Two-Stage Stochastic Variational Inequalities:
an ERM-solution procedure

Xiaojun Chen[1]   Ting Kei Pong[2]   Roger J-B Wets[3]

Abstract.   We propose a two-stage stochastic variational inequality model to deal with random variables in variational inequalities, and formulate this model as a two-stage stochastic programming with recourse by using an expected residual minimization (ERM) solution procedure. The solvability, differentiability and convexity of the two-stage stochastic programming and the convergence of its sample average approximation are established. Examples of this model are given, including the optimality conditions for stochastic programs, a Walras equilibrium problem and Wardrop flow equilibrium. We also formulate stochastic traffic assignments on arcs flow as a two-stage stochastic variational inequality based on Wardrop flow equilibrium and present numerical results of the Douglas-Rachford splitting method for the corresponding two-stage stochastic programming with recourse.

Keywords: stochastic variational inequalities, stochastic program with recourse, Wardrop equilibrium, expected residual minimization, regularized gap function, splitting method.

AMS Classification: 90C33, 90C15

Date: March 6, 2016

1   Introduction

All stochastic variational models involve inherently a "dynamic" component that takes into account decisions taken over time, or/and space, where the decisions depend on the information that will become available as the decision process evolves. So far, the models proposed for *stochastic variational inequalities* have either bypassed or not made explicit this particular feature(s). Various "stochastic" extensions of variational inequalities have been proposed in the literature but so far relatively little concern has been paid to the 'dynamics' of the decision, or solution, process that is germane to all stochastic variational problems: stochastic programs, stochastic optimal control, stochastic equilibrium models in economics or finance, stochastic games, and so on. The "dynamics" of the model considered here are of limited scope. What is essential is that it makes a distinction between two families of variables: (i) those that are of the "here-and-now" type and cannot depend on the outcome of random events to be revealed at some future time or place and (ii) those that are allowed to depend on these outcomes. Our restriction to the two-stage model allows for a more detailed exposition and analysis as well as the development of computational guidelines, implemented here in a specific instance. By empathy with the terminology used for stochastic programming models, one might be tempted to refer to such a class of problems as stochastic variational inequalities with recourse but, as we shall see from the formulation and examples, that would not quite catch the full nature of the variables, mostly because the decision-variables aren't necessarily chosen sequentially. We shall refer to this "short horizon" version as a *two-stage stochastic variational inequality*. In principle, the generalization to a multistage model is not challenging, at least conceptually, notwithstanding that a careful description of the model might become rather delicate, involved and technically, not completely trivial; a broad view of multistage models, as well as some of their canonical features, is provided in [58].

We consider the *two-stage stochastic variational inequality*: Given the (induced) probability space $(\Xi \subset I\!\!R^N, \mathcal{A}, P)$, find $\big(x \in I\!\!R^{n_1}, u : \Xi \to I\!\!R^{n_2}, \mathcal{A}\text{-measurable}\big)$ such that the following collection of variational inequalities is satisfied[4]:

$$-I\!\!E[G\big(\boldsymbol{\xi}, x, u_{\boldsymbol{\xi}}\big)] \in N_D(x)$$
$$-F\big(\boldsymbol{\xi}, x, u_{\boldsymbol{\xi}}\big) \in_{a.s.} N_{C_{\boldsymbol{\xi}}}\big(H(x, u_{\boldsymbol{\xi}})\big)$$

with

- $G : (\Xi, I\!\!R^{n_1}, I\!\!R^{n_2}) \to I\!\!R^{n_1}$ a vector-valued function, continuous with respect to (x, u) for all $\xi \in \Xi$, $\mathcal{A}$-measurable and summable with respect to ξ.

- $N_D(x)$ the normal cone to the closed-convex set $D \subset I\!\!R^{n_1}$ at $x \in I\!\!R^{n_1}$.

- $F : (\Xi, I\!\!R^{n_1}, I\!\!R^{n_2}) \to I\!\!R^{n_2}$ a vector-valued function, continuous with respect to (x, u) for all $\xi \in \Xi$ and $\mathcal{A}$-measurable with respect to ξ.

- $N_{C_\xi}\big(v\big)$ the normal cone to the closed-convex set $C_\xi \subset I\!\!R^{n_2}$ at $v \in I\!\!R^{n_2}$, the random set $C_{\boldsymbol{\xi}}$ is $\mathcal{A}$-measurable.

---

[4]Bold face $\boldsymbol{\xi}$ is reserved to denote the random vector of parameters whereas ξ refers to a specific realization of $\boldsymbol{\xi}$.

- $H : (I\!\!R^{n_1}, I\!\!R^{n_2}) \to I\!\!R^{n_2}$ a continuous vector-valued function.

The definition of the normal cone yields the following, somewhat more explicit, but equivalent formulation:

$$\text{find } \bar{x} \in D \text{ and } \bar{u} : \Xi \to I\!\!R^{n_2}, \ \mathcal{A}\text{-measurable, such that } H(\bar{x}, \bar{u}_{\boldsymbol{\xi}}) \in_{a.s.} C_{\boldsymbol{\xi}} \text{ and}$$
$$\langle I\!\!E[G(\boldsymbol{\xi}, \bar{x}, \bar{u}_{\boldsymbol{\xi}})], x - \bar{x} \rangle \geq 0, \ \forall\, x \in D,$$
$$\langle F(\boldsymbol{\xi}, \bar{x}, \bar{u}_{\boldsymbol{\xi}}), v - H(\bar{x}, \bar{u}_{\boldsymbol{\xi}}) \rangle \geq 0, \ \ \forall v \in C_{\boldsymbol{\xi}}, \ \ P\text{-}a.s.$$

The model assumes that the uncertainty can be described by a random vector $\boldsymbol{\xi}$ with known distribution P and a two-stage decision process: (i) x to be chosen before the value ξ of $\boldsymbol{\xi}$ is revealed (observed) and (ii) u to be selected with full knowledge[5] of the realization ξ. From the decision-maker's viewpoint, the problem can be viewed as choosing a pair $(x, \xi \mapsto u_\xi)$ where u depends on the events that might occur, or in other words, is $\mathcal{A}$-measurable. It is noteworthy that in our formulation of the stochastic variational inequality this pair is present, in one way or another, in each one of our examples. We find it advantageous to work here with this slightly more explicit model where the first collection of inclusions suggest the presence of (expected) *equilibrium constraints*; in [58], taking a somewhat more comprehensive approach, it is shown how these "equilibrium constraints" can be incorporated in a global, possibly more familiar, variational inequality of the same type as the second inclusion.

This paper is organized as follows. In §2, we devote a review on some fundamental examples and applications that are special cases of this model. In the last two examples in §2, we concentrate our attention on the stochastic traffic flow problems with the accent being placed on getting implementable solutions. To do this we show how an alternative formulation, based on the Expected Residual Minimization (ERM) might actually be more adaptable to coming with solutions that are of immediate interest to the initial design or capacity expansion of traffic networks. In §3, we develop the basic properties of such a model, lay out the theory to justify deriving solutions via a sample average approximation approach in §4 and finally, in §5, describe an algorithm procedure based on the Douglas-Rachford splitting method which is then illustrated by numerical experimentation involving a couple of "classical-test" networks. One of the basic goals of this article was to delineate the relationships between various formulations of stochastic variational inequalities as well as to show how the solution-type desired might also lead us to work with some variants that might not fit perfectly the canonical model.

2 Stochastic variational inequalities: Examples

When $\boldsymbol{\xi}$ is discretely distributed with finite support, i.e., Ξ is finite, one can write the problem as:

$$\text{find } \left(x, (u_\xi, \xi \in \Xi) \right) \text{ such that } \begin{cases} -I\!\!E[G(\xi, x, u_\xi)] & \in N_D(x). \\ -F(\xi, x, u_\xi) & \in N_{C_\xi}(H(x, u_\xi)), \forall\, \xi \in \Xi. \end{cases}$$

---

[5] A further refinement of the model would allow for the possibility that only partial observation is available in which case one would have to change $\mathcal{A}$-measurability to measurability with respect to a sub-sigma field.

When expressed in this form, we are just dealing with a, possibly extremely large, deterministic variational inequality. How large will depend on the cardinality $|\Xi|$ of the support and this immediately raises the question of the design of solution procedures for large scale variational inequalities. The difficulty in finding a solution might also depend on how challenging it is to compute $E[G(\boldsymbol{\xi}, x, u_\xi)]$, even when G doesn't depend on u.

Of course, both the choice of x and the realization ξ of the random components of the problem influence the 'upcoming' environment so we can also think as the *state* of the system being determined by the pair (ξ, x) and occasionally it will be convenient, mostly for technical reasons, to view the u-decision as a function of the state, i.e., $(\xi, x) \mapsto u(\xi, x)$, cf. §3.

On the other hand, various generalizations are possible:

(a) One could also have D depend on $\boldsymbol{\xi}$ and x, in which case we are dealing with a random convex-valued mapping: $D : \Xi \times R^{n_1} \rightrightarrows R^{n_1}$ and one would, usually, specify the continuity properties of D with respect to $\boldsymbol{\xi}$ and x; the analysis then enters the domain of stochastic generalized equations and can get rather touchy [62, 45, 46]. Here, we restrict our analysis to the case when D is independent of $\boldsymbol{\xi}$ and x.

(b) Another extension is the case when there are further global type constraints on the choice of the functions u, for example, constraints involving $E[u_\xi]$ or equilibrium-type constraints [47, 9].

(c) The formulation can be generalized to a two-stage stochastic quasi variational inequality or generalized Nash equilibrium. For example the second-stage variational inequality problem can be defined as follows:

$$-F\big(\boldsymbol{\xi}, x, u_{\boldsymbol{\xi}}\big) \in_{a.s.} \ N_{C_{\boldsymbol{\xi}}(x)}(H(x, u_{\boldsymbol{\xi}})),$$

where the set $C_{\boldsymbol{\xi}}$ depends on x [21]. However, the analysis of the generalization is not trivial and deserves a separate analysis. In this paper, we restrict our analysis to the case when $C_{\boldsymbol{\xi}}$ is independent of x.

In order to illustrate the profusion of problems that are covered by this formulation we are going to go through a series of examples including some that are presently understood, in the literature, to fall under the "stochastic variational inequality" heading.

2.1 One-stage examples

2.1 Example Single-stage problems. *This is by all means the class of problems that, so far, has attracted the major interest in the literature. In terms of our notation, it reads*

$$\text{find } \bar{x} \in D \subset R^n \text{ such that } -E[G(\boldsymbol{\xi}, \bar{x})] \in N_D(\bar{x}),$$

where D is a closed convex set, possibly bounded and often polyhedral, for all $x \in D$, the vector-valued mapping $\xi \mapsto G(\xi, x) \in R^n$ is summable and $x \mapsto E[G(\boldsymbol{\xi}, x)]$ is continuous. Especially in the design of solution procedures, it is convenient to rely on the alternative formulation,

$$\text{find } \bar{x} \in D \subset R^n \text{ such that } \ \forall x \in D, \langle E[G(\boldsymbol{\xi}, \bar{x})], x - \bar{x} \rangle \geq 0.$$

Detail. It's only with some hesitation that one should refer to this model as a "stochastic variational equality." Although, certain parameters of the problem are random variables and the solution will have to take this into account, this is essentially a deterministic variational inequality with the added twist that evaluating $I\!E[G(\boldsymbol{\xi}, x)]$, usually a multidimensional integral, requires relying on quadrature approximation schemes. Typically, P is then approximated by a discrete distribution with finite support, obtained via a cleverly designed approximation scheme or as the empirical distribution derived from a sufficiently large sample. So far, no cleverly designed approximation scheme has been proposed although the approach used by Pennanen and Koivu in the stochastic programming context might also prove to be effective [49, 48] in this situation. To a large extent the work has been focused on deriving convergence of the solutions of approximating variational inequalities where P has been replaced by a sequence of empirical probability measures P^{ν}, generated from independent samples of $\boldsymbol{\xi}$: $\xi^1, \xi^2, ..., \xi^{\nu}$. The approximating problem:

$$\text{find } x^{\nu} \in D \subset I\!R^n \text{ such that } -\nu^{-1} \sum\nolimits_{k=1}^{\nu} G(\xi^k, x^{\nu}) \in N_D(x^{\nu}).$$

Two basic questions then become:

(i) Will the solutions x^{ν} converge to a solution of the given problem when the number of samples gets arbitrarily large?

(ii) Can one find bounds that characterize the error, i.e., can one measure the distance between x^{ν} and the (set of) solution(s) to the given variational inequality.

There is a non-negligible literature devoted to these questions which provides satisfactory answers under not too stringent additional restrictions, cf. [25, 31, 28, 35, 32, 60, 5, 40, 41, 38, 65, 30]. □

In [27] Gürkan, Özge and Robinson rely on solving a variational inequality of this type to price an American call option with strike price K for an instrument paying a dividend d at some time $\tau \leq T$. The expiration date comes down to calculating the expectation of the option-value based on whether the option is exercised, or not, at time τ^- just before the price of the (underlying) instrument drops by d which otherwise follows a geometric Brownian motion, cf. [27, Section 3] and the references therein. The authors rely on the first order optimality conditions, generating a one-stage stochastic variational inequality where the vector-valued mapping $G(\xi, \cdot)$ corresponds to the gradient of the option-value along a particular price-path. It turns out that contrary to an approach based on a sample-path average approximation of these step functions [27, Figure 1] the sample average approximation of their gradients is quite well-behaved [27, Figure 2]. □

2.2 Example Stochastic linear complementarity problem. *The stochastic linear complementarity version of the one-stage stochastic variational inequality,*

$$0 \leq_{a.s.} M_{\xi}x + q_{\xi} \perp_{a.s.} x \geq 0,$$

where some or all the elements of the matrix M and vector q are potentially stochastic, was analyzed by Chen and Fukushima [8] suggesting in the process a solution procedure based on "ERM: Expected Residual Minimization."

Their *residual* function can be regarded as a relative of the *gap function* used by Facchinei and Pang to solve deterministic variational inequalities [21]. More specifically,

$$\mathbb{E}[\|\Phi(\boldsymbol{\xi}, x)\|^2] \text{ with } \Phi(\xi, x) = \begin{pmatrix} \varphi\big((M_\xi x + q_\xi)_1, x_1\big) \\ \vdots \\ \varphi\big((M_\xi x + q_\xi)_n, x_n\big) \end{pmatrix}$$

is minimized for $x \in \mathbb{R}^n_+$, where $\varphi : \mathbb{R}^2 \to \mathbb{R}$ is such that $\varphi(a, b) = 0$ if and only if $(a, b) \in \mathbb{R}^2_+$, $ab = 0$; for example, the min-function $\varphi(a, b) = \min\{a, b\}$ and the Fisher-Burmeister function $\varphi(a, b) = (a + b) - \sqrt{a^2 + b^2}$ satisfy these conditions. Quite sharp existence results are obtained, in particular, when only the vector q is stochastic. They are complemented by convergence result when the probability measure P is replaced by discrete empirical measures generated via (independent) sampling. The convergence of the stationary points is raised in [8, Remark 3.1].

The stochastic linear complementarity problem gets recast as a (stochastic) optimization problem where one is confronted with an objective function defined as a multi-dimensional integral. Convergence of the solutions of the discretized problems can then be derived by appealing to the law of large numbers for random lsc (lower semicontinuous) functions [2, 37].

The solution provided by the ERM-reformulated problem

$$\min_{x \geq 0} \mathbb{E}[\|\Phi(\boldsymbol{\xi}, x)\|^2]$$

doesn't, strictly speaking, solve the originally formulated complementarity problem for every realization ξ; this could only occur if the optimal value turns out to be 0, or equivalently, if one could find a solution that satisfies for (almost) all ξ, the system $0 \leq_{a.s.} M_\xi x + q_\xi \perp_{a.s.} x \geq 0$.

The residual function $\|\Phi(\xi, x)\|$ can be considered as a cost function which measures the loss at the event ξ and decision x. The ERM formulation minimizes the expected values of the loss for all possible occurrences due to failure of the equilibrium. Recently Xie and Shanbhag construct tractable robust counterparts as an alternative way to using the ERM approach [66]. □

2.2 Two-stage examples

Our first two examples in this subsection haven't been considered, so far, in the literature but in some ways best motivates our formulation, cf. §1, in the same way that optimality conditions for (deterministic) linear and nonlinear optimization problems lead us to a rich class of variational inequalities [21].

2.3 Example Optimality conditions for a stochastic program. *Not unexpectedly, the optimality conditions for a stochastic program with recourse (two-stage) lead immediately to a two-stage stochastic variational inequality. Here, we only develop this for a well-behaved simple (linear) recourse problem; to deal with more general formulations one has to rely on the full duality theory developed in [53, 58]. The class of stochastic programs considered are of the following (classical) type:*

$$\min \langle c, x^1 \rangle + \mathbb{E}[Q(\boldsymbol{\xi}, x^1)] \quad \text{subject to} \quad Ax^1 \geq b, \ x^1 \geq 0,$$

with

$$Q(\xi, x^1) = \inf \left\{ \langle q_\xi, x^2 \rangle \,\middle|\, W_\xi x^2 \geq d_\xi - T_\xi x^1, \ x^2 \geq 0 \right\},$$

where the matrices and vectors subscripted by ξ indicate that they depend on the realization of a (global) random variable $\boldsymbol{\xi}$ with support Ξ; for any fixed ξ: $x^1 \in I\!\!R^{n_1}, x_\xi^2 \in I\!\!R^{n_2}, A \in I\!\!R^{m_1 \times n_1}, b \in I\!\!R^{m_1}, q_\xi \in I\!\!R^{n_2}, W_\xi \in I\!\!R^{m_2 \times n_2}, d_\xi \in I\!\!R^{m_2}$ and $T_\xi \in I\!\!R^{m_2 \times n_1}$.

Detail. Let's assume,

relatively complete recourse: for all x^1 satisfying the (explicit) constraints $C^1 = \{ x^1 \in I\!\!R^{n_1} \,|\, A x^1 \geq b, \ x^1 \geq 0 \}$ and all $\xi \in \Xi$, one can always find a feasible recourse x_ξ^2, i.e., $Q(\xi, x^1) < \infty$ on $\Xi \times C^1$, and

strict feasibility: for some $\varepsilon > 0$, arbitrarily small, one can find $x^1 \in C^1$ and $\tilde{x}^2 \in \mathcal{L}_{n_2}^\infty$ such that $\tilde{x}^2 \geq_{a.s.} 0$ and for almost all ξ: $W_\xi \tilde{x}_\xi^2 > \tilde{\varepsilon} + d_\xi - T_\xi x^1$, where $\tilde{\varepsilon}$ is simply an m_2 dimensional vector of ε's.

For a problem of this type, $(\bar{x}^1, \bar{x}^2) \in I\!\!R_+^{n_1} \times \mathcal{L}_{n_2,+}^\infty$ is an optimal solution [54, 52] if and only if

(a) it is feasible, i.e., satisfies the constraints,

(b) $\exists$ multipliers $(y^1 \in I\!\!R^{m_1}, y^2 \in \mathcal{L}_{m_2}^1)$ such that

$$0 \leq y^1 \perp A\bar{x}^1 - b \geq 0, \quad 0 \leq_{a.s.} y_\xi^2 \perp_{a.s.} T_\xi \bar{x}^1 + W_\xi \bar{x}_\xi^2 - d_\xi \geq_{a.s.} 0,$$

(c) $(\bar{x}^1, \bar{x}^2)$ minimize

$$I\!\!E[\langle c - A^\top y^1 - T_\xi^\top y_\xi^2, x^1 \rangle + \langle q_\xi - W_\xi^\top y_\xi^2, x_\xi^2 \rangle]$$

for $x^1 \in I\!\!R_+^{n_1}$ and $x^2 \in \mathcal{L}_{n_2,+}^\infty$.

This means that under these assumptions, the double pairs

$$(x^1, x^2), \ (y^1, y^2) \in \left(I\!\!R_+^{n_1} \times \mathcal{L}_{n_2,+}^\infty \right) \times \left(I\!\!R_+^{m_1} \times \mathcal{L}_{m_2,+}^1 \right)$$

must satisfy the stochastic variational inequality:

$$0 \leq y^1 \perp A x^1 - b \geq 0,$$
$$0 \leq x^1 \perp c - A^\top y^1 - I\!\!E[T_\xi^\top y_\xi^2] \geq 0$$

and

$$0 \leq_{a.s.} y_\xi^2 \perp_{a.s.} T_\xi x^1 + W_\xi x_\xi^2 - d_\xi \geq_{a.s.} 0,$$
$$0 \leq_{a.s.} x_\xi^2 \perp_{a.s.} q_\xi - W_\xi^\top y_\xi^2 \geq_{a.s.} 0.$$

In terms of our general formulation in §1, the first pair of inequalities define the function G and the set $D = I\!\!R_+^{n_1} \times I\!\!R_+^{m_1}$ whereas the second pair define F and the random convex set $C_{\boldsymbol{\xi}}$ with (x^1, y^1) corresponding to x and (x^2, y^2) corresponding to u; $C_\xi = I\!\!R_+^{n_2} \times I\!\!R_+^{m_2}$. Of course, this can also be viewed as a stochastic complementarity problem albeit, in general, an infinite dimensional one. When, the probability distribution P has finite support, one can remove the "*a.s.*" in the second pair of inequalities and it's a problem involving only a finite number of variables and inequalities but this finite number might be truly considerable. $\qquad\square$

2.4 Example A Walras equilibrium problem. *In some ways, this example is an extension of the preceding one except that it doesn't lead to a stochastic complementarity problem but to a stochastic variational inequality that might not have the wished-for monotonicity properties for $E[G(\boldsymbol{\xi}, \cdot)]$ and F.*

Detail. We consider a stripped down version of the GEI-model, (General Equilibrium with Incomplete Markets), but even this model has a variety of immediate applications in finance, international commodity trading, species interaction in ecological models, ... The major difference with the extensive (economic) literature devoted to the GEI-model is the absence of a so-called financial market that allows agents to enter into contracts involving the delivery of goods at some future date.

Again, $\boldsymbol{\xi}$ provides the description of the uncertainty about the future events. We are dealing with a finite number of (individual) agents $i \in \mathcal{I}$. Each agent, endowed with vectors of goods $e_i^1 \in I\!\!R^L$ (here-and-now) and $e_{\xi,i}^2$ (in the future), choose its consumption plan, $c^{1\star}$ here-and-now and $c_\xi^{2\star}$ after observing ξ, so as to maximize their expected utilities $v_i^1(c^1) + E[v_i^2(c_\xi^2)]$, where the utility functions (v_i^1, v_i^2) are continuous, concave functions in (c^1, c^2) on closed convex sets $C_i^1 \subset I\!\!R_+^{n_1}$ and $C_i^2 \subset I\!\!R_+^{n_2}$ respectively. One often refers to C_i^1 and C_i^2 as agent-i's *survival* sets; in some situations it would be appropriate to let C_i^2 also depend on ξ, this wouldn't affect significantly our ensuing development. Each agent can also engage in here-and-now activities $y \in I\!\!R^{m_i}$ that will use up a vector of goods $T_i^1 y$ which, in turn, will generate, possibly depending on ξ, goods $T_{\xi,i}^2 y$ in the future; a simple example could be the saving of some goods and a more elaborate one involving "home production." The market process will allow each agent to exchange its (modified) endowments $(e_i^1 - T_i^1 y, e_{\xi,i}^2 + T_{\xi,i}^2 y)$ for its consumption at the prevalent market prices p^1 in the here-and-now market and p_ξ^2 in the future market. Thus, each agent has to find $(c_i^{1\star}, (c_{\xi,i}^{2\star}, \xi \in \Xi))$ that solves the following stochastic program with recourse:

$$\max \; v_i^1(c^1) + E[v_i^2(c_{\boldsymbol{\xi}}^2)]$$
$$\text{subject to } \langle p^1, c^1 + T_i^1 y \rangle \leq \langle p^1, e_i^1 \rangle,$$
$$\langle p_{\boldsymbol{\xi}}^2, c_{\boldsymbol{\xi}}^2 \rangle \leq_{a.s.} \langle p_{\boldsymbol{\xi}}^2, e_{\boldsymbol{\xi},i}^2 + T_{\boldsymbol{\xi},i}^2 y \rangle,$$
$$c^1 \in C_i^1, \; y \in I\!\!R_+^{m_i}, \; c_{\boldsymbol{\xi}}^2 \in_{a.s.} C_i^2, \; i \in \mathcal{I}.$$

Given that all agents will aim, selfishly, to maximize their expected rewards taking only into account their survival and budgetary limitations, the *Walras equilibrium* problem is to find a (nonnegative) price system $(p^1, (p_\xi^2, \xi \in \Xi))$ that will *clear the here-and-now market and the future markets*, i.e., such that the total demand does not exceed the available supply:

$$\sum\nolimits_{i \in \mathcal{I}} (e_i^1 - c_i^{1\star}) \geq 0, \qquad \sum\nolimits_{i \in \mathcal{I}} (e_{\xi,i}^2 - c_{\xi,i}^{2\star}) \geq 0, \; \forall_{a.s.} \, \xi \in \Xi.$$

It's common, as we do here, to assume[6], that the endowments of the agents are such that $e_i^1 \in \text{int} \, C_i^1$ and, for all $\xi \in \Xi$, $e_{\xi,i}^2 \in \text{int} \, C_i^2$. This will guarantee that this Walras equilibrium problem has a solution [33] and, in particular, that the agents' problems are stochastic programs

---

[6]This assumption, usually made to be able to rely on differential topology tools to prove existence of an equilibrium price system, is less than desirable from an applications viewpoint, cf. [34] for a workable assumption which would unfortunately require a significantly more detailed description of the basic model.

with relatively complete recourse and, consequently, the following optimality conditions apply: the pair $\left(c_i^{1\star}, y_i^\star, (c_{\xi,i}^{2\star}, \xi \in \Xi)\right)$ is an optimal solution for agent-i if and only if

(a) it satisfies the budgetary constraints,

(b) $\exists$ multipliers $\left(\lambda_i^1 \in I\!R, (\lambda_{\cdot,i}^2 \in \mathcal{L}^1)\right)$ such that

$$0 \le \lambda_i^1 \perp \langle p^1, e_i^1 - T_i^1 y_i^\star - c_i^{1\star}\rangle \ge 0, \qquad 0 \le_{a.s.} \lambda_{\xi,i}^2 \perp_{a.s.} \langle p_\xi^2, e_{\xi,i}^2 + T_{\xi,i}^2 y_i^\star - c_{\xi,i}^{2\star}\rangle \ge_{a.s.} 0,$$

(c) and

$$c_i^{1\star} \in \operatorname{argmax}_{c^1 \in C_i^1} v_i^1(c^1) - \lambda_i^1 \langle p^1, c^1\rangle,$$
$$c_{\xi,i}^{2\star} \in \operatorname{argmax}_{c^2 \in C_i^2} v_i^2(c^2) - \lambda_{\xi,i}^2 \langle p_\xi^2, c^2\rangle, \ \forall_{a.s.} \ \xi \in \Xi,$$
$$y_i^\star \in \operatorname{argmax}_{y \in R_+^{m_i}} -\lambda_i^1 \langle p^1, T_i^1 y\rangle + I\!E[\lambda_{\xi,i}^2 \langle p_\xi^2, T_{\xi,i}^2 y\rangle].$$

Assuming the utility functions are differentiable, these last conditions can be translated in terms of the first order optimality conditions for these programs:

$$\left(\nabla v_i^1(c_i^{1\star}) - \lambda_i^1 p^1\right) \in N_{C_i^1}(c_i^{1\star})$$
$$\left(\nabla v_i^2(c_{\boldsymbol{\xi},i}^{2\star}) - \lambda_{\boldsymbol{\xi},i}^2 p_{\boldsymbol{\xi}}^2\right) \in_{a.s.} N_{C_i^2}(c_{\boldsymbol{\xi},i}^{2\star})$$
$$0 \le y_i^\star \perp \left(\lambda_i^1 (T_i^1)^\top p^1 - I\!E[\lambda_{\boldsymbol{\xi},i}^2 (T_{\boldsymbol{\xi},i}^2)^\top p_{\boldsymbol{\xi}}^2]\right) \ge 0.$$

We can regroup these conditions so that fit the pattern of our two-stage formulation in §1: find $x = \left((p^1, c_i^1, y_i, \lambda_i^1), i \in \mathcal{I}\right)$ and $u_{\boldsymbol{\xi}} = \left((p_{\boldsymbol{\xi}}^2, c_{\boldsymbol{\xi},i}^2, \lambda_{\boldsymbol{\xi},i}^2), i \in \mathcal{I}\right)$ such that for all $i \in \mathcal{I}$:

$$\langle p^1, e_i^1 - c^1 - T_i^1 y_i\rangle \ge 0, \ \text{(feasibility)}$$
$$0 \le \lambda_i^1 \perp \langle p^1, e_i^1 - T_i^1 y_i - c_i^1\rangle \ge 0, \ \text{(multipliers complementarity)}$$
$$\left(\nabla v_i^1(c_i^1) - \lambda_i^1 p^1\right) \in N_{C_i^1}(c_i^1) \ \ (c^1\text{-optimality})$$
$$0 \le y_i \perp \left(\lambda_i^1 (T_i^1)^\top p^1 - I\!E[\lambda_{\boldsymbol{\xi},i}^2 (T_{\boldsymbol{\xi},i}^2)^\top p_{\boldsymbol{\xi}}^2]\right) \ge 0 \ \ (y\text{-optimality})$$
$$\sum_{i \in \mathcal{I}}(e_i^1 - c_i^1) \ge 0 \ \ \text{(clearing the market)}$$

and P-a.s.,

$$\langle p_{\boldsymbol{\xi}}^2, e_{\boldsymbol{\xi},i}^2 + T_{\boldsymbol{\xi},i}^2 y_i - c_{\boldsymbol{\xi},i}^2\rangle \ge_{a.s.} 0, \ \ \text{(feasibility)}$$
$$0 \le_{a.s.} \lambda_{\boldsymbol{\xi},i}^2 \perp_{a.s.} \langle p_{\boldsymbol{\xi}}^2, e_{\boldsymbol{\xi},i}^2 + T_{\boldsymbol{\xi},i}^2 y_i - c_{\boldsymbol{\xi},i}^2\rangle \ge_{a.s.} 0 \ \ \text{(multipliers complementarity)}$$
$$\left(\nabla v_i^2(c_{\boldsymbol{\xi},i}^2) - \lambda_{\boldsymbol{\xi},i}^2 p_{\boldsymbol{\xi}}^2\right) \in_{a.s.} N_{C_i^2}(c_{\boldsymbol{\xi},i}^2), \ \ (c^2\text{-optimality})$$
$$\sum_{i \in \mathcal{I}}(e_{\boldsymbol{\xi},i}^2 - c_{\boldsymbol{\xi},i}^2) \ge_{a.s.} 0. \ \ \text{(clearing the market)}$$

One approach in designing solution procedures for such a potentially whopping stochastic variational inequality is to attempt a straightforward approach relying, for example, on PATH Solver [18]. Notwithstanding, the capabilities of this excellent package, it is bound to be quickly overwhelmed by the size of this problem even when the number of agents and potential ξ-event is still

quite limited[7]. An approach based on decomposition is bound to be indispensable. One could rely on a per-agent decomposition first laid out in [24] and used in a variety of applications, cf. for example [50]. Another approach is via scenarios (events) based decomposition, relying on the Progressive Hedging algorithm [55, 56] and an approximation scheme as developed in [17]. Finally, one should also be able to expand on a splitting algorithm elaborated in §5 to allow for an agent/scenario decomposition expected to be reasonably competitive. □

The two last examples are stochastic variational inequalities that arise in connection with transportation/communication problems whose one must take into account uncertainties about some of the parameters of the problem. To fix terminology and notation, we begin with a brief description of the deterministic canonical model, the so called *Wardrop equilibirum model*; for excellent and thorough surveys of this model, see [14, 44] and as far as possible we follow their overall framework. There is no straightforward generalization to *the* "stochastic version" which is bound to very much depend on the motivation, or in other words, on the type of solution one is interested in. Here, we are going to be basically interested in problems where the uncertainty comes from the components of the underlying structure (network) or demand volume. Cominetti [13] and the references therein consider an interesting set of alternative questions primarily related to the uncertainty in the users' information and behavior.

Given an oriented network $\mathcal{N} = \big(\mathcal{G} \text{ (nodes)}, \mathcal{A} \text{ (arcs)}\big)$ together with for $c_a \geq 0$ the maximum flow capacity for each arc (a) and demand h_{od} for each origin(o)-destination(d) pairs. R_{od} are all (acyclic) routes r connecting o to d with N being the arcs(a)/routes(r) incidence matrix, i.e., $N_{a,r} = 1$ if arc $a \in r$. A route-flow $f = \big\{f_r, r \in \cup_{od} R_{od}\big\}$ results in an arc-flow $x = \big\{x_a = \langle N_a, f \rangle, a \in \mathcal{A}\big\}$. The travel time on route r, assumed to be additive, $\sum_{a \in r} t_a(x_a)$ where $t_a(\cdot)$, a congestion dependent function, specifies the travel time on arc a. Let

$$C = \Big\{(x_a = \langle N_a, f \rangle \leq c_a, a \in \mathcal{A}) \Big| f \geq 0, \sum_{r \in R_{od}} f_r = h_{od} \ \forall \ od\text{-pairs}\Big\}$$

be the polyhedral set of the arc-flows satisfying the flow conservation constraints. One refers to $x^\star = N f^\star \in C$ as a *Wardrop equilibrium* if

$$\forall \, od\text{-pairs}, \forall \, r \in R_{od} \text{ with } f_r^\star > 0, \ \sum_{a \in r} t_a(x_a^\star) \text{ is the minimal } od\text{-travel-time},$$

i.e., flow travels only on shortest routes. A minor variation, due to the capacity constraints on the arcs, of the authoritative argument of Beckmann *et all* [4], shows that these feasibility and equilibrium can be interpreted as the first order optimality condition of the convex program

$$\min \sum_{a \in \mathcal{A}} \int_0^{x_a} t_a(z) \, dz, \ x = (x_a, a \in \mathcal{A}) \in C.$$

Indeed, $x^\star$ is optimal if and only if it satisfies the *variational inequality*:

$$\sum_{a \in A} t_a(x_a^\star)(x_a - x_a^\star) \geq 0, \qquad \forall \, x \in C.$$

---

[7]Also homotopy-continuation method [16, 7, 11, 15] would also quickly run out of steam. A hybrid model, where first and second-period decisions are chosen simultaneously, is considered in [59], which because of its special structure attenuates to some extent the size issue and renders it computationally more accessible.

2.5 Example Prevailing flow analysis. *The main motivation in [10], which by the way is probably the first article to introduce a(n elementary) formulation of a two-stage stochastic variational inequality, was to determine the* steady flow *f that would prevail given an established network but where the origin(o)-destination(d) demand as well as the arcs capacities are subject to (stochastic) perturbations. The individual users would make adjustments to their steady-equilibrium routes, depending on these perturbations ξ, by choosing a recourse decision u_ξ that ends up to be the "nearest feasible" route to their steady-route. Although, one would actually like to place as few restrictions as possible on the class of recourse functions $(\xi, f) \mapsto u(\xi, f)$, from a modeling as well as a computational viewpoint, one might be satisfied with some restrictions or the application might dictate these restrictions as being the only really implementable "recourse" decisions; in [10] "nearest feasible" was interpreted as meaning the projection on the feasible region.*

Detail. Recasting the problem considered in [10] in our present notation, it would read: find (f, u_ξ) such that for all $\xi \in \Xi$,

$$-F(\xi, f, u_\xi) \in_{a.s.} N_{C_\xi}(u_\xi) \qquad \text{with} \quad u_\xi = \text{prj}_{C_\xi}(f)$$

and

$$C_\xi = \left\{ u \in \mathbb{R}^{n_2} \,\middle|\, Au = b_\xi, u \geq 0 \right\}.$$

This problem comes with no additional variational inequality, i.e., of the type $-\mathbb{E}[G(\cdot)] \in N_D(\cdot)$. Actually, in [10], it's assumed that one can restrict the choice of f to a set $D \subset \mathbb{R}_+^n$ which will guarantee that for all ξ, one can find u_ξ so that the corresponding variational inequality is solvable. This is a non-trivial assumption and it is only valid if we expect the perturbations of both the *od*-demand and those modifying the capacities to be relatively small, i.e., can essentially be interpreted as "noise" affecting these uncertain quantities. □

2.6 Example Capacity expansion. *Arc capacity expansion, $(c_a \to c_a + x_a, a \in \mathcal{A})$, is being considered for an existing, or to be designed, network (traffic, data transmission, high-speed rail, . . .). Only a limited pool of resources is available and thus, determine a number of constraints on the choice of x. To stay in tune with our general formulation of the two-stage model and provide a wide margin of flexibility in the formulation of these x-constraints, we express them as a variational inequality $\langle G(x), x' - x \rangle \geq 0$ for all $x' \in D$ where D is a closed convex subset of $\mathbb{R}^L$, $G : \mathbb{R}^L \to \mathbb{R}^L$ is a continuous vector-valued mapping and $L = |\mathcal{A}|$. The overall objective is to determine the arcs where capacities expansion will be most useful, i.e., will as well as possible respond to the "average" needs of the users of this network (minimal travel times, rapid connections, . . .). We interpret this to mean that the network flow will be at, or at least seek, a Wardrop equilibrium based on the information available: each od-pair demand and arcs' capacity both of which are subject to stochastic fluctuations. Taking our clue from the deterministic version described earlier, given a capacity expansion x and an environment $\xi \in \Xi$ affecting both demands and capacities, a solution $(u_{\xi,a}^\star, a \in \mathcal{A})$ of the following variational inequality would yield a Wardrop equilibrium:*

$$\sum_{a \in \mathcal{A}} t_a(\xi, u_{\xi,a}^\star)(u_a - u_{\xi,a}^\star) \geq 0, \quad \forall\, u = (u_a, a \in \mathcal{A}) \in C_\xi$$

with

$$C_\xi = \left\{ u \le c_\xi \ \middle| \ u = Nf, \ f \ge 0, \ \sum\nolimits_{r \in R_{od}} f_r = h_{\xi,od} \ \forall \ od\text{-pairs} \right\}.$$

Our two-stage stochastic variational inequality, cf. §1, would take the form:

$$\text{find } \left(x^\star, u^\star : \Xi \times D \to I\!\!R^L, \ \mathcal{A}\text{-measurable in } \xi \right) \text{ such that}$$
$$-G(x^\star) \in N_D(x^\star)$$
$$-F(\boldsymbol{\xi}, u^\star(\boldsymbol{\xi}, x^\star)) \in_{a.s.} N_{C_{\boldsymbol{\xi}}}(u^\star(\boldsymbol{\xi}, x^\star)),$$

where

$$F(\xi, u) = \left(t_a(\xi, u) \right)_{a \in \mathcal{A}}.$$

Detail. However, finding the solution to this particular stochastic variational inequality might not provide us with a well thought out solution to the network design problem: find optimal arc-capacities extension x that would result in expected minimal changes in the traffic flow when there are arbitrary, possibly significant, perturbations in the demand and the capacities. These concerns lead us to a modified formulation of this problem which rather than simply asking for a (feasible) solution of the preceding stochastic variational inequality wants to take into account a penalty cost associated with the appropriate recourse decisions when users are confronted with specific scenarios $\xi \in \Xi$. A "recourse cost" needs to be introduced. It's then also natural, and convenient computationally, to rely on an approach that allows for the possibility that in unusual circumstances some of the preceding variational inequalities might actually fail to be satisfied. This will be handled via a residual function. This application oriented reformulation results in a recasting of the problem and brings it closer to what has been called an Expected Residual Minimization or ERM-formulation. The analysis and the proposed solution procedure in the subsequent sections is devoted to this re-molded problem. Although this ERM-reformulation, see (4) below and §5, might usually[8] only provide an approximate solution of the preceding variational inequality, it provides a template for the design of optimal networks (traffic, communication,...) coming with both structural and demand uncertainties as argued in the next section. ☐

3 An Expected Residual Minimization formulation

We proceed with an ERM-formulation that in some instances might provide a solution which better suits the overall objective of the decision maker, cf. §5. Our focus will be on a particular class of stochastic variational inequalities which, in particular, include a number of traffic/communication problems (along the lines of Example 2.6):

$$\text{find } \left(\bar{x} \in I\!\!R^n, \bar{u} : \Xi \times D \to I\!\!R^n, \ \mathcal{A}\text{-measurable in } \xi \right) \text{ such that}$$
$$-G(\bar{x}) \in N_D(\bar{x}),$$
$$-F(\boldsymbol{\xi}, \bar{u}(\boldsymbol{\xi}, \bar{x})) \in_{a.s.} N_{C_{\boldsymbol{\xi}}}(\bar{u}(\boldsymbol{\xi}, \bar{x})),$$

---

[8]However, strict compliance can still be enforced in the ERM-reformulation by selecting the parameter λ in the objective of (4), below, sufficiently large.

where
$$\forall \xi \in \Xi: \ C_\xi \subset I\!\!R^n, \text{ closed, convex}, \quad \xi \mapsto C_\xi \ \mathcal{A}\text{-measurable},$$

D is closed and convex, and $G : I\!\!R^n \to I\!\!R^n$ and $F : \Xi \times I\!\!R^n \to I\!\!R^n$, as in §1, are continuous vector-valued functions in x and u respectively for all $\xi \in \Xi$, and F is $\mathcal{A}$-measurable in ξ for each $u \in I\!\!R^n$.

The dependence of F on x is assimilated in u by letting it depend on the state of the system determined by $(\boldsymbol{\xi}, x)$; it is thus convenient in the ensuing analysis to view F as just a function of $\boldsymbol{\xi}$ and u. A convenient interpretation is to think of x_a as the "average" flow that would pass through arc a when uncertainties aren't taken into account whereas, when confronted with scenario ξ, the actual flow would turn out to be $u_{\xi,a}$. Let's denote by $y_\xi = \big(y_{\xi,a} = u_{\xi,a} - x_a, \ a \in \mathcal{A}\big)$ the "recourse" required to adapt x_a to the events/scenario ξ, this will come at an expected cost (time delay) to the users. Since $\bar{x}$ is conceivably a Wardrop equilibrium, these adjustments will usually result in less desirable solutions. Such flow adjustments will come at a cost, say $\frac{1}{2}\langle y_\xi, Hy_\xi \rangle$, i.e. a generalized version of a least square cost proportional to the recourse decisions. This means that H, usually but not necessarily a diagonal matrix, would be positive definite. So, rather than viewing our decision variables as being u_ξ, we can equivalently formulate the recourse decisions in terms of the y_ξ, but now at some cost, and define $u_\xi = x + y_\xi$. In fact, let us go one step further and enrich our model by treating these y_ξ as activities[9] that result in a flow adjustment Wy_ξ and, hence, $u_\xi = x + Wy_\xi$; whenever $W = I$, the y_ξ just compensate for deviations from x.

When dealing with a deterministic variational inequality $-G(x) \in N_D(x)$, it is well known that its solution can be found by solving the following optimization problem

$$\min \theta(x) \text{ subject to } x \in D,$$

where θ is a *residual (gap-type) function*, i.e., such that $\theta \geq 0$ on D and $\theta(x) = 0$ if and only if $x \in D$ solves the variational inequality. In this article, we work with the following residual function, to which one usually refers as the regularized gap function [26]

$$\theta(x) = \max_{v \in D}\langle x - v, G(x)\rangle - \frac{\alpha}{2}\|x - v\|^2 \tag{1}$$

for some $\alpha > 0$. In terms of the overall framework of §1, this residual function-type will be attached to the inclusion $-G(x) := -I\!\!E[G(\xi, x)] \in N_D(x)$. The corresponding optimization problem reads

$$\min_{x \in D} \Big[\theta(x) = \max_{v \in D}\langle x - v, I\!\!E[G(\xi, x)]\rangle - \frac{\alpha}{2}\|x - v\|^2\Big]. \tag{2}$$

When dealing with the second inclusions $-F(\xi, u(\xi, x)) \in N_{C_\xi}(u(\xi, x))$ for P-almost all $\xi \in \Xi$, one could similarly introduce a collection of residual functions $\tilde{\theta}(\xi, u)$ whose properties would be similar to those of θ and ask that with probability 1, $\tilde{\theta}(\xi, u(\xi, x)) = 0$ if and only if the function $(\xi, x) \mapsto u(\xi, x)$ satisfies the corresponding stochastic variational inequality. This certainly might be appropriate in certain instances, but in a variety of problems one might want to relax this condition and replace it by a weaker criterion, namely that the *Expected Residual* $I\!\!E[\tilde{\theta}(\xi, u(\xi, x))]$ be

---

[9]Regulations introduced by the transportation-authority, the closing of certain links, etc.

minimal. A somewhat modified definition of a residual function will be more appropriate when dealing with this "collection of variational inequalities", we adopt here the one introduced in [10].

3.1 Definition (SVI-Residual function). *Given a closed, convex set $D \subset I\!\!R^n$ and the random vector $\boldsymbol{\xi}$ defined on $(\Xi, \mathcal{A}, P)$, let us consider the following collection of variational inequalities (SVI):*

$$\text{find } \left(\bar{x} \in I\!\!R^n, \bar{u} : \Xi \times D \to I\!\!R^n, \ \mathcal{A}\text{-measurable in } \xi \right) \text{ such that}$$
$$-F\left(\boldsymbol{\xi}, \bar{u}(\boldsymbol{\xi}, \bar{x})\right) \in_{a.s.} N_{C_{\boldsymbol{\xi}}}(\bar{u}(\boldsymbol{\xi}, \bar{x})).$$

A function $r : \Xi \times I\!\!R^n \to I\!\!R_+$ is a residual function for these inclusions if the following conditions are satisfied:

(i) *P-almost all $\xi \in \Xi$, $r(\xi, u) \geq 0$ for all $u \in C_\xi$,*

(ii) *For any $u : \Xi \times D \to I\!\!R^n$, it holds that*

$$-F\left(\xi, u(\xi, x)\right) \in_{a.s.} N_{C_\xi}(u(\xi, x)) \ \Leftrightarrow \ r(\xi, u(\xi, x)) =_{a.s.} 0 \text{ and } u(\xi, x) \in_{a.s.} C_\xi.$$

The stochastic variational inequality in this definition is in line with the second $a.s.$-inclusion in Example 2.6. We will work with a concrete example of SVI-residual functions: the regularized gap residual function with $\alpha > 0$,

$$r(\xi, u) = \max_{v \in C_\xi} \left\{ \langle u - v, F(\xi, u) \rangle - \frac{\alpha}{2} \|u - v\|^2 \right\}; \tag{3}$$

We will show in Theorem 3.3 that the above function satisfies the two conditions in Definition 3.1. The use of residual functions leads us to seeking a solution of the stochastic program

$$\begin{array}{ll} \min_{x \in D} & \theta(x) + \lambda I\!\!E[r(\xi, u(\xi, x)) + Q(\xi, x)] \\ \text{where} & u(\xi, x) = x + W y_\xi^*, \quad Q(\xi, x) = \frac{1}{2} \langle y_\xi^*, H y_\xi^* \rangle, \quad \forall_{a.s.} \ \xi \in \Xi, \\ & y_\xi^* = \operatorname{argmin} \left\{ \frac{1}{2} \langle y, H y \rangle \, \big| \, x + W y \in C_\xi \right\}. \end{array} \tag{4}$$

With the positive definite property of H and the convexity of C_ξ, $u(\xi, x)$ is uniquely defined by the unique solution y_ξ^* of the second stage optimization problem in (4). Moreover, with the residual functions θ and r defined in (1) and (3), respectively, the positive parameter λ in (4) allows for a further adjustment of the weight to ascribe to the required recourse decisions and residuals; with λ relatively large, and adjusting $H \succ 0$ correspondingly, one should end up with a solution that essentially avoids any violation of the collection (SVI) of variational inequalities[10].

---

[10] Problem (4) includes the expected value (EV) [28, 32, 41, 61] and the expected residual minimization (ERM) [1, 8, 10, 12, 22, 42, 69] for stochastic variational VI and stochastic complementarity problems as special cases. In particular, if we set $\lambda = 0$, then problem (4) is equivalent to the problem: find $x^* \in D$ such that

$$\langle (x - x^*), I\!\!E[G(\xi, x^*)] \rangle \geq 0, \ \forall \, x \in D.$$

On the other hand, if our variational inequality only involves inclusions of the (second) type $-F(\boldsymbol{\xi}, x) \in_{a.s.} N_C(x)$ with $C = C^\dagger$ independent of ξ, then problem (4) reduces to the (pure) ERM-formulation

$$\min_{x \in D = C^\dagger} I\!\!E[r(\boldsymbol{\xi}, x)];$$

for complementarity problems one has $C^\dagger \equiv I\!\!R_+^n$.

3.2 Assumption *Assume*

(i) W has full row rank;

(ii) for almost all ξ, $C_\xi \subset C^\dagger$, a compact convex set.

For almost all ξ, since C_ξ is convex and compact, it is easy to show that for these ξ, $u \mapsto r(\xi, u)$ is continuous. On the other hand, for each fixed u, it follows from [57, Theorem 14.37] that $\xi \mapsto r(\xi, u)$ is measurable. This means that r is a Carathéodory function. In addition, consider for each $\xi \in \Xi$,

$$v(\xi, u) = \text{prj}_{C_\xi}\left(u - \frac{1}{\alpha}F(\xi, u)\right). \tag{5}$$

Recall that $v(\xi, u)$ attains the maximum in (3). Clearly, $u \mapsto v(\xi, u)$ is continuous, and for each fixed u, the measurability of $\xi \mapsto v(\xi, u)$ again follows from [57, Theorem 14.37]. Consequently, also v is a Carathéodory function.

3.3 Theorem *(Residual function) When Assumption 3.2 is satisfied, r is a residual function for our collection (SVI) and for any $x \in D$ and almost every $\xi \in \Xi$, the function $r(\xi, u(\xi, x)) + Q(\xi, x)$ in (4) is finite, nonnegative with*

$$v(\xi, u(\xi, x)) = \text{prj}_{C_\xi}\left(u(\xi, x) - \frac{1}{\alpha}F(\xi, u(\xi, x))\right) \tag{6}$$

as the unique maximizer of the maximization problem in (3).

Proof. Let $x \in D$ and $u : \Xi \times D \to \mathbb{R}^n$ be $\mathcal{A}$-measurable in ξ. We now check the two conditions in Definition 3.1. First of all, we see that $r(\xi, u)$ is nonnegative for all $u \in C_\xi$ from the definition. Moreover, from the property of the regularized gap function for VI, we have $r(\xi, u(\xi, x)) = 0$ and $u(\xi, x) \in C_\xi$ if and only if $u(\xi, x)$ solves the (deterministic) variational inequality $-F(\xi, u(\xi, x)) \in N_{C_\xi}(u(\xi, x))$ for fixed $\xi \in \Xi$. Thus, it follows that r is a residual function of this collection of variational inequalities $-F(\xi, u(\xi, x)) \in_{a.s.} N_{C_\xi}(u(\xi, x))$, $\xi \in \Xi$.

We next show that for any $x \in D$, the y_ξ^* in (4) is, in fact, uniquely defined. To this end, consider $y = W^\top(WW^\top)^{-1}(\text{prj}_{C_\xi} x - x)$, where $WW^\top$ is invertible by Assumption 3.2. Then $Wy = \text{prj}_{C_\xi} x - x$ and $Wy + x \in C_\xi$ which means that the feasible set of the second stage optimization problem in (4), i.e., $\{y \mid x + Wy \in C_\xi\}$, is nonempty. Moreover, for almost all $\xi \in \Xi$, this set is closed and convex since C_ξ is compact and convex. Consequently, the second stage optimization problem in (4) has a strongly convex objective and a nonempty closed convex feasible set. Therefore, it has a unique minimizer y_ξ^*, and $\xi \mapsto y_\xi^*$ is measurable thanks to [57, Theorem 14.37].

Finally, for $u(\xi, x) = x + Wy_\xi^*$, we recall the well-known fact that the problem

$$\max_{v \in C_\xi} \langle u(\xi, x) - v, F(\xi, u(\xi, x)) \rangle - \frac{\alpha}{2}\|u(\xi, x) - v\|^2$$

has a solution, with the unique maximizer given by the projection (6). Thus, the function $r(\xi, u(\xi, x))$ in (4) is also finite-valued. Furthermore, since H is positive definite and r is nonnegative, $r(\xi, u(\xi, x)) + Q(\xi, x)$ has a finite nonnegative value at any $x \in D$ for almost every $\xi \in \Xi$. This completes the proof. $\qquad\square$

Theorem 3.3 means that under Assumption 3.2, problem (4) is a stochastic program with complete recourse, that is, for any x and almost all ξ, the second stage optimization problem in (4) has a solution. In general, this is not an innocuous assumption but in the context that we are considering it only means that the set D has singled-out a set of possible traffic network layouts that will guarantee that traffic will proceed, possibly highly disturbed and arduously, whatever be the perturbations resulting from the stochastic environment ξ.

We need the following to guarantee the objective of (4) is finite-valued:

3.4 Assumption *The functions $F(\xi, \cdot)$ and $G(\cdot)$ are continuously differentiable for all $\xi \in \Xi$. Moreover, for any compact set Ω, there are functions $d, \rho : \Xi \to I\!\!R_+$ such that*

$$\|F(\xi, u)\| \leq d_\xi \text{ and } \|\nabla F(\xi, u)\| \leq \rho_\xi$$

for all $u \in \Omega$, where $d \in \mathcal{L}_1^\infty$ and $\rho \in \mathcal{L}_1^1$.

3.5 Lemma *Suppose Assumptions 3.2 & 3.4 hold. Then for almost all ξ, r is continuously differentiable and its gradient is given by*

$$\nabla r(\xi, u) = F(\xi, u) - (\nabla F(\xi, u) - \alpha I)(v(\xi, u) - u). \tag{7}$$

Moreover, for any measurable $u_\xi \in_{a.s.} C_\xi$, both $\xi \mapsto r(\xi, u_\xi)$ and $\xi \mapsto \nabla r(\xi, u_\xi)$ are not only measurable but actually summable uniformly in u_ξ. In particular, this means that the objective function in (4) is well defined at any $x \in D$, and the optimal value of (4) is finite.

Proof. From Theorem 3.2 in [26], we know that $r(\xi, \cdot)$ is continuously differentiable and its gradient is given by (7). Moreover, notice that $r(\xi, u)$ and its gradients with respect to u are both Carathéodory functions. Hence, the measurability of $r(\xi, u_\xi)$ and $\nabla r(\xi, u_\xi)$ for any measurable function u_ξ follows. We now establish uniform summability.

Since $C^\dagger$ is compact, for an $u \in C^\dagger$, $\|u\| \leq \gamma$ for some $\gamma > 0$. This together with Assumption 3.4 yields $r(\xi, u_\xi) \leq 2\gamma d_\xi + \frac{\alpha}{2}(2\gamma)^2$ and

$$\|\nabla r(\xi, u_\xi)\| \leq \|F(\xi, u_\xi)\| + \|\nabla F(\xi, u_\xi) - \alpha I\|\|v(\xi, u_\xi) - u_\xi\| \leq d_\xi + 2\gamma(\rho_\xi + \alpha) =: d_\xi^r.$$

This proves uniform summability.

Finally, let x be feasible for (4) and consider the corresponding y_ξ^* (exists and is measurable according to the proof of Theorem 3.3). Set $\hat{y} = W^\top(WW^\top)^{-1}(\text{prj}_{C_\xi} x - x)$. Then $W\hat{y} + x \in C_\xi$ and hence we have from the definition of y_ξ^* and $C^\dagger$ that $\langle y_\xi^*, Hy_\xi^* \rangle \leq \langle \hat{y}, H\hat{y} \rangle \leq c$ for some constant $c > 0$ (that depends on x but is independent of ξ). Hence

$$\theta(x) + \lambda I\!\!E[r(\xi, u(\xi, x)) + Q(\xi, x)] \leq \theta(x) + 2\lambda\gamma I\!\!E[d_\xi] + 2\lambda\alpha\gamma^2 + \frac{\lambda c}{2},$$

which implies the well-definedness of the objective in (4) and the finiteness of the optimal value. This completes the proof. ☐

Lemma 3.5 establishes the differentiability of r and the finiteness of optimal value of (4). However, the objective function of problem (4) involves minimizers of constrained quadratic programs

for $\xi \in \Xi$ and is not necessarily differentiable even when the sample is finite, despite the fact that the function $r(\xi, \cdot)$ in (3) is continuously differentiable for almost all $\xi \in \Xi$.

Below, we adopt the strategy of the L-shaped algorithm for two-stage stochastic optimization problems [64, 6, 36] to obtain a relaxation of (4), whose objective function will be smooth when the sample is finite. We start by rewriting the recourse program. First, observe that the second stage problem is the same as

$$\min \frac{1}{2}\langle y, Hy \rangle \text{ subject to } Wy = u - x, \;\; u \in C_\xi.$$

With the substitution $z = H^{\frac{1}{2}}y$, the above problem is equivalent to

$$\min \frac{1}{2}\|z\|^2 \text{ subject to } WH^{-\frac{1}{2}}z = u - x, \;\; u \in C_\xi. \tag{8}$$

Observe that for each fixed u and x, the minimizer of $\min\left\{\frac{1}{2}\|z\|^2 \,\middle|\, WH^{-\frac{1}{2}}z = u - x\right\}$ is

$$z = H^{-\frac{1}{2}}W^\top B(u - x) \text{ where } B = (WH^{-1}W^\top)^{-1}. \tag{9}$$

Plugging this expression back in (8), the second stage problem is further equivalent to

$$\min \frac{1}{2}\langle u - x, B(u - x)\rangle \text{ subject to } u \in C_\xi,$$

whose unique solution u_ξ^* can be interpreted as a weighted projection of x onto C_ξ for each $\xi \in \Xi$. From this and (9), for each $\xi \in \Xi$, the unique solution y_ξ^* of the second stage problem is given by

$$y_\xi^* = \operatorname{argmin}\left\{\frac{1}{2}\langle y, Hy \rangle \,\middle|\, x + Wy \in C_\xi\right\} = H^{-1}W^\top B(u_\xi^* - x). \tag{10}$$

Combining the preceding reformulation of the second stage problem with the idea of the L-shaped algorithm, we are led to the following problem, whose objective is smooth when the sample is finite:

$$\begin{aligned} \min \quad & \theta(x) + \lambda \mathbb{E}[r(\xi, u_\xi) + \tfrac{1}{2}\langle u_\xi - x, B(u_\xi - x)\rangle] \\ \text{subject to} \quad & x \in D, \quad u_\xi \in_{a.s.} C_\xi. \end{aligned} \tag{11}$$

It is not hard to see that the optimal value of (11) is smaller than that of (4) since fewer restrictions are imposed on u_ξ, or, equivalently on y_ξ in (10). Hence, it follows from Lemma 3.5 that the optimal value of (11) is also finite.

Unlike (4) which only depends explicitly on the finite dimensional decision variable x, problem (11) depends also explicitly on the measurable function $\xi \mapsto u_\xi$. However, notice that (11) can be equivalently rewritten as

$$\min_{x \in D} \left\{\theta(x) + \lambda \min_{u \in \mathcal{L}_n^\infty} \mathbb{E}\left[r(\xi, u_\xi) + \frac{1}{2}\langle u_\xi - x, B(u_\xi - x)\rangle + \delta_{C_\xi}(u_\xi)\right]\right\}, \tag{12}$$

where δ_{C_ξ} is the indicator function of the set C_ξ which is zero inside the set and is infinity otherwise. We next recall the following result, which allows interchange of minimization and integration. This is a consequence of the finiteness of the optimal value of the inner minimization in (12) (thanks to Lemma 3.5), Exercise 14.61 and Theorem 14.60 (interchange of minimization and integration) in [57].

3.6 Lemma *Under Assumptions 3.2 & 3.4, for any fixed $x \in D$, we have*

$$\min_{u \in \mathcal{L}_n^\infty} \mathbb{E}[\Phi(\xi, u_\xi)] = \mathbb{E}[\min_{u \in R^n} \Phi(\xi, u)], \tag{13}$$

where

$$\Phi(\xi, u) = r(\xi, u) + \frac{1}{2}\langle u - x, B(u - x)\rangle + \delta_{C_\xi}(u).$$

Moreover, for $\bar{u} \in \mathcal{L}_n^\infty$,

$$\bar{u} \in \underset{u \in \mathcal{L}_n^\infty}{\mathrm{argmin}}\, \mathbb{E}[\Phi(\xi, u_\xi)] \quad \Leftrightarrow \quad \bar{u}_\xi \in \underset{u \in R^n}{\mathrm{argmin}}\, \Phi(\xi, u), \quad \xi \in \Xi, \ a.s. \tag{14}$$

Using the above result, one can further reformulate (11) equivalently as

$$\min_{x \in D} \varphi(x) := \theta(x) + \lambda \mathbb{E}[\psi(\xi, x)]$$
$$\psi(\xi, x) := \min_{u \in C_\xi} r(\xi, u) + \frac{1}{2}\langle u - x, B(u - x)\rangle, \tag{15}$$

which is an optimization problem that depends only explicitly on the finite dimensional decision variable x. Moreover, for each $x \in D$, from (14), we have $u^* \in \mathcal{L}_n^\infty$ attaining the minimum of the inner minimization problem in (12) if and only if its value at ξ, i.e., u_ξ^*, is a minimizer of the minimization problem defining $\psi(\xi, x)$ for almost all $\xi \in \Xi$. We note that $\psi(\xi, x)$ is also a Carathéodory function: the measurability with respect to ξ for each fixed x follows from [57, Theorem 14.37], while the continuity with respect to x for almost all ξ follows from the compactness of C_ξ and the continuity of $(u, x) \mapsto r(\xi, u) + \frac{1}{2}\langle u - x, B(u - x)\rangle$. In addition, thanks to Lemma 3.5, it is not hard to check that the objective in (15) is finite at any $x \in D$ under Assumptions 3.2 & 3.4.

We show next that both (4) and (11) (and hence (15)) are solvable, and discuss their relationship.

3.7 Theorem *(Solvability) Suppose Assumptions 3.2 & 3.4 hold. Then problems (4) and (11) are solvable. Let ν_1 and ν_2 be the optimal values of (4) and (11), respectively. Then $\nu_1 \geq \nu_2$. Moreover, if for any $x \in D$ and $x + Wy$, $x + Wz \in C_\xi$, we have*

$$|r(\xi, x + Wy) - r(\xi, x + Wz)| \leq \frac{1}{2}|\langle y, Hy \rangle - \langle z, Hz \rangle|, \quad \xi \in \Xi, \quad a.s. \tag{16}$$

then the two problems have the same optimal values.

Proof. According to Lemma 3.5 and the preceding discussions, the optimal value of both problems (4) and (11) are finite. We show that the values are attained.

We first show that the optimal value of problem (4) is attained. To this end, consider $\|x\| \to \infty$. Then from the boundedness of $C^\dagger \supseteq C_\xi$, it follows that $\|y_\xi^*\|$ in (4) must go to ∞ uniformly in ξ except for a set of measure zero, which implies that $\mathbb{E}[Q(\xi, x)]$ goes to ∞ in (4). This together with the nonnegativity of r and θ shows that the objective function of (4), as a function of x, is level bounded. Next, we show that the objective is lower semicontinuous. To this end, consider a sequence $x_k \to x_0$. From the discussions in (8) to (10) and using the continuity of weighted projections, we see that $u(\xi, x_k) \to u(\xi, x_0)$ for almost all $\xi \in \Xi$. This implies the continuity of $x \mapsto r(\xi, u(\xi, x)) + Q(\xi, x)$, which is a nonnegative function, and Fatou's lemma gives the lower semicontinuity of $\mathbb{E}[r(\xi, u(\xi, x)) + Q(\xi, x)]$. The lower semicontinuity of the objective of (4) now follows upon recalling that $\theta(x)$ is continuous. Hence (4) has a solution x^*, from which one can easily construct the second stage solution y_ξ^*.

We now show that the optimal value of problem (11) is attained. Note that from the discussion preceding this lemma, one can equivalently consider problem (15). For this latter problem, observe that

$$\varphi(x) \geq \lambda \mathbb{E}[\psi(\xi, x)] \geq \frac{\lambda}{2} \mathbb{E}[\min_{u \in C_\xi} \langle u - x, B(u - x) \rangle] \geq \frac{\lambda}{2} \min_{u \in C^\dagger} \langle u - x, B(u - x) \rangle,$$

where the first two inequalities follow from the nonnegativity of the residual functions, and the third inequality follows from $C_\xi \subseteq C^\dagger$. The above relation together with the positive definiteness of B and the compactness of $C^\dagger$ shows that the objective function of (15) is level bounded. We note also that the objective is lower semicontinuous, which is a consequence of the continuity of $\theta(x)$, the continuity of $\psi(\xi, x)$ in x and Fatou's lemma. Hence an optimal solution x^* exists, from which the corresponding u_ξ^* that minimizes the subproblem defining $\psi(\xi, x)$ can be obtained easily. From this and the relationship between the solutions of (11) and (15), one concludes that a solution of problem (11) exists.

Next, we consider the relationship between the optimal values of (4) and (11), which we denote by ν_1 and ν_2, respectively. Obviously, $\nu_1 \geq \nu_2$.

Suppose in addition that (16) holds, then $\nu_1 = \nu_2$. Indeed, to this end, let (x^2, u_ξ^2) be a solution of (11) and x^1 be a solution of (4). Set

$$y_\xi^{2*} = \operatorname{argmin}\left\{ \frac{1}{2}\langle y, Hy \rangle \,\middle|\, x^2 + Wy \in C_\xi \right\}$$

and $y_\xi^2 = H^{-1}W^\top B(u_\xi^2 - x^2)$. Since $u_\xi^2 = x^2 + Wy_\xi^2 \in C_\xi$, we have

$$\langle y_\xi^{2*}, Hy_\xi^{2*} \rangle - \langle y_\xi^2, Hy_\xi^2 \rangle \leq 0.$$

Using this inequality, condition (16) and $\langle y_\xi^2, Hy_\xi^2 \rangle = \langle u_\xi^2 - x^2, B(u_\xi^2 - x^2) \rangle$ yields

$$
\begin{aligned}
\nu^1 &= \theta(x^1) + \lambda E[r(\xi, u(\xi, x^1)) + Q(\xi, x^1)] \\
&\leq \theta(x^2) + \lambda E[r(\xi, u(\xi, x^2)) + Q(\xi, x^2)] \\
&= \theta(x^2) + \lambda E[r(\xi, x^2 + Wy_\xi^{2*}) + \frac{1}{2}\langle y_\xi^{2*}, Hy_\xi^{2*} \rangle] \\
&\leq \theta(x^2) + \lambda E[r(\xi, u_\xi^2) + \frac{1}{2}\langle y_\xi^2, Hy_\xi^2 \rangle] \\
&\quad + \lambda E[|r(\xi, x^2 + Wy_\xi^{2*}) - r(\xi, x^2 + Wy_\xi^2)| - \frac{1}{2}|\langle y_\xi^{2*}, Hy_\xi^{2*} \rangle - \langle y_\xi^2, Hy_\xi^2 \rangle|] \\
&\leq \theta(x^2) + \lambda E[r(\xi, u_\xi^2) + \frac{1}{2}\langle u_\xi^2 - x^2, B(u_\xi^2 - x^2) \rangle] = \nu^2.
\end{aligned}
$$

Therefore, problems (4) and (11) have the same optimal values. $\qquad\square$

Similarly to Lemma 3.5, we now study the differentiability of the objective function of problem (11). Let

$$
f(\xi, x, u) = r(\xi, u) + \frac{1}{2}\langle u - x, B(u - x) \rangle,
$$

where r is given by (3). The proof of the following lemma is similar to that of Lemma 3.5 and will thus be omitted.

3.8 Proposition *Suppose that Assumptions 3.2 & 3.4 hold. Then for almost all ξ, f is continuously differentiable and its gradient is given by*

$$
\nabla f(\xi, x, u) = \begin{pmatrix} \nabla_u f(\xi, x, u) \\ \nabla_x f(\xi, x, u) \end{pmatrix} = \begin{pmatrix} \nabla r(\xi, u) + B(u - x) \\ B(x - u) \end{pmatrix}. \tag{17}
$$

Moreover, $f(\xi, x, u_\xi)$ and $\nabla f(\xi, x, u_\xi)$ are measurable in ξ. Furthermore, for any compact set $\bar{D} \subseteq D$, there are $d^f, \rho^f \in \mathcal{L}_1^1$ such that $\|f(\xi, x, u_\xi)\| \leq d_\xi^f$ and $\|\nabla f(\xi, x, u_\xi)\| \leq \rho_\xi^f$ whenever $u_\xi \in C_\xi$ and $x \in \bar{D}$.

In general, a two-stage stochastic VI cannot be reformulated as a convex two-stage stochastic optimization problem. Below, we show that convexity is inherited partially in f from a linear two-stage stochastic optimization problem. This proposition will be used in the next section to establish, for this particular class of problems, a stronger convergence result for linear two-stage stochastic optimization problems.

3.9 Proposition *Under Assumptions 3.2 & 3.4, if for $\xi \in \Xi$, $F(\xi, x) = M_\xi x + q_\xi$ and $\inf_{\xi \in \Xi} \lambda_{\min}(B + M_\xi + M_\xi^\top) > \alpha$, then $f(\xi, x, \cdot)$ is strongly convex for each $x \in D$, with a strong convexity modulus independent of x and ξ. Moreover, if $\inf_{\xi \in \Xi} \lambda_{\min}(M_\xi + M_\xi^\top) \geq \alpha$, then $f(\xi, \cdot, \cdot)$ is convex.*

Proof. For a fixed ξ, notice that the function

$$
f(\xi, x, u) = \max_{v \in C_\xi} \langle u - v, M_\xi u + q_\xi \rangle - \frac{\alpha}{2}\|u - v\|^2 + \frac{1}{2}\langle u - x, B(u - x) \rangle
$$

19

is convex in u if the function

$$\varphi(\xi, x, u, v) = \langle u - v, M_\xi u + q_\xi \rangle - \frac{\alpha}{2}\|u - v\|^2 + \frac{1}{2}\langle u - x, B(u - x)\rangle$$

is convex in u for any fixed $v \in I\!R^n$ and $x \in I\!R^n$. Since

$$\nabla_{uu}^2 \varphi(\xi, x, u, v) = M_\xi + M_\xi^\top - \alpha I + B$$

is positive definite, $\varphi(\xi, x, \cdot, v)$ and thus $f(\xi, x, \cdot)$ is convex. The independence of the modulus on x can be seen from the fact that $\nabla_{uu}^2 \varphi(\xi, x, u, v)$ does not depend on x, while the independence on ξ can be seen from the assumption on the eigenvalue. Finally, when $\inf_{\xi \in \Xi} \lambda_{\min}(M_\xi + M_\xi^\top) \geq \alpha$, the convexity of $f(\xi, \cdot, \cdot)$ in (x, u) follows from the positive semi-definiteness of the matrix

$$\begin{pmatrix} \nabla_{uu}^2 \varphi(\xi, x, u, v) & \nabla_{ux}^2 \varphi(\xi, x, u, v) \\ \nabla_{xu}^2 \varphi(\xi, x, u, v) & \nabla_{xx}^2 \varphi(\xi, x, u, v) \end{pmatrix} = \begin{pmatrix} M_\xi + M_\xi^\top - \alpha I + B & -B \\ -B & B \end{pmatrix} \in I\!R^{2n \times 2n}.$$

This completes the proof. □

3.10 Remark *In [1], Agdeppa, Yamashita and Fukushima defined a convex ERM formulation for the stochastic linear VI by using the regularized gap function under the assumption that $\inf_{\xi \in \Xi} \lambda_{\min}(M_\xi + M_\xi^\top) > \alpha$. In [10], Chen, Wets and Zhang defined a convex ERM formulation for the stochastic linear VI by using the gap function under a weaker assumption that $E[M_\xi + M_\xi^\top]$ is positive semi-definite. In this paper, Lemma 3.9 shows that for any stochastic linear VI, we can choose a positive definite matrix B to define a convex optimization problem in the second stage of the generalized ERM formulation (11).*

To end this section, we include a corollary concerning the entire objective function of (11).

3.11 Corollary *Suppose that Assumptions 3.2 & 3.4 hold. Then the objective function $\theta(x) + \lambda I\!E[f(\xi, x, u_\xi)]$ is well defined at any $x \in D$ and any measurable function $u_\xi \in C_\xi$. Moreover, if $G(x) = I\!E[M_\xi]x + I\!E[q_\xi]$, then the objective function is convex under all the assumptions of Proposition 3.9.*

4 Convergence analysis

In this section, we discuss a sample average approximation (SAA) for (15) (and hence, equivalently, (11)) and derive its convergence. Let $G(x) = I\!E[F(\xi, x)]$ and $\xi^1, \ldots, \xi^\nu$ be iid (independent and identically distributed) samples of $\boldsymbol{\xi}$ and

$$G^\nu(x) = \frac{1}{\nu}\sum_{i=1}^{\nu} F(\xi^i, x)$$

$$\theta^\nu(x) = \max_{v \in D}\langle x - v, G^\nu(x)\rangle - \frac{\alpha}{2}\|x - v\|^2$$

$$\psi(\xi^i, x) = \min_{u \in C_{\xi^i}} r(\xi^i, u) + \frac{1}{2}\langle u - x, B(u - x)\rangle.$$

20

Note that G^ν and θ^ν are continuous P-a.s. for all ν, and recall that ψ is a Carathéodory function; cf. the discussion following (15). We consider the following SAA of (15) (and hence, equivalently, (11)):

$$\min_{x\in D} \varphi^\nu(x) := \theta^\nu(x) + \frac{\lambda}{\nu}\sum_{i=1}^{\nu}\psi(\xi^i, x). \tag{18}$$

Let X^* and X^ν be the solution sets of problems (15) and (18). We will give sufficient conditions for that X^* and X^ν to be nonempty and bounded, and for any cluster point of a sequence $\{x^\nu\}_{x^\nu\in X^\nu}$ to be in X^*.

4.1 Lemma *Under Assumptions 3.2 & 3.4, X^* and X^ν are nonempty P-a.s. for all ν. Moreover, there exists a compact convex set $\bar{D} \subseteq D$ so that $X^* \subseteq \bar{D}$ and $X^\nu \subseteq \bar{D}$ P-a.s. for all ν.*

Proof. From the nonnegativity of $r(\xi, u)$, we have for almost all $\xi \in \Xi$,

$$
\begin{aligned}
\psi(\xi, x) &= \min_{u\in C_\xi} r(\xi, u) + \frac{1}{2}\langle u - x, B(u - x)\rangle \\
&\geq \min_{u\in C_\xi} \frac{1}{2}\langle u - x, B(u - x)\rangle \\
&\geq \frac{\lambda_B}{2}\min_{u\in C^\dagger}\|u - x\|^2 \\
&= \frac{\lambda_B}{2}\|\operatorname{prj}_{C^\dagger} x - x\|^2,
\end{aligned}
$$

where $\lambda_B > 0$ is the smallest eigenvalue of the matrix B. On the other hand, recall from Assumption 3.2 that there exists $\tau > 0$ so that $\|u\| \leq \tau$ for all $u \in C^\dagger$, and from Assumption 3.4 that $\|F(\xi, u)\| \leq d_\xi$ for some $d \in \mathcal{L}_1^\infty$. Hence, for a fixed $x_0 \in D$, we have

$$\psi(\xi^\nu, x_0) \leq \sup_{u\in C^\dagger} r(\xi^\nu, u) + \frac{\|B\|}{2}(\tau + \|x_0\|)^2 \leq 2\tau d_{\xi^\nu} + \frac{\alpha}{2}(2\tau)^2 + \frac{\|B\|}{2}(\tau + \|x_0\|)^2 \leq \gamma_1$$

for some constant γ_1 P-a.s. for all ν, since $d \in \mathcal{L}_1^\infty$. Furthermore, we have

$$\theta^\nu(x_0) = \left\langle x_0 - \operatorname{prj}_D\left(x_0 - \frac{1}{\alpha}G^\nu(x_0)\right), G^\nu(x_0)\right\rangle - \frac{\alpha}{2}\left\|x_0 - \operatorname{prj}_D\left(x_0 - \frac{1}{\alpha}G^\nu(x_0)\right)\right\|^2 \leq \gamma_2$$

P-a.s. for all ν, since $F(\cdot, x_0) \in \mathcal{L}_1^\infty$.

Hence we have

$$
\begin{aligned}
\{x \in D \mid \varphi^\nu(x) \leq \varphi^\nu(x_0)\} &\subseteq \left\{x \in D \,\middle|\, \frac{\lambda}{\nu}\sum_{i=1}^{\nu}\psi(\xi^i, x) \leq \varphi^\nu(x_0)\right\} \\
&\subseteq \left\{x \in D \,\middle|\, \frac{\lambda}{\nu}\sum_{i=1}^{\nu}\psi(\xi^i, x) \leq \lambda\gamma_1 + \gamma_2\right\} \subseteq \{x \in D \mid \lambda\cdot\lambda_B\|\operatorname{prj}_{C^\dagger} x - x\|^2 \leq 2(\lambda\gamma_1 + \gamma_2)\}
\end{aligned}
\tag{19}
$$

P-a.s. for all ν, and that

$$\{x \in D \mid \varphi(x) \leq \varphi(x_0)\} \subseteq \{x \in D \mid \lambda\cdot\lambda_B\|\operatorname{prj}_{C^\dagger} x - x\|^2 \leq 2\varphi(x_0)\}. \tag{20}$$

These, together with the boundedness of $C^\dagger$, imply that the level sets of $\varphi^\nu(\cdot)$ and $\varphi(\cdot)$, are bounded and are included in a compact set that is independent of ξ, P-a.s. for all ν. Moreover, the objective functions of (15) and (18) are all lower semicontinuous, and θ, θ^ν, $\psi(\xi, x)$ are nonnegative. Hence X^* and X^ν are nonempty P-a.s. for all ν. Finally, from (19) and (20), one can choose for $\bar{D} = \{x \in D \mid \lambda \cdot \lambda_B \| \operatorname{prj}_{C^\dagger} x - x \|^2 \le 2 \max\{\varphi(x_0), \lambda\gamma_1 + \gamma_2\}\}$. $\qquad\square$

4.2 Theorem *(Convergence theorem) Suppose Assumptions 3.2 & 3.4 hold. Then φ^ν converges to φ a.s.-uniformly on the compact set $\bar{D}$ specified in Lemma 4.1. Let $\{x^\nu\}$ be a sequence of minimizers of problems (18) generated by iid samples. Then $\{x^\nu\}$ is P-a.s. bounded and any accumulation point x^* of $\{x^\nu\}$ as $\nu \to \infty$ is P-a.s. a solution of (15).*

Proof. From Lemma 4.1, the convex compact set $\bar{D} \subseteq D$ is such that $X^* \subseteq \bar{D}$,

$$\min_{x \in \bar{D}} \varphi(x) = \min_{x \in D} \varphi(x) = \min_{x \in R^n} \varphi(x) + \delta_D(x)$$

and for P-a.s. all ν,

$$\min_{x \in \bar{D}} \varphi^\nu(x) = \min_{x \in D} \varphi^\nu(x) = \min_{x \in R^n} \varphi^\nu(x) + \delta_D(x).$$

Note that $\varphi^\nu(x) = \theta^\nu(x) + \frac{\lambda}{\nu} \sum_{i=1}^\nu \psi(\xi^i, x)$. To analyze the convergence, we first show that $\theta^\nu(x)$ converges to $\theta(x)$ P-a.s. uniformly on $\bar{D}$. Since $\bar{D}$ is a nonempty compact subset of $\mathbb{R}^n$, $F(\xi, \cdot)$ is continuous at x for almost every $\xi \in \Xi$, $\|F(\xi, x)\| \le d_\xi$ for some $d \in \mathcal{L}_1^\infty$ due to Assumption 3.4, and the sample is iid, we can apply Theorem 7.48 in [61] to claim that $G^\nu(x)$ converges to $G(x)$ a.s.-uniformly on $\bar{D}$, that is, for any $\varepsilon > 0$, there is $\hat{\nu}$ such that for any $\nu \ge_{a.s.} \hat{\nu}$, one has

$$\sup_{x \in \bar{D}} |G(x) - G^\nu(x)| < \varepsilon.$$

Moreover, from the definition of θ in (1), we have

$$\theta(x) = \langle x - v(x), G(x) \rangle - \frac{\alpha}{2} \| x - v(x) \|^2$$

$$\theta^\nu(x) = \langle x - v^\nu(x), G^\nu(x) \rangle - \frac{\alpha}{2} \| x - v^\nu(x) \|^2,$$

where

$$v(x) = \operatorname{prj}_D \left(x - \frac{1}{\alpha} G(x) \right) \quad \text{and} \quad v^\nu(x) = \operatorname{prj}_D \left(x - \frac{1}{\alpha} G^\nu(x) \right).$$

Obviously, one has for $\nu \ge_{a.s.} \hat{\nu}$,

$$\sup_{x \in \bar{D}} \| v(x) - v^\nu(x) \| \le \sup_{x \in \bar{D}} \frac{1}{\alpha} \| G(x) - G^\nu(x) \| < \frac{\varepsilon}{\alpha}.$$

From the boundedness of $\bar{D}$ and the continuity of G, there is $\gamma > 0$ such that $\max\{\|z\|, \|\operatorname{prj}_D(z)\|, \|G(z)\|\} \le \gamma$ for any $z \in \bar{D}$. Hence we obtain for $\nu \ge_{a.s.} \hat{\nu}$,

$$\begin{aligned}
\sup_{x \in \bar{D}} |\theta(x) - \theta^\nu(x)| &\le \sup_{x \in \bar{D}}(\| v^\nu(x) - x \| \| G(x) - G^\nu(x) \| + \| v(x) - v^\nu(x) \| \| G(x) \| \\
&\qquad + 2\gamma\alpha \| v(x) - v^\nu(x) \|) \\
&< (4 + \frac{1}{\alpha})\gamma\varepsilon.
\end{aligned}$$

Hence, $\theta^\nu(x)$ converges to $\theta(x)$ a.s.-uniformly on $\bar{D}$. Similarly, again by using Theorem 7.48 in [61], we can show that $\frac{1}{\nu}\sum_{i=1}^{\nu}\psi(\xi^i, x) \to \mathbb{E}[\psi(\xi, x)]$ a.s.-uniformly on $\bar{D}$. Consequently, φ^ν converges to φ a.s.-uniformly on $\bar{D}$ as claimed.

Combining the convergence result with Theorem 7.11, Theorem 7.14 and Theorem 7.31 in [57], we obtain further that

$$\limsup_{\nu\to\infty} \mathrm{argmin}_{x\in\bar{D}}\varphi^\nu(x) \subseteq \mathrm{argmin}_{x\in\bar{D}}\varphi(x).$$

Since $\bar{D}$ is compact, we conclude further that $\{x^\nu\}$ is a.s. bounded and any accumulation point of $\{x^\nu\}$, as $\nu \to \infty$, is a.s. a solution x^* of (15). $\qquad\square$

It is worth noting that the function

$$\varphi(x) = \big(\max_{v\in D}\langle x - v, \mathbb{E}[G(\xi, x)]\rangle - \frac{\alpha}{2}\|x - v\|^2\big) + \lambda\mathbb{E}[\psi(\xi, x)]$$

is the sum of an EV formulation for the first stage and an ERM formulation for the second stage of the two stage stochastic VI. The convergence analysis of SAA for the EV formulation of a single stage stochastic VI in [61] cannot be directly applied to φ^ν.

We next present a result in the presence of convexity: when $F(\xi, \cdot)$ is an affine mapping, we can obtain further results if a first-order growth condition is satisfied. For the same reason explained above, the corresponding results in [61] cannot be directly applied.

4.3 Theorem *(Convergence theorem for the convex problem) In addition to Assumptions 3.2 & 3.4, suppose that for $\xi \in \Xi$, $F(\xi, u) = M_\xi u + q_\xi$. Let $\theta(x)$ be defined by (1) with $G(x) = \mathbb{E}(F(\xi, x))$ and $\inf_{\xi\in\Xi}\lambda_{\min}(M_\xi + M_\xi^\top) > \alpha$. Then φ and φ^ν are strongly convex P-a.s. for all ν. Let $\{x^\nu\}$ be a sequence of minimizers of problems (18) and the samples be iid. Then φ^ν converges to φ a.s.-uniformly on the compact set $\bar{D}$ specified in Lemma 4.1. Moreover, let x^* be the unique solution of (15) such that*

$$\varphi(x) \geq \varphi(x^*) + c\|x - x^*\|, \quad \forall x \in \bar{D}, \tag{21}$$

where $c > 0$ is a constant.[11] *Then a.s., we have $x^\nu = x^*$ for all sufficiently large ν.*

Proof. From Theorem 3.2 in [26], the functions θ and θ^ν defined by the regularized gap function are strongly convex and continuously differentiable P-a.s. for all ν. Moreover, we also have the convexity of $\psi(\xi, \cdot)$ and $\psi^\nu(\xi, \cdot)$ from the second conclusion in Proposition 3.9. Hence φ and φ^ν are strongly convex P-a.s. for all ν.

The a.s.-uniform convergence of φ^ν to φ on $\bar{D}$ is obtained from Theorem 4.2. Alternatively, we have a much simpler proof thanks to the presence of convexity: since $\varphi^\nu(x) \to \varphi(x)$ at each $x \in \bar{D}$ a.s., we have $\varphi^\nu(x) \to \varphi(x)$ for every x in any countable dense subset of $\bar{D}$ a.s., and consequently, using Theorem 10.8 from [51], we can also conclude that φ^ν converges to φ a.s.-uniformly on $\bar{D}$.

Next, assume in addition that (21) holds at the unique solution x^* of (15). We first recall from [61, Theorem 7.54] that using the convexity and the differentiability of $\varphi^\nu(\cdot)$ and $\varphi(\cdot)$, we

---

[11]The solution x^* is called a sharp solution in [61, Definition 5.22]. A sufficient condition for (21) is that $\min_{1\leq i\leq n}|\nabla\varphi(x^*)|_i > c$ and $\min_{1\leq i\leq n}(\nabla\varphi(x^*))_i(x - x^*)_i \geq 0, \forall x \in \bar{D}$.

have $\nabla\varphi^\nu(x^*)$ converging to $\nabla\varphi(x^*)$ a.s.-uniformly on the unit sphere. Combining this with (21), we have for large enough ν

$$\langle \nabla\varphi^\nu(x^*), h\rangle \geq \frac{c}{2}\|h\|, \qquad \forall\, h \in \mathcal{T}_{\bar{D}}(x^*),$$

where $\mathcal{T}$ is the tangent cone to $\bar{D}$ at x^*. This implies that x^* is the unique minimizer of φ^ν on $\bar{D}$ and hence $x^\nu = x^*$. $\qquad\square$

We close this section with the following result concerning a Lipschitz continuity property of ψ when $F(\xi, \cdot)$ is affine. The conclusion might be useful for future sensitivity analysis.

4.4 Proposition *In addition to Assumptions 3.2 & 3.4, if for $\xi \in \Xi$, $F(\xi, u) = M_\xi u + q_\xi$ and $\inf_{\xi \in \Xi} \lambda_{\min}(B + M_\xi + M_\xi^\top) > \alpha$, then for any arbitrary compact set $\bar{D} \subseteq D$, there is a measurable function $\kappa : \Xi \to \mathbb{R}_+$ with $\kappa \in \mathcal{L}_1^1$ such that for any $x, z \in \bar{D} \subseteq D$, we have*

$$|\psi(\xi, x) - \psi(\xi, z)| \leq \kappa(\xi)\|x - z\|.$$

Proof. Let $\bar{D} \subseteq D$ be an arbitrary compact set. For any $x, z \in D$, and $\xi \in \Xi$, let

$$\bar{u} = \arg\min_{u \in C_\xi} f(\xi, x, u) \quad \text{and} \quad \bar{w} = \arg\min_{w \in C_\xi} f(\xi, z, w). \tag{22}$$

In particular, we have from the definition of ψ that

$$\psi(\xi, x) = r(\xi, \bar{u}) + \frac{1}{2}\langle \bar{u} - x, B(\bar{u} - x)\rangle \quad \text{and} \quad \psi(\xi, z) = r(\xi, \bar{w}) + \frac{1}{2}\langle \bar{w} - z, B(\bar{w} - z)\rangle. \tag{23}$$

From Proposition 3.9, $f(\xi, x, \cdot)$ is strongly convex. Using this, Proposition 3.8, (22) and the first-order optimality conditions, we have

$$\langle \nabla r(\xi, \bar{u}) + B(\bar{u} - x), \bar{w} - \bar{u}\rangle \geq 0 \;\; \text{and} \;\; \langle \nabla r(\xi, \bar{w}) + B(\bar{w} - z), \bar{u} - \bar{w}\rangle \geq 0.$$

Adding these two inequalities, we obtain that

$$-\langle \nabla r(\xi, \bar{u}) - \nabla r(\xi, \bar{w}) + B(\bar{u} - \bar{w}), \bar{u} - \bar{w}\rangle + \langle x - z, B(\bar{u} - \bar{w})\rangle \geq 0.$$

Combining this with the strong convexity of f with respect to u established in Proposition 3.9, we have further that

$$\sigma\|\bar{u} - \bar{w}\|^2 \leq \langle \nabla r(\xi, \bar{u}) - \nabla r(\xi, \bar{w}) + B(\bar{u} - \bar{w}), \bar{u} - \bar{w}\rangle \leq \|B\|\|x - z\|\|\bar{u} - \bar{w}\|,$$

where $\sigma > 0$ is independent of x, z and ξ. Hence, we obtain that the solution is Lipschitz continuous, that is,

$$\|\bar{u} - \bar{w}\| \leq \frac{\|B\|}{\sigma}\|x - z\|, \tag{24}$$

whenever $x, z \in D$. Next, by the definition of r and Theorem 3.3, we have

$$r(\xi, u) = \left\langle u - \text{prj}_{C_\xi}\left(u - \frac{1}{\alpha}F(\xi, u)\right), F(\xi, u)\right\rangle - \frac{\alpha}{2}\left\|u - \text{prj}_{C_\xi}\left(u - \frac{1}{\alpha}F(\xi, u)\right)\right\|^2.$$

Since the projection and F are Lipschitz continuous, $\|F(\xi, u)\| \leq d_\xi$ and $\|\nabla F(\xi, u)\| \leq \rho_\xi$ for some $d \in \mathcal{L}_1^\infty$ and $\rho \in \mathcal{L}_1^1$, and $C_\xi \subseteq C^\dagger$ for almost all $\xi \in \Xi$, there is $\kappa_1 \in \mathcal{L}_1^1$ such that whenever $u, w \in C_\xi$,

$$|r(\xi, u) - r(\xi, w)| \leq \kappa_1(\xi)\|u - w\|. \tag{25}$$

Finally, since $\bar{D}$ and $C^\dagger$ are bounded, there is $L > 0$ such that for any $y \in \bar{D} \cup C^\dagger$, we have $\|y\| \leq L/2$.

Combining this with (23), (24) and (25), we obtain the following Lipschitz continuity property of ψ for any $x, z \in \bar{D}$:

$$|\psi(\xi, x) - \psi(\xi, z)| \leq \left[\kappa_1(\xi)\frac{\|B\|}{\sigma} + L\|B\|\left(1 + \frac{\|B\|}{\sigma}\right)\right]\|x - z\|.$$

This completes the proof. $\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\square$

Proposition 4.4 provides sufficient conditions for the Lipschitz continuity of ψ with modulus $\kappa(\xi)$, which can be used for deriving convergence rate. Indeed, if we have in addition that

$$\mathbb{E}[e^{t\kappa(\xi)}] < \infty \quad \text{for all } t \text{ close to } 0,$$

and for any $x \in \bar{D}$,

$$\mathbb{E}[e^{t(\psi(\xi, x) - E[\psi(\xi, x)])}] < \infty \quad \text{for all } t \text{ close to } 0,$$

then for any $\varepsilon > 0$, there exist positive constants $c(\varepsilon)$ and $\beta(\varepsilon)$ such that

$$\text{Prob}\left(\sup_{x \in \bar{D}}\left|\frac{1}{\nu}\sum_{i=1}^{\nu}\psi(\xi^i, x) - \mathbb{E}[\psi(\xi, x)]\right| \geq \varepsilon\right) \leq c(\varepsilon)e^{-\nu\beta(\varepsilon)}$$

for all ν. See [61, Theorem 7.65] for reference.

5   Implementation and experimentation

We begin with a more detailed reformulation of Example 2.6, more compatible with our computational approach, in particular it provides a more explicit version of the flow-conservation equations; in the traffic transportation community all possible node pairs are implicitly included in the *od*-collection even those with $d = o$ and $h_{oo} = 0$ when the o-node is simply a transhipment node. We follow the Ferris-Pang multicommodity formulation [23] which associates a (different) commodity with each destination node $d \in \mathcal{D} \subset \mathcal{N}$. In their formulation, a commodity is associated with each destination node in $\mathcal{D} \subseteq \mathcal{N}$ and $x^j \in \mathbb{R}^{|\mathcal{A}|}$ representing the flows of the commodities $j = 1, 2, \ldots, |\mathcal{D}|$ with x_a^j denoting the flow of commodity j on arc $a \in \mathcal{A}$.

Let $V = (v_{ij})$ denote the node-arc incidence matrix with entries $v_{ij} = 1$ if node i has outgoing flows on arc j, $v_{ij} = -1$ if node i has incoming flows on arc j, and $v_{ij} = 0$ otherwise. The following condition represents conservation of flows of commodities,

$$V x^j = d^j, \quad x^j \geq 0, \; j \in \mathcal{D}, \tag{26}$$

where d_i^j is demand at node $i \in \mathcal{N}$ for commodity j.

Let

$$A = \begin{pmatrix} V & & \\ & \ddots & \\ & & V \end{pmatrix} \in I\!\!R^{|\mathcal{D}||\mathcal{N}| \times |\mathcal{D}||\mathcal{A}|}, \quad x = \begin{pmatrix} x^1 \\ \vdots \\ x^{|\mathcal{D}|} \end{pmatrix} \in I\!\!R^{|\mathcal{D}||\mathcal{A}|}, \quad b = \begin{pmatrix} d^1 \\ \vdots \\ d^{|\mathcal{D}|} \end{pmatrix} \in I\!\!R^{|\mathcal{D}||\mathcal{N}|}.$$

Then (26) can be written as

$$Ax = b, \quad x \geq 0.$$

If we add the constraints on capacity c_a of travel flows on each arc a, then the constraints on the arc flows are presented as follows

$$Ax = b, \quad 0 \leq x, \quad Px \leq c,$$

where $P = (I, \ldots, I) \in I\!\!R^{|\mathcal{A}| \times |\mathcal{D}||\mathcal{A}|}$ and I is the $|\mathcal{A}| \times |\mathcal{A}|$ identity matrix.

Traffic equilibrium models are built based on travel demand, travel capacity on each arc and travel flows via nodes. The demand and capacity depend heavily on various uncertain parameters, such as weather, accidents, etc. Let $\Xi \subseteq I\!\!R^N$ denote the set of uncertain factors. Let $(d_\xi^j)_i > 0$ denote the stochastic travel demand at the ith node for commodity j and $(c_\xi)_a$ denote the stochastic capacity of arc a.

For a realization of random vectors $d_\xi^j \in I\!\!R^{|\mathcal{N}|}$ and $c_\xi \in I\!\!R^{|\mathcal{A}|}$, $\xi \in \Xi$, an assignment of flows to all arcs for commodity j is denoted by the vector $u_\xi^j \in I\!\!R^{|\mathcal{A}|}$, whose component $(u_\xi^j)_a$ denotes the flow on arc a for commodity j.
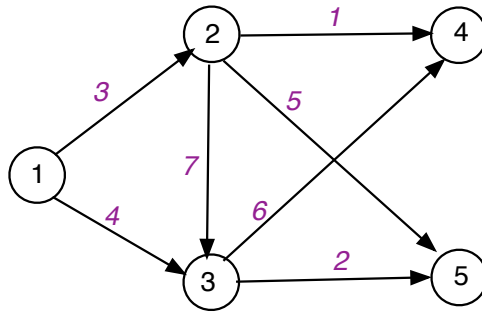

Figure 1: The 7-links, 6-paths network

The network in Figure 1 from [67] has 5 nodes, 7 arcs and 2 destination nodes $\{4, 5\}$. The

node-arc matrix for the network in Figure 1 is given as follows.

$$V = \begin{pmatrix} 0 & 0 & 1 & 1 & 0 & 0 & 0 \\ 1 & 0 & -1 & 0 & 1 & 0 & 1 \\ 0 & 1 & 0 & -1 & 0 & 1 & -1 \\ -1 & 0 & 0 & 0 & 0 & -1 & 0 \\ 0 & -1 & 0 & 0 & -1 & 0 & 0 \end{pmatrix}$$

and the matrices and vectors are as follows

$$A = \begin{pmatrix} V & 0 \\ 0 & V \end{pmatrix} \in I\!\!R^{10\times14}, \quad u_\xi \in I\!\!R^{14}, \quad b_\xi \in I\!\!R^{10}, \quad c_\xi \in I\!\!R^7.$$

For a forecast robust arc flows x, let the feasible set for each realization ξ be

$$C_\xi = \{u_\xi \mid Au_\xi = b_\xi,\ 0 \le u_\xi,\ Pu_\xi \le c_\xi\},$$

where $Pu_\xi = \sum_{j=1}^{|\mathcal{D}|} u_\xi^j$ is the total travel flows.

The arc travel time function $h(\xi,\cdot) : I\!\!R^{|\mathcal{D}||\mathcal{A}|} \to I\!\!R^{|\mathcal{A}|}$ is a stochastic vector and each of its entries $h_a(\xi, u_\xi)$ is assumed to follow a generalized Bureau of Public Roads (GBPR) function,

$$h_a(\xi, u_\xi) = \left(\eta_a + \tau_a\left(\frac{(Pu_\xi)_a}{(\gamma_\xi)_a}\right)^{n_a}\right), \quad a = 1, \ldots, |\mathcal{A}|, \tag{27}$$

where $\eta_a, \tau_a, (\gamma_\xi)_a$ and n_a are given positive parameters, and $(Pu_\xi)_a$ is the total travel flows on each arc $a \in \mathcal{A}$. Let

$$F(\xi, u_\xi) = P^T h(\xi, u_\xi).$$

Then

$$\nabla_u F(\xi, u_\xi) = P^T \text{diag}\left(\tau_a n_a \frac{(Pu_\xi)_a^{n_a-1}}{(\gamma_\xi)_a^{n_a}}\right) P,$$

which is symmetric positive semi-definite for any $u_\xi \in C_\xi \subseteq I\!\!R_+^{|\mathcal{D}||\mathcal{A}|}$. One commonly considered special case is when $n_a = 1$, for all $a \in \mathcal{A}$. In this case, $F(\xi, u_\xi) = M_\xi u_\xi + q$, where

$$M_\xi = P^T \text{diag}\left(\frac{\tau_a}{(\gamma_\xi)_a}\right) P \quad \text{and} \quad q = P^T(\eta_1, \ldots, \eta_{|\mathcal{A}|})^T. \tag{28}$$

Here, we have $\text{rank}(P) = |\mathcal{A}|$, and that for any $\xi \in \Xi$, $M_\xi \in I\!\!R^{|\mathcal{D}||\mathcal{A}|\times|\mathcal{D}||\mathcal{A}|}$ is a positive semi-definite matrix. Thus, $I\!\!E[M_\xi]$ is positive semi-definite. Another commonly considered case is when $n_a = 3$, for all $a \in \mathcal{A}$; see [23] and our numerical experiments in Section 5.2.

To define a here-and-now solution x, let

$$D = \{x \mid Ax = I\!\!E[b_\xi],\ 0 \le x,\ Px \le I\!\!E[c_\xi]\}$$

and $G(x) = P^T \bar{h}(x)$ where the components of $\bar{h}$ is defined by

$$\bar{h}_a(x) = \eta_a + \tau_a(Px)_a^{n_a} I\!\!E[(\gamma_\xi)_a^{-n_a}], \quad a = 1, \ldots, |\mathcal{A}|.$$

The deterministic VI formulation for Wardrop's user equilibrium, seeks a forecast arc flows $x \in D$ satisfying

$$-G(x) \in N_D(x). \tag{29}$$

On the other hand, the stochastic VI formulation for Wardrop's user equilibrium, seeks an equilibrium arc flow $u_\xi \in C_\xi$ for a known event $\xi \in \Xi$, such that

$$-F(\xi, u_\xi) \in_{a.s.} N_{C_\xi}(u_\xi). \tag{30}$$

In general, the solution sets of the variational inequalities (29) and (30) can have multiple solutions. Naturally, a here-and-now solution x should have minimum total distances to the solution set of (30) for almost all observations $\xi \in \Xi$. It can be written as a mathematical programming with equilibrium constraints [43] as the following

$$\begin{array}{ll} \min & \mathbb{E}[\|u_\xi - x\|^2] \\ \text{subject to} & -G(x) \in N_D(x), \quad -F(\xi, u_\xi) \in_{a.s.} N_{C_\xi}(u_\xi), \quad \xi \in \Xi. \end{array} \tag{31}$$

Recalling the definitions of the residual functions induced by the regularized gap function given in §3, we have

$$-G(x) \in N_D(x) \quad \Leftrightarrow \quad \theta(x) = 0 \text{ and } x \in D,$$
$$-F(\xi, u_\xi) \in N_{C_\xi}(u_\xi) \quad \Leftrightarrow \quad r(\xi, u_\xi) = 0 \text{ and } u_\xi \in C_\xi.$$

Note that $\theta(x) \geq 0, r(\xi, u_\xi) \geq 0$ for $x \in D$ and $u_\xi \in C_\xi$, and they are continuously differentiable. It is natural to consider the following l_1 penalty problem which trades-off optimization of total distance and the violation of the constraints in (31):

$$\begin{array}{ll} \min & \frac{1}{\lambda \rho}\theta(x) + \frac{1}{\rho}\mathbb{E}[r(\xi, u_\xi)] + \mathbb{E}[\|u_\xi - x\|^2] \\ \text{subject to} & x \in D, \quad u_\xi \in_{a.s.} C_\xi, \ \xi \in \Xi, \end{array} \tag{32}$$

for some positive numbers λ and ρ. Notice that this is a special case of (11), which was derived in §3 as a relaxation of (4). Our discussion above provided an alternative interpretation of this problem as an l_1 penalty problem.

Using Lemma 3.6, we see that (32) is further equivalent to the following problem

$$\begin{array}{ll} \min_{x \in D} & \theta(x) + \lambda \mathbb{E}[\psi(\xi, x)] \\ \text{where} & \psi(\xi, x) = \min_{u_\xi \in C_\xi} r(\xi, u_\xi) + \rho\|u_\xi - x\|^2. \end{array} \tag{33}$$

In this problem, for each $\xi \in \Xi$, the decisions u_ξ are dependent on a forecast arc flows x. The two-stage stochastic program (33) uses x as the first stage decision variable, and u_ξ as the second stage variable.

In a stochastic environment, ξ belongs to a set Ξ representing future states of knowledge. The stochastic optimization approach (32) (equivalently, (33)) is to find a vector x^* which minimizes the expected residual values with recourse cost. The main role of traffic model is to provide a forecast for future traffic states. The solution of the stochastic optimization approach (32) is a "here-and-now" solution which provides a robust forecast and has advantages over other models for long term planning. Stochastic traffic assignment on path flow has been formulated as stochastic complementarity problems and stochastic variational inequalities in [1, 10, 69]

5.1 The Douglas-Rachford splitting method

In this section, we focus on problem (32), which is a special case of (11) with $B = 2\rho I$ and $D = \{x \in \mathbb{R}_+^{|\mathcal{D}||\mathcal{A}|} \,|\, Ax = \mathbb{E}[b_\xi], \, Px \leq \mathbb{E}[c_\xi]\}$. We will discuss an algorithm for solving the following SAA of problem (32), for any fixed ν:

$$
\begin{aligned}
&\min && \theta(x) + \frac{\lambda}{\nu} \sum_{i=1}^{\nu} \left[r(\xi^i, u_{\xi^i}) + \rho \| u_{\xi^i} - x \|^2 \right] \\
&\text{subject to} && x \in D, \; u_{\xi^i} \in C_{\xi^i} \quad \forall i = 1, \dots, \nu,
\end{aligned}
\tag{34}
$$

where $\{\xi^1, \dots, \xi^\nu\}$ is an iid sample of Ξ of size ν.

Notice that the objective of (34) is highly structured: the decision variables are only related due to the quadratic term $\frac{\lambda\rho}{\nu} \sum_{i=1}^{\nu} \| u_{\xi^i} - x \|^2$; if this quadratic term was absent, then the problem would be decomposed into $\nu + 1$ smaller independent optimization problems that can be solved in parallel. This observation leads us to consider splitting methods in which the objective is decoupled into two parts and minimized separately. One such method is the Douglas-Rachford (DR) splitting method. This method was first proposed in [19] for solving feasibility problems and has been extensively studied in the convex scenario; see, for example, [3, 20, 29]. Moreover, the global convergence of the method for some class of problems with nonconvex objectives has been recently studied and established in [39].

To apply the DR splitting method, we first note that (34) can be equivalently written as a minimization problem that minimizes the sum of the following two functions:

$$
f(U, x) := \frac{\lambda\rho}{\nu} \sum_{i=1}^{\nu} \| u_{\xi^i} - x \|^2, \quad g(U, x) := \theta(x) + \delta_D(x) + \frac{\lambda}{\nu} \sum_{i=1}^{\nu} (r(\xi^i, u_{\xi^i}) + \delta_{C_{\xi^i}}(u_{\xi^i})), \tag{35}
$$

where g is a proper closed function and f is a quadratic function whose Lipschitz continuity modulus is $2\lambda\rho \left(1 + \frac{1}{\nu}\right)$, and U is the vector in $\mathbb{R}^{\nu|\mathcal{D}||\mathcal{A}|}$ formed by stacking u_{ξ^i}, i.e.,

$$
U = \begin{bmatrix} u_{\xi^1}^T & \cdots & u_{\xi^\nu}^T \end{bmatrix}^T.
$$

Each iteration of the DR splitting method then involves solving regularized optimization problems related to f, g and the current iterates, as well as updating the "dual" variables. The DR splitting method applied to solving (34) is presented in (36), where we denote by V (resp., Z) the vector in $\mathbb{R}^{\nu|\mathcal{D}||\mathcal{A}|}$ formed by stacking v_i (resp., z_i) for notational convenience, i.e.,

$$
V = \begin{bmatrix} v_1^T & \cdots & v_\nu^T \end{bmatrix}^T, \quad Z = \begin{bmatrix} z_1^T & \cdots & z_\nu^T \end{bmatrix}^T.
$$

The global convergence of the sequence generated by this method to stationary points of (34) follows immediately from [39, Theorem 1], [39, Theorem 4] and our choice of step-size parameter μ. However, notice that the subproblems for the (U, x)-update are nonconvex in general. Consequently, in practice, it is hard to compute *global* minimizers for these subproblems and typical optimization solvers are only guaranteed to return a stationary point satisfying the first-order optimality conditions. Fortunately, a closer look at the proofs of [39, Theorem 1] and [39, Theorem 4] reveals that, for the convergence proofs to remain valid, one only needs to update (U, x) so that

---

Douglas-Rachford splitting method for (34)

Step 0. Input $\{\zeta^0, Z^0\}$, $0 < \mu < \frac{\nu}{\nu+1}\frac{\sqrt{1.5}-1}{2\lambda\rho}$.

Step 1. Set

$$
\begin{cases}
(V^{k+1}, y^{k+1}) = \underset{V,y}{\operatorname{argmin}} \left\{ \frac{1}{2\mu}\|V - Z^k\|^2 + \frac{1}{2\mu}\|y - \zeta^k\|^2 + \frac{\lambda\rho}{\nu}\sum_{i=1}^{\nu}\|v_i - y\|^2 \right\} \\
x^{k+1} \in \underset{x\in D}{\operatorname{argmin}} \left\{ \theta(x) + \frac{1}{2\mu}\|2y^{k+1} - \zeta^k - x\|^2 \right\} \\
\text{For } i = 1, \ldots, \nu, \\
\quad u_{\xi^i}^{k+1} \in \underset{u_{\xi^i}\in C_{\xi^i}}{\operatorname{argmin}} \left\{ \frac{\lambda}{\nu}r(\xi^i, u_{\xi^i}) + \frac{1}{2\mu}\|2v_i^{k+1} - z_i^k - u_{\xi^i}\|^2 \right\}, \\
\zeta^{k+1} = \zeta^k + (x^{k+1} - y^{k+1}), \\
Z^{k+1} = Z^k + (U^{k+1} - V^{k+1}).
\end{cases}
$$

$$(36)$$

Step 2. If a termination criterion is not met, go to Step 1.

---

1. (U^{k+1}, x^{k+1}) is a first-order stationary point of the subproblems in the kth iteration; and

2. the subproblem objective values in the kth iteration at (U^{k+1}, x^{k+1}) are smaller than those at (U^k, x^k).

Thus, in practice, one can apply any optimization solver that employs a descent algorithm for solving the nonconvex subproblems corresponding to the (U, x)-update approximately. On the other hand, the (V, y)-updates for this algorithm can be solved explicitly and efficiently by exploiting the arrow-shaped structure of the Hessian of the quadratic optimization problem.

Finally, we discuss a possible termination criterion for the algorithm. To this end, observe from the definitions of f and g in (35) and the optimality conditions for the (V, y) and (U, x)-subproblems in (36) that

$$
0 = \nabla f(V^{k+1}, y^{k+1}) + \frac{1}{\mu}[(V^{k+1}, y^{k+1}) - (Z^k, \zeta^k)],
$$

$$
0 \in \partial g(U^{k+1}, x^{k+1}) + \frac{1}{\mu}[(U^{k+1}, x^{k+1}) - 2(V^{k+1}, y^{k+1}) + (Z^k, \zeta^k)].
$$

From these, it is not hard to deduce that

$$
-\frac{1}{\mu}[(U^k, x^k) - (V^k, y^k)] + (\nabla f(U^k, x^k) - \nabla f(V^k, y^k))
$$

$$(37)$$

is an element of $\nabla f(U^k, x^k) + \partial g(U^k, x^k)$. The algorithm can thus be terminated with an approximate stationary point (U^k, x^k) when the quantity in (37) is small in norm. For instance,

one can terminate when

$$\left(L + \frac{1}{\mu}\right) \frac{\|(U^k, x^k) - (V^k, y^k)\|}{\max\{\|(U^k, x^k)\|, \|(V^k, y^k)\|, 1\}} < tol \tag{38}$$

for some $tol > 0$; here, L stands for the Lipschitz continuity modulus of ∇f.

5.2 Numerical simulations

We now report simulation results based on the networks described at the beginning of this section. All codes were written in MATLAB and run in MATLAB R2011b.

We consider the case when $n_a = 3$ and compare the solution of our model against the solution of the EV formulation (2). The latter problem is a nonlinear programming problem with the smooth objective (1). In our experiments below, we choose $\alpha = 10$ in (1) for the EV formulation and solve the problem approximately by the MATLAB function fmincon[12] using the default tolerance, initialized at the projection of the origin onto the feasible set. The projections onto D (as well as C_ξ considered below) are computed approximately by the MATLAB function lsqlin,[13] using the default tolerance.

For our model, we choose $\alpha = 10$ in (3), $\lambda = 20$ and $\rho = 5$. To initialize the algorithm, we set $\{z_i^0\}$ to be the approximate solutions to $\min_{u_{\xi^i} \in C_{\xi^i}} r(\xi^i, u_{\xi^i})$ for each $i = 1, \ldots, \nu$ and ζ^0 to be the approximate solution to (2); these problems are solved approximately by fmincon using the same settings as described above. On the other hand, we terminate when

$$\left(2\lambda\rho\frac{\nu + 1}{\nu} + \frac{1}{\mu}\right) \frac{\|(U^k, x^k) - (V^k, y^k)\|}{\max\{\|(U^k, x^k)\|, \|(V^k, y^k)\|, 1\}} < tol, \tag{39}$$

for some $tol > 0$. This termination criterion is motivated by (38). To speed up the convergence, we use the same kind of heuristic for choosing μ as described in [39, Remark 4] and [63, Remark 2.1]: we set $\mu = 150\mu_0$, where $\mu_0 := \frac{0.99\nu}{\nu+1}\frac{\sqrt{1.5}-1}{2\lambda\rho}$, and update $\mu = \max\{\mu_0, \mu/2\}$ when either $\|(V^k, y^k)\| > 10^{10}$ or

$$\|(V^k, y^k) - (V^{k-1}, y^{k-1})\| > \frac{10^6}{k}.$$

In our experiment below, we use the same free travel time η in (27) as in [67], i.e.,

$$\eta = \begin{bmatrix} 6 & 4 & 3 & 5 & 6 & 4 & 1 \end{bmatrix}^T,$$

and $\tau = 0.15\eta$. The demand vector and the link capacity both have a beta distribution, as in [10, Example 4.1]. The lower bound for link capacity γ_ξ is $\begin{bmatrix} 10 & 10 & 20 & 20 & 10 & 10 & 10 \end{bmatrix}^T$ with a mean of $\overline{\gamma} := \begin{bmatrix} 15 & 15 & 30 & 30 & 15 & 15 & 15 \end{bmatrix}^T$, and the parameters for the beta distribution are $(\alpha, \beta) = (2, 2)$. We then set c_ξ in C_ξ to be $10\overline{\gamma}$, independent of $\xi \in \{\xi^1, \ldots, \xi^\nu\}$. For the demand vector b_ξ, the lower bound $\underline{b}$ is $\begin{bmatrix} 150 & 180 \end{bmatrix}^T$ and we set $b_\xi = \underline{b} + \begin{bmatrix} 120 & 96 \end{bmatrix}^T \circ beta(\alpha, \beta)$, with

---

[12]In the options, we set optimset('Display','off','Algorithm','interior-point','GradObj','on').
[13]In the options, we set optimset('Display','off','LargeScale','off').

the first entry corresponding to the OD pair $1 \to 4$. The corresponding parameters for the beta distribution are $(\alpha, \beta) = (5, 1)$ in this case. Furthermore, we set $tol = 5 \times 10^{-5}$ in (39) in our experiments for this network.

The computational results are given in Table 1, where we present results for sample sizes $\nu = 50$, 100 and 150. We denote by x^* the approximate solution to our model obtained from the DR splitting method and x_{EV} the solution to (2). For these solutions, we compute their projections w_ξ onto C_ξ for a larger sample Υ_ν of size $20 \times \nu$ that contains $\{\xi^1, \dots, \xi^\nu\}$, and report $\mathbf{vio} := \frac{1}{20\nu} \sum_{\xi \in \Upsilon_\nu} \max_{v \in C_\xi} (Pw_\xi - Pv)^T h(\xi, w_\xi)$,[14] the CVaR given by

$$\mathrm{CVaR} = \inf_\alpha \alpha + \frac{1}{2\nu} \sum_{\xi \in \Upsilon_\nu} \left[\max_{v \in C_\xi} (Pw_\xi - Pv)^T h(\xi, w_\xi) - \alpha \right]_+ ,$$

and the corresponding minimizer α^* at which the above infimum is attained. Clearly, the smaller these three quantities are, the closer w_ξ is to being optimal for the corresponding instance. Our computational results show that the w_ξ constructed from our x^* are better optimal solutions on average for $\xi \in \Upsilon_\nu$.

Size	x^*			x_{EV}		
ν	$\mathbf{vio}$	α^*	CVaR	$\mathbf{vio}$	α^*	CVaR
50	1.834e4	3.259e4	4.046e4	1.821e4	3.207e4	3.994e4
100	1.856e4	3.409e4	4.138e4	1.913e4	3.520e4	4.326e4
150	1.791e4	3.162e4	3.934e4	1.833e4	3.290e4	4.079e4

Table 1: Computational result for the network from [67].

We next consider the Nguyen and Dupuis network, which contains 13 nodes, 19 arcs, 2 destinations. See Figure 2.
We use the same free travel time η in (27) as in [68, Table 1], i.e.,

$$\eta = [\; 7 \quad 9 \quad 9 \quad 12 \quad 3 \quad 9 \quad 5 \quad 13 \quad 5 \quad 9 \quad 9 \quad 10 \quad 9 \quad 6 \quad 9 \quad 8 \quad 7 \quad 14 \quad 11 \;]^T,$$

and $\tau = 0.15\eta$. Also, as in [10, Example 4.2], our demand vector has a beta distribution. The lower bound $\underline{b}$ for the demand vector is $\begin{bmatrix} 300 & 700 & 500 & 350 \end{bmatrix}^T$, which corresponds to the OD pairs $1 \to 2$, $1 \to 3$, $4 \to 2$ and $4 \to 3$. The parameters for the beta distribution are $(\alpha, \beta) = (50, 10)$. The demand vector b_ξ is then generated according to $\underline{b} + 120\,beta(\alpha, \beta)$. On the other hand, we generate the link capacity γ_ξ exactly as in [10, Example 4.2]. We also set c_ξ in C_ξ to be $10\overline{\gamma}$, where $\overline{\gamma}$ is the mean of the three possible link capacities in [10, Example 4.2] with the probability specified there, independent of ξ. Furthermore, we set $tol = 5 \times 10^{-4}$ in (39) in our experiments for this network.

---

[14]The maximum value in the summand is computed using MATLAB function linprog, with the option set to be optimset('Display','off','Algorithm','active-set').
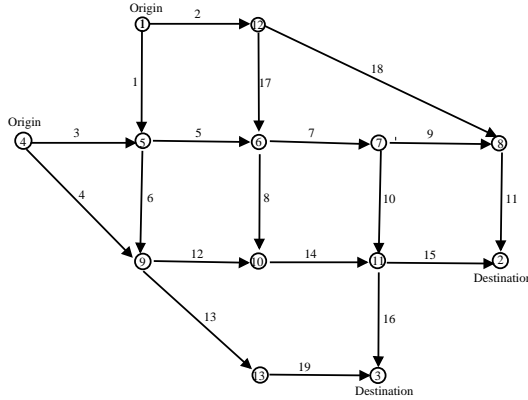
Figure 2: Nguyen and Dupuis Network with 19 links and 25 paths

The computational results are given in Table 2, where we present results for various sample sizes $\nu = 50$, 100 and 150 as before. We also report the same quantities as in Table 1: **vio**, the CVaR and the corresponding α^*. Our computational results again show that the w_ξ constructed from our x^* are better optimal solutions on average for $\xi \in \Upsilon_\nu$.

Size	x^*			x_{EV}		
ν	**vio**	α^*	CVaR	**vio**	α^*	CVaR
50	7.126e3	1.689e4	1.759e4	9.176e3	2.594e4	2.643e4
100	7.628e3	1.829e4	1.894e4	1.076e4	3.086e4	3.110e4
150	7.139e3	1.682e4	1.755e4	9.059e3	2.520e4	2.557e4

Table 2: Computational result for the Nguyen and Dupuis network.

References

[1] R. Agdeppa, N. Yamashita, and M. Fukushima. Convex expected residual models for stochastic affine variational inequality problems and its application to the traffic equilibrium problem. *Pacific Journal on Optimization*, 6:3–19, 2010.

[2] H. Attouch and R. Wets. Epigraphical processes: laws of large numbers for random lsc functions. *Séminaire d'Analyse Convexe*, 13:1–29, 1990.

[3] H. Bauschke and P. Combettes. *Convex Analysis and Monotone Operator Theory in Hilbert Spaces.* Springer, 2011.

[4] M. Beckmann, C. McGuire, and C. Winsten. *Studies in the Economics of Transportation.* Yale University Press, 1956.

[5] Y. Bello Cruz and A. Iusem. Convergence of direct methods for paramonotone variational inequalities. *Computational Optimization and Applications*, 46:247–263, 2010.

[6] J. Birge and F. Louveaux. *Introduction to Stochastic Programming*. Springer, 1997.

[7] D. Brown, X. DeMarzo, and B.C. Eaves. Computing equilibria when asset markets are incomplete. *Econometrica*, 64:1–27, 1996.

[8] X. Chen and M. Fukushima. Expected residual minimization method for stochastic linear complementarity problems. *Mathematics of Operations Research*, 30:1022–1038, 2005.

[9] X. Chen, H. Sun, and R. Wets. Regularized mathematical programs with stochastic equilibrium constraints: estimating structure demand models. *SIAM Journal on Optimization*, 25:53–75, 2015.

[10] X. Chen, R. Wets, and Y. Zhang. Stochastic variational inequalities: Residual minimization smoothing. *SIAM Journal on Optimization*, 22:649–673, 2012.

[11] X. Chen and Y. Ye. On homotopy-smoothing methods for box constrained variational inequalities. *SIAM Journal on Control and Optimization*, 37:589–616, 1999.

[12] X. Chen, C. Zhang, and M. Fukushima. Robust solution of monotone stochastic linear complementarity problems. *Mathematical Programming*, 117:51–80, 2009.

[13] R. Cominetti. Equilibrium routing under uncertainty. *Mathematical Programming*, 151:117–151, 2015.

[14] J.R. Correa and N.E. Stier-Moses. Wardrop equilibria. Wiley Online Library, 2011.

[15] C. Dang, Y. Ye, and Z. Zhu. An interior-point path-following algorithm for computing a leontief economy equilibrium. *Computational Optimization and Applications*, 50:213–236, 2011.

[16] P. Demarzo and B.C. Eaves. Computing equilibria of GEI by relocalization on a grassmann manifold. *Journal of Mathematical Economics*, 26:479–497, 1996.

[17] J. Deride, A. Jofré, and R. Wets. Solving deterministic and stochastic equilibrium problems via augmented Walrasian. *Computational Economics*, 2015 (submitted).

[18] S. Dirkse, M. Ferris, and T. Munson. http://pages.cs.wisc.edu/ ferris/path.html, 2015 – the PATH solver. Technical report, University of Wisconsin, 2015.

[19] J. Douglas and H. Rachford. On the numerical solution of heat conduction problems in two or three space variables. *Transactions of the American Mathematical Society*, 82:421–439, 1956.

[20] J. Eckstein and D. Bertsekas. On the Douglas-Rachford splitting method and the proximal point algorithm for maximal monotone operators. *Mathematical Programming*, 55:293–318, 1992.

[21] F. Fachinei and J.-S. Pang. *Finite-Dimensional Variational Inequalities and Complementarity Problems. I and II.* Springer, 2003.

[22] H. Fang, X. Chen, and M. Fukushima. Stochastic R_0 matrix linear complementarity problems. *SIAM Journal on Optimization*, 18:482–506, 2007.

[23] M. Ferris and J.-S. Pang. Engineering and economic applications of complementarity problems. *SIAM Review*, 39:669–713, 1997.

[24] M. Ferris and R. Wets. MOPEC: A computationally amiable formulation of multi-agent optimization problems with global equilibrium constraints. Department of Mathematics, University of California-Davis, August 2012.

[25] M. Fukushima. A relaxed projection method for variational inequalities. *Mathematical Programming*, 35:58–70, 1986.

[26] M. Fukushima. Equivalent differentiable optimization problems and descent methods for asymmetric variational inequality problems. *Mathematical Programming*, 53:99–110, 1992.

[27] G. Gürkan, A. Özge, and S. Robinson. Sample-path solution of stochastic variational inequalities with applications to option pricing. In J. Charnes, D. Morrice, D. Brunner, and J. Swain, editors, *Proceedings of the 1966 Winter Simulation Conference*, pages 337–344. INFORMS, 1966.

[28] G. Gürkan, A. Özge, and S. Robinson. Sample-path solution of stochastic variational inequalities. *Mathematical Programming*, 84:313–333, 1999.

[29] B. He and X. Yuan. On the $O(1/n)$ convergence rate of the Douglas-Rachford alternating direction method. *SIAM Journal on Numerical Analysis*, 50:700–709, 2012.

[30] A. Iusem, A. Jofré, and P. Thompson. Approximate projection methods for monotone stochastic variational inequalities. Technical report, IMPA, Rio de Janeiro, 2015.

[31] A. Iusem and B. Svaiter. A variant of Kopelevich's method for variational inequalities with a new search strategy. *Optimization*, 42:309–321, 1997.

[32] H. Jiang and H. Xu. Stochastic approximation approaches to stochastic variational inequalities. *IEEE Transaction on Automatic Control*, 53:1462–1475, 2008.

[33] A. Jofré, R.T. Rockafellar, and R. Wets. Variational inequalities and economic equilibrium. *Mathematics of Operations Research*, 32:32–50, 2007.

[34] A. Jofré, R.T. Rockafellar, and R. Wets. General economic equilibirum with financial market and retainability. *Economic Theory (submitted)*, 2014.

[35] A. Juditsky, A. Nemirovski, and C. Tauvel. Solving variational inequalities with stochastic mirror-prox algorithm. *Stoch. Syst.*, 1:17–58, 2011.

[36] P. Kall and S. Wallace. *Stochastic Programming*. J. Wiley & Sons, 1995.

[37] L. Korf and R. Wets. Random lsc functions: an ergodic theorem. *Mathematics of Operations Research*, 26:421–445, 2001.

[38] M. Lamm and S. Lu. Confidence interval computation for stochastic variational inequalities and stochastic cournot–nash equilibria. *Preprint*, 2015.

[39] G. Li and T. K. Pong. Douglas-Rachford splitting for nonconvex feasibility problems. to appear in *Mathematical Programming*, DOI: 10.1007/s10107-015-0963-5, 2015.

[40] S. Lu. Confidence regions for stochastic variational inequalities. *Optimization*, 63:1431–1443, 2012.

[41] S. Lu. Symmetric confidence regions and confidence intervals for normal map formulations of stochastic variational inequalities. *SIAM Journal on Optimization*, 24:1458–1484, 2014.

[42] M. Luo and G. Lin. Expected residual minimization method for stochastic variational inequality problems. *Journal of Optimization Theory and Applications*, 140:103–116, 2009.

[43] Z. Luo, J.-S.Pang, and D. Ralph. *Mathematical Programs with Equilibrium Constraints*. Cambridge University Press, 1996.

[44] P. Marcotte and P. Patriksson. *Transportation, Handbooks in Operations Research and Management Science*, volume 14 of *Handbooks in Operations Research and Management Science*, chapter Traffic Equilibrium, pages 623–713. Elsevier, 2007.

[45] V. Norkin and R. Wets. On strong graphical law of large numbers for random semicontinuous mappings,. *Vestnik of Saint-Petersburg University*, 10:102–110, 2013.

[46] V. Norkin, R. Wets, and H. Xu. Graphical convergence of sample average random set-valued mappings. Mathematics, University of California, Davis, 2010.

[47] J.-S. Pang, C.-L. Su, and Y. Lee. A constructive approach to estimating pure characteristics demand models with pricing. *Operation Research*, 63:639–659, 2015.

[48] T. Pennanen. Epi-convergent discretizations of multistage stochastic programs. *Mathematics of Operations Research*, 30:245–256, 2005.

[49] T. Pennanen and M. Koivu. Epi-convergent discretization of stochastic programs via integration quadratures. *Numerische Mathematik*, 100:141–163, 2005.

[50] A. Philpott, M. Ferris, and R. Wets. Equilibrium, uncertainty and risk in hydro-thermal electricity systems. *Preprint*, 2015.

[51] R.T. Rockafellar. *Convex Analysis*. Princeton University Press, 1970.

[52] R.T. Rockafellar and R. Wets. Stochastic convex programming: Kuhn-Tucker conditions. *Journal of Mathematical Economics*, 2:349–370, 1975.

[53] R.T. Rockafellar and R. Wets. Nonanticipativity and $\mathcal{L}^1$-martingales in stochastic optimization problems. *Mathematical Programming Study*, 6:170–187, 1976.

[54] R.T. Rockafellar and R. Wets. Stochastic convex programming: basic duality. *Pacific Journal of Mathematics*, 62:173–195, 1976.

[55] R.T. Rockafellar and R. Wets. Scenarios and policy aggregation in optimization under uncertainty. *Mathematics of Operations Research*, 16:119–147, 1991.

[56] R.T. Rockafellar and R. Wets. A dual strategy for the implementation of the aggregation principle in decision making under uncertainty. *Applied Stochastic Models and Data Analysis*, 8:245–255, 1992.

[57] R.T. Rockafellar and R. Wets. *Variational Analysis*. Springer, 1998.

[58] R.T. Rockafellar and R. Wets. Stochastic variational inequalities: single-stage to multistage. *Mathematical Programming (accepted)*, 2015.

[59] U. Shanbhag, G. Infanger, and P. Glynn. A complementarity framework for forward contracting under uncertainty. *Operations Research*, 59:810–834, 2011.

[60] A. Shapiro. *Lectures on Stochastic Programming*, chapter 5: Statistical Inference, pages 155–252. SIAM, 2009.

[61] A. Shapiro, D. Dentcheva, and A. Ruszczyński. *Lectures on Stochastic Programming – Modeling and Theory*. MPS-SIAM Series on Optimization. SIAM Publisher, 2009.

[62] A. Shapiro and H. Xu. Uniform laws of large numbers for set-valued mappings and subdifferentials of random functions. *Journal of Mathematical Analysis and Applications*, 325:1390–1399, 2007.

[63] D. Sun, K.-C. Toh, and L. Yang. A convergent proximal alternating direction method of multipliers for conic programming with 4-block constraints. Preprint, 2014.

[64] R. Van Slyke and R. Wets. L-shaped linear programs with application to optimal control and stochastic programming. *SIAM Journal on Applied Mathematics*, 17:638–663, 1969.

[65] M. Wang and D. Bertsekas. Incremental constraint projection methods for variational inequalities. *Mathematical Programming*, 150:321–363, 2015.

[66] Y. Xie and U.V. Shanbhag. On robust solutions to uncertain linear complementarity problems and their variants. *http://arxiv.org/abs/1503.03490*, 2015.

[67] H. Yang. Sensitivity analysis for the elastic-demand network equilibrium problem with applications. *Transportation Research B*, 31:55–70, 1997.

[68] Y. Yin, S. Madanat, and X.-Y. Lu. Robust improvement schemes for road networks under demand uncertainty. *European Journal on Operations Research*, 198:470–479, 2009.

[69] C. Zhang, X. Chen, and A. Sumalee. Robust Wardrop's user equilibrium assignment under stochastic demand and supply: expected residual minimization approach. *Transportation Research B*, 45:534–552, 2011.