\mathcal{C}$, but because $\mathcal{N}$ is a subspace, any $x'(\cdot) \in \mathcal{N}$ also has $-x'(\cdot) \in \mathcal{N}$. The inequality turns then into the requirement that $0 = \langle w(\cdot), y(\cdot) \rangle$ for all $y(\cdot) \in \mathcal{N}$ (inasmuch as having $x'(\cdot)$ range over all of $\mathcal{N}$ is the same as having $y(\cdot) = x'(\cdot) - x(\cdot)$ range over all of $\mathcal{N}$). Thus, the elements of $N_{\mathcal{N}}(x(\cdot))$ are the elements of $\mathcal{L}_n$ that are orthogonal to $\mathcal{N}$, which by (2.8) are the elements of $\mathcal{M}$:

$$N_{\mathcal{N}}(x(\cdot)) = \mathcal{M} \text{ for any } x(\cdot) \in \mathcal{N}. \qquad (2.14)$$

From (2.13) and (2.14) and the definition (2.5) of $\mathcal{F}(x(\cdot))$, we see that

$$-\mathcal{F}(x(\cdot)) \in N_{\mathcal{C}}(x(\cdot)) + N_{\mathcal{N}}(x(\cdot)) \iff$$
there exists $w(\cdot) \in \mathcal{M}$ such that $-F(x(\xi), \xi) - w(\xi) \in N_{C(\xi)}(x(\xi))$ for all $\xi \in \Xi$.
$$\qquad (2.15)$$

According to (2.11), this always implies (2.6), so the claim that any solution $x(\cdot)$ to (2.9) is a solution to (2.6) is verified.

The claim in the opposite direction rests on the inclusion in (2.11) being an equation, which through application of (1.13) holds if $\operatorname{ri} \mathcal{C} \cap \operatorname{ri} \mathcal{N} \neq \emptyset$, but this condition is unnecessary when the convexity is polyhedral. We have

$$\operatorname{ri} \mathcal{N} = \mathcal{N} \text{ and } \operatorname{ri} \mathcal{C} = \big\{ x(\cdot) \,\big|\, x(\xi) \in \operatorname{ri} C(\xi) \text{ for all } \xi \in \Xi \big\}, \qquad (2.16)$$

the first because $\mathcal{N}$ is a subspace of $\mathcal{L}_n$ (subspaces are by definition their own relative interiors) and the second because $\mathcal{C}$ is essentially the product of the sets $C(\xi)$ with respect to the identification of $\mathcal{L}_n$ with the product of copies of $\mathbb{R}^n$, one for each ξ; cf. [18, page 49]. This product description makes clear also that $\mathcal{C}$ is polyhedral if and only if every $C(\xi)$ is polyhedral. Of course $\mathcal{N}$, as a subspace, is polyhedral in particular. The criterion for having the inclusion in (2.11) hold as an equation comes down that way to the conditions given in the theorem, and the proof is complete. $\square$

Stochastic decomposition Because having $x(\cdot) \in \mathcal{N}$ refers to the existence of an $x \in \mathbb{R}^n$ such that $x(\xi) \equiv x$, one might think that such a simplification ought to have been incorporated in the statement of (2.9), making the problem come out as

find $x \in \mathbb{R}^n$ and $w(\cdot) \in \mathcal{M}$ such that $-F(x, \xi) - w(\xi) \in N_{C(\xi)}(x)$ for all $\xi \in \Xi$.
$$\qquad (2.17)$$

That, however, would obscure the *decomposition* that the last part of (2.9) reveals, as explained next.

Consider a situation in which an $w(\cdot) \in \mathcal{M}$ is at hand and we want to see whether an $x(\cdot) \in \mathcal{N}$ can be associated with it in order to have a solution to (2.9). This can be tackled by solving, for each scenario ξ, the variational inequality subproblem in $\mathbb{R}^n$ for $C(\xi)$ and the function $F(\cdot, \xi)$ to get $x(\xi)$ and then checking whether $x(\cdot) \in \mathcal{N}$, i.e., whether the scenario solutions all turn out to be the same (or in the presence of nonuniqueness can be selected to be the same). If we somehow had the right $w(\cdot)$ to start with, and the subproblem had a unique solution for one of the scenarios ξ, that $x(\xi)$ would have to give the constant value desired for $x(\cdot)$!

More realistically one can envision computational schemes in which tentative nonanticipativity multipliers $w(\cdot)$ are tried out and adjusted while the corresponding response functions $x(\cdot)$ get closer to being constant functions. That is the mechanism of the progressive hedging algorithm in multistage stochastic programming [24]. Such schemes will not be explored here, because of an already overloaded agenda. Nevertheless, the efforts we put into problem formulation in this paper are definitely aimed also at laying a foundation for possible numerical developments.

Relation to the ERM and EV approaches Through (2.9), as a restatement of (2.6), the connection between stochastic variational inequalites as proposed here and the problems studied in the ERM and EV approaches becomes much clearer. Obviously (2.9), in its simplified expression (2.17), differs from the ERM problem in (1.18) through the presence of the vectors $w(\xi)$. Those vectors, in modifying the functions $F(\cdot, \xi)$ make a huge difference. Although there is little hope that an x solving (1.18) even exists, aside from rare circumstances, the existence of $x(\xi) \equiv x$ solving (2.6)/(2.9) is readily ensured.[8] This indicates that (2.10) might be seen as the "fix" needed to bring viability to (1.18) by ensuring the existence of an exact solution.

At the same time, however, the EV problem in (1.16) emerges via (2.9) as the problem to which our basic stochastic variational inequality in (2.6) reduces when $C(\xi) \equiv C$. The elementary rule clarifying that is the following:[9]

$$\text{for a convex cone } K \text{ and a function } z(\cdot) \in \mathcal{L}_n, \text{ one has } E[z(\xi)] \in K$$
$$\Longleftrightarrow \ \exists \, w(\cdot) \text{ with } E[w(\xi)] = 0 \text{ such that } z(\xi) - w(\xi) \in K \text{ for all } \xi.$$
(2.18)

Applying this to $K = N_C(x)$ and $z(\xi) = -F(x, \xi)$ turns (2.9) into (1.16).

It's worth noting that the condition (1.14) which we associated with the case of ξ being known before a decision has to be made, while seemingly a collection of separate variational inequalities indexed by ξ, can also be viewed through (2.13) as a single variational inequality in $\mathcal{L}_n$:

$$- \mathcal{F}(x(\cdot)) \in N_{\mathcal{C}}(x(\cdot))$$
(2.19)

---

[8] An immediate criterion, as indicated in the background discussion, is the boundedness of $\mathcal{C}$, which corresponds to the boundedness of the sets $C(\xi)$. Later, in the multistage development, broader criteria will be presented.

[9] Get the second condition from the first by taking $w(\xi) = z(\xi) - E[z(\xi)]$; get the first from the second by taking expectations.

The collection of conditions for each ξ in (2.9) corresponds likewise to $-\mathcal{F}(x(\cdot)) - w(\cdot) \in N_{\mathcal{C}}(x(\cdot))$, and indeed this is how it was derived in the proof of Theorem 2.3. This affords another way of looking at the SVI in extensive form as translating the SVI in basic form into a mode where, by modifying the given $\mathcal{F}(\cdot) : \mathcal{L}_n \to \mathcal{L}_n$ by adding $w(\cdot)$ to it, the constraint of "deciding" before "observing" is relaxed to allow "decision" to follow "observation."

The need for more than single-stage models Although (2.6) furnishes the basic form that a stochastic variational inequality should have, in our opinion, with respect to uncertain $F(x, \xi)$ and $C(\xi)$, it only covers the case of a single x having to be fixed before the realization of a single scenario ξ. Our goal in this paper lies beyond just this. We want to encompass situations where "decisions" in time can alternate with "observations" in time in a multistage process. For that purpose nonanticipativity constraints on response functions are ever more essential.

Two-stage stochastic optimization can serve as to preview the issues behind the approach we will undertake in Sect. 3. Suppose we have a pattern in which an initial decision $x_1 \in \mathbb{R}^{n_1}$ must be taken before ξ is known, but afterward a recourse decision $x_2(\xi) \in \mathbb{R}^{n_2}$ can be taken which is able to respond to the information in ξ. There are many variants of this, but in keeping to the bare essentials let us suppose that there is a convex set $C(\xi)$ in $\mathbb{R}^{n_1} \times \mathbb{R}^{n_2}$ to which $(x_1, x_2(\xi))$ must belong (how it might be specified need not enter for the moment), and we are concerned with minimizing the expected value of a cost expression $g(x_1, x_2(\xi), \xi)$. What would correspond to first-order optimality?

We are dealing here with a mixed case of nonanticipativity: the second decision can depend on ξ but the first one can't. In line with the developments explained above, a good way to approach this is to pose it in terms of function pairs $(x_1(\cdot), x_2(\cdot))$ in $\mathcal{L}_{n_1+n_2} = \mathcal{L}_{n_1} \times \mathcal{L}_{n_2}$, a set $\mathcal{C}$ in that space and a nonanticipativity subspace $\mathcal{N}$ which restricts the response of $x_1(\cdot)$ without restricting that of $x_2(\cdot)$. The optimization problem then is to minimize a function over $\mathcal{C} \cap \mathcal{N}$ in $\mathcal{L}_{n_1} \times \mathcal{L}_{n_2}$, namely

$$\mathcal{G}(x_1(\cdot), x_2(\cdot)) = E_\xi[g(x_1(\xi), x_2(\xi), \xi)]. \tag{2.20}$$

The gradient of $\mathcal{G}$ can be calculated as the function $\nabla\mathcal{G} : \mathcal{L}_{n_1} \times \mathcal{L}_{n_2} \to \mathcal{L}_{n_1} \times \mathcal{L}_{n_2}$ given by

$$\nabla\mathcal{G}(x_1(\cdot), x_2(\cdot)) = \text{ the function in } \mathcal{L}_{n_1} \times \mathcal{L}_{n_2} \text{ that takes } \xi \text{ to } \nabla_{x_1,x_2} g(x_1(\xi), x_2(\xi), \xi) \tag{2.21}$$

(details will be presented later more generally). First-order optimality is characterized then by

$$- \nabla\mathcal{G}(x_1(\cdot), x_2(\cdot)) \in N_{\mathcal{C}\cap\mathcal{N}}(x_1(\cdot), x_2(\cdot)), \tag{2.22}$$

which is another "stochastic variational inequality in basic form" with $\mathcal{F} = \nabla\mathcal{G}$, but no longer single-stage. Even when $C(\xi)$ doesn't really depend on ξ, the variational inequality (2.22) can't be reduced to an EV form, but insights can come anyway from computing the normal cones to the new $\mathcal{C} \cap \mathcal{N}$.

There is now a different space $\mathcal{M}$ of nonanticipativity multipliers, consisting of the function pairs $(w_1(\xi), w_2(\xi)) \in \mathcal{L}_{n_1} \times \mathcal{L}_{n_2}$ such that $E[w_1(\xi)] = 0$ but $w_2(\xi) \equiv 0$. The stochastic variational inequality in extensive form associated with the one in basic form in (2.22) comes out then as the condition that

$$x_1(\xi) \equiv x_1 \text{ for some } x_1 \text{ such that } (x_1, x_2(\xi)) \in C(\xi) \text{ for all } \xi \in \Xi,$$
$$\text{and there exists some} w_1(\cdot) \in \mathcal{L}_{n_1} \text{ having } E[w_1(\xi)] = 0 \text{ such that}$$
$$-\nabla_{x_1, x_2} g(x_1(\xi), x_2(\xi), \xi) - (w_1(\xi), 0) \in N_{C(\xi)}(x) \text{ for all } \xi \in \Xi. \quad (2.23)$$

This illustrates that the special models researchers have been occupied with in a single-stage setting don't offer an adequate springboard for formulating stochastic variational inequalities in the realm of multistage stochastic optimization and equilibrium.[10] It furthermore offers hints of how nonanticipativity multipliers can be interpreted. In (2.23), $w_1(\xi)$ appears as a "shadow price for information" because it allows the constraint on $x_1(\cdot)$ being constant to be relaxed as if the future ξ could already be known.

3 Multistage modeling with nonanticipativity constraints

We proceed now from single-stage stochastic variational inequalities as in (2.6) and (2.9), as prototypes, to the formulation of general *multistage* versions in which "decisions" can respond to increasing availability of information. We adopt an N-stage pattern

$$x_1, \xi_1, x_2, \xi_2, \ldots, x_N, \xi_N \text{ where } x_k \in \mathbb{R}^{n_k}, \ \xi_k \in \Xi_k, \quad (3.1)$$

in which x_k is the decision to be taken at the kth stage and ξ_k stands for the information revealed after that decision, but before the next.[11] The previous x and ξ are replaced by

$$x = (x_1, \ldots, x_N) \in \mathbb{R}^{n_1} \times \cdots \times \mathbb{R}^{n_N} = \mathbb{R}^n \text{ for } n = n_1 + \cdots + n_N,$$
$$\xi = (\xi_1, \ldots, \xi_N) \in \Xi_1 \times \cdots \times \Xi_N, \text{ so that } \Xi \text{ is a subset of } \Xi_1 \times \cdots \times \Xi_N. \quad (3.2)$$

The exact nature of the sets Ξ_k doesn't matter (they could consist of vectors in some $\mathbb{R}^{\nu_k}$, for instance, or be boolean); all that concerns us is that Ξ is a *finite* set furnished with probabilities $p(\xi)$ that are positive and add to 1.[12]

---

[10] In a two-stage precedent of sorts for going beyond a pure single-stage model is present in in [3], where the "approximate" x in the ERM approach is treated as a first-stage decision compared to a ξ-dependent hindsight solution. This can still be viewed as an error minimization model rather than solving an actual variational inequality.

[11] Here we terminate with an observation, but we could instead terminate with a decision as in the two-stage preview. That alternative will be taken up later.

[12] This way of treating information serves us here as the being the simplest for purpose at hand. It fits as a special case of a more sophisticated treatment of finitely many scenarios that was laid out in [24]. This information structure could also be rendered in the form of a "scenario tree" with transition probabilities.

A crucial aspect of the sequencing in (3.1) is that the choice of x_k for $k > 1$ will be allowed to be influenced by the observations made before it but not by the observations made after it. This is the multistage version of *nonanticipativity*. A straightforward way of handling it is through response rules that express x_k as a function of $(\xi_1, \ldots, \xi_{k-1})$ only:

$$x(\xi) = (x_1, x_2(\xi_1), x_3(\xi_1, \xi_2), \ldots x_N(\xi_1, \xi_2, \ldots, \xi_{N-1})), \tag{3.3}$$

A better way though, as in Sect. 2, will be to articulate it as a constraint imposed within the space $\mathcal{L}_n$ of general functions $x(\cdot) : \xi \mapsto (x_1(\xi), x_2(\xi), \ldots, x_N(\xi))$ by restricting them to

$$\mathcal{N} = \big\{ x(\cdot) = (x_1(\cdot), \ldots, x_N(\cdot)) \,\big|\, x_k(\xi) \text{ does not depend on } \xi_k, \ldots, \xi_N \big\}. \tag{3.4}$$

This is the *nonanticipativity subspace* for our extended pattern of information. It henceforth replaces the single-stage $\mathcal{N}$ in Sect. 2. Corresponding *nonantipativity multipliers* will again come from a subspace $\mathcal{M}$ of $\mathcal{L}_n$, but defined now by

$$\mathcal{M} = \big\{ w(\cdot) = (w_1(\cdot), \ldots, w_N(\cdot)) \,\big|\, E_{\xi_k, \ldots, \xi_N}[w_k(\xi_1, \ldots, \xi_{k-1}, \xi_k, \ldots, \xi_N)] = 0 \big\}, \tag{3.5}$$

where the expectation is the conditional expectation knowing the initial components $\xi_1, \ldots, \xi_{k-1}$ of $\xi = (\xi_1, \ldots, \xi_{k-1}, \xi_k, \ldots, \xi_N)$. Once more there is underlying orthogonality with respect to the expectational inner product (2.2), which now expands to

$$\langle x(\cdot), w(\cdot) \rangle = \sum_{\xi \in \Xi} p(\xi) \sum_{k=1}^{N} \langle x_k(\xi), w_k(\xi) \rangle. \tag{3.6}$$

An elementary calculation using (3.6) confirms that

$$\mathcal{M} = \big\{ w(\cdot) \in \mathcal{L}_n \,\big|\, \langle x(\cdot), w(\cdot) \rangle = 0 \text{ for all } x(\cdot) \in \mathcal{N} \big\},$$
$$\mathcal{N} = \big\{ x(\cdot) \in \mathcal{L}_n \,\big|\, \langle x(\cdot), w(\cdot) \rangle = 0 \text{ for all } w(\cdot) \in \mathcal{M} \big\}, \tag{3.7}$$

as in the single-stage setting in (2.8). It follows from this mutual orthogonality relationship that

$$w(\cdot) \in N_{\mathcal{N}}(x(\cdot)) \iff x(\cdot) \in N_{\mathcal{M}}(w(\cdot)) \iff x(\cdot) \in \mathcal{N} \text{ and } w(\cdot) \in \mathcal{M}. \tag{3.8}$$

Variational inequality formulation and decomposition Along with nonanticipativity we constrain responses by scenario-dependent conditions $x(\xi) \in C(\xi)$ for nonempty closed convex sets $C(\xi)$ which now lie in the "product version" of $\mathbb{R}^n$ in (3.2) but need not themselves be products of sets in the component spaces $\mathbb{R}^{n_k}$. We refer to these restrictions as *basic* stochastic constraints, making room that way for eventual consideration of other "more advanced" stochastic constraints, perhaps involving risks

or expectations and not necessarily imposed scenario by scenario. In terms of the nonempty closed convex subset of $\mathcal{L}_n$ defined by

$$\mathcal{C} = \big\{ x(\cdot) \in \mathcal{L}_n \,\big|\, x(\xi) \in C(\xi) \text{ for all } \xi \in \Xi \big\}, \tag{3.9}$$

the basic constraints can be written as $x(\cdot) \in \mathcal{C}$.

The continuous mapping $\mathcal{F}$ from $\mathcal{L}_n$ to $\mathcal{L}_n$ that will enter the variational inequalities we are headed toward comes as before from the vectors $F(x, \xi) \in \mathbb{R}^n$. Now have the additional structure that

$$F(x, \xi) = (F_1(x, \xi), \ldots, F_N(x, \xi)) \quad \text{with } F_k(x, \xi) \in \mathbb{R}^{n_k},$$

where each $F_k(x, \xi)$ is continuous in $x \in \mathbb{R}^n$, so that $\mathcal{F}$ assigns to $x(\cdot)$ the function

$$\mathcal{F}(x(\cdot)) : \xi \mapsto F(\xi, x(\xi)) = (F_1(x(\xi), \xi), \ldots, F_N(x(\xi), \xi)). \tag{3.10}$$

Definition 3.1 (*SVI basic and extensive forms, multistage*) With respect to $F(x, \xi)$ and $C(\xi)$, the condition

$$- \mathcal{F}(x(\cdot)) \in N_{\mathcal{C} \cap \mathcal{N}}(x(\cdot)), \tag{3.11}$$

is the associated *multistage stochastic variational inequality in basic form*, whereas the condition

$$\begin{aligned} &x(\cdot) \in \mathcal{N} \text{ and there exists } w(\cdot) \in \mathcal{M} \text{ such that} \\ &-F(x(\xi), \xi) - w(\xi) \in N_{C(\xi)}(x(\xi)) \text{ for all } \xi \in \Xi. \end{aligned} \tag{3.12}$$

is the associated *multistage stochastic variational inequality in extensive form*.

With this double SVI formulation we get a result identical to that of Theorem 2.3 in the single-stage case except that a multistage information pattern of nonanticipativity is now covered.

Theorem 3.2 (basic-extensive equivalence, multistage). *If $x(\cdot)$ solves (3.12), then $x(\cdot)$ solves (3.11). Conversely, if $x(\cdot)$ solves (3.11), then $x(\cdot)$ is sure also to solve (3.12) if*

$$\text{there exists some } \hat{x}(\cdot) \in \mathcal{N} \text{ such that } \hat{x}(\xi) \in \text{ri } C(\xi) \text{ for all } \xi \in \Xi. \tag{3.13}$$

This constraint qualification is superfluous if the sets $C(\xi)$ are all polyhedral.

Proof The argument is identical to that of Theorem 2.3, with the only difference being the replacement of the earlier $\mathcal{N}$ and $\mathcal{M}$ by the current ones in (3.4) and (3.5), related by (3.7) and (3.8). □

The prime motivation for the extensive form is the stochastic decomposition it provides and the potential for utilizing that in computational methodology. The discussion of this matter in Sect. 2 for the single-stage case carries over fully and need

not be repeated in the present notation, but more can be added about the ways that the problem in extensive form can be interpreted.

Theorem 3.3 (primal-dual articulation of the extensive form). *The multistage stochastic variational inequality* (3.12), *as a condition in* $\mathcal{L}_n$ *on* $x(\cdot)$ *with auxiliary element* $w(\cdot)$, *is equivalent to a primal-dual variational inequality in* $\mathcal{L}_n \times \mathcal{L}_n$ *on the pair* $(x(\cdot), w(\cdot))$, *namely*

$$-\Phi(x(\cdot), w(\cdot)) \in N_{\mathcal{C} \times \mathcal{M}}(x(\cdot), w(\cdot)) \ \text{ for } \ \Phi(x(\cdot), w(\cdot)) = (F(x(\cdot)) + w(\cdot), -x(\cdot)).$$
(3.14)

Proof This is an easy consequence of the usual rules for computing normal cones. The conditions $-F(x(\xi), \xi) - w(\xi) \in N_{C(\xi)}(x(\xi))$ can be consolidated through (2.13) as $-\mathcal{F}(x(\cdot)) - w(\cdot) \in N_{\mathcal{C}}(x(\cdot))$, while the conditions $x(\cdot) \in \mathcal{N}$ and $w(\cdot) \in \mathcal{M}$ can be written on the basis of (3.8) as $x(\cdot) \in N_{\mathcal{M}}(w(\cdot))$. On the other hand, by (1.11) as translated to the current setting, the product of $N_{\mathcal{C}}(x(\cdot))$ and $N_{\mathcal{M}}(w(\cdot))$ is the normal cone to $\mathcal{C} \times \mathcal{M}$ at $(x(\cdot), w(\cdot))$. □

Specialization to simple basic constraints Still more can be said about the extensive form in the case of what we will call *simple* basic constraints, namely where $C(\xi)$ is a product of closed convex sets in the spaces $\mathbb{R}^{n_k}$ in the pattern

$$C(\xi) = D_1 \times D_2(\xi_1) \times D_3(\xi_1, \xi_2) \times \cdots \times D_N(\xi_1, \xi_2, \ldots, \xi_{N-1}), \quad (3.15)$$

because then

$$x(\cdot) \in \mathcal{C} \cap \mathcal{N} \iff x(\xi) \text{ has form (3.3) with}$$
$$x_1 \in D_1, \quad x_2(\xi_1) \in D_2(\xi_1), \ldots \ldots, x_N(\xi_1, \ldots, \xi_{N-1}) \in D_N(\xi_1, \ldots, \xi_{N-1}).$$
(3.16)

Theorem 3.4 (expectational case of the extensive form). *For simple basic constraints, the multistage stochastic variational inequality* (3.12) *in extensive form can be expressed in terms of* $x(\cdot)$ *alone as the condition that*

$x(\cdot)$ *has form (3.5), and* $\begin{cases} -E_{\xi_1, \xi_2, \ldots, \xi_N}[\, F_1(x(\xi), \xi)\,] \in N_{D_1}(x_1), \\ -E_{\xi_2, \ldots, \xi_N}[\, F_2(x(\xi), \xi)\,] \in N_{D_2(\xi_1)}(x_2(\xi_1)), \\ \quad\quad\vdots \\ -E_{\xi_N}[F_N(x(\xi), \xi)] \in N_{D_N(\xi_1, \ldots, \xi_{N-1})}(x_N(\xi_1, \ldots, \xi_{N-1})). \end{cases}$

(3.17)

Proof Utilizing the reduction in (3.16), this iteratively applies the expectation rule in (2.18). □

In the single-stage version, which we get from Theorem 3.4 by setting $N = 1$, this result reduces to the earlier one about the EV approach, namely that for constant C the

extensive form reduces to $-E_\xi F(x, \xi) \in N_C(x)$. However, for $N > 1$ the reduction is not as elementary and requires conditional expectations at different levels as in (3.17).

Adapting to models with terminal response It is possible and in many situations desirable to allow in (3.1) for a terminal decision $x_{N+1} \in \mathbb{R}^{n_{N+1}}$ which is able to respond to the final information input from ξ_N. From a mathematical standpoint, of course, this can equally well be regarded as already implicit in the model above as the case in which ξ_N is "trivialized," but that amounts to truncating the scenarios ξ and causes trouble with the meaning of $x(\xi)$ and $w(\xi)$. It is helpful instead, for examples considered later in this paper, to have explicit notation for this $N + 1$ extension, in which both $C(\xi)$ and $F(x, \xi)$ would acquire an $N + 1$ component. Since the point is to allow $x_{N+1}(\xi)$ to depend on all of $\xi = (\xi_1, \dots, \xi_N)$, the corresponding augmentation of $\mathcal{N}$ and $\mathcal{M}$ in (3.4) and (3.5) takes the form

$$
\mathcal{N}^+ = \{ x^+(\cdot) = (x_1(\cdot), \dots, x_N(\cdot), x_{N+1}(\cdot)) \,|\, (x_1(\cdot), \dots, x_N(\cdot)) \in \mathcal{N}, \; x_{N+1}(\cdot) \in \mathcal{L}_{n_{N+1}} \},
$$
$$
\mathcal{M}^+ = \{ w^+(\cdot) = (w_1(\cdot), \dots, w_N(\cdot), w_{N+1}(\cdot)) \,|\, (w_1(\cdot), \dots, w_N(\cdot)) \in \mathcal{M}, \; w_{N+1}(\xi) \equiv 0 \}.
$$
(3.18)

Then in the "simple" case of (3.15), for instance, the new $N + 1$ condition would have no expectation but just ask that $-F_{N+1}(x^+(\xi), \xi) \in N_{D_{N+1}(\xi)}(x_{N+1}(\xi))$ for all ξ.

The case of this with $N = 1$, in the pattern of x_1, ξ, $x_2(\xi)$ (response after a single observation, with no more information still to come) as previewed at the end of Sect. 2, deserves closer attention because of its prevalence in two-stage stochastic programming and potential game-like extensions at this level. There, having

$$
x(\cdot) = (x_1(\cdot), x_2(\cdot)) \in \mathcal{N}^+ \iff x_1(\cdot) \equiv \text{const} \in \mathbb{R}^{n_1}, \quad x_2(\cdot) \in \mathcal{L}_{n_2},
$$
$$
w(\cdot) = (w_1(\cdot), w_2(\cdot)) \in \mathcal{M}^+ \iff w_1(\cdot) \in \mathcal{L}_{n_1}, \; E_\xi[w_1(\xi)] = 0, \quad w_2(\xi) \equiv 0,
$$
(3.19)

the stochastic variational inequality in extensive form amounts to

$$
-(F_1(x_1, x_2(\xi), \xi), F_2(x_1, x_2(\xi), \xi)) + (w_1(\xi), 0) \in N_{C(\xi)}(x_1, x_2(\xi))
$$
$$
\text{with } E_\xi[w_1(\xi)] = 0.
$$
(3.20)

When $C(\xi) = D_1 \times D_2(\xi)$ this reduces to

$$
-E_\xi[F_1(x_1, x_2(\xi), \xi)] \in N_{D_1}(x_1), \; -F_2(x_1, x_2(\xi), \xi)) \in N_{D_2(\xi)}(x_2(\xi)) \text{ for all } \xi.
$$
(3.21)

Stochastic variational inequalities beyond basic and extensive A stochastic variational inequality that draws on nonanticipativity should ultimately be a condition of the form

$$
-\mathcal{F}(x(\cdot)) \in N_{\mathcal{K} \cap \mathcal{N}}(x(\cdot)) \text{ for some convex set } \mathcal{K} \subset \mathcal{C}, \tag{3.22}
$$

either directly or as elaborated in its expression, say, by Lagrange multiplier elements. The role of the basic form and its partner in extensive form is to provide a stripped-

down target to which such more general forms of stochastic variational inequalities may be reduced, e.g., for purposes of computing solutions. Examples with $\mathcal{K}$ specified by additional *expectation* constraints will be discussed below.

Another feature that could of course be relaxed, without stopping (3.18) from being called a stochastic variational inequality, is the special form of $\mathcal{F}(x(\cdot))$ as comprised of separate elements $F(x(\xi), \xi)$ for each ξ.

Monotonicity and the existence of solutions The concept of monotonicity of a mapping F from $\mathbb{R}^n$ to $\mathbb{R}^n$, defined in (1.10) relative to C, extends to mappings $\mathcal{F}$ from $\mathcal{L}_n$ to $\mathcal{L}_n$ relative to $\mathcal{C}$ as the requirement that

$$\langle \mathcal{F}(x'(\cdot)) - \mathcal{F}(x(\cdot)), x'(\cdot) - x(\cdot) \rangle \geq 0 \text{ for all } x(\cdot), \quad x'(\cdot) \in \mathcal{C} \qquad (3.23)$$

in terms of the expectational inner product (3.6). Strict monotonicity requires strict inequality when $x'(\cdot) \neq x(\cdot)$.

Theorem 3.5 (monotonicity of stochastic variational inequalities). *The mapping $\mathcal{F}$: $\mathcal{L}_n \to \mathcal{L}_n$ in (3.10) is monotone relative to $\mathcal{C}$ when the mapping $F(\cdot, \xi) : \mathbb{R}^n \to \mathbb{R}^n$ is monotone relative to $C(\xi)$ for every $\xi \in \Xi$, and likewise for strict monotonicity. Under monotonicity the set of solutions to the stochastic variational inequality (3.11) in basic form, if any, is convex. Under strict monotonicity, if a solution exists at all, it must be unique.*

Under monotonicity of $\mathcal{F}$ the mapping Φ in the primal-dual variational inequality (3.14) is monotone as well, implying that the set of solution pairs $(x(\cdot), w(\cdot))$ is convex.

Proof This just applies to our $\mathcal{L}_n$ setting a well known fact about solutions to monotone variational inequalities in general; cf. [25, 12.48]. □

The monotonicity in Theorem 3.5 is important because of its potential consequences for solution methodology, but ascertaining its availability in specific applications is separate challenge, of course. Monotonicity has consequences for existence as well as uniqueness. In the following, we indicate the *recession cone* of a convex set by superscript ∞.

Theorem 3.6 (existence of solutions to stochastic variational inequalities). *The set of solutions to the multistage stochastic variational inequality (3.11) in basic form is always closed. It is sure to be bounded and nonempty if $\mathcal{C} \cap \mathcal{N} \neq \emptyset$ and the sets $C(\xi)$ are bounded, or even if they are not all bounded as long as*

$$\nexists \text{ nonzero } y(\cdot) \in \mathcal{N} \text{ such that } y(\xi) \in C(\xi)^\infty \text{ (recession cone) for all } \xi \in \Xi.$$
$$(3.24)$$

When $\mathcal{F}$ is monotone as in Theorem 3.5, a criterion beyond such boundedness is available with respect to any $\hat{x}(\cdot)$ satisfying the conditions in the constraint qualification (3.13), namely

$$\text{if } \langle \mathcal{F}(x(\cdot), x(\cdot) - \hat{x}(\cdot) \rangle \geq 0 \text{ for all } x(\cdot) \in \mathcal{C} \cap \mathcal{N} \text{ with } ||x(\cdot) - \hat{x}(\cdot)|| > \rho,$$
$$\text{then there must exist a solution } x(\cdot) \text{ such that } ||x(\cdot) - \hat{x}(\cdot)|| \leq \rho. \qquad (3.25)$$

Proof As pointed out in the introduction, a variational inequality has at least one solution when the underlying convex set is bounded. For the SVI in basic form, that underlying set is $\mathcal{C} \cap \mathcal{N}$. A closed convex set is bounded if and only if its recession cone consists only of the zero vector [18, Theorem 8.4]. The recession cone of $\mathcal{C} \cap \mathcal{N}$ in $\mathcal{L}_n$ is the intersection of the recession cones of $\mathcal{C}$ and $\mathcal{N}$, with the recession cone of the subspace $\mathcal{N}$ being $\mathcal{N}$ itself [18, Corollary 8.3.3]. Thus, its elements are the functions $y(\cdot) \in \mathcal{N}$ such that $y(\cdot) \in \mathcal{C}^{\infty}$, and the latter comes down to having $y(\xi) \in C(\xi)^{\infty}$ for all ξ. The boundedness of $\mathcal{C} \cap \mathcal{N}$, when nonempty, is equivalent therefore to (3.19).

The extra criterion for the monotone case comes from applying [25, Theorem 12.51(a)] to the mapping

$$T : y(\cdot) \mapsto \mathcal{F}(\hat{x}(\cdot) + y(\cdot)) + N_{\mathcal{C} \cap \mathcal{N}}(\hat{x}(\cdot) + y(\cdot)),$$

which is "maximal monotone" when $\mathcal{F}$ is monotone [25, 12.48]. According to that result, a sufficient condition for the existence of $y(\cdot)$ with $0 \in T(y(\cdot))$ and $||y(\cdot)|| \leq \rho$ (which corresponds to a solution $x(\cdot)$ with $||x(\cdot) - \hat{x}(\cdot)|| \leq \rho$) is having

$$\langle v(\cdot), y(\cdot) \rangle \geq 0 \text{ whenever } v(\cdot) - F(\hat{x}(\cdot) + y(\cdot)) \in N_{\mathcal{C} \cap \mathcal{N}}(\hat{x}(\cdot) + y(\cdot)) \text{ with } ||y(\cdot)|| > \rho.$$

However, because the conditions on $\hat{x}(\cdot)$ in (3.13) make $\hat{x}(\cdot)$ belong to ri $(\mathcal{C} \cap \mathcal{N})$ (as explained in the proof of Theorem 3.2), the elements of $N_{\mathcal{C} \cap \mathcal{N}}(\hat{x}(\cdot) + y(\cdot))$ are orthogonal to $y(\cdot)$, so that $\langle v(\cdot), y(\cdot) \rangle$ reduces to $\langle \mathcal{F}(\hat{x}(\cdot) + y(\cdot)), y(\cdot) \rangle$. The criterion thereby translates to having the latter inner product be ≥ 0 when $\hat{x}(\cdot) + y(\cdot) \in \mathcal{C} \cap \mathcal{N}$ with $||y(\cdot)|| > \rho$. Replacing $y(\cdot)$ by $x(\cdot) - \hat{x}(\cdot)$ we arrive then at the condition in (3.25). □

In invoking (3.13), the criterion (3.25) under monotonicity also guarantees, through Theorem 3.2, the solvability of the corresponding multistage stochastic variational inequality in extensive form.

Lagrangian elaboration of basic constraints We next look at what can be gained when the sets $C(\xi)$ are specified by a system of function constraints. Specifically, we suppose that the sets $C(\xi)$ are specified by

$$x \in C(\xi) \iff x \in B(\xi) \text{ and } f_i(x, \xi) \begin{cases} \leq 0 & \text{for } i = 1, \ldots, r, \\ = 0 & \text{for } i = r + 1, \ldots, m, \end{cases} \quad (3.26)$$

where $B(\xi)$ is a nonempty closed convex set, and $f_i(x, \xi)$ is differentiable and convex in x for $i = 1, \ldots, r$, but affine in x for $i = r + 1, \ldots, m$. Then, in particular, $C(\xi)$ is a closed convex set.

Our aim is to obtain, and apply, a formula for the normal cones to the basic constraint set $\mathcal{C} \subset \mathcal{L}_n$ in terms of Lagrange multipliers for the conditions in (3.26). With ξ and x fixed for the moment, those multipliers should form a vector

$$y = (y_1, \ldots, y_m) \in Y = [0, \infty)^r \times (-\infty, \infty)^{m-r} \quad (3.27)$$

which, in coordination with x in (3.26), satisfies

$$y_i \begin{cases} \geq 0 & \text{for } i \leq r \text{ having } f_i(x, \xi) = 0, \\ = 0 & \text{for } i \leq r \text{ having } f_i(x, \xi) < 0, \end{cases}$$

or equivalently together with (3.26):

$$f(x, \xi) \in N_Y(y) \text{ for } f(x, \xi) = (f_1(x, \xi), \ldots, f_m(x, \xi)). \tag{3.28}$$

Theorem 3.7 (multiplier representation of basic constraints). *With respect to the constraint system* (3.26) *as given, the normal cone* $N_C(x(\cdot)) \subset \mathcal{L}_n$ *at an* $x(\cdot) \in C$ *contains all* $v(\cdot)$ *having, in the notation above, the representation*

$$\exists \, y(\cdot) \in \mathcal{L}_m, \ z(\cdot) \in \mathcal{L}_n, \ such \ that$$
$$\begin{cases} v(\xi) = \sum_{i=1}^m y_i(\xi) \nabla_x f_i(x(\xi), \xi) + z(\xi) \ with \\ f(x(\xi), \xi) \in N_Y(y(\xi)), \ z(\xi) \in N_{B(\xi)}(x(\xi)) \end{cases} \tag{3.29}$$

(in utilizing the fact that $N_Y(y(\xi)) \neq \emptyset$ *entails* $y(\xi) \in Y$*). This furnishes a complete description of* $N_C(x(\cdot))$ *under the constraint qualification that*

$$\exists \, \hat{x}(\cdot) \ such \ that, \ for \ all \ \xi \in \Xi, \ \hat{x}(\xi) \in \mathrm{ri} \, B(\xi) \ and \ f_i(\hat{x}(\xi), \xi) \begin{cases} < 0 & for \ i \leq r, \\ = 0 & for \ i > r. \end{cases} \tag{3.30}$$

Proof In view of the breakdown in (2.13), determining the elements of $N_C(x(\cdot))$ comes down to determining the elements of $N_{C(\xi)}(x(\xi))$ in $\mathbb{R}^n$ for individual ξ. The Lagrange multiplier representation in (3.29) for elements $v(\xi)$ of $N_{C(\xi)}(x(\xi))$ derives from well known rules for the calculus of normal cones to convex sets in $\mathbb{R}^n$, for instance in [18, Section 23]; in that context constraint qualifications can rely on relative interiors of convex sets as in (3.30). □

Stochastic variational inequalities of Lagrangian basic form When the representation (3.29) for $N_C(x(\cdot))$ is substituted into the condition in the stochastic variational inequality (3.12) in extensive form, the resulting relation, involving $y(\cdot) \in \mathcal{L}_m$ and $z(\cdot) \in \mathcal{L}_n$, is

$$-F(x(\xi), \xi) + w(\xi) = \sum_{i=1}^m y_i(\xi) \nabla_x f_i(x(\xi), \xi) + z(\xi)$$
$$\text{with } f(x(\xi), \xi) \in N_Y(y(\xi)), \ z(\xi) \in N_{B(\xi)}(x(\xi)). \tag{3.31}$$

We can thereby pass from (3.12) to a condition jointly on $x(\cdot)$ and $y(\cdot)$ instead of just $x(\cdot)$, which we call the associated stochastic variational inequality in *Lagrangian basic form*:

$x(\cdot) \in \mathcal{N}, \ y(\cdot) \in \mathcal{L}_m, \ \text{and} \ \exists w(\cdot) \in \mathcal{M} \ \text{such that}$

$$-\left(F(x(\xi), \xi) + \sum_{i=1}^{s} y_i(\xi) \nabla_x f_i(x(\xi), \xi), \right.$$

$$\left. -f(x(\xi), \xi) \right) + (w(\xi), 0) \in N_{B(\xi) \times Y}(x(\xi), y(\xi)). \qquad (3.32)$$

The really interesting thing about this representation is that (3.32) actually constitues a *stochastic variational inequality in extensive form* for $x^+(\cdot) = (x(\cdot), y(\cdot))$, with $y(\cdot)$ interpreted as a final response $x_{N+1}(\xi)$ *after the observation of* ξ_N. This follows the pattern for terminal response explained above with $\mathcal{N}^+$ and $\mathcal{M}^+$ as in (3.18). In terms of

$$\mathcal{D}^+ = \left\{ x^+(\cdot) = (x(\cdot), y(\cdot)) \,\middle|\, x^+(\xi) \in B^+(\xi) \right\} \text{ with } B^+(\xi) = B(\xi) \times Y,$$

$$F^+(x^+(\xi), \xi) = \left(F(x(\xi), \xi) + \sum_{i=1}^{s} y_i(\xi) \nabla_x f_i(x(\xi), \xi), -f(x(\xi), \xi) \right), \qquad (3.33)$$

we can express (3.32) as

$$x^+(\cdot) \in \mathcal{N}^+ \text{ and } \exists \, w(\cdot)^+ \in \mathcal{M}^+ \text{ such that } - F^+(x^+(\xi), \xi) + w^+(\xi) \in B^+(\xi). \qquad (3.34)$$

This is clearly again a stochastic variational inequality in extensive form for which the corresponding stochastic variational inequality in basic form is

$$- \mathcal{F}^+(x^+(\cdot)) \in N_{\mathcal{B}^+ \cap \mathcal{N}^+}(x^+(\cdot)). \qquad (3.35)$$

That provides confirmation of our underlying idea that the basic and extensive forms can serve as models to which more complicated stochastic variational inequalities can be reduced.

4 Some examples utilizing expectation functions and constraints

As explained in our introduction, stochastic variational inequalities ought to be broad enough in concept to assist in characterizing solutions to problems of stochastic optimization or equilibrium, even when initial decisions can be followed by recourse decisions in later stages. We'll illustrate here how our formulation achieves that coverage in a fundamental setting.

The examples to be presented will involve smooth *expectation functions* $\mathcal{G} : \mathcal{L}_n \to \mathbb{R}$, by which we mean expressions of the type

$$\mathcal{G}(x(\cdot)) = E_\xi[\, g(x(\xi), \xi)\,] = \sum_{\xi \in \Xi} p(\xi) g(x(\xi), \xi) \text{ for } g : \mathbb{R}^n \times \Xi \to \mathbb{R} \quad (4.1)$$

under the assumption that $g(x, \xi)$ is continuously differentiable in x for each ξ. That assumption makes $\mathcal{G}$ be continuously differentiable on $\mathcal{L}_n$, but it's worth looking at that in detail because of the special setting with an expectational inner product. From (4.1) it's clear that

$$\lim_{\varepsilon \to 0+} \frac{\mathcal{G}(x(\cdot) + \varepsilon u(\cdot)) - \mathcal{G}(x(\cdot))}{\varepsilon} = E_\xi \langle \nabla_x g(x(\xi), \xi), u(\xi) \rangle.$$

In other words $\mathcal{G}$ has directional derivatives

$$d\mathcal{G}(x(\cdot); u(\cdot)) = \langle \nabla \mathcal{G}(x(\cdot)), u(\cdot) \rangle \tag{4.2}$$

in terms of the gradient mapping $\nabla \mathcal{G} : \mathcal{L}_n \to \mathcal{L}_n$ which takes $x(\cdot)$ to the function

$$\nabla \mathcal{G}(x(\cdot)) : \xi \to \nabla_x g(x(\xi), \xi). \tag{4.3}$$

The assumed continuity of $\nabla_x g(\cdot, \xi) : \mathbb{R}^n \to \mathbb{R}^n$ for each ξ ensures the continuity of $\nabla \mathcal{G}$ as a mapping from $\mathcal{L}_n$ to $\mathcal{L}_n$ and confirms that we can rightly say that $\mathcal{G}$ is continuously differentiable with $\nabla \mathcal{G}(x(\cdot))$ as its gradient at $x(\cdot)$.

Optimality in minimizing an expectation in multistage optimization The elementary "optimization case" that was helpful in Sect. 1 as motivation for single-stage stochastic variational inequalities can now be expanded to a multistage setting and supplied with technical details.

For this we consider a smooth expectation function as above and for optimization turn to the problem

$$\text{minimize } \mathcal{G}(x(\cdot)) = E_\xi[\, g(x(\xi), \xi) \,] \text{ over all } x(\cdot) \in \mathcal{C} \cap \mathcal{N} \subset \mathcal{L}_n \tag{4.4}$$

with $\mathcal{C}$ and $\mathcal{N}$ as in Sect. 3. Multistage stochastic programming is covered by this, even without convexity in the objective; the linear programming subcase corresponds to the sets $C(\xi)$ being polyhedral.

Variational inequalities can capture conditions for optimality, and that is our target here for (4.4). From the differentiability that has been confirmed for $\mathcal{G}$, along with the convexity of $\mathcal{C} \cap \mathcal{N}$, we see that, if a local minimum in problem (4.4) occurs at $x(\cdot) \in \mathcal{C} \cap \mathcal{N}$, then $\langle \nabla \mathcal{G}(x(\cdot)), x'(\cdot) - x(\cdot) \rangle \leq 0$ for all $x'(\cdot) \in \mathcal{C} \cap \mathcal{N}$. Thus

$$\text{local min at } x(\cdot) \implies -\nabla \mathcal{G}(x(\cdot)) \in N_{\mathcal{C} \cap \mathcal{N}}(x(\cdot)), \tag{4.5}$$

which is the stochastic variational inequality in basic form in (3.11) for $\mathcal{F}(x(\cdot)) = \nabla \mathcal{G}(x(\cdot))$, corresponding to $F(x, \xi) = \nabla_x g(x, \xi)$.

This necessary condition becomes a sufficient condition for global optimality when $\mathcal{G}$ is convex, and then one has a *monotone* stochastic variational inequality. Under strict convexity a solution, if any, would have to be unique.

The corresponding stochastic variational inequality in extensive form requires that

$$x(\cdot) \in \mathcal{N} \text{ and there exists } w(\cdot) \in \mathcal{M} \text{ such that}$$
$$-\nabla_x g(x(\xi), \xi) - w(\xi) \in N_{C(\xi)}(x(\xi)) \text{ for all } \xi \in \Xi, \qquad (4.6)$$

where the condition for each scenario ξ means that $x(\xi)$ satisfies the first-order optimality condition for minimizing $g(\cdot, \xi) + \langle \cdot, w(\xi) \rangle$ over $C(\xi)$. *The linear term with $w(\xi)$ adds extra costs to the given costs so as to achieve stochastic decomposition by relaxing nonanticipativity and passing to a deterministic minimization problem for each ξ.* To drive that point home, in the case of strict convexity of the functions $g(\cdot, \xi)$ there could be only one $x(\xi)$ giving the minimum of $g(\cdot, \xi) + \langle \cdot, w(\xi) \rangle$ over $C(\xi)$. By determining it for each scenario ξ from knowledge of the right $w(\cdot)$, one would necessarily obtain a function $x(\cdot) \in \mathcal{N}$; in other words, nonanticipativity would be achieved automatically. It is for this reason that $w(\cdot)$ is said to provide shadow prices for future information.

To make use of (4.6) in solving (4.4) it isn't necessary, however, to determine the right $w(\cdot)$ in one miraculous step. With convexity in the objective, but not necessarily strict convexity, the decomposition in (4.6) can be achieved iteratively using the Progressive Hedging Algorithm in [24].

Basic constraint structure as in (3.26) would of course allow the condition in extensive form in (4.6) to be augmented to a stochastic variational inequality in Lagrangian form in the manner explained at the end of Sect. 3.

Equilibrium in a corresponding multistage game model A gamelike extension of this optimization example can be built around agents $j = 1, \ldots, J$ who select strategies $x^j(\cdot) \in \mathcal{C}^j \cap \mathcal{N}^j$ with $\mathcal{C}^j$ coming from sets $C^j(\xi)$ in the manner of (3.9). (The dimensions could be different for different agents, and this is why we are writing $\mathcal{N}^j$: a different $\mathcal{L}_{n^j}$ space may be involved.)

Agent j is concerned with minimizing with respect to the choice of $x^j(\cdot)$ an expectation function

$$\mathcal{G}^j(x^1(\cdot), \ldots, x^J(\cdot)) = E_\xi[\, g^j(x^1(\xi), \ldots, x^J(\xi), \xi)\,]. \qquad (4.7)$$

The complication is that this "cost" to agent j depends also on the strategies chosen by the other players. On the other hand, through the structure of stages of uncertainty and the availability of recourse decisions as information develops, *the agents can interact repeatedly over time* through their choices of

$$x^j(\xi) = (x_1^j, x_2^j(\xi_1), x_3^j(\xi_1, \xi_2), \ldots x_N^j(\xi_1, \xi_2, \ldots, \xi_{N-1})). \qquad (4.8)$$

As understood from the preceding optimization, the first-order optimality condition for agent j is

$$-\mathcal{F}^j(x^1(\cdot), \ldots, x^J(\cdot)) \in N_{C^j \cap \mathcal{N}^j}(x^j(\cdot)) \text{ for the mapping}$$
$$\mathcal{F}^j(x^1(\cdot), \ldots, x^J(\cdot)) : \xi \mapsto \nabla_{x^j} g^j(x^1(\xi), \ldots, x^J(\xi), \xi). \qquad (4.9)$$

The situation in which all these conditions in (4.9) are satisfied simultaneously constitutes a kind of equilibrium reflecting "stationarity" in the perceptions of the agents. It is a true Nash equilibrium if the functionals $\mathcal{G}^j$ in (4.7) are convex, which corresponds to each $g^j(x^1, \ldots, x^J, \xi)$ being convex with respect to x^j.

The most important fact here for our purposes of illustration is that the simultaneous satisfaction of the conditions in (4.9) can be unified into a *single* stochastic variational inequality in basic form:

$$-(\mathcal{F}^1(x^1(\cdot), \ldots, x^J(\cdot)), \ldots, \mathcal{F}^J(x^1(\cdot), \ldots, x^J(\cdot))) \in N_{\hat{\mathcal{C}} \cap \hat{\mathcal{N}}}(x^1(\cdot), \ldots, x^J(\cdot))$$
$$\text{where } \hat{\mathcal{C}} = \mathcal{C}^1 \times \cdots \times \mathcal{C}^J \text{ and } \hat{\mathcal{N}} = \mathcal{N}^1 \times \cdots \times \mathcal{N}^J. \tag{4.10}$$

The associated stochastic variational inequality in extensive form achieves decomposition into a separate problem for each scenario ξ by appealing to nonanticipativity multipliers

$$w(\cdot) = (w^1(\cdot), \ldots, w^J(\cdot)) \text{ in } \hat{\mathcal{M}} = \mathcal{M}^1 \times \cdots \times \mathcal{M}^J. \tag{4.11}$$

As in the single-agent optimization case, these multipliers provide shadow prices of information tailored to the individual agents, which allow nonanticipativity to be relaxed. For each ξ, then, one has a multistage game problem of deterministic character.

Having g^j be convex in its jth component would not be enough, in general, to make the variational inequality (4.10) be monotone, however. Monotonicity appears to be elusive in such an equilibrium setting, but future investigations may bring more understanding to this issue.

Incorporation of expectation constraints An example of a stochastic variational inequality of the more general form in (3.22), with the basic constraint set $\mathcal{C}$ replaced by a smaller set $\mathcal{K}$, comes from adding constraints of the type

$$\mathcal{G}_i(x(\cdot)) = E_\xi[\, g_i(x(\xi), \xi)\,] \begin{cases} \leq 0 & \text{for } i = 1, \ldots, r, \\ = 0 & \text{for } i = r+1, \ldots, m. \end{cases} \tag{4.12}$$

where each $\mathcal{G}_i$ is a smooth expectation functions as above. In working with such *expectation constraints* we will be able to exploit the fact that $\mathcal{G}_i$ is differentiable with gradients as in (4.2)–(4.3):

$$\nabla \mathcal{G}_i(x(\cdot)) \in \mathcal{L}_n \text{ for } \nabla \mathcal{G}_i(x(\cdot)) : \xi \mapsto \nabla_x g_i(x(\xi), \xi). \tag{4.13}$$

For the present purpose we also assume that $g_i(x, \xi)$ is convex on x for $i = 1, \ldots, r$ but affine in x for $i = r+1, \ldots, m$, so that $\mathcal{G}_i$ is convex for $i = 1, \ldots, r$ but affine for $i = r+1, \ldots, m$. This is needed to guarantee that the set

$$\mathcal{K} = \left\{ x(\cdot) \in \mathcal{C} \mid (4.12) \text{ holds} \right\} \tag{4.14}$$

is not just closed in $\mathcal{L}_n$ (through the continuity of each $\mathcal{G}_i$) but also convex, as is appropriate for the condition

$$- \mathcal{F}(x(\cdot)) \in N_{\mathcal{K} \cap \mathcal{N}}(x(\cdot)) \tag{4.15}$$

to legitimately be a *stochastic variational inequalty with expectation constraints*.

We can then try to understand how the normal cones $N_{\mathcal{K}}(x(\cdot))$ and $N_{\mathcal{K} \cap \mathcal{N}}(x(\cdot))$ relate to the previous normal cones $N_C(x(\cdot))$ and $N_{C \cap \mathcal{N}}(x(\cdot))$. We'll be interested in Lagrange multiplier vectors

$$y = (y_1, \ldots, y_m) \in Y = [0, \infty)^r \times (-\infty, \infty)^{m-r}, \tag{4.16}$$

which, in coordination with $x(\cdot)$ and (4.12), should satisfy

$$y_i \begin{cases} \geq 0 & \text{for } i \leq r \text{ having } \mathcal{G}_i(x(\cdot)) = 0, \\ = 0 & \text{for } i \leq r \text{ having } \mathcal{G}_i(x(\cdot)) < 0, \end{cases}$$

or equivalently when combined with (4.12):

$$\bar{\mathcal{G}}(x(\cdot)) \in N_Y(y) \text{ with } \bar{\mathcal{G}}(x(\cdot)) = (\mathcal{G}_1(x(\cdot)), \ldots, \mathcal{G}(x(\cdot))), \text{ i.e.,}$$
$$\bar{\mathcal{G}}(x(\cdot)) = E_\xi[\bar{g}(x(\xi), \xi)] \text{ for } \bar{g}(x, \xi) = (g_1(x, \xi), \ldots, g_m(x, \xi)). \tag{4.17}$$

Theorem 4.1 (multiplier representation for expectation constraints). *For $\mathcal{K}$ in (4.14), the normal cone $N_{\mathcal{K} \cap \mathcal{N}}(x(\cdot))$ at an $x(\cdot) \in \mathcal{K} \cap \mathcal{N}$ contains all $v(\cdot)$ having, in the notation above, the representation*

$$\exists \, y \in \mathbb{R}^m, \ z(\cdot) \in \mathcal{L}_n, \ such \ that \begin{cases} v(\cdot) = \displaystyle\sum_{i=1}^{m} y_i \nabla \mathcal{G}_i(x(\cdot)) + z(\cdot) \ with \\ \bar{\mathcal{G}}(x(\cdot)) \in N_Y(y), \ \ z(\cdot) \in N_{C \cap \mathcal{N}}(x(\cdot)) \end{cases} \tag{4.18}$$

(in utilizing the fact that $N_Y(y) \neq \emptyset$ implies $y \in Y$). This furnishes a complete description of $N_{\mathcal{K} \cap \mathcal{N}}(x(\cdot))$ under the constraint qualification that

$$\exists \, \hat{x}(\cdot) \in C \cap \mathcal{N} \ such \ that \ x(\xi) \in \text{ri} \, C(\xi) \ and \ \mathcal{G}_i(x(\cdot)) \begin{cases} < 0 & for \ i = 1, \ldots, r, \\ = 0 & for \ i = r+1, \ldots, m. \end{cases} \tag{4.19}$$

Moreover the same results hold with $\mathcal{K} \cap \mathcal{N}$ and $C \cap \mathcal{N}$ replaced everywhere by just $\mathcal{K}$ and C.

Proof Standard rules of convex analysis, e.g. [18, 23.8], support the assertions made about Lagrange multipliers. This takes into account that $\mathcal{N}$ is a subspace, so that having $\hat{x}(\cdot) \in \mathcal{N}$ along with $x(\xi) \in \text{ri} \, C(\xi)$. which corresponds to $x(\cdot) \in \text{ri} \, C$, implies $x(\cdot) \in \text{ri} \, (C \cap \mathcal{N})$. □

Stochastic variational inequalities of Lagrangian expectation form Again, as with the Lagrangian representation for basic constraints, we are able to pass to a wider expression of these conditions as a single variational inequality in the Lagrange multipliers and the other variables jointly. This emerges from consolidating the condition on the right of (4.18) to

$$v(\cdot) - \sum_{i=1}^{m} y_i \nabla \mathcal{G}_i(x(\cdot)) \in N_{\mathcal{C} \cap \mathcal{N}}(x(\cdot)) \text{ with } \bar{\mathcal{G}}(x(\cdot)) \in N_Y(y), \qquad (4.20)$$

having $\mathcal{G}$ as in (4.17), and going on to rewrite the variational inequality $-\mathcal{F}(x(\cdot)) \in N_{\mathcal{C} \cap \mathcal{N}}(x(\cdot))$ in (3.11) as a condition jointly on $y \in \mathbb{R}^m$ and $x(\cdot) \in \mathcal{L}_n$:

$$-\left(-\bar{\mathcal{G}}(x(\cdot)), \mathcal{F}(x(\cdot)) + \sum_{i=1}^{m} y_i \nabla \mathcal{G}_i(x(\cdot)) \right) \in N_{Y \times \mathcal{C} \cap \mathcal{N}}(y, x(\cdot)). \qquad (4.21)$$

Although this is truly a variational inequality in $(y, x(\cdot))$, it isn't one of basic form in our terminology. Likewise, the corresponding stochastic variational inequality in (3.12), similarly rewritten as

$x(\cdot) \in \mathcal{N}$ and there exist $w(\cdot) \in \mathcal{M}$ and $y \in \mathbb{R}^m$ such that

$$-\left(-\bar{\mathcal{G}}(x(\cdot)), \mathcal{F}(x(\cdot)) + \sum_{i=1}^{m} y_i \nabla \mathcal{G}_i(x(\cdot)) \right) + (0, w(\cdot)) \in N_{Y \times \mathcal{C}}(y, x(\cdot))$$

$$(4.22)$$

is not exactly one of extensive form. That doesn't stop us from referring to both of them as stochastic variational inequalities in *Lagrangian expectation form*, but much more can be said in this direction.

Reduction to extensive form with augmented nonanticipativity For one thing, we can interpret y as a "decision component" fixed in advance of observing $\xi_1, \ldots, \xi_N$. That way, it can be adjoined to the first-stage component of $x(\cdot)$ as a function $y(\cdot)$ constrained to be constant. In notation for that, we can think of augmenting $x(\xi)$ to

$\tilde{x}(\xi) \in \mathbb{R}^m \times \mathbb{R}^n$ with $\tilde{x}_1(\xi) = (y(\xi), x_1(\xi))$ but $\tilde{x}_k(\xi) = x_k(\xi)$ for $k = 2, \ldots, N$,

and in parallel augmenting $\mathcal{N}$ to

$$\tilde{\mathcal{N}} = \left\{ \tilde{x}(\cdot) = (y(\cdot), x(\cdot)) \,\middle|\, y(\cdot) \equiv \text{const}, \ x(\cdot) \in \mathcal{N} \right\}.$$

Then (4.21) comes out as the following variational inequality in $\tilde{x}(\cdot)$:

$$-\mathcal{F}_0(\tilde{x}(\cdot)) \in N_{(Y \times \mathcal{C}) \cap \tilde{\mathcal{N}}}(\tilde{x}(\cdot)) \text{ for } \mathcal{F}_0(\tilde{x}(\cdot))$$

$$= \left(-\bar{\mathcal{G}}(x(\cdot)), \mathcal{F}(x(\cdot)) + \sum_{i=1}^{m} y_i(\cdot) \nabla \mathcal{G}_i(x(\cdot)) \right). \qquad (4.23)$$

This is closer to being a stochastic variational inequality of basic type, but it still misses the mark because $\mathcal{F}_0(\tilde{x}(\cdot))$ doesn't fit our prescription of being a function $\xi \to F_0(\tilde{x}(\xi), \xi)$, due to the nature of the vector $\bar{\mathcal{G}}(x(\cdot))$ in (4.16).

We can do better however with the corresponding interpretation of (4.22). For that we have to recognize the necessity of an additional multiplier element $u(\cdot)$ with $E_\xi[u(\xi)] = 0$ to take care of the constancy constraint on $y(\cdot)$. Introducing

$$\tilde{C}(\xi) = Y \times C(\xi), \quad \text{yielding } \tilde{\mathcal{C}} = \left\{ (y(\cdot), x(\cdot)) = \tilde{x}(\cdot) \in \mathcal{L}_n \,\middle|\, \tilde{x}(\xi) \in \tilde{C}(\xi) \right\}, \tag{4.24}$$

we can express the variational inequality in (4.22) *equivalently* as

$x(\cdot) \in \mathcal{N}$ and there exist $w(\cdot) \in \mathcal{M}, \quad y \in \mathbb{R}^m, \quad u(\cdot)$ with $E_\xi[u(\xi)] = 0$, such that

$$-\left(-\bar{\mathcal{G}}(x(\xi)\xi)), \mathcal{F}(x(\xi)) + \sum_{i=1}^m y_i(\xi)\nabla\mathcal{G}_i(x(\xi)) \right)$$
$$+ (u(\xi), w(\xi)) \in N_{\tilde{\mathcal{C}}(\xi)}(y(\xi), x(\xi)), \quad \forall \xi \tag{4.25}$$

Why an *equivalence*? For $\mathcal{G}(x(\cdot)) = E_\xi[g(x(\xi), \xi)]$ we are invoking yet again the rule that

$$E_\xi[\bar{g}(x(\xi), \xi)] \in N_Y(y) \iff \exists u(\cdot), \ E_\xi[u(\xi)] = 0,$$
$$\text{with } \bar{g}(x(\xi), \xi) + u(\xi) \in N_Y(y(\xi)), \quad \forall \xi.$$

Taking this further, we can augment $w(\cdot) \in \mathcal{M}$ to

$\tilde{w}(\xi) \in \mathbb{R}^m \times \mathbb{R}^n$ with $\tilde{w}_1(\xi) = (u(\xi), w_1(\xi))$ but $\tilde{w}_k(\xi) = x_k(\xi)$ for $k = 2, \ldots, N$,

and introduce

$$\tilde{\mathcal{M}} = \left\{ \tilde{w}(\cdot) = (u(\cdot), w(\cdot)) \,\middle|\, E_\xi[u(\xi)] = 0, \ w(\cdot) \in \mathcal{M} \right\}$$

along with

$$\tilde{F}(\tilde{x}, \xi) = \left(-\bar{g}(x, \xi)), F(x, \xi) + \sum_{i=1}^m y_i \nabla_x g_i(x, \xi)) \right). \tag{4.26}$$

With this in hand we can express the variational inequality in (4.25) as

$\tilde{x}(\cdot) \in \tilde{\mathcal{N}}$ and $\exists\, \tilde{w}(\cdot) \in \tilde{\mathcal{M}}$ such that $- \tilde{F}(\tilde{x}(\xi), \xi)) + \tilde{w}(\xi) \in N_{\tilde{C}1(\xi)}(\tilde{x}(\xi)), \ \forall \xi.$ (4.27)

We have then reached a genuine stochastic variational inequality of extensive type, demonstrating once more that such a model can serve as a target for the reduction of more complicated variational inequalities.

Risk instead of expectation This section has focused on objectives and constraints involving expectations, but expectation could also be replaced by risk expressions like CVaR in utilizing the theory in [21]. This topic, needing more space for development than is available here, will be taken up separately later.

5 Potential applications to models with nonconvex constraints

The stochastic variational inequalities of *Lagrangian* form in Sects. 3 and 4 emerged from multiplier rules invoked for *convex* constraint systems of basic type or expectation type. However, multiplier rules utilizing other constraint qualifications for *nonconvex* constraint systems the same kinds of variational inequalities can be reached, This offers more fertile ground for our subject than might have been recognized so far.

A prominent example in the original history of variational inequalities is their capability of representing first-order optimality conditions even in optimization without constraint convexity. The Karush–Kuhn–Tucker conditions in nonlinear programming were expressed in that manner by Robinson [17] for the sake of studying how the solutions to a problem may depend on the data in the problem—a program that has continued to this day with ever-wider reach, cf. [7]. That avenue of research could well be followed also for stochastic models of optimization or equilibrium, even with nonconvexity.

The main idea is that many applications could lead to a condition of the same appearance but with $\mathcal{C}$ or $\mathcal{K}$ nonconvex and normal cones being defined in the broader sense of variational analysis instead of just convex analysis, as for instance in [25]. Such a stochastic condition with $\mathcal{C}$ or $\mathcal{K}$ nonconvex couldn't properly be called a variational inequality, but still, multiplier rules might be invoked to reduce it to much in the same way as above to a true variational inquality.

References

1. Agdeppa, R.P., Yamshita, N., Fukushima, M.: Convex expected residual models for stochastic affine variational inequality problems and its application to the traffic equilibrium models. Pac. J. Optim. **6**, 3–19 (2010)
2. Chen, X.J., Fukushima, M.: Expected residual minimization method for stochastic linear complementary problems. Math. Oper. Res. **30**, 1022–1038 (2005)
3. Chen, X.J., Wets, R.J.-B., Zhang, Y.: Stochastic variational inequalities: Residual minimization smoothing/sample average approximations. SIAM J. Optim. **22**, 649–673 (2012)
4. Chen, X.J., Zhang, Y., Fukushima, M.: Robust solution of monotone stochastic linear complementarity problems. Math. Program. **117**, 51–80 (2009)
5. Cottle, R.W., Giannessi, F., Lions, J.-L.: Variational Inequalities and Complementary Problems. Wiley, New York (1980)
6. Dontchev, A.L., Rockafellar, R.T.: Convergence of inexact Newton mehods for generalized equations. Math. Program. B **139**, 115–137 (2013)
7. Dontchev, A.L., Rockafellar, R.T.: Implicit Functions and Solution Mappings: A View From Variational Analysis, 2nd edn. Springer, New York (2014)
8. Facchinei, F., Pang, J.S.: Finite-Dimensional Variational Inequalities and Complementarity Problems. Springer, New York (2003)
9. Fang, H., Chen, X.J., Fukushima, M.: Stochastic R_0 matrix linear complementarity problems. SIAM J. Optim. **18**, 482–506 (2007)

10. Ferris, M., Pang, J.S.: Engineering and economic applications of complementarity problems. SIAM Rev. **39**, 669–713 (1997)
11. Gürkan, G., Özge, A.Y., Robinson, S.M.: Sample path solution of stochastic variational inequalities. Math. Program. **84**, 313–333 (1999)
12. Iusem, A., Jofré, A., Thompson, P.: Approximate projection methods for monotone stochastic variational inequalities. (preprint 2014)
13. Jiang, H., Xu, H.: Stochastic approximation approaches to the stochastic variational inequality problem. IEEE Trans. Autom. Control **53**, 1462–1475 (2008)
14. Jofré, A., Rockafellar, R.T., Wets, R.J.-B.: Variational inequalities and economic equilibirum. Math. Oper. Res. **32**, 32–50 (2007)
15. Ling, C., Qi, L., Zhou, G., Caccetta, L.: The SC^1 property of an expected residual function arising from stochastic complementarity problems. Oper. Res. Lett. **36**, 456–460 (2008)
16. Luo, M.J., Lin, J.H.: Expected residual minimization method for stochastic variational inequality problems. J. Optim. Theory Appl. **140**, 103–116 (2009)
17. Robinson, S.M.: Generalized equations and their solutions, I. Basic theory. Mathem. Program. Study **10**, 128–141 (1979)
18. Rockafellar, R.T.: Convex Analysis. Princeton University Press, Princeton (1970)
19. Rockafellar, R.T.: Monotone operators and the proximal point algorithm. SIAM J. Control Optim. **14**, 547–589 (1976)
20. Rockafellar, R.T.: Extended nonlinear programming. In: Di Pillo, G., Giannessi, F. (eds.) Nonlinear Optimization and Related Topics, pp. 381–399. Kluwer, Alphen aan den Rijnpp (1999)
21. Rockafellar, R.T., Uryasev, S.: The fundamental risk quadrangle in risk management, optimization and statistical estimation. Surv. Oper. Res. Manag. Sci. **18**, 33–53 (2013)
22. Rockafellar, R.T., Wets, R.J.-B.: Nonanticipativity and L^1-martingales in stochastic optimization problems. Math. Program. Study **6**, 170–187 (1976)
23. Rockafellar, R.T., Wets, R.J.-B.: The optimal recourse problem in discrete time: L^1-multipliers for inequality constraints. SIAM J. Control Optim. **16**, 16–36 (1978)
24. Rockafellar, R.T., Wets, R.J.-B.: Scenarios and policy aggregation in optimization under uncertainty. Math. Oper. Res. **16**, 119–147 (1991)
25. Rockafellar, R.T., Wets, R.J.-B.: Variational Analysis, No. 317 in the Series Grundlehren der Mathematischen Wissenschaften. Springer, New York (1997)
26. Ruszczyński, A., Shapiro, A.: Stochastic Programming, Handbooks in Operations Research and Management Science. Elsevier, Philadelphia (2003)
27. Xu, H.: Sample average approximation methods for a class of stochastic variational inequality problems. Asia Pac. J. Oper. Res. **27**, 103–109 (2010)
28. Zhang, C., Chen, X.: Smoothing projected gradient method and its application to stochastic linear complementary problems. SIAM J. Optim. **20**, 627–649 (2009)
29. Zhang, C., Chen, X., Sumalee, A.: Robust Wardrop user equilibrium assignment under stochastic demand and supply: expected residual inimization approach. Transp. Res. B **45**, 534–552 (2011)